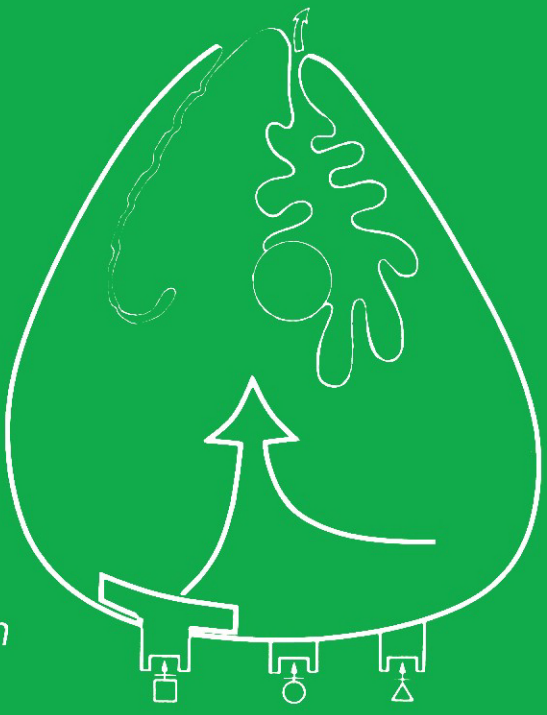
TOPICS IN GASTROENTEROLOGY • Series Editor: Howard M. Spiro

Modern Concepts in Gastroenterology

Volume 2



Edited by
Eldon Shaffer
and
Alan B. R. Thomson

Modern Concepts in Gastroenterology

Volume 2

TOPICS IN GASTROENTEROLOGY

Series Editor: **Howard M. Spiro, M.D.**
Yale University School of Medicine

COLON
Structure and Function
Edited by Luis Bustos-Fernandez, M.D.

KEY FACTS IN GASTROENTEROLOGY
Jonathan Halevy, M.D.

MEDICAL ASPECTS OF DIETARY FIBER
Edited by Gene A. Spiller, Ph.D., and Ruth McPherson Kay, Ph.D.

MODERN CONCEPTS IN GASTROENTEROLOGY, Volume 1
Edited by Alan B. R. Thomson, M.D., Ph.D., L. R. DaCosta, M.D., and William C. Watson, M.D.

MODERN CONCEPTS IN GASTROENTEROLOGY, Volume 2
Edited by Eldon Shaffer, M.D., and Alan B. R. Thomson, M.D., Ph.D.

NUTRITION AND DIET THERAPY IN GASTROINTESTINAL DISEASE
Martin H. Floch, M.S., M.D., F.A.C.P.

PANCREATITIS
Peter A. Banks, M.D.

Modern Concepts in Gastroenterology
Volume 2

Edited by
Eldon Shaffer, M.D.
University of Calgary
Calgary, Alberta, Canada

and
Alan B. R. Thomson, M.D., Ph.D.
The University of Alberta
Edmonton, Alberta, Canada

PLENUM MEDICAL BOOK COMPANY
NEW YORK AND LONDON

The Library of Congress cataloged the first volume of this title as follows:

Modern concepts in gastroenterology.

(Topics in gastroenterology)
Includes bibliographies and index.
1. Gastrointestinal system—Diseases. I. Thomson, A. B. R. (Alan Bryan Robert), 1943- . II. DaCosta, L. R. III. Watson, William C. IV. Series.
RC801.M58 1986 616.3'3 87-102291

ISBN-13: 978-1-4612-8079-8 e-ISBN-13: 978-1-4613-0781-5
DOI: 10.1007/978-1-4613-0781-5

Softcover reprint of the hardcover 1st edition 1989
233 Spring Street, New York, N.Y. 10013

Plenum Medical Book Company is an imprint of Plenum Publishing Corporation

This work is dedicated to
Beryl, Andrea, Emily, and Ali
and to
Jeannette, James, Matthew,
Jessica, and Benjamin

Contributors

L. M. Blendis, Department of Medicine, Toronto General Hospital, University of Toronto, Toronto, Ontario, Canada M5G 2C4

Jared M. Diamond, Department of Physiology, University of California–Los Angeles School of Medicine, Los Angeles, California 90024

Martin F. Kagnoff, Laboratory of Mucosal Immunology, Department of Medicine, University of California–San Diego, La Jolla, California 92093

Young S. Kim, Gastrointestinal Research Laboratory, Veterans Administration Medical Center, and Department of Medicine, University of California–San Francisco, San Francisco, California 94121

B. Langer, Department of Surgery, University of Toronto, Toronto, Ontario, Canada MG5 1L5

Michael D. Levitt, Research Service, Veterans Administration Medical Center, and Department of Medicine, University of Minnesota, Minneapolis, Minnesota 55417

Varocha Mahachai, Division of Gastroenterology, Department of Medicine, University of Alberta, Edmonton, Alberta, Canada T6G 2C2

Paul N. Maton, Digestive Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

George B. McDonald, Gastroenterology/Hepatology Section, Fred Hutchinson Cancer Research Center, and University of Washington School of Medicine, Seattle, Washington 98104

Gerald Y. Minuk, Department of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada R3A 1R9

Bernard J. Perey, Department of Surgery, Dalhousie University, Halifax, Nova Scotia, Canada B3H 2Y9

Pierre Poitras, Department of Medicine, University of Montreal, and Hôpital Saint-Luc, Montreal, Quebec, Canada H2X 3J4

Linda Rabeneck, Department of Medicine, University of British Columbia, and St. Paul's Hospital, Vancouver, British Columbia, Canada V6Z 1Y6; *present address:* Robert Wood Johnson Clinical Scholars Program, Yale University School of Medicine, New Haven, Connecticut 06510

N. W. Read, Sub-Department of Human Gastrointestinal Physiology and Nutrition, Royal Hallamshire Hospital, Sheffield, United Kingdom S1O 2JF

Claude C. Roy, Department of Pediatrics, Hôpital Sainte-Justine, and University of Montreal, Montreal, Quebec, Canada H3T 1C5

Fergus Shanahan, Division of Gastroenterology, University of California–Los Angeles, School of Medicine, Center for the Health Sciences, Los Angeles, California 90024

Sheila Sherlock, Department of Surgery, Royal Free Hospital School of Medicine, London, England NW3 2QG

Stephan Targan, Division of Gastroenterology, University of California–Los Angeles, School of Medicine, Center for the Health Sciences, Los Angeles, California 90024

Beatriz Tuchweber, Pediatric Research Center, Hôpital Sainte-Justine, and Department of Nutrition, University of Montreal, Montreal, Quebec H3T 1C5

Andrée M. Weber, Department of Pediatrics, Hôpital Sainte-Justine, and University of Montreal, Montreal, Quebec, Canada H3T 1C5

Ibrahim M. Yousef, Pediatric Research Center, Hôpital Sainte-Justine, and Department of Pharmacology, University of Montreal, Montreal, Quebec, Canada H3T 1C5

Foreword

Once again the gastroenterologists of northwestern Canada have come out with a series of essays advancing important modern concepts. As indefatigable as ever, they bring together the latest in clinical–pathophysiological considerations for the clinician. The topics run the gamut from the lovely liver to HIV infection, from the pathophysiology of bile flow to comments on antigens in colorectal cancer.

It is always easy to publish a first volume, for enthusiasm is high and everyone is anxious to win a place. Bringing out a second volume, the carrying forward of a good idea, is so much harder. In many ways this second volume in the series marks the coming of age of Canadian gastroenterology, as 11 of the 21 contributors currently reside in Canada. Still, the contributions from the United Kingdom and the United States make this an Anglo-American festival.

Reading the chapters will make you appreciate the care with which the contributors have garnered recent references to give up-to-date practical information for us all. I enjoyed all that I read.

My congratulations to Drs. Shaffer and Thomson for a job well done.

Howard M. Spiro, M.D.

New Haven, Connecticut

Preface

This book is based on papers presented at the third symposium on Recent Advances in Gastroenterology held by the Canadian Association of Gastroenterology. The proposed audience for this volume is the internist and the general surgeon, as well as those in the specialties of gastroenterology and hepatology. In addition, this publication will be of use and benefit to senior medical residents preparing for subspecialty examinations in internal medicine, general surgery, and gastroenterology.

The rate of change of medical practice and the growth of its scientific and information base are intimidating. This is particularly true of gastroenterology, which has major specialty divisions of its own. Through the generous support of Glaxo Canada Ltd., and with the organizational assistance of the Royal College of Physicians and Surgeons of Canada, the Canadian Association of Gastroenterology is pleased to undertake its commitment to the advancement of science and the improvement in the care of patients with diseases of the gastrointestinal tract. We are fortunate in having a distinguished group of contributors from North America and Europe. Their subjects are diverse, important, and topical.

It gives us special pleasure to acknowledge the considerable assistance from Glaxo Canada, and the support of Dr. Geoffrey Houlton, Vice President of Medical Affairs.

We would also like to express our appreciation to Mr. Frank M. Sabatino of Glaxo Canada for his enthusiastic support of this project.

Alan B. R. Thomson
Eldon A. Shaffer

Edmonton and Calgary

Contents

1

The Clinician Looks at Fatty Liver

Sheila Sherlock

1. INTRODUCTION

Fatty liver is now being diagnosed more often than in the past. The diagnosis may follow a clinically suspicious condition, e.g., unexplained hepatomegaly, but increasingly it accompanies the use of special techniques such as computerized tomography (CT), ultrasound, and liver biopsy.[1] In some instances the cause of fatty liver is obvious, such as alcoholism or obesity. In other instances the fatty change is a small part of another condition such as non-A, non-B hepatitis or Wilson's disease. It may be an integral part of a widespread metabolic abnormality such as Reye's syndrome. Finally, there remains a large group for whom the cause of the fatty liver is never determined.

2. MECHANISMS[2] (FIG. 1)

Fatty liver may be defined as an accumulation of fat, largely triglyceride, exceeding 5% of liver weight. Fat metabolism will be summarized to allow the clinician to understand how fatty liver develops.

Triglyceride fat is formed in the liver from fatty acids. These may come from dietary fat (triglyceride), which forms chylomicrons which are hydrolyzed in adipose tissue and the fatty acids returned to the liver. Some triglyceride is hydrolyzed at the hepatocyte membrane. Medium-chain triglycerides in the intestinal contents are hydrolyzed and the fatty acids so produced travel to the

Sheila Sherlock • Department of Surgery, Royal Free Hospital School of Medicine, London, England NW3 2QG.

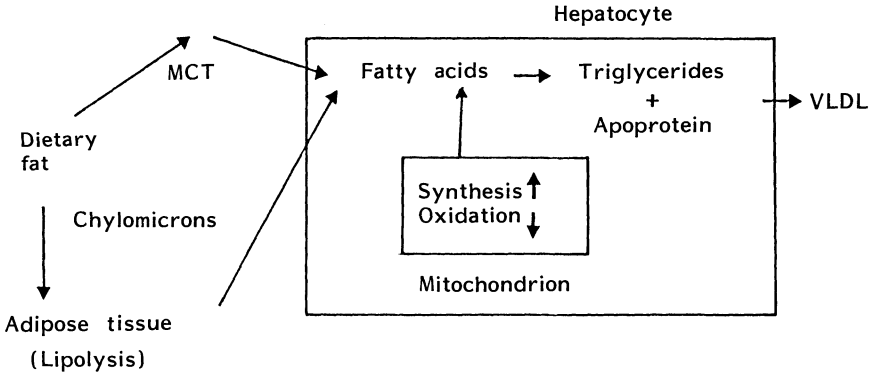


Figure 1. Factors in hepatic steatosis (fatty liver).

liver in the portal vein. Adipose tissue is the major source of fatty acids which are carried to the liver bound to albumin.

Within the liver, fatty acids may be synthesized from carbohydrate. The primary step takes place in the mitochondrion with the formation of acetyl CoA. Later stages take place in the cytoplasm.

Fatty acid oxidation takes place in the mitochondrion by β oxidation. Ultimately, the fatty acids are degraded to acetyl CoA, which is oxidized to carbon dioxide and water within the mitochondrion via the citric acid cycle.

The fatty acids are esterified to triglyceride within the smooth endoplasmic reticulum of the hepatocyte. This is packaged with an apoprotein formed in the rough endoplasmic reticulum, phospholipid and cholesterol, to form very-low-density lipoproteins (VLDL). These are transported to the cell surface as secretory vesicles, ultimately to be excreted into the sinusoid.

Thus, on theoretical grounds, fatty liver could result from increased delivery of fatty acids to the liver, increased synthesis of hepatic fatty acids, decreased oxidation of hepatic fatty acids, or impaired removal of hepatic triglyceride as VLDL particles, often due to the failure of synthesis of apoprotein.

3. DEMONSTRATION OF FATTY LIVER

3.1. Ultrasound

This shows fat as bright areas of increased echogenicity but of normal attenuation. The sensitivity of the technique is 94% and specificity 84%.[3]

Even mild degrees of fatty change may be recognized. The identification of fibrosis is not nearly as satisfactory.

3.2. CT Scanning

This shows the liver with a lower radiological density than normal (Fig. 2). Attenuation values are considerably less than those of the spleen (Table 1). The portal vein radicals appear particularly prominent. An enhanced, dynamic CT scan (after intravenous iodinated contrast material) is of particular value in diagnosing fatty liver and focal hepatic masses.[4] Attenuation values are reduced in acute fatty liver pregnancy,[5] but the technique is relatively insensitive to small amounts of fat.

Monoenergetic CT may be used to assess liver fat content.[6] Results agree with chemical and hepatic histological assessment of liver fat. However, this technique is not generally available.

The spin-echo technique of proton spectroscopic imaging may be useful in detecting fatty liver and in excluding metastases.[7] In the normal liver, the lipid signal is less than 10% whereas with fatty liver it exceeds 10% and is usually greater than 20%.

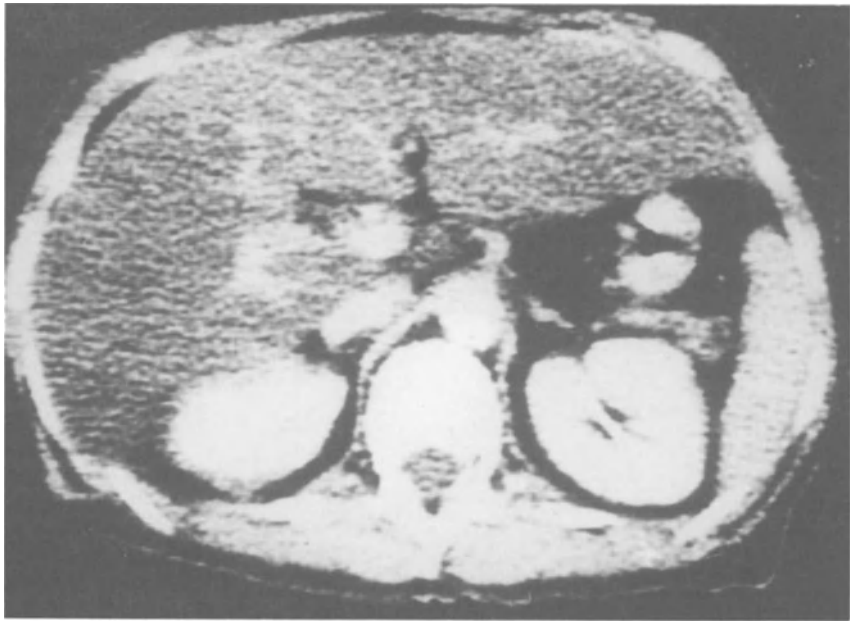


Figure 2. CT scan shows fatty liver. The liver is much less dense than the spleen and kidneys. In a scan unenhanced by contrast the blood vessels stand out as much more dense than the liver.

Table 1. CT Attenuation Units: Relationship of Normal Fatty Liver to Fatty Infiltration

Scan	Attenuation values (HU)	
	Normal liver (SD) ($n = 103$)	Liver with fatty infiltration (SD) ($n = 23$)
Unenhanced		
Liver	55.2 (8.1)	18.5 (21.8)
Spleen	48.2 (7.4)	46.0 (8.0)
Enhanced		
Liver	97.4 (16.4)	45.4 (27.9)
Spleen	109.4 (22.5)	95.5 (14.6)
Spleen minus liver attenuation differences		
Precontrast	7.0 (5.8)	27.5 (19.3)
Postcontrast	12.0 (11.7)	50.1 (26.0)

Note: Values are expressed in Hounsfield units. Significant differences (SD) are shown in parentheses.

3.3. Liver Biopsy

Histologically, two types of fatty change may be recognized: macrovesicular (large droplet) and microvesicular (small droplet)[8] (Fig. 3). Occasionally, the two types may be combined.

In formol-fixed hematoxylin-stained sections, fat appears as punched-out, empty vacuoles. It is essential, especially with minor degrees of fatty change, to confirm the presence of fat by the use of a specific stain, such as oil red O, on frozen sections. Connective tissue stains are also essential to the determination of the existence of zone III fibrosis or of cirrhosis (Fig. 4).

4. MACROVESICULAR FAT DISEASES (TABLE 2)

Histopathology[9] (Fig. 5)

The macrovesicular fat is present as one or two large globules, pushing the nucleus of the hepatocyte to one pole with margination of the cytoplasm. The fat deposited is neutral triglyceride and is usually most marked in zone III

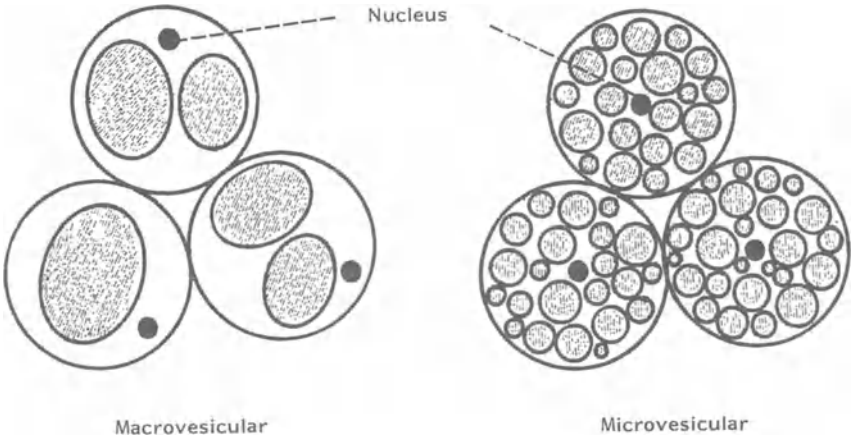


Figure 3. Diagram showing the differences between macrovesicular and microvesicular fatty liver.

(centrizonal) hepatocytes. Organelles in the displaced cytoplasm do not show ultrastructural damage.[10]

Some patients will also show steatohepatitis (fatty liver hepatitis). This is marked by zone III, perisinusoidal, pericellular, and perivenular sclerosis, with or without intracellular deposits of Mallory's hyaline. Micronodular cirrhosis is a very rare accompaniment. The relationship of the fatty change to steatohepatitis and to cirrhosis is obscure. Steatohepatitis is the essential lesion if cirrho-

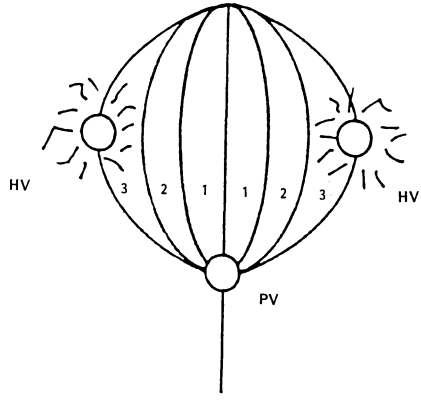


Figure 4. Zone 3 (centrizonal) collagenosis is associated with severe liver injury and is precirrhotic.

Table 2. Etiology of Large Droplet (Macrovesicular) Fatty Liver

Nutritional conditions
Kwashiorkor
Gastrointestinal disease
Pancreatic disease
Obesity
Intestinal bypass
Prolonged parenteral nutrition
Metabolic diseases
Diabetes mellitus
Galactosemia
Glycogenoses
Fructose intolerance
Wilson's disease
Tyrosinemia
Hyperlipidemias
Abetalipoproteinemia
Wolman's disease
Weber–Christian disease
Drug-related conditions
Alcoholism
Corticosteroid toxicity
Direct hepatotoxicity
Viral infections
Non-A, non-B hepatitis
Fever
Systemic disease
Cryptogenic

sis is going to develop. It does not complicate all forms of macrovesicular fatty change. It does not seem simply to be due to steatosis, and an additional factor causing hepatocyte injury seems necessary.

Lipogranulomas are ill-defined cellular accumulations due to rupture of fatty cysts in adjacent hepatocytes with a surrounding cellular reaction.[11] Fatty liver may be symptomless or the patient may complain of dragging pain on the right side of the abdomen, presumably due to the hepatomegaly. Biochemical changes include a modest increase in serum transminases, alkaline phosphatase, bilirubin, and γ-glutamyl transpeptidase. Portal venous pressure may be slightly increased, presumably due to interruption with sinusoidal blood flow by the overdistended hepatocytes.

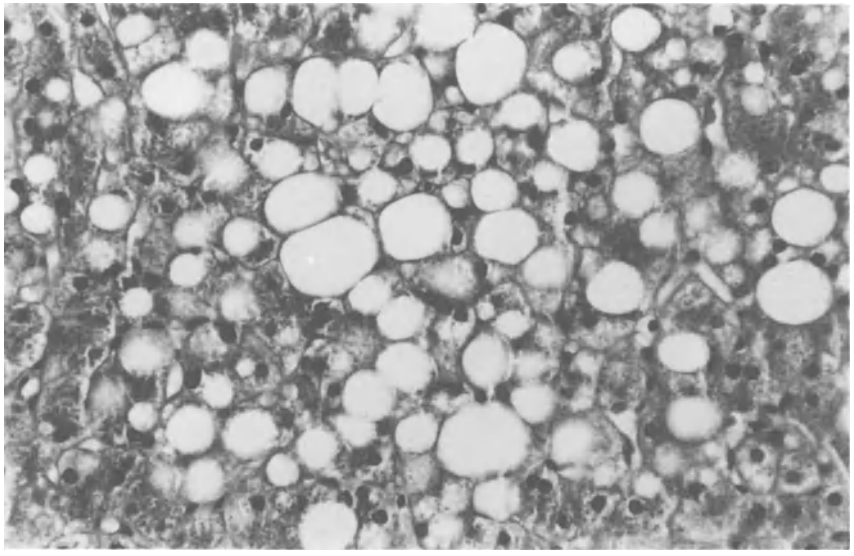


Figure 5. Liver biopsy showing macrovesicular fatty change. The large globules are displacing the hepatocyte nucleus. Stained HE ×425.

5. ALCOHOLIC FATTY LIVER

5.1. Histopathology

The picture is of macrovesicular fat. Occasionally, some microvesicular steatosis may also be seen and this is believed to represent fat in transition to large droplets. The alcoholic "foamy fat" syndrome will be discussed later. Alcoholic hepatitis may be seen and the cellular reaction always includes polymorphs.

5.2. Clinical Features

Clinically, the patient with simple fatty liver cannot be distinguished from one with mild alcoholic hepatitis. Needle liver biopsy is essential for the distinction.

The prognosis is good. If the patient is abstinent, within about 6 weeks the fat disappears from the hepatocyte. However, some 15% of alcoholic patients initially with noncirrhotic liver damage will develop cirrhosis during a follow-up period of at least 10 years.[12] Those with the liver biopsy appearances

of perivenular sclerosis (steatohepatitis) have a worse prognosis than those with simple fatty change.

Sudden death may be associated with fatty liver.[13] The possible causes are multiple and include alcohol withdrawal, pulmonary fat embolism, and hypoglycemia combined with lower serum magnesium levels.

5.3. Treatment

Treatment consists of stopping alcohol and implementing a good balanced diet. No drug is of any proven value.

6. OBESITY

Steatosis is found in 68–94% of persons with body weight more than 73% above ideal.[14] In one series, 59.7% of obese patients showed steatosis; in 28% of these, disease was moderate to severe, involving more than half the hepatocytes.[15] In addition, 8.7% showed an alcoholic hepatitis-like picture and 2.7% cirrhosis. Other workers have confirmed steatohepatitis and cirrhosis.[16] The fibrosis and cirrhosis have been attributed to factors other than obesity, such as alcohol excess, bacterial overgrowth, and concomitant use of drugs.[14] However, as hepatic fibrosis and steatonecrosis have been reported in obese children,[17] a simple relationship to obesity seems likely.

Analysis of liver biopsies shows fatty acids and triglycerides to be high, indicating a preferential incorporation of unsaturated fatty acids into triglycerides.[18] Increased lipolysis from depots increases triglyceride synthesis. There is also an imbalance between protein and calorie intake, further increasing lipogenesis. Hepatic steatosis is proportional to body weight and reflects increased total body stores of fat.[19]

Clinically, there may be mild hepatomegaly with modest changes in biochemical tests including an increase in serum transaminases, alkaline phosphatase, and γ-glutamyl transpeptidase.[20] Treatment is by weight reduction.

7. DIABETES MELLITUS

Approximately 50% of patients with type II (maturity onset) diabetes will show steatosis. It is unusual with type I unless diabetic control is very poor.

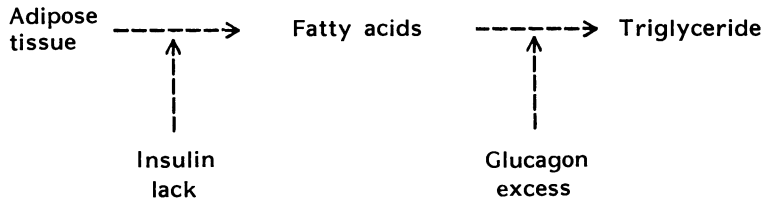


Figure 6. Fatty liver in type II diabetes is related to insulin deficiency and glucagon excess.

Insulin deficiency or glucagon excess, the essential defects in diabetes mellitus, inhibit glucose uptake and enhance triglyceride lipolysis, so increasing plasma free fatty acids (Fig. 6). These fatty acids are taken up by the liver. Within the hepatocyte, insulin deficiency or glucagon excess increase glycogen degradation and gluconeogenesis while inhibiting glucose utilization.[21] As a result, the requirement for hepatic oxidation of fat is enhanced. The liver can dispose of the fat only by forming triglyceride and this excess is stored in the hepatocyte. Obesity adds to the problem by increasing lipolysis and fatty acid transfer to the liver.

Occasionally, in addition to fatty liver, steatohepatitis and cirrhosis may be present.[22] It is unknown why and how often steatohepatitis develops and how frequently it is followed by cirrhosis. In addition to steatosis and steatonecrosis, hepatocytes may show nuclear vacuolation and polymorphs are absent.[22–24] These give clues to the diabetic association.

Clinically the liver is enlarged and may be tender. Biochemical changes include increases in transaminases, alkaline phosphatase, and γ-glutamyl transpeptidase which may rarely be found with an increase in serum bilirubin.

The hepatomegaly and fatty liver hepatitis may precede the development of glucose intolerance.[25] Prediabetes should always be suspected in a patient with unexplained fatty liver. Treatment is by better control of the diabetes and attention to maintaining ideal body weight.

8. JEJUNO-ILEAL BYPASS

Two percent of patients having a jejuno-ileal bypass to control morbid obesity will die from the procedure, usually with liver failure.[26] Steatohepatitis is the commonest underlying hepatic lesion. Toxic products of bacterial metabolism absorbed from the blind loop may be responsible. The condition is much

more common if the obese patient had pericellular, zone III fibrosis preoperatively.[27]

9. COMPLICATIONS OF PARENTERAL NUTRITION

Hepatic steatosis is usually related to protein–carbohydrate imbalance. Thus, fatty infiltration has been observed in patients receiving intravenous infusions with large amounts of dextrose as the only calorie source. Rates of carbohydrate infusion above the hepatic oxidative capacity promote fat synthesis.[28] Increased activities of rate-controlling enzymes of triglyceride synthesis are combined with diminished apolipoprotein production and consequent failure of triglyceride export from the hepatocyte.[29] Infusion of fat emulsions also increase triglyceride, cholesterol, and phospholipid in liver.[29] If total parenteral nutrition is balanced as regards protein, carbohydrate, and fat, then abnormalities of liver function tests, particularly serum transaminase increases, are common but fatty infiltration is not seen.[30]

10. KWASHIORKOR

The fatty deposition is gross and most marked in zone III. Analysis of hepatic lipid shows an increase in triglyceride, phospholipid, and a normal cholesterol content. Kwashiorkor is marked by dietary protein depletion with adequate calories, largely provided as carbohydrate.

Two factors are involved in the hepatic steatosis. There is increased mobilization of fatty acid from depots and decreased synthesis of VLDL apoproteins preventing excretion of triglyceride from the liver[31] (Fig. 7).

In children, the fatty change is reversible and steatohepatitis and cirrhosis are not complications.[32]

11. ABETALIPOPROTEINEMIA

This rare, autosomal recessive disease is marked by the accumulation of large quantities of lipid droplets of all sizes within the hepatocyte. Fibrosis and steatohepatitis are not seen. Scanning electron microscopy shows acanthocytes in sinusoids. The serum shows mild increases in transaminases.

In abetalipoproteinemia, defective apolipoprotein B synthesis prevents assembly of lipoproteins that can be excreted through the intracellular communi-

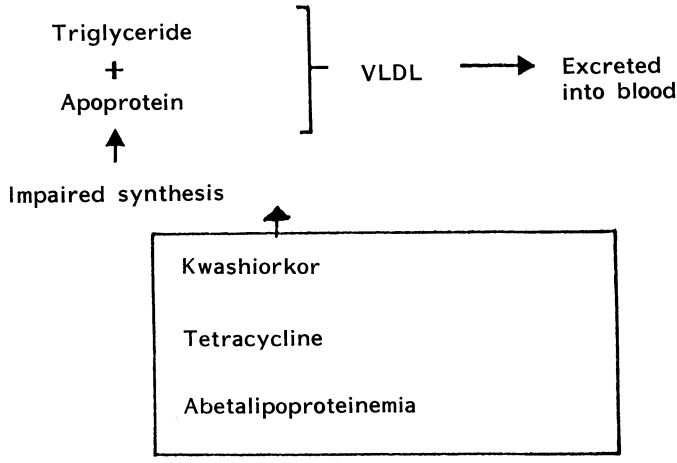


Figure 7. In these conditions fatty change is related to impaired synthesis of VLDL apoproteins preventing excretion of triglyceride from the liver.

cating tubular network. Consequently, fat accumulates in the cytoplasm of the hepatocytes due to failure of export from the liver.[33]

12. DRUGS

Macrovesicular fat in the liver is a common accompaniment of many hepatic drug reactions.[1] It is seen particularly in the direct, dose-dependent, metabolite-related hepatotoxicity group. Causes include carbon tetrachloride or acetaminophen overdose and occasionally as a reaction to halothane.

Some drugs produce an hepatic histological picture closely resembling that of viral hepatitis. However, fat, an unusual accompaniment of viral hepatitis, is usually present and this gives an important clue to drugs as the cause.[8]

Some drugs cause a histological picture in the liver identical to that of acute alcoholic hepatitis. These drugs lead to a striking increase in lysosomal phospholipid not only in the hepatocytes but also in Kupffer cells and peripheral nerves.[34,35] The relation of the phospholipidosis to the fatty liver is not known. Perhexiline maleate, an antianginal drug, now withdrawn, has been associated with this picture in the liver.[34,35] Patients developing the reaction lack a specific gene, the debrisoquine gene necessary for oxidation of the antihypertensive drug Debrisoquine.[36] Somewhat similar hepatic reaction is associated with Amiodarone, an effective antiarrhythmic drug.[35,37] Amiodarone and its major metabolite, N-desethyl amiodarone, are present in the liver for several months after stopping the drug.[35] Cirrhosis can develop. Electron microscopy

shows pleomorphic lysosomal inclusions in the liver, resembling those described in phospholipidosis.[37]

Massive doses of synthetic estrogens, given long term to treat prostatic cancer, may lead to hepatomegaly, increased serum γ-glutamyl transpeptidase, and a histological picture of steatonecrosis.[38]

The reaction to cytotoxic drugs such as methotrexate is largely fibrotic; however, they can all lead to steatosis, maximal in zone I. Corticosteroid therapy can lead to a fatty liver with droplets scattered in all zones. This is due to mobilization of fatty acids from adipose tissue.

13. MISCELLANEOUS DISEASES

It is not unusual to find fatty change in the liver in patients with miscellaneous infections and wasting diseases, and it is a frequent finding at autopsy. Multiple factors are involved, including obesity, infections, starvation, systemic disease, and unsuspected alcoholism. Fat may even develop in the liver after death, for in *congestive heart failure* liver biopsies show little fat, whereas at autopsy the typical yellow (fat) and red (zone III hemorrhage) nutmeg appearance is noted.[1] Mechanisms include depressed mitochondrial function with inhibition of fatty acid oxidation and failure of apoprotein synthesis. *Starvation* leads to increased delivery of fatty acids from adipose tissue to the liver. *Nutritional factors* may contribute. In the rat, choline- and methionine-deficient diets lead to deficiencies of lipotropes (labile, methyl groups) and to zone III fat fibrosis, fatty cysts, and Mallory's hyaline fatty cysts (fatty liver hepatitis).[39,40]

In *fulminant ulcerative colitis,* steatosis may be marked. The incidence is higher when autopsy, rather than biopsy, material is used for diagnosis.[41] Wedge liver biopsies from patients undergoing surgery for ulcerative colitis show 77% with significant histological changes, 35% with fatty liver.[42] Severe steatosis, resembling kwashiorkor, may develop in adults with *intestinal disease,* usually with a blind loop or after partial gastrectomy.[43] This is related to ineffective utilization of dietary protein due to increased enteric bacterial colonization.

Steatosis may accompany *pancreatic disease* such as cystic fibrosis. This is analogous to the situation in the pancreatectomized dog. Deficiencies of insulin and glucagon contribute.

Many inherited *metabolic diseases* of the liver are accompanied by hepatic steatosis. This is exemplified by the glycogenoses, galactosemia, glutaric aciduria (type II) and hereditary fructose intolerance. Steatohepatitis is a feature of Weber–Christian disease.[44]

In many instances, steatosis forms part of the histological pattern of that particular disease. The presence of fat, however, may give away the diagnosis. In *Wilson's disease,* fatty change is usual with accompanying ballooned cells,

clumped glycogen, and glycogenic vacuolation. Such appearances in a young person with chronic active hepatitis or cirrhosis should always suggest Wilson's disease.

Steatosis with acute hepatitis suggests the non-A, non-B type.[45] Occasionally, severe steatosis may be associated with *mild chronic hepatitis*.[46,47] Serum transaminases may be increased. The fatty change is so marked it tends to obscure the underlying chronic hepatitis with round cell infiltration in the portal tracts, piecemeal necrosis, and some lobular necrosis. The course is benign.

Familial cirrhosis has been described in association with steatosis.[48] Steatosis was also found in family members who did not have cirrhosis.

14. MICROVESICULAR FAT DISEASES

14.1. Histopathology

The liver architecture is normal. The hepatocytes are swollen. The fat accumulation is maximal in zone III and, in hematoxylin and eosin-stained sections, it has a foamy appearance (Fig. 8). Frozen sections stained for fat may be necessary for diagnosis if the fat deposition is minimal. Nuclei with

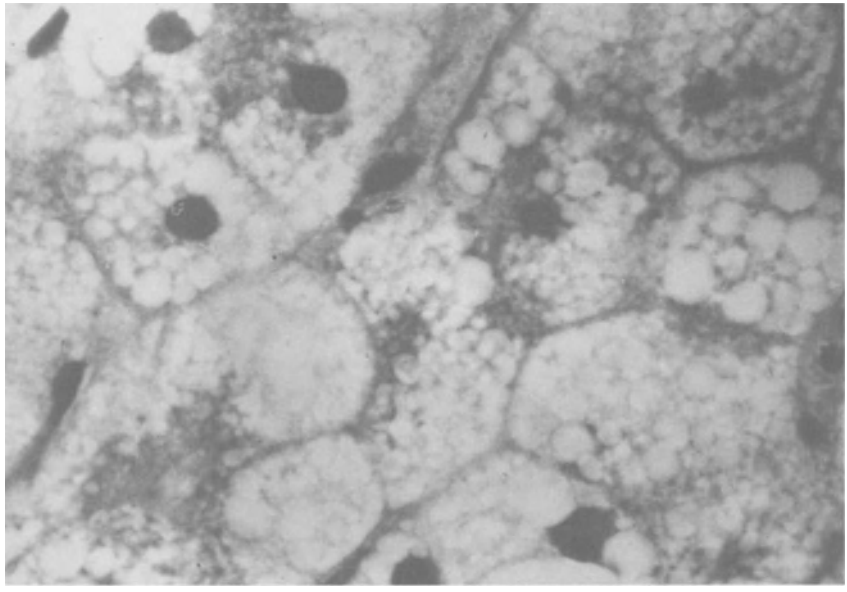


Figure 8. Microvesicular fat. The hepatocytes show a foamy appearance with prominent centrally placed hepatocytes. Stained HE ×350.

prominent nucleoli are centrally placed in the hepatocyte. Inflammation and cell necrosis are absent or minimal.

Electron microscopy shows mitochondria to be swollen, pleomorphic, and varying in shape. The smooth endoplasmic reticulum may be fragmented and lost. The rough endoplasmic reticulum is distorted and dilated, and ribosomes are detached.

14.2. Mechanism

The fat is deposited as fatty acid in acute fatty liver of pregnancy (Fig. 9) and as triglyceride in Reye's syndrome and other members of the group.

The fatty change is probably related to mitochondrial injury with depression of fatty acid oxidation. In addition, depression of synthesis of the apoprotein of very-low-density lipoprotein impairs the exit of lipid from the liver and triglyceride accumulates.

Increases in blood ammonia combined with low citrulline values can be related to reduction of mitochondrial Krebs cycle enzymes. The serum amino

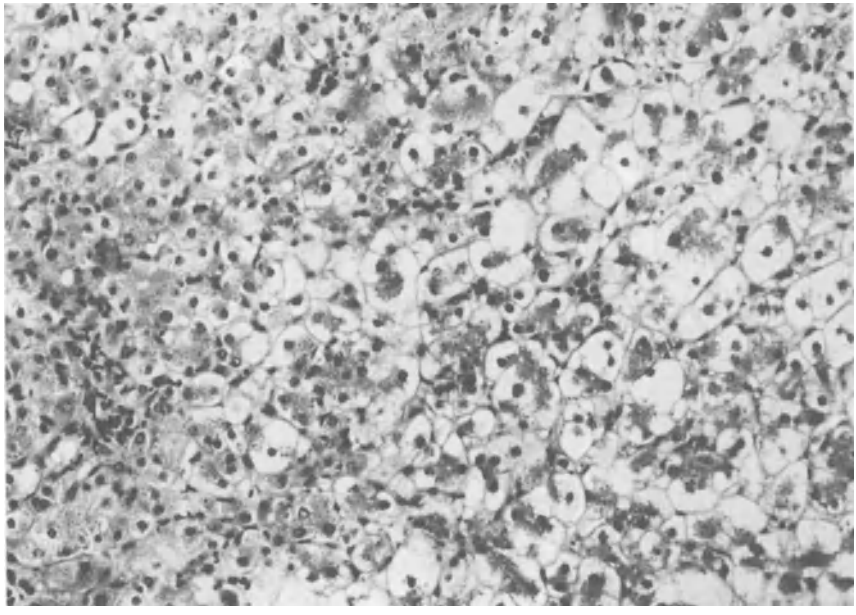


Figure 9. Acute fatty liver of pregnancy. The appearances are those of microvesicular fat. Hepatocytes appear as empty, resembling plant cells. Stained HE × 150.

Table 3. Microvesicular Fat Diseases

Acute fatty liver of pregnancy
Reye's syndrome
Vomiting disease of Jamaica
Sodium valproate toxicity
Congenital defects of urea cycle enzymes
Alcoholic foamy fat syndrome
Delta hepatitis in northern South America

acid profile shows high glutamine, alanine, and lysine levels. Hypoglycemia is frequent and inhibition of the citric cycle enzymes may contribute.

The mode of initiation of this family of diseases is diverse and in most instances not fully understood. Viral, toxic, and nutritional factors have been implicated.

Certain diseases may be grouped together on the basis of finding microvesicular fat in the hepatocytes. These diseases can be regarded as a family that has similarities but differs in various respects among its members.[49] The diseases include acute fatty liver of pregnancy, Reye's syndrome, sodium valproate toxicity, tetracycline toxicity, congenital defects in urea cycle enzymes, the foamy fat cholestatic syndrome of alcoholics, and South American hepatitis delta virus infection (Table 3). They all show the same general pattern (Table 4). The onset is marked by fatigue, nausea, vomiting (often severe and persistent), with variably jaundice, impairment of consciousness, and coma. Seizures may occur. Renal failure and disseminated intravascular coagulation may be

Table 4. Features of the Microvesicular Fat Diseases

Vomiting
Variable jaundice
Coma
Disseminated intravascular coagulation
Renal failure
Rise in blood ammonia values
Hypoglycemia
Rise in serum fatty acids
Liver biopsy
Microvesicular fat
Necrosis and cellular infiltration not prominent
Electron microscopy
Mitochondrial abnormalities

complications. The liver is not usually the only organ involved and triglyceride accumulation may be found in the renal tubules and occasionally in myocardium and pancreas. Liver failure is not the usual cause of death. Coma may be related to increases in blood ammonia levels or to cerebral edema. Bleeding is due to disseminated intravascular coagulation rather than to failure of the liver to synthesize blood-clotting factors.

15. ACUTE FATTY LIVER OF PREGNANCY (TABLE 5)

Acute fatty liver should be suspected in a woman presenting between the 30th and 38th week of pregnancy with nausea, repeated vomiting, and abdominal pain followed by jaundice.[50] In some patients, hypertension, edema, and proteinuria suggest an overlap with toxemias of pregnancy and eclampsia. Raised serum uric acid levels and a characteristic blood picture with neutrophilia, thrombocytopenia, normoblasts, giant platelets, and basophilic stippling are points of distinction from preeclampsia and viral hepatitis. The mortality, both for mothers and babies, is diminishing with recognition of less severe cases[51] and with earlier termination of pregnancy.[52]

CT scanning may be a useful noninvasive alternative to liver biopsy as an indicator of hepatic fat accumulation.[5,6]

Histopathology

In addition to the microvesicular fat, the liver may show lobular disarray, patchy hepatocellular necrosis, lobular inflammation, and reticulin condensation. Abnormal fat is not seen in other organs.

Table 5. Diagnosis of Acute Fatty Liver of Pregnancy

Vomiting last trimester
High serum uric acid
Blood film
Neutrophilia
Thrombocytopenia
Normoblasts
Liver biopsy
? Transjugular
Frozen sections
CT scan for fat

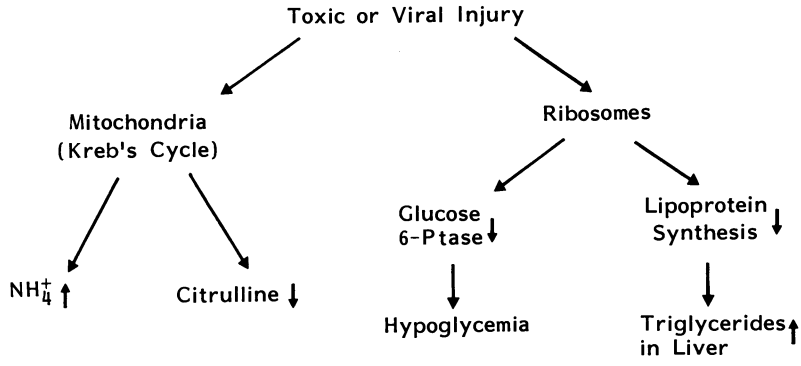


Figure 10. Biochemical changes in Reye's syndrome.

Electron microscopy shows the mitochondria to be swollen, pleomorphic, and varying in shape.[53] Dense bodies are seen; these are absent in Reye's syndrome. The ribosomal pattern in the rough endoplasmic reticulum is also abnormal. The cause is unknown. The condition does not recur in subsequent pregnancies and an inherited metabolic defect seems unlikely.

16. REYE'S SYNDROME

The sufferer is usually less than 14 years old. Three to seven days after an acute viral-type illness, the child develops vomiting and progressive neurological deterioration.[54] In severe cases, medullary coning and brain death result within 4–60 hr. Milder cases may remain undiagnosed.[55] The syndrome has followed almost any known viral disease, and salicylate administration has been incriminated.[56]

Blood ammonia and transaminase values are increased (Fig. 10). Hypoglycemia is a feature of the seriously ill, especially those less than 2 years old.

The disease is not confined to the liver and, indeed, death is usually due to cerebral edema. The liver shows microvesicular fat which is also seen in proximal renal tubules and myocardium.

Electron microscopy shows widespread mitochondrial changes with expansion of the matrix thickness, flocculation, and granularity. Mitochondrial dense bodies are not seen.[53]

The fatty change is related to depression of apoprotein synthesis. Mitochondrial injury would also depress oxidation of fatty acids. A block in the β oxidation of fatty acids has indeed been shown in the liver of patients with Reye's syndrome.[57] Plasma free fatty acids are increased. The activated form

of the long-chain fatty acids in the liver, the coenzyme A (CoA) esters, exhibits inhibitory effects on mitochondrial function.[58]

17. TETRACYCLINE TOXICITY

Large doses of intravenous tetracyclines produce a microvesicular fat syndrome. In the 1960s this was believed to be the explanation of acute fatty liver of pregnancy.[59] However, it was soon realized that tetracycline had the same effect on the liver and other organs in the nonpregnant.[60]

Tetracycline binds to transfer RNA in the hepatocyte and impairs protein synthesis including that of the apoprotein of VLDL. Mitochondrial oxidation of fatty acids, formation of triglyceride, and hepatic uptake of fatty acids may also be impaired.

18. SODIUM VALPROATE TOXICITY

This drug reaction is one of the few to affect children, with 69% of sufferers less than 10 years old.[61] The patient has usually been taking the drug for more than 1 month and less than 6 months. Onset is with vomiting, abdominal discomfort, convulsions, and sometimes hypoglycemia. Jaundice is late. Disseminated intravascular coagulation may be a complication.

Hepatic histology shows microvesicular fat with hepatocellular (usually zone III) necrosis of variable degree, a finding unusual in the other microvesicular fat diseases.[62]

Multiple metabolic defects have been seen in experimental animals and humans taking valproate. Toxic metabolites of valproate include 1-propyl-4-pentenoic acid, which is structurally related to hypoglycin A, a toxic metabolite of Ackee apples that has been incriminated in Jamaican vomiting sickness, another disease associated with microvesicular fat.[61]

19. CRANIOCEREBRAL TRAUMA

Microvesicular fat may be seen with sudden death in childhood, e.g., from head injury or drowning.[63]

20. UREA CYCLE ENZYME DEFECTS

Congenital defects in urea cycle enzymes are associated with high blood ammonia values and microvesicular fatty change in the liver. It has even been

suggested that sufferers from acute fatty liver of pregnancy may have bile defects in these enzymes. This seems unlikely since acute fatty liver does not recur in subsequent pregnancies. However, in one family with carbamoyl transferase deficiency, a fatality seemed to be associated with valproate therapy.[64]

21. ALCOHOLIC FOAMY FAT DEGENERATION

Occasionally, the alcoholic patient presents a cholestatic picture with deep jaundice, hepatomegaly, and increased serum alkaline phosphatase, transaminases, and cholesterol.[65,66] Functional renal failure is usual. This is often the first episode of hepatic decompensation. The prognosis is very variable.

Liver biopsy histology shows a normal liver architecture, microvesicular fat predominantly in zone III, giant mitochondria, cholestasis, foci of liver cell necrosis, mild perisinusoidal fibrosis, and variable portal zone edema with some neutrophil inflammation.

Electron microscopy[66] shows changes in the smooth and rough endoplasmic reticulum and mitochondrial changes varying from giant forms to atrophy and loss of matrix and cristae. The Golgi apparatus is altered. Disse's space contains some collagen fibers and sometimes loss of bile canaliculi.

22. DELTA HEPATITIS INFECTION IN NORTHERN SOUTH AMERICA

Intense microvesicular steatosis of the hepatocyte has been reported with fetal delta virus hepatitis in indigent populations of hepatitis B carriers in northern areas of South America, particularly the Amazon Basin. The outbreaks have involved Yucpa Indians in Venezuela.[67,68] In Brazil, the condition is termed Labrea fever,[69] and in northern Colombia, Santa Marta hepatitis.[70] Steatosis is accompanied by conspicuous eosinophilic necrosis of hepatocytes and portal inflammation. Hepatic delta antigen can be detected in the liver. Delta hepatitis seems to differ from hepatitis B in that virtual cytotoxicity is more prominent than lymphocytotoxicity. The microvesicular steatosis is best explained by a severe cytotoxic reaction.

23. FOCAL FATTY LIVER

Focal deposits of fat in an otherwise normal, or nearly normal, liver have been described as a clinical,[71] histological,[72] and radiological[73] entity. The increased use of imaging has led to its more frequent recognition. The focal fat may simulate mass lesions such as metastases.[74] The lesions are multiple, non-

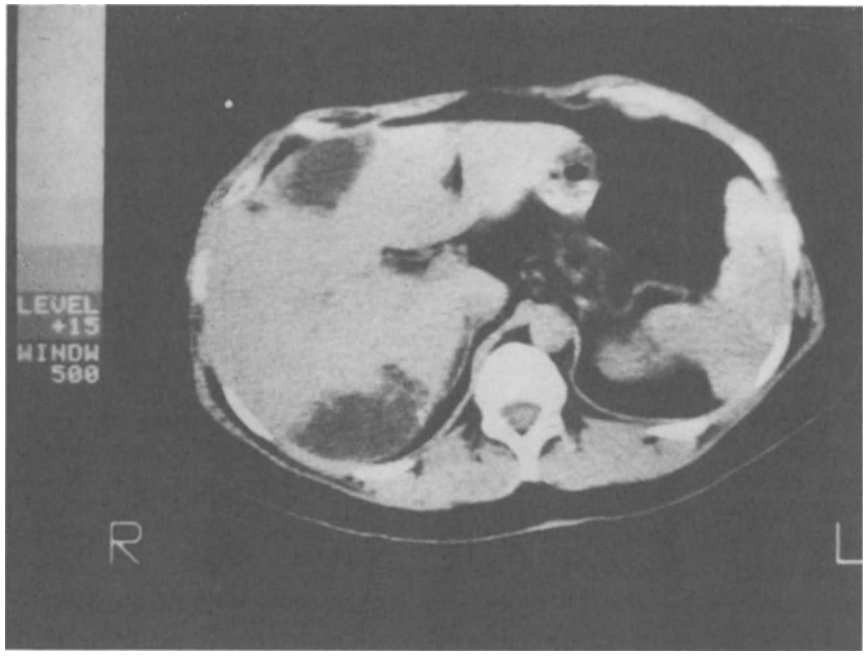


Figure 11. Focal fatty liver. CT scan shows two angular subcapsular lesions of decreased attenuation.

spherical, and often angular and subcapsular; they vary in number (Fig. 11). They show decreased attenuation on a CT scan[75] and are hyperechoic on sonography.[76] Occasionally they have a segmental or lobular distribution. Sometimes there is a central core of higher density.[77] The lesions may persist or disappear only to return to other areas of the liver.

Radionuclide scans show appearances that do not differ from normal. The xenon-133 scan shows the isotope to be taken up by the fat and the picture recorded matches that of the CT scan.[78] If necessary, the nature of the lesion can be confirmed by guided liver biopsy.

A vascular basis for the lesions is suggested by the distribution which matches the lobar and segmental divisions of the hepatic arteries and portal veins.[79]

Focal fatty liver has many associations, particularly with diabetes, alcoholism, obesity, hyperalimentation, Cushing's syndrome, and corticosteroid therapy. These are associations similar to those of diffuse fatty change in the liver.

Occasionally, fatty change extends throughout the liver with the exception

of small flattened portions of less affected parenchyma called "skip areas."[80] These areas are often subcapsular and may be supplied by systemic venous blood from the capsule of the liver and the gallbladder rather than by splanchnic venous blood through the portal vein.

24. "CRYPTOGENIC" FATTY LIVER

Liver biopsy and imaging procedures such as ultrasound and CT scanning are resulting in many patients being identified with excess fat within the liver (steatosis). When common causes such as obesity, alcoholism, and diabetes have been excluded, a hard core remains with no obvious etiology. Some patients may be prediabetic. In the absence of steatohepatitis, this is probably of little clinical significance. Patients usually have no symptoms other than anxiety. Serum transaminases and γ-glutamyl transpeptidase levels may be slightly increased. Treatment is by reassurance and avoiding overinvestigation.

REFERENCES

1. Sherlock S. *Diseases of the Liver and Biliary System,* 7th ed. Oxford: Blackwell Scientific, 1985; pp. 381–385.
2. Hoyumpa AM, Green HL, Dunn GD, Schenker S. Fatty liver: Biochemical and clinical considerations. *Dig Dis Sci* 1975;20:1142–1170.
3. Saverymuttu SH, Joseph AEA, Maxwell JD. Ultrasound scanning in the detection of hepatic fibrosis and steatosis. *MJ* 1986;292:13–15.
4. Scatarige JC, Scott WW, Donovan PJ et al. Fatty infiltration of the liver: Ultrasonographic and computed tomographic correlation. *J Ultrasound Med* 1984;3:9–14.
5. McKee CM, Weir PE, Murnaghan GA, Callender ME. Acute fatty liver of pregnancy and diagnosis by computed tomography. *Br Med J* 1986;292:291–292.
6. Bydder GM, Kreel L, Chapman RWG, Harry D, Sherlock S. Accuracy of computed tomography in diagnosis of fatty liver. *Br Med J* 1980;281:1042–1044.
7. Heiken JP, Lee JKT, Dixon WT. Fatty infiltration of the liver: Evaluation by proton spectroscopic imaging. *Radiology* 1985;157:707–710.
8. Scheuer PJ. *Liver Biopsy Interpretation,* 3rd ed. London: Bailliere-Tindall, 1980.
9. Baptista A, Bianchi L, DeGroote J et al. Alcoholic liver disease: Morphological manifestations. Review by an international group. *Lancet* 1981;1:707–711.
10. Uchida T, Kao H, Quispe-Sjogren M et al. Alcoholic foamy degeneration: A pattern of acute alcoholic injury of the liver. *Gastroenterology* 1983;84:683–692.
11. Christoffersen P, Braendstrup O, Juhl E et al. Lipogranulomas in human liver biopsies with fatty change. *Acta Pathol Microbiol Scand Sec A* 1971;19:150–158.
12. Sorensen TIA, Orholm M, Bentsen KD et al. Prospective evaluation of alcohol abuse and alcoholic liver injury in men as predictors of the development of cirrhosis. *Lancet* 1984;2:241–243.
13. Randall BR. Fatty liver and sudden death: A review. *Hum Pathol* 1980;11:147–153.

14. Braillon A, Capron JP, Herve MA, Degott C, Quenu MC. Liver in obesity. *Gut* 1985;26:133–139.
15. Nasrulla SM, Wills CE Jr, Galambos JT. Hepatic morphology in obesity. *Dig Dis Sci* 1981;26:325–327.
16. Adler M, Schaffner F. Fatty liver, hepatitis and cirrhosis in obese patients. *Am J Med* 1979;67:811–816.
17. Moran JR, Ghishan FK, Halter SA, Greene HL. Steatohepatitis in obese children: A cause of chronic liver dysfunction. *Am J Gastroenterol* 1983;78:374–377.
18. Mavrelis PG, Ammon HV, Gleysteen JJ et al. Hepatic free fatty acids in alcoholic liver disease and morbid obesity. *Hepatology* 1983;3:226–231.
19. Sherlock S, Bearn AG, Billing B. The effect of insulin on the liver in normal and diabetic man. In *Trans 10th Conf Liver Injury*. New York: Josiah Macy Jr. Foundation.
20. Bray GA. Complications of obesity. *Ann Intern Med* 1985;103:1052–1062.
21. Cherrington AD, Steiner KE. The effects of insulin on carbohydrate metabolism in vivo. *Clin Endocrinol Metab* 1982;11:307–328.
22. Robbers H, Strohfeldt P, Kruger CH. Differential diagnosis between diabetic and alcoholic fatty liver: A study of 171 diabetics and 100 chronic alcoholics. *German Med Monthly* 1968;13:124–125.
23. Itoh S, Tsukada Y, Motomura Y, Ichinoe A. Five patients with nonalcoholic diabetic cirrhosis. *Acta Hepatogastroenterol* 1979;26:90–97.
24. Falchuk KR, Fiske SC, Haggitt RC, Federman M, Trey C. Pericentral hepatic fibrosis and intracellular hyalin in diabetes mellitus. *Gastroenterology* 1980;78:535–541.
25. Batman PA, Scheuer PJ. Diabetic hepatitis preceding the onset of glucose intolerance. *Histopathology* 1985;9:237–243.
26. McGill DB, Hympherys SR, Baggenstoss AH, Dickson ER. Cirrhosis and death after jejunoileal shunt. *Gastroenterology* 1972;63:872–877.
27. Haines NW, Baker AL, Boyer JL et al. Prognostic indicators of hepatic injury following jejunoileal bypass performed for refractory obesity: A prospective study. *Hepatology* 1981;1:161–167.
28. Wolfe RR, O'Donnell TF, Stone MD et al. Investigation of factors determining the optimal glucose infusion rate in total parenteral nutrition. *Metabolism* 1980;29:892–900.
29. Kaminski DL, Adams A, Jellinek M. The effect of hyperalimentation on hepatic lipid content and lipogenic enzyme activity in rats and man. *Surgery* 1980;88:93–100.
30. Bengoa JM, Hanauer SB, Sitrin MD, Baker AL, Rosenberg IH. Pattern and prognosis of liver function test abnormalities during parenteral nutrition in inflammatory bowel disease. *Hepatology* 1985;5:79–84.
31. Truswell AS, Hansen JDL, Watson CE, Wannenburg P. Relation of serum lipids and lipoproteins to fatty liver in kwashiorkor 1,2. *Am J Clin Nutr* 1969;22:568–576.
32. Cook GC, Hutt MSR. The liver after kwashiorkor. *Br Med J* 1967;3:454–457.
33. Avigan MI, Ishak KG, Gregg RE, Hoofnagle JH. Morphologic features of the liver in abetalipoproteinemia. *Hepatology* 1984;4:1223–1226.
34. Beaugrant M, Poupon R, Levy VG et al. Lesions hepatiques dues au maleate de perhexiline. Etude clinique, biologique, histopathologique, ultrastructurale et biochemique de 7 cas. *Gastroenterol Clin Biol* 1978;2:579–588.
35. Poucell S, Ireton J, Valencia-Mayoral P et al. Aminodarone-associated phospholipidosis and fibrosis of the liver: Light, immunohistochemical and electron microscopic studies. *Gastroenterology* 1984;86:926–936.
36. Morgan MY, Reshef R, Shah RR, Oates NS, Smith RL, Sherlock S. Impaired oxidation of debrisoquine in patients with perhexiline liver injury. *Gut* 1984;25:1057–.
37. Simon JB, Manley PN, Brien JF, Armstrong PW. Amiodarone hepatoxicity simulating alcoholic liver disease. *N Engl J Med* 1984;311:167–172.

38. Seki K, Minami Y, Nishikawa M et al. Non-alcoholic steatohepatitis induced by massive doses of synthetic estrogen. *Gastroenterol Japan* 1983;18:197–203.
39. Gyorgy P, Goldblatt H. Hepatic injury on nutritional basis in rats. *J Exp Med* 1939;70:185–192.
40. Hartroft WS. Intracellular (''pseudo-alcoholic'') lyalin in experimental dietary cirrhosis of rats and mice. *Am J Pathol* 1958;34:603.
41. Dordal E, Glagov S, Kirsner JB. Hepatic lesions in chronic inflammatory bowel disease. 1. Clinical correlations with liver biopsy diagnosis in 103 patients. *Gastroenterology* 1967;52:239–253.
42. Eade MN. Liver disease in ulcerative colitis. 1. Analysis of operative liver biopsy in 138 consecutive patients having colectomy. *Ann Intern Med* 1970;72:475–487.
43. Neale G, Antcliff AC, Welbourn RB, Mollin DL, Booth CC. Protein malnutrition after partial gastrectomy. *Q J Med* 1967;36:469–94.
44. Kimura H, Kako M, Yo K, Oda T. Alcoholic hyalins (Mallory bodies) in a case of Weber-Christian disease: Electron microscopic observations of liver involvement. *Gastroenterology* 1980;78:807–812.
45. Kryger P, Christoffersen P. Light microscopic morphology of acute hepatitis non-A, non-B: A comparison with hepatitis type A and B. *Liver* 1982;2:200.
46. Alvarez SJ. Chronic hepatitis with fatty changes. *Hepatology* 1983;3:1058 (abstract).
47. Bruguera M, Zambon D, Ros E, Sanchhez Tapias JM, Rodes J. Chronic hepatitis: A possible etiology of fatty liver. *Liver* 1985;5:111–116.
48. Altman AR, Gottfried EB, Paronetto F, Lieber CS. Idiopathic familial cirrhosis and steatosis in adults. *Gastroenterology* 1979;77:1211–1216.
49. Sherlock S. Acute fatty liver of pregnancy and the microvesicular fat diseases. *Gut* 1983;24:265–269.
50. Burroughs AK, Seong NH, Dojcinov DM, Scheuer PJ, Sherlock SVP. Idiopathic acute fatty liver of pregnancy in 12 patients. *Q J Med* 1982;481–497.
51. Pockros PJ, Peters RL, Reynolds TB. Idiopathic fatty liver of pregnancy: Findings in ten cases. *Medicine* (Baltimore) 1984;63:1–11.
52. Kaplan MM. Current concepts. Acute fatty liver of pregnancy. *N Engl J Med* 1985;313:367–370.
53. Weber FL Jr, Snodgrass PJ, Powell DE, Rao P, Huffman SL, Brady PG. Abnormalities of hepatic mitochondrial urea-cycle enzyme activities and hepatic ultrastructure in acute fatty liver of pregnancy. *J Lab Clin Med* 1979;94:27–41.
54. Partin JC. Reye's syndrome (encephalopathy and fatty liver): Diagnosis and treatment. *Gastroenterology* 1975;69:511–518.
55. de Vivo DC. How common is Reye's syndrome? *N Engl J Med* 1983;309:179–181.
56. Halpin TJ, Holtzhauer FJ, Campbell RJ et al. Reye's syndrome and medication use. *J Am Med Assoc* 1982;248:687–691.
57. Pande SV, Blanchaer MC. Reversible inhibition of mitochondrial adenosine diphosphate phosphorylation by long chain acyl coenzyme A esters. *J Biol Chem* 1971;246:402–411.
58. Kang ES, Capaci MT, Korones DN, Tekade N. Liver coenzyme A ester content: comparison between Reye's syndrome and control subjects. *Clin Sci* 1982;63:455–460.
59. Kunelis CT, Peters RL, Edmondson HA. Fatty liver of pregnancy and its relationship to tetracycline therapy. *Am J Med* 1965;38:359–377.
60. Schiffer MA. Fatty liver associated with administration of tetracycline in pregnant and non-pregnant women. *Am J Obstet Gynecol* 1966;96:326–332.
61. Zimmerman HJ, Ishak KG. Valproate-induced hepatic injury: Analysis of 23 fatal cases. *Hepatology* 1982;2:591–597.
62. Zanfrani ES, Berthelot P. Sodium valproate in the induction of unusual heaptotoxicity. *Hepatology* 1982;2:648–649.

63. Bonnell HJ, Beckwith JB. Fatty liver in sudden childhood death. Implications for Reye's syndrome? *Am J Dis Child* 1986;140:30–33.
64. Hjelm M, de Silva LYK, Seakins JWT, Oberholzer VG. Evidence of inherited urea cycle defect in a case of fatal valproate toxicity. *Br Med J* 1986;292:23–24.
65. Morgan MY, Sherlock S, Scheuer PJ. Acute cholestasis, hepatic failure and fatty liver in the alcoholic. *Scand J Gastroenterol* 1978;13:299–303.
66. Uchida T, Kao H, Quispe-Sjogren M et al. Alcoholic foamy degeneration: A pattern of acute alcoholic injury of the liver. *Gastroenterology* 1983;84:683–692.
67. Histologic studies of severe delta agent infection in Venezuelan indians. *Hepatology* 1983;3:906–912.
68. Hadler SC, De Monzon M, Ponzetto A et al. Delta virus infection and severe hepatitis: An epidemic in the Yucpa indians of Venezuela. *Ann Intern Med* 1984;100:339–344.
69. Bensebath G. Hepatite de Labrea (febre negra de Labrea) e outras hepatites fulminantes en Sena Madureira, Acre e Boca do. *Inst Med Trop Sao Paulo* 1983;25:182–194.
70. Buitrago B, Popper H, Hadler SC et al. Specific histologic features of Santa Marta hepatitis: a severe form of hepatitis delta virus infection in northern South America. *Hepatology* 1986; in press.
71. Clain JE, Stephens DH, Charboneau JW. Ultrasonography and computed tomography in focal fatty liver. *Gastroenterology* 1984;87:948–952.
72. Brawer MK, Austin GE, Lewin KJ. Focal fatty change of the liver, a hitherto poorly recognized entity. *Gastroenterology* 1980;78:247–252.
73. Mulhern CB Jr, Arger PH, Coleman BG, Stein GN. Non-uniform attenuation in computed tomography study of the cirrhotic liver. *Radiology* 1979;132:399–402.
74. Yates CK, Streight RA. Focal fatty infiltration of the liver simulating metastatic disease. *Radiology* 1986;159:83–84.
75. Alpern MB, Lawson TL, Foley WD et al. Focal hepatic masses and hepatic infiltration detected by enhanced dynamic CT. *Radiology* 1986;158:45–49.
76. Swobodnik W, Wechsler JG, Manne W, Ditschuneit H. Multiple regular circumscript fatty infiltrations of the liver. *J Clin Ultrasound* 1985;13:577–580.
77. Flournoy JG, Potter JL, Sullivan BM, Gerza BM, Ramzy I. CT appearance of multifocal hepatic steatosis. *J Comput Assist Tomogr* 1984;8:1192–1194.
78. Patel S, Sandler CM, Rauschkolb EN, McConnell BJ. 131 Xe uptake in focal hepatic fat accumulation: CT correlation. *Am J Roentgenol* 1982;138:1192–1194.
79. Gale ME, Gerzof SG, Robbins AH. Portal architecture: A differential guide of fatty infiltration of the liver on computed tomography. *Gastrointest Radiol* 1983;8:231–236.
80. Marchal G, Tshibwabwa-Tumba E, Verbeken E et al. "Skip-areas" in hepatic steatosis: A sonographic-angiographic study. *Gastrointest Radiol* 1986;11:151–157.

2

Surgical Treatment of Portal Hypertension

B. Langer

1. INTRODUCTION

Portal hypertension occurs commonly in patients with liver disease and by itself is of little clinical significance. Only when the major life-threatening complication—bleeding esophageal varices—occurs does it become clinically important. At that point, the attending physician must have a good understanding of the pathophysiology of elevated portal venous pressure, the associated complicating factors in patients with chronic liver disease, the therapeutic choices available (both for the acute and for the long-term management of the bleed), and the results of treatment one can expect so that the best decisions can be made in the individual patient's situation. This chapter will deal mainly with the surgical options but I will also touch on nonsurgical therapy (in as objective a fashion as possible) in order to put surgery into what I consider its proper perspective.

2. PATHOPHYSIOLOGY

Portal hypertension occurs when there is obstruction to flow in the portal venous system. It may result from postsinusoidal disease (hepatic vein thrombosis or veno-occlusive disease), presinusoidal disease (portal vein thrombosis,

B. Langer • Department of Surgery, University of Toronto, Toronto, Ontario, Canada M5G 1L5.

portal triaditis from schistosomiasis), or hepatocellular disease, including hepatic fibrosis or cirrhosis. The most common cause of cirrhosis is either alcohol or viral infection. Although even minor degrees of alcoholic liver injury may increase intrahepatic and thereby portal pressure, it is usually only after considerable hepatocellular injury and fibrosis have occurred that the portal pressure becomes high enough for esophageal varices to form and for bleeding to occur.

Normally, the portal vein provides 70–75% of total hepatic blood flow at low pressure (under 10 mm Hg). The hepatic artery provides the remaining 25–30% of the hepatic blood flow at systemic pressure (120 mm Hg). As progressive fibrosis in the liver occurs, the resistance to flow increases and portal pressure rises. Two other phenomena also occur: decrease in total hepatic blood flow and development of collateral portal systemic communications.

In the early stages of cirrhosis, the decrease in portal inflow may be compensated for by an increased hepatic arterial flow. This compensatory phenomenon varies between individuals and is usually lost late in the disease, so that significant decreases in total hepatic blood flow and loss of hepatic function may be present in advanced cirrhosis.

Collateral circulation around the liver occurs at many sites (retroperitoneal, umbilical, inferior mesenteric-pelvic), but the most important in terms of bleeding complications are collaterals in the submucosa of the gastric fundus and distal esophagus. The specific initiating factor(s) in rupture of these veins is not known, but bleeding appears to be more likely when pressures are very high.

3. GENERAL PATIENT CONSIDERATIONS

The general condition of the patient and the associated liver function is a major influencing factor in management and in predicting outcome following a variceal hemorrhage. Unlike the patient with a bleed from an ulcer who may otherwise be in perfect health, the cirrhotic patient often has a number of complicating conditions that add to the risk from both the acute bleed and any interventional treatments that might be used. These include:

1. Abnormal blood coagulation as a result of impaired hepatic synthetic function. The platelet count may also be low from congestive splenomegaly, but patients uncommonly bleed from thrombocytopenia alone.
2. Poor nutritional state from advanced liver disease associated with a decreased ability to deal with infection and resulting in impaired wound healing.
3. Impaired respiratory function from tense ascites and poor muscle function. These patients are also more susceptible to postanesthetic and postoperative atelectasis and pneumonia.

Table 1. Clinical and Laboratory Classification of Patients with Cirrhosis in Terms of Hepatic Functional Reserve

	A	B	C
Serum bilirubin (μg%)	<2.0	2.0–3.0	>3.0
Serum albumin (g%)	<3.5	3.0–3.5	<3.0
Ascites	None	Easily controlled	Poorly controlled
Neurologic symptoms	None	Minimal	Advanced
Nutritional state	Excellent	Good	Poor

4. Impaired renal function. Advanced liver disease can be associated with sodium retention and oliguria, with renal failure occurring despite the kidney being anatomically normal, termed "hepatorenal failure." The latter may be related to decreased renal blood flow especially to the cortex, but there is also increased susceptibility to acute tubular necrosis.
5. Hepatic encephalopathy resulting from decreased hepatic function and spontaneous portasystemic shunting, and aggravated by a gastrointestinal bleed that may precipitate coma.

The degree of impairment of liver function has been shown to influence both the risk of dying from any given bleeding episode and the risk of treatment interventions, particularly surgical procedures. The Child classification[1] or one of its modifications, is widely used to stage such patients when selecting their treatment.

4. CLINICAL TRIALS VERSUS CLINICAL EXPERIENCE VERSUS CLINICAL BIAS

Hardly any field of medicine has available more treatment options, more opinions and biases, and more controversy than the management of bleeding varices. Fortunately, the need to carry out rigorous clinical trials in this disease was recognized over 20 years ago. In the intervening years, this ultimate tool has been widely applied so that many of the proposals for treatment have been subjected to rigorous scrutiny.

In reading the literature it is important to keep the following in mind:

1. The treatment considerations are different in different groups of clinical situations, such as (a) the patient with proven esophageal varices who has never bled (prophylactic therapy), (b) the patient who is in the middle of a significant variceal hemorrhage (emergency therapy), and (c) the patient who has survived a bleed and is currently stable (elec-

tive). This last classification may be further subdivided into (i) urgent—immediately postbleed, and stable, (ii) true elective fully recovered from bleeding, with regained optimum prehemorrhage liver function and general status.

2. Liver status (Child classification plus or minus another classification) and general status have an important bearing on outcome. In comparing treatment options, stratification for severity of liver disease is essential.
3. Etiology of the liver disease is important. There is some evidence that alcoholic liver disease carries a worse prognosis than nonalcoholic cirrhosis following certain types of shunt.[2] Therefore, comparing populations of patients that differ as to etiology may produce bias in analyzing the results of different treatments.
4. General supportive care, especially the development of ICU care for patients with bleeding varices, has improved over the last two decades, thus creating a significant bias if one compares the results of one treatment in the 1980s with another in the 1960s.

5. SURGICAL TREATMENT OPTIONS

Operative approaches to management of portal hypertension and bleeding varices can be classified into two broad groups:

1. Portasystemic shunting, which is aimed at diverting portal blood around the critical esophagogastric area which bleeds, and
2. Devascularization procedures, which include either directly or indirectly interrupting the blood flow to the region which bleeds, but without shunting

All of these procedures involve the general risks of operating on patients with compromised liver function in addition to certain specific risks related to each operation.

6. END-TO-SIDE PORTACAVAL SHUNT

This operation was the earliest procedure aimed at dealing with the problem of portal hypertension. Based on experimental work done by Pavlov and Eck in the late 1800s, it involves ligation of the hepatic end of the portal vein and anastomosis of the distal end to the inferior vena cava. Following the work of Whipple and others in the 1940s,[3,4] it became the standard operation in the 1950s and 1960s. The operation was used not only in the emergency management of bleeding and as elective therapy for patients whose bleeding had stopped, but because of its obvious lifesaving effect in some patients who would other-

wise have exsanguinated, end-to-side portacaval shunt came into use as prophylactic therapy in cirrhotic patients with varices who had never bled.[5]

At about this time, a number of centers began to evaluate the portacaval shunt prospectively. Randomized controlled trials were set up to examine its effectiveness in both elective therapeutic and prophylactic settings. Regretfully, no randomized trials comparing portacaval shunt to standard medical therapy in the emergency situation were done then or since.

The results of these trials are now well known. In the elective therapeutic studies, portacaval shunt proved to be an effective method of preventing recurrent variceal bleeding, but all three North American trials failed to show a statistically significant improvement in survival following portacaval shunting, although each trial showed a trend toward an advantage to shunt surgery.[6–8] Combining the data from all three trials just reached statistical significance. If one includes the French trial,[9] the combined data do not show an advantage for shunting. It should be noted, however, that in each trial shunting remained a bailout option for patients who failed medical therapy by rebleeding.

The trials of prophylactic shunting were uniformly disappointing.[10,11] Again, portacaval shunting was effective in reducing the incidence of subsequent variceal bleeding, but there was no significant difference in survival between the shunted and nonshunted patients; the leading cause of death in the nonshunted patients, bleeding, was replaced by hepatic failure in the shunted patients. Shunted patients also had a higher incidence of portasystemic encephalopathy than nonshunted patients, presumably as a result of total diversion of portal venous perfusion. This change in portal circulation was also felt to be responsible for the higher incidence of late hepatic failure.

The conclusions drawn from these prophylactic and therapeutic trials were as follows: (1) portacaval shunting is effective in lowering portal pressure and reducing the incidence of subsequent variceal hemorrhage; (2) portacaval shunting is associated with increased incidence of postshunt encephalopathy and late hepatic failure; and (3) further improvement in surgical therapy would have to be brought about by either (a) identifying factors that predisposed to postshunt encephalopathy and excluding those patients from that treatment, or (b) developing other operations that are not associated with the same problem of postshunt encephalopathy.

Because of its ease of construction in most patients, the end-to-side portacaval shunt has been widely used in the emergency control of variceal bleeding. In a sequential study, Orloff in 1967[12] showed a marked reduction in mortality rate when using this operation in all bleeding patients rather than persisting with conservative therapy. Our own experience[13] suggests that many patients can be controlled during the acute bleed without major surgical intervention. In the course of this, their general and hepatic status improves so that should they require surgery later, the operative risk would be less. When acute bleeding could not be controlled, we have used this shunt as a lifesaving mea-

sure, as have many groups including some doing prospective trials of other therapies.[14,15]

7. SIDE-TO-SIDE PORTASYSTEMIC SHUNTS

A number of other operations were developed that did not involve a complete interruption of the main portal vein. Some were developed because of technical ease of execution, others because of the belief that leaving the portal vein patent would allow continued portal perfusion, while at the same time permitting shunting of blood to the systemic circulation and lowering the portal pressure. Three of these operations are worth mentioning: side-to-side portacaval shunt, proximal splenorenal shunt, and interposition mesocaval shunt.

(1) *Side-to-side portacaval shunt* is usually a more difficult technical procedure than end-to-side portacaval shunt. It has a theoretical advantage in patients with troublesome ascites because it allows for reversal of blood flow in the portal vein and hepatic sinusoidal decompression. Whether this portal perfusion, which is achieved by hepatic areterial inflow exiting via the portal vein in the cirrhotic liver, provides effective hepatocellular perfusion is as yet unproven. The studies of Orloff and others have, however, convincingly shown that this operation is very effective in the treatment of both cirrhotic ascites and the Budd–Chiari syndrome.[16–18]

(2) The *proximal splenorenal shunt* was preferred by authors such as Linton[19] because it allowed splenectomy and relief of thrombocytopenia as well as portal decompression. No controlled trials of this operation have been carried out. It is now rarely done. The reasons for the good survival and low incidence of portasystemic encephalopathy in previous reports may relate more to patient selection and a higher incidence of shunt occlusion than to any physiological advantage of the shunt itself.

(3) The *interposition mesocaval shunt* was first described by Clatworthy *et al.*[20] in infants with portal vein thrombosis. Originally, it involved ligation and division of the vena cava in order to reach the superior mesentery vein. This was modified by Lord[21] using an interposition graft. Drapanas[22] reported enthusiastically on this procedure in cirrhotic patients with bleeding varices. In an uncontrolled study, he observed a low incidence of encephalopathy and good survival, and concluded that the design of the shunt permitted both continued portal perfusion of the liver plus portal venous decompression. These observations have not been borne out by other studies. Fulenwider *et al.*[23] showed how postshunt angiograms can be misinterpreted, and described the phenomenon of "pseudoperfusion" in which portal vein filling could be obtained by hepatic artery injection after mesocaval shunting and represented hepatofugal rather than hepatopedal flow. In our institution,[24] careful postoperative angiographic studies in mesocaval shunt patients have shown that in the vast majority of

patients portal perfusion is lost following this operation, thus supporting Warren's contention that the mesocaval shunt totally diverts portal flow. Furthermore, comparison of the incidence of postshunt encephalopathy, rebleeding rates, and long-term survival in nonrandomized but comparable groups of patients undergoing either portacaval or mesocaval shunts has failed to show any significant difference in any of these outcome measurements.[24] A prospective, randomized controlled trial of the therapeutic interposition mesocaval shunt using the saphenous vein was carried out by Stipa *et al.*[25] This trial too failed to show any significant difference in either encephalopathy or survival.

A very small trial of the mesocaval shunt compared to portacaval shunt carried out in the emergency situation showed a slight advantage to portacaval shunt, but the operative mortality reported was higher than in most other reported series and may not be representative.[26] Our own experience with emergency mesocaval shunting[27] in uncontrolled studies shows no difference in operative mortality, rebleeding, or encephalopathy rates.

A number of authors have suggested that ascites is less frequent after any type of side-to-side shunting than end-to-side portacaval shunt. Mesocaval shunt has therefore also been used as the preferred method of total shunting in patients with massive ascites associated with bleeding varices. Not all reports support this contention.

The *Dadron mesocaval shunt* has been reported to have a higher occlusion rate than the portacaval shunt, up to 30% at 2 years.[28–30] Not all authors report this high occlusion rate, and there are no good data to suggest that it has a significant effect on either late rebleeding or survival. It has been suggested by Rypins *et al.*[31] that deliberately using a narrow diameter mesocaval shunt results in less encephalopathy, but these studies are uncontrolled and have been disputed by other authors.

In summary, there does not appear to be any major advantage of mesocaval over portacaval shunt. Both function as total shunts and are associated with the same problems of encephalopathy and late hepatic failure because of loss of hepatic portal perfusion. Where a total shunt is indicated, the choice would seem to rest on the individual surgeon's experience, local technical considerations, and possibly the presence of severe ascites or demonstrated preoperative hepatofugal flow in the portal vein.

8. SELECTIVE SHUNTS

Two operations have been devised to provide portasystemic shunting of blood from the gastroesophageal region where bleeding occurs while at the same time maintaining pressure and normal flow in the main portal system to permit continued portal perfusion of the liver. These are the Inokuchi shunt anastomosing the distal end of the left gastric vein to the cava[32] and the Warren

shunt anastomosing the distal end of the splenic vein to the renal vein. The former has been little used outside of Japan, whereas the Warren shunt has been widely used and carefully studied.

The *Warren procedure* (or distal splenorenal shunt) involves careful dissection of the retropancreatic splenic vein, ligation of its proximal end, and anastomosis of its distal end to the left renal vein. This allows portasystemic shunting from the spleen and upper gastric and esophageal region. Careful interruption of communications between this left side of the portal system and the main system by ligating the gastroepiploic, left gastric, umbilical, and pancreatic veins helps maintain high pressure in the main portal system and portal perfusion. Both angiographic and functional studies have shown that this operation can be carried out with a high degree of technical success and that it is effective in protecting individuals against subsequent variceal hemorrhage.[34–36] The operation is more difficult to do and therefore has a longer learning curve and higher start-up cost in terms of complications and operative mortality. In experienced hands, however, the complication rate and operative mortality and the incidence of postshunt encephalopathy are significantly lower after distal splenorenal shunt than total shunts. This is associated with improvement in the quality of life as measured by functional status in postshunt patients as well as frequency of readmission to hospital.[34] Controlled trials of this operation, however, have not yet demonstrated a significant improvement in long-term survival compared to that from total shunts. The data from the Atlanta group, however, stratified for alcoholic and nonalcoholic patients has shown that survival of nonalcoholic patients treated by the distal shunt is not only better than those patients treated by total shunt, but is also better than alcoholic patients treated by distal splenorenal shunt.[35]

Unique complications of this operation are as follows:

1. Portal vein thrombosis, which may be partial or complete, occurs in about 10% of such patients and is occasionally the cause of postoperative death.[37,38] Usual manifestations are ascites with or without abdominal pain.
2. Ascites may also occur as a result of division of retroperitoneal lymphatics with leakage of lymph into the peritoneal cavity. The fact that venous hypertension continues in the portal mesenteric system also favors the formation of ascitic fluid rather than its absorption, as might occur with total shunting.
3. Reformation of collateral circulation with the development of late encephalopathy has been described.[39,40] The usual sites are in the retroperitoneum, particularly around the pancreatic veins. Warren and colleagues have described both angiographic and surgical techniques for interruption of these collaterals with reversal of encephalopathy,[41] and

currently recommend a modification of the distal splenorenal shunt that includes division of pancreatic collateral veins at the time of the original operation.[42]

9. OTHER SHUNTS

A large number of other shunt operations have been described, limited only by the imagination of the surgeon and availability of vessels. None of these have any specific benefits over those described above. Attempts to counteract the deleterious effects of total shunting by arterialization of the remaining intrahepatic portal system have had occasional enthusiastic supporters[43,44] but have not met with sufficient success to be widely applied.

Studies of either hepatic function, hepatic perfusion, or other liver or patient factors have failed to find any reliable predictors of postshunt encephalopathy, so that a selection process might be applied to shunt only those patients whose compensatory mechanisms will allow them to tolerate portal diversion without adverse effects.

10. DEVASCULARIZATION PROCEDURES

The general term "devascularization procedure" is applied to operations that attempt to control variceal bleeding by either a direct surgical obliteration of the distal esophageal veins or an indirect reduction of blood flow to the area by techniques other than creating portasystemic anastomoses.

Transesophageal varix ligation was first described by Boerama in 1949[45] and Crile in 1950.[46] This operation is carried out through a short thoracotomy, and involves opening the esophagus longitudinally and oversewing the esophageal veins under direct vision to induce their obliteration. It is a straightforward, simple procedure that can be carried out by most surgeons and does not require the special technical skills needed to perform shunt surgery. For this reason it was widely used in the 1940s and 1950s in the emergency situation, and some, like Linton,[19] considered it a first step in a two-step approach to the control of variceal bleeding, the definitive operation being a portasystemic shunt. This approach indicates what is recognized as a major problem with transesophageal ligation. Although it is effective in stopping acute bleeding, there is a high incidence of reformation of varices and rebleeding. This was well shown by Marshall Orloff in his sequential studies of transesophageal ligation and portacaval shunt in the emergency treatment of bleeding varices.[12] Taking the approach of using surgery as the primary treatment of virtually all comers who presented with variceal hemorrhage, Orloff reported no significant difference in operative mortality and early survival between portacaval shunt and transeso-

phageal ligation. However, transesophageal ligation had a much higher rebleeding rate and poorer long-term survival. Most surgeons now no longer consider transesophageal ligation a useful procedure since its operative risk is equivalent to emergency shunting and its rebleeding rate is high. In addition, newer techniques, such as injection sclerotherapy, offer similar benefits at much less risk.

11. ESOPHAGEAL AND GASTRIC TRANSECTION

Esophageal transection is an extension of the principle of transesophageal ligation but guarantees that all submucosal veins are obliterated at least at the site of transection. This is an old operation, championed in 1964 by Milnes Walker.[47] It can be carried out through a transthoracic or transabdominal approach. Originally, it carried a high mortality, but the technique was modified in the 1970s using the end-to-end stapler. Current advocates of the operation, including Johnston[48] and Wexler,[49] use the procedure as an alternative to portasystemic shunting. They emphasize its advantage of producing less portasystemic encephalopathy than total shunts and of being performable by surgeons without special training in shunt surgery. However, the complication of esophageal stenosis and leak is higher than simple transesophageal ligation, and the available data, though uncontrolled, suggest that the operative mortality of esophageal transection is at least as high as that of portacaval shunt,[50] the rebleeding rate is much higher, and there is no good evidence that long-term survival is any better.

Gastric transection was proposed by Norman Tanner in the 1950s[51] as an alternative to shunting and involved an abdominal approach in which the stomach was transected and reanastomosed just below the cardia. A more difficult operation than esophageal transection, it appears to have no advantage in terms of operative risk or outcome, and currently it is not recommended.

12. SPLENECTOMY

Splenectomy is the operation of choice in patients with portal hypertension due to splenic vein thrombosis with a patent portal vein. In such cases where there is regional portal hypertension only (also called "left-sided portal hypertension"), removal of the spleen lowers blood flow to the esophagogastric venous system sufficiently to prevent rebleeding. Splenectomy is also the treatment of choice in cases of primary portal hypertension due to primary splenic abnormalities resulting in high portal inflow through the spleen and the resulting high portal pressure (Banti's syndrome). This diagnosis is uncommon and is difficult to make since there may be associated histological abnormalities in the liver secondary to the high inflow portal hypertension.

Removing the spleen reverses the hypersplenism commonly seen in patients with portal hypertension from any cause. It is virtually never necessary, however, to carry out splenectomy to treat hypersplenism by itself. Splenectomy has been advocated by the Wittes[52] for the treatment of portal hypertension and bleeding in some cirrhotics. Although splenectomy does reduce blood flow to the entire portal system, it does not adequately lower portal pressure to provide long-term protection from variceal bleeding.

13. EXTENSIVE DEVASCULARIZATION

Because of the shortcomings of esophageal transection and splenectomy by themselves, attempts have been made to more extensively devascularize the gastroesophageal region. Hassab[53] used a combination of splenectomy and gastric devascularization to treat a population of patients who have mainly presinusoidal portal hypertension due to schistosomiasis and who are particularly susceptible to postshunt encephalopathy after total shunts. Because liver function in these patients is usually good, the operative mortality is low, but there is a significant incidence of late rebleeding.

Another method of extensive devascularization has been proposed by Sugiura.[54] The *Sugiura operation* includes esophageal transection, splenectomy, and gastric devascularization. It can be carried out in two stages—(1) transthoracic esophageal transection, and (2) abdominal splenectomy and gastric devascularization—or, alternatively, in one stage using a thoracoabdominal approach. Sugiura has also included vagotomy as part of the operation and emphasized the importance of leaving the paraesophageal veins intact to prevent reformation of varices and bleeding. The value of these latter two components of the operation is as yet unproven.

Sugiura and Futagawa[55] in uncontrolled studies in a Japanese population of 671 patients, mainly with postnecrotic cirrhosis, reported an operative mortality of 4.3%, lower than the 16.5% mortality in his portasystemic shunt experience done in a similar population. Postoperative encephalopathy rates were also extremely low. It must be noted, however, that these patients were largely a very good risk group of patients (36% Child A and 37% Child B) and that 30% of the operations were carried out as prophylactic therapy in patients who had never bled from their varices.

The role of the Sugiura operation in the predominantly alcoholic cirrhotic Western population of portal hypertension remains to be determined. Ginsberg and colleagues,[56] using a one-stage thoracoabdominal operation, found a low operative mortality in good risk patients when done electively, but 100% mortality in Child C patients operated on as an emergency. Gouge and Ranson[57] reported 9.5% operative mortality in the complete operation in good-risk patients, but 53% mortality in urgent, emergency, and high-risk patients, which

was higher than their previous experience with shunt procedures. In addition, they had a 37% late rebleeding rate. Our own experience is limited to its use in noncirrhotic, nonshuntable patients with portal vein thrombosis in whom early results are good.

14. SELECTION OF SURGICAL PROCEDURE

If there is a role for major surgical procedures in the treatment of portal hypertension, which operative procedure should be used? The answer to that question is in part based on controlled trials, but must also reflect experience and bias since not all the necessary trials have been carried out to date.

(1) *Prophylactic surgery.* controlled trials of the portacaval shunt have shown convincingly that this operation has no place in cirrhotics with varices who have never bled. Another recent report of a prospective trial of surgery versus medical therapy[58] showed no difference in survival but claimed improved quality of life in operated patients. The study, however, is flawed by mixing a variety of procedures in the operated group and ignoring sclerotherapy. There are controlled trials of other prophylactic therapy which have shown a decrease in the expected rate of variceal bleeding after prophylactic sclerotherapy and a suggested decrease in the late mortality resulting from this intervention.[59] At present, however, we must conclude that major surgery has no place in prophylactic therapy.

(2) *Emergency surgery.* In patients with active bleeding, nonsurgical therapy has improved in recent years with the addition of injection sclerotherapy, so that most acute episodes of bleeding can be controlled by methods short of major surgery. If the acute bleeding episode can be controlled in this way, patients can often achieve improved liver function and general condition so that subsequent surgery, if necessary, can be carried out at a lower risk. The approach still advocated by Orloff, emergency shunt surgery for all bleeding patients, has little to recommend it. If emergency surgery is to be used for patients who fail to stop bleeding despite excellent conservative treatment including sclerotherapy, which operation is to be preferred? The uncontrolled data available suggest that the operative mortality of any major operation in acutely bleeding patients, most of whom are alcoholics classified as Child C, is 40 to 50% and sometimes higher. The Sugiura procedure and the Warren operation will likely have a higher mortality in most surgeons' hands because of the complexity and length of the procedure. Esophageal transection alone is a fairly simple procedure which most surgeons can do, but its operative mortality rate is similar to total shunting, and the rebleeding rate much higher. Our preference, therefore, is for a total shunt in this situation, which can be carried out quickly and with a mortality rate of 30% to 50%, a low rebleeding rate, and acceptable long-term survival.

(3) *Elective surgery.* Trials of elective portacaval shunt in the 1960s showed only a slight advantage of portacaval shunt over continued medical therapy in patients who had survived a variceal bleed. There have been several trials of the distal splenorenal shunt compared to total shunt,[34,35,60,61] which have confirmed the superiority of distal splenorenal shunt over total shunt in terms of postshunt encephalopathy and quality of life, but have as yet failed to show a survival advantage, except for the small group of nonalcoholic patients. The current interest in injection sclerotherapy has meant that one can no longer carry out trials of shunt surgery in comparison with no treatment but rather with the best nonsurgical treatment, which currently includes sclerotherapy. The Atlanta group has recently reported such a trial[62] and concluded that overall sclerotherapy provides better survival than distal splenorenal shunt, providing that shunt surgery is used liberally as an alternative for patients who rebleed in the sclerotherapy arm (30% in the Atlanta trial). Furthermore, distal splenorenal shunt may be the preferred treatment in nonalcoholics and in those alcoholics who have life-threatening first bleeds and in whom there are other considerations that make long-term sclerotherapy impractical.

Other trials comparing sclerotherapy to shunt surgery are in progress. Cello *et al.*[63] compared portacaval shunts and sclerotherapy as emergency treatment and showed no survival difference, but did show an advantage for sclerotherapy in terms of total health care costs. Sclerotherapy, however, is here to stay, and it is likely to remain the primary method of management of most patients with variceal bleeding, both acute and long-term, simply because of its relative safety, wide availability, and what is already known about its effectiveness. Surgery, especially shunt surgery, still has a limited but important role to play in the emergency situation for the management of persistent life-threatening bleeding and as an alternative for sclerotherapy failures, as well as the primary treatment for certain patients such as nonalcoholics and those in whom long-term follow-up may be difficult.

REFERENCES

1. Child CG. *The Liver and Portal Hypertension.* Philadelphia: W.B. Saunders, 1964.
2. Henderson MJ, Milikan WJ, Wright-Bacon L et al. Hemodynamic differences between alcoholic and nonalcoholic cirrhotics following distal splenorenal shunt-effect on survival? *Ann Surg* 1983;325–334.
3. Whipple AO. The problem of portal hypertension in relation to the hepatosplenopathics. *Ann Surg* 1945;122:449–475.
4. Blakemore AH, Lord JW Jr. Technique of using vitallium tubes in establishing portacaval shunts for portal hypertension. *Ann Surg* 1945;122:476–489.
5. Palmer ED, Brick IB, Jahnke EJ. Esophageal varices without hemorrhage in cirrhosis: proper indication for shunting procedures. *N Engl J Med* 250:863–865.
6. Resnick RH, Iber FL, Ishihara AM et al. A controlled study of the therapeutic portacaval shunt. *Gastroenterology* 1974;67:843–857.

7. Jackson FC, Perrin EB, Felix WR et al. A clinical investigation of the portacaval shunt versus surgical analysis of the therapeutic operation. *Ann Surg* 1971;174:672–701.
8. Reynolds TB, Donovan AJ, Mikkelsen WP et al. Results of a 12-year randomized trial of portacaval shunt in patients with alcoholic liver disease and bleeding varices. *Gastroenterology* 1981;80:1005–1011.
9. Rueff B, Degos F, Dego JD et al. A controlled study of therapeutic protacaval shunt in alcoholic cirrhosis. *Lancet* 1976;1:655–659.
10. Resnick RH, Chalmers TC, Ishihara AM et al. A controlled study of the prophylactic portacaval shunt. *Ann Intern Med* 1969;70:675–688.
11. Conn HO, Lindenmuth WW. Prophylactic portacaval anastomosis in cirrhotic patients with esophageal varices. *N Engl J Med* 1968;279:725–732.
12. Orloff MJ. Emergency portacaval shunt: A comparative study of shunt, varix ligation and nonsurgical treatment of bleeding esophageal varices in unselected patients with cirrhosis. *Ann Surg* 1967;166:456–478.
13. Langer B, Patel SC, Stone RM et al. Selection of operation in patients with bleeding esophageal varices. *Can Med Assoc J* 1978;118:369–372.
14. Korula J, Balart LA, Radvan G et al. A prospective randomized controlled trial of chronic esophageal variceal sclerotherapy. *Hepatology* 1985;5:584–589.
15. Conn HO, Grace ND. Portal hypertension and sclerotherapy of esophageal varices: a point of view. *Endoscopy Rev* 1985; May/June, 39–53.
16. Orloff MJ, Chandler JG, Charters AC et al. Comparison of end-to-side and side-to-side portacaval shunts in dogs and human subjects with cirrhosis and portal hypertension. *Am J. Surg* 1974;128:195–201.
17. Burchell AR, Rousselot LM, Panke WF. A seven-year experience with side-to-side portacaval shunt for cirrhotic ascites. *Ann Surg* 1968;168:655–670.
18. Orloff MJ, Johansen KH. Treatment of Budd–Chiari syndrome by side-to-side portacaval shunt: Experimental and clinical results. *Ann Surg* 1978;188:494–512.
19. Linton RR. The surgical treatment of bleeding esophageal varices. *Western J Surg* 1955;63:366–377.
20. Clatworthy HW, Wall T, Waltman RN. A new type of porta-systemic venous shunt for portal hypertension. *Arch Surg* 1955;71:588–596.
21. Lord JW, Rossi G, Daliana M et al. Mesocaval shunt modified by the use of a teflon prosthesis. *Surg Gynecol Obstet* 1970;130:525–526.
22. Drapanas T. Interposition mesocaval shunt for treatment of portal hypertension. *Ann Surg* 1972;176:435–448.
23. Fulenwider JT, Nordlinger BM, Millikan WJ et al. Portal pseudo-perfusion, an angiographic illusion. *Ann Surg* 1979;189:257–268.
24. Reznick RK, Langer B, Taylor BR et al. Results and hemodynamic changes after interposition. *Mesocaval Shunt Surg* 1984;95:275–279.
25. Stipa S, Zipara V, Anza M et al. A randomized controlled trial of mesentericocaval shunt with autologous jugular vein. *Surg Gynecol Obstet* 1981;153:353–356.
26. Malt RA, Abbott WM, Warshaw AL et al. Randomized trial of emergency mesocaval and portacaval shunts for bleeding esophageal varices. *Am J Surg* 1978;135:584–588.
27. Soutter DI, Langer B, Taylor B et al. Results of emergency portasystemic shunting in cirrhotics with bleeding varices. *Clin Invest Med* 1986;9:A50.
28. Smith RB, Warren WD, Salam AA et al. Dacron interposition shunts for portal hypertension. *Ann Surg* 1980;192:9–17.
29. Fletcher MS, Dawson JL, Williams R. Long-term follow-up of interposition mesocaval shunting in portal hypertension. *Br J Surg* 1981;68:485–487.
30. Dennis MA Jr, Monson RC, O'Leary JP. Interposition mesocaval shunt: A less than ideal procedure. *Am J Surg* 1978; 44:734–738.

31. Rypins EB, Mason GR, Conroy RM et al. Predictability and maintenance of portal flow patterns after small-diameter portacaval H-grafts in man. *Ann Surg* 1984;200:706–710.
32. Inokuchi K, Kobayashi M, Kusaba A et al. New selective decompression of esophageal varices. *Arch Surg* 1970;100:157–162.
33. Warren WD, Zeppa R, Fomon JJ. Selective transplenic decompression of gastroesophageal varices by distal splenorenal shunt. *Ann Surg* 1967;166:437.
34. Langer B, Taylor BR, MacKenzie DR et al. Further report of a prospective randomized trial comparing distal splenorenal shunt with end-to-side portacaval shunt. *Gastroenterology* 1985;88:424–429.
35. Millikan WJ, Warren WD, Henderson JM et al. The Emory prospective randomized trial: Selective versus nonselective shunt to control variceal bleeding. *Ann Surg* 1985;201:712–722.
36. Hutson DG, Pereiras R, Zeppa R et al. The fate of esophageal varices following selective distal splenorenal shunt. *Ann Surg* 1976;183:496–501.
37. Rotstein LE, Makowka L, Langer B et al. Thrombosis of the portal vein following distal splenorenal shunt. *Surg Gynecol Obstet* 1979;149:847–851.
38. Henderson JM, Millikan WJ, Chipponi J et al. The incidence and natural history of thrombus in the portal vein following distal splenorenal shunt. *Ann Surg* 1982;196:1–7.
39. Tylen U, Simert G, Vang J. Hemodynamic changes after distal splenorenal shunt studies by sequential angiography. *Diag Radiol* 1976;585–589.
40. Maillard JN, Flamant YM, Hay JM et al. Selectivity of the distal splenorenal shunt. *Surgery* 1979;86:663–671.
41. Hannah SS, Smith RS, Henderson JM et al. Reversal of hepatic encephalopathy after occlusion of portasystemic shunts. *Am J Surg* 1981;142:285–289.
42. Warren WD, Millikan WJ, Henderson JM et al. Selective variceal decompression after splenectomy of splenic vein thrombosis. *Ann Surg* 1983;694–702.
43. Maillard JN, Benhamou JP, Rueff B. Arterialization of the liver with portacaval shunt in the treatment of portal hypertension due to intrahepatic block. *Surgery* 1970;67:883–890.
44. Adamsons RJ, Butt K, Iyer S et al. Portacaval shunt with arterialization of the portal vein by means of a low flow arteriovenous fistula. *Surg Gynecol Obstet* 1978;146:869–876.
45. Boerama I. Bleeding varices of the esophagus in cirrhosis of the liver and Banti's syndrome. *Arch Chir Neerl* 1949;3:1.
46. Crile G Jr. Transesophageal ligation of bleeding esophageal varices: A preliminary report of 7 cases. *Arch Surg* 1950;61:654.
47. Walker RM. Esophageal transection for bleeding varices. *Surg Gynecol Obstet* 1964;323–329.
48. Johnston GW. Treatment of bleeding varices by oesophageal transection with the SPTU gun. *Ann RCPS Engl* 1977;59:404–408.
49. Wexler MJ. Treatment of bleeding esophageal varices by transabdominal esophageal transection with the EEA stapling instrument. *Surgery* 1980;88:406–416.
50. Durtschi MB, Carico CJ, Johansen KH. Esophageal transection fails to salvage high-risk cirrhotic patients with variceal bleeding. *Ann J Surg* 1985;150:18–23.
51. Tanner NC. The late results of porta-azygos disconnection in the treatment of bleed from oesophageal varices. *Ann RCPS Engl* 1961;29:153–174.
52. Witte CL, Witte MH, Krone CL. Contrasting hemodynamic patterns of portal hypertension. *Ann Surg* 1971;176:68–79.
53. Hassab MA. Nonshunt operations in portal hypertension without cirrhosis, Hassab. *Gynecol Obstet Surg* 1970;648–654.
54. Sugiura M, Futagawa S. A new technique for treating esophageal varices. *J Thorac Card Surg* 1973;66:677–685.
55. Sugiura M, Futagawa S. Esophageal transection with paraesophagogastric devascularizations (the Sugiura procedure) in the treatment of esophageal varices. *World J Surg* 1984;8:673–682.

56. Ginsberg RJ, Waters PF, Zeldin RA et al. A modified Sugiura procedure. *Ann Thorac Surg* 1982;34:258–264.
57. Gouge TH, Ranson JHC. Esophageal transection and paraesophageal devascularization for bleeding esophageal varices. *Ann J Surg* 1986;151:47–54.
58. Inokuchi K, Cooperative Study Group of Portal Hypertension of Japan. Prophylactic portal nondecompression surgery in patients with esophageal varices: Interim study. *Ann J Surg* 1984;200:61–65.
59. Paquet KJ. Prophylactic endoscopic sclerosing treatment of the esophageal wall in varices: A prospective controlled randomized trial. *Endoscopy* 1982; 13:4–5.
60. Reichle RA, Fahmy WF, Golsorkhi M. Prospective comparative clinical trial with distal splenorenal and mesocaval shunts. *Am J Surg* 1979;137:13–21.
61. Fischer JE, Bower RH, Atamian S et al. Comparison of distal and proximal splenorenal shunts: A randomized prospective trial. *Ann J Surg* 1981;194:531–544.
62. Warren WD, Henderson JM, Millikan WJ et al. Distal splenorenal shunt versus endoscopic sclerotherapy for long-term management of variceal bleeding: Preliminary report of a prospective randomized trail. *Ann Surg* 1986;203(5):454–462.
63. Cello JP, Grendell JH, Crass RA et al. Endoscopic sclerotherapy versus portacaval shunt in patients with severe cirrhosis and variceal hemorrhage. *N Engl J Med* 1984;311:982–989.

3

Update on Bile Formation and on Mechanism of Bile Acid-Induced Cholestasis

Claude C. Roy, Beatriz Tuchweber, Andrée M. Weber, and Ibrahim M. Yousef

1. INTRODUCTION

In its broadest sense cholestasis may be defined as a decrease or failure of bile formation and/or secretion. This results in the hepatic and systemic accumulation of certain components that are normally secreted into bile, such as bilirubin, bile acids, cholesterols, and liver enzymes.[1–4] Although bile flow is affected when the biliary tree is obstructed, the term cholestasis is increasingly reserved for those conditions in which anatomic and functional integrity of the biliary tree and gallbladder is preserved.

Normal bile is a complex mixture of bile acids, unesterified cholesterol, phospholipids (mainly lecithin), bile pigments, vitamins of the lipid-soluble (vitamin D_2) and water-soluble (vitamin B_{12}, folic acid, pyridoxine) variety, steroid hormones, inorganic ions, and a variety of proteins (enzymes, transport proteins, immunoglobulins, and hormones).[5] The exact metabolic role played

Claude C. Roy and Andrée M. Weber • Department of Pediatrics, Hôpital Sainte-Justine, and University of Montreal, Montreal, Quebec, Canada H3T 1C5. *Beatriz Tuchweber* • Pediatric Research Center, Hôpital Sainte-Justine, and Department of Nutrition, University of Montreal, Montreal, Quebec, Canada H3T 1C5. *Ibrahim M. Yousef* • Pediatric Research Center, Hôpital Sainte-Justine, and Department of Pharmacology, University of Montreal, Montreal, Quebec, Canada H3T 1C5.

by some of these components remains to be elucidated. However, it is obvious that bile secretion is not only a means of transfer and removal of a number of unwanted molecules, but also conservation of others: note the highly efficient enterohepatic circulation of bile salt molecules that leave the biliary tree, traverse the gut, and return via the portal circulation to the liver. The recycling of bile salts ensures that these biological detergents perform a number of important solubilization, transport, and regulatory functions.[6] Interruption of this continuous flow of bile salt molecules through a decrease in the formation and/or secretion of bile has far-reaching effects.

This chapter will provide an overview of current understanding of bile formation in the light of information generated during the past decade by research carried out in our unit on bile acid-induced cholestasis.

2. BILE FORMATION

2.1. Bile Acid Secretion and Bile Flow

Bile acids are synthesized in the liver, conjugated, and secreted into bile before reaching the small intestine through the biliary tree. Reabsorption of bile acids and of their bacterially modified forms occurs in the jejunum, but predominantly in the ileum and, to a limited extent, in the colon. After passage across enterocytes, bile acids return to the liver via the portal circulation. Since >95% of bile acids remains in the enterohepatic circulation with each cycle, about 100 mmoles is removed by the human liver. These molecules are the major determinant (or stimulus) of bile production.[7] From experiments performed in bile fistula dogs 100 years ago, it was recognized that bile contains various choleretic agents.

Under normal circumstances, bile acids are the principal organic component of bile. Sperber[8] originally proposed that the active transport of bile acids across the hepatocyte establishes an osmotic gradient that promotes the flow of water into the canaliculus. Although the osmotic activity of bile acids may be reduced by self-aggregating to form micelles, accompanying ions, especially the counterion Na^+, promote bile flow.[9] Infusion of bile acids promotes bile flow. There is a linear relationship between the amount of bile salt and the rate of bile flow, thereby suggesting that bile salts promote canalicular bile flow. Recently, doubt was cast on the reliability of erythritol secretion as a measure of canalicular bile flow, but in most animal species the relationship between erythritol clearance and bile salt secretion rate is linear.[5,7]

When the line relating erythritol clearance and bile flow is extrapolated to the ordinate from the lowest rates of secretion obtained, the intercept defines the canalicular bile flow that would be present in the theoretical situation of

zero bile acid secretion. This flow is termed "canalicular bile acid-independent flow" (BAIF). The size of this component varies between animal species from 2 to 70 μl/min/kg body weight.[5,7] Thus canalicular bile flow is influenced in part by bile acids, bile acid-dependent flow (BADF), as well as by the BAIF. However, both components of canalicular bile flow are closely interrelated. The uptake of bile acids is dependent on Na^+ cotransport by the hepatocyte secondary to the sodium gradient generated by the "sodium pump" Na^+, K^+-ATPase. This transport membrane protein actively pumps Na^+ out of the cell to maintain the high extracellular Na^+ and the low intracellular Na^+ concentrations.[5] Canalicular secretion of other ions or even organic compounds may contribute to the BAIF, which has also been attributed to the activity of Na^+, K^+-ATPase.

2.2. Lobular Uptake of Bile Acids (Fig. 1)

Bile acids reach the liver predominantly via the portal vein and to a much lesser extent via the hepatic artery. The hepatocytes are arranged in sheets one cell thick between the venous and arterial supply (afferent supply) and the venous drainage (efferent supply). The cells close to the portal vein radicals are

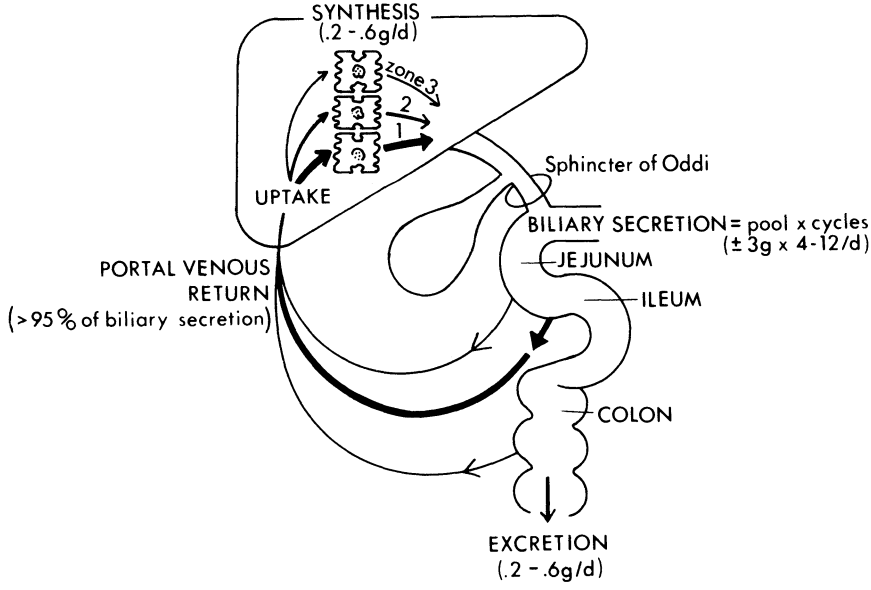


Figure 1. Quantitative aspects of the enterohepatic circulation of bile acids that are more extensively taken up by hepatocytes closest to the apparent blood supply (zone 1). Modified after Paumgartner G., Enterohepatic circulation of bile acids. In: *Hepatology: A Festschrift for Hans Popper* (Brunner H., Thaler H., eds.), Raven Press, New York, 1985, pp. 329–334.

exposed to incoming blood containing higher oxygen tension than the cells close to the central vein. This concept led to the suggestion that hepatocytes can be classified into three zones: zone 1, close to the portal vein; zone 3, close to the central vein; and zone 2, between zones 1 and 3.[10] Morphological and biochemical evidence suggest that the roles of these zones in biliary secretion differ.[11] After biliary drainage, periportal canalicular size is reduced compared to controls or to bile acid-infused animals, suggesting a predominant role of the cells in zone 1 for bile acid uptake. Layden and Boyer[12,13] found that the diameter of canaliculi in the vicinity of the portal vein was 0.84 μm versus 0.70 μm in the zone around the central vein. Infusion of bile acids increased the diameter of the canaliculi around the central vein, so that the lobular differences no longer existed. They concluded that at physiological concentrations, bile acids are largely removed by hepatocytes in the portal vein area (zone 1). When large quantities of bile acids are presented to the liver, more molecules are removed closer to the central vein, thereby suggesting that these zone 3 cells constitute a reserve area for bile acid uptake and may be more involved in BAIF. Using methods for the selective destruction of zone 1 by alkyl alcohol and zone 3 by bromobenzene, Gumucio *et al.*[14] reached similar conclusions.

The lobular gradient of bile acid uptake has also been demonstrated using a bile salt analog.[15] Of interest is the recent demonstration that the BAIF is very low in newborn animals[16] and that this is associated with the absence of the lobular gradient.[17]

2.3. Hepatocyte Uptake of Bile Acids

The various events leading to the uptake of bile acids by liver cells involve the anatomical structure of the liver (lobular gradient uptake of bile acids), the binding of bile acids to albumin and to other plasma proteins, secondary active Na^+-dependent transport as demonstrated for taurocholic acid (TCA) as well as Na^+-independent uptake as shown when the bile acid K_m is exceeded, and, finally, interaction of bile acids with membrane receptors.[1–5,7] Although some of these steps have been demonstrated experimentally using various techniques, many are still speculative.

The uptake of bile acids is rapid with a half-time as short as 2–12 min.[18–20] There is a small difference between the different bile acids, the dihydroxy bile acid, chenodeoxycholic acid (CDCA), being taken up more rapidly than trihydroxycholic acid (CA). However, there is no difference between the free form and the different conjugates of the same bile acid (TCA = GCA = CA). In humans, the uptake of bile acids varies between 12 μmol/min in the fasting state and 57 μmol/min postprandially. There is experimental evidence that the uptake is saturable with a greater V_{max} for trihydroxy than for dihydroxy bile

acids.[21] The uptake is competitive in that taurine-conjugated CDCA inhibits TCA uptake.[22] Furthermore, countertransport has been shown since TCA accumulates in the cell against an electrochemical gradient and requires energy. All characteristics of the carrier-mediated transport have therefore been shown: saturation kinetics, competitive inhibitions, and countertransport.[7]

At physiological concentrations, there is little uptake of cholate by erythrocytes[23] and thus presumably more than 95% of bile acids in blood are normally present in plasma. Bile acids in plasma are associated with serum proteins in order to reduce their membrane detergent properties, which can bring about lysis of erythrocytes. The binding of bile acids to plasma proteins depends on the structure of the bile acids: CA < CDCA < LCA.[24] Albumin is the major carrier of bile acids[25,26] but there is a disagreement as to the number of bile acid-binding sites on each albumin molecule. The K_m value for the binding of conjugated cholate is in the order of 1, though some workers have detected other sites of low affinity. More recently, other plasma-binding proteins have been identified, in particular apoprotein B.[27] Cholate appears to be more than 95% bound to plasma albumin leaving only 5% in free solution. It has been assumed that this small percentage of free bile acids determines the rate of uptake.[28,29] However, recent studies suggest that albumin exerts an important role in hepatic uptake and that uptake does not depend on the concentration of free bile acids in plasma.[30] In these studies, the amount of free bile acid was reduced by increasing the albumin concentration, but this had no significant effect on uptake. The interaction of albumin with cell membranes in the immediate vicinity where bile acid uptake is taking place results in conformational changes in the albumin molecule that reduce the binding affinity of bile acid to albumin. This increases the concentration of free bile acids available for uptake at the cell membrane level. In hepatocytes, the membrane-binding capacity of bovine serum albumin to bile acids may be a saturable process.[31] A similar mechanism may exist for the transport of bilirubin and BSP.[32]

Once bile acid molecules reach the sinusoidal membrane, they may be transported through the sinusoidal membrane to the cellular cytoplasm by what is known as secondary active transport. This transport process is energy-dependent and requires the presence of a chemical gradient of ions (mainly Na^+) that cotransport with bile acids. Thus the uptake of bile acids, although saturable, depends on cotransport with Na^+.[33] The primary energy source is Na^+, K^+-ATPase, located on the sinusoidal membrane (68 sites/μm^2), directing the movement of Na^+ out of the cell.[34] Because uptake is saturable, bile acid translocation across the membrane likely involves a specific carrier as identified by Accatino and Simon.[35] This specific carrier is reduced by cycloheximide treatment[36] and is increased through the recruitment of unused carrier sites by increasing the bile acid pool.[37]

2.4. Cellular Transport

The vectorial transport of bile acids through the hepatocyte to the canalicular membrane is not well understood. However, binding to cytosolic proteins and to cellular membranes has been proposed as well as vesicular transport of bile acids. Because the preparation of cytosol destroys any *in vivo* compartmentation, transcellular transport is more difficult to study than membrane transport. Substantial amounts of bile acids are present in liver homogenates (~200 μM).[38–40] For each bile acid tested, more than 50% of that in the homogenate was found in the cytoplasm. These studies may not be accurate because of methodological problems in obtaining cellular fractions without bringing about a redistribution of molecules between organelles or cytoplasm. They do, however, suggest that bile acids are not all located in the cytoplasm.

Strange *et al.*[40,41] proposed the following model for distribution of bile acids in the cytoplasm:

	Microsomes	
Mitochondria	Bile acid in free solution	Cytosolic protein
	Nuclei	

Using this model, they showed experimentally that 22% of CA, 33% of CDCA, and 70% of LCA in the cytoplasm was bound to cytosolic proteins, whereas 11% of CA, 4% of CDCA, and 1% of LCA were in free solution. The rest was bound to mitochondria, microsomes, and nuclei. This model indicates that the relative amounts of free and protein-bound bile acids differ from one bile acid to another. The cellular proteins involved in bile acid binding have been studied by Strange *et al.*[42–44] The binding of GCA, CA, CDCA, and LCA to cytosolic proteins by equilibrium dialysis shows, as expected, that the less polar bile acids, CDCA and LCA, demonstrated more binding capacity than CA and GCA. It was further shown that bile acids were bound to proteins of the same molecular weight as the X and Y fractions.[45] However, the affinity of the binding to X was significantly lower than to Y (1:100). Other experiments showed that in the Y fraction, the glutathione-*S*-transferases were responsible for bile acid binding.[46,47] It is therefore possible that binding of bile acids to glutatione blocks and back-diffusion of bile acids into plasma rather than modulating the transport of bile acids in the cell. A similar suggestion has been made with regard to the binding of glutathione to bilirubin.[48]

Recently, two proteins unrelated to glutathione-*S*-transferase were found in the Y fraction of rat liver and shown to bind a variety of bile acids.[49,50] Binding of bile acids to cytosolic proteins reduces free bile acids in the hepatocytes to concentrations similar to those observed in plasma. This would result in a gradient favorable for the uptake of bile acids. Other studies with labeled

bile acids showed binding of bile acids to cellular fractions *in vitro* and suggested that bile acids through their hydrophobic–hydrophilic properties are inserted in the bilayers of the membranes[39] rather than bound to membrane proteins. However, cautious interpretation is in order in view of difficulties that can occur during the formation of bile acid–lipid micelles. Some of the bile acids could precipitate during the binding studies. This difficulty can be circumvented by reducing the binding through the inclusion of protein and cholesterol in phospholipid vesicles.[51]

The evidence that subcellular organelles are involved in vectorial vesicular transport of bile acids has been suggested by morphological studies.[52–58] Both the smooth endoplasmic reticulum and the Golgi apparatus have been implicated.

2.5. Canalicular Secretion of Bile Acids

2.5.1. The Carriers

Bile salts must cross canalicular membranes for biliary secretion to take place. It is doubtful that the electrochemical gradient is the only factor controlling bile acid secretion since the concentration of bile acids in liver is around 0.2 mM and greater than 2 mM in the canaliculus. Because the resting cell membrane potential is about −40 mV, there will be substantial movement of negative bile acids down this electrical gradient into the canaliculus, but against a significant concentration gradient. Thus, there is a need for an energy-requiring secretory system and/or a carrier transport system. Recently, Meier *et al.* identified bile acid protein carriers in purified canalicular fractions.[59] Moreover, pericanalicular microfilaments may have a mechanical function in canalicular bile formation.[60–64]

2.5.2. Role of Biliary Phospholipids on Bile Acid Secretion

Recent evidence has suggested that the secretion of bile acids may depend on the output of biliary phospholipids. A hypothetical model was proposed in which biliary lipid-containing vesicles are transported from the hepatocytic intracellular site to the canalicular membrane. These vesicles could then fuse with the membrane forming highly fluid lipid microdomains that would preferentially be solubilized by bile acids.[65,66] Further experimental evidence showed that if the supply of biliary phospholipids is exhausted, bile acids then solubilize the bile canalicular membrane and cause a defect in the secretion of bile acids as well as in bile flow. We therefore suggested that the supply of biliary phospholipids could be an important determinant of the maximum secretion rate of bile acids.[67]

2.6. Bile Salt-Independent Bile Flow

2.6.1. Na^+, K^+-ATPase

It would be hazardous to rely on bile acids as the only generating system for bile flow because of the variability of bile acid flux through the liver. Bile acid-independent bile flow (BAIF) has been demonstrated in many species. Na^+, K^+-ATPase was originally suggested as playing a role in the formation of this component of bile flow.[68] In general, there is a good correlation between bile flow and Na^+, K^+-ATPase activity. The concept is based on the ion pump being situated in the canalicular membrane and resulting in active sodium movement into the canaliculus or lateral spaces with passive movement of water and electrolytes to maintain osmotic and electrical neutrality. This theory came into question after histochemical studies showed that in hepatocytes Na^+, K^+-ATPase was located on the sinusoidal (basolateral) part of the membrane.[69,70] Thus, various hypotheses were put forward to explain the correlation between Na^+, K^+-ATPase activity and bile flow. It was suggested that the enzyme on the sinusoidal membrane maintains high-extracellular, low-intracellular Na^+ necessary to drive and accumulate substances into the cell prior to their metabolism or secretion in, as noted earlier,[5] bile. This may also be true for other substances, such as Cl^- and HCO_3^- or organic substances.[34] Second, extrusion of sodium into the lateral space would be followed by passive diffusion through the tight junction (paracellular pathway).[71] Finally, accumulation of substrate energized by the Na^+ gradient would be followed by extrusion of that substrate into bile. Although these suggestions are probably correct, Na^+, K^+-ATPase has been found by monoclonal and ferritin antibody conjugates to be also located on the canalicular membrane and even at much higher concentration, in terms of units per micrometers than on the sinusoidal and basolateral membrane.[71–74] Thus, the direct role of Na^+, K^+-ATPase in the BAIF is tenable in addition to its secondary role on bile acid uptake.[75]

2.6.2. Bicarbonate and Bile Salt-Independent Bile Flow

The high content of bicarbonate in bile can be explained by the presence of a bicarbonate transport system similar to that found in other organs.[76–80] The Na^+/H^+ antiport may provide sufficient anions to combine with CO_2 produced via carbonic anhydrase activity within the cell and produce HCO_3^- for secretion in bile. It has been suggested that the Na^+/H^+ antiport is located on the sinusoidal membrane.[81] The contribution of Cl^- to the formation of BAIF remains unclear. Although a Na^+/Cl^- transport system has been postulated, studies showed that neither the rate of entry nor the equilibrium concentration of Cl^- was affected by the intracellular Na^+ or by the Na^+ gradient. It can therefore be concluded that Cl^- passively distributes across the plasma membrane and is

not involved in the production of the BAIF. Klassen[82] suggested the existence of a putative choleretic anion in rat bile. He found that the sum of Na^+ and K^+ was greater than that of Cl^- and HCO_3^-. In addition the cation–anion gap varied with changes of bile flow and could reach 30 mM. The fact that it exceeded the bile acid concentration raises the possibility of another osmotically active anion in rat bile. Klos *et al.*[83] suggested the presence of at least two choleretic components in bile: one is bile acid and the other an anion of molecular weight <1000.

2.7. Role of Cytoskeletal Proteins

The high pressure necessary for driving bile through the ductular system has been difficult to explain simply on the basis of hydrostatic pressure. It must result from coordination of contractile forces analogous to the contractile subunits of muscle. This phenomenon has recently been established by identifying a microfilamental network surrounding the canalicular pole of the hepatocyte.[84,85] The role of these microfilaments in bile formation was suggested by experiments with agents such as cytochalasin B or phalloidin which cause microfilamental disorganization and reduce bile flow.[86,87] Using isolated couplets of hepatocytes, Oshio *et al.* observed that bile canalicular structures contract and that this contractile activity is increased when cholic acid is added.[88,89] Furthermore, for contractions to take place, the coordinated activity of two neighboring hepatocytes is required. Contractions are coordinated and segmental, suggesting that they are involved in the forward movement of bile in the canalicular system.[90]

2.8. The Paracellular Pathway

Differences in the biliary tree permeability for water and ions may account for differences in bile flow. The permeability to water and ions of the canaliculus is controlled by the permeability of the canalicular membrane itself and that of the tight junction. The permeability of the junctional complex between cells is often estimated by measuring the bile-to-plasma ratio of large solutes (such as sucrose, inulin, polyethylene, or glycol), which are presumed to be excluded from hepatocytes and other cells and thus enter bile solely by the paracellular pathway. It can also be examined morphologically by following the penetration of lanthanum into the canalicular lumen through the intercellular space.[91]

The paracellular pathway depends on the permeability and integrity of the tight junction. If Na^+, K^+-ATPase is located on the basolateral membrane, then a degree of permeability or leakage is necessary physiologically in a sinusoidal to canalicular direction, if Na^+ is important in bile secretion. The neg-

ative electrical charge of the junctional complex is an important factor in the transport selectivity of the complex and in exclusion of bile acids. The amiloride-sensitive Na^+/H^+ antiport found in the sinusoidal domain of the cell could modulate intracellular pH, influence the function of the gap junction coupling on the lateral domain of the membrane, and thus affect cell-to-cell communication.[92] Ca^{2+} ions could also influence the tight junction by a direct effect on gap junction structure or microfilamental attachments.[93]

2.9. Cholehepatic Pathway

Hofmann *et al.* recently proposed that free bile acids after secretion into the ductular system could be protonated, reabsorbed from the biliary tract, and resecreted into bile. Such multiple cycles through a cholehepatic shunt pathway would stimulate canalicular bile flow at each cycle. Bile acids known to have a high choleretic potential serve as proof for this "cholehepatic shunt" pathway.[94–96] However, since most of these bile acids have a high critical micelle concentration, another explanation for their choleresis is the fact that they can be secreted as monomers.

2.10. Fluid Phase Endocytosis

The isolated perfused rat liver and *in vitro* rat hepatocyte cultures have been used to quantitate and characterize pathways of fluid phase endocytosis. These studies showed that each hour the hepatocyte can endocytose the equivalent of at least 20% of its cell volume and five times its plasma membrane surface area. Of the total amount of inulin endocytosed, 80% is regurgitated back into plasma, 18% is transported in sequestration compartments, and only 2% is secreted in bile. These vesicular pathways do not require vesicle acidification and are unaffected by the presence of taurocholate. Intact microtubular function is necessary for the transport of endocytosed substrates to bile or to sequestered compartments, but not for their initial uptake.[97]

2.11. Bile Ductules, Ducts, and Ductular Plexus

Canalicular bile is modified in its travels through the biliary system through secretion and reabsorption of water and ions. Evidence of distal bile water reabsorption comes mainly from the bile/plasma ratio of some inert markers, such as erythritol and mannitol, which are above 1. This indicates reabsorption of water and ions in view of the apparent inpermeability of ductular epithelium for these markers. In recent experiments with dogs, Tavoloni[98] calculated that bile water is reabsorbed in the ductular system at a rate of 0.2 μl/min/g of

liver tissue, possibly through a sodium-coupled chloride transport mechanism similar to that of the gallbladder.

Furthermore, Yamamoto and Phillips[99] using scanning electron microscopy of biliary casts identified a complex of portal bile ductular channels. They suggested that these structures (ductular plexus) could be involved in modification of canalicular bile by secretion or reabsorption of solutes and water.

3. BILE ACID-INDUCED CHOLESTASIS

3.1. Mechanisms of Cholestasis

Since cholestasis is a defect in bile formation and/or secretion, an abnormality in any of the steps involved in bile formation theoretically could induce cholestasis. Therefore cholestasis can be due to either impaired solute and water transport or enhanced bile "reflux" (Table 1).

3.2. Clinical Relevance of Experimental Bile Acid-Induced Cholestasis (Table 2)

Since bile acids appear to be the major driving force in bile formation, it is not surprising that an abnormality in bile acids transport leads to cholestasis. The exact role of bile acids as a cause of some cholestatic conditions in humans is controversial. However, in neonates, toxic bile acids may play an important primary role in the pathogenesis of cholestasis.[100] There is some evidence that in patients on total parenteral nutrition (TPN) lithocholic acid, a well-studied cholestatic agent, may be responsible.[101] This monohydroxy bile acid, found in

Table 1. Mechanisms of Cholestasis

Impaired solute and water transport
Altered uptake (carriers, permeability, viscosity, Na^+, K^+ATPase)
Impaired BADF (altered bile acid synthesis or cellular transport)
Defective BAIF (ion pumps, BCM, and tight junction permeability)
Interference with formation of biliary lipids (bile acids, hypolipidemic agents)
Alterations of cytoskeleton (cytochalasin, phalloidin, N.A. Indian cirrhosis)
Lack of cellular ATP (hypoxia, TPN)
Alterations of BCM (bile acids, drugs, obstruction)
Enhanced bile "reflux"
Altered membrane and tight junction permeability (drugs, sex, hormones, bile acids)
Injury to bile ducts (inflammation, hypoplasia, atresia, SC, PBC, stones) resulting in the abnormal reabsorption or secretion of water and/or electrolytes

Table 2. Clinical Relevance of Experimental Bile Acid-Induced Cholestasis

Neonates
↑ Serum bile acids
Synthesis of monohydroxy bile acids
Inborn error of bile acid metabolism
THCA
TPN
↑ LCA
↑ Sulfated LCA
Medical dissolution of gallstones
Liver injury related to LCA and to the extent of its sulfation
Cystic fibrosis
Increased LCA with a high glycine/taurine ratio of conjugated bile acids

high concentrations, correlates with the liver injury which can be prevented by antibiotics suppressing the intestinal microflora and hence LCA production.[102] Of patients receiving CDCA for dissolution of gallstones, LCA-type morphological lesions have been identified in a small percentage.[103] Finally, there is a familial type of cholestasis in which an abnormal (possibly cholestatil) bile acid (THCA) has been identified.[104] In addition, abnormal bile acid composition, especially an increase in LCA content and an increase in the glycine/taurine ratio of conjugated bile acids, was found in patients with cystic fibrosis.[105] However, unknown is what relation, if any, these abnormalities may have with the focal and multilobular biliary cirrhosis seen with increased frequency in these patients.[106] Furthermore, in all forms of cholestasis, an increased concentration of bile acids is noted with an increased proportion of sulfated bile acids.[107–109] Therefore an understanding of bile acid-induced cholestasis may lead to the prevention and/or treatment of many cholestatic conditions in humans.

3.3. LCA-Induced Cholestasis

Lithocholic acid is a natural bile acid present in trace amounts in the bile of humans,[110] rats,[111–113] and several other mammalian species.[114,115] The major route for the formation of LCA is 7_1-dehydroxylation of chenodeoxycholic acid,[116] but it can be formed directly from cholesterol in the liver by a pathway involving 26-hydroxylation of cholesterol. The latter pathway has been demonstrated in animals[117,118] and in humans.[119]

Lithocholic acid causes cholestasis in all experimental animals.[120–124] The

mechanism originally proposed for LCA cholestasis was the formation of precipitates in the bile canalicular lumen and the subsequent mechanical obstruction of the lumen.[122] However, electron microscopic studies failed to document the presence of crystalline precipitates as a uniform observation in this form of cholestasis. Thus, the obstruction of bile flow by precipitation of LCA may be only a contributing factor to the overall picture.[124] However, all microscopic studies were uniform in demonstrating specific ultrastructural changes in the bile canalicular membrane (BCM) and the pericanalicular region with no significant morphological alterations in the subcellular organelles.[123–127] Thereafter, the suggestion was made that the BCM is the primary target organelle in this form of cholestasis. About 10 years ago Yousef, Kakis, et al.[128–131] in a series of experiments, demonstrated that LCA cholestasis is associated with a dramatic increase in cholesterol content of the BCM which resulted in an increase in the cholesterol-to-phospholipid ratio. These authors suggested that changes in the composition of the membrane are the cause of the cholestasis, since the increase in the cholesterol phospholipid ratio is expected to influence membrane viscosity, permeability, and enzymatic activities.[132–134] Further studies showed that the cholestasis was associated with increased cholesterol synthesis in the hepatocytes[135] and increased transport of cholesterol from the microsomes to the BCM.[136] Interestingly, newborn guinea pigs were less susceptible to LCA cholestasis. This observation appeared to be due to the absence of any increase in cholesterol and consequently of less incorporation of LCA and cholesterol in the BCM as a consequence of less binding of LCA to cytosolic proteins.[137] Thus, understanding these mechanisms provided strategies for possibly preventing the development of this form of cholestasis, either by inhibiting cellular binding of LCA to cytosolic proteins,[138] or by reducing cholesterol synthesis[139] or its cellular transport to the BCM[140] (Fig. 2). For example, use of cycloheximide to inhibit protein synthesis and consequently reduced LCA binding to cytosolic protein prevented LCA cholestasis.[138] When mevinolin (a cholesterol synthesis inhibitor) or clofibrate (a hypolipidemic agent), which also reduces cholesterol synthesis, were used, cholestasis was also prevented.[139] Finally, colchicine, a microtubule-disrupting agent which inhibits cellular transport of cholesterol, also prevented LCA cholestasis.[140] In all cases where prevention of cholestasis was observed, the cholesterol content of BCM and the cholesterol phospholipid ratio in the membrane were normal, suggesting that normal membrane composition is necessary for preservation of its functional and normal bile flow.

3.4. LCA-Sulfates

Some have claimed that LCA-induced cholestasis may not be significant in humans because the human liver is capable of efficiently sulfating and hence

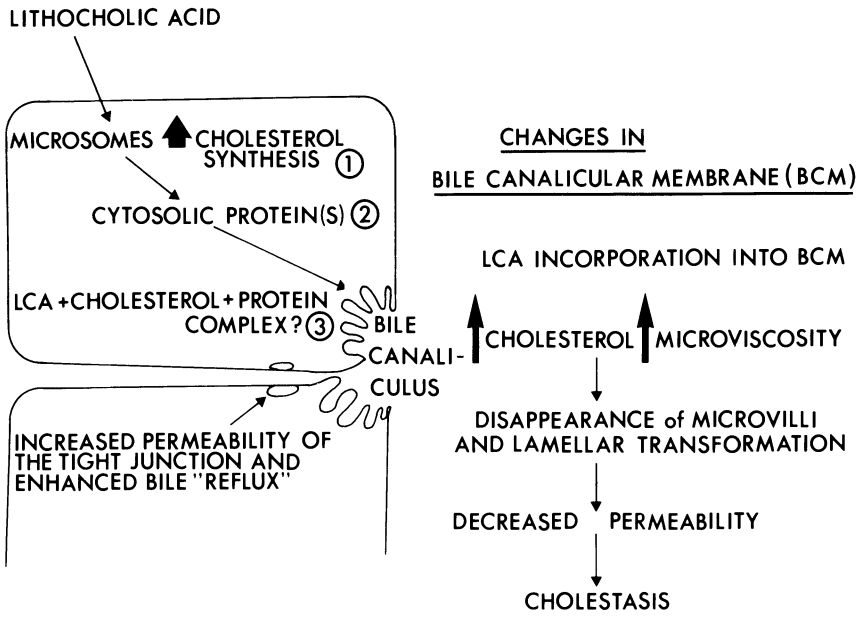


Figure 2. Model of proposed mechanism of lithocholic acid-induced cholestasis. (1) Inhibition of cholesterol synthesis (mevinolin, clofibrate) offers partial protection against cholestasis. (2) Inhibition of protein synthesis (cycloheximide) completely prevents cholestasis. (3) Cholic acid infusion inhibits the incorporation of LCA into the canalicular membrane and then increases cholesterol content, which leads to an altered cholesterol/phospholipid ratio. [Modified from Tuchweber B., Weber A., Roy C. C., Yousef I. M. Mechanisms of experimentally induced intrahepatic cholestasis. In: *Progress in Liver Diseases* (Popper M., Schaffner F., eds.), Vol. 8, Grune and Stratton, Orlando, 1986, pp. 164–178.]

detoxifying this compound,[141,142] increasing its secretion in urine and feces. However, there is little support for increased secretion in urine and feces acting as a significant protective mechanism. First, urinary excretion of sulfated bile acids is small relative to the bile acid pool, even if 50–85% of urinary bile acids are sulfates.[143] Second, fecal excretion is not a very effective sequestering mechanism. Increased levels of sulfated bile acids have been demonstrated in plasma after meals. Finally, a recent study has shown that the intestinal absorption of lithocholic acid sulfate is as efficient as that of nonsulfated LCA.[144] Detoxification of LCA by sulfation was assumed to be related to the increased solubility of sulfated bile acids. However, Carey *et al.*[145] demonstrated that the solubility of sulfated glyco-LCA (G-LCA-S) was similar to that of the nonsulfated G-LCA and only the solubility of tauro-LCA-S was better than that of tauro-LCA. Since sulfated bile acids such as LCA are well absorbed from the intestine, are not secreted in the urine in appreciable amounts, and form con-

jugates in the liver with glycine and/or taurine, it was imperative to test experimentally the effect of LCA-S on bile flow and to determine any cholestatic potential.

Experiments performed in our laboratory demonstrated that, in fact, G-LCA-S is cholestatic, but tauro-LCA-S is not.[146] Although the mechanism of G-LCA-S cholestasis is not yet fully understood, the evidence shows that it involves the cellular transport of this compound[146] and may also bring about changes in the permeability of the tight junction.[147] However, the observation that tauro-LCA-S does not induce cholestasis provides a mean of possible prevention of LCA-S cholestasis by conjugation of the bile acid with taurine. Taurine is a preferred substrate for conjugation of bile acids[148]; thus supplementation of taurine in the diet would increase the proportion of tauro-conjugated bile acids.[149,150]

The guinea pig conjugates its bile acids mainly with glycine.[151] Thus infusion of LCA-S induced cholestasis because the majority of the infused bile acids were conjugated with glycine. However, supplementation of taurine in the drinking water of these animals changed the conjugation to taurine and prevented LCA-S-induced cholestasis.[152] Taurine prevented the cholestatic effect of LCA-S in guinea pigs. Patients with cystic fibrosis lose large quantities of bile acids in their feces and have bile acids predominantly conjugated with glycine, which are precipitated in the acid milieu of their duodenums. Feeding taurine to these patients improved fat absorption by improving the duodenal bile acid glycine/taurine profile.[153,154] However, the role of taurine in preventing liver disease in cystic fibrosis is still under investigation. In addition to the role of taurine in prevention of LCA cholestasis, the influence of taurine on bile acid secretory maximum was investigated.[155] Taurine supplementation given I.V. or P.O. increased the secretory maximum (SRm) in guinea pigs, in correlation with the extent of conjugation with taurine. Thus we concluded that taurine modified the conjugation pattern and increased the SRm through its effect on canalicular secretion.[155]

The amino acid content of TPN solutions has been entertained as a cause of TPN-associated cholestasis.[156] The addition of taurine may prevent the decrease in bile flow and in bile acid secretion rate.[157] The protective effect of taurine remains to be elucidated, but could be mediated by its effect on the conjugating pattern as well as by its role in membrane protection.[158]

3.5. Cholestasis Induced by High Concentration of Bile Acid and the Control of Secretory Rate Maximum

When bile acids are infused into experimental animals in stepwise increasing doses, further administration of the bile acid above its maximum secretory rate results in a decline of both bile flow and bile acid secretion.[159–162] This

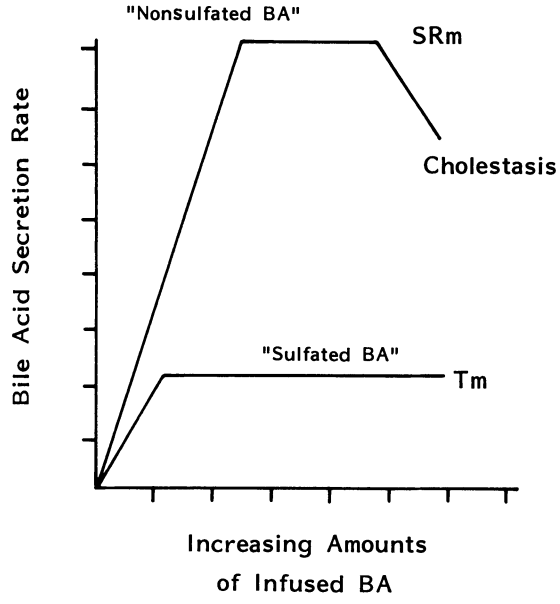


Figure 3. The concept of a secretory rate maximum for bile acids only applies to nonsulfated bile acids. After reaching a plateau (SRm), there is a progressive decrease in the secretion rate as opposed to the sulfated acids where a transport maximum (Tm) is achieved.

differs from the accepted classical definition of "transport maximum" (Tm) in which the secretory rate remains constant despite an increasing delivery of a particular substance above this level. The secretory rate maximum (SRm) of bile acids differ from one bile acid to the other, being inversely related to the detergent strength of the bile acid.[162] The cause of the difference between the SRm and the classical Tm is not known. The reduction in bile acid secretion and bile flow that occurs after reaching the SRm of normal bile acids cannot be simply explained by general systemic toxicity.[161]

A recent study with various common bile acids has confirmed the drop in bile flow achieved after reaching the SRm. The dynamics of secretion differ when similar experiments are carried out with the sulfated forms of these bile acids. As noted in Fig. 3, with increased bile acid infusion, sulfates are excreted at a constant rate (Tm), as opposed to nonsulfated bile acids which peaked at a SRm.[163]

Recent evidence has suggested that the secretion of bile acids may also depend on the output of biliary phospholipids.[164] A hypothetical model was proposed in which biliary lipids are transported from specialized hepatocytic intracellular sites to the canalicular membranes. These lipids fuse with the can-

alicular membrane forming microdomains that are preferentially solubilized by bile acids[164,165] (Fig. 4). Since the SRm is related to the detergent strength of the bile acids, it was speculated that bile acids, when infused in higher concentration, would lead to depletion of the intracellular and/or canalicular biliary phospholipids. Through this mechanism, membrane structure and function could be altered to modulate the SRm and eventually interfere with bile acid secretion. During the initial stages of bile acid infusion, bile flow with the secretory rate of bile acids, phospholipid, and cholesterol were increased. Phospholipid secretion reached a maximum secretion first, followed by the SRm of bile acids coupled with cholesterol. The SRm of bile acids correlated inversely with their detergent strength CA > CDCA > DOCA. In contrast, maximum secretion of phospholipid was similar for all bile acids. After reaching the maximum secretion, the secretion of phospholipids decreased and the decline became more marked after the bile acids reached their SRm. Although the major phospholipid species secreted was phosphatidylcholine, the secretion rate of phosphatidylethanolamine and sphingomyelin was increased after the bile acids reached their SRm. This change in the biliary phospholipids secreted was also reflected

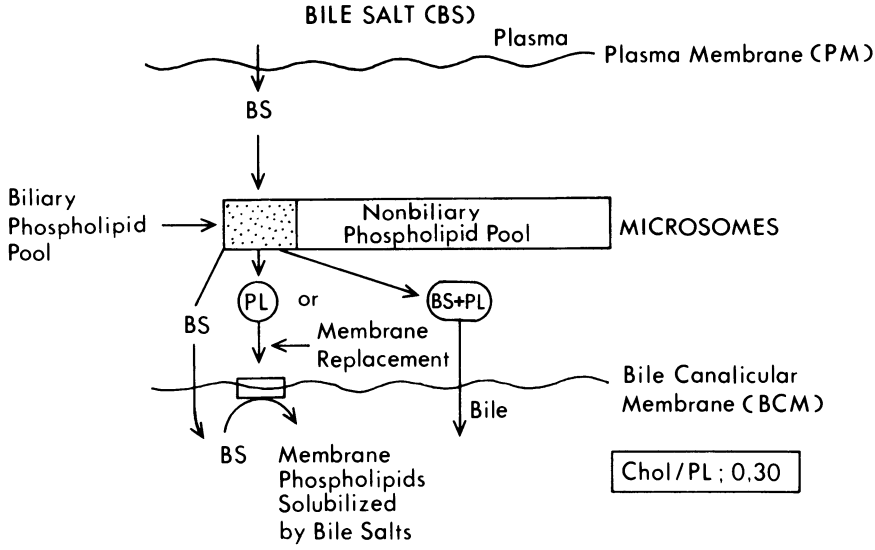


Figure 4. Phospholipids and the SRm of bile acids. In situations where large amounts of bile acids deplete a hypothetical biliary phospholipid pool, membrane and particularly BCM phospholipids are solubilized leading to an increase in the cholesterol/phospholipid (Chol/Pl) ratio, a decrease in the fluidity of the BCM, and a drop in the bile acid secretion rate. (Modified after Lowe P. J., Barnwell S. G., Coleman R. Rapid kinetic analysis of bile salt-dependent secretion of phospholipids, cholesterol, and plasma membrane enzymes into bile. *Biochem J* 1984;222:631–637.)

in an increased percentage of long-chain fatty acids, which are characteristic of the fatty acid constituents sphingomyelin and phosphatidylethanolamine.

These findings suggest that the available pool of biliary phosphatidylcholine becomes depleted with increasing doses of bile acid infusion. The depletion of phosphatidylcholine (biliary) results in the secretion of abnormal phospholipids into bile. The origin of these phospholipids was investigated in preliminary studies. After cholic acid infusions at 1, 2, 3, and 4 μ mol/min/100 g body weight (each dose infused for 30 min), the total phospholipid in canalicular membranes was significantly reduced without affecting cholesterol content. As a result, there was an alteration in the cholesterol-to-phospholipid ratio which, as previously shown, was also observed with LCA-induced cholestasis.[130] This suggested that the sphingomyelin and phosphatidylethanolamine secreted in the bile after their SRm may originate from canalicular membranes as well as other cellular membranous phospholipids. The increase in cholesterol-to-phospholipid ratio in these membranes may well influence the viscosity and fluidity of the membranes, and could result in dysfunction of membrane transport and the reduction in bile formation noted after the SRm of the bile acids had been exceeded. Thus solubilization of cellular membrane phospholipids could regulate the SRm of bile acids and the SRm of bile acids would be inversely related to their detergent strength, since deoxycholic acid (DOCA) solubilizes more phospholipids than CDCA and CA,[163] and consequently DOCA has the lowest SRm of the bile acids tested. In view of the fact that 99% of circulating bile acids have to be taken up and secreted by hepatocytes with higher concentrations after and during cholestasis after meals, investigating the protective role of sulfation proved interesting.

Sulfation has been regarded as a detoxification mechanism for bile acids.[168,169] However, little information is available on the physiochemical and physiological properties of sulfated bile acids. We prepared sulfated bile acids of CA, CDCA, and DOCA using sulfur trioxide trimethylamine as the sulfur donor and investigated their physiochemical properties as well as their effect on bile formation. The complete sulfation abolishes the detergent properties of the bile acids in that they cannot solubilize phospholipid or cholesterol *in vitro*. In addition, all bile acids showed a pK_a of 4.8 and a very high CMC between 20 and 40 mM, a 5- to 10-fold increase compared to the nonsulfated form.[168] When the bile acid sulfates were infused in stepwise increasing doses they exhibited Tm characteristics, but a plateau was lower than for nonsulfated bile acids (Fig. 3). In addition, they significantly reduced the secretion of phospholipid and cholesterol. The "choleretic potential," defined as the amount of bile secreted per unit of bile acid secreted, is significantly higher than for nonsulfates.[59] Thus, sulfation may protect from the cholestatic effect of high concentrations of bile acids by virtue of their high choleretic potential as well as by reducing membrane damage resulting from the absence of detergent properties.[163]

4. CONCLUSION

Intense study of the many different aspects of bile formation using a variety of probes and experimental models has resulted in an enormous amount of information. Important leads offer the prospect of considerable improvement in our understanding of bile formation and secretion. Much progress has been achieved by the creation of cholestasis in animals by a variety of endogenous and exogenous compounds, particularly monohydroxy bile acids.

Bile acid-induced cholestasis may have clinical relevance because of the abnormality in bile acid synthesis in newborns and the presence of cholestatic bile acids in patients receiving TPN and CDCA or ursodeoxycholic acid for the gallstones. In addition, bile acid abnormalities are associated with various liver diseases; a high concentration of bile acids is a hallmark of cholestasis. Investigations on lithocholic acid-induced cholestasis have uncovered several potential means of prevention, i.e., taurine supplementation, hypolipidemic drugs, and sulfation. Whether these measures will prove of beneficial use in the high-risk groups of patients described previously remains unknown.

ACKNOWLEDGMENT. The authors thank Sylvie Tasse and Danielle St-Cyr Huot for their excellent preparation of this manuscript.

REFERENCES

1. Sherlock S. Cholestasis. In: *Diseases of the Liver and Biliary System*, 7th ed. Oxford: Blackwell Scientific, 1985, pp. 214–245.
2. Tuchweber B, Weber A, Roy CC, Yousef IM. Mechanisms of experimentally induced intrahepatic cholestasis. In: *Progress in Liver Disease*, Vol 3 (Popper m, Schaffner F, eds). Orlando: Grune and Stratton, 1986, pp. 164–178.
3. Oelberg DG, Lester R. Cellular mechanisms of cholestasis. *Ann Rev Med* 1986;37:297–317.
4. Phillips MJ, Poucell S, Oda M. Mechanisms of cholestasis. *Lab Invest* 1986;54:593–608.
5. Boyer J. Mechanism of bile secretion and transport. In: *Progress in Liver Disease*, Vol 8 (Popper M, Schaffner F, eds). Orlando: Grune and Stratton, 1986, pp. 609–636.
6. Carey MC. The enterohepatic circulation. In: *The Liver: Biology and Pathobiology* (Arias I, Popper H, Schacter D, Shafritz DA, eds). New York: Raven press, 1982, p. 429.
7. Strange RC. Hepatic bile flow. *Physiol Rev* 1984;64:1055–1102.
8. Sperber I. Biliary secretion of organic anions and its influence on bile flow. In: *The Biliary System* (Taylor W, ed). Oxford: Blackwell, 1965, pp. 457–467.
9. Graft J, Peterlik M. Mechanism of transport of inorganic ions into bile. In: *The Hepatobiliary System: Fundamental and Pathological Mechanisms* (Taylor W, ed). New York: Plenum Press, 1975, pp. 43–58.
10. Rappaport AM. The microcirculatory hepatic unit. *Microvasc Res* 1973; 6:212–228.
11. Gumucio JJ, Miller DL. Functional implications of liver cell heterogeneity. *Gastroenterology* 1981;80:393–403.

12. Layden TJ, Boyer JL. Influence of bile acids on bile canalicular size. *Lab Invest* 1978;39:110–119.
13. Layden TJ, Boyer JL. Recruitment of central lobular hepatocytes for bile acid-dependent bile secretion. In: *Biological Effects of Bile Acids* (Paumgartner G, Strehl A, Gerok W, eds). Lancaster, UK: MTP, 1979, pp. 3–9.
14. Gumucio JJ, Balabaud C, Miller DL, Demason LJ, Appelman HD, Stoecker TJ, Franzblaud DR. Bile secretion and liver cell heterogeneity in the rat. *J Lab Clin Med* 1978;91:350–362.
15. Jones AL, Hradek GT, Renston RH, Wong KW, Karlaganis G, Paumgartner G. Autoradiographic evidence for hepatic lobular concentration gradient of bile acid derivative. *Am J Physiol* 1980;238:G233–G237.
16. Shaffer EA, Zahavi I, Gall G. Postnatal development of hepatic bile formation in the rabbit. *Dig Dis Sci* 1985;30:558–563.
17. Suchy FJ, Balistreri WF, Schockey JR, Garfield SA. Absence of a hepatic lobular gradient for bile acid uptake in the suckling rat. *Hepatology* 1983;3:847.
18. O'Maille ERL, Richards TG, Short AH. The influence of conjugation of cholic acid on its uptake and secretion: Hepatic extraction of taurocholate and cholate in the dog. *J Physiol London* 1967;189:337–350.
19. Smallwood RA, Iser JH, Hoffman NE. The hepatic transport of conjugated and unconjugated bile acid studied simultaneously in the rat. In: *Advances in Bile Acid Research* (Matern S, Hackenschmidt J, Back P, Gerok W, eds). Stuttgart: Schattauer, 1975, pp. 229–232.
20. Strange RC, Nimmo IA, Percy-Robb IW. Studies in the rat on the hepatic subcellular distribution and biliary excretion of lithocholic acid. *Biochim Biophys Acta* 1979;588:70–80.
21. Schwarz LR, Burr R, Schwenk M, Pfaff E, Greim H. Uptake of taurocholic acid into isolated rat-liver cells. *Eur J Biochem* 1975;55:617–623.
22. Reichen J, Paumgartner G. Uptake of bile acids by perfused rat liver. *Am J Physiol* 1976;231:734–742.
23. Strange RC, Hume R, Eadington DW, Nimmo IA. Distribution of glycocholate in blood from human fetuses and adults. *Pediatr Res* 1981;15:1425–1428.
24. Rudman D, Kendall FE. Bile acid content of human serum. II. The binding of cholanic acids by human plasma proteins. *J Clin Invest* 1957;36:538–542.
25. Burke CW, Lewis B, Panveliwalla D, Tabaqchali S. The binding of cholic acid and its taurine conjugate to serum proteins. *Clin Chim Acta* 1971;32:207–214.
26. Roda A, Cappelleri G, Aldini R, Roda E, Barbara L. Quantitative aspects of the interaction of bile acids with human serum albumin. *J Lipid Res* 1982;23:490–495.
27. Delahunty T, Feldkamp T. A study of endogenous N-cholylglycine distribution among serum proteins using radioimmunoassay. *Steroids* 1980;36:439–449.
28. Shand DG, Gotham RH, Wilkinson GR. Perfusion-limited effects of plasma drug binding on hepatic drug extraction. *Life Sci* 1976;19:125–130.
29. Wilkinson GR, Shand DG. A physiological approach to hepatic drug clearance. *Clin Pharmacol Ther* 1975;18:377–390.
30. Forker EL, Luxon BA. Albumin helps mediate removal of taurocholate by rat liver. *J Clin Invest* 1981;67:1517–1522.
31. Ockner R, Weisigner R, Lysenko N. Specific and saturable binding of ^{125}I-albumin to rat hepatocytes: Further evidence for a surface membrane albumin receptor (abstract). *Gastroenterology* 1980;79:1041.
32. Weisigner R, Gollan J, Ockner R. An albumin receptor on the liver cell may mediate hepatic uptake of sulfobromophthalein and bilirubin: Bound ligand, not free, is the major uptake determinant. *Gastroenterology* 1980;79:1065.
33. Ruifrock PG, Meijer DKF. Sodium ion-coupled intake of taurocholate by rat liver plasma membrane vesicles. *Liver* 1982;2:28–34.

34. Arias IM. Mechanisms and consequences of ion transport in the liver. In: *Progress in Liver Diseases,* Vol 3 (Popper H, Shaffner F, eds). Orlando: Grune and Stratton, 1986, pp. 145–159.
35. Accatino L, Simon FR. Identification and characterization of a bile acid receptor in isolated liver surface membranes. *J Clin Invest* 1976;57:496–508.
36. Gonzalez MC, Sutherland E, Simon FR. Regulation of hepatic transport of bile salts and their binding to liver surface membrane fractions. *J Clin Invest* 1979; 63:684–694.
37. Simon FR, Sutherland EM, Gonzalez M. Regulation of bile salt transport in rat liver. *J Clin Invest* 1982;70:401–411.
38. Okishio T, Nair PP. Studies on bile acids. Some observations on the intracellular localization of major bile acids in rat liver. *Biochemistry* 1966;5:3662–3667.
39. Strange RC, Beckett GJ, Percy-Robb IW. Nuclear and cytosolic distribution of conjugated cholic acid and radiolabelled glycocholic acid in rat liver. *Biochem J* 1979;178:71–78.
40. Strange RC, Chapman BT, Johnston JD, Nimmo IA, Percy-Robb IW. Partitioning of bile acids into subcellular organelles and the in vivo distribution of bile acids in rat liver. *Biochem Biophys Acta* 1979;573:535–545.
41. Strange RC. Hepatic bile salt transport. A review of subcellular binding sites. *Biochem Soc Trans* 1981;9:170–174.
42. Strange RC, Cramb R, Hayes JD, Percy-Robb IW. Partial purification of two lithocholic acid-binding proteins from rat liver 100000g supernatants. *Biochem J* 1977;165:425–429.
43. Strange RC, Nimmo IA, Percy-Robb IW. Equilibrium-dialysis studies of the interaction between cholic acid and 100000g-supernatant preparations from the rat liver. *Biochem J* 1976;156:427–433.
44. Strange RC, Nimmo IA, Percy-Robb IW. Binding of bile acids by 100,000g supernatants from rat liver. *Biochem J* 1977;162:659–664.
45. Levi AJ, Gatmaitan Z, Arias IM. Two hepatic cytoplasmic protein fractions, Y and Z, and their possible role in the hepatic uptake of bilirubin, sulfobromophthalein and other anions. *J Clin Invest* 1969;48:2156–2167.
46. Pattinson N. Purification by affinity chromatography of glutathione S-transferases A and C from rat liver cytosol. *Anal Biochem* 1981;115:424–427.
47. Pattinson NR. Isolation of two cholic acid-binding proteins from rat liver cytosol using affinity chromatography. *Biochim Biophys Acta* 1981;667:70–76.
48. Wolkoff AW, Ketley JN, Waggoner JG, Berk PD, Jakoby WB. Hepatic accumulation and intracellular binding of conjugated bilirubin. *J Clin Invest* 1978;61:142–149.
49. Sugiyama Y, Yamada T, Kaplowitz N. Newly identified organic anion-binding proteins in rat liver cytosol. *Biochem Biophys Acta* 1982;709:342–352.
50. Sugiyama Y, Yamada T, Kaplowitz N. Newly identified bile acid binders in rat liver cytosol. *J Biol Chem* 1983;258:3602–3607.
51. Conrad MJ, Singer SJ. The solubility of amphipathic molecules in biological membranes and lipid bilayers and its implications for membrane structure. *Biochemistry* 1981;20:808–818.
52. Jones AL, Schmucker DL, Mooney JS, Adler RD, Ockner RK. Morphometric analysis of rat hepatocytes after total biliary obstruction. *Gastroenterology* 1976;71:1051–1061.
53. Jones AL, Schmucker DL, Mooney JS, Adler R, Ockner RK. A quantitative analysis of hepatic ultrastructure in rats during enhanced bile secretion. *Anat Rec* 1978;192:227–228.
54. Jones AL, Schmucker DL, Mooney JS, Ockner RK, Adler RD. Alterations in hepatic pericanalicular cytoplasm during enhanced bile secretory activity. *Lab Invest* 1979;40:512–517.
55. Boyer JL, Istbashi M, Hruban Z. Formation of pericanalicular vacuoles during sodium dehydrocholate choleresis: A mechanism for bile acid transport. In: *The Liver; Quantitative Aspects of Structure and Function* (Preisig R, Bircher J, Paumgartner G, eds). Bern: Cantor-Aulendorf, 1979, pp. 163–178.

56. Suchy FJ, Balistreri WF, Hung J, Miller P, Garfield SA. Intracellular bile acid transport in rat liver as visualized by electron microscope radioautography using a bile acid analogue. *Am J Physiol* 1983;(*Gastrointest Liver Physiol* 8):G681–G689.
57. Goldsmith MA, Huling S. Effect of estradiol on hepatocyte handling of horseradish peroxidase (HRP) and bile salts (abstract). *Gastroenterology* 1982;82:1255A.
58. Reuben A, Allen RM, Boyer JL. Intrahepatic sources of "biliary-like" bile acid (BA)-phospholipid (PL)-cholesterol (CH) micelles (abstract). *Gastroenterology* 1982;82:1241A.
59. Meier PJ, Ruetz SH, Fricker G, Landmann L. Characterization of canalicular bile acid transport system of rat liver (abstract). In: *IX International Bile Acid Meeting Falk Symposium*, 1986, p. 32.
60. Korn ED. Biochemistry of actomyosin-dependent cell motility. *Proc Natl Acad Sci USA* 1978;75:588–599.
61. Tuchweber B, Gabbiani C. Phalloidin-induced hyperplasia of actin microfilaments in rat hepatocytes. In: *The Liver: Quantitative Aspects of Structure and Function* (Preisig R, Bircher J, Paumgartner G, eds). Bern: Cantor-Aulendorf, 1976, pp. 84–90.
62. Phillips MJ, Oda M, Mak E, Fisher MM, Jeejeebhoy KN. Microfilament dysfunction as a possible cause of intrahepatic cholestasis. *Gastroenterology* 1975;69:48–58.
63. Phillips MJ, Oshio C, Miyairi M. Microfilament dysfunction as a mechanism in intrahepatic cholestasis; evidence from time lapse cinemicrophotography (abstract). *Gastroenterology* 1980;79:1120.
64. Oshio C, Phillips MJ. Contractility of bile canaliculi: implications for liver function. *Science* 1981;212:1041–1042.
65. Lowe PJ, Barnwell SG, Coleman R. Rapid kinetic analysis of the bile-salt-dependent secretion of phospholipid, cholesterol and a plasma-membrane enzyme into bile. *Biochem J* 1984;222:631–637.
66. Barnwell SG, Lowe PJ, Coleman R. The effects of colchicine on the secretion into bile of bile salts, phospholipids, cholesterol and plasma membrane enzymes: Bile salts are secreted unaccompanied by phospholipids and cholesterol. *Biochem J* 1984;220:723–731.
67. Yousef IM, Barnwell S, Gratton F, Tuchweber B, Weber A, Roy CC. Liver cell membrane solubilization may control maximum secretory rate of cholic acid in the rat. *Am J Physiol* 1987;252:684–696.
68. Erlinger S. Does Na^+K^+ATPase have any rate in bile secretion? *Am J Physiol* 1982;243:6243–6247.
69. Blitzer BL, Boyer JL. Cytochemical localization of Na^+K^+ATPase in the rat hepatocyte. *J Clin Invest* 1978;62:1104–1108.
70. Latham PS, Kashgarian M. The ultrastructural localization of transport ATPase in the rat liver at non-bile canalicular plasma membrane. *Gastroenterology* 1979;76:988–996.
71. Layden TJ, Elias E, Boyer JL. Bile formation in the rat. The rate of the paracellular shunt pathway. *J Clin Invest* 1978;62:1375–1385.
72. Schenk DB, Leffert HL. Monoclonal antibodies to rat Na^+,K^+ATPase block enzymatic activity. *Proc Natl Acad Sci USA* 1983;80:5281–5285.
73. Takemura S, Omori K, Tanaka K, Omori K, Matsumura S, Tashiro Y. Quantitative immunoferritin localization of Na^+,K^+ATPase on the canine hepatocyte cell surface. *J Cell Biol* 1984;99:1502–1512.
74. Leffert HL, Schenk DB, Hubert JJ, Skelly H, Schumacher M, Ariyasu R, Ellisman M, Koch KS, Keller GA. Hepatic Na^+,K^+ATPase: A current view of its structure, function and localization in rat liver as revealed by studies with monoclonal antibodies. *Hepatology* 1985;5:501–507.
75. Yousef IM, Tuchweber B, Weber AM, Roy CC. Where is Na^+,K^+ATPase located in the liver plasma membrane. *Gastroenterology* 1984;86:1632–1633.

76. Diamond JM. Transport of salt and water in rabbit and guinea pig gall bladder. *J Gen Physiol* 1964;48:1–14.
77. Schulz I, Strover F, Ullrich KJ. Lipid soluble weak organic buffers as "substrate" for pancreatic secretion. *Pfluegers Arch* 1971;323:121–140.
78. Swanson CG, Solomon AK. Evidence for Na-H exchange in the rabbit pancreas. *Nature New Biol* 1972;236:183–184.
79. Turnberg LA, Bieberdorf FA, Morawski SG, Fordtran JS. Interrelationship of chloride, bicarbonate, sodium and hydrogen transport in human ileum. *J Clin Invest* 1970;49:557–567.
80. Turnberg LA, Fordtran JS, Carter MW, Rector FC Jr. Mechanism of bicarbonate absorption and its relationship to sodium transport in the human jejunum. *J Clin Invest* 1970;49:548–556.
81. Moseley H, Meier PJ, Knickelbein R, Aronson PS, Boyer JL. Evidence for Na^+H^+ exchange in rat liver basolateral but not canalicular membrane vesicles (abstract). *Hepatology* 1984;4:1040.
82. Klassen CD. Does bile acid secretion determine canalicular bile production in rats? *Am J Physiol* 1971;220:667–673.
83. Klos C, Paumgartner G, Reichen J. Cation-anion gap and choleretic properties of rat bile. *Am J Physiol* 1979; (*Endocrinol Metab Gastrointest Physiol* 5):E434–E440.
84. French SW. Role of canalicular contraction in bile flow. *Lab Invest* 1985;53:245–249.
85. Denk H, Lackinger E, Vennigerholz F. Pathology of cytoskeleton of hepatocytes. In: *Progress in Liver Diseases,* Vol 8 (Popper H, Schaffner F, eds). Orlando: Grune and Stratton, 1986, pp. 237–251.
86. Phillips MJ, Oda M, Mak E, Fisher MM, Jeejeebhoy KN. Microfilament dysfunction as a possible cause of intrahepatic cholestasis. *Gastroenterology* 1975;69:48–58.
87. Vonk RJ, Yousef IM, Corriveau JP, Tuchweber B. Phalloidin-induced morphological and functional changes of rat liver. *Liver* 1982;2:133–140.
88. Oshio C, Phillips MJ. Contractility of bile canaliculi: Implication for liver function. *Science* 1981;212:1041–1042.
89. Miyairi M, Oshio C, Watanabe S, Smith CR, Yousef IM, Phillips MJ. Taurocholate accelerates bile canalicular contractions in isolated rat hepatocytes. *Gastroenterology* 1984;87:788–792.
90. Watanabe S, Smith CR, Phillips MJ. Coordination of the contractile activity of bile canaliculi, evidence from calcium microinjection of triplet hepatocytes. *Lab Invest* 1985;531:275–280.
91. Boyer JL. Tight junction in normal and cholestatic liver. Does the paracellular pathway have a significance? *Hepatology* 1983;3:614–617.
92. Spray DC, Ginzberg RD, Morales EA, Gatmaitan Z, Aris IM. Electrophyisological properties of gap junctions between dissociated pairs of rat hepatocytes. *J Cell Biol* 1986;103:135–144.
93. Lindbald L, Schersten T. Influence of cholic and chenodeoxycholic acid on canalicular bile flow in man. *Gastroenterology* 1976;70:1121–1124.
94. O'Maille ERL, Kozmary SV, Hofmann AF, Gurantz D. Differing effects of norcholate and cholate on bile flow and biliary lipid secretion in the rat. *Am J Physiol* 1984;246:G67–G71.
95. Danzinger RG, Nakagaki M, Hofmann AF. Differing effects of hydroxy-7-oxotaurine conjugated bile acids on bile flow and biliary lipid secretion in dogs. *Am J Physiol* 1984;246:G166–G173.
96. Yoon YB, Hagey LR, Hofmann AF, Gurantz D, Michelotti EL, Steinbach JH. Effect of side-chain shortening on the physiologic properties of bile acids: Hepatic transport and effect on biliary secretion of 23-nor-urso-deoxycholate in rodents. *Gastroenterology* 1986;90:837–852.
97. Scharschmidt BF, Lake JR, Licho V, Wong MA, Van Dyke RW. Fluid phase endocytosis by cultured rat hepatocytes. Implications for plasma membrane remodeling. *Hepatology* 1985;5: 1011.
98. Tavoloni N. Role of ductular bile water reabsorption in canine bile secretion. *J Lab Clin Med* 1985;106:154–161.

99. Yamamoto K, Phillips MJ. A hitherto unrecognized bile ductular plexus in normal rat liver. *Hepatology* 1984;4:381–385.
100. Javitt NB. Current status of cholestasis induced by nomohydroxy bile acids. In: *Jaundice* (Goresky GA, Fisher MM, eds). New York: Plenum Press, 1975, pp. 401–409.
101. Fouin-Fortunet H, Le Quernec L, Erlinger S, Lerebours E, Collin R. Hepatic alterations during total parenteral nutrition in patients with inflammatory bowel disease: A possible consequence of lithocholate toxicity. *Gastroenterology* 1982;82:932–937.
102. Capron JP, Gineston JL, Herve MA, Braillon A. Metronidazole in prevention of cholestasis associated with parenteral nutrition. *Lancet* 1983;1:446–447.
103. Fisher RL, Anderson DW, Boyer JL, Ishak K, Klatskin G, Lachin JM, Phillips MJ and the Steering Committee for the National Cooperative Gallstone Study Group. A prospective morphologic evaluation of hepatic toxicity of cohnodeoxycholic acid in patients with cholelithiasis. The National Cooperative Gallstone Study. *Hepatology* 1982;2:187–201.
104. Hanson RF, Isenberg JN, Williams GC, Hachey D, Szczepanik P, Klein PD, Scharp HL. The metabolism of 3α, 7α, 12α-trihydroxy-5β-cholestane-26-oic acid in two siblings with cholestasis due to intrahepatic bile duct anomalies. An apparent inborn error of cholic acid synthesis. *J Clin Invest* 1975;56:557–587.
105. Roy CC, Weber AM, Morin CL, Combes JC, Nussle D, Megevand A, Lasalle R. Abnormal biliary lipid composition in cystic fibrosis. *N Engl J Med* 1977;297:1301–1305.
106. Roy CC, Weber AM, Morin CL, Lepage G, Brisson G, Yousef I, Lasalle R. Hepatobiliary disease in cystic fibrosis: A survey of current issues and concepts. *J Ped Gastroenterol Nutr* 1982;1:469–478.
107. Korman MG, Hofmann AF, Summerskill WH. Assessment of activity in chronic liver disease: Serum bile acids compared with conventional tests and histology. *N Engl J Med* 1974;290:1399–1402.
108. Makino I, Hashimoto H, Shinozaki K, Yoshino K, Nakagawa S. Sulfated and nonsulfated bile acids in urine, serum and bile of patients with hepatobiliary diseases. *Gastroenterology* 1975;68:545–553.
109. Stiehl A, Ast E, Czygan P, Frohling W, Raedsch R, Kommerell B. Pool size, synthesis and turnover of sulfated and nonsulfated cholic acid and chenodeoxycholic acid in patients with fibrosis of the liver. *Gastroenterology* 1978;74:572–577.
110. Fisher MM, Yousef IM. Sex differences in the bile acid composition of human bile: studies in patients with and without gallstones. *Can Med Assoc J* 1973;109:190–193.
111. Yousef IM, Kakis G, Fisher MM. Bile acid metabolism in mammals: III. Sex differences in the bile acid composition of rat bile. *Can J Biochem* 1972;50:402–414.
112. Subbiah MT, Kuksis A, Mookerjea S. Secretion of bile salts by isolated and intact rat liver. *Can J Biochem* 1969;47:847–854.
113. Fisher MM, Kakis G, Yousef IM. Bile acid pool in Wistar rats. *Lipids* 1976;11:93–96.
114. Yousef IM, Yousef MK, Bradley WG. Bile acid composition in some desert rodents. *Proc Soc Exp Biol Med* 1973;143:596–601.
115. Yousef IM, Bradley WG, Yousef MK. Bile acid composition of some lizards from Southwestern United States. *Proc Soc Exp Biol Med* 1977;154:22–26.
116. Danielsson H, Einarsson K. Formation and metabolism of bile acids. In: *Biological Basis of Medicine,* Vol 5 (Bittar EE, Bittar N, eds). New York: Academic Press, 1969, pp. 279–325.
117. Mitropoulos KA, Myant NB. The formation of lithocholic acid, chenodeoxycholic acid and alpha and beta-muricholic acids from cholesterol incubated with rat liver mitochondria. *Biochem J* 1967;103:472–479.
118. Mitropoulos KA, Myant NB. The formation of lithocholic acid chenodeoxycholic acid and other bile acids from 3-beta-hydroxychol-5-enoic acid in vitro and in vivo. *Biochim Biophys Acta* 1967;144:430–439.

119. Anderson K, Kok E, Javitt NB. Bile acid synthesis in man: metabolism of 7 hydroxycholesterol-^{14}C and 26-hydroxycholesterol-^{3}H. *J clin Invest* 1972;51:112–117.
120. Fisher MM, Magnusson R, Miyai K. Bile acid metabolism in mammals: I. Bile acid induced intrahepatic cholestasis. *Lab Invest* 1971;21:88–91.
121. Javitt NB. Cholestasis in rats induced by taurolithocholate. *Nature* 1966; 210:1262–1263.
122. Javitt NB, Emerman S. Effect of sodium taurolithocholate on bile flow and bile acid excretion. *J Clin Invest* 1968;47:1002–1014.
123. Miyai K, Mayr WW, Richardson AL. Acute cholestasis induced by lithocholic acid in rats. A freeze-fracture replica and a thin secretion study. *Lab Invest* 1975;32:527–535.
124. Miyai K, Richardson AL, Mayr WW, Javitt NB. Subcellular pathology of rat liver in cholestasis and choleresis induced by bile salts. I. Effects of lithocholic, 3 hydroxy-5-cholenoic, cholic and dehydrocholic acids. *Lab Invest* 1977;36:249–258.
125. Palmer RH, Bolt MG. Bile acid sulfates. I. Synthesis of lithocholic acid sulfate and their identification in human bile. *J Lipid Res* 1971;12:671–679.
126. Cowen AE, Korman MG, Hofmann AF, Cass OW, Coffin SB. Metabolism of lithocholate in healthy men. II. Enterohepatic circulation. *Gastroenterology* 1975;69:67–76.
127. Layden TJ, Schwarz, Boyer JL. Scanning electron microscopy of rat liver. Studies of the effect of taurolithocholate and other models of cholestasis. *Gastroenterology* 1975;69:724–738.
128. Yousef IM, Kakis G, Fisher MM. Lithocholate induced intrahepatic cholesteasis (abstract). *Gastroenterology* 1976;70:996.
129. Kakis G, Yousef IM. Pathogenesis of lithocholate and taurolithocholate induced intrahepatic cholestasis in rats. *Gastroenterology* 1978;75:595–607.
130. Kakis G, Phillips MJ, Yousef IM. The respective roles of membrane cholesterol and of sodium potassium adenosine triphosphatase in the pathogenesis of lithocholate induced cholestasis. *Lab Invest* 1980;43:73–81.
131. Kakis G, Yousef IM. Mechanism of cholic acid protection in lithocholate induced intrahepatic cholestasis in rats. *Gastroenterology* 1980;78:1402–1411.
132. Cooper RA. Abnormalities of cell membrane fluidity in the pathogenesis of disease. *N Engl J Med* 1977;297:371–377.
133. Papahadjopoulos D, Cowden M, Kimelberg H. Role of cholesterol in membranes: effects on phospholipid-protein interactions membrane permeability and enzymatic activity. *Biochim Biophys Acta* 1973;330:8–26.
134. Simon FR, Gonzales M, Davis R. Studies on the pathogenesis of cholesterol gallstone formation: Alterations of bile acid transport and liver surface membrane lipid structure by estrogen. In: *Gallstones* (Fisher MM, Goresky CA, Shaffer EA, Strasberg SM, eds). New York: Plenum Press, 1979; 251–265.
135. Yousef IM, Tuchweber B. Effect of lithocholic acid on cholesterol synthesis and transport in the rat liver. *Biochim Biophys Acta* 1984;796:336–344.
136. Yousef IM, Lewittes M, Tuchweber B, Roy CC, Weber A. Lithocholic acid cholesterol interactions in rat liver plasma membrane fractions. *Biochim Biophys Acta* 1984;796:345–353.
137. Lewittes M, Tuchweber B, Weber A, Roy CC, Yousef IM. Resistance of the suckling guinea pig to lithocholic acid induced cholestasis. *Hepatology* 1984;4:486–491.
138. Yousef IM, Tuchweber B, Weber A. Prevention of lithocholate induced cholestasis by cycloheximide, an inhibitor of protein synthesis. *Life Sci* 1983;33:103–110.
139. Kugelmass R, Tuchweber B, Weber A, Roy CC, Yousef IM. Effect of inhibition of cholesterol synthesis on lithocholic induced cholestasis in rats (abstract). *Hepatology* 1982;2:730.
140. Barnwell SG, Yousef IM, Tuchweber B. The effect of colchicine on the development of lithocholic acid induced cholestasis: a study of the role of microtubules in intracellular cholesterol transport. *Biochem J* 1986;236:345–350.

141. Palmer RH. Bile acids, liver injury, and liver disease. *Arch Intern Med* 1972;130:606–617.
142. Cowen AE, Korman MG, Hofmann AF, Cass OW. Metabolism of lithocholate in healthy man. I. Biotransformation and biliary excretion of intravenously administered lithocholate, lithocholylglycine, and their sulfates. *Gastroenterology* 1975;69:59–66.
143. Raedsch R, Stiehl A, Gundert-Remy U, Kommerell B. Hepatic secretion of sulfated and glucuronidated bile acids in man. In: *Enterohepatic Circulation of Bile Acids and Sterol Metabolism* (Paumgartner G, Stiehl A, Gerok W, eds). England: MTP Press, 1984, pp. 225–229.
144. Kuipers F, Heslinga H, Havinga R, Vonk RJ. Intestinal absorption of lithocholic acid sulfates in the rat: Inhibitory effects of calcium. *Am J Physiol* 1986;251:G189–G194.
145. Carey MC, Wy SFJ, Watkins JB. Solution properties of sulfated monohydroxy bile salts relative insolubility of the disodium salt of glycolithocholate sulfate. *Biochim Biophys Acta* 1979;575:16–26.
146. Yousef IM, Tuchweber B, Vonk RJ, Masse D, Audet M, Roy CC. Lithocholate cholestasis-sulfated glycolithocholate induced intrahepatic cholestasis in rats. *Gastroenterology* 1981;80:233–241.
147. Roy CC, Fournier LA, Dorvill NP, Perea A, Weber A, Tuchweber B, Yousef IM. The pattern of bile acid conjugation: A critical determinant of the cholestatic potential of sulfated lithocholate (abstract). *Pediatr Res* 1983;17:198A.
148. Vessey DA. The biochemical basis for the conjugation of bile acids with either glycine or taurine. *Biochem J* 1978;147:621–626.
149. Sjovall J. Dietary glycine and taurine on bile acid conjugation in man. *Proc Soc Exp Biol Med* 1959;100:676–678.
159. Truswell AS, McVeigh S, Mitchell WD. Effect in man of feeding taurine on bile acid conjugation and serum cholesterol levels. *J Artheroscler Res* 1965;5:526–529.
151. Kibe A, Wake C, Kuramoto T, Hoshita T. Effect of dietary taurine on bile acid metabolism in guinea pig. *Lipids* 1980;15:224–229.
152. Dorvil NP, Yousef IM, Tuchweber B, Roy CC. Taurine prevents cholestasis induced by lithocholic acid sulfate in guinea pigs. *Am J Clin Nutr* 1983;37:221–232.
153. Darling PB, Lepage G, Leroy C, Masson P, Roy CC. Effect of taurine supplements on fat absorption in cystic fibrosis. *Pediatr Res* 1985;19:578–582.
154. Belli DC, Levy E, Darling PB, Leroy C, Lepage G, Giguere P, Roy CC. Taurine improves the absorption of a fat meal in cystic fibrosis patients. *Pediatrics* 1987;80:517–523.
155. Belli DC, Fournier LA, Lepage G, Tremblay P, Yousef IM, Roy CC. The influence of taurine on the bile acid maximum secretory rate in the guinea pig. *Pediatr Res* 1988;24:34–37.
156. Roy CC, Belli DC. Hepatobiliary complications associated with TPN: An enigma. *J Am Coll Nutr* 1985;4:651–660.
157. Guertin F, Roy CC, Tuchweber B, Yousef IM. Taurine prevents cholestasis associated with short term amino acid-dextrose infusion in guinea pigs (abstract). *Hepatology* 1986:6:1196.
158. Wright CE, Tallan HH, Lin YY, Gaull GE. Taurine: Biological update. *Ann Rev Biochem* 1986;55:427–453.
159. Drew R, Priestly GB. Cholestatic properties of various bile salts. *Experientia* 1979;35:809–811.
160. Hall TJ, Baker AL, Cooper MJ, Moosa AR. Choleresis and cholestasis produced by infusion of taurocholic acid or taurodehydrocholic acid combined with BSP in the rhesus monkey. *Dig Dis Sci* 1979;24:350–357.
161. Herz R, Paumgartner G, Preisig R. Inhibition of bile formation by high doses of taurocholate in the perfused rat liver. *Scand J Gastroenterol* 1975:11:741–761.
162. Hardison WGM, Hatoff DE, Miyai K, Weinger RG. Nature of bile acid maximum secretion rate in the rat. *Am J Physiol* 1981;241:337–343.

163. Yousef IM, Barnwell SA, Tuchweber B, Weber A, Roy CC. Effect of complete sulfation of bile acids on bile formation in rats. *Hepatology* 1987;7:535–542.
164. Billington D, Coleman R. Plasma membrane vesiculation, phospholipid solubilization and their possible relationship to bile secretion. *Biochim Biophys Acta* 1978;509:33–47.
165. Rahman K, Hammond TG, Lowe PJ, Barnwell SG, Clark B, Coleman R. Control of biliary phospholipid secretion. *Biochem J* 1986;234:421–427.
166. Stiehl A. Bile salt sulfates in cholestasis. *Eur J Clin Invest* 1974;4:59–63.
167. Summerfield JA, Cullen J, Barnes S, Billing BH. Evidence for re-control of urinary excretion of bile acid and bile acid sulfate: the cholestatic syndrome. *Clin Sci Mol Med* 1977;52:51–65.
168. Bartholomew TC, Summerfield JA, Billing BH, Lawson AM, Setchell KD. Bile acid profiles of human serum and skin interstitial fluid and their relationship to pruritus studied by gas chromatography-mass spectrometry. *Clin Sci* 1982:63:65–73.
169. Donovan JM, Yousef IM, Carey MC. Complete sulfation of the common bile salts of man. Synthesis, properties and interactions with lecithin (abstract). *Gastroenterology* 1984;86:1064.

4

Causes and Management of Ascites

L. M. Blendis

1. INTRODUCTION

Ascites is a major complication of cirrhosis, the other two complications being variceal hemorrhage and portasystemic encephalopathy. Yet it receives relatively little attention in the literature, being mistakenly regarded as less of a threat to life. Although the patient with bleeding varices and encephalopathy can present more dramatically, ascites and edema represent as great, if not more insidious, a threat to life. Indeed, Powell and Klatskin[1] showed that the appearance of ascites in a cirrhotic patient was associated with a significantly reduced 5-year survival rate.

Ascites and edema also form an important prognostic index for alcoholic liver disease.[2] The two major causes of death directly related to ascites are spontaneous bacterial peritonitis[3] and the hepatorenal syndrome,[4] which is often precipitated by overaggressive diuretic therapy. Ascites has also been implicated in predisposing to variceal hemorrhage by increasing gastroesophageal reflux, but this remains controversial.[5] Certainly overly aggressive therapy can also precipitate portasystemic encephalopathy. In addition, ascites results in a generalized deterioration in health, reduced mobility, and decreased appetite leading to malnutrition and loss of muscle mass. This is exacerbated by unpleasant or frankly intolerable low-salt diets. The patient becomes cachectic and the quality of life deteriorates. Thus, the pathogenesis of this complication of cirrhosis is worth eliciting since its prevention would result in a prolonged and better life.

L. M. Blendis • Department of Medicine, Toronto General Hospital, University of Toronto, Toronto, Ontario, Canada M5G 2C4.

2. PATHOGENESIS

2.1. The Affector Side

2.1.1. Portal Hypertension

The derangement of Starling forces within the peritoneal capillary bed is responsible for the localizing of fluid in the peritoneal compartment. Yet the peritoneal capillary and lymphatic system is extremely efficient in reabsorbing peritoneal fluid, especially in the region of the diaphragm.[6] So why does the system fail in cirrhotics? In short, because the patient becomes waterlogged as a result of being in prolonged positive salt and water balance, the pathogenesis of which remains complicated and almost certainly multifactorial.

Over the last few years it has become clearer that portal hypertension is a key factor in the pathogenesis of salt and water retention. The classical explanation is that portal hypertension causes pooling of blood in the splanchnic circulation, resulting in a decreased effective circulating blood volume or "underfilling" of the circulation. This in turn leads to a decreased renal blood flow and glomerular filtration rate, elevated plasma renin activity and serum aldosterone levels, sodium and fluid retention, and ascites. This in turn causes an incremental rise in portal pressure and further splanchnic pooling, creating a self-perpetuating vicious cycle.[7] The arguments against the so-called underfilling theory include the fact that total blood volume is invariably increased in these patients (so that the patients are never underfilled). There are therefore no grounds for invoking a reduction in effective circulating blood volume. Crucial to this argument is that a deceased effective circulating volume should result in a reduction of renal blood flow. Certainly in the early stages of ascites, renal blood flow is usually normal[8,9] together with plasma renin activity and serum aldosterone levels,[10,11] which may actually be reduced.[12]

In canine models of cirrhosis, Levy *et al.* elegantly demonstrated what the sequence of events was: cirrhosis with portal hypertension and normal plasma volume; positive sodium balance with an increased plasma volume; and finally ascites, but without any decrease in renal blood flow and therefore effective circulating blood volume.[13,14] Further studies showed that cirrhosis, in the absence of portal hypertension as a result of a side-to-side portacaval anastomosis, did not result in ascites. Such animals were able to escape from DOCA in contrast to cirrhotics with portal hypertension or those whose portacaval shunts had blocked.[15] These authors confirmed in a chronic model, while Campbell *et al.* confirmed in an acute model of hepatic vein constriction,[16] the well-recognized clinical observation that sinusoidal rather than presinusoidal is associated with ascites and sodium retention.

Thus, sinusoidal portal hypertension appears to send a message to the kid-

neys to retain sodium. How this message is sent is still unknown. However, experimental evidence in animals indicates that an increase in hepatic vein pressure results in increased nervous activity from both the portal vein and the kidneys. Thus, the possibility exists of a nervous reflex arc from the hepatic sinusoids to the nephron, stimulated by sinusoidal hypertension, and resulting in sodium retention.[17] Only a minority of cirrhotic patients have ascites at any one time. This is presumably due to the fact that the majority maintain the ability to "escape" from the tendency to retain fluid. When the patient loses this normal physiological response, ascites develops.

Is the height of the portal pressure a factor in the development of ascites? That is, do the sodium-retaining signals increase as the pressure increases? The answer is probably "no." In a large group of patients with alcoholic liver disease and ascites, the mean portal pressure (as measured by the wedged hepatic portal vein pressure) was a figure not significantly different from that of patients without ascites. Furthermore, the small increase in pressure in the ascitic patients could be abolished by paracentesis (Blendis, unpublished observations, 1987).

2.1.2. Other Factors

What other factors may enhance the localization of fluid into the peritoneal cavity? Hypoalbuminemia is theoretically important, reducing the oncotic pressure in the peritoneal capillaries. Yet no significant relationship between the plasma albumin concentration and the presence of ascites has been demonstrated.[18] Lymphatic congestion certainly occurs in postsinusoidal portal hypertension. Increased lymph formation drains increased plasma, and diffusion into the space of Disse is the result of the elevated pressure.[19] This raised lymphatic pressure causes increased lymph-rich fluid to weep from the liver surface and will affect the final ascitic protein concentration. However, there is no evidence that this is a major factor. The most important surface for the parietal peritoneum is probably in the subdiaphragmatic region,[6] yet the ability of this capillary system in ascites to reabsorb excess fluid is virtually unknown.

2.2. The Effector Side (Fig. 1)

The patient fails to "escape" from the tendency to retain salt and water. Let us consider what factors are involved in renal sodium and water handling in cirrhotic patients.

2.2.1. Glomerular Filtration Rate and Renal Blood Flow

A decrease in glomerular filtration rate will lead to salt and water rentention. Yet ascites formation occurs in the presence of normal filtration rates and

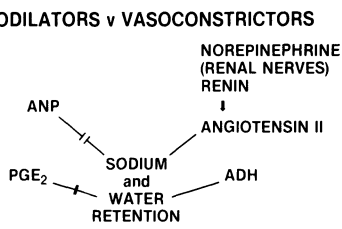


Figure 1. Pathogenesis of salt and water retention in cirrhotic ascites.

renal blood flow. This suggests that decreased filtration is a late rather than early additional factor in the development of ascites.

2.2.2. Renin–Angiotensin–Aldosterone System

This system has been the prime candidate for the cause of sodium retention ever since the discovery that ascitic cirrhotics have secondary hyperaldosteronism.[20] Yet up to one-third of patients with ascites may have normal plasma renin activity and serum aldosterone levels.[10,11] Furthermore, a study of patients with "early" ascites found that the mean plasma renin activity and serum aldosterone levels were significantly *lower,* rather than higher, than normal, even in the presence of an increased plasma volume.[12] This finding strongly supports the overflow theory.

An important method of studying the problem is to perturbate the system, reversing the tendency to salt and water retention either temporarily via head-out water immersion, or "permanently" by peritoneovenous shunting. The pressure of the water on the body immersed in a tub of warm water (temperature 34.5°C) causes a centralization of interstitial fluid into the intravascular compartment. In normal individuals this results in an increase in plasma volume, with a resultant 5% fall in venous hematocrit and a significant diuresis and natriuresis.[21] This is associated with a fall in plasma renin activity and serum aldosterone. In ascitic cirrhotics the effects are similar but exaggerated, with a greater (66%) fall in plasma renin activity and serum aldosterone from higher baselines and in some patients a greater natriuresis and diuresis. However, some patients have a blunted response with virtually no natriuresis.[22] Furthermore, similar studies performed during aldosterone blockade with spironolactone failed to blunt the natriuretic response.[23] It has been suggested that there is a critical level of serum aldosterone (50 ng/dl) below which the fall in serum aldosterone results in a natriuresis, but above which no natriuresis occurs.[24]

The permanent improvement in sodium and water balance that occurs following peritoneovenous shunting is also associated with near-normalization of plasma renin and serum aldosterone.[25] However, it is important to note that the

urinary changes following the peritoneovenous shunt are immediate, i.e., sodium excretion increases from 1 μeq/min to more than 100 μeq/min within 2 hr. Yet serum aldosterone levels, although decreasing by this time, are still markedly elevated and do not normalize until 24–72 hr postop.[25] Furthermore, follow-up studies at 1 year in ascites-free patients with normal plasma renin and serum aldosterone levels showed persistent sodium retention when challenged with 100 meq sodium diet.[26]

2.2.3. Serum Catecholamines and Sympathetic Nerve Activity

In 1968, Epstein and co-workers[27] demonstrated decreased superficial renal arterial opacification following renal arteriography in patients with hepatorenal syndrome. This appeared to be reversible at autopsy. The absence of any histological abnormality led to the suggestion that the changes were due to superficial vasoconstriction of the renal arteries.[27] This appeared to be confirmed by the observation that the changes were reversible, either transiently (by infusion of vasodilators directly into the renal artery with temporary improvement of renal function)[28] or following transplantation of the kidney into a noncirrhotic recipient.[29] Further confirmation came from isotopic renography that showed decreased superficial blood flow in the initial excretory phase.[30,31] These changes could be demonstrated early in the natural history of cirrhosis before deterioration in renal function. Caution is warranted as certain facets of these observations remain controversial. First, due to the instability of renal blood flow, the isotopic observations must be treated with caution.[32] Second, there is still no convincing evidence that redistribution of renal blood flow away from the superficial renal cortex leads to changes in salt and water handling by the kidney.[33,34] Nevertheless, these changes were likely due to increased sympathetic nerve activity for whatever reason and there is considerable evidence that the renal sympathetic nervous system directly influences sodium excretion with increased activity resulting in sodium retention.[35]

Confirmation of the increased renal sympathetic nerve activity in cirrhosis had to wait 15 years until improved assay systems became available. The advent of radioimmunoassays for catecholamines revealed that the plasma catecholamine (principally norepinephrine) levels in ascitic cirrhotics were significantly elevated above those of nonascitic cirrhotics. This was a curious finding since it has been known for 40 years that such patients have a hemodynamic circulation with increased cardiac output and low normal blood pressure,[39] and therefore a decreased systemic vascular resistance. That renal vein catecholamine levels were higher than renal arterial levels in ascitic cirrhotics indicated that such patients had increased renal sympathetic activity,[36] although not all studies agreed.[38] Such catecholamine data therefore suggested that these patients possessed a vascular resistance to circulating vasoconstrictors, especially

as angiotensin levels were also elevated.[11] Confirmation for this decreased responsiveness had already been provided, both in patients[40] and in experimental animal models of liver disease,[41,42] as a decreased hypertensive response to increased endogenous catecholamine secretion[43] or exogenous vasoconstrictor infusion.[40–42] The mechanism of this decreased response, whether due to abnormalities in receptor number or function,[44] or to circulating vasoactive competitors, such as the false neurotransmitters,[45] awaits clarification. Furthermore, the role of systemic vasodilation in the pathogenesis of sodium retention and renal vasoconstriction remains intriguing.

Perturbation of the system by water immersion resulted in either a decrease[46,47] or no change[48] in plasma catecholamine levels, and a variable natriuresis. However, combining water immersion with exogenous intravenous norepinephrine infusion resulted in a universal rise in blood pressure and a natriuresis, indicating that moderate intravascular volume expansion plus reversal of the peripheral vasodilation results in a significant reversal of sodium retention. Similarly, patients with resistant ascites undergoing peritoneovenous shunt have highly elevated preoperative levels of both free and conjugated forms of all the circulating catecholamines: norepinephrine, epinephrine, and dopamine.[49] Thus, there is increased renal production of catecholamines. Immediately after establishing a peritoneovenous shunt, postop during the initial massive natriuresis and diuresis, renal production of norepinephrine decreased.

This suggested a possible reduction in renal sympathetic nervous activity and therefore vasoconstriction, yet plasma catecholamine levels remained unchanged. Plasma catecholamine levels only normalized 2 weeks after the surgery.

In summary, although increased sympathetic nerve activity (indicated by elevated circulating catecholamine levels) is associated with hepatic ascites formation, acute or chronic perturbation of the volume status of such patients has demonstrated a dissociation between this system and salt and water retention in cirrhosis. Further, although these studies suggest that the catecholamine abnormalities occur in response to alteration in volume, they have not clarified the respective roles of volume and capacity due to vascular unresponsiveness.

2.2.4. *Arginine Vasopressin*

Although antidiuretic hormone (arginine vasopressin) does not have a direct role in renal sodium handling, inappropriate elevations of serum arginine vasopressin have been shown to occur in ascitic cirrhotics, particularly in those with hyponatremia.[50] Indeed, Bichet and co-workers[50] demonstrated that such patients could be separated into two groups depending on their response to a water load. Patients who excreted greater than 80% of an I.V. load of 20 ml/kg of 5% dextrose followed by 200 ml of water hourly within 5 hr were defined

as "excretors." In contrast, those who failed to excrete were defined as "non-excretors." Such patients had significantly lower serum sodium levels, lower 24-hr urine sodium excretion, and higher plasma renin, aldosterone, and norepinephrine levels. Both groups had elevated plasma arginine vasopressin levels, but in nonexcretors the levels were significantly higher and did not suppress with a water load. Furthermore, this group showed a highly significant correlation between plasma norepinephrine and arginine vasopressin levels and sodium excretion.[37] Nevertheless, following perturbation of the system either with water immersion[51] or peritoneovenous shunt,[52] diuresis and natriuresis occurred in the absence of any change[51] or in the presence of an actual increase in plasma arginine vasopressin levels.[52] This suggests that the dominant factor in fluid retention in hepatic ascites is a decreased filtered load to the distal segment, which is corrected by volume expansion. This also suggests that inappropriately high levels of plasma arginine vasopressin are of secondary importance.

2.2.5. Urinary Prostaglandins

The demonstration of increased urinary prostaglandin E (PGE) excretion in ascitic cirrhotics suggested an important role of these prostaglandins in salt and water retention in patients with liver disease. Again there is little evidence for a major role of prostaglandins in renal sodium handling, as opposed to renal water excretion.[53] The mechanism action of PGE is thought to be both by renal vasodilation[54] and a direct effect on distal nephron tubule function.[53] The importance of the PGE group in the maintenance of renal function was indicated first by the deterioration in renal function following administration of prostaglandin synthetase inhibitors (such as the nonsteroidal, antiinflammatory drugs, termed NSAID),[55,56] and second, by the demonstration of a decrease in PGE excretion in patients with hepatorenal syndrome.[56]

Again, perturbation of the system by water immersion significantly raises urinary PGE excretion along with the diuresis and natriuresis.[57] Recently, studies in patients undergoing peritoneovenous shunt again confirmed the elevated PGE_2 excretion in patients with hepatic ascites.[58] Immediately postoperative there was a transient increase in PGE_2 excretion. However, by day 3 after surgery, excretion rates were lower than preoperative, and by 2 weeks they had decreased to within the normal range.[58] These later changes occurred in the presence of a persistently increased sodium and water excretion. Thus these results suggest that increased urinary PGE_2 excretion occurs as a renal response to increased vasoconstrictor action on the kidney and that increased renal PGE production with consequent vasodilation helps to preserve renal function. Thus, a decrease in renal PGE production (either temporarily following NSAID ingestion or "permanently") results in a significant deterioration in renal function,

as in hepatorenal syndrome. That is, renal PGE production is a secondary phenomenon; with improvement in the renal circulation, renal PGE production returns to normal.

2.2.6. Atrionatriuretic Peptides

Atrionatriuretic peptides, found principally in the wall of the right atrium,[59] have a direct natriuretic action[60] and vasodilate muscular arterioles[61] including the efferent arterioles of the renal glomeruli.[62] Thus, the role of atrionatriuretic peptides in renal sodium and water retention in cirrhosis with and without ascites is likely to be important. Thus far, plasma atrionatriuretic peptides levels have been reported to be normal, elevated, or decreased in both groups of patients.[63,64] However, in a recent study by Gines and co-workers,[65] the levels were increased and sampling from the coronary sinus showed highly elevated levels, indicating increased production and excretion of atrionatriuretic peptides from the atria of the heart in ascitic cirrhotics. Elevated levels have also been demonstrated by Campbell *et al.*[66] in patients with resistant ascites. Furthermore, following peritoneovenous shunt there is a massive six fold rise in atrionatriuretic peptide levels immediately postop, which was related to an increased mean atrial pressure. Such changes would help explain the immediate massive natriuresis within 2 hr in the presence of both a slowly falling serum aldosterone and unchanged plasma catecholamine levels together with further elevation of plasma arginine vasopressin levels. Such changes would also help explain the further vasodilation and lowering of systemic vascular resistance found in these patients.[49] Thus far, the findings indicate that deficiency of atrionatriuretic peptides is not the cause of hepatic ascites; rather, a resistance to atrionatriuretic peptides is responsible. Such resistance could be due to the circulating vasoconstrictors causing renal vasoconstriction and thus preventing the vasodilatory action of atrionatriuretic peptides on the kidneys. Such patients in the late postoperative phase should be studied following the normalization of the circulating vasoconstrictors, but while renal sodium retention remains preserved.

2.2.7. Summary

In summary, the cause of sodium and water retention resulting in ascites in cirrhotic patients is still unclear, though obviously multifactorial. More and more evidence points to a perturbation in both systemic and renal hemodynamics (due to an imbalance of vasoconstrictor and vasodilator activity) as an important, if not the major, underlying abnormality.

3. MANAGEMENT OF ASCITES

3.1. Medically Responsive Ascites

The basic management of ascites is salt and water restriction, guided by clinical and biochemical findings. When the patient first presents, the physician will establish the cause of the liver disease and the degree of liver dysfunction. As far as the management of ascites is concerned, the presence or absence of leg edema will be seen to be important, as are the serum electrolytes, blood urea nitrogen (BUN), and serum creatinine. A 24-hr urine collection for total volume and sodium excretion on a known sodium intake (e.g., a no-added-salt diet will contain approximately 100 mM sodium) will also provide useful information.

3.2. Mild Ascites

In a patient who presents with mild to moderate ascites for the first time, the likelihood is that the serum BUN and creatinine will be normal, and that the 24-hr urine sodium excretion will be between 20 and 50 mM/day. Such patients should respond to mild sodium restriction, i.e., no-added-salt diet, no fluid restriction, and small doses of "distal" diuretics. Such patients when challenged with a water-load-test have both a prompt diuresis and natriuresis. This response suggests a distal physiological abnormality. The inability to generate "free water" indicates that (1) the filtered fluid load is reaching the distal nephron, (2) both the diluting, thick ascending loop of Henle and the suppression of arginine vasopressin are functioning normally, and (3) the causes of sodium retention in such patients lie in the more distal parts of the distal convoluted tubule and collecting duct. Diuretic therapy in these patients should begin with either spironolactone 100 mg or with amiloride 10 mg each morning.[67] Spironolactone because of its long half-life takes 3–4 days before the accumulated dose of this slow-acting diuretic becomes effective. In contrast, the effect of amiloride occurs within 24–48 hr.

Adjustment of the doses of these diuretics may be required. When there is an unsatisfactory response, 25–50 mg of hydrochlorothiazide may be combined with 25 or 50 mg of spironolatone in the form of Aldactazide® or 50 mg with 5 mg amiloride in Moduret® for an enhanced effect. The major side-effect of spironolactone is painful gynecomastia, which may necessitate stopping the medication. Rarely it causes a tubular acidosis. Both aldactone and amiloride can cause hyperkalemia. It is important not to advise taking salt substitutes (which are usually potassium salts) with distal diuretic therapy since this can result in serious hyperkalemia.

3.3. Moderate Ascites

In patients with greater amounts of ascites, serum BUN and creatinine levels may be at the upper limit of normal, or may be slightly elevated due to a minor reduction in creatinine clearance. Twenty-four-hour sodium excretion is usually further reduced to 10–20 mM sodium per day. A water-loading test may result in satisfactory water excretion with the formation of free water but with grossly impaired sodium excretion. This suggests the presence of increased reabsorption of chloride and sodium in the diluting segment. Such patients will likely require increasing amounts of distal diuretics to maximum doses, i.e., spironolactone 400–600 mg daily[68] or amiloride 30 mg daily. It is doubtful if further increases ever result in a significantly greater natriuresis. Instead, it is necessary to add a "loop diuretic" such as furosemide, starting at a dose of 40 mg daily. This powerful diuretic will result in an acute diuresis within 6–8 hr. Since the mobilization of 500 ml of ascitic fluid back into the circulation takes up to 24 hr,[69] the net fluid loss and therefore weight loss should not exceed 0.5 kg/day. The response of such nonazotemic patients may depend on the extent of the elevation of renin and aldosterone: patients with higher levels do not respond to furosemide.[70] The exception is patients with associated peripheral edema who mobilize fluid more rapidly. These patients appear to be at much less risk of developing the complications of powerful diuretic therapy.

3.4. Resistant Ascites

Patients who develop massive ascites even on maximum diuretic therapy are said to have "resistant" ascites. Admitting these patients and keeping them in bed for a few days on strict 20 mM sodium and 1-liter fluid restriction may often result in a diuresis, possibly related to improved renal hemodynamics and glomerular filtration, and decreased plasma renin, serum aldosterone, and plasma norepinephrine levels.[71] In patients resistant to maximum diuretic therapy, 24-hr urine sodium excretion is less than 10 mM/day, regardless of sodium intake. The patient fails to produce either a natriuresis or diuresis following a water load, indicating maximum reabsorption in the proximal tubule. What are the options for further management in such patients?

3.5. Additional Diuretic Therapy

More powerful loop diuretics such as ethacrynic acid have been used but have a high incidence of serious side effects.[72] The powerful diuretic Zaroxylen® 5–10 mg daily acts on the proximal tubule as well as the cortical diluting segment, and may be effective in some patients.[73] Again, this agent must be used with caution, and with careful monitoring of renal function.

3.6. Therapeutic Abdominal Paracentesis

Removal of large amounts of ascitic fluid by paracentesis is the oldest form of therapy, practiced for hundreds of years. Until the 1950s it was still the main method of therapy. Then, with the advent of the newer, less toxic diuretics such as the mercurials and then the thiazides, paracentesis disappeared from the therapeutic armory in cirrhosis, possibly because of injudicious use. It crept back in over the years, being used in many Veterans Administration hospitals, and was given respectability again by careful studies monitoring systemic and renal hemodynamics and function as well as hormone profiles.[74] These have shown that the removal of 5 liters or more of ascitic fluid with replacement of 1 unit of plasma or salt-free albumin results in no change in any of the above functional parameters. In contrast, with no intravenous colloid replacement, large-volume paracentesis results in a significant deterioration in serum creatinine and creatinine clearance, but a rise in right atrial pressure, serum atrionatriuretic peptide levels, and cardiac output.[75] The latter is presumably due to an improvement in venous return with reduction in intraabdominal pressure. In fact, a controlled trial of repeated abdominal paracentesis resulted in significantly less morbidity than diuretic therapy.[76] Thus, repeated abdominal paracentesis (which can be continued on an outpatient basis) could become the routine therapy for refractive ascites. However, a small group of patients require increasingly frequent paracentesis, resulting in a deterioration of quality of life.

3.7. Peritoneovenous Shunting

The peritoneovenous shunting operation is not without complications. Therefore, patients should be carefully selected and be clearly resistant to other forms of therapy for their ascites. For example, patients with decompensated liver disease do poorly and peritoneovenous shunting is contraindicated. A recent history of variceal bleeding is also a relative contraindication. In these patients, the varices should first be sclerosed and then the patient should undergo shunting. In patients with a recent history of spontaneous bacterial peritonitis, diagnostic paracentesis is necessary to be secure that the infection has completely resolved. Indeed, all patients should undergo a diagnostic paracentesis preoperatively to exclude an asymptomatic spontaneous bacterial peritonitis. In addition, any evidence of myocardial or ischemic heart disease requires investigation.

At operation, great care must be taken to correctly position the shunt. In addition, removal of most of the ascites and replacement with warmed Ringer's lactate or normal saline minimizes postoperative coagulopathy problems. Most patients have an asymptomatic rise in prothrombin time and a decrease in plate-

let count. Unfortunately, 5–20% of patients develop a disseminated intravenous coagulation (DIC)-like syndrome.[77,78] This can occur within the first few days postoperatively and can manifest clinically with the appearance of spontaneous bruising and bleeding. This DIC-like syndrome is thought to be due to a factor in the ascitic fluid (such as a procollagen[79] or endotoxin) which results in the stimulation of the coagulation cascade and absorption of platelets.

Following a successful peritoneovenous shunt, the patients either have a spontaneous natriuresis or diuresis, or else require 10 mg furosemide I.V. to initiate this mobilization. The resultant disappearance of ascites, with the liberation of the salt content of their diet and decreased diuretic requirements, causes an improvement in their quality of life; their appetite improves and they gain true weight or lean body mass.[80]

Unfortunately, there are many possible complications with this shunt, including infection, shunt blockage, as well as the DIC-like syndrome, which can result in shunt dysfunction. Modification of the LeVeen shunt with preset loops in parts (Hakim) or with a manual pumper for the patient (Denver) have not significantly affected the incidence of complications.[81] Further, the peritoneovenous shunt does not materially affect liver function and therefore does not improve the overall mortality rate.

More experimental therapeutic maneuvers, such as head-out water immersion or infusions of pharmacological doses of atrionatriuretic peptides, do not show any signs of being useful routine therapeutic modalities.

4. MANAGEMENT OF THE COMPLICATIONS OF ASCITES

4.1. Spontaneous Bacterial Peritonitis[82]

Spontaneous bacterial peritonitis is a primary infection of ascitic fluid in cirrhotics that is potentially lethal. The incidence appears to have increased over the years, especially with asymptomatic infections.[83] This is related to an increased recognition of the condition as well as the increasing numbers of paracentesis now performed, in which approximately 10% of patients are found to be infected.[84] The major infecting organisms are *E. coli,* streptococcus group D, and viridans.[85] The patients may present with fever, chills, and specific abdominal symptoms such as pain, tenderness, and decreased bowel sounds (Table 1). Alternatively, patients may present with nonspecific symptoms, such as unexplained encephalopathy or deteriorating liver and/or renal function. Suspect the diagnosis, then confirm it by examining the ascitic fluid for the criteria shown in Table 2.[85–89] Start treatment with antibiotics (an aminoglycoside plus either ampicillin or cephalosporin) while awaiting the cultures. If there is a significant increase in ascitic polymorphonuclear count in the diagnostic tap

Table 1. Clinical Findings in Patients with Spontaneous Bacterial Peritonitis

Manifestation	Percent of patients	
	Study 1	Study 2
Significant fever	54	68
Chills	14	
Abdominal pain	51	79
Abdominal tenderness	51	79
Rebound tenderness	47	42
Reduced bowel sounds	21	54
Decreased mentation	51	54
Asterixis	46	
Peripheral edema	72	
Hypotension	5	14

(Table 2), the antibiotics should be continued for a full 2 weeks, even if the culture is negative. Why? Because the prognosis in culture-negative neutrocytic ascites may be just as grave as spontaneous bacterial peritonitis.[90]

4.2. Hepatorenal Syndrome[4]

Hepatorenal syndrome is a rare complication of cirrhosis, but it occurs most commonly in decompensated alcoholic cirrhotics with ascites. It is usually

Table 2. Examination of Ascitic Fluid

Examination	Finding	Usefulness
Appearance	Turbid	−
pH 7.31	+	
Lactate	15 mg/dl	+
WBC count		
Total	500/mm^3	+
Polymorphs		
Total	250/mm^3	++++
% WBC count	25%	++++
Glucose	Low	−
Protein	High	−
Arterial-ascitic fluid gradient		
pH	0.1	−
Lactate	−20 mg/dl	−

precipitated by one of a number of factors including overaggressive diuresis, variceal bleeding, and spontaneous bacterial peritonitis. Hepatorenal syndrome is defined as a development of renal failure in a cirrhotic patient in the absence of intravascular depletion (i.e., excluding prerenal azotemia) or parenchymal renal disease (e.g., acute tubular necrosis). Hepatorenal syndrome is therefore characterized by a steadily rising serum creatinine, low urinary sodium excretion (10 mM/day), a benign urinary sediment, and a normal jugular venous pressure.

The prognosis in such a patient with hepatorenal syndrome is poor, with a mortality rate of more than 80%. There is no effective medical therapy. Various renovasodilators have been tried (either intravenously or via direct infusion into the renal artery), but only transient improvements occur.[28,91] The use of peritoneovenous shunt has been expounded for years, with some positive anecdotal results.[92–94] Yet a recent controlled trial showed no significant improvement with shunting over medical therapy on the dismal mortality rate.[95] In contrast, liver transplantation does hold out real hope for the complete reversal of this fatal condition, with a reported long-term survival in four patients of 1–4 years.[96]

5. CONCLUSION

The development of ascites in a cirrhotic patient, who until that point might well have had a good life expectancy, bodes ill. A decrease in the quality of life from dietary restriction, the necessity of diuretic medication, plus the possibility of life-threatening complications are all consequences of this development. It thus behooves us to continue our quest for an understanding of ascites so that treatment can improve until prevention becomes a reality.

REFERENCES

1. Powell WJ, Klatskin G. Duration of survival in patients with alcoholic cirrhosis. Influence of alcohol withdrawal and possible effects of recent changes in management of the disease. *Am J Med* 1968;44:406–412.
2. Orrego H, Israel Y, Blake JE, Medline A. Assessment of prognostic factors in alcoholic liver disease; toward a global quantitative expression of severity. *Hepatology* 1983;3:896–905.
3. Crossley IR, Williams R. Spontaneous bacterial peritonitis. *Gut* 1985;26:325–331.
4. Levy M. Hepatorenal syndrome. In: *The Kidney: Physiology and Pathophysiology* (Seldin DW, Giebisch G, ed). New York: Raven Press, 1985, pp. 1945–1961.
5. Eckardt VF, Grace ND, Kantrowitz PA. Does lower esophageal sphincter incompetency contribute to esophageal variceal bleeding? *Gastroenterology* 1976;71:185–189.

6. Lill SR, Parsons RH, Buhac I. Permeability of the diaphragm and fluid resorption from the peritoneal cavity in the rat. *Gastroenterology* 1979;76:997–1001.
7. Better OS, Schrier RW. Disturbed volume homeostasis in patients with cirrhosis of the liver. *Kidney Int* 1983;23:303–311.
8. Epstein FH, Lesser GT, Berger EY. Renal function in decompensated cirrhosis of the liver. *Proc Soc Exp Bio Med* 1950;75:822–824.
9. Klinger EL, Vaamonde CA, Vaamonde LS, Lancastremere RG, Morosi HJ, Frisch E, Papper S. Renal function changes in cirrhosis of the liver, a prospective study. *Arch Intern Med* 1970;125:1010–1015.
10. Wilkinson SP, Jowett TP, Slater JDH, Arroya V, Moddie H, Williams R. Renal sodium retention in cirrhosis: Relation to aldosterone and nephron site. *Clin Sci* 1979;56:169–177.
11. Wernze H, Spectz HG, Muller G. Studies on the activity of the renin-angiotensin-aldosterone system (RAAs) in patients with cirrhosis of the liver. *Klin Wochensch* 1978;56:389–397.
12. Bernardi M, Trevisani F, Santini C, De Palma R, Gasbarrini G. Aldosterone related blood volume expansion in cirrhosis before and after the early phase of ascites formation. *Gut* 1983;24:761–766.
13. Levy M. Sodium retention and ascites formation in dogs with experimental portal cirrhosis. *Am J Physiol* 1977;233F:572–585.
14. Levy M, Allotey JBK. Temporal relationships between urinary salt retention and altered systemic hemodynamics in dogs with experimental cirrhosis. *J Lab Clin Med* 1978;92:560–569.
15. Unikowsky B, Wexler MJ, Levy M. Dogs with experimental cirrhosis of the liver but without intrahepatic hypertension do not retain sodium or form ascites. *J Clin Invest* 1983;1594–1604.
16. Campbell V, Greig PD, Cranford J, Langer B, Silverman M, Blendis LM. A comparison of acute reversible pre and post sinusoidal portal hypertension on salt and water retention in the dog. *Hepatology* 1982;2:54–58.
17. Kostreva DR, Castaner A, Kampine JP. Reflex effects of hepatic baroreceptors on renal and cardiac sympathetic nerve activity. *Am J Physiol* 1980;238:R390–R395.
18. Cherrick GR, Kerr DNS, Read AE, Sherlock S. Colloid osmotic pressure and hydrostatic pressure relationships in the formation of ascites in hepatic cirrhosis. *Clin Sci* 1960;19:361–375.
19. Witte MH, Witte CL, Dumont AE. Progress in liver disease; physiological factors involved in the causation of cirrhotic ascites. *Gastroenterology* 1971;61:742–750.
20. Coppage WS, Island DP, Cooner AE, Liddle GW. The metabolism of aldosterone in normal subjects and in patients with hepatic cirrhosis. *J Clin Invest* 1962;41:1672–1680.
21. Epstein M. Renal effects of head-out water immersion in man. Implications for an understanding of volume homeostasis. *Physiol Rev* 1978;58:529–581.
22. Epstein M, Levinson R, Sancho J, Haber E, Re R. Characterization of the renin-aldosterone system in decompensated cirrhosis. *Circ Res* 1977;41:818–829.
23. Epstein M, Pins DS, Schneider N, Levinsion R. Determinants of deranged sodium and water homeostasis in decompensated cirrhosis. *J Lab Clin Med* 1976;87:822–839.
24. Nicholls KM, Shapiro MD, Kluge R, Chung HM, Bichet DG, Schrier RW. Sodium excretion in advanced cirrhosis: Effect of expansion of central blood volume and suppression of plasma aldosterone. *Hepatology* 1986;6:235–238.
25. Blendis LM, Greig PD, Langer B, Baigrie RS, Ruse J, Taylor BR. The renal and hemodynamic effects of the peritoneovenous shunt for intractable hepatic ascites. *Gastroenterology* 1979;77:250–257.
26. Greig PD, Blendis LM, Langer B, Taylor BR, Colapinto RF. Renal and hemodynamic effects of the peritoneovenous shunt. II. Long term effects. *Gastroenterology* 1981;80:119–125.
27. Epstein M, Berk DP, Hollenberg NK, Adams DF, Chalmers TC, Abrams HL, Merrill JP.

Renal failure in the patient with cirrhosis. The role of active vasoconstriction. *Am J Med* 1970;49:175–185.
28. Barnardo DE, Baldus WP, Maher FT. Effects of dopamine on renal function in patients with cirrhosis. *Gastroenterology* 1970;58:524–531.
29. Koppel MH, Coburn JW, Mims MM, Goldstein H, Boyle JD, Rubini ME. Transplant of cadaveric kidneys from patients with hepatorenal syndrome. *N Engl J Med* 1969;280:1367–1371.
30. Kew MC, Brunt PW, Varma RR, Hourigan KJ, Williams HS, Sherlock S. Renal and intrarenal blood flow in cirrhosis of the liver. *Lancet* 1971;2:504–509.
31. Ring-Larsen H. Renal blood flow in cirrhosis; relation to systemic and portal hemodynamics and liver function. *Scand J Clin Lab Invest* 1977;37:635–642.
32. Epstein M. Schneider N, Befeler B. Relationship of systemic and intrarenal hemodynamics in cirrhosis. *J Lab Clin Med* 1977;89:1175–1187.
33. Stein JH, Mauk RC, Boonjarern S, Ferris T. Differences in the effect of furosemide and chlorothiazide on the distribution of renal cortical blood flow in the dog. *J Lab Clin Med* 1972;79:995–1003.
34. Stein JH, Boonjarern S, Wilson CB, Ferris TF. Alterations in intrarenal blood flow distribution. Methods of measurement and relationship to sodium balance. *Circ Res* 1973;32–33 (Suppl. 1) 61–72.
35. DiBona GF. Neurogenic regulation of renal tubular sodium re-absorption. *Am J Physiol* 1977;233:F73–F81.
36. Ring-Larsen H, Hesse B, Henriksen JH, Christensen NJ. Sympathetic nervous activity and renal and systemic hemodynamics in cirrhosis: Plasma norepinephrine concentration, hepatic extraction and renal release. *Hepatology* 1982;2:304–310.
37. Bichet DG, VanPutten VH, Schrier RW. Potential role of increased sympathetic activity in impaired sodium and water excretion in cirrhosis. *N Engl J Med* 1982;307:1552–1557.
38. Willet I, Esler M, Burke F. Total and renal sympathetic nervous activity in alcoholic cirrhosis. *J Hepatol* 1985;1:639–648.
39. Kowalski HJ, Abelmann WH. The cardiac output at rest in Laennec's cirrhosis. *J Clin Invest* 1953;32:1025–1033.
40. Lunzer MR, Manghani KK, Newman SP, Sherlock S, Bernard AG, Ginsburg J. Impaired cardiovascular responsiveness in liver disease. *Lancet* 1975;2:382–389.
41. Finberg JPM, Syrop HA, Better OS. Blunted pressor response to angiotensin and sympathomimetic amines in the bile duct ligated dogs. *Clin Sci* 1981;61:535–359.
42. Bomzon A, Rosenberg M, Gali D, Binah O, Mordechovitz D, Better OS, Greig PD, Blendis LM. Systemic hypotension and decreased pressor response in dogs with chronic bile duct ligation. *Hepatology* 1986;6:595–600.
43. Lunzer MR, Newman SP, Sherlock S. Skeletal muscle blood flow and neurovascular reactivity in liver disease. *Gut* 1973;14:354–359.
44. Gerbes AL, Remien J, Jungst D, Saverbruch T, Daum Gartner G. Evidence for down regulation of beta-2-adrenoreceptors in cirrhotic patients with severe ascites. *Lancet* 1986;1:1409–1411.
45. Bernardi M, Trevisani F, Landini C. Plasma norepinephrine, weak neurotransmitters and renin activity during active tilting in liver cirrhosis: Relationship with cardiovascular homeostasis and renal function. *Hepatology* 1983;3:56–64.
46. Nicholls KM, Shapiro MD, VanPutten VJ. Elevated plasma norepinephrine concentrations in decompensated cirrhosis: Association with increased secretion rates, normal clearance rates and suppressibility by central blood volume expansion. *Circ Res* 1985;56:457–461.
47. Bichet D, Groves B, Schrier RW. Mechanisms of improvement of water and sodium excretion by immersion in decompensated cirrhotic patients. *Kidney Int* 1983;24:788–794.

48. Epstein M, Larios O, Johnson G. Effects of water immersion on plasma catecholamines in decompensated cirrhosis. *Min Electr Metab* 1985;11:25–34.
49. Blendis LM, Sole MJ, Campbell P, Lossing AG, Greig PD, Taylor BR, Langer B. The effect of peritoneovenous shunting on catecholamine metabolism in patients with hepatic ascites. *Hepatology* 1987;1:143–148.
50. Bichet DG, Szatalowicz V, Chaimovitz C, Schrier RW. Role of vasopressin in abnormal water excretion in cirrhotic humans. *Ann Intern Med* 1982;96:413–417.
51. Epstein M, Weitzman RE, Preston S, Denunzio AG. Role of ADH as a determinant of impaired excretion in patients with decompensated cirrhosis (abstr). *Clin Res* 1982;30:539.
52. Reznick RK, Langer B, Taylor BR, Greig PD, Blendis LM. Hyponatremia and arginine vasopressin secretion in patients with refactory ascites undergoing peritoneovenous shunting. *Gastroenterology* 1983;84:713–718.
53. Dunn MJ, Hood VL. Prostaglandins and the kidney. *Am J Physiol* 1977;233:F169–F184.
54. Lonigro AJ, Itskovitz HD, Crowshaw K, McGriff JC. Dependency of renal blood flow on protaglandin synthesis in the dog. *Circ Res* 1973;32:712–717.
55. Boyer TD, Zia PK, Reynolds TB. Effect of indomethacin and prostaglandin A on renal function and plasma renin activity in alcoholic liver disease. *Gastroenterology* 1979;77:215–222.
56. Zipser RD. The role of renal prostaglandins and the effects of nonsteroidal anti-inflammatory drugs in patients with liver disease. *Am J Med* 1986;81:(Suppl 28) 95–103.
57. Epstein M, Lifschitz M, Ramachandran M, Rappaport K. Characterization of renal prostaglandin E responsiveness in decompensated cirrhosis: Implications for renal sodium handling. *Clin Sci* 1982;62:555–563.
58. Shaw-Stiffel T, Campbell PJ, Sole MJ, Greig P, Wong PY, Blendis LM. Renal prostaglandin E2 and other vasoactive modulators in refractory hepatic ascites; response to peritoneovenous shunting. *Gastroenterology* 1988.
59. DeBold AJ, Borenstein HB, Veress AT, Sonnenberg H. A rapid and potent natriuretic response to intravenous injection of atrial myocardial extract in rats. *Life Sci* 1981;28:89–94.
60. Sonnenberg H. Mechanisms of release and renal tubular action of atrial natriuretic factor. *Fed Proc* 1986;45:2106–2110.
61. Winquist RJ, Faison EP, Walman SA, Schwartz K, Murad F, Rapoport RM. Atrial natriuretic factor elicits an endothelium-independent relaxation and activates particulate granylate cylase in vascular smooth muscle. *Proc Natl Acad Sci USA* 1984;7661–7664.
62. Maack T, Atlas SA, Camargo MJF, Cogan MG. Renal hemodynamics and natriuretic effects of atrial natriuretic factor. *Fed Proc* 1986;45:2128–2132.
63. Shenker Y, Sider RS, Ostafin A, Grekin RJ. Plasma levels of immunoreactive atrial natriuretic factor in healthy subjects and in patients with edema. *J Clin Invest* 1985;76:1681–1687.
64. Fernandez-Gruz A, Marco J, Cuadrado ML. Plasma levels of atrial natriuretic peptide in cirrhotic patients. *Lancet* 1985;1:1439–1440.
65. Gines P, Jimenez W, Narasa M, Sanz G, Tito LL, Lopez C, Rivera F, Arroya V, Rodes J. Atrial natriuretic factor (ANF) in cirrhosis plasma levels, cardiac release and splanchnic extraction. *J Hepatol* 1986;3 (Suppl.):S30.
66. Campbell PJ, Skorecki KL, Logan AG, Wong PY, Leung WM, Grieg P, Blendis LM. The acute effects of peritoneovenous shunting (PVS) on plasma atrial natriuretic peptide in cirrhotics with massive refractory ascites. *Am J Med* 1988;84:112–119.
67. Yamada S, Reynolds TB. Amiloride (MK-870). A new antikaluretic diuretic. *Gastroenterology* 1970;59:833–841.
68. Campra JL, Reynolds TB. Effectiveness of high-dose spironolactone therapy in patients with chronic liver disease and relatively refractory ascites. *Dig Dis* 1978;23:1025–1030.
69. Shear L, Ching S, Gabuzda G. Compartmentalization of ascites and edema in patients with hepatic cirrhosis. *N Engl J Med* 1970;282:1391–1396.

70. Perez-Ayuso RM, Arroyo V, Planas R, Gaya J, Bory F, Rimola A, Rivera F, Rodes J. Randomized comparative study of efficacy of furosemide versus spironolactone in nonazotemic cirrhosis with ascites. *Gastroenterology* 1983;84:961–968.
71. Ring-Larsen H, Henriksen JH, Wilken C, Clausen J, Pals H, Christensen NJ. Diuretic treatment in decompensated cirrhosis and congestive heart failure: effect of posture. *Br Med J* 1936;292:1351–1353.
72. Sherlock S, Senewirantne B, Scott A, Walker JG. Complications of diuretic therapy in hepatic cirrhosis. *Lancet* 1966;1:1049–1052.
73. Lowenthal DT, Sherar L. Use of a new diuretic agent (Metolazone) in patients with edema and ascites. *Arch Intern Med* 1973;132:38–42.
74. Kao HW, Rakov NE, Savage E, Reynolds TB. The effect of large volume paracentesis on plasma volume: A cause of hypovolemia. *Hepatology* 1985;5:403–407.
75. Simon DM, McCain JR, Bonkovsky HL, Wells JO, Hartle DK, Galambos JT. Effects of therapeutic paracentesis on systemic and hepatic hemodynamics and on renal and hormonal function. *Hepatology* 1987;7:423–429.
76. Gines P, Arroyo V, Quintero E, Planas R, Bory F, Cabrera J, Rimola A, Viver J, Camps J, Jimenez W, Mastai R, Gaya J, Rodes J. Comparison of paracentesis and diuretics in treatment of cirrhotics with tense ascites. *Gastroenterology* 1987;93:234–241.
77. Harmon DC, Demirzian Z, Ellman L, Fisher JE. Disseminated intravascular coagulation with the peritoneovenous shunt. *Ann Intern Med* 1979;90:774–779.
78. Greig PD, Langer B, Blendis LM, Taylor BR, Glynn MFX. Complications after peritoneovenous shunting for ascites. *Am J Surg* 1980;139:125–131.
79. Salem HH, Dudley FJ, Merrett A, Perkin J, Firkin BG. Coagulopathy of peritoneovenous shunt. Studies on the pathogenic role of ascitic fluid collagen and value of anti-platelet therapy. *Gut* 1983;24:412–417.
80. Blendis LM, Harrison JE, Russell KM, Miller C, Taylor BR, Greig PD, Langer B. The effects of peritoneovenous shunting on body composition. *Gastroenterology* 1986;90:127–134.
81. Blendis LM. LeVeen vs. Denver. Peritoneovenous shunts: Inadequate numbers of techniques? Mini-editorial. *Hepatology* 1987;7:405–406.
82. Crossley IR, Williams R. Spontaneous bacterial peritonitis. *Gut* 1985;26:325–331.
83. Hoefs JC, Canawti HN, Sapico FL, Hokins RR, Weiner J, Montgomerie JZ. Spontaneous bacterial peritonitis. *Hepatology* 1982;2:399–407.
84. Pinzello G, Simonetti RG, Craxi A, de Piazza S, Spano C, Pagliaro L. Spontaneous bacterial peritonitis: A prospective investigation in predominantly nonalcoholic cirrhotic patients. *Hepatology* 1983;3:545–549.
85. Bar-Meir S, Lerner E, Conn HO. Analysis of ascitic fluid in cirrhosis. *Dig Dis Sci* 1979;24:136–144.
86. Gitline N, Stauffer JL, Silverstri RC. The pH of ascitic fluid in the diagnosis of spontaneous bacterial peritonitis in alcoholic cirrhosis. *Hepatology* 1982;4:408–411.
87. Scemama-Clergue J, Doutrellet-Phillipon C, Metreau JM, Tessiere B, Capron D, Dhumeaux D. Ascitic fluid pH in alcoholic cirrhosis: A re-evaluation of its use in the diagnosis of spontaneous bacterial peritonitis. *Gut* 1985;26:332–335.
88. Garcia-Tsao G, Conn HO, Lerner E. The diagnosis of spontaneous bacterial peritonitis: Comparison of pH, lactate concentration and leucocyte count. *Hepatology* 1985;5:91–96.
89. Brook I, Altmann RS, Loebman WW, Seef LB. Measurement of lactate in ascitic fluid. An aid to the diagnosis of peritonitis with particular reference to spontaneous bacterial peritonitis of the cirrhotic. *Dig Dis Sci* 1981;26:1089–1094.
90. Runyon BA, Hoefs JC. Culture-negative neutrocytic ascites: a variant of spontaneous bacterial peritonitis. *Hepatology* 1984;4:1209–1211.

91. Cohn JN, Tristani FE, Khatri IM. Renal vasodilator therapy in the hepatorenal syndrome. *Med DC* 1970;39:1–7.
92. Fullen WD. Hepatorenal syndrome: Reversal by peritoneovenous shunt. *Surgery* 1977;82:337–341.
93. Schroeder ET, Anderson GH, Smulyan H. Effects of a portocaval or peritoneovenous shunt on renin in the hepatorenal syndrome. *Kidney Int* 1979;15:54–61.
94. Kinney MJ, Schneider A, Wapnick S, Grosberg S, LeVeen H. The hepatorenal syndrome and refractory ascites. *Nephron* 1979;23:228–232.
95. Lina S, Schaefer JW, Moore EE, Good JT, Giansiracusa R. Peritoneovenous shunt in the management of the hepatorenal syndrome. *Kidney Int* 1986;30:736–740.
96. Wood RP, Ellis D, Starzl T. The reversal of the hepatorenal syndrome in four pediatric patients following successful orthotopic liver transplantation. *Ann Surg* 1987;205:415–419.

5

Hepatic Encephalopathy

Gerald Y. Minuk

1. INTRODUCTION

Hepatic encephalopathy is a complex neurological disorder associated with advanced liver disease of either acute or chronic nature. The etiology of this disorder has confounded investigators since it was first described by Hippocrates in 460–370 B.C. One reason why the pathogenesis has remained so elusive for so long is that our understanding of normal consciousness remains relatively incomplete. Another is that much of the research past and present has been carried out in animal models of acute or, less commonly, chronic liver disease and whether the results of these studies can be extrapolated to humans remains to be determined. Finally, there is tremendous compartmentation and functional heterogeneity throughout the brain and the sophisticated techniques required to study specific regions or, more precisely, the specific synapses responsible for disturbed levels of consciousness have yet to be developed. Despite these limitations, sufficient experimental data are available to state that normal brain function in humans is dependent on at least three important factors: (1) adequate oxygenation, (2) adequate glucose or nutritional support, and (3) intact neurotransmitter systems. Thus, disturbances in any one of these three areas could conceivably result in an alteration in normal consciousness.

The purpose of this chapter is to provide an updated account of the leading hypotheses for the pathogenesis of hepatic encephalopathy. In addition, new approaches to the diagnosis and treatment of patients with hepatic encephalopathy will be discussed.

Gerald Y. Minuk • Department of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada R3A 1R9.

2. Pathogenesis

Very early results in experimental animals and humans revealed that hepatic encephalopathy could occur in the presence of adequate arterial oxygenation.[1] Similarly, early studies employing cross-circulation experimental protocols (systemic-portal versus systemic-systemic vascular shunts in liverless rats) and infusion experiments in cirrhotic patients demonstrated that neither a deficiency in glucose nor other nutritional factors produced by the liver such as uridine or cytidine could be implicated in the pathogenesis of hepatic encephalopathy.[2,3] Thus, according to these early studies, hepatic encephalopathy did not appear to be secondary to either hypoxemia or inadequate nutritional support. Rather, the agent or agents responsible appeared to be nitrogenous compounds that under normal conditions are efficiently cleared and metabolized by the liver. With these findings in mind, more recent investigators have offered a variety of suggestions as to the nature of those nitrogenous compound(s). The most popular and well studied of these are ammonia, aromatic amino acids (the false neurotransmitter hypothesis), and γ-aminobutyric acid (GABA).

2.1. The Ammonia Hypothesis

2.1.1. Homeostatic Mechanisms

Ammonia is an ubiquitous compound that is an essential nutrient and component of all proteins. Bacterial ureases (which convert urea to ammonia) and deamination of glutamine are two important mechanisms whereby ammonia is formed in the gut and then absorbed into the portal circulation.[4] Having reached the liver, ammonia then serves as a substrate in the resynthesis of urea and glutamine.[5] Urea is subsequently excreted in the urine where it constitutes 75% of the body's loss of ammonia. Smaller amounts of ammonia are lost as free ammonia in the urine and stool; an even smaller proportion is exhaled from the lungs. Ammonia in the systemic circulation contributes significantly to the process of alanine synthesis by skeletal muscle.[6] In advanced liver disease of either acute or chronic nature, ammonia extraction from the portal circulation by the liver is impaired due to either functional intrahepatic shunting of blood past hepatocytes or extrahepatic anatomical shunts through venous collaterals. The net result of the shunting process is an elevation in plasma and ultimately brain ammonia levels.[6] In cachectic patients, plasma levels are further contributed to by the decrease in ammonia extraction and alanine formation due to limited skeletal muscle mass.

2.1.2. Ammonia and Brain Function

How elevated plasma ammonia levels might interfere with central nervous system function is not entirely clear. It is known, however, that ammonia up-

take and/or formation by the brain is increased in the presence of advanced liver disease as reflected by an increase in cerebral spinal fluid glutamine levels.[5,6] It is also known that ammonia has a direct depressant effect on neuronal membrane transfer of electrolytes and water.[7] In particular, ammonia inhibits chloride ion extrusion from postsynaptic neurons causing hyperpolarization of the resting membrane potential. Ammonia also alters the cytoplasmic/mitochondrial NADH/NAD ratio and interferes with the malate–aspartate shuttle thereby decreasing cell energy stores and diminishing the concentrations of excitatory neurotransmitters (glutamate and aspartate).[7] Which, if any, of these effects are important in the pathogenesis of hepatic encephalopathy remains to be determined.

2.1.3. Supportive Evidence

The ammonia hypothesis is supported by the following findings: (1) Artificial elevations of plasma ammonia levels in both animals and humans (by the ingestion of high protein meals) or congenital abnormalities of the urea cycle which lead to significant hyperammonemia in children[8,9] reproduce many of the personality and neuromuscular manifestations of hepatic encephalopathy. (2) Repeated infusions of ammonia salts result in Alzheimer type II astrocyte changes that are similar to the histopathological findings observed in the autopsied brains of patients with chronic hepatic encephalopathy. (3) Cerebral uptake of ammonia is increased in hepatic encephalopathy as are concentrations of ammonia in the central nervous system.[6] (4) Therapeutic manipulations of the bowel known to alter ammonia absorption are often associated with an improvement in the encephalopathic state.[10,11]

2.1.4. Nonsupportive Evidence

Over the past 10 years, a sufficient body of evidence has accumulated to suggest that ammonia does not play a pivotal role in the pathogenesis of hepatic encephalopathy. Some of that evidence includes the following: (1) The finding that hepatic encephalopathy can occur in low ammonia conditions such as germ-free rats.[4] (2) Electroencephalographic abnormalities in humans or animals with hyperammonemia do not resemble those found in humans or animals with well-documented hepatic encephalopathy.[12] (3) One of the more important features of hepatic encephalopathy is the marked fluctuation in signs or symptoms and the total reversibility of the disorder once hepatic function is restored. Ammonia, on the other hand, is a dose-dependent neurotoxin that results in nonfluctuant and, if long-standing, irreversible impairment of central nervous system function.[13] (4) Pure acute ammonia intoxication as seen in children with defects in intermediary metabolism is a hyperkinetic excitatory state often leading to convulsions, whereas hepatic encephalopathy is typically a hypokinetic state

characterized by drowsiness, somnolence, and coma.[14] (5) Finally, blood ammonia concentrations do not correlate particularly well with the existence or severity of hepatic encephalopathy. Indeed, plasma ammonia levels can increase or remain unchanged while the patient's clinical status improves. Alternatively, plasma levels might decrease while the patient's clinical status deteriorates.[15]

Thus, current evidence indicates that a plasma ammonia disturbance is more likely to be indicative of deranged nitrogen metabolism than have etiological importance in the pathogenesis of hepatic encephalopathy.

2.2. False Neurotransmitter Hypothesis

2.2.1. Homeostatic Mechanisms

The branched chain amino acids consist of leucine, isoleucine, and valine. The aromatic amino acids are tyrosine, phenylalanine, and tryptophan. Branched chain amino acids are normally metabolized by skeletal muscle (in the presence of insulin). Their uptake by the liver is limited. Aromatic amino acids, on the other hand, are rapidly and efficiently cleared by the liver and metabolized within hepatocytes to form glucose. Both branched chain and aromatic amino acids compete at the level of the blood–brain barrier for access into the brain via the same neutral amino acid transport system.[16]

The paradoxical finding of elevated serum insulin and glucagon levels is common in patients with advanced liver disease.[17] One of the many effects of the hyperinsulinemic component of the hormonal disturbance is a significant increase in branched chain amino acid uptake by skeletal muscle, and thus a decrease in their serum concentrations. Conversely, the hyperglucagonemia results in enhanced aromatic amino acid release from skeletal muscle, and thus an increase in their serum concentrations. At the same time, hepatic clearance of aromatic amino acids is impaired, resulting in a further increase in their serum concentrations.[18,19] The net result of these changes is a marked increase in aromatic amino acid uptake by the neutral amino acid transport system at the blood–brain barrier.

2.2.2. Aromatic Amino Acids and Brain Function

Within the brain, increased levels of phenylalanine impair the normal conversion that takes place from tyrosine to dopamine and norepinephrine (''true'' neurotransmitters). Instead, tyrosine is decarboxylated to tyramine and then further shunted toward the formation of octopamine, a relatively weak neurotransmitter compound.[20] The false neurotransmitter hypothesis suggests that elevated brain octopamine levels successfully compete with dopamine and

norepinephrine for their respective neuronal receptor sites and thus decrease "true" neurotransmission.[16] The increased levels of tryptophan in the brain may independently contribute to the encephalopathic state by virtue of its role as a precursor to serotonin which has inherent neuroinhibitory properties.[17]

2.2.3. Supportive Evidence

Support for the false neurotransmitter hypothesis is derived from the following findings. First, the amino acid profile discussed previously (elevated aromatic amino acids and decreased branched chain amino acids) is common in patients with hepatic encephalopathy.[20] Second, data in experimental animals with coma generally show a decrease in norepinephrine and an increase in brain serotonin levels.[21] Third, pharmacological experiments in normal animals suggest that increased serotonin-receptor binding is associated with drowsiness and sleep, whereas decreased binding is associated with arousal.[22]

2.2.4. Nonsupportive Evidence

Unfortunately for proponents of the false neurotransmitter hypothesis, artificially induced increases in brain octopamine levels to values 200,000 times normal have not been associated with a decreased level of consciousness, nor has norepinephrine and dopamine depletion in the brains of experimental animals been documented.[23] Moreover, norepinephrine and dopamine levels have been normal in most human autopsy studies.[24] Finally, controlled clinical trials of L-dopa, a precursor of dopamine, have failed to consistently demonstrate a beneficial therapeutic response to this agent in patients with a chronic hepatic encephalopathy.[25,26]

2.3. GABA Hypothesis

γ-Aminobutyric acid (GABA) was first formally proposed as an etiological agent in the pathogenesis of hepatic encephalopathy by Schafer and Jones in 1982.[27] Their hypothesis was based on similarities between the electroencephalographic (EEG) findings of hepatic encephalopathy and the EEG changes observed when agents that bind to GABA receptor sites were administered to experimental animals.[28] Moreover, by this time, GABA was widely acknowledged as the principal inhibitory neurotransmitter in the mammalian brain with approximately 25–45% of neuronal activity in the brain having been identified as being under GABAergic control mechanisms.[29] Because the GABA hypothesis is the most recently advanced of the three hypotheses under discussion, it will be considered in somewhat more detail than the others.

2.3.1. Homeostatic Mechanisms

2.3.1a. Sources of Systemic GABA. Systemic GABA, like ammonia and the aromatic amino acids, is largely gut-derived.[30,31] Both aerobic and anaerobic bacteria of the intestine produce amounts of GABA that are proportionate to the luxuriance of their growth. Recently, it was also demonstrated that the intestinal wall is capable of producing large amounts of GABA, a finding that may be of particular relevance to the encephalopathy seen in germ-free rats with liver failure.[32]

2.3.1b. Systemic GABA Clearance. The liver is the organ primarily responsible for GABA clearance and metabolism.[33,34] The kidney contributes only minimally and in a limited number of animal species.[33] Portal venous GABA concentrations are approximately twice that of systemic blood.[30,35] A specific GABA transport system has been identified on the surface of isolated hepatocytes that was not present on either Kupffer cells or endothelial cells of the liver.[34] This transport system is sodium- and pH-dependent[35] with decreased GABA uptake in the presence of hyponatremia and systemic alkalosis (two metabolic disturbances commonly found in patients with hepatic encephalopathy).[36]

2.3.1c. Serum GABA Levels in Liver Failure. Three clinical studies have demonstrated an elevation in serum GABA concentrations in the presence of encephalopathy and advanced liver disease.[37–39] In the first, which included patients with acute or chronic liver disease, GABA concentrations were markedly elevated and significantly higher than those described subsequently in the two other reports. In that study, however, serum GABA levels were determined by a radioreceptor, competitive inhibition assay.[37] It is conceivable that other amino acids that bind to the same GABA receptor site would artificially increase the recorded GABA concentrations. Indeed, preliminary data would suggest that this was the case.[40]

In the remaining two studies, one in adults with advanced chronic liver disease[38] and the other in young children with Reye's syndrome,[39] serum GABA levels were 5–20 times the upper limit of normal, but were within the normal range in patients with liver disease and no evidence of hepatic encephalopathy. In each of these studies, serum GABA concentrations were measured by ion exchange chromatography with fluorometric enhancement.[41] Determinations of serum GABA concentrations in animal models of hepatic encephalopathy have led to conflicting results with some studies reporting increased values while others report no change.[40–42]

2.3.1d. GABA Access to the Brain. Based on experiments in which bolus infusions of GABA were administered intravenously into the systemic circula-

tion of healthy experimental animals, it is generally believed that little if any GABA from the systemic circulation gains access into the central nervous system. Yet when GABA is administered by a constant intravenous infusion in healthy dogs (an animal with cerebral blood flow that more closely approximates that of man), a stepwise increase in cerebral spinal fluid GABA concentrations has been observed.[43] The amounts of GABA in the systemic circulation required to achieve these results, however, far exceed the levels reported in experimental animals or humans with hepatic encephalopathy.

More recent studies suggest that in the presence of advanced liver disease, systemic GABA entry into the central nervous system is significantly enhanced. In one such study, using the galactosamine model of acute fulminant hepatitis in rabbits, cerebral uptake of the slowly metabolized, radiolabeled, GABA isomer [^{14}C]α-aminoisobutyric acid was increased 5- to 10-fold in gray matter areas when compared to healthy controls.[44] In other experiments, intracarotid infusions of radiolabeled GABA in the same animal model were associated with significant increases in the calculated brain uptake index in all regions of the brain studied.[45]

2.3.2. GABA and Brain Function

At least two specific binding sites exist for GABA on the GABA receptor complex: a high- and a low-affinity binding site.[46] The low-affinity site appears to be the more physiologically relevant of the two.[47] The role of the high-affinity site (if any) has yet to be determined.

γ-Aminobutyric acid exerts its neuroinhibitory effects in the central nervous system by increasing neuronal membrane permeability to chloride ions through chloride ionophores located on postsynaptic membranes.[28] This results in increased chloride influx and hyperpolarization of the resting cell membrane potential. The hyperpolarized cell is thereby rendered refractory to subsequently conducted action potentials. Benzodiazepines and barbiturates also bind to specific sites on the GABA receptor complex, resulting (in the case of benzodiazepines) in an increase in the frequency that the chloride channel is opened or (in the case of barbiturates) an increase in the duration that the channel remains open.[28] Neither benzodiazepines nor barbiturates exert these effects in the absence of GABA.

In animal models of hepatic encephalopathy, various claims have been made regarding the status of these macromolecular GABA receptors. For example, some investigators contend that at some time during the encephalopathic process, perhaps even prior to the onset of encephalopathy, the density (but not affinity) of high- and/or low-affinity GABA receptors in the brain is increased.[48–50] The increase in GABA receptor density provides an attractive explanation for the clinical observation that patients with hepatic encephalopathy

are unusually sensitive to the effects of benzodiazepine and barbiturate compounds.[51] Conversely, others have been unable to demonstrate such changes in GABA receptor density.[49] This conflict may in part be resolved by recent findings that indicate that anxiety and/or physical handling of experimental animals prior to sacrifice may cause GABA receptor density changes similar to those previously described in encephalopathic animals.[52]

2.3.3. Supportive Evidence

Support for the GABA hypothesis can be derived from the following findings: (1) GABA is synthesized by gut bacteria as well as the intestinal wall (which, as discussed earlier, may be of particular importance in the explanation of the encephalopathy that accompanies liver failure in germ-free animals). (2) The liver is the organ primarily responsible for systemic GABA homeostasis. (3) Serum concentrations of GABA are increased in humans with hepatic encephalopathy. (4) GABA entry into the central nervous system is enhanced in liver failure, and (5) GABA receptor changes may occur during the course of hepatic encephalopathy. (6) small amounts of GABA, when instilled into the brains of experimental animals, result in clinical and electroencephalographic changes compatible with the encephalopathic state; (7) the administration of GABA receptor agonists induce the same visual evoked patterns as those seen in animals or humans with hepatic encephalopathy[53,54]; and (8) the results of new therapeutic approaches for hepatic encephalopathy wherein GABA receptor antagonists have been used therapeutically with some initial success[55–58] (to be discussed in Section 4).

2.3.4. Nonsupportive Evidence

Not all studies have provided support for the GABA hypothesis. For example, in one, rats treated with galactosamine did not show an increase in endogenous blood GABA levels (although hepatic GABA clearance was impaired).[42] Moreover, following intravenous GABA administration, concentrations of GABA in the brain were lower in comatose than in the control rats. Recent work in humans suggests that elevated serum GABA concentrations are not a uniform finding.[59] In other studies, brain GABA concentrations in animals and humans who had died with acute or chronic liver failure and encephalopathy were either normal or even below normal when compared to healthy controls.[60–62] Finally, instillation of muscimol (a specific GABA receptor agonist) into the cerebral ventricles of rats with chronic portal caval shunts did not induce coma (although it did hasten the development of encephalopathy in rats with acute liver damage following portal caval shunt plus hepatic artery ligation).[63] The same study also showed that in the absence of muscimol, neither

bicuculline nor picrotoxin (GABA receptor and chloride ionophore antagonists, respectively) influenced the course of encephalopathy in animals with acute hepatic failure.

In short, sufficient nonsupportive or contradictory evidence has accrued to suggest that much further work is required before the GABA hypothesis can be accepted as being relevant to the pathogenesis of hepatic encephalopathy.

3. DIAGNOSIS

3.1. Clinical Features

There are numerous clinical findings associated with the encephalopathic state, none of which are specific for hepatic encephalopathy.[36] Perhaps the most common findings are hypersomnolence, disturbed level of consciousness, reversal of normal sleep patterns, decrease in spontaneous movements, fixed stare, and apathy. Personality changes may also occur, with some patients occasionally appearing childish or extremely irritable. Many of these findings, particularly those associated with personality changes, are relatively early signs of hepatic encephalopathy and may only be apparent to the patient's family or close friends. As the encephalopathy becomes more advanced, however, the signs become more marked, confusion develops, and the encephalopathy becomes apparent to even the casual observer. By this time, abnormalities in spatial gnosis, as reflected by construction apraxia, are also evident. Other common physical signs include fetor hepaticus (an odd, sweetish smell of the breath thought to be secondary to increased mercaptan concentrations), hyperactive stretch reflexes, paratonia, and, ultimately, decerebrate posturing. The speech and cadence of the encephalopathic patient is slow, slurred, and monotonous. Asterixis, an often sought and important clinical feature of hepatic encephalopathy, is tested by fixed extension of the forearms, dorsiflexion of the wrists, and fanning of the digits with eyes opened and then closed. A rapid, forward flexion of the hands with further fanning of the fingers reflects decreased afferent (proprioceptive) input to the reticular activating system. When arms or hands are not available or when it is inappropriate to test these areas, as in the case of trauma to these regions, the neck, tongue, or eyelids can be used in a similar manner. Once the encephalopathy has advanced to the stage of coma, asterixis, like many of the other motor abnormalities, is no longer present. The Glasgow coma scale is used by many physicians to describe the more advanced clinical stages of hepatic encephalopathy.[64]

Often patients with advanced liver disease present sufficiently early in their course that symptoms or signs of encephalopathy are not yet apparent to any observer. In this setting, more sensitive tests such as psychometric testing,

standard electroencephalograms, or visual evoked potential responses may be useful diagnostic aids.

3.2. Laboratory Tests

There is no one laboratory test that is specific for the diagnosis of hepatic encephalopathy. Rather, a combination of the mode of presentation, clinical features, and the presence of certain laboratory findings are usually required. The laboratory findings that are most useful in establishing the diagnosis of hepatic encephalopathy and at the same time correlate well with the clinical course of the patient are an increase in glutamine and ketoglutarate levels in the cerebral spinal fluid.[65] However, these tests are rarely carried out due in part to concerns regarding the performance of a lumbar puncture in patients with potentially severe bleeding diatheses and/or increased intracranial pressure. More readily available, but at the same time less helpful, are a series of systemic biochemical abnormalities that are less specific and correlate better with the extent of liver disease than the stage of encephalopathy. These include elevated plasma ammonia, aromatic amino acids, short-chain fatty acids, and mercaptan levels.[24] A decrease in branched chain amino acid concentrations also falls into this category.[24]

Recently, much interest has been generated by the use of various psychometric tests in diagnosing early (preclinical) hepatic encephalopathy.[66–68] These tests have also been found to be useful in following the course of the disease, a feature that is becoming increasingly important as new therapeutic modalities are being considered. Of the many psychometric tests that can be applied, the Reitan trail-making or number connection test has achieved the most popularity.[69] In this test, patients are timed as they connect a series of numerical dots. The number patterns of the dots are changed frequently in order to minimize the learning effect of repetitive testing. The test is quick, inexpensive, and accurate. It can be carried out in the hospital or office. It is more sensitive than the majority of other psychometric tests including the five-pointed-star construction test, signature test, or subtraction-of-serial-sevens test and is easier to quantitate. Unfortunately, like the biochemical markers, the test suffers from a lack of specificity.

The EEG abnormalities observed in patients with hepatic encephalopathy are also nonspecific. Similar if not identical changes are seen in patients with encephalopathy secondary to advanced renal disease or other severe metabolic disturbance. Nonetheless, the presence of low-frequency, high-amplitude waveforms in a patient with advanced liver disease, a decreased level of consciousness, and no alternative explanation for the latter finding is strong supportive evidence for the diagnosis of hepatic encephalopathy.[70,71] Moreover, the extent of the abnormality on EEG correlates well with the patient's clinical status.[72]

Thus, serial EEGs may be useful in following the course of the patient's illness.

Although problems with specificity occasionally arise, EEGs are relatively sensitive tests of cognitive function. Indeed, abnormalities compatible with hepatic encephalopathy may be present prior to the appearance of clinical features and may remain abnormal once clinical features have resolved.[72]

EEG tracings may also help identify alternative causes of a decreased level of consciousness such as seizure activity or changes suggestive of a space-occupying lesion.

Repeated stimulation of the visual cortex with controlled flashes of light elicits both excitatory and inhibitory neuronal activity within that region of the brain. A visual evoked response is a computerized analysis of brain activity in the visual cortex that averages these responses to provide a specific waveform with both positive and negative inflections[73]. Disturbances in the character and dynamics of the waveform may occur early (prior to the onset of clinical features) and seem to correlate well with the clinical stage of encephalopathy.[74,75] Although the pattern of the visual evoked response is not entirely specific for hepatic encephalopathy, it is more so than an EEG. Other conditions that can produce similar wave disturbances include postictal somnolence and the administration of benzodiazepines or barbituates.[76] Indeed, the changes associated with these disorders are so similar to hepatic encephalopathy that a common pathogenetic mechanism has been suggested. Somatosensory and auditory evoked responses can be used in a similar manner.

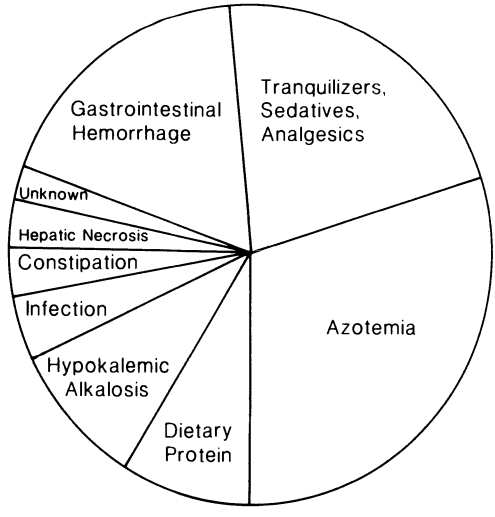


Figure 1. Common precipitating causes of hepatic encephalopathy.

3.3. Differential Diagnosis

The following disorders can be associated with clinical features that are indistinguishable from hepatic encephalopathy: uremia, respiratory failure, congestive heart failure, and hyponatremia.[36] Fortunately, laboratory and/or radiological investigations will often identify their presence. Other disorders that can mimic hepatic encephalopathy include drug overdose, Wernicke's encephalopathy, the presence of a subdural hematoma, occasionally alcohol withdrawal, and hypoglycemia. Here again, a high index of suspicion is often all that is required to exclude these possibilities.

4. TREATMENT

After ensuring that glucose levels and blood oxygen saturation are adequate, the treatment of hepatic encephalopathy consists of a search for precipitating factors, minimizing protein absorption from the gut, and supporting or maximizing liver function.

The three most common precipitating factors of hepatic encephalopathy are prerenal azotemia (secondary to aggressive use of diuretics and/or persistent vomiting/diarrhea); the injudicious use of tranquilizers, sedatives, or analgesics; and overt or covert gastrointestinal hemorrhage (either variceal, peptic ulceration, or gastritis)[77] (Fig. 1). Other precipitating factors include normal or excess dietary protein intake, hypokalemic alkalosis (often secondary to diuretics), bacterial infection (pneumonia and bacterial peritonitis are especially common), and prolonged constipation.[77]

The removal of proteinaceous compound(s) from the gut is an important aspect of treatment for hepatic encephalopathy.[24,36] This can be achieved by (1) reducing (when the liver failure is chronic) or eliminating (when acute) protein from the diet[78–80]; (2) administering nonabsorbable antibiotics, such as neomycin, that are effective against intestinal bacteria[81,82]; and (3) using purgatives such as lactulose (at a dose necessary to achieve two or three bowel actions daily)[83–86] or magnesium sulfate/lactulose enemas.[87,88]

Whether the coadministration of neomycin and lactulose has a synergistic or detrimental effect on the course of encephalopathy remains unclear.[89,90] In contrast to theoretical concerns regarding what effect neomycin will have on the bacterial flora required for lactulose metabolism, at least one clinical study has shown a significant improvement in plasma ammonia concentration when both agents were administered simultaneously as compared to when provided on an individual basis.[91] Testing of stools to ensure that the pH is less than 5 may help in identifying patients in whom combination therapy is unlikely to be of benefit.[11,90,92]

Initial enthusiasm was generated by the results of early uncontrolled stud-

ies in which branched chain amino acids were administered either parenterally or enterally to patients with hepatic encephalopathy.[93–95] Subsequently, prospective, randomized, controlled trials, with the possible exception of studies using ketoanalogs, were unable to confirm these early findings.[26,96] Although not effective for hepatic encephalopathy, the use of branched chain amino acids is safe and beneficial (in terms of nutrition) in the setting of an encephalopathic patient with severe malnutrition who requires protein calories.

Controlled clinical trials have failed to demonstrate that dopaminergic drugs such as L-dopa and bromocriptine consistently improve the encephalopathic state.[25,26] The use of these agents should therefore be restricted to the rare patient who is refractory to all other measures and is not a candidate for liver transplantation.

Metronidazole, a potent antibiotic against gram-negative anaerobic bacteria of the gut, is as effective as neomycin in treating patients with mild to severe hepatic encephalopathy.[97,98] The two agents may even have a synergistic effect when used together.[97] Metronidazole should generally be used in the setting of chronic encephalopathy where the nephrotoxic and ototoxic potential of neomycin preclude the use of this agent beyond 5 days and lactulose alone is insufficient to maintain a functional sensorium.

Sorbitol and lactitol are two synthetic compounds that are better tolerated than lactulose, more predictable in their effect, and cause less osmotic diarrhea and hypernatremia. Preliminary studies in 40 cirrhotic patients with acute episodes of hepatic encephalopathy indicate that these agents may be as effective as lactulose in improving the clinical grade of encephalopathy, EEG grade, number connection times, and venous blood ammonia concentrations.[99] Lactitol enemas were also shown to be as effective as lactulose enemas in the treatment of 45 episodes of acute hepatic encephalopathy.[87] Of interest, the same study reported that tap water enemas had no effect on the mental state grade, number connection test, asterixis grade, plasma ammonia level, or EEG grade.

Sodium benzoate, sodium phenylacetate, and zinc all increase nitrogen excretion. Zinc also alters ammonia and GABA metabolism. When sodium benzoate and sodium phenylacetate were given in a double-blind crossover fashion to eight male patients with chronic hepatic encephalopathy, all patients either improved or maintained their mental status comparable to that achieved with conventional therapy.[100] Moreover, blood ammonia levels decreased and the encephalopathy index improved in all but one treated patient. Similarly, in eight cirrhotic patients with chronic hepatic encephalopathy, oral zinc therapy was associated with a significant improvement in trail-making test results while no improvement was seen in 12 placebo recipients.[101] Whether the positive results obtained with these agents will be confirmed by future studies (which hopefully will involve larger patient populations) remains to be determined.

Perhaps the most exciting new development in the treatment of hepatic encephalopathy is the use of benzodiazepine antagonists, such as Ro 15-1788

and CGS-8216, which interfere with the binding of endogenous, benzodiazepinelike ligands to their receptor sites on the GABA receptor complex. These agents have been shown to reverse hepatic encephalopathy in rabbits[55] and rats.[56] They have also been reported to induce marked and rapid, albeit transient, improvements in humans with chronic liver disease and encephalopathy.[57,58,102] The results of controlled clinical trials with these agents are eagerly awaited.

5. SUMMARY

In 1978 it was stated that "when one has finished reviewing all of the postulated mechanisms for the biochemical pathogenesis of portasystemic encephalopathy, one is left with no completely satisfying answer and a number of quite unsatisfactory ones." Although much of this statement remains true today, our understanding of the pathogenesis of hepatic encephalopathy is certainly more complete now than it was ten years ago, as is our ability to diagnose and treat patients with this disorder. Specifically, the GABA hypothesis has led to a more complete understanding of normal brain physiology in both animals and humans. It has also resulted in extensive documentation of the effects of advanced liver disease on amino acid neurotransmitter mechanisms.

With respect to developments in diagnosis, the Reitan trail-making test and visual evoked responses have evolved as useful and convenient tests in the diagnosis and monitoring of encephalopathic patients. Regarding treatment, the emergence of new agents designed to reverse the underlying pathophysiological disturbance in hepatic encephalopathy marks the beginning of a new and exciting era of providing specific treatment for this potentially lethal disorder.

ACKNOWLEDGMENTS. The author wishes to acknowledge the support of the Manitoba Medical Services Foundation and the Medical Research Council of Canada. Ms. D. Byron is to be thanked for her prompt and accurate typing of the manuscript.

REFERENCES

1. Maiolo AT, Bianchi Porro G, Galli C et al. Brain energy metabolism in hepatic coma. *Adv Exp Biol Med* 1971;4:52–70.
2. Geiger A, Yamasaki S. Cytidine and uridine requirement of the brain. *J Neurochem* 1956;1:93–100.
3. Shafer WH, Isselbacher KJ. Uridine metabolism in chronic liver disease. *Gastroenterology* 1961;40:782–784.

4. Nance FV, Batson RC, Kline DG. Ammonia production in germ-free ECK fistula dog. *Surgery* 1971;70:169–174.
5. Summerskill WHJ, Wolfe SJ. The metabolism of ammonia and αketo acids in liver disease and hepatic coma. *J Clin Invest* 1957;36:361–372.
6. Lockwood AH, McDonald JM, Reiman RE, Gelbard AS, Laughlin JS, Duffy TE, Plum F. The dynamics of ammonia metabolism in man. Effect of liver disease and hyperammonemia. *J Clin Invest* 1979;63:449–460.
7. Hindfelt B, Plum F, Duffy TE. Effect of acute ammonia intoxication or cerebral metabolism in rat with portacaval shunts. *J Clin Invest* 1977;59:386–396.
8. Conn HO, Lieberthal MM. *The Hepatic Coma Syndromes and Lactulose*. Baltimore: Williams and Wilkins, 1979.
9. Flannery DB, Hsia YE, Wolf B. Current status of hyperammonemic syndromes. *Hepatology* 1982;2:495–506.
10. Ficher CJ, Faloon WW. Blood ammonia levels in hepatic cirrhosis: Their control by oral administration of neomycin. *N Engl J Med* 1957;256:1030–1035.
11. Crossley IR, Williams R. Progress in the treatment of chronic portasystemic encephalopathy. *Gut* 1984;25:85–98.
12. Pappas SC, Ferenci P, Schafer DF, Jones EA. Visual evoked potentials: hepatic encephalopathy resembles the post-ictal state but differs from hyperammonemia. *Hepatology* 1982 (abstract) 2:708.
13. Plum F, Hindfelt B. The neurological complication of liver disease. In: *Metabolic and Deficiency Diseases of the Nervous System,* Part I (Vinken PJ, Bruyn GW, eds). New York: Elsevier, 1976, pp. 346–377.
14. Zieve L. The mechanism of hepatic coma. *Hepatology* 1981;1:360–365.
15. Zieve L, Nicoloff DM. Pathogenesis of hepatic coma. *Ann Rev Med* 1975;26:143–157.
16. Fischer JE, Baldessarini RJ. False neurotransmitters and hepatic failure. *Lancet* 1971;2:75–79.
17. Soeters P, Fischer JE. *Lancet* 1976;2:880–882.
18. Vilstrup H, Bucher D, Krog B, Damgard SE. Amino acid clearance in cirrhosis. *Eur J Clin Invest* 1979;12:197.
19. Fossi-Fanelli F, Angelico M, Cangiano C, Cascino A, Capocaccia R, Deconciliis D, Riggio O, Capocaccia L. Effect of glucose and/or branched chain amino acid infusion on plasma amino acid inbalance in chronic liver failure. *J Parenter Enteral Nutr* 1981;5:414.
20. James JH, Ziparo V, Jepson B, Fischer JE. Hyperammonaemia plasma amino acid inbalance and blood–brain aminoacid transport: A unified theory of portal-systemic encephalopathy. *Lancet* 1979;2:772–775.
21. Zieve L, Olsen EL. Can hepatic coma be caused by a reduction of brain nonadrenaline or dopamine? *Gut* 1977;18:688–691.
22. Curzon G, Knott PJ. Environmental, toxicological and related aspects of tryptophan metabolism with particular reference to the central nervous system. *CRC Crit Rev Toxicol* 1977; September:145–187.
23. Fischer JE, Baldessarini RJ. Pathogenesis and therapy of hepatic coma. In: *Progress in Liver Diseases,* Vol 5 (Popper H, Schaffner F, eds). New York: Grune and Stratton, 1975, pp. 363–387.
24. Hoyumpa AM, Jr, Schenker S. Perspectives in hepatic encephalopathy. *J Lab Clin Med* 1982;100:477–487.
25. Michel H, Solere M, Granier P et al. Treatment of cirrhotic hepatic encephalopathy with L-dopa. A controlled trial. *Gastroenterology* 1980;79:207–211.
26. Uribe M, Farca A, Maquez MA et al. Treatment of chronic portal systemic encephalopathy with bromocriptine. A double blind controlled trial. *Gastroenterology* 1979;76:1347–1351.

27. Schafer DF, Jones EA. Hepatic encephalopathy and γ-amino-butyric acid neurotransmitter system. *Lancet* 1982;1:18–20.
28. Schafer DF, Jones EA. Potential neural mechanisms in the pathogenesis of hepatic encephalopathy. In: *Progress in Liver Diseases,* Vol 7 (Popper H, Schaffner F, eds). Orlando: Grune and Stratton, 1982, pp. 615–627.
29. Hoyumpa AM. The unfolding GABA story. *Hepatology* 1986;6(5):1042–1044.
30. Schafer DF, Fowler JM, Jones EA. Colonic bacteria: A source of γ-aminobutyric acid in blood. *Proc Soc Exp Biol Med* 1981;167:301–303.
31. Minuk GY. Gamma-aminobutyric acid (GABA) production by eight common bacterial pathogens. *Scand J Infect Dis* 1986;18:465–467.
32. van Berlo CLH, deJonge HR, van den Bogaard AEJM, van Eijk HMH, Janssen MA, Soeters PB. γ-Aminobutyric acid production in small and large intestine of normal and germ-free Wistar rats. *Gastroenterology* 1987;92:472–479.
33. Ferenci P, Schafer DF, Schrager R, Jones EA. Metabolism of the inhibitory neurotransmitter γ-aminobutyric acid in a rabbit model of acute hepatic failure. *Hepatology* 1981;1:509–512.
34. Minuk GY, Vergalla J, Ferenci P, Jones EA. Identification of an acceptor system of gamma-aminobutyric acid on isolated rat hepatocytes. *Hepatology* 1984;4:180–185.
35. Minuk GY, Sarjeant EJ. The effect of (a) neomycin and lactulose treatment on systemic and portal serum GABA levels in rats and (b) pH changes on [^{3}H]GABA binding to isolated rat hepatocytes. Proceedings of Sixth International Symposium on Ammonia, Amino Acids and Hepatic Encephalopathy, Maastricht, The Netherlands (in press).
36. Zieve L. Hepatic encephalopathy. In: *Diseases of the Liver,* 5th ed (Schiff L, Schiff ER, eds). Philadelphia: J.B. Lippincott, Co., 1982, p. 433.
37. Ferenci P, Schafer DF, Kleinberger G, Hoofnagle JH, Jones EA. Serum levels of gamma-aminobutyric acid-like activity in acute and chronic hepatocellular diseases. *Lancet* 1983;1:811–814.
38. Minuk GY, Winder A, Burgess ED, Sarjeant EJ. Serum gamma-aminobutyric acid (GABA) levels in patients with hepatic encephalopathy. *Hepatolgastroenterology* 1985;32:171–174.
39. Minuk GY, Sarjeant EJ, Buchan K. Elevated serum γ-amino-butyric acid levels in children with Reye's syndrome. *J Paed Gastroenterology Nutr* 1985;4:528–531.
40. Maddison JE, Dodd PR, Morrison M, Johnston GAR, Farrell GC. Plasma GABA, GABA-like activity and the brain GABA benzodiazepine receptor complex in rat with chronic hepatic encephalopathy. *Hepatology* 1987;7:621–628.
41. Hare TA, Manyam NVB. Rapid and sensitive ion-exchange fluorometric measurement of γ-aminobutyric acid in physiological fluids. *Anal Biochem* 1980;101:349–355.
42. Zeneroli ML, Iuliano E, Racagni G et al. Metabolism and brain uptake of γ-minobutyric acid in galactosamine-induced hepatic encephalopathy in rats. *J Neurochem* 1982;38:1219–1222.
43. Loscher W, Frey HH. Transport of GABA at the blood–CSF interface. *J Neurochem* 1982;38:1072–1079.
44. Horowitz ME, Schafer DF, Molnar P et al. Increased blood brain transfer in a rabbit model of acute liver failure. *Gastroenterology* 1983;84:1003–1011.
45. Bassett ML, Mullen KD, Scholz B et al. Increased brain GABA uptake in pre-coma encephalopathy in a rabbit model of fulminant hepatic failure. *Hepatology* 1985 (abstract); 5:1032.
46. Tallman JF, Gallager DW. The GABAergic system: A locus of benzodiazepine action. *Ann Rev Neurosci* 1985;8:21–44.
47. Harris AR, Allan AM. Functional coupling of γ-aminobutyric acid receptors to chloride channels in brain membranes. *Sciences* 1985:228:1108–1110.
48. Jones EA, Schafer DF, Ferenci P et al. The neurobiology of hepatic encephalopathy. *Hepatology* 1984;4:1235–1242.

49. Baraldi M, Zeneroli ML. Experimental hepatic encephalopathy: Changes in the binding of γ-aminobutyric acid. *Sciences* 1982;216:427–428.
50. Zeneroli ML, Baraldi M, Pinelli G et al. Brain receptor changes in portal-systemic encephalopathy in dogs. *Hepatology* 1985 (abstract);5:953.
51. Jones EA. Hepatic encephalopathy update. *Viewpoint Dig Dis* 1986;18:1–4.
52. Maddison JE, Dodd PR, Johnston GAR, Farrell GC. Brain γ-aminobutyric acid receptor binding is normal in rats with thioacetamide-induced hepatic encephalopathy despite elevated plasma γ-aminobutyric acid-like activity. *Gastroenterology* 1987;93:1062–1068.
53. Zeneroli ML, Ventura E, Baraldi M et al. Visual evoked potentials in encephalopathy induced by galactosamine, ammonia, dimethyldisulfide and octanoic acid. *Hepatology* 1982;2:532–538.
54. Pappas SC, Jones EA. Methods for asserting hepatic encephalopathy. *Semin Liver Dis* 1983;3:298–307.
55. Bassett ML, Mullen KD, Skolnick P et al. GABA and benzodiazepine receptor antagonist ameliorate hepatic encephalopathy in a rabbit model of fulminant hepatic failure. *Hepatology* 1985(abstract); 5:1032.
56. Baraldi M, Zeneroli ML, Ventura E et al. Supersensitivity of benzodiazepine receptors in hepatic encephalopathy due to fulminant hepatic failure in the rat: reversal by a benzodiazepine antagonist. *Clin Sci* 1984;67:167–175.
57. Scollo-Lavizzari G, Steinmann E. Reversal of hepatic coma by benzodiazepine antagonist (Ro 15-1788). *Lancet* 1985;1:1324.
58. Bansky G, Meier PJ, Ziegler WH et al. Reversal of hepatic coma by benzodiazepine antagonist (Ro 15-1788). *Lancet* 1985;1:1324–1325.
59. Moroni F, Riggio O, Carla V, Festuccia V et al. Hepatic encephalopathy: Lack of changes of γ-aminobutyric acid content in plasma and cerebrospinal fluid. *Hepatology* 1987;7:816–820.
60. Pomier-Layrargues, Bories P, Mirouze D, Giordan J, Feneyrou B, Bellet-Hermann H, Marchal G, Michel H. Brain gamma amino-butyric acid in acute hepatic encephalopathy in dogs following hepatectomy with or without abdominal evisceration. In: *Hepatic Encephalopathy in Chronic Liver Failure* (Capocaccia L, Fischer JE, Rossi-Fanelli F, ed). New York: Plenum Press, 1984, pp. 127–134.
61. Lavoie J, Giguere JF, Layrargues GP, Butterworth RF. Amino acid changes in autopsied brain tissue from cirrhotic patients with hepatic encephalopathy. *J Neurochem* 1987;49:692–697.
62. Butterworth RF, Giguere JF. Cerebral aminoacids in portasystemic encephalopathy: Lack of evidence for altered γ-aminobutyric acid (GABA) function. *Metabol Brain Dis* 1986;1:221–228.
63. Rzepczynski D, Zieve L, Lindblad S, LaFontaine D. In vivo studies of GABAergic effect in experimental hepatic encephalopathy. *Hepatology* 1986;6:902–905.
64. Levy DE, Bates D, Caronna JJ et al. Prognosis in non-traumatic coma. *Ann Intern Med* 1981;94:293–301.
65. Giguere JF, Butterworth RF. Amino acid changes in regions of the CNS in relation to function in experimental portal-systemic encephalopathy. *Neurochem Res* 1984;9:1307–1319.
66. Gilbertstadt S, Gilbertstadt H, Buegel B, Collier R, McClain CF, Zieve L. Defective intellectual function in alcoholic cirrhotics and non-cirrhotic alcoholics: Relationship to severity of liver disease. *Gastroenterology* 1978 (abstract); 74:1037.
67. Elsass P, Lund Y, Ranek L. Encephalopathy in patients with cirrhosis of the liver: a neuropsychological study. *Scand J Gastroenterol* 1978;13:241–247.
68. Rikkers L, Jenko P, Rudman D, Freides D. Subclinical hepatic encephalopathy: Detection, prevalence and relationship to nitrogen metabolism. *Gastroenterology* 1978;75:462–469.
69. Conn HO. Trailmaking and number-connection test in the assessment of mental state in portal systemic encephalopathy. *Dig Sci Dis* 1977;22:541–550.

70. Parsons-Smith BG, Summerskill WHJ, Dawson AM, Sherlock S. The electroencephalograph in liver disease. *Lancet* 1957;2:867–871.
71. Yaar I, Shapiro MB, Pottala EW. Spectral analysis of the EEG in hepatic encephalopathy treated with levodopa. *Electroencephalog Clin Neurophysiol* 1981;52:617–625.
72. MacGillivray BB. The EEG in liver diseases. In: *Handbook of EEG and Clinical Neurophysiology,* Vol 15 (MacGillivray BB, ed). Amsterdam: Elsevier, pp. 26–50.
73. Cobb WA, Dawson GD. The latency and form in many of the occipital potentials evoked by bright flashes. *J Physiol* 1960;152:108–121.
74. Schafer DF, Pappas SC, Brody LE, Jacobs R, Jones EA. Visual evoked potentials in a rabbit model of hepatic encephalopathy. I. Sequential changes and comparisons with drug-induced comas. *Gastroenterology* 1984;86:540–545.
75. Zeneroli ML, Ventura E, Baraldi M et al. Visual evoked potentials in encephalopathy induced by galactosamine, ammonia, dimethyldisulfide and octanoic acid. *Hepatology* 1982;2:532–538.
76. Pappas C, Ferenci P, Schafer DF, Jones EA. Visual evoked potential in a rabbit model of hepatic encephalopathy. II. Comparisons of hyperammonemic encephalopathy, postictal coma, and coma induced by synergistic neurotoxins. *Gastroenterology* 1984;86:546–551.
77. Conn HO. The hepatic coma syndrome. *Warren-Teed GI Tract* 1974;4:18–23.
78. Greenberger NJ, Carley J, Schenker S, Bettinger I, Stamnes C, Beyer P. Effect of vegetable and animal protein diets in chronic hepatic encephalopathy. *Am J Dig Dis* 1977;22:845.
79. Sherlock S, Summerskill WHJ, White LP, et al. Portal systemic encephalopathy. Neurological complications of liver disease. *Lancet* 1954;2:453–457.
80. Swart GR, Frenkel M, van de Berg JWO. Minimum protein requirements in advanced liver disease; a metabolic study of the effect of oral branched chain amino acids. In: *Metabolism and Clinical Implications of Branched Chain Amino and Keto Acids* (Walser M, Williamson JR, eds). New York: Elsevier, 1981, pp. 427–432.
81. Dawson AM, McLaren J, Sherlock S. Neomycin in the treatment of hepatic coma. *Lancet* 1957;2:1263–1268.
82. Fast BB, Wolfe SJ, Stormont J, Davidson CS. Antibiotic therapy in the management of hepatic coma. *Arch Intern Med* 1958;101:467–475.
83. Bircher J, Muller J, Guggenheim P, Haemmerli UP. Treatment of chronic portal systemic encephalopathy with lactulose. *Lancet* 1966;1:890–892.
84. Elkington SG, Floch MH, Conn HO. Lactulose in the treatment of chronic portal-systemic encephalopathy: A double blind clinical trial. *N Engl J Med* 1969;281:408–412.
85. Simmons F, Goldstein H, Boyle JD. A controlled clinical trial of lactulose in hepatic encephalopathy. *Gastroenterology* 1970;59:827–832.
86. Bircher J, Haemmerli UP, Scolio-Lavizzari G, Hoffman K. Treatment of chronic portal-systemic encephalopathy with lactulose. Report of six patients and review of the literature. *Am J Med* 1971;51:148–159.
87. Uribe M, Campollo O, Vargas F, Ravelli GP, Mundo F, Zapata L, Gil S, Garcia-Ramos G. Acidifying enemas (lactitol and lactulose) vs nonacidifying enemas (tap water) to treat acute portal-systemic encephalopathy: A double-blind randomized clinical trial. *Hepatology* 1987;7(4):639–643.
88. Weber FL. Therapy of portal-systemic encephalopathy: The practical and the promising. *Gastroenterology* 1981;81:174–177.
89. Pirotte J, Guffens JM, Devos J. Comparative study of basal arterial ammonaemia and of orally induced hyperammonaemia in chronic portal systemic encephalopathy treated with neomycin, lactulose and association of neomycin and lactulose. *Digestion* 1974;10:435–444.
90. Conn HO, Lieberthal MM. Lactulose and neomycin: Combined therapy. In: *The Hepatic Coma Syndromes and Lactulose.* Baltimore: Williams and Wilkins, 1978, pp. 340–345.

91. Weber FL, Fresard KM, Lally BR. Effect of lactulose and neomycin on urea metabolism in cirrhotic subjects. *Gastroenterology* 1982;82:213–217.
92. Castell DO, Moore EW. Ammonia absorption from the human colon. The role of non-ionic diffusion. *Gastroenterology* 1971;60:33–42.
93. Herlong HF, Maddrey WC, Walser M. The use of orthinine salt of branched-chain keto acids in portal-systemic encephalopathy. *Ann Intern Med* 1980;93:545–550.
94. Eriksson LS, Persson A, Wahren J. Branched-chain amino acids in the treatment of chronic hepatic encephalopathy. *Gut* 1982;23:801–806.
95. McGhee A, Henderson JM, Milliken WJ, Bleir J. Comparison of the effect of hepaticaid and casein modular diets on encephalopathy, plasma amino acids and nitrogen balance in cirrhotic patients. *Ann Surg* 1983;197:288–293.
96. Michel H, Solere M, Granier P, Cavvet G, Bali JP, Pons F, Bellet-Hermann H. Treatment of cirrhotic hepatic encephalopathy with L-dopa. A controlled trial. *Gastroenterology* 1980;79:207–211.
97. Morgan MH, Read AE, Speller DCE. Treatment of hepatic encephalopathy with metronidazole. *Gut* 1982;23:1–7.
98. Loft S, Sonne J, Dossing M, Andreasen PB. Metronidazole pharmacokinetics in patients with hepatic encephalopathy. *Scand J Gastroenterol* 1987;22:117–123.
99. Heredia D, Caballeria J, Arroyo V, Ravelli G, Rodes J. Lactitol versus lactulose in the treatment of acute portal systemic encephalopathy (PSE). *J Hepatol* 1987;4:293–298.
100. Mendenhall CL, Rouster S, Marshall L, Weesner R. A new therapy for portal systemic encephalopathy. *Am J Gastroenterol* 1986;81(7):540–543.
101. Reding P, Duchateau J, Bataille C. Oral zinc supplementation improves hepatic encephalopathy: Results of a randomized controlled trial. *Lancet* 1984;493.
102. Singh HK, Gulati A, Srimal RC, Dhawan BN. Effect of Ro 15-1788 on dizepam, GABA and pentobarbitone induced EEG changes in rabbits. *Indian J Med Res* 1986;83:633–641.

6

Recent Advances in the Management of Islet Cell Tumors

Paul N. Maton

1. INTRODUCTION

Islet cell tumors are relatively rare. They can occur as a sporadic disease and most of them have also been described as part of multiple endocrine neoplasia type 1. One of the best available estimates of frequency of the occurrence of these tumors comes from Ireland where the incidence of insulinomas is 0.9 case per million per year, gastrinomas 0.4 cases, and all other islet cell tumors less than 0.2 case per million per year.[1] Islet cell tumors are usually described as slow growing and some patients with extensive metastatic disease live for many years. Nevertheless, many of these tumors release biologically active products into the circulation that can produce symptoms out of all proportion to tumor bulk. Indeed, these symptoms may be life threatening in a patient who otherwise has a relatively good prognosis from the point of view of tumor growth and spread. Conversely, the abolition of paraneoplastic symptoms can leave the patient feeling well despite the presence of a considerable tumor burden.

All islet cell tumors of the pancreas have similar histological appearances and the type of clinical syndrome depends on the nature of the products that are released into the circulation. These tumors are therefore classified functionally into those that produce principally vasoactive intestinal peptide (VIPomas), insulin (insulinomas), glucagon (glucagonomas), gastrin (gastrinomas), somatostatin (somatostatinomas), and growth hormone releasing factor (GRFomas).[2]

Paul N. Maton • Digestive Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892.

Islet cell tumors may also release ACTH causing atopic Cushing's syndrome,[3] amines and other products giving rise to an "atypical" carcinoid syndrome[2] or humoral hypercalcemic factors.[2] Rarely, tumors can produce solely neurotensin or pancreatic polypeptide, neither of which apparently gives rise to a clinical syndrome.[2] These latter two tumor types are usually grouped with those tumors that produce no detectable products and are referred to as nonfunctioning tumors. A proportion of islet cell tumors contain multiple peptides as detected by immunocytochemistry and a small proportion release more than one peptide into the circulation. In these cases, the tumors are usually classified by the dominant clinical syndrome, though on occasion this classification may be somewhat arbitrary.

When considering the management of islet cell tumors, it is necessary to consider both the control of the paraneoplastic symptoms and control of the tumor. Clearly, in patients with metastatic disease who are not curable, the control of symptoms is the major therapeutic goal. In patients without metastatic disease, if the primary tumor can be identified, there is a possibility of surgical cure.

2. Control of Symptoms

2.1. Introduction

Traditional modes of therapy for symptoms caused by islet cell tumors include methods designed to reduce tumor bulk, such as surgery, chemotherapy, or hepatic arterial embolization (see below). But even if appropriate, such forms of therapy remain ineffective in a proportion of patients, particularly those with metastatic disease. An alternative approach for symptomatic relief in patients with islet cell tumors is to block the function of the target tissue on which circulating peptides act. This approach has been most successful in patients with gastrinomas (Zollinger–Ellison syndrome) where the use of histamine H_2-receptor antagonists reduces acid secretion by parietal cells and thus effectively blocks the action of circulating gastrin on acid secretion. No comparably specific and effective drugs have been developed for other islet cell tumors, though a variety of measures to control symptoms have been used (Table 1). Recently, however, new drugs have been developed that can control symptoms due to islet cell tumors. For gastrinoma, omeprazole has become available, which is capable of controlling acid secretion even more effectively than the H_2-receptor antagonists, and for other islet cell tumors a somatostatin analog, SMS 201-995, has proved promising.*

*SMS 201-995 has been renamed octreotide (generic name) and Sandostatin (Sandoz) (trade name).

Table 1. Therapy for Symptoms Produced by Islet Cell Tumors

VIPoma	Codeine, loperamide, lithium, indomethacin, metaclopramide, chlorpromazine, corticosteroids, fluid and electrolyte supplements
Gastrinoma	Cimetidine, ranitidine ± anticholinergics
Insulinoma	Carbohydrate-rich meals, diazoxide, verapamil, diphenylhydantoin
Glucagonoma	Topical and oral zinc, protein supplements, insulin, anticoagulants
Somatostatinoma	Cholecystectomy
GRFoma	Bromocriptine, hypophysectomy

2.2. Omeprazole

Omeprazole is the first of a new class of agents, substituted benzimidazoles, which function as specific inhibitors of the hydrogen-potassium ATPase found on the luminal surface of the gastric parietal cell. Omeprazole apparently binds irreversibly to the ATPase and thus inhibits gastric acid secretion.[4] Onset of inhibition occurs within 4 hr and the inhibition lasts 2 or 3 days.[4] A previous publication from our unit demonstrated that short-term use of omeprazole was safe and effective for controlling gastric acid secretion in patients with gastrinomas.[5] However, in 1984 animal toxicity data demonstrated that long-term omeprazole in high dosage caused the development of gastric carcinoids in rats.[6] All trials in humans were stopped, the only exception being in patients with Zollinger–Ellison syndrome. Accordingly, we have conducted a long-term study of the safety and efficacy of this drug in patients with gastrinoma.[7]

Twenty-seven patients have been treated with omeprazole for up to 3 years. Each patient was monitored before treatment and after 6, 12, 24, and 36 months after starting the drug. On each occasion gastric acid secretion was determined, and a complete blood count, biochemical profile, and urinalysis were performed. In addition, each patient underwent upper intestinal endoscopy and gastric biopsies were taken. The dose requirement of omeprazole was determined as that which reduced gastric acid secretion to less than 10 meq/hr in the last hour before the next dose. This criterion for determining the dose of antisecretory medication had been shown to be safe in previous studies with H_2-receptor antagonists.[8] All patients were treated with omeprazole, initially with one dose per 24 hr, and if the dose required was greater than 120 mg/24 hr, the dose was split into two equal doses administered every 12 hr.

Of the 27 patients, 25 required only one dose per 24 hr, and two required

two doses per 24 hr. The mean dose required per day was 75 mg with a range of 20–120 mg. The dose of omeprazole was not related to the basal or maximal gastric acid secretion rate, nor to the plasma concentration of gastrin, but was directly related to the previous H_2-receptor antagonist dose.

Of the 27 patients treated with omeprazole, 4 have been treated for 3–6 months, 8 for 6–12 months, 7 for 12–24 months, and 8 for more than 24 months. In every patient, symptoms related to gastric acid hypersecretion were controlled completely and there were no endoscopic features of peptic ulcer disease. During therapy only two patients have required an increase in dose as judged by acid secretion measurements. General toxicity studies showed no side effects and no changes in complete blood count, plasma electrolytes, liver function tests, thyroid function tests, or urinalysis. Furthermore, there was no change in plasma concentrations of gastrin. Follow-up endoscopy has revealed no gastric carcinoid tumors and there have been no microscopic tumors observed in gastric biopsies taken at the time of endoscopy.

Thus, omeprazole offers certain advantages over H_2-receptor antagonists for patients with gastrinomas. All patients taking omeprazole preferred omeprazole because of the need for fewer doses and fewer tablets per day, and none of the patients needed anticholinergics while taking omeprazole. Gastric acid secretion could be controlled in all patients taking omeprazole, including two in whom acid secretion was not controlled by 8 g 24 hr of ranitidine. Furthermore, none of the patients with metastatic gastrinoma taking omeprazole required intravenous antisecretory medication during vomiting caused by chemotherapy. If long-term studies confirm the lack of toxicity of omeprazole, it will become the drug of first choice in patients with gastrinoma. Since the animal studies first demonstrated the development of carcinoids in rats, new evidence has appeared suggesting that this is not a direct tumorigenic effect of the drug itself, but rather is due to the fact that omeprazole is capable of producing achlorhydria, which leads to hypergastrinemia. Prolonged hypergastrinemia in turn produces an increase in mucosal enterochromaffin-like cells (ECL cells) and the ECL hypertrophy on occasion leads to the development of gastric carcinoid tumors.[9] Not only omeprazole but also ranitidine, if given in sufficient dosage, will increase the number of ECL cells in the gastric mucosa of rats.[10] The ECL cell hyperplasia is prevented by antrectomy, which lowers circulating plasma gastrin, and the degree of ECL cell hyperplasia is directly related to plasma gastrin concentration.[10] The rat responds to omeprazole-induced achlorhydria with a 15-fold increase in plasma gastrin. Such a rise in plasma gastrin does not occur when omeprazole in therapeutic dosage is given to humans. Plasma concentrations of gastrin remain in the normal range.[11] Nevertheless, continued vigilance during the use of omeprazole in humans is clearly mandatory.

2.3. Somatostatin Analogue-SMS 201-995

2.3.1. Introduction

When administered to normal volunteers, the naturally occurring peptide somatostatin reduces basal and stimulated plasma concentrations of many circulating peptides, and inhibits gastric and pancreatic secretion and intestinal absorption and motility by direct action.[12] Somatostatin has been infused into patients with islet cell tumors and has been shown to produce biochemical and sometimes symptomatic improvement in patients with VIPoma, insulinoma, gastrinoma, and glucagonoma.[13] However, natural somatostatin has a half-life in the circulation of about 3 min and therefore its therapeutic usefulness is limited. Recently, a long-acting somatostatin analog was developed with a half-life of approximately 100 min when administered by subcutaneous injection. This compound, SMS 201-995, has offered a new approach for symptom relief in patients with islet cell tumors. Because of the rarity of these tumors, and because different islet cell tumors have different clinical characteristics and natural histories, no prospective trials of SMS 201-995 have been performed. The strategy adopted by the drug company was to supply SMS 201-995 for individual patients through requesting investigators with general recommendations. These recommendations were that the dose be titrated for each patient, starting at 50 μg subcutaneously once or twice a day and increasing the dose on the basis of tolerability and therapeutic effects. If a therapeutic effect was not obtained on the maximum dose after 1 week, the drug would be stopped. Originally the maximum dose used was 600 μg/day, but this has now been changed to 1500 μg/day.

2.3.2. VIPomas

VIPomas give rise to watery diarrhea, hypokalemia, hypochlorhydria, hypophosphatemia, and sometimes hypercalcemia. Case reports describe 18 patients aged 43–75, all with metatastic VIPoma, who have been treated with SMS 201-995.[13] All had elevated plasma concentrations of VIP and all had had previous therapy including surgery, chemotherapy, hepatic arterial embolization radiotherapy, and a variety of drugs prescribed for symptomatic relief. Treatment with SMS 201-995 was initiated at 100–450 μg/24 hr, given subcutaneously in divided dosage every 8 or 12 hr. All 18 cases responded initially with a reduction of diarrhea. In those cases where the rate of onset of the drug was described, the beneficial effects of SMS 201-995 were evident within 12 hr. In some patients the effects of SMS 201-995 were very striking (Fig. 1), with a complete abolition of diarrhea, normalization of biochemistry, reduction in the need for fluid and electrolyte supplements, and transformation of the

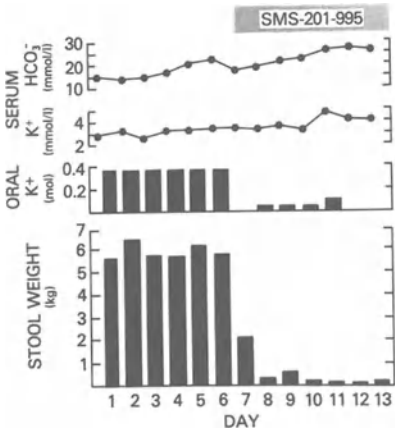


Figure 1. The effect of SMS 201-995 on daily stool weight, oral potassium supplementation (oral K^+), serum bicarbonate (HCO_3^-), and serum potassium (K^+) concentrations. (From Ref. 16.)

patient's existence from being in permanent hospital residence to living a normal life outside the hospital.[14] Not all cases had such a good response. In three cases, the effect only lasted a few days and increasing the dose of SMS 201-995 as high as 1200 μg/day without beneficial effect. Many patients continued to be controlled on their initial dose of SMS 201-995, but in two patients the diarrhea was only controlled with an increase in dosage and in two others only with the addition of steroids. One patient became resistant to the drug after treatment for more than 1 year.

Plasma concentrations of VIP fell during therapy in 16 of the 18 patients but fell to the normal range in only 6.[13] Symptomatic relief of the diarrhea was not always related to changes in plasma concentrations of VIP. One patient whose plasma concentration of VIP did not change during therapy had a complete, sustained reduction in diarrhea, and in five patients SMS 201-995 continued to be efficacious despite a subsequent rise in plasma concentration of VIP. In some patients plasma concentrations of VIP during therapy were in the range that causes a secretory diarrhea when VIP is infused into normal volunteers.[15] In at least one case it was demonstrated that the molecular size of circulating VIP changed to a larger, presumably less biologically active form during therapy with SMS 201-995.[16] However, it is also probable that, like natural somatostatin, SMS 201-995 has a direct effect on the gut.

SMS 201-995 seems effective in ameliorating symptoms in more than 80% of patients with VIPomas, many of whom have symptoms that are resistant to other modes of therapy. This drug clearly represents a significant advance in therapy of the symptoms due to VIPoma.

2.3.3. Glucagonomas

The glucagonoma syndrome is characterized by a unique rash, anorexia, weight loss, mild glucose intolerance, and, on occasion, diarrhea, neuropsychiatric symptoms, and venous thrombosis. Twelve patients with this syndrome were given SMS 201-995.[13] The patients were aged 31–76, all had metastatic disease, and at least nine had suffered from the rash typical of the syndrome. All 12 patients had elevated plasma concentrations of glucagon and all had received other forms of therapy including surgery, chemotherapy, hepatic arterial embolization, and symptomatic therapy for the skin rash including local or systemic zinc.

Treatment with SMS 201-995 was instituted at 100–450 μg/24 hr, in divided dosage. The rash resolved in the eight patients who had the rash at the time of the starting therapy, although two patients were also taking zinc. The improvement of the rash occurred over several days, slower than the response of diarrhea in patients with VIPomas. Weight loss, abdominal pain, and diarrhea improved in the patients with those symptoms, but SMS 201-995 apparently had little or no effect on the diabetes. Plasma concentrations of glucagon fell in nine of the 12 patients, but fell into the normal range in only one patient, and in one patient the rash resolved with no change in plasma concentrations of glucagon, amino acids, or zinc.[13]

Eight of the patients continued with SMS 201-995 for more than 6 months. Two of the patients required an increase in dose because of the recurrence of rash during therapy and in both cases the rash responded to the increased dose. The drug apparently continues to be effective in all patients though concentrations of glucagon have returned to pretreatment levels in at least two patients.

SMS 201-995 seems to be effective in controlling the rash and possibly other symptoms in patients with glucagonomas, though the response of the rash is sometimes difficult to assess because it may improve simply on admission to the hospital. The fact that the rash responds slowly and in some patients the response is independent of the effects on plasma concentrations of glucagon suggests that the drug may have a direct effect on the skin. This is supported by the finding that SMS 201-995 improves the skin rash in some patients with psoriasis.[17]

2.3.4. Insulinomas

Insulinomas produce transient episodes of neuropsychiatric disturbance often associated with palpitations and sweating as a consequence of low blood glucose. There are reports of some 20 patients with insulinomas given SMS 201-995 in the literature.[13] The patients were aged 33–82. In eight cases the tumors

were benign, in six malignant, and in six not stated. All 20 patients had had symptoms of hypoglycemia and at least six had received therapy with diazoxide and glucose, and five had received chemotherapy. All 20 patients had elevated plasma concentrations of insulin. Therapy was initiated with 50 μg or more every 12 hr. Thirteen of the 20 patients received the drug for short periods of up to a week. Effects on plasma insulin were not consistent. Insulin fell by more than 20% in 12 patients, was unchanged in three, rose in one, and data were not given in the other two. Similarly, effects on plasma concentrations of glucose were variable and did not always reflect changes in plasma concentrations of insulin. In two patients SMS 201-995 did not improve the hypoglycemia, and in at least one case it made the hypoglycemia worse. The reasons for this are not clear, but in normal volunteers SMS 201-995 is more potent at suppressing the release of growth hormone than the release of insulin.[18] If this were to occur in a patient with insulinoma, SMS 201-995 therapy could result in worsening of the hypoglycemia.

SMS 201-995 was given to seven patients with insulinoma for more than 1 month in doses between 100 and 1500 μg/24 hr. Hypoglycemia was well controlled in six despite a return of insulin concentrations to pretreatment levels in one of these. Symptoms recurred in two of the six patients, and one but not the other responded to an increased dose of SMS 201-995.

Although therapy with SMS 201-995 seems to be of benefit in about 50% of patients with insulinoma, this may be an overestimate. There is little evidence for significant clinical benefit in short-term studies, and even if SMS 201-995 is effective in maintaining normoglycemia, 85% of insulinomas are localized and benign and therefore, unlike other islet tumors, amenable to surgical cure. Moreover, in patients not cured by surgery, diazoxide may be useful, and it is not yet clear what role SMS 201-995 has in the management of such patients.

2.3.5. Gastrinomas

Gastrinomas produce Zollinger–Ellison syndrome characterized by gastric acid hypersecretion giving rise to peptic ulceration of diarrhea or both. There are published reports describing 49 patients with gastrinomas treated with SMS 201-995.[13] Many of these patients had a variety of previous therapies including surgery, chemotherapy, arterial embolization, and antisecretory therapy. Four of the patients had localized tumor, 24 had metastases, 2 had no tumor found on laparotomy, and in 19 the extent of the tumor was not stated. All 49 patients had elevated plasma concentrations of gastrin, and 23 patients had documented elevated basal acid output.

Short-term studies established that SMS 201-995 lowers plasma gastrin concentrations and gastric acid secretion. The time course and relative magni-

tude of the inhibition of plasma gastrin and gastric acid secretion suggests that SMS 201-995 inhibits acid secretion at least in part by direct action on the stomach, and this is true in studies carried out on normal volunteers given pentagastrin.[19] Thirty-two patients have been treated in long-term studies, 24 for between 1 and 6 months, and 8 for greater than 6 months. No systematic data are available on symptomatic responses in this group of patients because acid hypersecretion, the cause of most symptoms in this syndrome, was usually controlled by histamine H_2-receptor antagonists. However, some patients when given SMS 201-995 were able to reduce their intake of antisecretory medication.

Despite the finding that SMS 201-995 reduces gastric acid secretion and plasma concentrations of gastrin in patients with gastrinoma, the role of SMS 201-995 in the management of these patients is likely to remain limited. For long-term suppression of gastric acid hypersecretion, both histamine H_2-receptor antagonists and omeprazole can be taken orally and will render these patients asymptomatic. Furthermore, as parenteral histamine H_2-receptor antagonists are safe and effective, it is unlikely that SMS 201-995 will be useful in circumstances when oral antisecretory medication cannot be taken.

2.3.6. GRFomas

Islet cell tumors producing growth hormone releasing factor (GRF) stimulate the normal pituitary to release excessive amounts of growth hormone and give rise to acromegaly. Three patients with GRFomas—all women, aged 15, 25, and 60 years, and all with metastatic disease—received SMS 201-995.[13] Two of these patients had typical acromegaly and the 15-year-old patient had gigantism and primary amenorrhea. One patient had pituitary surgery, two had abdominal surgery, and one had hepatic arterial embolization. All three patients had elevated plasma concentrations of GRF and of growth hormone. Interestingly, two of these three patients initially presented with the gastrinoma syndrome and only subsequently developed signs of acromegaly.

All three patients were treated for more than 1 month and two for more than 6 months. All three patients had a good symptomatic response with reduction in hyperhidrosis in the two adult patients and arrest of growth and onset of menarche in the prepubescent female with gigantism. In one patient the pituitary became smaller as judged by CT scan.

In all three patients SMS 201-995 therapy reduced plasma concentrations of GRF and growth hormone, but these did not fall to within normal range in any. In one patient SMS 201-995 suppressed plasma concentrations of growth hormone while plasma concentrations of GRF were still above normal, leading the authors to suggest that SMS 201-995 acts directly on the pituitary to reduce growth hormone secretion as well as on the tumor to decrease GRF secretion.

This suggestion is consistent with the finding that SMS 201-995 suppresses growth hormone in normal volunteers and in patients with pituitary acromegaly.[20]

SMS 201-995 produced clinically significant symptomatic relief in patients with GRFomas who were resistant to other modes of therapy. This drug seems to be a significant advance in the therapy for such patients.

2.3.7. Nonfunctioning and Other Tumors

Four patients with nonfunctioning islet cell tumors were treated with SMS 201-995.[13] In two there were no tumor products detectable in plasma or urine and the drug was given only for a potential antitumor effect (Section 3.4.2). The other patients had elevated plasma concentrations of pancreatic polypeptide and glucagon. All the patients were treated for more than 3 months. Two of the patients noted an improvement in fatigue and abdominal pain and a reduction in plasma concentration of glucagon and pancreatic polypeptide. The other two patients had no symptomatic or biochemical response.

There are no reports of patients with pancreatic islet cell tumors producing solely neurotensin or somatostatin who received SMS 201-995. However, in one patient with an islet cell tumor producing a parathyroid hormone-like peptide, SMS 201-995 reduced plasma concentrations of calcium. Furthermore, one patient with a gastrinoma that also produced ACTH giving rise to Cushing's syndrome showed a marked clinical and biochemical response when given SMS 201-995.[21]

2.3.8. Side Effects

About half the patients receiving SMS 201-995 have noted side effects, but these have not been serious enough to stop therapy in any patient. Pain at the site of the injection can be minimized by warming the solution prior to administration and the hyperglycemia that occurs when starting the drug has been transient except in one case.[22] The most common significant side effects are gastrointestinal and include nausea, vomiting, heartburn, abdominal pain, constipation, and diarrhea. This is usually most marked at the beginning of therapy but may recur on occasion, and in some patients is a persistent problem. Some patients given the drug experience steatorrhea. In most patients this is mild and transient, though occasionally gross steatorrhea occurs, and in a series of patients given 1500 μg of SMS 201-995 a day, moderate persistent steatorrhea was a consistent finding.[23] Inhibition of pancreatic secretion may contribute to this malabsorption.

A few patients with islet cell tumors who received infusions of natural somatostatin experienced a rebound phenomenon when the drug was stopped. Plasma concentrations of peptides rose above pretreatment values and in some

cases there was an exacerbation of the symptoms.[13] Such a finding has been rare in patients given SMS 201-995, although there are reports of a patient with a VIPoma who developed an exacerbation of diarrhea[24] and a patient with a gastrinoma who developed intractable vomiting after SMS 201-995 was stopped.

2.3.9. Unresolved Issues

The current data on SMS 201-995 remain incomplete. The most appropriate dose of drug has not been determined and it is not known whether the dose should be the same for all tumor types. The mechanism of action of SMS 201-995 in those cases where circulating concentrations of peptide do not fall within the normal range is not clear, nor is it understood why some patients seem to require increasing doses of SMS 201-995 with time whereas others do not. Patients with tumors producing somatostatin develop constipation, mild malabsorption, gallstones, and mild diabetes.[25] It seems likely that long-term administration of the somatostatin analog, SMS 201-995, will produce an iatrogenic form of this syndrome, and indeed the fact that the drug causes steatorrhea suggests that this may be the case. However, the mechanism, frequency, persistence, and magnitude of the steatorrhea in relationship to drug dose has not been adequately documented. Furthermore, no study has examined prospectively the possible occurrence of gallstones during SMS 201-995 therapy. Whether the development of antibodies to SMS 201-995 as a mechanism of acquired resistance to this drug will be clinically significant is unknown, though preliminary reports suggest that this is not likely to be a major problem. Furthermore, there are questions as to whether SMS 201-995 has any antitumor effects (Section 3.4.2).

2.3.10. Conclusions

On the basis of the published data, it seems that there is an established role for SMS 201-995 in both the short- and long-term management of symptoms caused by VIPomas, glucagonomas, GRFomas, and unresectable insulinomas. However, the use of SMS 201-995 is likely to be limited in the management of symptoms caused by gastrinomas because of the availability of other medications and in the management of symptoms caused by the majority of insulinomas that are benign and are best treated surgically.

3. CONTROL OF THE TUMOR

3.1. Introduction

In considering management of the tumor in patients with islet cell tumors, the major determinant of therapy is whether or not the patient has metastatic

disease. The percentage of patients with metastatic disease at the time of diagnosis varies widely among the different islet cell tumors. Eighty to ninety percent of all patients with insulinomas and about 50% of patients with gastrinoma do not have metastases. Thus, in the two commonest islet cell tumors a significant number of patients are potentially curable by surgery. This is not true of the other types of islet cell tumor. About 70% of patients with VIPomas, 80% of patients with somatostatinomas, 100% of patients with GRFomas, and more than 50% of patients with glucagonomas have metastatic disease at the time of diagnosis. Assessment of tumor site and extent is thus an important and necessary step in determining the therapeutic approach.

3.2. Tumor Localization

In localizing islet cell tumors, upper GI series, intestinal endoscopy, endoscopic retrograde cholangiopancreatography, ultrasound examination of the abdomen, and liver-spleen scans are not generally useful. Those investigations of at least potential usefulness are CT scan, selective abdominal angiography, and portal venous sampling. We have recently evaluated methods of tumor localization in patients with Zollinger–Ellison syndrome.[26–28] Each patient had an abdominal CT scan and selective angiography and those who proceeded to surgery for potential cure also had portal venous sampling. In each case tumor extent was subsequently assessed at laparotomy, at autopsy, or, in a few cases of patients with metastatic disease, by percutaneous biopsy. Both CT scanning and selective angiography were examined for their ability to detect both the primary tumor and metastatic disease in the liver.

Fifty-eight patients had their liver assessed for the presence or absence of tumor. In 18 patients tumor was found in the liver at laparotomy, autopsy, or on liver biopsy and in 40 patients there was no tumor found. Of those patients with tumor in the liver, CT scan detected tumor in 13, the angiogram in 16, and the combination of both techniques in all 18. Of those with no tumor in the liver, CT scanning produced one false positive result but angiography produced no false positives. Thus, the combination of these techniques was both sensitive and specific for identifying metastatic gastrinoma.[27] Similar findings have been noted for insulinomas.[29]

Fifty-eight patients were assessed for the presence of primary tumor outside the liver. In 37 patients tumor was found at laparotomy or autopsy. The CT was positive in 21, angiography in 26, and the combination in 28 of these patients. Thus, the combination of these imaging techniques only detected 73% of extrahepatic tumors. Twenty-one patients had no tumor found. The CT and angiogram were both positive in one case in which tumor was not found at surgery. Thus, the combination of CT scan and angiography in detecting extrahepatic gastrinomas was specific but not very sensitive. In a group of 19 pa-

tients with insulinoma seen at the National Institutes of Health (NIH), angiography detected the extrahepatic tumor in 16 cases (84%).[29]

Portal venous sampling was carried out only in those patients who proceeded to potentially curative surgery. Since venous sampling is only performed as a secondary investigation following the other imaging techniques, its usefulness has been assessed as an adjunct to the other imaging studies. Of the 12 patients with tumor detected by CT scan or angiography, there was a gradient in plasma concentration of gastrin detected by portal venous sampling in 11. Subsequently, at surgery, tumor was found in all 12 patients. Thus, in those patients in whom imaging studies had detected a tumor, portal venous sampling did not provide the surgeon with additional useful information. Fifteen of the patients who proceeded to surgery had a negative CT scan and angiogram. Portal venous sampling detected a gradient in eight. In only two of those eight patients was the surgeon able to find the tumor and in those two patients the tumor was evident on immediate exposure of the pancreas. In the other six patients a detailed, lengthy examination of the area of interest in relation to the portal venous sampling result along with Kocherization of the duodenum and the pancreatic head, together with intraoperative ultrasound,[30] failed to reveal a tumor.[31] Thus, the finding of a gradient when imaging studies were negative did not help the surgeon find the gastrinoma. Presumably, these patients had microscopic tumors that would be detectable only on serial section of total pancreatic resection specimens.[32] In contrast to these results, in patients with insulinoma in whom CT and angiography did not produce a positive result, portal venous sampling has been more successful in identifying tumors that could subsequently be resected at surgery.[33]

The current scheme of localization of tumor in patients with the sporadic form Zollinger–Ellison syndrome at the NIH is shown in Fig. 2. A similar

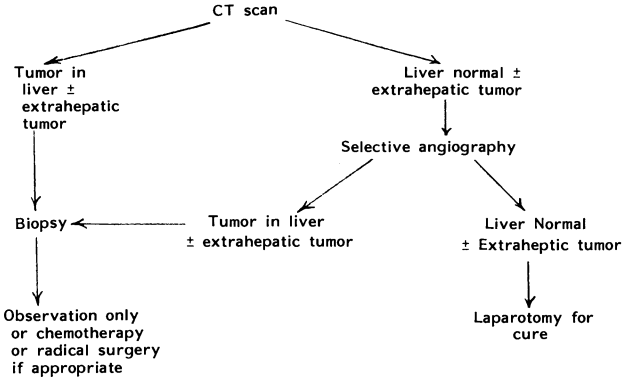


Figure 2. Scheme for tumor localization and management in patients with the sporadic form of gastrinoma.

scheme is used for insulinoma. Patients with gastrinoma and negative imaging studies are operated on as in up to 27% of patients with negative studies the surgeon finds tumor (see above). This scheme has proved useful in saving some patients from an unnecessary laparotomy, while allowing selection of others suitable for curative surgery.

3.3. Treatment of Localized Tumor

Management of localized tumor is surgical resection. Usually tumors will be found in or near the pancreas, but gastrinomas have been found in the liver and ovary. Tumors in the tail of the pancreas can be removed as a distal pancreatectomy. Lesions in the head of the pancreas are best enucleated as this procedure can result in cure[31] and avoids the morbidity and mortality of a Whipple procedure. As a result of appropriate imaging and patient selection, the surgical cure rate in patients with Zollinger–Ellison syndrome is now as high as 43% immediately postoperatively and 30% long term.[31] Insulinomas are curable by surgical resection in more than 80% of cases.[33]

Patients with multiple endocrine neoplasia type 1 frequently have multiple primary tumors. Insulinomas in these patients can be resected resulting in clinical and biochemical cure,[33] but resection of gastrinomas together with resection of two-thirds of the pancreas have not led to any reports of a cure for Zollinger–Ellison syndrome. Presumably these patients have multiple microscopic tumors throughout the pancreas. As patients with Zollinger–Ellison syndrome and multiple endocrine neoplasia type 1 cannot be cured and may have tumors of low potential for metastasis,[8] it is arguable as to whether such patients should have surgery.

3.4. Treatment of Metastatic Tumor

3.4.1. Introduction

A variety of approaches have been taken to the therapy of patients with metastatic islet cell tumor. Many patients have undergone partial or attempted complete resection of tumor, both to control tumor growth and extent and in an attempt to produce resolution of symptoms. This approach is only appropriate for patients who are relatively fit and in whom the bulk of the disease is relatively localized.[34] Other patients have undergone hepatic arterial embolization. No formal trials of this procedure have been performed in patients with islet cell tumors, but reports suggest good symptomatic results with an acceptable morbidity and low mortality.[35] However, there is currently no evidence that embolization prolongs survival and further studies are required.

Chemotherapy has been used for a variety of islet cell tumors and has

usually included streptozotocin, often with 5-fluorouracil and sometimes other agents.[8,33,36,37] One study of 37 patients with a variety if metastatic islet cell tumors given streptozotocin and 5-fluorouracil noted a response rate of between 60 and 85% with a duration of response lasting 17 months.[36] However, in this study, the definition of a response included patients in whom there was a reduction only in plasma peptide concentration with no change in tumor size. The response rate expressed as reduction in tumor size was not stated. In a recent study in 10 patients with actively growing metastatic gastrinoma, the combination of streptozotocin, 5-fluorouracil, and doxorubicin caused a reduction in size of tumor in only 40% and the response lasted for a mean of 7 months with a range of 6–18 months.[37] Thus, the results of chemotherapy in islet cell tumors are not dramatic.

Recently, two other potential therapies for the control of tumor growth in patients with metastatic islet cell tumors have become available: the somatostatin analog SMS 201-995 and α-interferon.

3.4.2. SMS 201-995

SMS 201-995 was given for more than 2 months and effects on tumor size were monitored, usually by serial CT scans, in 10 patients with VIPoma, 8 with glucagonoma, 5 with insulinoma, 16 with gastrinoma, 3 with GRFoma, and 4 with nonfunctioning or other tumors.[13] There was no standardization of previous therapy or SMS 201-995 dosage. Patients received between 200 and 1500 μg of SMS 201-995 a day. Of the 46 patients treated, it was stated that metastases increased in size in 20, stayed unchanged in 18, and decreased in size in 8 (Table 2). Reduction in size of metastases was claimed in four patients

Table 2. Effect of SMS 201-995 on Tumor Size[a]

Tumor	Patients (*N*)	Tumor size on follow-up: Larger (*N*)	No change (*N*)	Smaller (*N*)
Insulinoma	5	1	4	0
Gastrinoma	16	12	1	3
VIPoma	10	1	5	4
Glucagonoma	8	4	4	0
GRFoma	3	1	1	1
Nonfunct[b]/other	4	1	3	0
Total	46	20	18	8
% of Total	100	44	39	17

[a]Data taken from reference 13.
[b]Abbreviation: "nonfunctioning" tumor.

with VIPoma, three with gastrinoma, and one with GRFoma. However, the claim that metastases decreased in size has been partially retracted in the case of one patient with VIPoma. In one patient with gastrinoma, one metastasis decreased in size while another metastasis grew during SMS 201-995 therapy. This has been classified as tumor growth.

All cases where reduction in size of tumors was observed during therapy have probably been reported, while negative responses are less likely to have been published. Thus a reduction in size of metastasis in 30% of cases of islet cell tumors is probably an overestimate. This is supported by data from one study in which 22 patients were studied prospectively and tumor reduction was not observed in any patient.[23] Nevertheless, it is clear that in animal models of tumors, including islet cell tumors, pituitary tumors, and a chondrosarcoma, somatostatin or somatostatin analogs are capable of inhibiting tumor growth.[38] Should it prove that SMS 201-995 has significant antitumor activity in human islet cell tumors, then the indications for use of the drug would change. However, it is not clear from the reports whether there is a significant antitumor effect in human islet cell tumors or, if there is an effect, what the dose range is for this effect. These questions can only be answered by further studies.

3.4.3. Interferon

α-Interferons have been used for a variety of tumors with some success reported in renal cell carcinoma and malignant melanoma. One recent report has described the use of leukocyte interferon therapy in 22 patients with islet cell tumors.[39] All these patients had progressive metastatic disease as judged by serial scanning techniques. Fourteen patients had had previous surgery, 6 hepatic arterial embolization or ligation, 6 radiotherapy, 18 chemotherapy, and 2 SMS 201-995. All patients had become resistant to their previous therapy. They were given 3–6 million IU α-interferon subcutaneously a day. Of the 22 patients treated, tumors got larger in 2, did not change in size in 14, and were smaller in 6 (Table 3). The tumors got smaller in 2 of 4 patients with gastrinoma, 2 of 7 patients with VIPoma, and 2 of 9 patients with nonfunctioning tumors.

It is not clear whether these results are specific to this particular center or to the type of interferons used. Further studies from other centers will be required to fully assess the efficacy of interferon.

4. CONCLUSIONS

In recent years there have been a number of advances in the management of islet cell tumors extending to all aspects of tumor manifestations. With the

Table 3. Effect of Leukocyte Interferon on Tumor Size[a]

Tumor	Patients (*N*)	Tumor size on follow-up: Larger (*N*)	No change (*N*)	Smaller (*N*)
Insulinoma	1	1	0	0
Gastrinoma	4	0	2	2
VIPoma	7	0	5	2
Somatostatinoma	1	0	1	0
Nonfunctioning	9	1	6	2
Total	22	2	14	6
% of Total	100	9	64	27

[a]Data taken from reference 3.

advent of increasingly powerful gastric acid antisecretory medication and the development of SMS 201-995, it is now possible to completely control the paraneoplastic symptoms and signs in a greater proportion of patients with islet cell tumors than ever before. Even in those patients who are not rendered completely asymptomatic, a considerable number will be helped. Significant advances have been made in the localization of tumors and considerable knowledge gained about the limitations of the techniques used, leading to improved patient selection and the possibility for surgical cure in patients with localized tumors. Finally, in patients with metastatic disease, there is some hope that the new agents may be more effective than the traditional therapies.

ACKNOWLEDGMENT. The author thanks Audrey Pellegrino for preparing the manuscript for publication.

REFERENCES

1. Buchanan KD, Johnston CF, O'Hare MMT, Ardill JES, Shaw C, Collins JSA, Watson RCP, Atkinson AB, Hadden DR, Kennedy TL, Sloan JM. Neuroendocrine tumors: A European view. *Am J Med* 1986; 81 Suppl 6B:14–22.
2. Ch'ng JLC, Polak JM, Bloom SR. Miscellaneous tumors of the pancreas. In: *The Exocrine Pancreas: Biology, Pathobiology and Diseases* (Go VLW, Gardner JD, Brooks FP, Lebenthal E, DiMagno EP, Scheele GA, eds). New York: Raven Press, 1986, pp. 763–771.
3. Maton PN, Gardner JD, Jensen RT. The incidence and etiology of Cushing's syndrome in Zollinger–Ellison syndrome. *N Engl J Med* 1986;315:1–5.
4. Clissold SP, Campoli-Richards DM. Omeprazole: A preliminary review of the pharmacodynamic and pharmacokinetic properties, and therapeutic potential in peptic ulcer disease and Zollinger–Ellison syndrome. *Drugs* 1986;32:15–47.
5. McArthur KE, Collen MJ, Maton PN, Cherner JA, Howard JM, Gardner JD, Jensen RT.

Omeprazole: Effective convenient therapy for Zollinger–Ellison syndrome. *Gastroenterology* 1985;88:939–944.
6. Carlsson E, Larsson H, Mattson H, Ryberg B, Sundell G. Pharmacology and toxicology of omeprazole—with special reference to the effect on the gastric mucosa. *Scand J Gastroenterol* 1986;21 (Suppl 118):31–38.
7. Maton PN, McArthur KE, Wank SA, Slaff JI, Jensen RT, Gardner JD. Long term efficacy and safety of omeprazole in patients with Zollinger–Ellison syndrome. *Gastroenterology* 1986;90:1537.
8. Jensen RT, Doppman JL, Gardner JD. Gastrinoma. In: *The Exocrine Pancreas: Biology, Pathobiology and Disease* (Go VLW, Gardner JD, Brooks FP, Lebenthal E, DiMagno EP, Scheele GA, eds). New York: Raven Press, 1986, pp. 727–744.
9. Elder JB. Inhibition of acid and gastric carcinoids. *Gut* 1985;26:1279–1283.
10. Larsson H, Carlsson E, Mattsson H, Lundell L, Sundler F, Sundell G, Wallmark B, Watanabe T, Hakanson R. Plasma gastrin and gastric enterochromaffin-like cell activation and proliferation. Studies with omeprazole and ranitidine in intact and antrectomized rats. *Gastroenterology* 1986;90:391–399.
11. Pounder RE, Sharma BK, Walt RP. Twenty-four hour intragastric acidity during treatment with oral omeprazole. *Scand J Gastroenterol* 1986;21(Suppl 118):108–116.
12. Reichlin S. Somatostatin. *N Engl J Med* 1983;309:1495–1501, 1556–1563.
13. Maton PN, Gardner JD, Jensen RT. The use of the long-acting somatostatin analogue SMS 201-995 in patients with pancreatic islet cell tumors. *Dig Dis Sci* 1989;(in press).
14. Maton PN, O'Dorisio TM, Howe BA, McArthur KE, Howard JM, Cherner JA, Malarky WB, Collen MJ, Gardner JD, Jensen RT. Effect of long-acting somatostatin analogue (SMS 201-995) in a patient with pancreatic cholera. *N Engl J Med* 1985;312:17–21.
15. Kane ME, O'Dorisio TM, Krejs GJ. Production of secretory diarrhea by intravenous infusion of vascoactive intestinal polypeptide. *N Engl J Med* 1983;309:1482–1485.
16. Maton PN, O'Dorisio TM, O'Dorisio MS, Malarkey WB, Gower WJ Jr., Gardner JD, Jensen RT. Successful therapy of pancreatic cholera with the long acting somatostatin analogue SMS 201-995. Relation between plasma concentrations of drug and clinical and biochemical responses. *Scand J Gastroenterol* 1986;21(Suppl 119):181–186.
17. Camisa C, O'Dorisio TM, Maceyko R, Mekhjian HS, Malarkey WB. Improvement of psoriasis with chronic use of somatostatin (SMS) analogue SMS 201-995. *Clin Res* 1987;35:637A.
18. Johnston DG, Davies RR, Turner SJ. Effects of somatostatin and SMS 201-995 on carbohydrate metabolism in normal man. *Scand J Gastroenterol* 1986;21(Suppl 119):158–165.
19. Gyr KE, Whitehouse I, Beglinger C, Kohler E, Dettwiler S, Fried M. Human pharmacological effects of SMS 201-995 on gastric secretion. *Scand J Gastroenterol* 1986;21(Suppl 119):96–102.
20. Lamberts SWJ, Del Pozo E. Acute and long-term effects of SMS 201-995 in acromegaly. *Scand J Gastroenterol* 1986;21(Suppl 119):141–148.
21. Bonfils S, Ruszniewski P. Laucournet H, Costil V, Rene W, Mignon M. Long-term management of Zollinger–Ellison syndrome with SMS 201-995, a long-lasting somatostatin analog. *Can J Physiol Pharm: Program of the 6th International Symposium on Gastrointestinal Hormone,* Vancouver, Canada, July 6–10, 1986, p. 63.
22. Shepherd JJ, Senator GB. Regression of liver metastases in patients with gastrin-secreting tumor treated with SMS 201-995. *Lancet* 1986;2:574.
23. Kvols LK, Buck M, Moertel CG, Schutt AJ, Rubin J, O'Connell MJ, Hahn RG. Treatment of metastatic islet cell carcinoma with a somatostatin analogue (SMS 201-995). *Ann Intern Med* 1987; 107:162–168.
24. Edwards CA, Cann PA, Read NW, Holdsworth CD. The effect of somatostatin analogue SMS 201-995 on fluid and electrolyte transport in a patient with secretory diarrhea. *Scand J Gastroenterol* 1986;21(Suppl 119):259–261.

25. Krejs GJ, Orci LO, Conlon JM, Ravazzola M, Davis GR, Raskin P, Collins SM, McCarthy DM, Baetens D, Rubinstein A, Aldor TAM, Unger RH. Somatostatinoma syndrome: Biochemical, morphologic, and clinical features. *N Engl J Med* 1979;301:285–292.
26. Wank SA, Doppman JL, Miller DL, Collen MJ, Maton PN, Vinayek R, Slaff JI, Norton JA, Gardner JD, Jensen RT. Prospective study of the ability of computed axial tomography to localize gastrinomas in patients with Zollinger–Ellison syndrome. *Gastroenterology* 1987;92:905–912.
27. Maton PN, Miller DL, Doppman JL, Collen MJ, Norton JA, Vinayek R, Slaff JE, Wank SA, Gardner JD, Jensen RT. Role of selective angiography in the management of patients with Zollinger–Ellison syndrome. *Gastroenterology* 1987;92:913–918.
28. Cherner JA, Doppman JL, Norton JA, Miller DL, Krudy AG, Raufman J-P, Collen MJ, Maton PN, Gardner JD, Jensen RT. Selective venous sampling for gastrin to localize gastrinomas. *Ann Intern Med* 1986;105:841–847.
29. Dunnink NR, Long JA Jr., Krudy A, Shanker TH, Doppman JL. Localizing insulinomas with combined radiographic techniques. *Am J Roentgenol* 1980; 135:747–752.
30. Cromack DT, Norton JA, Sigel B, Shawker TH, Doppman JL, Maton PN, Jensen RT. The use of high resolution intraoperative ultrasound to localize gastrinomas. An initial report of a prospective study. *World J Surg* 1987;11:648–652.
31. Norton JA, Doppman JL, Collen MJ, Harmon JW, Maton PN, Gardner JD, Jensen RT. Prospective study of gastrinoma localization and resection in patients with Zollinger–Ellison syndrome. *Ann Surg* 1986;204:468–479.
32. Roche A, Raisonnier A, Gillon-Savouret MC. Pancreatic venous sampling and arteriography in localizing insulinomas and gastrinomas: Procedure and results in 55 cases. *Radiology* 1982;145:621–627.
33. Comi RJ, Gorden P, Doppman JL, Norton JL. Insulinoma. In: *The Exocrine Pancreas: Biology, Pathobiology and Diseases* (Go VLW, Gardner JD, Brooks FP, Lebenthal E, DiMagno EP, Scheele GA, eds). New York: Raven Press, 1986, pp. 745–761.
34. Norton JA, Sugarbaker PH, Doppman JL, Wesley RA, Maton PN, Gardner JD, Jensen RT. Aggressive resection of metastatic disease in selected patients with malignant gastrinoma. *Ann Surg* 1986;203:352–359.
35. Allison DJ. The nonsurgical management of metastatic endocrine tumours to the liver. In: *Interventional Radiology* (Wilkins RA, Viamonte M, eds). Oxford: Blackwell Scientific, 1982, pp. 191–208.
36. Moertel CG, Hanley JA, Johnson LA. Streptozotocin alone compared with streptozotocin plus fluorouracil in the treatment of advanced islet cell carcinoma. *N Engl J Med* 1980;303:1189–1194.
37. von Schrenk T, Howard J, Doppman JL, Norton JA, Maton PN, Smith F, Vinayek R, Frucht H, Wank SA, Gardner JD, Jensen RT. Prospective study of chemotherapy in patients with metastatic gastrinoma. *Gastroenterology* 1988;94:1326–34.
38. Reubi JC. Somatostatin analogue inhibits chondrosarcoma and insulinoma tumor growth. *Acta Endocrinol* 1985;109:108–114.
39. Eriksson B, Oberg K, Alm G, Karlsson A, Lindqvist G, Andersson T, Wilander E, Wide L. Treatment of malignant endocrine pancreatic tumors with human leukocyte interferon. *Lancet* 1986;2:1307–1309.

7

Clinical Pharmacokinetic Considerations in Gastroenterology

Varocha Mahachai

1. INTRODUCTION

Physiological changes in gastrointestinal and liver disease alter drug pharmacokinetics, although the alterations are of poorly defined clinical significance. Quite often these changes are unpredictable, depending on the nature of the drug and the clinical status of the patient. In the presence of liver disease, the dosage of a drug metabolized by the liver is often adjusted empirically. Specific guidelines are lacking because the available clinical and biochemical assessments of liver disease are inadequate. Even less is known about the effect of gastrointestinal disease on drug handling. Impaired drug absorption may become critical when a rapid onset of drug action is required or when elimination is so rapid that effective plasma concentrations cannot be maintained. Therefore, the adjustment of dosage and route of administration for some drugs may be necessary to attain optimal pharmacological effect.

2. PHARMACOKINETIC APPLICATION

Pharmacokinetics may be defined as the mathematical relationship between the dose of a drug and its concentration in the blood or plasma, as

Varocha Mahachai • Division of Gastroenterology, Department of Medicine, University of Alberta, Edmonton, Alberta, Canada T6G 2C2.

determined by the characteristic absorption, distribution, and elimination of the drug. *Absorption* is the process of transfer of a drug from its site of administration to the systemic circulation. *Distribution* is the process of reversible transfer of a drug to and from the site of measurement; distribution depends on drug binding to macromolecules in blood and tissue. *Elimination* represents the irreversible loss of drug from the site of measurement, either by metabolism or excretion. In this chapter, the focus will be on the absorption and hepatic elimination processes, since they are potentially impaired in gastrointestinal and liver disease and are therefore of major interest to all clinicians.

In the past, many of us have used an empirical approach in selecting a dosage regimen for our patients, based on subjective observations of the desired effects of the medication. Recent information on the pharmacokinetic profile of a particular medication and the relationship between efficacy and drug concentration has led to more optimal dosage regimens. These maximize pharmacological effects and lessen side effects. Ultimately, however, the value of a given dosage regimen must be assessed by the clinical end point, which may not always be clearly defined. Thus, an understanding of the pharmacokinetic properties of a particular agent facilitates more rapid achievement of a safer, more effective dosage regimen, and serves as a useful means of evaluating existing dosage regimens.

Several considerations have to be made in designing a proper dosage regimen: (1) most important, the intrinsic pharmacological properties and toxicology of the drug; (2) the pharmacokinetic properties of the drug, how the body handles it, and its route of administration; (3) the patient's individual profile including age, weight, disease state, genetic factors, and drug tolerance or dependence; and (4) convenience of dosage regimen and patient compliance.

3. PHYSIOLOGICAL CONSIDERATIONS

3.1. Gastrointestinal Absorption

Drugs are most commonly given orally, providing the gastrointestinal tract with a major role in determining the rate and extent of drug absorption. The efficiency with which an orally administered drug is absorbed is a function of many variables. The drug should be sufficiently water-soluble to dissolve in the aqueous gastrointestinal secretions and yet be sufficiently lipophilic to diffuse across the lipid membranes of the enterocytes to reach the splanchnic circulation. The absorption of a drug is a function of its physicochemical nature, its formulation, and its interactions with various luminal components of the gastrointestinal tract. Absorption of orally administered drug therefore depends on several gastrointestinal factors, including intraluminal pH, gastric and intestinal

transit times, gastrointestinal and pancreatic secretions, metabolic and enzymatic activities (both in the gut wall and in the liver), intestinal uptake, and the splanchnic and portal blood flow. Besides, other intraluminal factors such as bacteria, other drugs, or food can also influence the extent of drug absorption. Physiological variables can be affected in many conditions (Table 1). Drug absorption is therefore potentially altered in these conditions.

The stomach is generally not an important site of absorption, except for weakly acidic or nonionized forms of some drugs. For example, the absorption of aspirin and ethanol from the stomach has been estimated to be approximately 10 and 30% of the oral dose, respectively.[1,2] However, the rate of gastric emptying may influence the rate of drug absorption by determining the rate of input of drug into the small intestine where major absorption occurs. The small intestine is the most important site of drug absorption. This may be because of its large surface area.[3] The intestinal transit determines the contact time between drug and intestinal absorptive surface. Although the intestinal transit does not play a crucial role in the absorption of most drugs, it could become a major determinant of the absorption of enteric-coated and slow-release formulations. The colon usually serves as a reserve area for the absorption of drugs that have escaped absorption proximally, perhaps because of the physicochemical properties of the drug or its dosage forms. Some drugs have been specially formulated to act or be absorbed locally in the colon, either as enteric-coated or sustained-release preparations, or by complexing two agents requiring colonic flora for the release of active moieties (e.g., sulfasalazine or the dimer of 5-aminosalicylic acid). Colonic absorption takes place after rectal administration of some drugs given in the form of suppositories or enemas.

Most drugs used in clinical practice are administered orally and they must

Table 1. Conditions Affecting Physiological Variables

Physiological variables	Conditions affecting variables
Intraluminal events (pH, bacteria, secretion)	Food, drugs (antisecretory agents), diseases (achlorhydria, bacterial overgrowth, pancreatic disease)
GI motility	Food, drugs (narcotics, anticholinergics), diseases
Intestinal uptake	Food, diseases
Blood flow (portal and splanchnic)	Food, drugs, diseases, posture, cardiac output
Intestinal and hepatic metabolism	Genetically determined, drugs, food
Drug binding	Drug, diseases

traverse the gastrointestinal barrier before reaching the systemic circulation. The nature and characteristics of these gastrointestinal barriers are of considerable importance in drug absorption. Most drugs are absorbed from the intestine by a passive process, either through the lipid or aqueous channels in the brush border membranes, from a region of higher concentration in the lumen to one of lower concentration in the enterocyte. Absorption is usually favored when the drug is in the nonionized and more lipophilic form. Some drugs are absorbed by an active transport process whereby the substance is transported against a concentration gradient and utilizes metabolic energy. For example, methyldopa, levodopa, and penicillamine are absorbed by the amino acid active transport mechanisms.[4]

Drug absorption also depends on the physicochemical properties of the drug, such as solubility, particle size, and chemical form. The "pH partition theory" of drug absorption assumes that the gastrointestinal tract is a lipid barrier to the transport of drugs: the nonionized form of an acidic or basic drug is highly lipid-soluble and is therefore readily absorbed. Most acidic drugs remain predominantly un-ionized at the low pH of gastric contents and may be absorbed from the stomach as well as from the small intestine. For example, aspirin, which is a weak acid, remains largely nonionized in the acidic milieu of the stomach allowing ready absorption across the gastric mucosa. However, aspirin and most weak acids are still well absorbed in the small intestine due to the greater surface area and a relatively longer residence time in this intestinal location. Weakly basic drugs are essentially nonionized throughout the intestine and are readily absorbed. The gastrointestinal luminal pH therefore is an important factor in drug absorption because many drugs are either weak organic acids or bases. Gastric pH can be altered in certain disease states such as achlorhydria or hypersecretory conditions. Several drugs, especially those used in the treatment of peptic ulcer disease, can change intragastric pH and may influence the dissolution, stability, and/or absorption of certain drugs.

Lipid solubility also determines the rate of drug absorption, and this lipophilic nature is provided by the partition coefficient between a fatlike solvent and water or an aqueous buffer.[5] The unstirred water layer, adjacent to the luminal surface of the intestinal membrane, is a potential rate-determining barrier in the absorption of most drugs and can influence the absorption of many lipid-soluble compounds. The presence of such an unstirred layer explains the less-than-anticipated decreased dependence of the absorption rate of certain drugs on lipid solubility and ionization.[6–9]

Both splanchnic and portal blood flow determine the rate of delivery of drugs from the intestine to the liver, where further metabolism may take place. The interactions between the splanchnic blood flow and drug absorption have not been fully explored. The dependence of intestinal absorption of drugs on gastrointestinal blood flow is determined by the absorbability of the medica-

tions. Absorption of lipid-soluble molecules and molecules that are small enough to easily penetrate the aqueous phase is rapid and highly dependent on the rate of gastrointestinal blood flow.[10] In general, the rate of drug absorption is unaffected by normal variability in mesenteric blood flow. To significantly influence drug absorption, there must be dramatic changes in mesenteric blood flow from disease or drug effects. In contrast, hepatic clearance of drugs highly metabolized by the liver is affected by portal blood flow.

3.2. Presystemic Metabolism

After oral administration, a drug must pass sequentially from the gastrointestinal lumen, through the gut wall, and then through the liver before reaching the systemic circulation. A fraction of the amount absorbed may be eliminated before reaching the bloodstream. An oral dose of a drug may be completely absorbed, yet incompletely available to the systemic circulation. This process is often referred to as the *"presystemic"* or *"first-pass metabolism,"* which can occur in the intestine and/or the liver. Several drugs are known to undergo extensive first-pass metabolism thereby increasing the requirement of the oral dose in relation to its intravenous dose (Table 2). This extensive first-pass effect sometimes precludes the use of a drug such as nitroglycerin as an oral dosage form. The extent of first-pass metabolism can be determined in relation

Table 2. Drugs That Undergo Extensive First-Pass Metabolism

β-Adrenoreceptor blocking agents
Aprenolol
Labetalol
Metoprolol
Oxprenolol
Propranolol
Analgesics
Codeine
Morphine
Pentazocine
Pethidine
Phenacetin
Antiarrhythmic agents
Lidocaine
Calcium channel blocking agents
Verapamil
Nifedipine
Vasoactive drugs
Isoproterenol
Dihydroergotamine
Nitroglycerin

Table 3. Drugs That Undergo Intestinal Wall Metabolism

Levodopa	Morphine
Flurazepam	Phenacetin
Isoprenaline	Estrogen
Salicylamide	

to the fraction of the dose that is absorbed. Separate contributions of the gut wall and liver to presystemic metabolism have been studied in animal experiments using segmental intestinal perfusion and isolated liver perfusion techniques. The model of portacaval transposition has also been employed to study separate contributions of the intestine and the liver to this process in intact animals.[11] First-pass metabolism in the gut wall and liver can be differentiated by comparing plasma drug concentration after oral and intraportal or intraperitoneal administration. This allows an assessment of separate contributions of the intestinal wall and the liver to this process. Knowledge of intestinal first-pass metabolism is less well developed than our understanding of hepatic first-pass metabolism. Examples of drugs that undergo metabolism in the intestinal wall are shown in Table 3. A high proportion of absorbed ethinylestradiol undergoes presystemic metabolism in the gut wall.[12] It has also been found *in vitro* that ethinylestradiol is extensively metabolized by human jejunal mucosa to sulfate conjugate.[13] Gut wall metabolism also results in changes in metabolite excretion patterns. For example, intravenous isoproterenol is excreted largely unchanged in humans. On the other hand, the sulfated conjugate accounts for 80% of the drug in the urine after oral administration, suggesting that the presystemic metabolism of isoproterenol occurs in the gut wall.[14] Other drugs that undergo extensive intestinal first-pass metabolism are tributaline, morphine, and salicylamide. This is a fruitful area for future clinical research.

Intestinal drug metabolism occurs either in the intestinal lumen by microorganisms, or in the intestinal mucosa which contains metabolic enzymes.[15] Phase I and phase II reactions can take place in the intestinal epithelial cells similar to those occurring in the liver. Phase I reactions involve oxidation, reduction, and hydrolysis. The microsomal enzyme system involving cytochrome P-450 catalyzes various oxidative and reductive processes. Compounds then undergo phase II reactions in which they are conjugated with sulfate or glucuronic acid. The rate of phase I reactions and the contents of cytochrome P-450 are much less in the intestinal mucosa than those in the liver. The intestinal cells, however, have more capability of phase II reactions (conjugation and sulfation) which thereby reduce bioavailability of certain drugs such as sal-

icylamide,[16] morphine,[17] and isoprenaline.[18] Intestinal biotransformation can be affected by factors that influence activities of the metabolizing enzymes.[19] For example, the activity of cytochrome P-450-mediated enzymes can be induced by certain factors such as cigarette smoking or ingestion of charcoal-broiled meat.[20]

3.3. Physiological Significance of Intestinal Drug Metabolism

The intestinal mucosa is exposed to a high concentration of a wide variety of xenobiotics contained in food and drugs. It is amazing how the gut mucosa can resist the damage from commonly ingested foreign chemical substances. The colonic mucosa is somewhat more susceptible than the small intestine to damage by certain carcinogens. It has been speculated that the biological system for the resistance to chemical toxic injury may be located in the microsomal enzyme system of the enterocytes. As the intestine is the portal of entry of many foreign substances, it may play a most important role in the detoxification process.

It is important to determine the factors that influence this intestinal first-pass process, and it will be of interest to determine whether this process is altered by gastrointestinal diseases. Intraluminal bacteria are required for the metabolism of certain compounds such as levodopa and sulfasalazine.[21] The effect of altered intestinal flora (e.g., bacterial overgrowth) on drug absorption is unknown. Nonspecific enzymes such as penicillin esters and dopa-decarboxylase, also present in the intestinal cells, may play major roles in metabolism of certain drugs.

3.4. A Physiological Approach to Hepatic Drug Clearance

The liver is a major organ of drug biotransformation as it contains many enzymatic pathways. Lipophilic compounds must be metabolized to more polar, water-soluble forms in order to be eliminated. The majority of phase I oxidative reaction occurs in the endoplasmic reticulum mediated by the cytochrome P-450 system. These phase I reactions (oxidation, reduction, or hydrolysis) will render the compounds more soluble and available for excretion, or further metabolized by phase II reactions. The metabolites formed can either be pharmacologically active or inactive. The phase II reactions further metabolize the phase I compounds to more polar metabolites (by glucuronidation, sulfation, acetylation, and conjugation).

The efficiency of the liver in eliminating a drug is usually expressed in terms of hepatic clearance. *Clearance* is defined as the efficiency of any organ to irreversibly remove a drug from the perfusing medium. Clearance is expressed as the volume of blood from which drug is completely removed per

unit time. At steady state, clearance may be estimated from organ blood flow and arterial–venous concentration difference across an organ such as the liver. Hepatic clearance (Cl_H) may be calculated as follows:

$$Cl_H = Q\ (Ca - Cv)/Ca = QE \quad (1)$$

where Q is total liver blood flow, Ca is mixed portal venous and hepatic arterial concentration, Cv is hepatic venous total drug concentration, and E is the extraction ratio or the efficiency of drug removal by the liver. Thus, hepatic clearance is a function of liver blood flow which determines the rate of delivery of a drug to the hepatocyte mass and the ability of the liver to extract a drug as it perfuses through the hepatic sinusoids. In contrast to the commonly used pharmacological constant elimination half-life ($T_{½}$), clearance is a more *precise* physiological measurement of the efficiency of drug elimination. For many drugs, the liver is the major site of elimination, and thus systemic clearance equates hepatic clearance.

Biological determinants of hepatic drug elimination include the metabolic activity of the liver, hepatic blood flow, and the extent of drug binding in the plasma. Alterations in any of these factors that determine hepatic drug elimination (whether caused by disease state, drug interaction, or genetic or individual variation) will affect the blood concentration–time curve. The hepatic extraction reflects the overall activity of the elimination process as well as other variables such as blood flow and drug binding. These extraction ratios therefore provide a basis for the classification of drugs with respect to the disposition into high- and low-extraction compounds as shown in Table 4.[22]

If a drug has a low extraction ratio (due to a small intrinsic clearance relative to liver blood flow), hepatic clearance will be independent of changes in liver blood flow but will be highly sensitive to the hepatic metabolizing enzyme activity (intrinsic clearance) and proportional to the unbound drug concentration in plasma. There will be only a small first-pass effect after oral administration of drugs in this class, and the dose will reach systemic circulation mostly intact. On the other hand, for a drug that has a high extraction ratio, the hepatic clearance will predominantly reflect liver blood flow. The clearance and half-life will be sensitive to changes in blood flow and will be insensitive to alterations in metabolic activity. The compounds that have extraction ratios between 30 and 70% fall into the category of intermediate extraction compounds. They will have an intermediate mixture of properties, with their clearance being partly dependent on liver blood flow and hepatic enzyme activity.

For the majority of drugs, a plot of the blood concentration expressed as a natural logarithm versus time gives a straight line, indicative of a monoexponential process of elimination. The two independent physiological variables,

Table 4. Pharmacokinetic Classification of Drugs Eliminated Primarily by Hepatic Metabolism[a]

Drug class	Approximate extraction ratio
Flow-limited	
Lignocaine	0.83
Propranolol	0.60–0.80
Pethidine	0.60–0.95
Pentazocine	0.8
Nortriptyline	0.5
Morphine	0.5–0.75
Capacity-limited, binding-sensitive	
Phenytoin	0.03
Diazepam	0.03
Tolbutamide	0.02
Warfarin	0.003
Chlorpromazine	0.22
Clindamycin	0.23
Quinidine	0.27
Digitoxin	0.005
Capacity-limited, binding-insensitive	
Theophylline	0.09
Hexobarbitone	0.16
Amylobarbitone	0.03
Antipyrine	0.07
Chloramphenicol	0.28
Thiopentone	0.28
Paracentamol	0.43

[a]Modified from Ref. 22 with permission.

volume of distribution (V_d) and clearance (Cl), determine the slope or constant rate of elimination (K_{el}):

$$K_{el} = \frac{Cl}{V_d} \tag{2}$$

Half-life ($T_{½}$) is a commonly used pharmacokinetic parameter that relates to clearance and volume of distribution as follows:

$$T_{½} = \frac{V_d \times 0.693}{Cl} \tag{3}$$

Thus, $T_{1/2}$ is a dependent variable that can be influenced by any physiological changes in either volume of distribution or clearance. The volume of distribution is dependent on drug binding to macromolecules in blood and tissue. Total clearance (Cl) is customarily used as a measure of the efficiency of the elimination process. It can be estimated from

$$Cl = \frac{\text{dose}}{\text{AUC}} \tag{4}$$

where AUC is the area under the blood concentration time curve extrapolated to infinity. The pharmacokinetic profile of a given drug is useful in choosing the proper dosage and dosing interval by knowing the fate of the drug in the body: the extent of its absorption, the way it distributes in the body, and the rate of its elimination from the body.

4. DRUGS USED AS INDICATORS OF HEPATIC FUNCTION

Conventional liver function tests are often a poor measure of the severity of liver disease. Unlike creatinine clearance in the assessment of renal function, these biochemical tests do not often reflect the severity of liver disease. More dynamic functional tests are required for the assessment of liver disease. With the increased understanding of the physiological factors determining drug disposition, clearance of certain drugs has been used as a probe for the individual aspects of hepatic function. In choosing a drug for a test of hepatic function, the compound must be nontoxic, must be available for intravenous use (or, if given orally, its absorption must be complete and rapid), and, finally, must be measurable either as the parent compound or its metabolites. The drugs must be predominantly eliminated by the liver in order to reflect hepatic disease.

Antipyrine has been used widely as a model drug to assess hepatic enzyme activity. Because antipyrine is negligibly bound to plasma protein and tissue and because it is eliminated almost exclusively by the liver with a low hepatic extraction ratio, its clearance is dependent on oxidative metabolic activity. Routine medical biochemical tests seem to be poor indicators of the liver's ability to metabolize antipyrine. The problem of using antipyrine as a functional test is that there are overlaps between the clearance values in patients with liver disease versus normal subjects. These overlaps lead to the lack of prediction of the severity of liver disease at any time.[23] Despite these limitations, antipyrine clearance may serve as a quantitative measure of liver function and it may be useful as a prognostic indicator to follow the course of liver disease.[24]

Clearance of highly extracted compounds such as indocyanine green and galactose has been used to estimate liver blood flow.[25,26] This clearance tech-

nique is a simple and noninvasive means of indirectly estimating liver blood flow. It is important to understand the limitations of this method. The extraction ratio of indocyanine green and galactose is assumed to be unity (thus avoiding the invasive means of obtaining the true hepatic extraction by hepatic vein catheterization), but the extraction ratio may in fact be altered in the presence of liver disease, due to liver dysfunction and intrahepatic or extrahepatic shunting. Accordingly, a single measurement may not give as meaningful information for blood flow estimation as if one were to use the clearance of these compounds as an indicator to follow the course of liver disease or to monitor the outcome of therapy.

5. *BIOAVAILABILITY*

The rate and extent of absorption and delivery of a drug to the systemic circulation is usually expressed in terms of bioavailability. Bioavailability is often used to measure the extent of first-pass metabolism of an orally administered drug when the drug is completely absorbed, although such determinations cannot differentiate between biotransformation processes occurring in the liver and the intestine. The bioavailability of a drug may be defined as the fraction of unchanged drug present to the systemic circulation after the first pass of a drug through the intestine and the liver. In general, drug bioavailability is examined in healthy volunteers under carefully controlled conditions without any interference of other factors such as food, other drugs, or disease state. The data are used by regulatory agencies and clinicians to establish appropriate dosage regimens for their patients. Unfortunately, in the real clinical setting other factors are often involved such as food, coingestion of other drugs, and disease state for which the drug is being taken in the first place. Other than physicochemical and pharmacological properties of the drug, factors attributable to the specific patient such as age, sex, other diseases, and concomitant medication must also be considered. Although the clinical end point is the pharmacological response, the pharmacokinetic properties may be useful in the selection and design of proper dosage regimens. Bioavailability studies are particularly important with drugs that have a low therapeutic ratio in order to design optimal dosage regimens, to avoid unnecessary toxicity, and to attain adequate pharmacological effects. This is also particularly important for drugs that have a short half-life that often exhibit fluctuations of plasma concentration and pharmacological response, and for drugs with variable or poor absorption.

The first-pass process occurring in the intestine and/or liver must be considered in bioavailability estimation as it may be responsible for the variability of drug bioavailability, especially with drugs that undergo extensive first-pass metabolism. First-pass metabolism can become saturable depending on the dose

and quite often is variable. Intestinal metabolism is usually capacity-limited because of the saturation of enzymatic activity.[27] The intestinal first-pass metabolism is responsible for the reduced bioavailability of drugs undergoing intestinal biotransformation.[28] Because of quantitative considerations, the systemic bioavailability is largely dependent on the hepatic first-pass effect, since the liver contains a higher metabolic biotransformation capacity.

Absolute *availability* or systemic *availability* (F) of a drug is the fraction or percentage of the administered dose that actually reaches the systemic circulation. This is determined from blood levels or urinary excretion data after oral administration, with reference to similar measurements made after intravenous administration. This can be calculated from the area under the blood or plasma concentration time curve after oral administration (AUC_{po}) and that obtained after intravenous administration (AUC_{iv}).

$$F = \frac{AUC_{po}}{AUC_{iv}}$$

Bioavailability can also be determined on a relative basis by comparing the test drug in reference to the standard dosage form both given orally. Bioavailability studies usually are carried out in a crossover fashion in the same person to avoid the effects of intersubject variability in drug clearance. The standard procedure for the determination of the absolute or relative bioavailability of a drug requires the administration of different formulation in a random order crossover study on separate occasions. The principal assumption underlying this technique is that the drug is eliminated from the body at the same rate on different occasions. Unfortunately, large intraindividual variation in first-pass metabolism by the liver can occur with several high-clearance drugs. This problem has introduced the use of simultaneous administration of the test and standard dosage forms of the drug where one dosage form can be labeled with stable isotopes.[29,30] This technique allows simultaneous administration of different routes or different formulations. The compound can be labeled with ^{13}C, ^{2}H, or ^{15}N.

6. PHARMACOKINETIC VARIABILITY: EFFECTS OF DISEASE

Information on drug pharmacokinetics is often based on the studies carried out in healthy, young adult subjects. In real clinical situations, drugs are used in persons of all ages with high prevalence of disease. It would be logical to try to understand the pharmacokinetics of drugs in patient populations in which they will be used. Disease affects various organ systems of the body and affects drug pharmacokinetics or the way drugs are absorbed, distributed, excreted,

and metabolized. Drug elimination can be directly affected by renal disease. Metabolism of some drugs is affected by hepatic disease. Drug binding can be affected by both renal disease and hepatic disease. Cardiovascular disease can affect drug distribution and the transport of drugs to eliminating organs. The next section of this chapter will focus on how gastrointestinal and liver diseases can affect drug pharmacokinetics.

6.1. Effects of Gastrointestinal Disease on Drug Absorption

Many conditions involving the gastrointestinal tract are likely to cause changes in drug absorption because of altered physiological functions. Little attention has been paid to the effect of gastrointestinal disease on drug absorption.

6.1.1. Gastrointestinal Motility Disorders

Any factor that accelerates gastric emptying is likely to increase the rate of drug absorption. Conversely, any factor delaying emptying is likely to decrease the rate of drug absorption. Gastric emptying is delayed in certain disease states such as gastric atrophy, gastric carcinoma, pyloric stenosis, gastric ulcer, and postgastric surgery.[31] Interestingly, there is a reported case of a woman with hypertrophic pyloric stenosis who retained 61 out of 66 enteric-coated aspirin tablets in her stomach.[32]

Impaired drug absorption of drugs such as acetaminophen has been shown to occur in the presence of delayed gastric emptying.[33,34] On the other hand, rapid intestinal transit associated with diarrhea may decrease drug absorption, especially with enteric-coated and slow-release formulations. This effect is possibly due in part to the reduced time available for absorption and in part to the failure of drug dissolution in the gut lumen.[35] Failure of action of oral contraceptives can occur during episodes of gastroenteritis, again possibly due to impaired absorption.[36] During acute exacerbation of inflammatory bowel disease, an oral dosage form of corticosteroid usually is supplemented by the intravenous route on the unproven assumption that absorption is impaired. Indeed, any condition that alters gastrointestinal motility such as diabetic gastroparesis, hypothyroidism, intestinal pseudoobstruction, and diarrheal disorders can potentially influence drug absorption.

6.1.2. Diseases of the Stomach

Drug absorption is potentially impaired in achlorhydria because of the changes in intragastric acidity and gastric emptying. The absorption of aspirin is increased in achlorhydria, even though the high gastric pH would be expected to inhibit aspirin absorption (because more of the drug is in the ionized

form). The increase in absorption may be related to the delayed gastric emptying.[37] The absorption of cephalexin, penicillin V, tetracycline, or acetaminophen was not found to be altered in patients with achlorhydria.[38,39] Drug absorption may be altered following gastrointestinal surgery because of the reduction in absorptive surface area or changes in gut motility and secretory patterns. Absorption of acidic, basic, and neutral compounds is reduced in patients who had undergone antrectomy with gastroduodenostomy and selective vagotomy.[40] This altered absorption may be associated with delayed gastric emptying. Absorption of sulfonamide was found to be impaired in patients who had undergone Billroth II gastrectomy as the gastric emptying is altered,[41] but the absorption of digoxin was not altered.[42] The unpredictable effects of gastric surgery on drug absorption can perhaps be related to the fact that the absorptive surface of the small intestine still remains intact, and these effects also depend on the physicochemical properties of the drugs.

6.1.3. Diseases of the Intestine

Studies of drug absorption in patients with small intestinal disease have been limited and conflicting. Some have shown impaired drug absorption in the presence of malabsorption,[43] but others have found drug absorption to be normal in such patients.[44] Absorption of digoxin tablets is impaired in radiation-induced enteritis but, interestingly, the problem is overcome when using digoxin in an elixir form.[45] This suggests that impaired absorption may be the result of delayed tablet dissolution rather than a defect in the intrinsic absorption of the drug. Propranolol bioavailability was not found to be altered in patients with untreated celiac sprue,[46] but direct perfusion of propranolol into the proximal jejunum results in a 70% decrease in absorption in patients with celiac disease compared to normal subjects. This suggests that celiac disease (which commonly involves the proximal part of the bowel) affects the absorption but not the bioavailability of propranolol as the normal distal part of the intestine still can compensate and provide for adequate total absorption. Intestinal bypass surgery can also reduce bioavailability of some drugs such as oral contraceptives, hydrochlorothiazide, and phenytoin.[47–49] Impaired drug absorption occurs in radiation-induced enteropathy.[50]

In the presence of small intestinal diseases, drug absorption may be impaired because of the reduced effective surface area for absorption. The diffusion process may be impaired in the presence of inflamed bowel. The changes in bacterial ecology in the intestinal lumen (such as bacterial overgrowth syndrome) may alter the rate of drug metabolism in the intestinal lumen. The effect of Crohn's disease and celiac disease on drug absorption is variable depending on the activity and extent of the disease and the nature of a given drug.[51–53] The decreased bioavailability may be due to an absorptive defect and increased

bioavailability may be due to increased permeability.[54,55] Absorption of prednisolone was found to be reduced in patients with Crohn's disease,[56] and oral absorption may be delayed.[57]

The alteration in the absorption of propranolol in Crohn's disease may be due to the change in acid microclimate.[58] The bioavailability of the acidic lipid-soluble compound indomethacin is unaltered in patients with treated celiac disease, and the absorption of salicylate is slightly faster than in control subjects, possibly due to the more rapid gastric emptying that occurs in celiac disease.[59]

Absorption patterns of antibiotics vary in patients with celiac disease.[60] In patients with Crohn's disease, absorption of lincomycin is decreased, absorption of clindamycin or cotrimoxazole is increased, but absorption is unchanged with rifampicin and cephalexin.[61] Possible explanations for alterations in absorption of antibiotics include variations in gastric emptying, mucosal damage, impaired diffusion, increased permeability, and defective mucosal enzyme activity. It has been suggested that the decreased absorption of acetaminophen in patients with celiac or Crohn's disease may be caused by the altered gastric emptying in these patients rather than the absorptive defect.[62]

6.2. *Drug Metabolism in Liver Disease*

The liver is the major organ of drug biotransformation. Thus, it might be expected that liver disease would have significant effects on drug metabolism. The effect of liver disease on drug metabolism has been reviewed.[63] The clinical significance of altered drug metabolism in patients with liver disease is not entirely clear. However, it would seem logical and practical to cautiously administer drugs in patients with advanced liver disease as they are extremely vulnerable to the toxic effects of drugs, particularly those with low therapeutic ratio. Practically, dose administration in patients with liver disease is adjusted according to the patient response under close observation, as there are no definite predictors of response or guidelines for dosage administration.

The unpredictability of the effects of liver disease on pharmacokinetics could be due to the multiple effects of liver disease on drug-metabolizing enzyme activities, drug binding, and on hepatic blood flow. It also relates to the complexity of hepatic metabolism as some enzyme systems are far more sensitive to the effects of liver disease than others. The clinical significance of altered drug metabolism in liver disease depends on the type and severity of the underlying liver disease and on the pharmacokinetic properties of the drugs.

Drugs with high hepatic extraction ratio have the greatest potential for increased bioavailability in patients with liver disease. This is due to reduced hepatic extraction as a consequence of impaired metabolic capacity and the development of hepatic portasystemic shunt. This shunt allows drugs to bypass functioning hepatocytes, thus escaping metabolism. The reduced systemic

clearance and markedly increased systemic bioavailability often lead to increased toxicity. This therefore poses major problems in the administration of high extracted compounds such as propranolol,[64] metoprolol,[65] pethidine, pentazocine,[66] and nifedipine.[67] The elimination of lidocaine, for example, is highly blood flow-dependent and has a low therapeutic ratio, making dosage adjustment in patients with cirrhosis necessary. Its clearance is markedly impaired in chronic active hepatitis and cirrhosis,[68,69] but not in acute viral hepatitis.[70] Dosage modification of drugs with low hepatic extraction ratio (whose clearance is primarily dependent on hepatic metabolizing capacity) may be essential for certain drugs with low therapeutic index such as theophylline,[71] or diazepam.[72] Clearance of chlordiazepoxide is also prolonged in cirrhosis.[73] In contrast, the clearance of oxazepam and lorazepam, benzodiazepines that are glucuronidated to inactive metabolites, does not appear to be significantly altered in either cirrhosis or acute viral hepatitis.[74,75] This finding suggests that glucuronidation may be relatively spared in some forms of liver disease. Recent work suggests that *ester* glucuronidation but not *ether* glucuronidation may be impaired in cirrhosis.[76] Delayed clearance and decreased protein binding have been demonstrated in patients with chronic active liver disease. Plasma protein binding of drugs is often impaired in patients with liver disease, possibly because of their reduced serum albumin concentration.[77] However, the altered protein binding in liver disease does not always correlate with serum albumin concentration. A decrease in protein binding in liver disease will only substantially influence clearance of highly albumin-bound drugs with a low intrinsic clearance or hepatic extraction ratio, such as phenytoin or warfarin. This is because the clearance of both drugs depends on the unbound drug available for metabolism and the activity of hepatic metabolizing enzymes. The effects of liver disease on drug pharmacokinetics are complex and sometimes difficult to predict.

The clinical significance of the effects of liver disease on drug elimination (and, more specifically, whether the dosage regimen of a drug should be modified in patients with liver disease) is not clear, although the incidence of adverse effects to drugs is expected to be higher in this population. It would be ideal to have certain guidelines as to how the dosage should be modified in liver disease. The best advice at this present state of knowledge is to administer drugs to patients with liver disease carefully and to titrate the dose to the desired clinical response. It would appear prudent to use sedatives and analgesics cautiously in patients with liver disease: it is advised to titrate the dose in each patient to clinical response and, in selected instances, to monitor the plasma concentration.[78] A useful comment on safe prescriptions in liver disease states that "little change in prescribing is necessary when liver disease is inactive, though dosage should be kept low and particular care should be taken with sedative and antidepressant drugs. When active liver disease or signs of hepatic

decompensation are present, it is likely that drug metabolism is deranged and the greatest care, indeed, should be exercised in prescribing."[79]

7. DRUG–DRUG INTERACTION

When drugs are taken concurrently, drug interactions may occur on a pharmaceutical, pharmacodynamic, or pharmacokinetic basis. This is a common phenomenon in clinical practice where multiple drugs are often prescribed, particularly in the elderly who are often subjected to many disease processes. An increased awareness of the mechanisms of drug interactions will enable the clinicians to anticipate and cautiously prescribe multiple drugs to avoid undesired interactions. This awareness will prompt the clinicians to monitor for possible interactions and possibly to measure plasma concentrations.

Pharmacokinetic drug interactions can occur when one drug interferes with the absorption and disposition of another drug. Basic pharmacological knowledge of mechanisms of drug action on physiological variables can be applied to predict potential drug interactions (Table 5). An alteration in pharmacokinetics does not necessarily lead to an alteration in clinical response or to enhanced drug toxicity. Drug interactions become critical when drugs with narrowed therapeutic index or steep dose response curve are prescribed. Pharmacokinetic

Table 5. Mechanisms of Drug–Drug Interactions Affecting Bioavailability

Mechanism	Examples
Luminal binding	Charcoal, cholestyramine, antacid, ferrous sulfate
Altered GI motility	Anticholinergic, metoclopramide
pH effect	Antisecretory agents
Altered portal blood flow	Hydralazine, propranolol, digoxin, cimetidine
Altered metabolic enzyme activity (P-450)	
Inducer	Phenobarbital, phenytoin, carbamazepine, ethanol, rifampin, griseofulvin
Suppressor	Cimetidine, diltiazem

drug interactions can occur during the absorption, distribution, or elimination steps.

7.1. Drug Interactions Affecting Gastrointestinal Absorption

Absorbents such as cholestyramine and activated charcoal reduce the rate and extent of absorption of many drugs.[80–83] Gastric pH can be altered by antacids and various antisecretory agents used in the treatment of peptic ulcer disease. Ketoconazole absorption is impaired in patients with achlorhydria and in patients receiving antacids and H_2-receptor antagonists.[84] The effect of antacids on the absorption of other drugs is unpredictable.[85,86] It is unknown if the reduction in intragastric pH by any of the potent antisecretory agents including cimetidine, ranitidine, famotidine, synthetic prostaglandin, and substituted benzimidazole influences absorption of other drugs. It is unresolved as to whether antacids interfere with absorption of cimetidine.[87–91] Antacids can interact with other drugs by different mechanisms: urinary excretion of aspirin is favored by the urine-alkalinizing effect of antacids, antacids also chelate drugs such as tetracycline,[92] calcium carbonate can precipitate ferrous sulfate, and gastric emptying is delayed by aluminum hydroxide.[93] Most of these adverse interactions can be avoided by giving the drug and antacid 2 hr apart.

The gastric emptying rate influences the time for water-soluble drugs to dissolve. Drugs that promote gastric emptying, such as metoclopramide, tend to increase the rate of drug absorption, whereas many drugs such as anticholinergics, narcotics, analgesics, or antidepressents may delay gastric emptying and reduce the rate of drug absorption.[35,94] For example, propantheline has been found to significantly reduce the absorption rate of riboflavin, sulfamethoxasole, ethanol, and acetaminophen.[35,41,95,96] On the other hand, the promotility agent, metoclopramide, significantly increases absorption rate of ethanol, acetaminophen, tetracycline, and ampicillin.[94,96–98] Small-bowel transit time also plays a role in drug absorption as it determines the contact time between drug and absorption site. The absorption of drugs from enteric-coated preparations is also largely determined by the rate of gastric emptying. Intersubject variability in the rate of absorption often occurs with these preparations. Coated granule preparations have been designed so that they can be emptied gradually but continuously into the duodenum.

The effect of the change in gastrointestinal motility on the rate and extent of drug absorption may also be determined by drug formulation. For example, the bioavailability of long-acting propranolol is not affected by the coadministration of metoclopramide.[99] A combination of an H_2-receptor antagonist and a promotility agent or an anticholinergic agent is sometimes used in the treatment of peptic ulcer disease. They might also interact with each other as the anticholinergic agent may delay gastric emptying and the H_2-receptor antagonist,

cimetidine, is known to suppress hepatic oxidative metabolism.[100] When cimetidine is coadministered with pirenzepine (an antimuscarinic agent) in patients with peptic ulcer disease, there is no pharmacokinetic interaction between the two drugs.[101] Thus, pirenzepine in therapeutic doses may not have an effect on gastric emptying and the inhibition of metabolic pathways by cimetidine is not pertinent to the effect of pirenzepine.

7.2. Drug Interactions Affecting Hepatic Metabolism

Several drugs are known to induce hepatic microsomal enzyme activities of the cytochrome P-450-mediated pathways, including phenobarbital, phenytoin, carbamazepine, ethanol, rifampin, and griseofulvin.[102–109] This enzymatic pathway can be inhibited by cimetidine and diltiazem.[110,111] Cimetidine can therefore interact with chlordiazepoxide and diazepam.[112,113] Cimetidine also inhibits the metabolism of highly extracted drugs such as propranolol, perhaps also through its action in reducing hepatic blood flow.[114] Drugs that commonly interact with cimetidine are warfarin, lidocaine, diazepam, propranolol, labetalol, theophylline, and phenytoin. Cimetidine reduces the clearance of indocyanine green, a compound commonly used to estimate liver blood flow.[114] Cimetidine also reduces clearance and prolongs the half-life of antipyrine,[115] a compound commonly used to estimate hepatic enzyme activity. The inhibition of drug metabolism by cimetidine is not related to its H_2-receptor antagonist properties as ranitidine does not exert a similar effect on drug metabolism.[116–121] Hepatic blood flow, one of the important determinants of the clearance of drugs that are highly extracted by the liver, can be decreased by drugs that reduce cardiac output, such as propranolol which can in turn decrease its own clearance.[122] Propranolol also reduces the clearance of the highly extracted compound lidocaine, whose metabolism is highly blood flow-dependent.[123]

8. EFFECT OF FOOD ON DRUG ABSORPTION

In clinical situations, drugs are often given to patients in the presence of food to avoid their irritating effect on the stomach. The variable effects of food on drug absorption depend on the time, amount, and constituents of the meal and the nature of a drug (Table 6). The absorption rate of digoxin, acetaminophen, and sulfonamides can be reduced in the presence of food as the result of delayed gastric emptying.[124–126] Food can affect the rate and extent of drug absorption by several mechanisms. It can alter gastrointestinal functions: gastric emptying time, gastrointestinal motility, secretion, and blood flow. Food can also have a direct effect on drug absorption by forming complexes with the

Table 6. Effect of Food on Drug Absorption

Increased	Decreased	Delayed	No effect
Griseofulvin	Digoxin	Cephalosporins	Digoxin elixir
Hydrochlorothiazide	Acetaminophen	Sulphonamides	Chlorpropramide
Propranolol	Sulfonamide	Quinidine	Tolbutamide
Metoprolol	Ketoconazole		Ethambutol
Mebendazole	Captopril		
Nitrofurantoin	Tetracycline		
Spironolactone	Ampicillin		

compound or by acting as a physical barrier to drug diffusion. It can also affect the disintegration or dissolution of drugs.

Both the rate and the extent of absorption of many antibiotics are significantly decreased after a meal.[127–130] Absorption of tetracycline is reduced by about 50% by a meal,[131] and its absorption is reduced when taken with milk products because of the interaction of the drug with calcium resulting in a poorly absorbed complex.[132] On the other hand, absorption of certain drugs can be promoted when given with a meal.[133,134] This enhanced bioavailability after a meal may relate to a longer contact time between drug and the absorptive surface of the small intestine. Food can also alter both systemic and portal blood flow, and this may have an effect on the first-pass drug metabolism of selected compounds in the intestinal wall and the liver.[135,136]

A number of dietary factors influence drug pharmacokinetics. Metabolism of drugs is reduced in patients with vitamin C deficiency.[137] Charcoal-broiled beef, cabbage, and brussels sprouts have been shown to induce drug metabolism.[138]

9. EFFECT OF AGING ON PHARMACOKINETICS

Another important biological process that can influence pharmacokinetics is aging. Many drugs are used in older persons, even though most of the information on the pharmacokinetics of drugs is based on studies done in youths. Furthermore, multiple drugs are often used in the elderly and the prevalence of disease increases with age. Older persons are more susceptible to drug-related toxicity, partly because of altered physiological responses as well as drug interactions.

There are several possible mechanisms of the effects of aging on drug pharmacokinetics (Table 7). Drug absorption may be altered in the elderly because of changes in physiological functions. Intragastric pH tends to be higher in the elderly because of the reduction in parietal cell mass. Intestinal blood flow is potentially reduced in the presence of atherosclerosis occurring in the elderly. Gastric emptying time may also be altered in the elderly. Bacterial overgrowth is more common in the elderly. An impaired esophageal motility in the elderly may result in a failure of drug absorption if the tablet were to remain lodged in the esophagus. It is debatable if only active absorption processes are impaired with advancing age.[139] Increased plasma drug concentrations of several drugs were previously demonstrated in the elderly, although these changes may be a function of reduced elimination rather than a change in absorption.[140] Aging can also lead to an alteration in drug binding, metabolism, excretion, or distribution.[141,142]

Other major changes in the elderly include alteration in their volume of distribution as a result of the reduction of lean body mass in proportion to total body weight, or a diminuition of serum albumin concentration as a result of chronic illness and malnutrition. Drugs that are highly bound or have a low volume of distribution are more likely to be affected with the changes of aging. Changes of blood flow to different organs can also affect the distribution of drugs. Cardiac output can decrease by up to 40% in the elderly. Organ blood flows (including renal blood flow, hepatic blood flow) and peripheral blood flow are often decreased in the elderly and can result in decreased metabolic and renal clearance.[143]

Hepatic drug metabolism appears reduced in the elderly although there is a great deal of variation.[144,145] Both liver mass and hepatic blood flow decrease with age. The intrinsic hepatic clearance is highly variable with age.[146] Nutritional deficiency may be partly responsible for the decline in microsomal en-

Table 7. Possible Mechanisms of the Effects of Aging on Drug Pharmacokinetics

Changes in drug absorption
Reduction of hepatic metabolism
Reduction of liver blood flow
Reduction of GFR and renal blood flow
Reduced cardiac output
Reduced lean body mass
Reduced serum albumin

zyme function in the elderly.[147] Reduced renal function is important in drug administration in the elderly, especially for drugs excreted by the kidneys. Renal clearance is generally reduced in the elderly as the kidney undergoes physiological and anatomical changes with aging, including the reduction in glomerular filtration rate, tubular reabsorption, and secretion. Besides the potential alteration of pharmacokinetic properties of drugs in the elderly, this patient population tends to be more sensitive to the effects of various drugs, resulting in an altered pharmacological response. In most cases, the changes in pharmacological response are a result of pharmacokinetic alteration with aging.

10. CONCLUSION

In this chapter, factors that determine the relationship between drug dosage and plasma drug concentration (pharmacokinetics) have been reviewed. Most pharmacokinetic data are derived from studies performed on healthy, young volunteers without other interferences such as disease, food, other drugs, or the aging process. In real clinical situations, a number of factors contribute to intra- and interindividual variability in drug pharmacokinetics. This pharmacokinetic variability can lead to an alteration in pharmacological response that can pose major problems when drugs with narrow therapeutic index are given. The knowledge of basic pharmacokinetic principles provides the basis for efficacious and safe drug administration. Liver and gastrointestinal tract play important roles in determining drug availability to the systemic circulation after oral administration. The dosage regimen is customarily modified in patients with liver disease since metabolism of several drugs is impaired, although there are no specific guidelines for using drugs in these patients due to the lack of sensitive indication of the ability of the liver to metabolize drugs. Less is known of the effect of gastrointestinal disease on drug absorption. Impaired drug absorption could become critical for drugs with a short half-life and rapid elimination whose pharmacological effect is dependent on adequate plasma concentration. Additional factors including other drugs, food, and the aging process should be considered in the design of proper dosage regimen. The clinical significance of pharmacokinetic alteration depends on the nature of a given drug and the clinical status of the patients. Generalizations cannot be made for any disease or any given drug but drug administration must be individualized on the basis of several factors including disease state, concomitant drug administration, food, and aging. A suspicion of pharmacokinetic alteration should be a signal to initiate a request for plasma drug level monitoring for dosage modification in order to attain optimal pharmacological effect and minimize possible toxicities. Much work remains to be done in this area before specific guidelines

can be provided in terms of these factors. It is, however, important to appreciate the need to assess the dosage regimen on an individual basis.

REFERENCES

1. Cooke AR, Hunt JN. Absorption of acetylsalicylic acid from unbuffered and buffered gastric contents. *Am J Dig Dis* 1970;15:95–102.
2. Cooke AR, Birchall A. Absorption of ethanol from the stomach. *Gastroenterology* 1969;57:269–272.
3. Levine RR. Factors affecting gastrointestinal absorption of drugs. *Am J Dig Dis* 1970;15:171–188.
4. Wass M, Evered DF. Transport of penicillamine across the mucosa of the rat small intestine in vitro. *Biochem Pharmacol* 1970;19:1287–1295.
5. Schanker LS. On the mechanism of absorption of drugs from the gastrointestinal tract. *J Med Pharm Chem* 1960;2:343–359.
6. Winne D. The influence of unstirred layers on intestinal absorption. In: *Intestinal Permeation: Proceedings of the Fourth Workshop Conference,* Vol. 4 (Krumer K, Lauterbach F, eds). Hoechst, Schloss Reisenburg, October 1975, pp. 19–22. Amsterdam: Excerpta Medica International Congress Series No. 391, 1977, pp. 58–64.
7. Winne D. Dependence of intestinal absorption in vivo on the unstirred layer. *Naunyn Schmiedebergs Arch Pharmacol* 1978;304:175–181.
8. Suzuki A, Higuchi WI, Ho NFH. Theoretical model studies of drug absorption and transport in the gastrointestinal tract I. *J Pharm Sci* 1970;59:644–651.
9. Suzuki A, Higuchi WI, Ho NFH. Theoretical model studies of drug absorption and transport in the gastrointestinal tract II. *J Pharm Sci* 1970;59:651–659.
10. Winne D. Influence of blood flow in intestinal absorption of xenobiotics. *Pharmacology* 1980;21:1–15.
11. Effeney DJ, Pond SM, Man-Wai L, Silber BM, Riegelman S. A technique to study hepatic and intestinal drug metabolism separately in the dog. *J Pharm Exp Ther* 1982;221:507–511.
12. Back DJ, Breckenridge AM, Maciver M, Orme M, Purba HS, Rowe PH, Taylor I. The gut wall metabolism of ethinyloestradiol and its contribution to the pre-systemic metabolism of ethinyloestradiol in humans. *Br J Clin Pharmacol* 1982;13:325–330.
13. Back DJ, Bates M, Breckenridge AM, Ellis A, Hall JM, Maciver M, Orme M, Bowe PH. The in vitro metabolism of ethinyloestradiol mestranol and levonorgesterel by human jejunum mucosa. *Br J Clin Pharmacol* 1981;11:275–278.
14. Connolly ME, Davies DS, Dollery CT, Morgan CD, Paterson JW, Sandler M. Metabolism of isoprenaline in dog and man. *Br J Pharmacol* 1972;46:458–472.
15. Hartiala K. Metabolism of hormones, drugs and other substances by the gut. *Physiol Rev* 1973;53:496–534.
16. Barr WH, Riegelman S. Intestinal drug absorption and metabolism. II. Kinetic aspects of intestinal glucuronide conjugation. *J Pharm Sci* 1970;59:164–168.
17. Del Villar E, Sanchez E, Tephly TR. Morphine metabolism II. Studies on morphine glucuronyl transferase activity in intestinal microsomes of rats. *Drug Metab Dispos* 1974;2:370–374.
18. George CF, Blackwell EW, Davies DS. Metabolism of isoprenaline in the intestine. *J Pharm Pharmacol* 1974;26:265–267.

19. Hoensch HP, Hartmann F. The intestinal enzymatic bio-transformation system: Potential role in protection from colon cancer. *Hepatogastroenterology* 1981;28:221–228.
20. Pantuck EJ, Hsiao KC, Kuntzman R, Conney AH. Intestinal metabolism of phenacetin in the rat: effect of charcoal-broiled beef and rat chow. *Science* 1975;187:744–746.
21. Goldman P, Peppercorn MA, Goldin BR. Drugs metabolized by intestinal microflora. In: *Drug Interactions* (Morselli, Cohen, Garattini, eds.). New York: Raven Press, 1974, pp. 91–102.
22. Blaschke TF. Protein binding and kinetics of drugs in liver diseases. *Clin Pharmacokinet* 1977;2:32–44.
23. Branch RA, Herbert CM, Read AE. Determinants of serum antipyrine half-lives in patients with liver disease. *Gut* 1973;14:569–573.
24. Andreasen PB, Ranek L. Liver failure and drug metabolism. *Scand J Gastroenterol* 1975;10:293–297.
25. Caesar J, Shaldon S, Chiandussi L, Guevara L, Sherlock S. The use of indocyanine green in the measurement of hepatic blood flow and as a test of hepatic function. *Clin Sci* 1961;21:43–57.
26. Henderson JM, Hanna SS. Effective liver blood flow: Determination by galactose clearance. *Can J Surg* 1983;26:129–132.
27. Barr WH, Aceto T, Chung H, Shakur M. Dose-dependent drug metabolism during the absorptive phase. *Rev Can Biol* 1973;32:21–42.
28. Gibaldi M, Perrier D. Route of administration and drug disposition. *Drug Metab Rev* 1974;3:185–199.
29. Murphy PJ, Sullivan HR. Stable isotopes in pharmacokinetic studies. *Annu Rev Pharmacol Toxicol* 1980;20:609–621.
30. Eichelbaum M, Von Unruh G, Somogyi A. Application of stable labelled drugs in clinical pharmacokinetic investigations. *Clin Pharmacokinet* 1982;7:490–507.
31. Nimmo WS. Drugs, diseases and altered gastric emptying. *Clin Pharmacokinet* 1976;1:189–203.
32. Harris FC. Pyloric stenosis: Hold-up of enteric-coated aspirin tablets. *Br J Surg* 1973;60:979–981.
33. Heading RC, Nimmo J, Prescott LF, Tothill P. The dependence of paracetamol absorption on the rate of gastric emptying. *Br J Pharmacol* 1973;47:415–421.
34. Nimmo J, Heading RC, Tothill P, Prescott LF. Pharmacological modifications of gastric emptying: Effects of propantheline and metoclopramide on paracetamol absorption. *Br Med J* 1973;1:587–589.
35. Prescott LF. Gastrointestinal absorption of drugs. *Med Clin* NA 1974;58:907–916.
36. John AH, Jones AJ. Gastroenteritis causing failure of oral contraceptives. *Br Med J* 1975;3:207–208.
37. Pottage A, Nimmo J, Prescott LF. The absorption of aspirin and paracetamol in patients with achlorhydria. *J Pharmacol* 1974;26:144–145.
38. Davies JA, Holt JM. Absorption of cephalexin in diseased and aged subjects. *J Antimicrob Chemother* 1975;1 (supple):69–70.
39. Kramer PA, Chapron DJ, Benson J, Mercile SA. Tetracycline absorption in elderly patients with achlorhydria. *Clin Pharmacol Ther* 1978;23:467–472.
40. Venho VMK, Aukee S, Jussila J, Mattila MJ. Effect of gastric surgery on the gastrointestinal drug absorption in man. *Scand J Gastroenterol* 1975;10:43–47.
41. Antonioli JA, Schelling JL, Steininger E, Borel GA. Effect of gastrectomy and of an anticholinergic drug on the gastrointestinal absorption of a sulfonamide in man. *Int Z Klin Pharmakol Ther Toxikol* 1971;5:212–215.
42. Ochs H, Bodem G, Kodrat G, Savic B, Baur MP. Biologische verfugbarkeit von Digoxin bei

patienten mit und ohne magenresektion nach Billroth II. *Dtsch Med Wochenschr* 1975;100:2430–2434.
43. Hall WH, Doherty JE. Tritiated digoxin XXII. Absorption and excretion in malabsorption syndromes. *Am J Med* 1974;56:437–442.
44. Gerson CD, Lowe EH, Lindenbaum J. Bioavailability of digoxin tablets in patients with gastrointestinal dysfunction. *Am J Med* 1980;69:43–49.
45. Jusko WJ et al. Digoxin absorption from tablets and elixir. The effect of radiation-induced malabsorption. *J Am Med Assoc* 1974;230:1554–1555.
46. Sandle GI, Ward A, Rawlins MD, Record CO. Propranolol absorption in untreated coeliac disease. *Clin Sci* 1982;63:81–85.
47. Johansson EDB, Krai JG. Oral contraceptives after intestinal bypass operations. *J Am Med Assoc* 1976;236:2847.
48. Backman L, Beermann B, Groschinsky-Grind M, Hallberg D. Malabsorption of hydrochlorothiazide following intestinal shunt surgery. *Clin Pharmacokinet* 1979;4:63–68.
49. Kennedy MC, Wade DN. Phenytoin absorption in patients with ileojejunal bypass. *Br J Clin Pharmacol* 1979;7:515–518.
50. Sokol GH, Greenblatt DJ, Lloyd BL, Georgotas A, Allen MD, Harmatz JS, Smith TW, Shader RI. Effect of abdominal radiation therapy on drug absorption in humans. *J Clin Pharmacol* 1978;18:388–396.
51. Parsons RL. Drug absorption in gastrointestinal disease with particular reference to malabsorption syndromes. *Clin Pharmacokinet* 1977:2:45–60.
52. Parsons RL, Kaye CM, Raymond K, Trounce JR, Turner P. Absorption of propranolol and practolol in coeliac disease. *Gut* 1976;17:139–143.
53. Schneider RE, Babb J, Bishop H, Mitchard M, Hoare AM, Hawkins CF. Plasma levels of propranolol in treated patients with coeliac disease and patients with Crohn's disease. *Br Med J* 1976;2:794–795.
54. Melander A, Kahlmeter G, Kamme C, Ursing B. Bioavailability of metronidazole in fasting and non-fasting healthy subjects and in patients with Crohn's disease. *Eur J Clin Pharmacol* 1977;12:69–72.
55. Bergan T, Bjerke Pem, Fausa O. Pharmacokinetics of metronidazole in patients with enteric disease compared to normal volunteers. *Chemotherapy* 1981;27:233–238.
56. Shaffer JA, Williams SE, Turnberg LA, Houston JB, Rowland M. Absorption of prednisolone in patients with Crohn's disease. *Gut* 1983;24:182–186.
57. Milsap RL, George DE, Szefler SJ, Murray KA, Lebenthal E, Jusko WJ. Effects of inflammatory bowel disease on absorption and disposition of prednisolone. *Dig Dis Sci* 1983;28:161–168.
58. Kitis G, Lucas ML, Schneider RE, Bishop H, Sargent A, Blair JA, Allan RN. Jejunal acid microclimate and its effects on absorption of folic acid and propranolol. *Gut* 1983;24:A438.
59. Parsons RL, Kaye CM, Raymond K. Pharmacokinetics of salicylate and indomethacin in coeliac disease. *Eur J Clin Pharmacol* 1977;11:473–477.
60. Parsons RL, Paddock GM. Absorption of two antibacterial drugs, cephalexin and co-trimoxazole in malabsorption syndromes. *J Antimicrob Chemother* 1975;1:(Suppl):59–67.
61. Parsons RL, Paddock GM, Hossack GM, Hailey DM. Antibiotic absorption in Crohn's disease. In: *Chemotherapy, Vol 4, Pharmacology of Antibiotics* (Williams, Gebbes, eds). New York: Plenum Press, 1976, pp. 219–229.
62. Holt S, Heading RC, Clements JA, Tothill P, Prescott LF. Acetaminophen absorption and metabolism in celiac disease and Crohn's disease. *Clin Pharmacol Ther* 1981;30:232–238.
63. Wilkinson GR, Schenker S. Drug disposition and liver disease. *Drug Metab Rev* 1976;4:139–175.

64. Wood AJJ, Kornhauser DM, Wilkinson GR, Shand DG, Branch RA. The influence of cirrhosis on steady-state blood concentrations of unbound propranolol after oral administration. *Clin Pharmacokinet* 1978;3:478–487.
65. Regardh C-G, Jordo L, Ervik M, Lundborg P, Olsson R, Ronn O. Pharmacokinetics of metoprolol in patients with hepatic cirrhosis. *Clin Pharmacokinet* 1981;6:375–388.
66. Neal EA, Meffin PJ, Gregory PB, Blaschke TF. Enhanced bioavailability and decreased clearance of analgesics in patients with cirrhosis. *Gastroenterology* 1979;77:96–102.
67. Kleinbloesem, Van Harten J, Wilson JPH, Danhof M, Van Brummelen P, Breimer DD. Nifedipine: Kinetics and hemodymanic effects in patients with liver cirrhosis after intravenous and oral administration. *Clin Pharmacol Ther* 1986;40:21–28.
68. Adjepon-Yamoah KK, Nimmo J, Prescott LF. Gross impairment of hepatic drug metabolism in a patient with chronic liver disease. *Br Med J* 1974;4:387–388.
69. Thomson PD, Melmon KL, Richardson JA, Cohn K, Steinbrunn W, Cuolihee R, Rowland M. Lidocaine pharmacokinetics in advanced heart failure, liver disease, and renal failure in humans. *Ann Intern Med* 1973;78:499–508.
70. Williams RL, Blaschke TF, Meffin PJ, Melmon KL, Rowland M. Influence of viral hepatitis on the disposition of two compounds with high hepatic clearance: Lidocaine and indocyanine green. *Clin Pharmacol Ther* 1976;20:290–299.
71. Piafsky KM, Sitar DS, Rangno RE, Ogilvie RI. Theophylline disposition in patients with hepatic cirrhosis. *N Engl J Med* 1977;296:1495–1497.
72. Klotz U, Arant GR, Hoyumpa A, Schenker S, Wilkinson GR. The effects of age and liver disease on the disposition and elimination of diazepam in adult man. *J Clin Invest* 1975;55:347–359.
73. Greenblatt DJ, Harmatz JJ, Shader RI. Factors influencing diazepam pharmacokinetics: Age, sex, and liver disease. *Int J Clin Pharmacol Biopharm* 1978;16:177–179.
74. Shull HJ, Wilkinson GR, Johnson R, Schenker S. Normal disposition of oxazepam in acute viral hepatitis and cirrhosis. *Ann Intern Med* 1976;84:420–425.
75. Kraus JW, Marshall JP, Johnson R, Wilkinson GR, Schenker S. Lorazepam elimination in liver disease. *Gastroenterology* 1977 (abstract);73:1228.
76. Witassek F, Bircher J, Huguenin P, Preisig R. Abnormal glucuronidation of zomepirac in patients with cirrhosis of the liver. *Hepatology* 1983:3:415–422.
77. Thiessen JJ, Sellers EM, Denbeigh P, Dolman L. Plasma protein binding of diazapam and tolbutamide in chronic alcoholics. *J Clin Pharmacol* 1976;16:345–351.
78. Schenker S, Hoyumpa AM Jr, Wilkinson GR. The effect of parenchymal liver disease on the disposition and elimination of sedatives and analgesics. *Med Clin NA* 1975;59:877–896.
79. Anon. Safe prescribing in liver disease. *Br Med J* 1973;2:193–194.
80. Levy G, Tsuchiya T. Effect of activated charcoal on aspirin absorption in man. Part I. *Clin Pharmacol Ther* 1972;13:317–322.
81. Neuvonen PJ, Elonen E. Effect of activated charcoal on absorption and elimination of phenobarbitone, carbamazepine and phenylbutazone in man. *Eur J Clin Pharmacol* 1980;17:51–57.
82. Northcott RC, Stiel JN, Hollifield JW, Stant EG Jr. The influence of cholestyramine on thyroxine absorption. *J Am Med Assoc* 1969;208:1857–1861.
83. Robinson DS, Benjamin DM, McCormack JJ. Interaction of warfarin and nonsystemic gastrointestinal drugs. *Clin Pharmacol Ther* 1971;12:491–495.
84. Vander Meer JWH, Scheijgronel HW, Heykants J, Van Cutsem J, Brugmans J. The influence of gastric acidity on the bio-availability of ketoconazole. *J Antimicrob Chemo* 1980;6:552–554.
85. Hurwitz A. Antacid therapy and drug kinetics. *Clin Pharmacokinet* 1977;2:269–280.
86. Prescott LF. Gastric emptying and drug absorption. *Br J Clin Pharm* 1974;1:189–190.

87. Steinberg WM, Lewis JH. Mylanta II inhibits the absorption of cimetidine. *Gastroenterology* 1980;78:1269.
88. Steinberg WM, Lewis JH, Katz DM. Antacids inhibit absorption of cimetidine. *N Engl J Med* 1982;307:400–404.
89. Gugler R, Brand M, Somogyi A. Impaired cimetidine absorption due to antacids and metoclopramide. *Eur J Clin Pharmacol* 1981;20:225–228.
90. Russell WL, Lopez LM, Normann SA, Doering PL, Guild RT. Effect of antacids on predicted steady-state cimetidine concentrations. *Dig Dis Sci* 1984;29:385–389.
91. Mahachai V, Jamali F, Thomson ABR. Effect of combination of antacids and cimetidine on 24-hour intragastric acidity in patients with asymptomatic duodenal ulcer. *Clin Therapeut* 1984;6:808–823.
92. Harcourt RS, Hamburger M. The effect of magnesium sulfate in lowering tetracycline blood levels. *J Lab Clin Med* 1957;50:464–468.
93. Hurwitz A. Antacid therapy and drug kinetics. *Clin Pharmacokinet* 1977;2:269–280.
94. Manninen V, Apajalahti A, Melin J, Karesoja M. Altered absorption of digoxin in patients given propantheline and metoclopramide. *Lancet* 1973;1:398–400.
95. Levy G, Gibaldi M, Procknal JA. Effect of an anticholinergic agent on riboflavin absorption in man. *J Pharm Sci* 1972;61:798–799.
96. Gibbons DO, Lant AF. Effects of intravenous and oral propantheline and metoclopramide on ethanol absorption. *Clin Pharmacol Ther* 1975;17:578–584.
97. Nimmo J. The influence of metoclopramide on drug absorption. *Postgrad Med J* 1973;49(Suppl) 25–28.
98. Gothoni G, Pentikainen P, Vapaatalo HI, Hackman R, Afbjorksten K. Absorption of antibiotics: Influence of metoclopramide and atropine on serum levels of pivampicillin and tetracycline. *Ann Clin Res* 1972;4:228–232.
99. Charles BG, Renshaw PJ, Kay JJ, Ravenscroft PJ. Effect of metoclopramide on the bioavailability of long-acting propranolol. *Br J Clin Pharmacol* 1981;11:517–518.
100. Somogyi A, Gugler R. Drug interactions with cimetidine. *Clin Pharmacokinet* 1982;7:23–41.
101. Jamali F, Mahachai V, Reilly P, Thomson ABR. Lack of pharmacokinetic interaction between cimetidine and pirenzepine. *Clin Pharmacol Ther* 1985;38:325–330.
102. Alvan G, Piafsky k, Lind M, Von Bahr C. Effect of pentobarbital on the disposition of alprenolol. *Clin Pharmacol Ther* 1977;22:316–321.
103. Data JL, Wilkinson GR, Nies AS. Interaction of quinidine with anticonvulsant drugs. *N Engl J Med* 1976;294:699–702.
104. Neuvonen PJ, Penttila O. Interaction between doxycycline and barbiturates. *Br Med J* 1974;1:535–536.
105. Aitio M-L, Mansury L, Tala E, Haataja M, Aitio A. The effect of enzyme induction on the metabolism of disopyramide in man. *Br J Clin Pharmacol* 1981;11:279–285.
106. O'Reilly RA. Interaction of sodium warfarin and rifampin. Studies in man. *Ann Intern Med* 1974;81:337–340.
107. Syvalahti EKG, Pihlajamaki KK, Iisalo EJ. Rifampicin and drug metabolism. *Lancet* 1974;2:232–233.
108. Ahmad D, Mathur P, Ahuja S, Henderson R, Carruthers G. Rifampicin-quinidine interaction. *Br J Dis Chest* 1979;73:409–411.
109. Bennett PN, John VA, Witmarsh VB. Effect of rifampicin on metoprolol and antipyrine kinetics. *Br J Clin Pharmacol* 1982;13:387–391.
110. Carrum G, Egan JM, Abernethy DR. Diltiazem treatment impairs hepatic drug oxidation: Studies of antipyrine. *Clin Pharmacol Ther* 1986;40:140–143.

111. Speek KV, Patwardhan RV, Avant GR, Mitchell MC, Schenker S. Inhibition of microsomal drug metabolism by histamine H_2-receptor antagonists studied in vivo and in vitro in rodents. *Gastroenterology* 1982;82:89–96.
112. Klotz U, Anttila V, Reimann I. Cimetidine/diazepam interaction. *Lancet* 1979;2:699.
113. Desmond PV, Patwardhan RV, Schenker S, Speek KV Jr. Cimetidine impairs elimination of chlordiazepoxide (Librium) in man. *Ann Intern Med* 1980;93:266–268.
114. Feely J, Wilkinson GR, Wood AJJ. Reduction of liver blood flow and propranolol metabolism by cimetidine. *N Engl J Med* 1981;304:692.
115. Somogyi A, Gugler R. Drug interactions with cimetidine. *Clin Pharmacokinet* 1982;7:23–41.
116. Klotz U, Reimann I. Influence of cimetidine on the pharmacokinetics of desmethyldiazepam and oxazepam. *Eur J Clin Pharmacol* 1980;18:517–520.
117. Reimann IW, Klotz U, Siems B, Frolich JC. Cimetidine increases steady state plasma levels of propranolol. *Br J Clin Pharmacol* 1981;12:785–790.
118. Reimann IW, Klotz U, Froelich JC. Effects of cimetidine and ranitidine on steady state propranolol kinetics and dynamics. *Clin Pharmacol Ther* 1982;32:749–757.
119. Heagerty AM, Donovan MA, Castleleden CM, Pohl JF, Patel L. The influence of histamine (H_2) antagonists on propranolol pharmacokinetics. *Int J Clin Pharmacol Res* 1982;2:203–205.
120. Feely J, Guy E. Lack of effect of ranitidine on the disposition of lignocaine. *Br J Pharmacol* 1983;15:378–379.
121. Serlin MJ, Sibeon RG, Breckenridge AM. Lack of effect of ranitidine on warfarin action. *Br J Clin Pharmacol* 1981;12:791–794.
122. Nies AS, Evans GH, Shand DG. Regional hemodynamic effects of beta-adrenergic blockade with propranolol in the unanesthetized primate. *Am Heart J* 1973;85:97–102.
123. Ochs HR, Carstens G, Greenblatt DJ. Reduction in lidocaine clearance during continuous infusion and by coadministration of propranolol. *N Engl J Med* 1980;303:373–377.
124. White RJ, Chamberlain DA, Howard M, Smith TW. Plasma concentrations of digoxin after oral administration in the fasting and postprandial state. *Br Med J* 1971;1:380.
125. Jaffe JM, Colaizzi JL, Barry H. Effects of dietary components on GI absorption of acetaminophen tablets in man. *J Pharm Sci* 1971;60:1646–1650.
126. MacDonald H, Place VA, Falk H, Darken MA. Effect of food on absorption of sulfonamides in man. *Chemotherapy* 1967;12:282–285.
127. Kirby WMM, Roberts CE, Bardick RE. Comparison of two new tetracyclines with tetracycline and demethylchlortetracycline. *Antimicrob Agents Chemother,* October 31–November 2, 1961;286–292.
128. Klein JO, Finland M. The new penicillins. *N Engl J Med* 1963;269:1019–1025.
129. Welling PG, Huang H, Koch PA, Craig WA, Madsen PO. Bioavailability of ampicillin and amoxicillin in fasted and nonfasted subjects. *J Pharm Sci* 1977;66:549–552.
130. Siegler DI, Bryant M, Burley DM, Citron KM. Effect of meals on rifampicin absorption. *Lancet* 1974;2:197–198.
131. Welling PG, Koch PA, Lau CC, Craig WA. Bioavailability of tetracycline and doxycycline in fasted and nonfasted subjects. *Antimicrob Agents Chemother* 1977;11:462–469.
132. Rosenblatt JE, Barrett JE, Brodie JL, Kirby WM. Comparison of in vitro activity and clinical pharmacology of doxycycline and other tetracylines. *Antimicrob Agents Chemother* November 26–28, 1966;134–141.
133. Overdiek HWPM, Merkus FWHM. Influence of food on the bioavailability of spironolactone. *Clin Pharmacol Ther* 1986;40:531–536.
134. Melander A, Danielson K, Schersten B, Wahlin E. Enhancement of the bioavailability of propranolol and metoprolol by food. *Clin Pharmacol Ther* 1977;22:108–112.

135. Melander A, McLean A. Influence of food intake on presystemic clearance of drugs. *Clin Pharmacokinet* 1983;8:286–297.
136. Liedholm H, Melander A. Concomitant food intake can increase the bioavailability of propranolol by transient inhibition of its presystemic primary conjugation. *Clin Pharmacol Ther* 1986;40:29–36.
137. Basu TK. Effect of protein malnutrition and ascorbic acid levels on drug metabolism. *Can J Physiol Pharmacol* 1983;61:295–301.
138. Pantuck EJ, Pantuck CB, Garland WA, Min BH, Wattenberg LW, Anderson KE, Kappas A, Conney AH. Stimulatory effect of brussel sprouts and cabbage on human drug metabolism. *Clin Pharm Ther* 1979;25:88–95.
139. Castleden CM, Volans CN, Raymond K. The effect of aging on drug absorption from the gut. *Age Aging* 1977;6:138–143.
140. Massoud N. Pharmacokinetic considerations in geriatric patients. In: *Pharmacokinetic Basis for Drug Treatment* (Benet LZ, Massoud N, Gambertoglio JG, eds). New York: Raven Press, 1984, pp. 283–310.
141. Greenblatt DJ, Sellers EM, Shader RT. Drug disposition in old age. *N Engl J Med* 1982;306:1081–1088.
142. Schmucker DL. Aging and drug disposition: An update. *Pharmacol Rev* 1985;37:133–148.
143. Triggs EJ, Nation RL. Pharmacokinetics in the aged: A review. *J Pharmacokinet Biopharm* 1975;3:387–418.
144. Bach B, Hansen JM, Kampmann JP, Rasmussen SN, Skovsted L. Disposition of antipyrine and phenytoin correlated with age and liver volume in man. *Clin Pharmacokinet* 1981;6:389–396.
145. Farah F, Taylor W, Rawlins MD, James D. Hepatic drug acetylation and oxidation: effects of aging in man. *Br Med J* 1977;2:155–156.
146. Vestal RE, Norris AH, Tobin JD, Cohen BH, Shock NW, Andres R. Antipyrine metabolism in man: Influence of age, alcohol, caffeine and smoking. *Clin Pharmacol Ther* 1975;18:425–432.
147. Smithard DJ, Langman MJS. Drug metabolism in the elderly. *Br Med J* 1977;2:520–521.

8

Immunological Injury to the Intestine and Liver in Human Marrow Transplant Recipients

George B. McDonald

1. INTRODUCTION

Over 3000 marrow transplants are now carried out yearly, a number that will surely increase as long-term survival after marrow grafting improves. Patients are returning to their community physicians having been "cured" of their original diseases but suffering the sequelae of the marrow transplant procedure. On another plane, it is of great interest that marrow transplantation creates intestinal and liver diseases that are similar to several puzzling, naturally occurring human illnesses. For example, acute graft-versus-host disease (GVHD) causes a form of inflammatory bowel disease that is most prominent in the ileocecal region. Chronic graft-versus-host disease in the liver looks much like primary biliary cirrhosis, and in the esophagus, like epidermolysis bullosa. This chapter will summarize the clinical presentation and pathophysiology of intestinal and liver injury after transplantation, especially that mediated by the immune system.

George B. McDonald • Gastroenterology/Hepatology Section, Fred Hutchinson Cancer Research Center, and University of Washington School of Medicine, Seattle, Washington 98104.

2. MARROW TRANSPLANT METHODS

Protocols for marrow grafting depend on the disease being treated and on the degree of genetic match between donor and recipient.[1–5] There is also variability among the over 100 transplant centers as to choice and dosage of cytotoxic and immunosuppressive drugs, infection prophylaxis, and management of posttransplant problems. Table 1 lists terms and abbreviations commonly used in transplant centers.

2.1. Pretransplant Conditioning Therapy

Patients with nonmalignant conditions such as marrow aplasia are given pretransplant chemotherapy (usually cyclophosphamide) to ablate their remaining marrow in order to prevent rejection of the donor marrow graft. Patients with malignancy commonly receive high-dose chemoradiotherapy or combination chemotherapy to eliminate malignant cells and to prevent graft rejection. High-dose conditioning therapy may cause liver damage (venocclusive disease of the liver) and a diffuse enteritis. [6,7] The liver damage may be fatal, but the enteritis is usually transient.

2.2. Histocompatibility

Most patients receive *allogeneic* marrow from a genetically similar donor, usually a sibling who shares parental haplotypes with the patient. Human leukocyte antigen (HLA) identity is determined by serological microcytotoxicity assays (for HLA-A and B antigens) and by mixed leukocyte culture reactivity (for HLA-D antigens).[1,4] The closer the genetic match between patient and donor marrow, the less likely there will be a reaction against the patient by the grafted marrow (graft-versus-host disease). However, HLA-matched sibling transplants may still develop GVHD, most likely because of transplantation antigens not detected by current typing techniques. Alternatives to HLA-matched marrows are those from HLA-nonidentical family members, in whom one haplotype is genetically identical and the other haplotype shows phenotypic similarity for one or two but not all HLA determinants. Transplants have also been successfully performed between unrelated donors and recipients who are phenotypically identical for HLA-A, B, and D determinants.[4]

Others receive *autologous* (their own) marrow, aspirated and cryopreserved before intensive chemoradiation therapy is begun. Patients fortunate enough to have an identical twin can receive *syngeneic* marrow and are generally spared

Table 1. Commonly Used Terms and Abbreviations

Conditioning therapy	Preparative regimen given before marrow transplantation in order to prevent graft rejection, eradicate malignant cells, or both. Uses chemotherapy drugs alone or with total-body irradiation. See "Chemoradiotherapy."
Chemoradiotherapy	Pretransplant conditioning therapy combining chemotherapy (usually cyclophosphamide) with total-body irradiation.
TBI	Total-body irradiation, a form of immunosuppression and antitumor therapy.
Day 0	The day of marrow infusion.
Allogeneic transplant	Donor and host are of the same species but are not genetically identical. Most allogeneic transplants are from a sibling who shares parental haplotypes with the patient. Synonymous with allograft.
Syngeneic transplant	From an identical twin.
HLA	Abbreviation for human leukocyte antigen, a set of linked genes on the sixth chromosome. Class I antigens (HLA-A and B) are detected by serological microcytotoxicity assays.
HLA-matched	Donor and patient are HLA-identical. Matched sibling transplants may still develop GVHD in spite of HLA identity, most likely because of transplantation antigens not detected by current typing techniques.
HLA-mismatched	Donor and patient are HLA-nonidentical at one or more loci.
GVHD	Abbreviation for graft-versus-host disease, a symptom complex involving skin, liver, and intestinal mucosa (acute GVHD), and skin, liver, mucus membranes, and salivary glands (chronic GVHD).
Hyperacute GVHD	GVHD that occurs earlier and with greater severity than normal. Seen most often in HLA-mismatched patients and in those not receiving prophylaxis against GVHD.
GVHD prophylaxis	Drugs given after marrow infusion to prevent or lessen the severity of GVHD. Usually includes combinations of methotrexate, cyclosporine, and prednisone.

from severe graft-versus-host disease. There is controversy over whether some of these nonallogeneic graft recipients develop less severe skin GVHD.[8]

2.3. Posttransplant Immunoprophylaxis

Because of the presumed immunological etiology of GVHD, patients receiving allogeneic marrow grafts are usually given prophylactic immunosuppressive drugs.[1,4] In order to be effective, these drugs must be started before GVHD becomes apparent. Methotrexate, cyclosporine, and prednisone are currently used.[9–12] Deleting posttransplant immunosuppression is associated with frequent and severe "hyperacute" GVHD, i.e., GVHD that occurs earlier and with greater severity than normal.[13] Another approach to preventing GVHD involves treating the donor marrow *in vitro* before infusion, to remove immunocompetent cells while retaining hemopoietic stem cells.[1,4]

2.4. Marrow Reconstitution

Conditioning regimens destroy the patient's hemopoietic and immunological systems. Donor marrow cells repopulate the marrow cavities, and in 2–4 weeks peripheral blood counts rise.[1] Over time, all hemopoietic and immune cells are derived from the donor marrow, including plasma cells and macrophages.[14] Patients have pronounced immunological deficiencies for 4–5 months, the time required for host lymphoid tissues to be repopulated by functioning donor cells.[1,4,14] With chronic GVHD patients remain immunologically crippled, suffering from recurrent bacterial infections. Long-term survivors without GVHD lead normal lives without infection.

3. OVERVIEW OF INTESTINAL AND LIVER COMPLICATIONS

Some patients come to marrow transplantation with intestinal and liver diseases that may jeopardize posttransplant survival. Examples are viral hepatitis, peptic ulceration, intestinal infection, perianal cellulitis, and fungal or viral esophagitis. Posttransplant, there are four causes of intestinal and liver disease: (1) chemotherapy or chemoradiotherapy, which causes gut and liver damage that becomes apparent in the first weeks posttransplant; (2) graft-versus-host disease, the acute form of which usually appears between days 20 and 60, and the chronic form from 3 to 15 months posttransplant; (3) infections occurring before full marrow recovery; and (4) side effects of drugs and other therapy such as methotrexate, cyclosporine, antibiotics, or total parenteral nutrition. Chemoradiotherapy toxicity, infections, and the side effects of drugs are discussed in recent reviews.[6,7]

4. ACUTE GRAFT-VERSUS-HOST DISEASE

The syndrome of acute GVHD usually begins 3–6 weeks after transplantation in patients receiving prophylactic immunosuppressive therapy, but may occur as early as day 7 in patients not receiving these drugs.[9,13] A maculopapular skin rash appears on the trunk, soles, palms, and ears. Crampy abdominal pain, nausea, vomiting, watery diarrhea, and jaundice may then appear.[6,7,16,17] A widely used system for grading the overall severity of acute GVHD lists mild skin GVHD as grade 1; grade 2 is moderate disease in multiple organs; grade 3 is severe disease; and grade 4 is life-threatening GVHD.[16,18] The accuracy of this grading system is dependent on clinical findings, exclusion of drug toxicity and infection as causes of tissue damage, and biopsies of skin, gut, and liver. Twenty-five to fifty percent of engrafted allogeneic recipients develop grade 2–4 GVHD; 30–60% of these patients die of GVHD or its related infectious complications.[9,16,19]

4.1. Acute Intestinal GVHD

Patients with acute GVHD may have few abdominal complaints or may have unceasing pain, profuse diarrhea, nausea, protein loss, intestinal bleeding, and sepsis.[6] The watery diarrhea persists in high volume, up to 10 liters daily, even in the absence of oral intake. The diarrheal fluid usually contains cellular debris, occult blood, and large amounts of exuded serum proteins.[6,20] Serum proteins are correspondingly low. Abdominal pain increases after food ingestion but may occur spontaneously as large volumes of fluid course through the intestine. Anorexia, nausea, and vomiting are common but not specific, as they are also manifestations of herpesvirus infection or drug toxicity.[21]

Plain x rays of the abdomen and barium contrast studies of the intestine demonstrate widespread mucosal and submucosal edema, most prominent in the distal small intestine.[22,23] Transit is rapid, and the barium is often diluted with excessive luminal fluid. Mucosal ulcerations and pneumatosis intestinalis may be present. The high-volume diarrhea typical of acute GVHD probably results from both increased small intestinal secretions and reduced fluid absorption, particularly in the ileum and right colon. The endoscopic appearance of intestinal GVHD ranges from normal (with only histological abnormalities) to patchy erythema to extensive mucosal friability and sloughing. These severe changes are most prominent in the ileum and cecum, but may occasionally extend from stomach to rectum.[24] Biopsy specimens of grossly normal-appearing stomach, duodenum, and rectum often show histological changes of acute GVHD if there is extensive midgut disease.[21,25–27]

Histology of intestinal mucosa from patients with acute GVHD shows abnormalities that range from necrosis of individual crypt epithelial cells to total

denudation of epithelium.[24–27] Many of these histological changes, though characteristic enough of acute GVHD to be very useful diagnostically, are not specific. By light microscopy, the first change is necrosis of individual cells in the proliferative zone of intestinal and gastric crypts (apoptosis).[25,26] Debris from cell necrosis is enveloped in spaces at the base or side of crypts (apoptotic bodies). Apoptosis, as a mechanism of cell injury, is thought to be caused by immune mechanisms. Support for this hypothesis as it relates to GVHD comes from electron microscopic studies of human intestinal epithelium.[28] Viable lymphocytes were seen extending pseudopods toward nuclear membranes in crypt cells from mucosa involved with GVHD, but not from controls.[28] These morphological findings are consistent with T-cell-mediated cytotoxicity in which there is conjugation to the target membrane, activation of lymphotoxins, and lysis of target cells.[29] The cytotoxic lymphocyte may then detach to encounter other target cells. Isolated apoptotic bodies in acute GVHD are not accompanied by lymphocyte infiltration, in contrast to epithelial apoptosis seen in forms of inflammatory bowel disease.[25,30] Confounding the theory that mobile cytotoxic lymphocytes are the cause of apoptosis in GVHD are recent studies showing that isolated apoptotic bodies can be found in immunocompromised patients with acquired immune deficiency syndrome (AIDS) and severe T-cell deficiency[31,32] and in patients with cytomegalovirus enteritis.[33] Perhaps apoptosis in these situations is related to immune recognition of virally infected cells, as opposed to recognition of crypt cell-surface differentiation antigens in acute GVHD. This hypothesis remains untested.

The lymphocytes present in rectal biopsy lamina propria of patients with acute GVHD show different immunostaining properties, compared to both normal and non-GVHD posttransplant biopsy specimens. In marrow recipients without GVHD, there is a decrease in T lymphocytes in the lamina propria,[34] consistent with previous findings of lymphocyte depletion in Peyer's patches and mesenteric lymph nodes of allograft recipients.[35] Lymphoid atrophy is related to conditioning chemoradiotherapy, not GVHD.[36] In patients with GVHD, T lymphocytes were increased in both lamina propria and epithelium due to an increase in suppressor-cytotoxic ($T8^+$) cells. [34,37] Lymphocytes did not express activation markers,[34] consistent with diminished functional capacity of peripheral blood lymphocytes from patients with GVHD.[14] These studies, along with findings that show a lower incidence of acute GVHD among patients transplanted with T-lymphocyte-depleted donor marrow, [38,39] make cytotoxic-suppressor T cells the most likely effectors of crypt cell apoptosis.

In patients with acute GVHD, there is often a progression from isolated apoptotic lesions to dropout of whole crypts, then groups of crypts, and eventually denudation of the epithelium.[6,24,27,35] These severe changes are most prominent in the ileum and ascending colon. In these later stages, the lamina propria and smooth muscle may be infiltrated with lymphocytes and plasma

cells.[27,35] There is evidence that some of the intestinal damage in acute GVHD is due to the "innocent bystander" phenomenon, in which allogeneic reactions against host lymphoid cells cause the local release of toxic lymphokines.[40,41] There is morphological evidence for this phenomenon in humans. Electron microscopy of rectal mucosa and skin biopsy specimens from patients with acute GVHD shows damage to melanocytes and enteroendocrine cells in the vicinity of less differentiated cells, the probable targets of cytotoxic T cells.[28,42] Cytoplasmic granules of both natural killer and cytotoxic T cells lack antigenic specificity and may damage nontarget cells.[29]

The role of microorganisms in the pathogenesis of GVHD remains uncertain. One hypothesis states that immunodeficiency results from a donor–host lymphoid interaction, allowing microbial invasion of the mucosa.[43] Studies in animals and humans with GVHD show enhanced survival and reduced mucosal necrosis when the graft recipient is germ-free or decontaminated (reviewed in ref. 6). Although superinfection of intestinal mucosa already affected by GVHD is common, many cases of florid intestinal GVHD are without clinical or histological evidence of infection.[26,27,35] Nonetheless, luminal organisms may perpetuate enterocolitis by several mechanisms and contribute to continued disease.[44,45]

Treatment of acute GVHD is with immunosuppressive drugs, such as prednisone, antithymocyte globulin, and/or cyclosporine.[9,10,16,19] Survival correlates with the severity of GVHD: 88% of patients with aplasia who achieved allogeneic engraftment and grade 0–1 GVHD survive, compared to 45% survival in those developing grades 2–4 GVHD.[46]

4.2. Acute GVHD of the Liver

Hepatic dysfunction occurs in more than 80% of patients after allogeneic transplantation, of whom roughly half have acute GVHD.[7,47–49] Most patients with liver GVHD have skin and intestinal GVHD as well, but liver changes typical of GVHD can be seen without other organs being involved.

In acute GVHD, liver abnormalities include cholestasis and mild hepatocellular necrosis.[48,49] Serum alkaline phosphatase levels may increase to 20 times their normal upper limit, with parallel increases in total serum bilirubin.[17] Elevations of serum hepatocellular enzymes seldom exceed 10 times normal. Features of severe hepatic dysfunction such as clotting factor deficiency and encephalopathy are unusual, especially early in the course of acute GVHD.

Histology of liver biopsy specimens shows a mild, nonspecific lobular hepatitis in the early phases of acute GVHD,[48–50] findings that may not be helpful in differentiating acute GVHD from viral or drug hepatitis.[50] After GVHD has been present for 1–2 weeks, cholestasis and characteristic abnormalities of the small septal and interlobular bile ducts are found.[48–51] These changes in-

clude irregular nuclear enlargement, cytoplasmic swelling with eosinophilia, and segmental destruction of the duct wall.[49,50,52] Lymphocyte content in portal areas varies greatly and may not be increased. Immunohistological stains show that portal areas contain significant numbers of natural killer cells in biopsies from patients with GVHD compared to both normal and non-GVHD biopsies, and increased suppressor-cytotoxic ($T8^+$) cells with GVHD relative to graft recipients without GVHD.[53] Lymphocytes do not express activation markers or appear to infiltrate bile duct epithelium.[53] An electron microscopic study of human liver GVHD demonstrated lymphocytes with elongated processes in contact with cell membranes of bile duct epithelium.[54] Adjacent cells were abnormal, with swelling of the endoplasmic reticulum and mitochondria, and necrosis of basement membranes. There were also areas showing close contact between lymphocytes and hepatocytes.[54] These ultrastructure changes of acute GVHD are similar to those in pig liver allograft rejections.[55]

The "innocent bystander" phenomenon may obtain in the liver as well as in skin and gut. One proposed mechanism for the recruitment of cytotoxin-bearing cells involves the expression of class II major histocompatibility complex antigens by bile duct cells.[53,56–58] Expression of HLA-DR antigens is mediated by the lymphokine γ-interferon.[59,60] Normally, bile duct epithelium does not express class II antigens,[61] but in some inflammatory states, such as primary biliary cirrhosis, allograft rejection, and GVHD, bile duct cells become HLA-DR immunopositive.[51,53,62,63] However, expression of HLA-DR antigens in other epithelia is a nonspecific response to lymphoctye infiltration.[64] The mechanisms for bile duct injury in acute GVHD remain poorly understood.

Treatment of liver GVHD is the same as for GVHD in general.[9,10,16,19] Resolution of the jaundice and elevated alkaline phosphatase levels occurs slowly with successful treatment.

5. CHRONIC GRAFT-VERSUS-HOST DISEASE

Chronic GVHD affects 30–50% of long-term survivors of allogeneic transplantation.[19] In most, the signs and symptoms are an extension of acute GVHD, but in 25% of patients chronic GVHD develops de novo 3–15 months after transplantation, after a period of well-being.[16,19,65,66] Common clinical findings include skin disease, buccal mucositis, keratoconjunctivitis, esophageal and vaginal strictures, chronic liver disease, and polyserositis. Patients with chronic GVHD have severe immunological deficiency as long as the disease remains active, with an inability to mount a humoral antibody response to bacterial antigens.[14]

There are several independent risk factors for development of chronic

GVHD. These include moderate-to-severe acute GVHD, increasing age of the patient, and infusion of unirradiated donor buffy coat cells.[19,65,66]

Treatment of chronic GVHD with immunosuppressive drugs favorably affects the natural history of extensive chronic GVHD and may arrest the disease in one-third of cases.[9,19,65] The prognosis is better among patients with chronic GVHD limited to the skin and liver than among patients with multisystem disease. Late-onset chronic GVHD seen in older patients and in those given buffy coat cell infusions is generally less severe and more responsive to therapy than the chronic GVHD that develops after acute GVHD.[1,19]

5.1. Esophageal Abnormalities in Chronic GVHD

Esophageal lesions include generalized desquamation of the epithelium in the upper half of the esophagus, webs, ringlike narrowings, and upper esophageal strictures.[67,68] These changes are similar to the naturally occurring autoimmune diseases bullous pemphigoid and dystrophic epidermolysis bullosa.[69,70] The analogy between chronic GVHD and autoimmune disease is further supported by the finding of circulating autoantibodies and immunoglobulin, and of complement deposits at the dermal-epidermal junction of the skin.[65,71,72]

Symptoms of dysphagia, painful swallowing, retrosternal pain, aspiration, and weight loss occur in 15–20% of patients with extensive chronic GVHD.[67,68] Esophageal dysfunction may contribute to pulmonary disease in some patients who develop airway obstruction associated with chronic GVHD.[73] Manometric studies show that patients with esophageal desquamation clear acid poorly and have nonspecific motor abnormalities.[67] Acid peptic reflux may cause retrosternal pain, distal esophagitis, and strictures in these patients, who are at additional risk for reflux damage because of salivary gland dysfunction from chronic GVHD.

The diagnosis is made by X-ray examination of the upper esophagus, particularly to demonstrate webs, rings, and strictures.[68] The shallow epithelial peeling evident on endoscopy is often not seen by X ray. In some patients, the sloughed mucosa hangs in shreds into the lumen. Endoscopy and dilation of strictures and webs carry a high risk of perforation in these patients. Histological findings in esophageal biopsy specimens are similar to those in skin and oral mucosa—infiltration with lymphocytes, neutrophils, and eosinophils; necrosis of individual squamous cells in the basal layer; and desquamation of superficial layers.[27,67,68,74] Although the skin manifestations of chronic GVHD have been likened to scleroderma, the esophagus from patients with chronic GVHD does not show either muscle or neuronal fibrosis.[67]

Although esophageal lesions respond favorably to immunosuppressive drugs

in parallel to skin and oropharyngeal involvement, residual webs and strictures from submucosal fibrosis may persist, requiring periodic dilation.[67] Insidious weight loss, particularly in children, is one manifestation of progressive upper esophageal stenosis.

5.2. Chronic GVHD of the Liver

Liver disease occurs in 90% of patients with chronic GVHD, either as part of a multisystem illness or as a more limited disease involving just skin and liver.[16,19] Most patients have fluctuating jaundice, anorexia, and weight loss. Typical laboratory features include elevations of serum alkaline phosphatase to 5–10 times normal, and of serum hepatocellular enzymes to 3–6 times normal.

Histological changes of chronic GVHD of the liver are usually quite characteristic: portal enlargement and fibrosis, bile duct dropout, Kupffer cell proliferation, and profound cholestasis.[48–50] Lymphocytes in the portal areas are often more prominent than in acute GVHD, perhaps because of more complete lymphoid reconstitution. Parenchymal inflammation and acidophil bodies are usually less prominent.[49,50] The most striking changes in severe, long-standing chronic GVHD are a marked reduction in the number of small bile ducts, similar to the histological picture in long-standing primary biliary cirrhosis.[49,50] Micronodular cirrhosis has been described in a small number of patients with chronic GVHD.[48,75,76] Without the ability to exclude non-A, non-B hepatitis viruses in these cases, the cause of cirrhosis cannot be ascertained. Non-A, non-B (NANB) hepatitis virus must be prevalent among marrow graft recipients, given a 30% frequency of chronic hepatitis among patients before transplantation and average postgrafting blood product transfusion requirements in excess of 15 units.[77,78]

Segmental destruction of the bile duct epithelium in chronic GVHD is not a feature of NANB hepatitis, but bile duct cell multilayering, invasion by lymphocytes, and epithelial vacuolization have been described with NANB hepatitis.[79,80] Clinically, persistently elevated alkaline phosphatase is more consistent with chronic GVHD than chronic viral hepatitis. Other differential diagnoses include infiltrative liver diseases such as tuberculosis, drug-related liver injury (e.g., trimethoprim-sulfamethoxazole, cyclosporine, or azathioprine), and extrahepatic biliary obstruction from gallstones or biliary sludge.[7]

5.3. Intestinal Changes in Chronic GVHD

In some patients, extensive intestinal mucosal disease of acute GVHD persists beyond day 100, as patients develop signs of chronic GVHD. Acute GVHD in these patients is relatively refractory to immunosuppressive drugs. The intestinal disease may progress to stenotic segments of mid- and distal ileum, lead-

ing to partial obstruction.[24] We have resected involved segments of intestine in two patients, with relief of symptoms in one.[22,24]

Before immunosuppressive therapy for early chronic GVHD became standard, we saw four patients with extensive chronic GVHD and malabsorption, diarrhea, abdominal pain, and malnutrition.[27,71] At autopsy, there was extensive fibrosis in the submucosal and subserosal areas of the intestine, along with hyalinization of submucosal blood vessels.[27,71] More recently, we have seen several patients with chronic GVHD, malabsorption, and jejunal bacterial overgrowth. Bacterial overgrowth represents an intestinal manifestation of the immunodeficiency of chronic GVHD.[14,81]

6. SUMMARY

It is not intuitive that marrow grafting should have such profound effects on gut and liver, but toxicity from conditioning therapy, graft-versus-host diseases, and the consequences of immunodeficiency remain obstacles to long-term survival after marrow transplantation. Immunological mechanisms of damage in GVHD are not well understood, but both T-lymphocyte–epithelial cell interaction and nonspecific release of lymphokines are probably responsible for intestinal mucosal and biliary damage in acute GVHD. An understanding of the mechanisms of tissue damage in this milieu may provide insight into the pathogenesis of naturally occurring, immune-mediated disorders of the intestine and liver.

REFERENCES

1. Storb R, Thomas ED. Allogeneic bone marrow transplantation. *Immunol Rev* 1983;71:77–102.
2. Storb R, Thomas ED, Buckner CD et al. Marrow transplantation for aplastic anemia. *Semin Hematol* 1984;21:27–35.
3. O'Reilly RJ, Brochstein J, Dinsmore R, Kirkpatrick D. Marrow transplantation for congenital disorders. *Semin Hematol* 1984;21:188–221.
4. Martin PJ, Hansen JA, Storb R, Thomas ED. Human marrow transplantation: an immunological perspective. *Adv Immunol* 1987;40:379–438.
5. Parkman R. The application of bone marrow transplantation to the treatment of genetic diseases. *Science* 1986;232:1373–1378.
6. McDonald GB, Shulman HM, Sullivan KM, Spencer GD. Intestinal and hepatic complications of human bone marrow transplantations. *Gastroenterology* 1986;90:460–477, 770–784.
7. McDonald GB, Shulman HM, Wolford JL, Spencer GD. Liver disease after human marrow transplantation. *Semin Liver Dis* 1987;7:210–229.
8. Ferrara JLM. Syngeneic graft-vs-host disease. *Arch Dermatol* 1987;123:741–742.
9. Deeg HJ, Storb R. Acute and chronic graft-versus-host disease: clinical manifestations, prophylaxis, and treatment. *J Natl Cancer Inst* 1986;76:1325–1328.

10. Neudorf S, Filipovich A, Ramsay NKC, Kersey J. Prevention and treatment of acute graft-versus-host disease. *Semin Hematol* 1984;21:91–100.
11. Storb R, Deeg HJ, Whitehead J et al. Marrow transplantation for leukemia: A controlled trial of a combination of methotrexate and cyclosporine versus cyclosporine alone for prophylaxis of acute graft-versus-host disease. *N Engl J Med* 1986;314:729–735.
12. Storb R, Deeg HJ, Farewell V et al. Marrow transplantation for severe aplastic anemia: Methotrexate alone compared with a combination of methotrexate and cyclosporine for prevention of acute graft-versus-host disease. *Blood* 1986;68:119–125.
13. Sullivan KM, Deeg HJ, Sanders J et al. Hyperacute graft-versus-host disease in patients not given immunosuppression after allogeneic marrow transplantation. *Blood* 1986;67:1172–1175.
14. Witherspoon RP, Lum LG, Storb R. Immunologic reconstitution after human marrow grafting. *Semin Hematol* 1984;21:2–10.
15. Atkinson K, Farewell V, Storb R et al. Analysis of late infections after human bone marrow transplantation. Role of genotypic nonidentity between marrow donor and recipient and of nonspecific suppressor cells in patients with chronic graft-versus-host disease. *Blood* 1982;60:714–720.
16. Sullivan KM. Graft-versus-host disease. In: *Clinical Bone Marrow Transplantation* (Blume KG, Petz LD, eds). New York: Churchill Livingstone, 1983, pp. 91–129.
17. Glucksberg H, Storb R, Fefer A et al. Clinical manifestations of graft-versus-host disease in human recipients of marrow from HLA-matched sibling donors. *Transplantation* 1974;18:295–304.
18. Weiden PL, The Seattle Marrow Transplant Team. Graft-versus-host disease in allogeneic marrow transplantation. In: *Biology of Bone Marrow Transplantation* (Gale RP, Fox DF, eds). New York: Academic Press, 1980, pp. 37–48.
19. Sullivan KM. Acute and chronic graft-versus-host disease in man. *Int J Cell Cloning* 1986;4:42–93.
20. Weisdorf SA, Salati LM, Longsdorf JA et al. Graft-versus-host disease of the intestine: A protein losing enteropathy characterized by fecal $alpha_1$-antitrypsin. *Gastroenterology* 1983;85:1076–1081.
21. Spencer GD, Hackman RC, McDonald GB et al. A prospective study of unexplained nausea and vomiting after marrow transplantation. *Transplantation* 1986;42:602–607.
22. Fisk JD, Shulman HM, Greening RR, McDonald GB, Sale GE, Thomas ED. Gastrointestinal radiographic features of human graft-versus-host disease. *Am J Roentgenol* 1981;136:329–336.
23. Maile CW, Frick MP, Crass JR et al. The plain abdominal radiograph in acute intestinal graft-versus-host disease. *Am J Roentgenol* 1985;145:289–292.
24. Spencer GD, Shulman HM, Myerson D, Thomas ED, McDonald GB. Diffuse intestinal ulceration after marrow transplantation: A clinical-pathological study of 13 patients. *Hum Pathol* 1986;17:621–633.
25. Epstein RJ, McDonald GB, Sale GE, Shulman HM, Thomas ED. The diagnostic accuracy of the rectal biopsy in acute graft-versus-host disease: A prospective study of thirteen patients. *Gastroenterology* 1980;78:764–771.
26. Snover DC, Weisdorf SA, Vercellotti GM et al. A histopathologic study of gastric and small intestinal graft-versus-host disease following allogeneic bone marrow transplantation. *Hum Pathol* 1985;16:387–392.
27. McDonald GB, Sale GE. The human gastrointestinal tract after allogeneic marrow transplantation. In: *The Pathology of Bone Marrow Transplantation* (Sale GE, Shulman HM, eds). New York: Masson, 1984, pp. 77–103.
28. Gallucci BB, Sale GE, McDonald GB, Epstein R, Shulman HM, Thomas ED. The fine structure of human rectal epithelium in acute graft-versus-host disease. *Am J Surg Pathol* 1982;6:293–305.

29. Podack ER. The molecular mechanism of lymphocyte-mediated tumor cell lysis. *Immunol Today* 1985;6:21–27.
30. Otto HF. The interepithelial lymphocytes of the intestine. Morphological observations and immunological aspects of intestinal enteropathy. In: *Current Topics of Pathology*, Vol 57 (Altmann KW et al, eds). Berlin: Springer-Verlag, 1973, pp. 81–121.
31. Kotler DP,Gaetz HP, Lange M, Klein EB, Holt PR. Enteropathy associated with the acquired immunodeficiency syndrome. *Ann Intern Med* 1984;101:421–428.
32. Snover DC, Filipovich AH, Ramsay NKC, Weisdorf SA, Kersey JH. Graft-versus-host disease-like histopathological findings in pre-bone-marrow transplantation biopsies of patients with severe T cell deficiency. *Transplantation* 1985;39:95–97.
33. Snover DC. Mucosal damage simulating acute graft-versus-host-reaction in cytomegalovirus colitis. *Transplantation* 1984;39:669–670.
34. Dilly SA, Sloane JP. Changes in rectal leucocytes after allogeneic bone marrow transplantation. *Clin Exp Immunol* 1987;67:151–158.
35. Beschorner WE, Yardley JH, Tutschka P, Santos G. Deficiency of intestinal immunity with graft-versus-host disease in humans. *J Infect Dis* 1981;144:38–46.
36. Kolb H, Sale GE, Lerner KG, Storb R, Thomas ED. Pathology of acute graft-versus-host disease in the dog. An autopsy study of ninety-five dogs. *Am J Pathol* 1979;96:581–594.
37. Weisdorf SA, Platt JL, Snover DC, Kersey JH, Scharp HL. In situ analysis of T and killer lymphocyte subpopulations in rectal biopsies from bone marrow transplant patients. *Gastroenterology* 1983;84:1348 (abstract).
38. Filipovich AH, McGlave PB, Ramsay NKC et al. Pretreatment of donor bone marrow with monoclonal antibody OKT3 for prevention of acute graft-versus-host disease in allogeneic histocompatible bone-marrow transplantation. *Lancet* 1982;1:1266–1269.
39. Prentice HG, Blacklock HA, Janossy G et al. Use of anti-T-cell monoclonal antibody OKT3 to prevent acute graft-versus-host disease in allogeneic bone-marrow transplantation for acute leukaemia. *Lancet* 1982;1:700–703.
40. Grebe SC, Streilein JW. Graft-versus-host reactions: A review. *Adv Immunol* 1976;22:119–221.
41. Elson CO, Reilly RW, Rosenberg IH. Small intestinal injury in the graft-versus-host reaction: An innocent bystander phenomenon. *Gastroenterology* 1977;72:886–889.
42. Gallucci BB, Shulman HM, Sale GE, Lerner KG, Caldwell LE, Thomas ED. The ultrastructure of the human epidermis in chronic graft-versus-host disease. *AM J Pathol* 1979;95:643–662.
43. van Bekkum DW, Knaan S. Role of bacterial microflora in development of intestinal lesions from graft-versus-host reaction. *J Natl Cancer Inst* 1977;58:787–790.
44. Pollard M, Chang CF, Strivastava KK. The role of microflora in development of graft-versus-host disease. *Trans Proc* 1976;4:533–536.
45. Mee AS, McLaughlin JE, Hodgson HJF, Jewell DP. Chronic immune colitis in rabbits. *Gut* 179;20:1–5.
46. Storb R, Prentice RL, Buckner CD et al. Graft-versus-host disease and survival in patients with aplastic anemia treated by marrow grafts from HLA-identical siblings. Beneficial effect of a protected environment. *N Engl J Med* 1983;308:302–307.
47. Farthing MJG, Clark ML, Sloane JP et al. Liver disease after bone marrow transplantation. *Gut* 1982;23:465–474.
48. Shulman HM, McDonald GB. Liver disease after marrow transplantation. In: *The Pathology of Bone Marrow Transplantation* (Sale GE, Shulman HM, eds). New York: Masson, 1984, pp. 104–135.
49. Snover DC, Weisdorf SA, Ramsay NK et al. Hepatic graft-versus-host disease: A study of the predictive value of liver biopsy in diagnosis. *Hepatology* 1984;4:123–130.

50. Shulman HM, Sharma P, Amos D et al. A coded histologic study of hepatic graft-versus-host disease after human bone marrow transplantation. *Hepatology* 1988;8:463–470.
51. Tanaka M, Umihara J, Chiba S, Ishikawa E. Intrahepatic bile duct injury following bone marrow transplantation. Analysis of pathological features based on three-dimensional and histochemical observation. *Acta Pathol J* 1986;36:1793–1806.
52. Sloane JP, Farthing MJ, Powles RL. Histopathological changes in the liver after allogeneic bone marrow transplantation. *J Clin Pathol* 1980;33:344–350.
53. Dilly SA, Sloane JP. An immunohistological study of human hepatic graft-versus-host disease. *Clin Exp Immunol* 1985;62:545–553.
54. Bernuau D, Gisselbrecht C, Devergie A et al. Histological and ultrastructural appearance of the liver graft-versus-host disease complicating bone marrow transplantation. *Transplantation* 1980;29:236–244.
55. Searle J, Kerr JFR, Battersby C et al. An electron microscopic study of the mode of donor cell death in unmodified rejection of pig liver allografts. *Aust J Exp Biol Med Sci* 1977;55:401–406.
56. Shearer GM, Levy RB. Graft-versus-host-associated immune suppression is activated by recognition of allogeneic murine I-A antigens. *J Exp Med* 1983;157:936–946.
57. Suitters AJ, Lampert IA. Class II antigen induction in the liver of rats with graft-versus-host disease. *Transplantation* 1984;38:194–196.
58. Takacs L, Szende B, Rot AS et al. Expression of MHC Class II antigens on bile duct epithelium in experimental graft-versus-host disease. *Clin Exp Immunol* 1985;60:449–456.
59. Szetein MB, Steeg PS, Johnson HM, Oppenheim JJ. Regulation of human peripheral blood monocyte DR antigen expression in vitro by lymphokines and recombinant interferons. *J Clin Invest* 1984;73:556–565.
60. Walker EB, Maino V, Sanchez-Lanier M et al. Murine gamma interferon activates the release of a macrophage-derived Ia-inducing factor that transfers Ia inductive capacity. *J Exp Med* 1984;159:1532–1547.
61. Lautenschlager I, Taskinen E, Inkinen K et al. Distribution of the major histocompatibility complex antigens on different cellular components of human liver. *Cell Immunol* 1984;85:191–200.
62. Ballardini G, Bianchi FB, Doniach D et al. Aberrant expression of HLA-DR antigens on bile duct epithelium in primary biliary cirrhosis: Relevance to pathogenesis. *Lancet* 1984;2:1009–1113.
63. Takacs L, Szende B, Monostori E et al. Expression of HLA-DR antigens on bile duct cells of rejected liver transplant. *Lancet* 1983;2:1500 (letter).
64. McDonald GB, Jewell DP. Class II antigen (HLA-DR) expression by intestinal epithelial cells in inflammatory disease of the colon. *J Clin Pathol* 1984;40:312–317.
65. Sullivan KM, Shulman HM, Storb R et al. Chronic graft-versus-host disease in 52 patients: Adverse natural course and successful treatment with combination immunosuppression. *Blood* 1981;57:267–276.
66. Storb R, Prentice R, Sullivan KM et al. Predictive factors in chronic graft-versus-host disease in patients with aplastic anemia treated by marrow transplantation from HLA-identical siblings. *Ann Intern Med* 1983;98:461–466.
67. McDonald GB, Sullivan KM, Schuffler MD, Shulman HM, Thomas ED. Esophageal abnormalities in chronic graft-versus-host disease in humans. *Gastroenterology* 1981;80:914–921.
68. McDonald GB, Sullivan KM, Plumley TP. Radiographic features of esophageal involvement in chronic graft-versus-host disease. *Am J Roentgenol* 1984;142:501–506.
69. Johnston DE, Koehler RE, Balfe DM. Clinical manifestations of epidermolysis bullosa dystrophica. *Dig Dis Sci* 1981;26:1144–1149.

70. Agha FP, Raji MR. Esophageal involvement in pemphigoid: Clinical and roentgen manifestations. *Gastrointest Radiol* 1982;7:109–112.
71. Shulman HM, Sullivan KM, Weiden PL et al. Chronic graft-versus-host syndrome in man. A long-term clinico-pathologic study of 20 Seattle patients. *Am J Med* 1980;68:204–217.
72. Tsoi MS. Immunological mechanisms of graft-versus-host disease in man. *Transplantation* 1982;33:459–464.
73. Ralph DD, Springmeyer SC, Sullivan KM et al. Rapidly progressive air flow obstruction in marrow transplant recipients: Possible association between obliterative bronchiolitis and chronic graft-versus-host disease. *Am Rev Resp Dis* 1984;129:641–644.
74. Schubert MM, Sullivan KM, Morton TH et al. Oral manifestations of chronic graft-versus-host disease. *Arch Intern Med* 1984;144:1591–1595.
75. Yau JC, Zander AR, Srigley JR et al. Chronic graft-versus-host disease complicated by micronodular cirrhosis and esophageal varices. *Transplantation* 1986;41:129–130.
76. Knapp AB, Crawford JM, Rappeport JM, Gollan JL. Cirrhosis as a consequence of graft-versus-host disease. *Gastroenterology* 1987;92:513–519.
77. McDonald GB, Sharma P, Matthews DE et al. Venocclusive disease of the liver after bone marrow transplantation: Diagnosis, incidence, and predisposing factors. *Hepatology* 1984;4:116–122.
78. Wulff JC, Santner TJ, Storb R et al. Transfusion requirements after HLA-identical marrow transplantation in 82 patients with aplastic anemia. *Vox Sang* 1983;44:366–374.
79. Poulsen H, Christoffersen P. Abnormal bile duct epithelium in chronic aggressive hepatitis and cirrhosis. *Hum Pathol* 1972;3:217–225.
80. Bianchi L, Desmet VJ, Popper H et al. Histologic patterns of liver disease in hemophiliacs, with special reference to morphologic characteristics of non-A, non-B hepatitis. *Semin Liv Dis* 1987;7:203–209.
81. Izutsu KT, Sullivan KM, Schubert MM et al. Disordered salivary immunoglobulin secretion and sodium transport in human chronic graft-versus-host disease. *Transplantation* 1983;35:441–446.

9

New Therapeutic Approaches to Upper Gastrointestinal Motility Disorders

Pierre Poitras

1. INTRODUCTION

Motility of the gastrointestinal tract has long remained an obscure area of gastroenterological science. During the last decades, however, extensive investigations in animals and humans have led to a better understanding of the physiology of gastrointestinal motility and the pathophysiology and therapeutic approaches to related disorders. This chapter will focus on the motility of the stomach and small bowel.

2. PHYSIOLOGY

Motility of the human upper gastrointestinal tract may be divided into two phases: interdigestive (fasting) and digestive (fed). Motor activity during fasting is periodic, with cyclic contractions sweeping the bowel clean of residual debris. This interdigestive motor complex (IDMC) or migratory myoelectric complex is considered the basic organizational structure of this period. Four consecutive phases of electrical and mechanical activity constitute this com-

Pierre Poitras • Department of Medicine, University of Montreal, and Hôpital Saint-Luc, Montreal, Quebec, Canada H2X 3J4.

plex. Phase I is a period of quiescence when no activity occurs. During phase II, irregular and intermittent contractions mix and propulse the intestinal chyme. Phase III, though brief (5–10 min), is the most active and most characteristic component of this complex: a strong peristaltic contraction initiated from the lower esophagus moves down the gastrointestinal tract to the terminal ileum, forcefully propelling the luminal content distally. By the time phase III reaches the ileum, a new one has usually begun in the upper gut. Phase IV, lasting 0–10 min, constitutes a transient decrease of contractile activity until a new phase I marks the beginning of a new complex. This sequential motor pattern will therefore recur cyclically unless interrupted by feeding.

The duration of the IDMC is fairly constant (90–114 min) in the dog (where most physiological studies have been done), whereas a high degree of variability exists in humans (50–200 min). In fact, in Western man, due to frequent eating, the IDMC is often nonexistent during the day.

Perturbations of the IDMC have been described in many pathological situations[1] and most findings still support the notion that this complex (mainly its phase III) constitutes an important physiological defense against stasis of gastrointestinal content during fasting, termed the intestinal "housekeeper" by C. F. Code.

Control of gut motility is exerted at various levels, from the intestinal smooth muscle cell to the brain. This stratification of control includes neural mechanisms.[2] The enteric nervous system, including the submucosal and myenteric plexuses (intrinsic innervation), plays a major role in the organization and coordination of motility during fasting. Vagal and sympathetic (extrinsic) innervation can modulate this motor activity. Initiation of smooth muscle contraction is primarily due to release of acetylcholine, whereas adrenergic substances tend to have an inhibitory effect. Other transmitters (serotonin, purines, peptides, etc.) are probably also involved in neural control. Interdigestive motility is also influenced by hormonal mechanisms.[3] Motilin, a peptide synthesized in endocrine cells of the duodenojejunal region, appears important for the cyclic occurrence of phase III. In the dog, initiation of phase III from the gastroduodenal regional is associated with sharp elevation in plasma motilin and is inhibited following the administration of motilin-specific antagonists.[4] The hormonal role of motilin is not yet well established in humans. However, many authors reported a close association between plasma motilin level and gastric phase III. [5,6] Other peptides can probably influence the motility pattern of the gut but their physiological importance remains to be established.

After a meal, the IDMC disappears for a variable period of time, depending on what is ingested, and is replaced by a constant mixing and propulsing activity. This fed pattern, resembling phase II of the IDMC, propels the gastric content into the duodenum and along the small bowel. Ingested liquids empty first from the stomach. Later, digestible solids, when triturated to particle sizes

of less than 2 mm, pass through the pylorus. As for undigestible solids, size and density influence their rate of gastric emptying.[7] Plastic spheres with diameters larger than 1.6 mm empty more slowly than ^{99m}Tc-labeled liver; spheres with density <1 or >1 empty more slowly than spheres of the same size but with a density of 1. Mojaverian *et al.*[8] observed that a Heidelberg capsule (7×20 mm) was rapidly (30 min) emptied from the stomach of fasting individuals, whereas the frequent intake of food (every 2–3 hr) delayed its gastric emptying for an average of over 14.5 hr. This study supports the hypothesis that phase III of the IDMC is responsible for the propagation of nondigestible material along the gastrointestinal tract. Fed and interdigestive patterns of motility therefore appear both to be essential and to interact in the digestive process. From a therapeutic perspective, the role of the IDMC on the gastric emptying of large, nondigestible solids should be considered when designing drugs targeted for degradation or action in the intestine. This physiological concept also offers an interesting avenue to detect motility disorders. Feldman *et al.*,[9] using solid radiopaque markers to evaluate the gastric emptying of nondigestible solids, found it a more sensitive indicator of gastric motor dysfunction in diabetic patients than emptying studies of digestible solids or liquids using radionuclide scintigraphy. Undigestible alimentary fibers leave the stomach during the postprandial period at a slower rate than liquids.[10] These fibers, although often as long as 1–5 mm, are also thin and pliable; probably for these reasons, they do not behave like undigestible solids that retain their solid character, like the plastic or metal devices discussed above.

Duodenal receptors sensitive to osmotic and fat content will control, perhaps via neurohormonal mechanisms, the duration of the postprandial gastric motor activity. Recently, ileal receptors sensitive to fat have been shown to regulate the gastric emptying rate and the small-bowel transit time.[11] Whether the malabsorbed lipids associated with pancreatic or intestinal diseases could stimulate these ileal receptors, decreasing gastric emptying (''ileal brake''), is speculative but constitutes an attractive explanation for the easy satiety and vomiting sometimes found in these patients. The central nervous system also influences gastrointestinal motility as discussed later.

3. PATHOPHYSIOLOGY

Delayed gastric emptying, though present in several diseases, frequently has an unknown etiology.[12] In many patients with dyspepsia, the diagnosis remains unclear despite extensive investigation (endoscopy, X ray, ultrasound, etc.). A motility disorder may be suspected but often cannot be substantiated. In only 59% of 27 patients referred for functional dyspepsia suggestive of gastric stasis, Jian *et al.*[13] confirmed delayed gastric emptying of radionuclide

markers. Of practical importance, therapy with cisapride (a gastrokinetic agent) was more effective in patients with delayed than in those with normal gastric emptying.

Mechanisms responsible for abnormal emptying have been investigated by various authors using different techniques. Electrogastrograms sometimes reveal abnormal electrical rhythms. Tachyarrhythmia, tachygastria, and bradygastria were recorded in some patients with idiopathic delayed gastric emptying.[14] Recording mechanical activity, in most hands, is more suitable. Malagelada[15] studied gastroduodenal motility in 104 patients referred to the Mayo Clinic for functional dyspepsia. Postprandial manometric abnormalitites of the stomach in the form of antral hypomotility were found in 75 patients. During fasting, 32 exhibited an abnormal motor activity profile. Some had normal IDMCs with superimposed bursts of intense contractile activity; in others, either no recognizable phase III activity was observed or, if present, phase III had an abnormal configuration or propagation. Labo *et el.*[16] analyzed the interdigestive motility of 22 patients with idiopathic delay in gastric emptying rate. No phase III was observed in the stomach of 20 patients. In some, motilin was released abnormally, whereas in others a defect in the response to motilin was suspected since the normal cyclic increase in plasma motilin was present. Conversely, others[17] reported an increased frequency of phase III (often propagated in a retrograde direction) in five patients with idiopathic gastroparesis. In 18 of 22 patients diagnosed as having the irritable bowel syndrome, Kumar and Wingate[18] observed spontaneous or stress-induced irregular motor activity in the duodenum. Therefore, abnormal motility has been observed in many patients suspected of upper gastrointesinal dysmotility. Various dysmotility patterns have been found and can probably now provide the basis for objectively classifying these disorders, which previously could not be identified or separated on clinical grounds. Whether the documented abnormalities represent the cause of the disease or its consequence, whether they are an intrinsic disorder of the gut or manifestations of psychological alterations or stress, are still debatable questions. Moreover, investigations failed to identify any gastrointestinal dysfunction in a significant percentage of cases.

4. THERAPEUTICS

The classical concept that the gastrointestinal tract motility is regulated by a balance between stimulatory (cholinergic) and inhibitory (adrenergic) influences remains generally true. Bethanechol, a parasympathetic agonist, although useful for gastroesophageal reflux, rarely improves patients with a gastrointesinal motility disorder. Dopamine, a precursor of norepinephrine synthesis, in-

hibits gut motility, probably via specific receptor sites. Two dopamine antagonists, metoclopramide and domperidone, enhance the lower esophageal sphincter pressure and the gastric motor activity. Acting mostly on peripheral receptors, and therefore deprived of most unpleasant central effects often encountered with metoclopramide, domperidone appears a suitable therapy to try first in many patients with suspected or proven upper gastrointestinal motor dysfunction. Some recent evidence suggests, however, that the central action of metoclopramide could also provide some therapeutic benefit.[19] Other products, not yet commercially available, bear promise for the future. Hypothesizing that the retrograde or disorganized IDMCs found in some patients with functional dyspepsia could be due to decreased adrenergic control, Sninsky *et al.*[17] administered lidamidine, an α_2-adrenergic agonist, to their patients and resolved both the motor alterations and the clinical symptoms. Conversely, cisapride facilitates acetylcholine release by the myenteric plexus and enhances lower esophageal sphincter pressure, gastric emptying of solids and liquids, and their intestinal propagation. In preliminary studies, cisapride was useful for treatment of esophagitis,[20] corrected the small-bowel transit of chyme in chronic intestinal pseudo-obstruction,[21] and increased the defecation rate in patients with atonic constipation,[22] suggesting that it also had a stimulatory effect on the distal intestine.

Peptides may also play a role in the control of gastrointestinal motility. However, their physiological importance is rarely assessed, especially during the postprandial period.[3] Motilin is physiologically active during the interdigestive phase in the dog; enkephalins appear involved in the ileal brake[23] and in the effect of stress on duodenal motor activity.[24] Further development in basic knowledge will probably guide future therapeutic orientation with peptides. Few possibilities have been practically explored up to now, due partly to the difficulty in obtaining active and convenient pharmaceutical preparations of peptides. Presently, therapeutic assistance is provided mainly through opiate derivatives. Injections of morphine[25] or trimebutine[26] induce phase III and motilin release; dextromethorphan, a codeine analog, was superior to placebo for treatment of irritable bowel syndrome.[27] More studies are needed to define possible new therapeutic roles for these substances classically considered for treatment of diarrhea. Naloxone, an opiate antagonist, improved some patients with idiopathic pseudo-obstruction[28] or chronic constipation.[29] Vasopressin, administered by nasal spray, may be useful in some patients with postvagotomy gastric atonia and with idiopathic constipation (J. C. Schang, personal communication).

Electromechanical devices, comparable to cardiac pacemakers, were developed by the Mayo Clinic group to influence the intestinal electrical activity controlling muscle contraction. Such devices were useful in animals to decrease

the gastric emptying rate in dumping syndrome,[30] and to slow flow and enhance absorption of the small intestine.[31] They could also improve the gastric emptying delay observed after Roux-en-Y procedure.[32]

The last question concerns the importance of stress and psychological alterations in motility disorder. "Stress" has been shown to influence gut motility. Central stimulation by sham feeding in dogs inhibits phase III[33]; likewise labyrinthine stimulation after a meal inhibits the usual postprandial motility response in humans.[24] Painful stress (e.g., immersion of hands in cold water) delays gastric emptying of food.[34] Stress, induced experimentally, among other things, by arcade video games, significantly reduces the number of phase IIIs.[19]

The causal relation of stress to motility disorder or even its presence in symptomatic patients is very difficult to establish. Talley *et al.*[35] reported that patients with dyspepsia of unknown origin did not experience an excess of major life stress during the one year preceding diagosis when compared with a control population in the community. On the other hand, it is often postulated that some patients, in particular those with an anxious or neurotic disposition, may be more likely to show stress-related abnormalities. Talley *et al.*[36] observed that their patients with essential dyspepsia were more neurotic, anxious, and depressed than community controls. However, in Malagelada's study,[15] the only one with objective measures of gut dysfunction, just 23 and 3% of their patients with documented gastric or gastroduodenal malfunction, respectively, had an associated psychiatric diagnosis established, compared to a psychiatric problem in 63% of those with normal motility pattern.

Treatment by various forms of psychotherapy can benefit some individuals. Behavioral therapy caused one patient with idiopathic gastroparesis to resume normal gastric emptying function.[37] Therapeutic success has also been obtained in patients with suspected motility disorders of the lower gastrointestinal tract. In a controlled study, dynamically oriented individual psychotherapy significantly improved the symptoms of patients with irritable bowel syndrome.[38] In some patients suffering from chronic constipation of unknown origin, Devroede[39] observed that transit time of markers and the interval between defecations improved during analytical psychotherapy.

5. CONCLUSION

Many patients seen by the gastroenterologist are suspected of suffering from gastrointestinal motility disorders. In some, motility disorders can be identified but frequently this cannot be predicted on clinical grounds. Furthermore, it is probable that all patients will not suffer from a similar disease. The etiology of any observed dysmotility almost always remains unknown.

Potent pharmacological agents are constantly being developed, electrome-

chanical devices appear to be an interesting issue, while psychotherapy could be beneficial for some of these conditions. Additional knowledge concerning the physiology of gastrointestinal motility and the physiopathological basis for its disorders is essential to develop specific treatments for specific diseases.

REFERENCES

1. Collins SM, Valori RM. Disruption of small intestinal migrating myoelectric patterns in human disease states. In: *Small Intestinal and Colonic Motility* (Poitras P, ed). Proceedings of the first international symposium, Montreal, Jouveinal Lab., 1985, pp. 35–50.
2. Diamant NE. Neurological control of the IDMC. In: *Small Intestinal and Colonic Motility* (Poitras P, ed). Proceedings of the first international symposium, Montreal, Jouveinal Lab., 1985, pp. 3–13.
3. Poitras P. Hormonal control of the IDMC. In: *Small Intestinal and Colonic Motility* (Poitras P, ed). Proceedings of the first international symposium, Montreal, Jouveinal Lab., 1985, pp. 15–23.
4. Poitras P. Motilin is a digestive hormone in the dog. *Gastroenterology* 1984;87:909–913.
5. You CH, Chey WY, Lee KY. Studies on plasma motilin concentration and interdigestive motility of the duodenum in humans. *Gastroenterology* 1980;79:62–66.
6. Janssens J, Van Trappen G, Peeters TL. The activity front of migrating motor complex of the human stomach but not of the small intestine is motilin-dependent. *Regul Peptides* 1983;6:363–369.
7. Meyer JH, Dressman J, Fink A, Amidon G. Effect of size and density on canine gastric emptying of nondigestible solids. *Gastroenterology* 1985;89:805–813.
8. Mojaverian P, Ferguson RK, Vlasses PH et al. Estimation of gastric residence terms of the Heidelberg capsule in humans: Effect of varying food composition. *Gastroenterology* 1985;89:392–397.
9. Feldman M, Smith HJ, Simon TR. Gastric emptying of solid radiopaque markers: studies in healthy subjects and in diabetic patients. *Gastroenterology* 1984;87:895–902.
10. Malagelada JR, Robertson JS, Brown MC et al. Intestinal transit of solid and liquid components of a meal in health. *Gastroenterology* 1984;87:1255–1263.
11. Read NW, McFarlane A, Kinsman RI et al. Effect of infusion of nutrients solutions into the ileum on gastrointestinal transit and plasma levels of neurotensin and enteroglucagon. *Gastroenterology* 1984;86:274–280.
12. Minami H, McCallum RW. The physiology and pathophysiology of gastric emptying in humans. *Gastroenterology* 1984;86:1592–1610.
13. Jian R, Ruskone A, Ducrot F, Chaussade S, Rambaud JC, Bernier JJ. Radionuclide and therapeutic assessment of chronic idiopathic dyspepsia. *Gastroenterology* 1986;90:1477.
14. You CH, Lee KY, Chey WY, Menguy R. Electrogastrophic study of patients with unexplained nausea, bloating and vomiting. *Gastroenterology* 1980;79:311–314.
15. Malagelada JR, Stanghellini U. Manometric evaluation of functional upper gut symptoms. *Gastroenterology* 1985;88:1223–1231.
16. Labo G, Bortolotti M, Bezzadini P, Bonora G, Bersani G. Interdigestive gastroduodenal motility and serum motilin levels in patients with idiopathic delay in gastric emptying. *Gastroenterology* 1986;90:20–26.
17. Sninsky CA, Martin JC, Mathias JR. Effect of lidamidine hydrochloride, a proposed α_2-adrenergic agonist, in patients with gastroduodenal motor dysfunction. *Gastroenterology* 1983;84:1315.

18. Kumar D, Wingate DL. The irritable bowel syndrome: a paroxysmal motor disorder. *Lancet* 1985;2:973–977.
19. Valori RM, Kumar D, Wingate DL. Effect of different types of stress and of prokinetic drugs on the control of the fasting motor complex in humans. *Gastroenterology* 1986;90:1890–1900.
20. Janisch HD. A double-blind multi-center trial to compare the efficacy of cisapride and ranitidine in the acute treatment of gastroesophageal reflux disease. *Gastroenterology* 1986;90:1475.
21. Camilleri M, Brown ML, Zinsmeister AR, Malagelada JR. Cisapride corrects the impaired small bowel transit time of chyme in chronic intestinal pseudoobstruction. *Gastroenterology* 1985;88:1340.
22. Gomme L. Effect of R51619 on chronic constipation. In: *Cisapride Investigators Brochure* 1983; 20.
23. Kinsman RI, Read NW. Effect of naloxone on feedback regulation of small bowel transit by fat. *Gastroenterology* 1984;87:335–337.
24. Stanghellini V, Malagelada JR, Zinsmeister AR, Go VLW, Kao PC. Stress-induced gastroduodenal motor disturbances in humans: possible humoral mechanisms. *Gastroenterology* 1983;85:83–91.
25. Sarna S, Chey WY, Condon RE, Dodds WJ, Myers T, Chang TM. Cause and effect relationship between motilin and migrating myoelectric complexes. *Am J Physiol* 1983;245:G277–G289.
26. Poitras P, Honde C, Havrankova J et al. Effect of trimebutine on intestinal motility and plasma motilin in the dog. *Am J Physiol* 1988 (in press).
27. Meshkinpour H, Roberts GS, Sandler R, Snape W, Oren J, Baky SH. Dextromethorphan in treatment of irritable bowel syndrome. *Gastroenterology* 1985;88:1501.
28. Schang JC, Devroede G. Beneficial effects of naloxone in a patient with intestinal pseudoobstruction. *Am J Gastroenterol* 1985;80:407–411.
29. Kreek MJ, Hanh EF, Schaeffer RA, Fishman J. Naloxone, a specific opioid antagonist, reverses chronic idiopathic constipation. *Lancet* 1983; 261–262.
30. Becker JM, Sava P, Kelly KA, Shturman L. Intestinal pacing for canine postgastrectomy dumping. *Gastroenterology* 1983;84:383–387.
31. Collin J, Kelly KA, Phillips SF. Increased canine jejunal absorption of water glucose and sodium with intestinal pacing. *Am J Dig Dis* 1978;23:1121–1123.
32. Morrison P, Kelly K, Hocking M. Electrical dysrrhythmia in the Roux-en-Y jejunal limb and their correction by pacing. *Gastroenterology* 1985;88:1508.
33. Lemoyne M, Wassef R, Tasse D, Trudel L, Poitras P. Motilin and the vagus in dogs. *Can J Physiol Pharmacol* 1984;62:1092–1096.
34. Thompson DG, Richelson E, Malagelada JR. Perturbation of upper gastrointestinal function by cold stress. *Gut* 1983;24:277–283.
35. Talley NJ, Piper DW. Major life event stress and dyspepsia of unknown cause. *Gut* 1986;27:127–134.
36. Talley NJ, Fung LH, Gilligan IJ, McNeil D, Piper DW. Association of anxiety, neuroticism and depression with dyspepsia of unknown cause. *Gastroenterology* 1986;0:886–892.
37. Latimer, PR, Malmud LS, Fisher RS. Gastric stases and vomiting: behavioral treatment. *Gastroenterology* 1982;83:684–688.
38. Svedlund J, Ottoson JO, Sjodin I, Dotevall G. Controlled study of psychotherapy in irritable bowel syndrome. *Lancet* 1938;10:589–592.
39. Devroede G. Irritable bowel syndrome: Clinical and therapeutical aspects. In: *Small Intestinal and Colonic Motility* (Poitras P, ed). Proceedings of the First International Symposium, Montreal, Jouveinal Lab, 1985, pp. 129–139.

10

Is There a Place for Proximal Gastric Vagotomy?

Bernard J. Perey

1. INTRODUCTION

Selective denervation of the acid-secreting portion of the stomach was introduced in 1967 by Holle and Hart of Munich,[1] who combined the procedure with pyloroplasty. The operation was later popularized without the pyloroplasty by Johnston and Wilkinson in Leeds[2,3] and Amdrup and Jensen in Copenhagen.[4] Quite rapidly, this type of vagotomy gained favor throughout Europe and became widely used in several countries for the elective surgical treatment of duodenal ulcer.[5–14] In contrast, adoption of this operation has been slow in North America, particularly in the United States where caution has been expressed against the wide use of this vagotomy because of the fear that many surgeons may not perform the procedure adequately or that the incidence of recurrent ulceration may eventually turn out to be prohibitively high. Thus, it is not surprising that we should address once again the question of whether or not there is a place for proximal gastric vagotomy.

2. TECHNICAL CONSIDERATIONS

Much confusion has been created by the wide array of terms used by various authors to describe the same operation. Such terms include highly se-

Bernard J. Perey • Department of Surgery, Dalhousie University, Halifax, Nova Scotia, Canada B3H 2Y9.

lective vagotomy, hyperselective vagotomy, selective proximal vagotomy, supraselective vagotomy, ultraselective vagotomy, parietal cell vagotomy, fundic vagotomy, and proximal gastric vagotomy (PGV). Because of its explicit nature and relative popularity, only the last of these terms will be used in the present text. The operation should not be confused with selective vagotomy, which consists of transecting the two vagus nerves beyond their hepatic and celiac branches and has never gained much popularity.

Proximal gastric vagotomy is achieved by severing all branches of the vagus nerves to the entire abdominal portion of the esophagus and to the proximal stomach, down to the incisura at a point on the lesser curvature located about 7 cm from the pylorus. In the course of this denervation, all blood vessels and autonomic fibers traveling with the transected vagal branches are also cut, resulting in a complete devascularization of the vertical portion of the lesser curvature. Because of the rich vascular anastomotic network in the wall of the stomach, blood coming from the greater curvature and the antrum provides an ample supply to the devascularized stomach; fortunately, necrosis of the lesser curvature is a rare event.[6] To minimize the risk of recurrent ulceration, it is essential to identify and divide all vagal branches to the entire (5–7 cm) abdominal esophagus[15–18] and to the posterior wall of the upper fundus.[19] Because of the importance of these technical details in minimizing the risk of ulcer recurrence and of the tedious nature of the procedure, surgeons who perform this operation should be particularly meticulous.

To ensure completeness of the vagotomy, various intraoperative tests have been developed.[18–21] Such techniques compound the duration and complexity of the operation and are not used by most surgeons. A well-trained vagotomist should obtain good results without intraoperative testing.

Since innervation of the distal stomach is preserved, no pyloroplasty or gastroenterostomy is required, the gastrointestinal tract is not entered, and the risk of postoperative infection is very low.

Although a number of authors have used a drainage procedure in combination with PGV, this approach is no longer recommended since it provides no known benefit and increases postoperative morbidity.[14]

3. PHYSIOLOGICAL EFFECTS OF PROXIMAL GASTRIC VAGOTOMY

3.1. Gastric Acid Secretion

The reduction of acid secretion following PGV is similar to that observed after truncal vagotomy. The extent of the reduction varies from case to case. It is unrelated to the preoperative levels of acid secretion, but a modest reduction

usually indicates that the vagotomy was incomplete and forebodes a higher risk of ulcer recurrence.[11,22–26]

Comparing the effects of cimetidine and PGV on 24-hr acid secretion in duodenal ulcer patients, Feldman and Richardson[27] found that 400 mg of cimetidine twice a day brought secretion to the level of normal controls, whereas PGV reduced acid output to less than half of normal. This is consistent with the clinical observation that PGV can be expected to heal duodenal ulcers that are resistant to medical treatment. PGV reduces basal acid output by about 70–80%.[22,28–30] The reduction of maximal acid output (MAO) in response to pentagastrin is usually less pronounced, stabilizing between 50 and 70%.[22,28–30] The reduction of gastrin-stimulated MAO falls within the same range.[34]

Acid secretion in response to insulin is also reduced. A positive Hollander test was observed by Johnston in 3% of patients 1 week after PGV[35] while it was 16% after truncal vagotomy.[36] During the first postoperative year there is a great increase in the number of patients who have a positive Hollander test: 50–60% 1–2 years after PGV.[35,37] Even so, the positive response to insulin is slight, averaging only 3.5 mM/hr compared with 36 mM/hr preoperatively. A high percentage of positive Hollander tests in the immediate postoperative period correlates with a high recurrence rate and reflects errors in operative technique.[38]

A slight recovery of acid secretion in the first two postoperative years has been observed by some authors,[32,33,35,39] but not by all.[40] Surgical technique is critical in preventing significant hypersecretion and ulcer recurrence. Reports on patients operated prior to 1976 should be interpreted with caution since major technical improvements were introduced at about that time.[16] Debas proposed that recovery of secretion could occur as a result of the loss of a fundic inhibitory substance after PGV.[41] Others suggested gradual reinnervation of the parietal cell mass[42] or G-cell hyperplasis.[43]

Although cholinergic agents can restore preoperative MAO following truncal vagotomy,[44,45] such agents have little or no effect on MAO following PGV.[46,47] This suggests that a vagus-dependent antral mechanism contributes to the control of acid secretion.

Parietal cells have been found to be morphologically normal up to 3 years after PGV, although they showed ultrastructural evidence of diminished secretory capacity.[48]

3.2. Serum Gastrin

Serum gastrin concentration is elevated following PGV between meals, with meals, and in response to insulin, but these findings are not significantly different from what is observed with truncal vagotomy.[49,52] While postoperative serum concentrations are two to three times higher than in nonoperated duodenal ulcer patients, this only results in half of the acid output.[34]

3.3. Gastric Motility

There is some loss of gastric receptive relaxation after PGV.[53,54] This could explain the early satiety that is commonly experienced by patients in the first 2 or 3 postoperative weeks.[55]

PGV tends to enhance emptying of liquids,[56] which may explain the occasional mild dumping seen following this operation. However, emptying and dispersion of solid foods remain normal.[53,56–59] These findings are in contrast with the more severely disturbed gastric motility resulting from other surgical procedures used to treat duodenal ulcer.

3.4. Esophageal Motility

Despite the extensive dissection of the distal esophagus carried out as part of PGV, there is no resulting alteration in the resting pressure, function, or length of the lower esophageal sphincter.[60–62] This is consistent with the observation that proven reflux esophagitis rarely develops following PGV. Transient dilation of the body of the esophagus is occasionally seen, but the cause of this is unknown. Postoperative dysphagia is very common but disappears spontaneously within 2–3 months.[8,13,31,62] Its exact mechanism is not understood.

3.5. Biliary Tree

Truncal vagotomy with drainage and gastrectomy have been reported to be followed by an increased incidence of cholelithiasis.[63–65] Evidence has not yet been found that PGV has a similar effect. It is known that emptying of the human gallbladder in response to a meal is partly regulated by cholinergic stimulation and that vagal innervation of the gallbladder is an essential pathway for such stimulation.[66,67] Such innervation should be preserved with PGV and lost with truncal vagotomy. Gallbladder size and function are normal after PGV but are grossly abnormal after truncal vagotomy.[66–70] These findings suggest that PGV will not be followed by gallbladder disease to the same extent as truncal vagotomy.

3.6. Pancreas

Truncal vagotomy disturbs both the endocrine and exocrine functions of the pancreas as well as the release of secretin in response to acidification of the duodenum, but PGV produces little or no such effect.[71–77]

3.7. Fecal Fat Excretion

Fecal excretion of fat is significantly increased following truncal vagotomy but is near-normal after PGV.[78]

4. INDICATIONS

The great majority of proximal gastric vagotomies are performed for the control of chronic duodenal ulcers that have failed to respond adequately to medical treatment in the absence of complicating factors such as perforation, bleeding, stenosis, coexistent gastric ulceration, or high operative risk.[40,79,80] Although preoperative testing of gastric secretion is still being suggested to identify those patients best suited for PGV,[81] there is no relationship between preoperative levels of secretion and the postoperative reduction of acid secretion or the risk of ulcer recurrence.[11,22–26] Consequently, preoperative testing has not yet been proven to be of any practical help in selecting the best operation.

The use of PGV in emergency treatment of perforated duodenal ulcers has given excellent results.[13,30,82–86] In view of the high risk of ulcer recurrence requiring definitive surgery following simple closure of a perforated duodenal ulcer and the very low morbidity of PGV, the use of this acid-reducing procedure as an adjunct to the emergency closure of the perforation may be of considerable value in selected situations.

Experience has been less extensive with PGV in the acute management of bleeding duodenal ulcers.[13,87,88] Control of bleeding is carried out through a duodenotomy and care must be taken to spare the pylorus in order to minimize the risks of postoperative sequelae. There is little information on the long-term results of PGV for this indication.

Several authors have reported on the use of PGV combined with dilation or duodenoplasty in the presence of duodenal or pyloric stenosis.[13,84,89–91] Careful analysis of such reports should recognize that when the stenosis is at the pylorus or immediately adjacent to it, as is frequently the case, a plasty of the narrowed area would in effect result in a pyloroplasty, whereas dilation carries some risk of perforation and a high risk of failure from recurrence of the stenosis or the ulcer. One reason for the high ulcer recurrence following dilation may be that the original ulcer is often pyloric or prepyloric, a variant that does not respond well to PGV, as discussed below. Consequently, there is diminishing support for the use of PGV in the presence of stenosis. On the other hand, when the stenosis is well beyond an unobstructed pylorus, there is general agreement that duodenoplasty combined with PGV is an acceptable approach.

The results of PGV are unsatisfactory when used for pyloric and prepyloric ulcers, even in the absence of stenosis.[5,23–25,30,88] The explanation for the high risk of recurrence in this situation is not known. It is thus recommended that another type of procedure be selected in such cases.

Benign gastric ulcers of the body of the stomach without associated duodenal ulcer disease (gastric ulcer type I) may be treated by PGV and ulcer excision with results that are at least as good as with distal gastrectomy and

gastroduodenostomy.[92–94] However, the reported series are small and it is too early to make a definitive judgment regarding this particular indication. Conversely, there is much support for not performing PGV when duodenal ulcer disease coexists with a gastric ulcer.[31,87]

Proximal gastric vagotomy in combination with fundoplication has been advocated in the treatment of reflux esophagitis.[95,96] This may permit a better antireflux procedure, but its main value would be in the management of reflux combined with duodenal ulcer disease.

5. RESULTS

5.1. Operative Mortality

Proximal gastric vagotomy is clearly the safest operation available for the treatment of duodenal ulcer. In 1975, Johnston reported an operative mortality of 0.3% in more than 5000 cases.[6] Since then, the mortality has decreased further, in part because of the increasing rarity of postoperative necrosis of the lesser curvature of the stomach. In a review of 24 major series reported from around the world between 1975 and 1979, Hershlag and Argov[8] found only one operative death out of 4321 operations. Similar statistics are reported by other extensive reviews of the literature.[79,80] The safety of the operation is partly related to the fact that the gastrointestinal tract is not entered, but the impressive statistics are also explained by the selection of patients without complicating factors.

5.2. Morbidity

Vagotomists universally report that patients recover quite rapidly following PGV. Oral intake is usually resumed after 2 days and most patients leave the hospital on or before the sixth postoperative day. Wound complications are rare. Patients commonly report an early postprandial satiety, but this is less pronounced than following truncal vagotomy and disappears within 2 to 3 weeks.[8,37,55] Dysphagia is also common, but is rarely severe and disappears spontaneously within 2–3 months.[8,13,31,62]

Sequelae are much rarer than following truncal vagotomy or gastric resection. About 85% of patients have excellent or good results (Visick I or II) when operated for chronic duodenal ulcer without stenosis.[8,9,13,26,30,37,55,97] Dumping and diarrhea are particularly rare and always mild when they develop. The most attractive feature of PGV is the absence of early and late complications.

5.3. Recurrence of Ulceration

The major drawback of PGV is the risk of ulcer recurrence. It accounts for the majority of poor results and is the main reason why the operation has not been more widely adopted in North America. Reported recurrence rates have varied from 0 to 30%, but the reasons for such discrepancies are now reasonably well understood. There are a number of factors influencing the risk of recurrence.

The broad consensus among vagotomists is that surgical technique is the single most important consideration in preventing recurrence. It is also the simple explanation for the very high recurrence rates reported by some. Several years after the introduction of PGV, reports began to appear pointing out the need to extensively dissect the entire abdominal portion of the esophagus,[15–17] and to divide all vagal branches to the posterior wall of the upper fundus.[19] The relationship of these technical steps to the incidence of ulcer recurrence is now well established and series including operations performed in the early 1970s tend to have high recurrence rates. For similar reasons, operations performed by inadequately trained vagotomists will tend to lead to poor results.[5,10,25,98,99] Errors in technique not only lead to poor early results but also increase the risks of delayed recurrence. In an effort to ensure completeness of vagotomy, a number of intraoperative tests have been devised and some experts feel that such tests should at least be used during the training period of each vagotomist. Others feel confident that good results can be obtained without such tests. A reasonable compromise is to monitor acid secretion before and after surgery so that the surgeon can obtain an indication of the adequacy of his technique from the degree of reduction of the maximal acid output after PGV.[5,11,23,98] Surgeons tend to follow a learning curve with this operation,[11,13] a fact that should be kept in mind by both surgeon and referring physician. Since the number of duodenal ulcers requiring surgery has been steadily decreasing, it is probably desirable that fewer surgeons perform the majority of these procedures.

The duration of follow-up after PGV will also affect reported recurrence rates since recurrences have been observed several years after PGV.[11,23,87] However, the great majority of recurrences develop in the first two postoperative years and very few after 3 years. Considering the very large number of patients that have now been followed for 8 or 10 years, it is probably quite safe to assume that results at 10 years will be essentially the same as the results at 5 years.

The type of ulcer may also determine the risk of recurrence as previously mentioned; pyloric ulcers, prepyloric ulcers, and ulcers with pyloric stenosis are particularly prone to recur following PGV. Many have now opted for another surgical procedure in such cases.

Other factors have been studied such as age,[11,98] sex,[11,12,26,55,97,98] duration of symptoms,[11,22,98] history of previous complications,[11] and levels of preoperative acidity,[11,23,26,81] but none of these parameters were found to have any relationship to recurrence.

In summary, when PGV is performed by experienced vagotomists for the treatment of chronic duodenal ulcers refractory to medical management, in the absence of stenosis the long-term risk of recurrence should not be substantially different from that obtained with truncal vagotomy and should be about 10%.[5,11,13,30,37] But even if the recurrence rate turns out to be higher, particularly in the hands of less experienced surgeons, PGV still presents distinct advantages over other procedures because recurrences usually respond very well to nonoperative management.[25,30,31,88,100] A very small group of patients undergoing PGV will eventually require further surgery for recurrence and can be cured by antrectomy.[101] These will then be the only patients in the initial group to be exposed to the risk and morbidity of gastric resection.

6. CONCLUSIONS

When comparing the respective merits of PGV, truncal vagotomy and drainage, and vagotomy with antrectomy in the treatment of uncomplicated chronic duodenal ulcer, all authors agree that PGV is remarkable for its safety, freedom from side effects, and the relative ease with which recurrences can be treated. There is ample evidence that when well-trained vagotomists use PGV for this particular indication, the risk of long-term ulcer recurrence should be about 10% and certainly less than 15%. Similar results can also be obtained in well-selected patients with a complicated duodenal ulcer and even in type I gastric ulcers.

While truncal vagotomy with pyloroplasty may result in a slightly diminished ulcer recurrence rate, it leads to disabling side effects for which there is little or no effective treatment in more than 20% of the cases.[102]

Vagotomy with gastric resection is successful in achieving a permanent cure of the ulcer in practically every case, but this is accomplished at the cost of an increased operative risk and of the sequelae of truncal vagotomy.[31–103] The comparison between PGV and gastric resection has been reduced by many experienced surgeons to choosing between 10 easily manageable recurrences and one death.[5] Others have concluded that PGV would still be the preferable operation if the recurrence rate were even as high as 20%.[100]

Although there is little doubt that the average gastrointestinal surgeon can learn to perform PGV adequately, it is less well established that all abdominal surgeons will have the opportunity to master the fine technical points of the procedure. For this reason, one must expect that the application of PGV will never reach its full potential.

Proximal gastric vagotomy may undergo further refinements in its technique and indications. It may eventually be rendered obsolete by a new form of medical treatment. But it is fair to conclude that for the foreseeable future PGV is here to stay.[40,55,79,80,87]

REFERENCES

1. Holle F, Hart W. Neue wege der chirurgie des gastroduodenal ulkus. *Med Klin* 1967;62:441–450.
2. Johnston D, Wilkinson AR. Selective vagotomy with innervated antrum without drainage procedure for duodenal ulcer. *Br J Surg* 1969;56:626.
3. Johnston D, Wilkinson AR. Highly selective vagotomy without a drainage procedure in the treatment of duodenal ulcer. *Br J Surg* 1970; 57:289–295.
4. Amdrup E, Jensen HE. Selective vagotomy of the parietal cell mass preserving innervation of the undrained antrum. *Gastroenterology* 1970;59:522–527.
5. Harmon JW, Jordan PH, Jr. Verdict on vagotomy. *Gastroenterology* 1981;81:809–810.
6. Johnston D. Operative mortality and postoperative morbidity of highly selective vagotomy. *Br Med J* 1975;4:545–547.
7. Hautefeuille P, Picaud R. Les recidives apres vagotomies pour ulcere duodenal. New York: Masson, 1983.
8. Hershlag A, Argov S. Parietal cell vagotomy. II. The first decade. Clinical considerations. *Curr Surg* 1983;40:93–104.
9. Kennedy T. Long-term results of proximal gastric vagotomy. *Can J Surg* 1984;27:340–341.
10. Hoffmann J, Jensen HE, Schulze S, Poulsen PE, Christiansen J. Prospective controlled vagotomy trial for duodenal ulcer: Results after five years. *Br J Surg* 1984;71:582–585.
11. Adami H-O, Enander L-K, Enskog L, Ingvar C, Rydberg B. Recurrences 1 to 10 years after highly selective vagotomy in prepyloric and duodenal ulcer disease. *Ann Surg* 1984;199:393–399.
12. Emas S, Ferstrom M. Prospective randomized trial of selective vagotomy with pyloroplasty and selective proximal vagotomy with and without pyloroplasty in the treatment of duodenal, pyloric and prepyloric ulcers. *Am J Surg* 1985;149:236–243.
13. Gorey TF, Lennon F, Heffernan SJ. Highly selective vagotomy in duodenal ulceration and its complications. A 12-year review. *Ann Surg* 1984;200:181–84.
14. Aeberhard P, Walther M. Results of a controlled randomized trial of proximal gastric vagotomy with and without pyloroplasty. *Br J Surg* 1978;65:634–636.
15. Hedenstedt S, Lundquist G, Moberg S. Selective proximal vagotomy in the treatment of duodenal ulcer. *Acta Chir Scand* 1972;138:591–596.
16. Hallenbeck GA, Gleysteen JJ, Aldrete JS, Slaughter RL. Proximal gastric vagotomy: Effects of two operative techniques on clinical and gastric secretory results. *Ann Surg* 1976;184:434–442.
17. Kronborg O, Jorgensen PM, Holst-Christensen J. Influence of different techniques of proximal gastric vagotomy upon risk of recurrent duodenal ulcer and gastric acid secretion. *Acta Chir Scand* 1977;143:53–56.
18. Donahue PE, Bombeck CT, Yoshida J, Nyhus LM. The simplified endoscopic congo red test for completeness of vagotomy. *Surg Gynecol Obstet* 1986;163:287–289.
19. Grassi G. Anatomy of the "criminal branch" of the vagus and its surgical implications. In: *Surgery of the Stomach and Duodenum* (Nyhus LM, Wastell C, eds). Boston: Little, Brown, 1977, pp. 60–65.

20. Grassi G, Orecchia C. A comparison of intraoperative tests for completeness of vagal section. *Surgery* 1974;75:155–160.
21. Burge H, Vane JR. Method of testing for complete nerve section during vagotomy. *Br Med J* 1958;1:615–618.
22. Christiansen J, Jensen HE, Ejby-Poulsen P, Bardram L, Henriksen FW. Prospective controlled vagotomy trial for duodenal ulcer. Primary results, sequelae, acid secretion, and recurrence rates two to five years after operation. *Ann Surg* 1981;1983:49–55.
23. Andersen D, Amdrup E, Hostrup H, Sorensen FH. The Aarhus county vagotomy trial: Trends in the problem of recurrent ulcer after parietal cell vagotomy and selective gastric vagotomy with drainage. *World J Surg* 1982;6:86–92.
24. Amdrup E. Recurrent ulcer. *Br J Surg* 1981;68:679–681.
25. Graffner HO, Liedberg GF, Oscarson JEA. Recurrence after parietal cell vagotomy for peptic ulcer disease. *Am J Surg* 1985;150:336–340.
26. Koo J, Lam SK, Chan P et al. Proximal gastric vagotomy, truncal vagotomy with drainage, and truncal vagotomy with antrectomy for chronic duodenal ulcer. *Ann Surg* 1983;197:265–271.
27. Feldman M, Richardson CT. Total 24-hour gastric acid secretion in patients with duodenal ulcer. Comparison with normal subjects and effects of cimetidine and parietal cell vagotomy. *Gastroenterology* 1986;90:540–544.
28. Greenall MJ, Lyndon PJ, Goligher JC, Johnston D. Long-term effect of highly selective vagotomy on basal and maximal acid output in man. *Gastroenterology* 1975;68:1421–1425.
29. Saik RP, Greenburg AG, Peskin GW. Pros and cons of parietal cell versus truncal vagotomy. *Am J Surg* 1984;148:32–38.
30. Hollinshead JW, Smith RC, Gillett DJ. Parietal cell vagotomy: experience with 114 patients with prepyloric or duodenal ulcer. *World J Surg* 1982;6:596–602.
31. Donahue PE, Bombeck CT, Condon RE, Nyhus LM. Proximal gastric vagotomy versus selective vagotomy with antrectomy: Results of a prospective, randomized trial after four to twelve years. *Surgery* 1984;96:585–591.
32. Jepson K, Lari J, Humphrey CS, Smith RB, Wilkinson AR, Johnston D. A comparison of the effects of truncal, selective and highly selective vagotomy on maximal acid output in response to pentagastrin. *Ann Surg* 1973;178:769–772.
33. Johnston D, Wilkinson AR, Humphrey CS et al. Serial studies of gastric secretion in patients after highly selected (parietal cell) vagotomy without a drainage procedure for duodenal ulcer I. Effect of highly selective vagotomy on basal and pentagastrin-stimulated maximal acid output. *Gastroenterology* 1973;64:1–11.
34. Blair AJ, Richardson CT, Walsh JH, Chew P, Feldman M. Effect of parietal cell vagotomy on acid secretory responsiveness to circulating gastrin in humans. Relationship to postprandial serum gastrin concentration. *Gastroenterology* 1986;90:1001–1007.
35. Johnston D, Wilkinson AR, Humphrey CS et al. Serial studies of gastric secretion in patients after highly selective (parietal cell) vagotomy without a drainage procedure for duodenal ulcer. II. The insulin test after highly selective vagotomy. *Gastroenterology* 1973;64:1–11.
36. Johnston D, Goligher JC. The influence of the individual surgeon and of the type of vagotomy upon the insulin test after vagotomy. *Gut* 1971;12:963–967.
37. Liavag I, Roland MA. Seven-year follow-up of the proximal gastric vagotomy. Clinical results. *Scand J Gastroenterol* 1979;14:49–56.
38. Kronborg O, Madsen P. A controlled, randomized trial of highly selective vagotomy versus selective vagotomy and pyloroplasty in the treatment of duodenal ulcer. *Gut* 1975;16:268–271.
39. Van Heerden JA, Kelly KA, Dozoi RR et al. Proximal gastric vagotomy. Initial experience. *Mayo Clin Proc* 1980;55:10–13.

40. Donahue PE, Tsai H-S, Yoshida J, Nyhus LM. Proximal gastric vagotomy: The first 25 years. In: *Advances in Surgery,* Vol 19 (Mannick JA et al, eds). Chicago: Year Book, 1986 pp. 139–173.
41. Debas HT. Proximal gastric vagotomy interferes with a fundic inhibitory mechanism. A hypothesis for the high recurrence rate of peptic ulceration. *Am J Surg* 1983;146:51–56.
42. Joffe SN, Crocket A, Doyle D. Morphologic and functional evidence of reinnervation of the gastric parietal cell mass after parietal cell vagotomy. *Am J Surg* 1982;143:80–85.
43. Dunn DH, Decanini C, Bonsack ME, Eisenberg MM, Delaney JP. Gastrin cell populations after highly selective vagotomy in the dog. *Am J Surg* 1979;137:111–115.
44. Payne RA, Kay AW. The effect of vagotomy on the maximal acid secretory response to histamine in man. *Clin Sci* 1962;22:373–382.
45. Broome A. Mechanism of the vagotomy-induced suppression of the maximal acid response to histamine in antrectomized ulcer patients. *Scand J Gastroenterol* 1967;2:275–282.
46. Rowland M, Berstad A, Liavag I. Effect of urecholine and carbacholine on pentagastrin-stimulated gastric secretion after proximal gastric vagotomy in duodenal ulcer patients. *Scand J Gastroenterol* 1972;10:315–319.
47. Feldman M, Walsh JH. Effect of bethanechol on gastric acid secretion and serum gastrin concentration after proximal gastric vagotomy. *Ann Surg* 1982;196:14–17.
48. Romeo G, Sanfilippo G, Basile R, Catania G, Ianello A, Carnazza ML. Ultrastructural study of parietal cells before and after parietal cell vagotomy in patients with duodenal ulcer. *Surg Gynecol Obstet* 1981;153:61–64.
49. Jaffe BM, Clendinnen BJ, Clark RJ, Alexander-Williams J. The effect of selective and proximal vagotomy on serum gastrin. *Gastroenterology* 1974;66:944–953.
50. Stadil F, Rehfeld JF, Christiansen PM, Kronborg O. Gastrin response to food in duodenal ulcer patients before and after selective and highly selective vagotomy. *Br J Surg* 1974;61:884–888.
51. Stadil F, Rehfeld JF. Gastrin response to insulin after selective, highly selective and truncal vagotomy. *Gastroenterology* 1974;66:7–15.
52. Thompson JC, Lowder WS, Peurifoy JT, Swierczek JS, Rayford PL. Effects of selective proximal vagotomy and truncal vagotomy on gastric acid and serum gastrin responses to a meal in duodenal ulcer patients. *Ann Surg* 1978;188:431–438.
53. Geurts WJC, Winckers EKA, Wittebol P. The effects of highly selective vagotomy on secretion and emptying of the stomach. *Surg Gynecol Obstet* 1977;145:286–336.
54. Stadaas JO. Gastric motility 1 year after proximal gastric vagotomy. *Scand J Gastroenterol* 1980;15:799–804.
55. Herrington JL Jr, Davidson J III, Shumway SJ. Proximal gastric vagotomy. Follow-up of 109 patients for 6-13 years. *Ann Surg* 1986;204:108–113.
56. Kelly KA. Gastric emptying of liquids and solids: Roles of proximal and distal stomach (editorial review). *Am J Physiol* 1980;239:71–76.
57. Rosenquist CJ, Carrigg JW, Regal A-MY, Kohatsu S. Electrical, contractile, and radiographic studies of the stomach after proximal gastric vagotomy. *Am J Surg* 1977;134:338–342.
58. Dewar P, King R, Johnston D. Bile acid and lysolecithin concentrations in the stomach in patients with duodenal ulcer before operation and after treatment by highly selective vagotomy, partial gastrectomy, or truncal vagotomy and drainage. *Gut* 1982;23:569–577.
59. Jian R, Kilian D, Grall Y et al. Effect de la vagotomie fundique sur la vidange gastrique des solides et des liquides d'un repas chez des patients avec et sans recidive ulcereuse duodenale post-operatoire. *Gastroenterol Clin Biol* 1985;9:486–490.
60. Braasch JW, Sala LE, Ellis FH Jr, Crozier R. Parietal cell vagotomy: its effect on lower esophageal sphincter function. *Arch Surg* 1980;115:699–701.

61. Temple JG, Goodall RJR, Hay DJ, Miller D. Effect of highly selective vagotomy upon the lower esophageal sphincter. *Gut* 1981;22:368–370.
62. Guelrud M, Zambrano-Rincones V, Simon C et al. Dysphagia and lower esophageal sphincter abnormalities after proximal gastric vagotomy. *Am J Surg* 1985;149:232–235.
63. Anderson JR, McLean Ross AH, Din NA, Small WP. Cholelithiasis following peptic ulcer surgery: A prospective controlled study. *Br J Surg* 1980;618–620.
64. Tompkins RK, Kraft AR, Zimmerman E, Lichtenstein JE, Zollinger RM. Clinical and biochemical evidence of increased gallstone formation after complete vagotomy. *Surgery* 1972;71:196–200.
65. Clave RA, Gaspar MR. Incidence of gallbladder disease after vagotomy. *Am J Surg* 1969;118:169–174.
66. Fisher RS, Rock E, Malmud LS. Cholinergic effects on gallbladder emptying in humans. *Gastroenterology* 1985;89:716–722.
67. Fisher RS, Rock E, Malmud LS. Gallbladder emptying response to sham feeding in humans. *Gastroenterology* 1986;90:1854–1857.
68. Baxter JN, Grime JS, Critley M. Shields R. Effect of truncal vagotomy and pyloroplasty and of highly selective vagotomy alone, on gallbladder emptying dynamics: A prospective study. *Gut*1984;25:A578.
69. Wilber BG, Gomez FC, Tompkins RK. Canine gallbladder bile. Effects of proximal gastric vagotomy, truncal vagotomy, and truncal vagotomy with pyloroplasty on volume and composition. *Arch Surg* 1975;110:792–796.
70. Ihasz M, Griffith CA. Gallstones after vagotomy. *Am J Surg* 1981;141:48–50.
71. Kuzin MT, Postolov PM. Selective proximal vagotomy in the treatment of duodenal ulcer. *World J Surg* 1980;4:347–352.
72. Lavigne ME, Wiley ZD, Martin P et al. Gastric, pancreatic and biliary secretion and the rate of gastric emptying after parietal cell vagotomy. *Am J Surg* 1979;138:644–651.
73. Tatkan Y. Effets des vagotomies sur la sécrétion de la sécrétine. *J Chir* 1985;122:711–716.
74. Berstad A, Roland M, Petersen H, Liavac I. Altered pancreatic function after proximal gastric vagotomy in man. *Gastroenterology* 1976;71:958–960.
75. Brooks AM, Stenning GF, Grossman MI. Effect of gastric vagal denervation on inhibition of acid secretion by secretin. *Am J Dig Dis* 1971;16:193–202.
76. Lindskov J, Amtorp O, Larsen HR. The effect of highly selective vagotomy on exocrine pancreatic function in man. *Gastroenterology* 1976;70:545–549.
77. Ward AS, Bloom SR. The role of secretin in the inhibition of gastric secretion by intraduodenal acid. *Gut* 1974;15:889–897.
78. Edwards JP, Lyndon PJ, Smith RB, Johnston D. Faecal fat excretion after truncal, selective and highly selective vagotomy for duodenal ulcer. *Gut* 1974;15:521–525.
79. Stabile BE, Passaro E Jr. *Duodenal Ulcer: A Disease in Evolution,* Vol 21. Chicago: Year Book, 1984, pp. 1–79.
80. Thompson JC, Wiener I. Evaluation of surgical treatment of duodenal ulcer: short and long-term effects. *Clin Gastroenterol* 1984;13:569–600.
81. Corman J, Smeesters C, Beaudoin M et al. Valeur predictive des tests d'acidite preoperatoires dans la recidive ulcereuse apres vagotomie supra-selective. *Can J Surg* 1983;26:348–350.
82. Weerts JM, Jacquet N, Lombard R et al. La vagotomie supra-selective: une intervention de choix dans la maladie ulcereuse duodenale. *Rev Med Liege* 1985;40:742–746.
83. Boey J, Lee NW, Koo J, Lam PHM, Wong J, Ong GB. Immediate definitive surgery for perforated duodenal ulcers. A prospective controlled trial. *Ann Surg* 1982;196:338–344.
84. Ferraz EM, Filho HA, Bacelar TS, Lacerda CM, DeSouza AP, Kelner S. Proximal gastric vagotomy in stenosed or perforated duodenal ulcer. *Br J Surg* 1981;68:452–454.
85. Sawyers JL, Herrington JL. Perforated duodenal ulcer managed by proximal gastric vagotomy and suture application. *Ann Surg* 1977;185:656–660.

86. Devis A, Becker C, DeGraef J. Traitement par vagotomie supra-selective des ulceres pyloro-duodenaux perfores. *Acta Gastroent Belg* 1984; 47:595–602.
87. Knight CD Jr, Van Heerden JA, Kelly KA. Proximal gastric vagotomy. Update. *Ann Surg* 1983;197:22–26.
88. Rossi RL, Dial PF, Georgi B, Braasch JW, Shea JA. A five to ten year follow-up study of parietal cell vagotomy. *Surg Gynecol Obstet* 1986;162:301–306.
89. McMahon MJ, Greenall MJ, Johnston D, Goligher JC. Highly selective vagotomy plus dilatation of the stenosis compared with truncal vagotomy and drainage in the treatment of pyloric stenosis secondary to duodenal ulceration. *Gut* 1976;17:471–476.
90. Dunn DC, Thomas WEG, Hunter JO. Highly selective vagotomy and pyloric dilatation for duodenal ulcer with stenosis. *Br J Surg* 1981;68:194–196.
91. Hooks VH, Bowden TA Jr., Mansberger AR Jr., Sisley JF. Highly selective vagotomy with dilatation or duodenoplasty. A surgical alternative for obstructing duodenal ulcer. *Ann Surg* 1986;203:545–550.
92. Jordan PH Jr. Treatment of gastric ulcer by parietal cell vagotomy and excision of the ulcer. *Arch Surg* 1981;116:1320–1323.
93. Reid DA, Duthie HL, Bransom CJ, Johnson AG. Late follow-up of highly selective vagotomy with excision of the ulcer compared with Billroth T gastrectomy for treatment of benign gastric ulcer. *Br J Surg* 1982;69:605–607.
94. Emas S, Hammarberg C. Prospective, randomized trial of selective proximal vagotomy with ulcer excision and partial gastrectomy with gastroduodenostomy in the treatment of corporeal gastric ulcer. *Am J Surg* 1983;146:631–634.
95. Jordan PH Jr. Parietal cell vagotomy facilitates fundoplication in the treatment of reflux esophagitis. *Surg Gynecol Obstet* 1978;147:593–595.
96. Samana G, Mercier V, Mounier S. Traitement du reflux gastrooesophagien. Interet de la vagotomie acido-fundique associee a la fundoplicature de Nissen. *Presse Med* 1986;15:255–256.
97. Linhardt GE, Stoddard CG, Johnson AG. Do women do worse after proximal gastric vagotomy? *Br J Surg* 1982;69:321–322.
98. Blackett RL, Johnston D. Recurrent ulceration after highly selective vagotomy for duodenal ulcer. *Br J Surg* 1981;68:705–710.
99. Solhand JH, Bjerkeset T, Halvorsen JF. Highly selective vagotomy in the treatment of duodenal ulcer in a teaching hospital. A one to three year follow-up relating results to the number of operating surgeons, their surgical experience and training conditions. *Surgery* 1977;82:248–253.
100. Trout HH. Ulcer recurrence, morbidity, and mortality after operations for duodenal ulcer. *Am J Surg* 1982;144:570–572.
101. Hoffmann J, Meisner S, Jensen HE. Antrectomy for recurrent ulcer after parietal cell vagotomy. *Br J Surg* 1983;70:120–121.
102. Alexander-Williams J. The complications of vagotomy and pylorplasty. In: *Surgery of the Stomach and Duodenum* (Nyhus LM, Wastell C, eds). Boston: Little, Brown, 1977, pp. 272–274.
103. Dorricott NG, McNeish AR, Alexander-Williams J et al. Prospective randomized multicenter trial of proximal gastric vagotomy or truncal vagotomy and antrectomy for chronic duodenal ulcer: Interim results. *Br J Surg* 1978;65:152–54.

11

Altered Carbohydrate Antigen Expression In Colorectal Cancer And Polyps

Young S. Kim

1. INTRODUCTION

Cancer of the colon and rectum is the most common gastrointestinal malignancy in many Western countries. In the United States, it is estimated that in 1989 there will be 150,000 new cases of colorectal cancer diagnosed and that 65,000 deaths from colorectal cancer will occur. The overall 5-year survival rate for patients with colorectal cancer is only 40–45%, a figure that has changed very little over the last two decades.[1] Disease prevention appears to provide the best hope for improving the outcome of individuals with colorectal cancer. There are two approaches to disease prevention: *Primary prevention* involves identifying and eliminating the cause of the disease and also includes careful monitoring of populations at high risk for developing the disease. In the case of colorectal cancer, various environmental and host factors have been suggested to play a role in carcinogenesis, but the actual causative agents are not yet known. Recently, some progress has been made in identifying subjects at high risk for developing colorectal cancers among family members of familial polyposis coli or colon cancer family syndromes. The approach has been to investigate the biochemical or cell proliferative activities in normal appearing

Young S. Kim • Gastrointestinal Research Laboratory, Veterans Administration Medical Center, and Department of Medicine, University of California–San Francisco, San Francisco, California 94121.

colonic mucosa (i.e., "flat mucosa") in an effort to identify the genotype carrier state and to monitor the progression to severe dysplasia in affected patients.[2–5] *Secondary prevention* involves detection and interruption of neoplasms at an early stage to prevent metastasis. Of the approaches that have been used, removal of rectal polyps by rigid sigmoidoscopy seems to decrease the future development of rectal cancers and improve patient survival.[6] While sigmoidoscopy is useful, there are problems with patient (and even physician) compliance when used for screening asymptomatic individuals, and a large portion of the colon is not examined. The development of fecal occult blood tests has made colon cancer screening more manageable. Most studies have demonstrated that earlier stage neoplasms can be detected,[7] but no study has yet shown an improvement in survival in people screened for fecal occult blood. Although fecal occult blood testing is the best mass screening test available, of those who test positive, only half will be bleeding from a colorectal neoplasm, thereby causing many people to undergo extensive evaluations of the gastrointestinal tract for what ultimately proves to be benign disease. The development of new, more sensitive markers for colonic neoplasms could contribute significantly to the secondary prevention of this disease. So far, the most clinically useful tumor marker is the carcinoembryonic antigen (CEA) particularly in the "follow-up" phase after surgery, providing lead times for the detection of recurrences and metastases and to monitor therapy.[8]

When considering potential tumor-associated markers of colorectal cancer, multiple markers should be investigated because of tumor cell heterogeneity.[9] Furthermore, most tumor-associated markers used to date are absent or minimally expressed in poorly differentiated colon cancer cells. Therefore, phenotypic markers for these colon cancer cells need to be identified for this particularly ominous subgroup of colorectal cancers. Similarly, identification of markers for invasive and/or metastatic properties of colorectal cancer cells would prove useful in predicting prognosis and in therapy.

2. GLYCOPROTEIN AND GLYCOLIPID

Substantial evidence indicates that many of the known phenotypic markers for colon cancer are carbohydrate-containing molecules. In the colon, mucosal cells are rich in carbohydrates. Cellular carbohydrates can be broadly classified into two types: secretory and structural. An example of the former is intestinal mucin; the latter type includes glycoprotein and glycolipid components of cell surface membranes and of the membranes of various subcellular organelles (Fig. 1). The oligosaccharide moiety of glycoproteins and glycolipids has the potential for generating structural diversity since the information is based not only on the number of different monomeric units and their sequence, but also on the

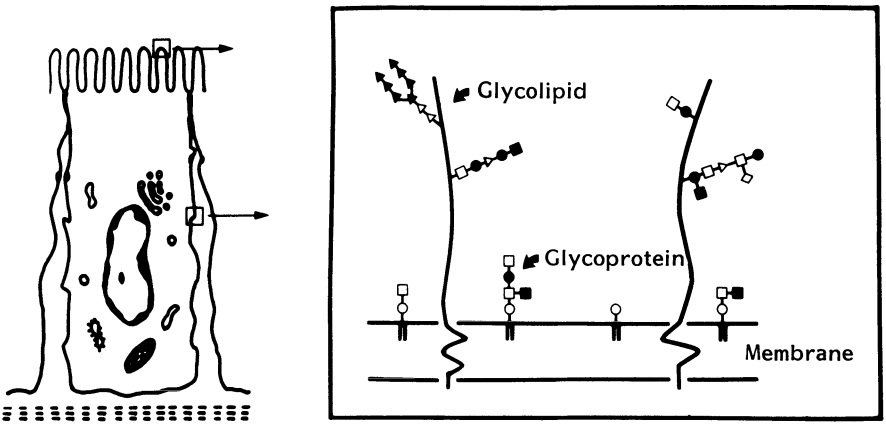


Figure 1. Schematic diagram of membrane-associated glycoproteins and glycolipids. Squares, circles, and triangles represent monosaccharide.

position, anomeric configuration of the glycosidic units, and the presence of multiple branch points. As a result, carbohydrate polymers can generate considerable information and have been implicated in many important biological functions such as receptor functions, growth regulation, cellular differentiation, cell–cell interaction, cell–substratum interactions, and diverse immunological functions.[10]

3. ALTERED EXPRESSION OF BLOOD GROUP SUBSTANCES IN COLORECTAL CANCER AND POLYPS

Increasingly over the last few years, carbohydrate structures related to blood group substances (BGS) are being recognized as important cancer-associated antigens in the gastrointestinal tract.[10–12] The carbohydrate epitopes reside on oligosaccharide side chains of glycolipids and glycoproteins. Their synthesis depends on the action of specific glycosyltransferase enzymes that sequentially add individual sugar residues in specific linkages to the oligosaccharide backbone.

Four main types of changes in carbohydrates have been observed in colorectal cancer. These are (1) expression of an antigen not usually found in normal mucosa, (2) deletion of an antigen usually expressed by normal mucosa with or without accumulation of precursors, (3) neosynthesis, which occurs either by modification of existing structures (such as side chain elongation or addition of sialic acid or fucose) or synthesis of new structures, and (4) carbohydrate antigens that are found in normal mucosa increase or decrease in cancer

cells. Although the biological significance of tumor-associated carbohydrate antigens is unknown, they may reflect developmental arrest or an aberrant program of differentiation.

Three main groups of tumor-associated carbohydrate antigens in colorectal cancer that have attracted attention in recent years will be discussed. These are (1) blood group antigens, (2) fucose-containing Le^x-related antigens, and (3) fucose-containing Le^y-related antigens. These latter two groups of antigens, Le^x and Le^y, have only recently been studied in colon cancer.

4. *ABH AND Le^a AND Le^b ANTIGENS AND CA-19-9 ANTIGEN (OR SIALYL-Le^a)*

A simplified version of carbohydrate structures responsible for the ABH Le^a and Le^b blood group antigen expression is shown in Table 1. Two types of carbohydrate backbone structure may occur in blood group antigens. The carbohydrate sequence of Gal (galactose) β1, 3 GlcNAc (*N*-acetylglucosamine) is called type 1 lactosamine backbone, while Gal β1, 4 GlcNAc is termed type 2 lactosamine backbone or chain. When fucose is added to galactose of the lactosamine backbone, blood group H antigen is formed. Addition of either GalNAc or galactose to H antigen converts it to A or B antigens. Thus ABH antigen could be expressed on both type 1 and type 2 chains. In contrast, Le^a and Le^b antigens can be expressed only on type 1 chain since the fourth carbon position of GlcNAc of type 2 chain is already occupied, so that fucose cannot be linked to GlcNAc at that position.

When the fucose is added in α1–4 linkage to the penultimate GlcNAc of

Table 1. Immunodeterminant Structure of ABH Blood Group Antigens

Antigens	Type 1	Type 2	Antigens
	$Gal\xrightarrow{\beta 1,3}GlcNAc-R$	$Gal\xrightarrow{\beta 1,4}GlcNAc-R$	
H_1	$Gal\xrightarrow{\beta 1,3}GlcNAc-R$ $\uparrow \alpha 1,2$ Fuc	$Gal\xrightarrow{\beta 1,4}GlcNAc-R$ $\uparrow \alpha 1,2$ Fuc	H_2
A_1	$GalNAc\xrightarrow{\alpha 1,3}Gal\xrightarrow{\beta 1,3}GlcNAc-R$ $\uparrow \alpha 1,2$ Fuc	$GalNAc\xrightarrow{\alpha 1,3}Gal\xrightarrow{\beta 1,4}GlcNAc-R$ $\uparrow \alpha 1,2$ Fuc	A_2
B_1	$Gal\xrightarrow{\alpha 1,3}Gal\xrightarrow{\beta 1,3}GlcNAc-R$ $\uparrow \alpha 1,2$ Fuc	$GalNAc\xrightarrow{\alpha 1,3}Gal\xrightarrow{\beta 1,4}GlcNAc-R$ $\uparrow a1,2$ Fuc	A_2

type 1 chain, Le^a antigen is formed. The addition of a second fucose to galactose converts Le^a to Le^b antigen. Gene-specified glycosyltransferases are responsible for the addition of each sugar to its acceptor in a specific anomeric linkage, thereby conferring the final antigenic structure. The addition or deletion of even one monosaccharide structure may completely alter the antigenicity of the molecule.

In the human embryo and fetus, major blood group antigens are expressed in the epithelial cells of both the proximal and distal colon. However, after birth, the expression of A, B, H, and Le^b antigens is lost from the distal colon while being preserved in more proximal segments.[13–15] Le^a antigen, however, remains uniformly distributed throughout the proximal and distal adult colon.

For many years, changes in the A, B, and H antigens expressed have been reported to occur in many epithelial cancer tissues including colon cancer.[15–17] Recently, using five monoclonal antibodies, we examined the pattern of blood group antigen expression in colon cancer tissues.[15] In colon cancer, the most frequent finding was the reexpression of blood group antigens (A, B, H, Le^b) in the distal segment (occurring in 90% of distal colorectal cancer), suggesting that an oncofetal gene function had been derepressed or reactivated in a situation analogous to that of CEA. Another intriguing finding was that over 60% of colon cancers of both proximal and distal colon expressed blood group antigens (A or B) incompatible with the host's blood type. For example, some blood group A patients expressed B antigen in the colon cancer tissues. Deletion of the host's blood group antigen from cancer cells was also observed in nearly half of the colon cancers occurring in the proximal colon.

ABH expression has also been examined in hyperplastic polyps, which are considered to have no malignant potential, and in adenomatous polyps, which may have malignant potential.[18] Hyperplastic polyps did not exhibit incompatibility and only rarely demonstrated blood group antigen reappearance but deletion phenomenon was observed in a few specimens obtained from right colon. Adenomatous polyps, however, frequently reexpressed blood group antigen in the distal colon, expressed incompatible blood group antigen A or B in 27% of polyps, and one specimen demonstrated blood group antigen deletion. Thus, some of the antigenic alterations that occur in colorectal cancer tissue also appear in the premalignant variant of polyps, often in early stages of the neoplastic process. The biochemical and molecular basis, as well as the biological significance for reexpression of fetal antigens, appearance of incompatible blood groups, or deletion of blood group substances, are not well understood.

Another blood group-related antigen that we have examined is the gastrointestinal cancer-associated antigen (GICA), which is recognized by MAb 19-9 (Table 2). This antigen, a sialylated Le^a blood group antigen, occurs as a ganglioside in tissues, but in serum and pancreatic juices it is found on mucin-type glycoproteins.[19,20] Most specimens of colon cancer and second trimester

Table 2. Immunodeterminant Structure of Le^a, Le^b, and Other Related Antigens

Antigens	Type 1 Gal$^{\beta 1,3}$GlcNAc–R	Type 2 Gal$^{\beta 1,4}$GlcNAc–R	Antigens
Le^a	Gal$^{\beta 1,3}_{\rightarrow}$GlcNAc→R ↑ α1,4 Fuc	Gal$^{\beta 1,4}_{\rightarrow}$GlcNAc→R ↑ α1,3 Fuc	Le^x(X)
Le^b	Gal$^{\beta 1,3}_{\rightarrow}$GlcNac→R ↑ α1,2 ↑ α1,4 Fuc Fuc	Gal$^{\beta 1,4}_{\rightarrow}$GlcNAc→R ↑ α1,2 ↑ α1,3 Fuc Fuc	Le^y(Y)
Sialyl-Le^a GICA or CA 19-9	Gal$^{\beta 1,3}_{\rightarrow}$GlcNAc–R ↑ α2,3 ↑ α1,4 NeuAc Fuc		

fetal colon express this antigen, whereas normal mucosa rarely express it.[21,22] Thus, a modification of Le^a epitope by the addition of sialic acid appears to produce an oncodevelopmental antigen in colon cancer tissues.

5. T ANTIGEN

Changes in the expression of another major class of blood group antigen have been recently reported in some epithelial tumors. This is the T antigen that is related to M and N blood group antigen, as shown in Fig. 2. The tetrasaccharide structure linked to threonine or serine, which contributes to M and N antigenicity, is made of up to two sialic acids and a galactose-linked β1–3 to GalNAc (*N*-acetylgalactosamine). When both sialic acids are removed (or not synthesized), a disaccharide appears that is the immunodominant structure for so-called T antigen. The lectin *(Arachis hypogoea)* peanut agglutinin or PNA recognizes this disaccharide structure. This disaccharide structure is also present in carbohydrates other than M and N antigens, e.g., in the core structure of mucin glycoproteins.

During the past several years, Springer and his associates as well as other investigators reported enhanced expression of T antigen in several epithelial cancer tissues.[23,24] Using fluorescein-conjugated (FITC) peanut lectin, we have observed that many colon cancer tissues stain intensely with FITC-peanut lectin while distant normal mucosa does not.[25,26] Recently, we extended these studies by comparing the reactivity of normal colonic mucosa, nonneoplastic diseased mucosa, colonic polyps (hyperplastic and adenomatous), colon cancer, and fe-

tal colonic mucosa to peanut lectin polyclonal and monoclonal anti-T antisera using a more sensitive immunoperoxidase method.[27] All three reagents showed intense staining of most colon cancer tissues in both proximal and distal colon. However, polyclonal and particularly monoclonal anti-T antibody showed much less reactivity with normal mucosa and hyperplastic polyps compared to peanut lectin. Furthermore, both polycolonal and monoclonal anti-T antibody staining appeared to correlate with the malignant potential of polyps as assessed by polyp size, histological type, and degree of dysplasia.

6. Le^x AND Le^y-RELATED ANTIGENS

Le^x and Le^y antigens are positional isomers of Le^a and Le^b, respectively, with fucose attached to type 2 lactosamine backbone as shown in Tables 3 and 4. The immunodeterminant structure for Le^x or X antigen is terminal trisaccharide structure with a fucose-linked α1,3 to the penultimate GlcNAc of lactosamine structure (Table 3). Alternatively, when an additional fucose is added in α1,2 to the terminal galactose of Le^x antigen, the immunodeterminant tetrasaccharide structure for the Le^y or *Y* antigen is formed. The epitopes for Le^x or Le^y antigens can reside on short or extended lactosamine backbones of carbohydrate side chains of glycolipids and glycoproteins with or without additional internal fucose.

Recently, several monoclonal antibodies (MAbs) have been established that recognize Le^x or Le^y determinants (Tables 3 and 4).[10,28–35] Among the MAbs having specificities for Le-related antigens, anti-SSEA-1 and AH8-183 MAbs recognize the Le^x trisaccharide present on either short or extended lac-

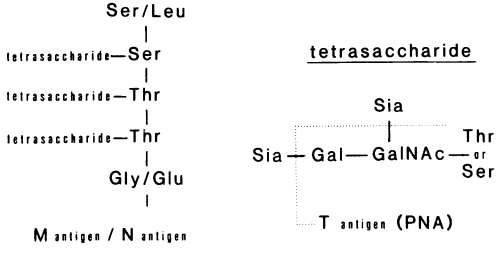


Figure 2. Structures of MN-related antigens. *N*-terminal structure of M & N antigens on the left and tetrasaccharide structures on the right. The disaccharide Galβ1,3GalNAc is the immunodeterminant structure for T antigen. Sia, sialic acid or *N*-acetyl neuraminic acid; Gal, galactose; GalNAc, *N*-acetylgalactosamine; Thr, threonine; Ser, serine; Leu, leucine; Gly, glycine; Glu, glutamic acid; PNA, peanut agglutinin.

Table 3. Le^x-Related Antigen Expression in Colonic Tissues

Antigens (MoAb)	Antigenic determinant	Antigen expression in tissues				
		N[a]	C	F	HP	AP
Le^x (anti-SSEA-1 AH8-183)	$Gal^{\beta1,4}_{\rightarrow}GlcNAc-R$ $\uparrow \alpha1,3$ Fuc	+	+	+	+	+
Extended Le^x (FHI)	$Gal^{\beta1,4}_{\rightarrow}GlcNAc^{\beta1,3}_{\rightarrow}Gal^{\beta1,4}_{\rightarrow}GlcNAc-R$ $\uparrow \alpha1,3$ Fuc	−	+	+	−	+
Difucosyl Le^x (FH4)	$Gal^{\beta1,4}_{\rightarrow}GlcNAc^{\beta1,3}_{\rightarrow}Gal^{\beta1,4}_{\rightarrow}GlcNAc-R$ $\uparrow \alpha1,3$ $\quad\quad$ $\uparrow \alpha1,3$ Fuc $\quad\quad$ Fuc	−	+	+	−	+
Sialyl difucosyl Le^x (FH6)	$Gal^{\beta1,4}_{\rightarrow}GlcNAc^{\beta1,3}_{\rightarrow}Gal^{\beta1,4}_{\rightarrow}GlcNAc-R$ $\uparrow \alpha2,3$ $\quad$ $\uparrow \alpha1,3$ $\quad\quad$ $\uparrow \alpha1,3$ NeuAc $\quad$ Fuc $\quad\quad$ Fuc	−	+	+[b]	−	+
Sialyl monofucosyl Le^x (IB9)	$Gal^{\beta1,4}_{\rightarrow}GlcNAc^{\beta1,3}_{\rightarrow}Gal^{\beta1,4}_{\rightarrow}GlcNAc-R$ $\uparrow \alpha2,3$ $\quad\quad$ $\uparrow \alpha1,3$ NeuAc $\quad\quad$ Fuc	−	+	+	−	+

[a]N, normal mucosa; C, cancer; F, fetal mucosa; HP, hyperplastic polyp; AP, adenomatous polyp.
[b]Negative in distal colon.

Table 4. Le^y-Related Antigen Expression in Colonic Tissues

Antigens (MoAb)	Antigenic determinant	Antigen expression in tissues				
		N[a]	C	F	HP	AP
Le^y (AH6)	$Gal^{\beta1,4}_{\rightarrow}GlcNAc-R$ $\uparrow \alpha1,2$ $\quad$ $\uparrow \alpha1,3$ Fuc $\quad$ Fuc	+	+	+	+	+
Extended Le^y (CC-1 CC-2)	$Gal^{\beta1,4}_{\rightarrow}GlcNAc^{\beta1,3}_{\rightarrow}Gal^{\beta1,4}_{\rightarrow}GlcNAc-R$ $\uparrow \alpha1,2$ $\quad$ $\uparrow \alpha1,3$ Fuc $\quad$ Fuc	±[b]	+	+[b]	−	+
Trifucosyl Le^y (KH-1)	$Gal^{\beta1,4}_{\rightarrow}GlcNAc^{\beta1,3}_{\rightarrow}Gal^{\beta1,4}_{\rightarrow}GlcNAc-R$ $\uparrow \alpha1,2$ $\quad$ $\uparrow \alpha1,3$ $\quad\quad$ $\uparrow \alpha1,3$ Fuc $\quad$ Fuc $\quad\quad$ Fuc	±[b]	+	+[b]	−	+

[a]N, normal mucosa; C, cancer; F, fetal mucosa; HP, hyperplastic polyp; AP, adenomatous polyp.
[b]Negative in distal colon.

tosamine (type 2 chain) backbones. The FH-1 and FH-4, FH-6, and KB-9 MAbs have greatest affinity for extended Le^y determinants. The FH-1 MAb simply recognizes an extended Le^x epitope, whereas FH-4 and FH-6 specifically recognize an extended Le^x epitope containing an additional internal fucose linked in α1,3 linkage to GlcNAc and thus the FH-4 determinant is referred to as "difucosyl Le^x." The FH-6 MAb has additional specificity because of *N*-acetyl neuraminic acid (sialic acid) (NeuAc) added to terminal galactose (α2,3) of "difucosyl Le^x." The MAb IB-9 has greatest affinity for extended Le^x structure with only one internal fucose linked to GlcNAc(α1,3) and NeuAc linked to terminal galactose (α2,3). Both FH-6 and KB-9 therefore recognize sialylated Le^x antigens.[33]

There are several MAbs that have specificities toward Le^y determinants. The AH-6 MAb recognizes that Le^y tetrasaccharide present on either short or extended lactosamine backbones. The other three MAbs recognize only extended Le^y determinants: CC-1 and CC-2 MAbs share similar epitope specificities and have the greatest affinity for extended Le^y antigens, and KH-1 MAb specifically recognizes an extended Le^y epitope containing an internal fucose linked in α1,3 linkage to GlcNAc (referred to as "trifucosyl Le^y").[35]

From our immunohistochemical data employing these MAbs, the following statements can be made (Tables 3 and 4)[34–36]:

1. Anti-SSEA-1 and AH8-183 (which recognize short or extended Le^x antigens) and AH-6 (which recognizes short or extended Le^y antigens) were expressed by virtually all colon cancers and polyps but also by most normal colonic mucosa.
2. MAbs that recognize extended Le^x antigens (e.g., FH-1) and extended Le^y antigens (e.g., CC-1 and CC-2) and MAbs that recognize polyfucosylated extended Le^x (e.g., FH-4) and polyfucosylated extended Le^y (e.g., KH-1) preferentially stained malignant colonic tissues and rarely stained normal mucosa. Furthermore, the expression of extended and polyfucosyl Le^x and Le^y antigens was limited exclusively to premalignant (adenomatous) polyps and was virtually absent from nonmalignant (hyperplastic) polyps. Moreover, among adenomatous polyps, extended and polyfucosylated antigen expression tended to correlate with three measures of malignant potential: larger polyp size, villous histology, and severe dysplasia.
3. The pattern of expression of sialylated Le^x antigen (defined by FH-6 and IB-9) was very similar to that exhibited by extended and polyfucosyl-Le^x and Le^y antigens.
4. All Le^x- and Le^y-related antigens were expressed in second-trimester fetal colon.

7. SUMMARY AND FUTURE PERSPECTIVE

Several carbohydrate structures behave as cancer-associated antigens in the human colon, although none is specific only for colon cancer. These antigens are expressed heterogeneously in colon cancer tissue due to tumor cell heterogeneity, and therefore the use of multiple MAbs will enhance the potential clinical usefulness of these reagents. The expression of some carbohydrate antigens by colon cancer is considered rather "cancer-specific" since normal colonic mucosa does not express these antigens as demonstrated by the following examples: (1) reappearance of ABH, Le^b in distal colon cancer, (2) incompatible A or B expression, (3) deletion of ABH, Le^a and Le^b from proximal colon cancer, (4) appearance of sialylated Le^x antigen in colon cancer, (5) appearance of T antigen, and (6) expression of extended and/or polyfucosylated or sialylated Le^a and Le^x antigens in colon cancer.

Most of these cancer-specific alterations may also be markers of premalignancy since adenomatous polyps but not hyperplastic polyps are capable of exhibiting these changes and antigenic expression often correlates with malignant potential in adenomatous polyps. These data raise a number of important questions, such as:

1. Why and how do malignant and premalignant colonic cells express these antigens? Are these changes cause–effect or epiphenomena associated with malignant transformation?
2. Do other preneoplastic colonic conditions exhibit similar antigenic alterations? For example, are those antigens expressed in the normal-appearing "flat mucosa" of affected individuals of familial polyposis coli or family cancer syndrome or in the dysplastic cells of ulcerative colitis patients?
3. What is the potential clinical role of these antigens in the diagnosis or management of patients with colorectal neoplasms?

Clearly, more work needs to be carried out in this exciting and promising area of investigation.

REFERENCES

1. Bresalier RS, Boland CR, Sack TL, Kim YS. Cancer of the gastrointestinal tract. In: *Gastroenterology Annual,* Vol 3 (Kern F, Blum AL, eds.). New York: Elsevier, 1986, pp. 413–469.
2. Luk GD, Baylin SB. Ornithine decarboxylase as a biologic marker in familial colonic polyposis. *N Engl J Med* 1984;311:80–83.
3. Lipkin M, Blattner WE, Fraumeni JF Jr, Lynch HT, Deschner E, Winawer S. Tritiated thy-

midine labeling distribution as a marker for hereditary predisposition to colon cancer. *Cancer Res* 1983;43:1899–1904.
4. Yonezawa S, Nakamura T, Tanaka S, Sato E. Binding of ulex europeus agglutinin-1 in polyposis coli: Comparative study with solitary adenoma in sigmoid colon and rectum. *J Natl Cancer Inst* 1983;71:19–23.
5. Boland CR, Lance P, Levin B, Riddell RH, Kim YS. Abnormal goblet cell glycoconjugates in rectal biopsies associated with an increased risk of neoplasia in patients with ulcerative colitis: Early results of a prospective study. *Gut* 1984;25:1364–1371.
6. Gilbertson VA. Proctosigmoidoscopy and polypectomy in reducing the incidence to rectal cancer. *Cancer* 1974;34:936–939.
7. Simon JB. Occult blood screening for colorectal carcinoma: A critical review. *Gastroenterology* 1985;88:820–837.
8. Neville AM. International union against cancer report: Workshop on immunodiagnosis. *Cancer Res* 1986;46:3744–3746.
9. Schlom J. Basic principles and applications of monoclonal antibodies in the management of carcinomas: The Richard and Hinda Rosenthal Foundation Award Lecture. *Cancer Res* 1986;46:3225–3238.
10. Hakomori S. Aberrant glycosylation in cancer cell membranes as focussed on glycolipids: overview and perspectives. *Cancer Res* 1985;45:2405–2414.
11. Hakomori S. Glycosphingolipids. *Sci Am* 1986;254:44–53.
12. Itzkowitz SH, Kim YS. New carbohydrate tumor markers. *Gastroenterology* 1986;90:491–494.
13. Szulman AE. The histological distribution of the blood group substances in man as disclosed by immunofluorescence, III. The ABO and H antigens in embryos and fetuses from 18 mm in length. *J Exp Med* 1964;119:503–515.
14. Szulman AE, Marcus DM. The histologic distribution of the blood group substances in man as disclosed by immunofluorescence. VI. The Le^a and Le^b antigens during fetal development. *Lab Invest* 1973;28:565–574.
15. Yuan M, Itzkowitz SH, Palekar A et al. Distribution of blood group antigens, A, B, H, Le^a and Le^b in normal, fetal and malignant colonic tissue. *Cancer Res* 1985;45:4499–4511.
16. Denk H, Tappeiner G, Holzmer HJ. Blood group substances (BGS) as oncofetal antigens in carcinoma of the distal colon. *Eur J Cancer* 1974;10:487–490.
17. Cooper HS, Haesler WE. Blood group substances as tumor antigens in the distal colon. *Am J Clin Pathol* 1978;69:594–598.
18. Itzkowitz SH, Yuan M, Ferrell LD, Palekar A, Kim YS. Cancer-associated alterations of blood group antigen expression in human colorectal polyps. *Cancer Res* 1986;46:5976–5984.
19. Magnani JL, Nilsson B, Brockhaus M. A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-N-funcopentose. II. *J Biol Chem* 1982;257:14365–145369.
20. Magnani JL, Steplewski Z, Koprowski H, Ginsburg V. Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin. *Cancer Res* 1983;43:5489–5492.
21. Atkinson BF, Ernst CS, Herlyn M, Steplewski Z, Sears HF, Koprowski H. Gastrointestinal cancer-associated antigen in immunoperoxidase assay. *Cancer Res* 1982;42:4820–4823.
22. Raux H, Labbe F, Fondaneche M, Koprowski H, Burtin P. A study of gastrointestinal cancer-associated antigen (GICA) in human fetal organs. *Int J Cancer* 1983;32:315–319.
23. Springer GF, Murthy SM, Desai PR, Fray WA, Tegtmeyer H, Scanlon EF. Patients immune response to breast and lung carcinoma-associated Thomsen-Friedenreich (T) specificity. *Klin Wochenschr* 1982;60:121–131.

24. Coon JS, Weinstein RS, Summers JL. Blood group precursor T-antigen expression in human urinary bladder carcinoma. *Am J Clin Pathol* 1982;77:692–699.
25. Boland CR, Montgomery CK, Kim, YS. Alterations in colonic mucin occurring with cellular differentiation and malignant transformation. *Proc Natl Acad Sci USA* 1982;79:2051–2055.
26. Boland CR, Montgomery CK, Kim YS. A cancer associated mucin alteration in benign colonic polyps. *Gastroenterology* 1982;82:644–672.
27. Yuan M, Itzkowitz SH, Boland CR et al. Comparison of T-antigen expression in normal premalignant and malignant human colonic tissue using lectin and antibody immunohistochemistry. *Cancer Res* 1986;46:4841–4847.
28. Fukushima K, Hirota M, Terasaki PI et al. Characterization of sialysylated Lewis as a new tumor-associated antigen. *Cancer Res* 1984;44:5279–5285.
29. Brown A, Ellis IO, Embleton MJ, Baldwin RW, Turner DR, Hardcastle JD. Immunohistochemical localization of Y hapten and the structurally related H type-2 blood group antigen on large bowel tumors and normal adult tissues. *Int J Cancer* 1984;33:727–736.
30. Hirota M, Fukushima K, Terasaki PI et al. Sialosylated Lewis in the sera of cancer patients detected by a cell-binding inhibition assay. *Cancer Res* 1985;45:1901–1905.
31. Sakamoto J, Furukawa K, Cordon-Cardo C et al. Expression of Lewis^a, Lewis^b, X and Y blood group antigens in human colonic tumors and normal tissue and in human tumor-derived cell lines. *Cancer Res* 1986;46:1553–1561.
32. Kannagi R, Fukushi Y, Tachikawa T et al. Quantitative and qualitative characterization of human cancer-associated serum glycoprotein antigens expressing fucosyl or sialyl-fucosyl type 2 chain polylactosamine. *Cancer Res* 1986;46:2619–2626.
33. Abe K, Hakomori S-I, Ohshiba S. Differential expression of difucosyl type 2 chain (Le^y) defined by monoclonal antibody AH6 in different locations of colonic epithelia, various histological types of colonic polyps, and adenocarcinomas. *Cancer Res* 1986;46:2639–2644.
34. Itzkowitz SH, Yuan M, Fukushi Y et al. Lewis^x and sialylated Lewis^y-related antigen expression in human malignant and nonmalignant colonic tissues. *Cancer Res* 1986;46:2627–2632.
35. Kim YS, Yuan M, Itzkowitz SH et al. Expression of Le^y and extended Le^y blood group-related antigens in human malignant, premalignant, and non-malignant colonic tissues. *Cancer Res* 1986;46:5985–5992.
36. Yuan M, Itzkowitz SH, Ferrell LD et al. Expression of Le^x and sialylated Le^x antigens in human colorectal polyps. *J Natl Cancer Inst* 1987;78:479–488.

12

Modern Concepts of Regulation of Intestinal Nutrient Transport

Jared M. Diamond

1. INTRODUCTION

Humans, like other animals, insult their intestines with wildly varying food inputs. We start out life as 7-lb babies, consuming small amounts of food, all of it in the form of milk, yet we end up as adults weighing between 40 and 140 kg, consuming quantities of food one or two orders of magnitude greater than the intake of babies, mostly in the form of solids. If we're lucky, we can eat whatever we want in whatever amounts we want: one day, a high-carbohydrate diet of elegant pastries; the next day, a no-carbohydrate, high-protein diet of filet mignon; the following day, neither pastry nor filet mignon, but a high-fat diet of rich cheese. If unlucky and poor, we may starve or occasionally get something to eat, but frequently suffer from deficiencies of vitamins, essential amino acids, or nitrogen. Even if living in affluence, we may be unlucky in other ways: we may develop diabetes or may require surgical resection of the intestine or else total parenteral nutrition. If working vigorously as lumberjacks, we may consume 7000 calories per day. Conversely, lying in bed all day requires only 800 calories. Half of us occasionally become pregnant, nurse

Jared M. Diamond • Department of Physiology, University of California–Los Angeles School of Medicine, Los Angeles, California 90024.

one or more infants, and must increase our calorie intake correspondingly (up to several-fold in some nursing mammals).

How does the intestine adapt to all these varying demands for extracting nutrients? By the beginning of this century, Pavlov discovered that pancreatic enzymes adapt to diet with amylase increasing on a high-starch diet and chymotrypsin on a high-protein diet. Dietary regulation of intestinal hydrolases, such as sucrase or aminooligopeptidase, has also been well studied. It is well known to developmental biologists that in most infant mammals the level of intestinal sucrase rises whereas lactase drops at the time of weaning.

But what about regulation of the final stage in the assimilation of food, i.e, nutrient transport from the intestine to the bloodstream? Beginning with the introduction of the everted sac technique for studying intestinal absorption in the 1950s, much progress has been made toward defining the cellular mechanisms of intestinal nutrient transport and, more recently, the molecular isolation of the transporters. However, little attention has been paid to regulation of these absorptive mechanisms, which instead have been tacitly assumed not to be regulated after the neonatal period. The reasons for neglecting the regulation of intestinal nutrient transport relate to difficulties in unraveling the physiological mechanisms themselves and to the incorrect assumption that regulation of intestinal nutrient transport would prove to be unnecessary and nonexistent because the human intestine normally absorbs virtually all simple carbohydrates and proteins in a meal and thus seemed to have an excessively large absorptive capacity.

There are three reasons why this assumption is incorrect. First, numerous clinical conditions, such as the short-gut syndrome,[1] are associated with malabsorption. Second, the normal absorptive capacity of the intestine is regularly exceeded in some common physiological conditions, such as pregnancy, lactation, and certain dietary variations. Finally, the focus on just the question of whether nutrients are completely absorbed assumes tacitly that the rate of absorption makes no difference, provided that absorption is complete. Yet the rate of absorption obviously determines the postprandial time course of blood nutrient concentrations, and that time course is itself important. For example, the symptoms of diabetes do not arise from incomplete intestinal absorption of carbohydrate but from a maladaptive time course of plasma glucose concentration. Similarly, dietary manipulations that vary the rate of intestinal amino acid absorption cause changes not only in the postprandial time course of blood amino acid concentrations but also in the concentration ratio for essential amino acids to nonessential amino acids in the blood (R. K. Buddington, personal communication).

This chapter will focus on the specific regulation of individual nutrient transporters in the intestine by dietary levels of their substrates. Three related topics will also be reviewed: the signals for such specific regulation, instances

of nonspecific regulation of absorption of many nutrients in parallel, and the ontogenetic regulation of intestinal nutrient absorption.

2. A SYNOPSIS OF INTESTINAL NUTRIENT TRANSPORTERS

Earlier studies of intestinal nutrient absorption usually measured transepithelial transport: the transport of nutrients from the intestinal lumen across the epithelium to the bloodstream or serosal surface. Familiar methods for measuring such transepithelial transport include transport measurements from the mucosal to serosal surface of the everted intestinal sac *in vitro,* and the measurement of solute appearance in the bloodstream after solutes were added to the lumen of the intestine *in vivo.* Such measurements of transepithelial transport lump together two separate membrane transport steps: nutrient uptake from the intestinal lumen, across the brush border (apical) membrane, into the enterocyte; and nutrient exit from the enterocyte, across the basolateral membrane, into the bloodstream.

The former step of transport from lumen to cell across the brush border membrane is the one that has received more attention. In the 1960s and 1970s this step was often studied by mounting an epithelial sheet in Ussing chambers, adding a radiolabeled nutrient to the mucosal bathing solution, and measuring the amount of radiolabel in the tissue after a short time. More recently, my colleagues and I studied this step by an everted sleeve preparation, by which the intestine is everted, a cylindrical sleeve 1 cm long is tied over a solid glass rod and mounted over a stirring bar, a radiolabeled nutrient is added to the bathing solution, and tissue label is again determined after a short time.[2] Both the chamber and the everted sleeve methods measure uptake into the intact tissue *in vitro.* An alternative procedure for studying the brush border transport

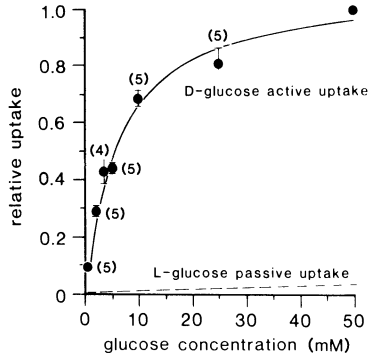


Figure 1. Uptake of glucose by everted sleeves of mouse jejunum, as a function of bath glucose concentration. The dashed straight line is for L-glucose, which is only taken up passively. The solid points are the active uptake of D-glucose, determined after correcting for passive uptake using L-glucose uptake. The solid curve is the best fit of the solid points to the Michaelis–Menten equation. Values are normalized to the active glucose uptake measured at 50 mM in an adjacent sleeve from the same mouse. (From Ref. 2.)

step is to prepare vesicles from the brush border membrane and to study uptake into these vesicles.

These three methods have been useful for demonstrating the existence of specific transporters for particular nutrients in the brush border membrane. For example, Fig. 1 illustrates results of our studies of brush border uptake of D-glucose in an everted sleeve of mouse jejunum.[2] The figure shows that uptake of glucose is stereospecific (because uptake of D-glucose is much faster than that of L-glucose). Uptake involves saturable kinetics with a Michaelis–Menten constant (K_m) around 6 mM. Other experiments with everted sleeves have demonstrated that brush border D-glucose uptake is Na^+-dependent, is inhibited competitively by other aldohexoses such as galactose or 3-*O*-methylglucose, and proceeds against a concentration gradient by building up a higher concentration of D-glucose in the tissue than in the mucosal bathing solution. Similar observations, made in other preparations, have been interpreted to mean that glucose absorption involves cotransport with Na^+, i.e., Na^+ enters the cell across the apical membrane by moving down its electrochemical gradient. This gradient is maintained by active extrusion of Na^+ at the basolateral membrane by the sodium pump, Na^+/K^+-ATPase. Transporters located in the brush border membrane couple the entry of Na^+ and glucose into the cell. The electrochemical gradient for Na^+ provides the driving force for the uphill movement of glucose into the enterocyte.

Similar experiments with other solutes have revealed the existence of at least 20 different transporters for nutrients in the intestinal brush border.[3] There are two separate transporters for glucose, one operating with a high capacity but low affinity, the other with a low capacity but high affinity.[4] Fructose is absorbed by a facilitated diffusion mechanism that does not convey fructose uphill and that is not Na^+-dependent.[5] There are at least four separate Na^+-dependent transporters for amino acids: the imino acids, the acidic amino acids, the basic amino acids, and the neutral amino acids. There may also be at least three separate Na^+-independent amino acid transporters.[6] Dipeptides have still another transporter,[7] as do short-chain fatty acids, and as do bile acids in the ileum. At least eight different water-soluble vitamins each have a specific transporter.[8,9] Specific transporters have been well documented for iron, calcium, and phosphate, and probably exist for numerous other essential trace minerals. In mammals (though not in some fish and other vertebrates), these nutrient transporters are mainly located in the small intestine; the large intestine contributes less than 10% of the small intestine's transport capacity.

The basolateral transport step from the enterocyte into the bloodstream has been much less well defined because of the inaccessibility of this membrane in the intact intestine. However, studies of basolateral vesicles have shown that this membrane does have specific transporters for sugars, amino acids, iron, and probably other nutrients.[10,11]

3. PATTERNS OF SPECIFIC REGULATION OF BRUSH BORDER TRANSPORTERS BY THEIR SUBSTRATES

If one increases the dietary input of some particular solute and then measures its absorption *in vivo,* there is an increased rate of absorption, simply because the solute's concentration in the intestinal lumen increases and the instantaneous uptake/concentration relation generally rises monotonically. Such observations in themselves provide no evidence for regulation of the transporter by diet. However, one can test for regulation by altering an animal's diet, then excising the intestine and measuring the uptake/concentration relation *in vitro.* This method allows comparison of transport rates from solutions with the same substrate concentration but in animals that had been maintained on different rations. A difference in uptake would then imply regulation of the transporter.

Figure 2 illustrates such an experiment demonstrating regulation of intestinal brush border glucose and proline transport.[12] Mice were maintained on a carbohydrate-free, high-protein ration, then switched to a high-carbohydrate, medium-protein ration. At various times after the ration switch, animals were sacrificed, their intestines excised, and uptake by an everted sleeve measured at a glucose or proline concentration high enough to saturate the transporters and yield a measure of the V_{max}. It turns out that both the D-glucose transporter and the L-proline transporter are up-regulated by increased dietary intake of their respective substrates, i.e., glucose transport increases on the high-carbo-

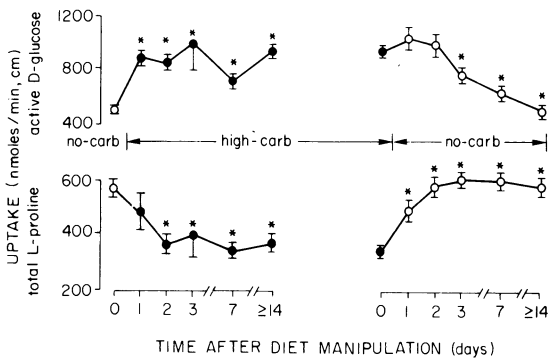


Figure 2. Time course for effect of diet change on active D-glucose uptake and total L-proline uptake measured at 50 mM in the proximal jejunum. Left half, mice are switched from no-carbohydrate (no-carb), high-protein diet (open circles) to high-carbohydrate (high-carb), low-protein (closed circles) at time $t=0$; subsequent points give uptake rates for glucose (top) and proline (bottom) after the indicated number of days on the high-carbohydrate, low-protein diet. Right half, mice switched from high-carbohydrate, low-protein diet to no-carbohydrate, high-protein diet at $t = 0$; subsequent points give uptake rates after the indicated number of days on the no-carbohydrate, high-protein diet. Vertical bars give ± SEM ($n=15$–24 mice for $t = 0$ points and 5–10 mice for other points). Asterisks indicate uptake values that differ significantly ($p<0.05$, t test) from the corresponding $t=0$ value. (From Ref. 12.)

hydrate ration, while proline transport increases on the high-protein diet. Up-regulation reaches a plateau about 1 day after the animal begins eating the high-substrate ration, whereas down-regulation is not complete until several days after the animal has returned to the low-substrate (or no-substrate) regimen. The opposite changes in glucose transport and proline transport suggest specific and independent regulation of each transporter, as does the observation that glucose transport also varies with dietary carbohydrate level when carbohydrate is replaced with fat and when dietary protein level is held constant. Specific regulation is further indicated by the fact that a high-fructose diet stimulates fructose transport without affecting proline transport.[5] Although Fig. 2 demonstrates dietary regulation only of the brush border transport step, studies of basolateral vesicles have also demonstrated regulation of the basolateral transport step for glucose.[13]

Specific regulation by dietary substrate levels has been demonstrated in recent years for many other nutrient transporters. However, the case of D-glucose remains one of the best studied. Kinetic analysis of glucose uptake as a function of glucose concentration in mice either on a high-carbohydrate or a carbohydrate-free ration reveals that the increased uptake is due to an increase in the maximal transport velocity (V_{max}). There is no convincing evidence of a change in K_m.[14] *Prima facie,* this suggests that the high-carbohydrate diet is associated with an increased number of copies of the transporter without any qualitative change in the binding constant of each copy. Such an interpretation has been confirmed by experiments in which the density of glucose transporter sites in the intact intestinal mucosa was measured by binding studies with phlorizin, a specific inhibitor of glucose transport.[15] Probably dietary carbohydrate ultimately regulates synthesis of the glucose transport protein and induces the transporter.

Variation in transporter site density also appears to be responsible for the familiar gradient of glucose absorption rate along the intestine. Glucose absorption, whether measured *in vivo* or *in vitro,* is maximal in the proximal jejunum and lower in both the duodenum and especially the ileum. This regional gradient of glucose transporter activity parallels the regional gradient in mean glucose concentration along the lumen of the intestine because glucose is released by hydrolysis from complex carbohydrates proximally in the small intestine and is absorbed in transit, resulting in maximal glucose concentrations in the jejunum. If the average luminal glucose concentration were to regulate the transporter site density at that point along the intestine, one would expect transporter density to be maximal in the jejunum, resulting in the highest absorption rates there. This interpretation is confirmed by phlorizin-binding studies, which show higher glucose transporter site density in the proximal intestine than in the distal intestine.[15] Reinforcing this interpretation is the fact[14] that the normal gradient of glucose transport activity along the intestine disappears in mice on

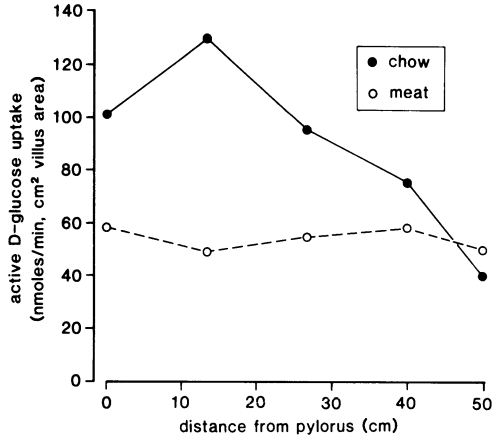


Figure 3. Active uptake of 50 mM D-glucose by everted sleeves of mouse intestine as a function of intestinal position (abscissa, distance from pylorus: proximal duodenum at far left, terminal ileum at far right). Glucose uptake values are normalized to surface area at the villus level. Solid upper curve with solid circles: mice maintained on a high-carbohydrate, medium-protein diet of chow. Lower dashed curve with open circles: mice maintained on a high-protein, low-carbohydrate diet of meat. Note that regional differences in glucose transport along the length of the intestine, observed on the high-carbohydrate diet, disappear on the meat diet that is nearly devoid of carbohydrate. (From Ref. 14.)

a carbohydrate-free diet, which would abolish the usual regional gradient of luminal glucose concentration (Fig. 3).

Dietary regulation of intestinal transport (Fig. 2) is reminiscent of the concept of adaptive regulation of enzyme synthesis.[16] Many bacterial enzymes, and a few enzymes of eukaryotic species, are regulated. In some cases substrates are found to induce their enzymes, whereas in other instances substrates repress their enzymes. For the intestine, Fig. 2 showed that oral glucose and proline can induce their own transporters. Does the same pattern apply to all other regulated intestinal nutrient transporters?

Figure 4 shows that the answer is no. Some substrates induce their intestinal transporter, others repress, and for still others the transporter activity decreases with decreasing dietary substrate level through a minimum or else to an asymptote.[3]

At an early stage in our studies of regulation of intestinal transport, before most of the results depicted in Fig. 4 were available, William Karasov and I used *a priori* reasoning to predict the pattern of regulation from three considerations.[17] First, we suggested that nutrients used for calories would tend to induce their transporters because the resulting payoff in available calories at high substrate levels would more than offset the biosynthetic costs of the transporter, whereas transporter costs would not be recouped at low substrate levels. Second, substrates that are nutritionally essential should tend to repress their transporters because the transporter is most needed to extract the minimum required daily amount of nutrient from a deficient diet. Conversely, the transporter might be safely repressed at high dietary substrate levels, as the same minimum needed amount could be provided by fewer transporter copies oper-

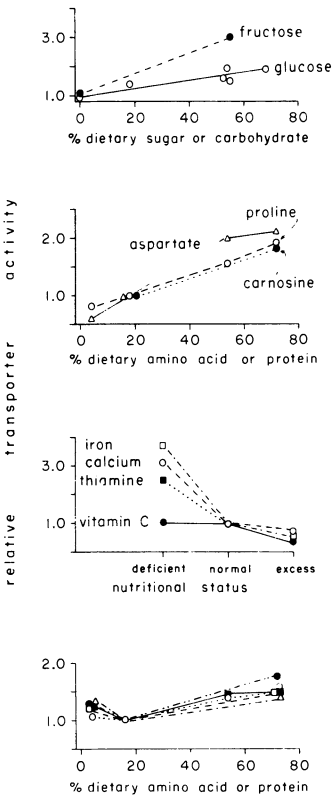


Figure 4. Summary of how activities of intestinal brush border nutrient transporters are regulated by dietary or body substrate levels. All panels except the second from the bottom show maximal transport rates in proximal and midintestine of mice that have consumed a diet containing the indicated substrate level for a period sufficient to allow complete adaptation (several days up to 2 weeks). Top panel: from Refs. 5 and 14. Second panel from top: patterns for the amino acids aspartate and proline, and the dipeptide carnosine, from Refs. 7 and 23. Second panel from bottom: iron brush border receptors in guinea pig proximal intestine,[18] calcium uptake in chick duodenum,[19] thiamine uptake in rat jejunum,[22] and vitamin C uptake in guinea pig ileum.[21] Bottom panel: patterns for the amino acids histidine (solid circles), lysine (open circles), leucine (solid squares), alanine (open squares), and methionine (open triangles). Ordinate values are uptakes at the dietary level indicated on the abscissa normalized to uptake in animals on a fructose- or carbohydrate-free ration (top panel), in mice on a 18% casein or 18% amino acid ration (second and bottom panels), or in animals on a ration with the indicated nutrient at adequate levels (third panel). (From Ref. 3.)

ating at a higher substrate concentration or even by passive diffusion. Finally, toxic substrates should tend to repress their transporters, so as to protect the animal against the risk of toxicity from absorption of too much substrate in the diet.

For the glucose transporter, all three considerations predict that glucose should induce its transporter, as is actually observed (Fig. 2, top panel of Fig. 4). Since glucose provides calories, high dietary carbohydrate levels should induce glucose transport; since glucose is nontoxic, this induction is safe; and since glucose is nonessential, the glucose transporter can safely be repressed on a carbohydrate-free diet. Thus, all three considerations favor induction. Further, other nutrients that resemble glucose in being calorigenic, nonessential, and nontoxic—namely, fructose, aspartic acid, proline, and possibly dipeptides—also induce their corresponding transporter (second panel of Fig. 4).

This reasoning also applies to iron, which is essential, yields no calories, and is toxic in large amounts. Because iron is essential, the iron transporter must have maximal activity on an iron-deficient diet; because iron is toxic, the

transporter should be repressed on a high-iron diet; and because iron yields no calories, repression on a high-iron diet costs the animal nothing and saves the cost of unnecessary transporter copies. Thus, all three considerations dictate repression as actually observed.[18] In addition, four other nutrients that resemble iron in being essential, toxic, and noncalorigenic—namely, calcium,[19] phosphate,[20] vitamin C,[21] and thiamine[22]—also resemble iron in repressing their own transporter (third panel of Fig. 4). The repression of the two vitamin transporters by their substrates is interesting because it shows that dietary megadoses of vitamins have two opposing effects on vitamin absorption: to increase the amount absorbed through increasing the luminal vitamin concentration, but to decrease the efficiency of absorption by repressing the transporter. Conceivably, vitamin megadoses might even have the self-defeating effect of reducing net vitamin absorption by the intestine.

For essential amino acids, however, dietary influences on their respective transporters are more complex, as the three considerations conflict with one another. Since essential amino acids can be used for calories, they should induce their transporters. Since essential amino acids are individually essential as specific nutrients and as a source of nitrogen, they should tend to repress their transporters. Also, since essential amino acids in large amounts tend to be much more toxic than nonessential amino acids, they should repress their transporters. The bottom panel of Fig. 4 shows that the body has managed to strike a compromise among these conflicting considerations.[23] It would be suicidal to repress essential amino acid transport on a low-protein diet, and in fact the transporter for essential neutral amino acids and the transporter for essential basic amino acids both maintain asymptotically constant activity or are even induced on a low-protein diet. On a high-protein diet, the activity of both transporters is again induced, but not nearly to the same degree as that observed for the transporters of the relatively nontoxic, nonessential amino acids aspartate and proline.

In short, the intestine exhibits diverse regulatory patterns for nutrients, but some adaptive sense can be made of this diversity.

4. SIGNALS FOR THE SPECIFIC REGULATION OF BRUSH BORDER TRANSPORTERS

How does a particular transporter learn that more of its substrate has arrived at the intestine via the diet and that it should therefore up-regulate or down-regulate itself? The simplest signaling mechanism would be for the enterocyte itself to directly perceive the concentration of its substrate in the intestinal lumen. The preferred substrate may only be an ultimate signal that releases intermediate signals; the best ultimate signal may not always be the

preferred substrate; and the ultimate signal, whatever it is, may be perceived at a location other than the intestinal lumen.

At least the identity of the ultimate signal is simple in the case of the fructose transporter and the aspartate transporter, which are induced most effectively by fructose[5] and aspartate,[24] respectively. However, that does not indicate whether fructose and aspartate levels are perceived in the intestinal lumen, within the enterocyte after the brush border transport step, or in the bloodstream after the basolateral transport step as well.

More disconcerting is the regulatory signaling of the glucose transporter: dietary fructose is as good an ultimate inducer as dietary glucose, although fructose cannot serve as a substrate of the glucose transporter.[5] Amino acids and peptides are even more complex. The basic amino acid transporter is most effectively induced by an acidic amino acid.[24] The transporter for proline is not particularly induced by proline but rather is induced by some as-yet-unidentified component of protein.[24] The dipeptide transporter is not specifically induced by peptides; free amino acids are just as effective.[7] These signaling puzzles raise the possibility that the familiar distinction between essential and nonessential amino acids involves not only the question of which amino acids the body is unable to synthesize itself (and thus requires in the diet), but also which amino acids are good at turning on transporters for other amino acids, even though the inducing amino acid itself is not a substrate.

All these statements about signals are based on substrates added to the diet. In addition, glucose or fructose infused into the bloodstream also induces glucose or fructose transport. Whether or not this effect involves an induction only of basolateral transporters remains uncertain.[13,25,26]

Finally, the possibility exists that some nutrients may regulate transport via hormonal intermediates. For example, glucose infusion in the distal intestine results in up-regulation of glucose transport in the proximal intestine, suggesting humoral mediation.[27] The possibility of a humoral mediator is also suggested by the up-regulation of sugar and amino acid transport in hen colon by a high-salt diet.[28]

5. NONSPECIFIC REGULATION

The previous sections have concerned regulation of individual transporters, proceeding independently of regulation of other transporters. In addition, there are situations in which the body needs to process more or less of many nutrients in parallel. For example, there is increased need for many different nutrients during pregnancy and lactation, at low environmental temperatures, and under conditions of chronic exercise, all of which are associated with hy-

perphagia. Conversely, starvation is associated with decreased inputs of all nutrients in parallel.

It has in fact been found that pregnancy, lactation, and other physiological conditions resulting in hyperphagia also increase uptake of multiple nutrients.[17] Unlike the specific and quite rapid up-regulation of transporter by dietary levels of its substrate, the generalized up-regulation associated with hyperphagic conditions is slow, requiring a week or more to develop even in small rodents. Conversely, starvation is associated with decreased uptake of nutrients when calculated per centimeter length of intestine.[14] Starvation is not associated with changes in uptake per milligram of tissue wight.[14,17] The weight of tissue per cm length of intestine varies with hyperphagia or starvation, increasing due to hyperplasia in pregnancy and other hyperphagic conditions, but decreasing in starvation. These changes in quantity of absorbing mucosa are presumably what cause increased or decreased transport of many nutrients in hyperphagia or starvation. The lack of effect on transport rates calculated per mg suggests no specific effects of these conditions on particular transporters, apart from producing more or less mucosa and hence more or fewer transporter copies.

It is generally assumed that the intestinal hyperplasia during pregnancy, lactation, and other conditions is mediated by trophic hormones, although their identity remains uncertain.

Parallel, nonspecific regulation of intestinal transporters appears in several clinically important conditions. Intestinal mass and hence transporter activities decrease in patients and animals maintained on total parenteral nutrition.[29] In the short-gut syndrome, intestinal resection results in hypertrophy and hence increased transporter activity of the remnant gut.[30] Experimental diabetes in animals leads within a few days to a specific up-regulation of glucose transport, followed by a slower up-regulation (probably mediated by hyperplasia) of other transporters.[31]

6. ONTOGENETIC DEVELOPMENT OF TRANSPORTERS

Humans and other mammals undergo two inevitable and dramatic shifts in intestinal function. The first shift occurs at birth, when the intestine suddenly takes over from the placenta the responsibility for supplying nutrients. The second shift occurs at the time of weaning, when the intestine switches from an intake of milk to solids. Does the intestine adapt to these dramatic shifts, or do we already as fetuses develop an "all-purpose intestine" that is equally suited to the fetal period, nursing period, and adulthood? The quantity of intestinal mucosal surely increases from birth until adulthood; are there also any qualitative changes in mucosal properties? If ontogenetic development is ac-

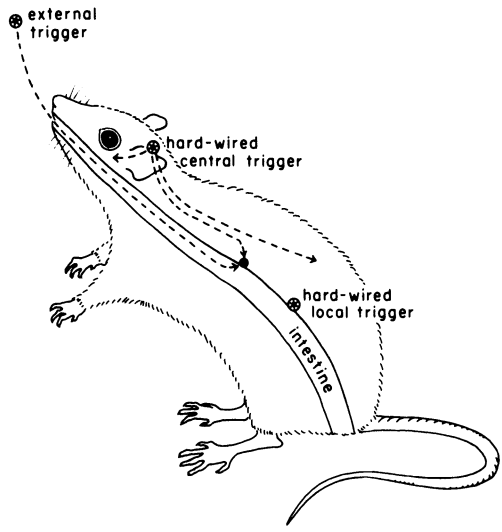


Figure 5. Developmental changes in rat intestine may be triggered by an external signal (e.g., weaning or a change in diet), by a hard-wired central trigger (e.g., a pituitary, thyroid, or adrenal hormone), or by a local trigger (e.g., a timing mechanism within the enterocytes themselves). (From Ref. 36.)

companied by adaptive responses of the intestine, are those responses on an "open program," i.e., dependent on external signals such as the act of birth or of weaning? Alternatively, might the developmental responses be independent of external signals and instead be 'hard-wired" through an internal timing mechanism within the body? Is the timer a local one in the intestine itself, or is it a central timer operating through hormones or nerves (Fig. 5)?

Much is known about the development of other gut functions in infant mammals.[32,33] In particular, the decline of intestinal lactase and rise of sucrase toward the time of weaning have been much studied, as have the effects of corticosteroid and thyroid hormones on these changes. However, comparatively little is known about the development of intestinal nutrient transporters.

How do the nutrient transporters keep pace with the growing needs of the growing animal? One can calculate the total absorptive capacity of the intestine for a particular nutrient by measuring uptake per unit intestinal length at various positions along the intestine and then integrating over its entire length. As illustrated in Fig. 6, the intestine's uptake capacity for proline or glucose in young rabbits increases with age, largely because the intestine itself grows. Unknown is whether the young intestine always has a comfortable excess of uptake capacity over physiological needs, or whether uptake capacity may in some instances impose a bottleneck on the animal's ability to grow. Infant rats

that are artificially fed quantities of milk in excess of normal do not show increased growth rates from 4 to 10 days of life, raising the possibility that quantity of food alone is not the determinant of growth and that some bottleneck exists.

Figure 6 deals with the quantative development of the transporters. In addition, how do the transporters adjust qualitatively to changes in diet? For example, in rats the weaning-associated transition from milk to solid food is accompanied by a shift from fat and protein to carbohydrate, and from a simple sugar (lactose) to complex carbohydrates. One might expect, although we have little information as yet, that weaning will be accompanied by the up-regulation of some transporters and the down-regulation of others. If so, one could then inquire whether those changes were triggered by weaning itself or by a hard-

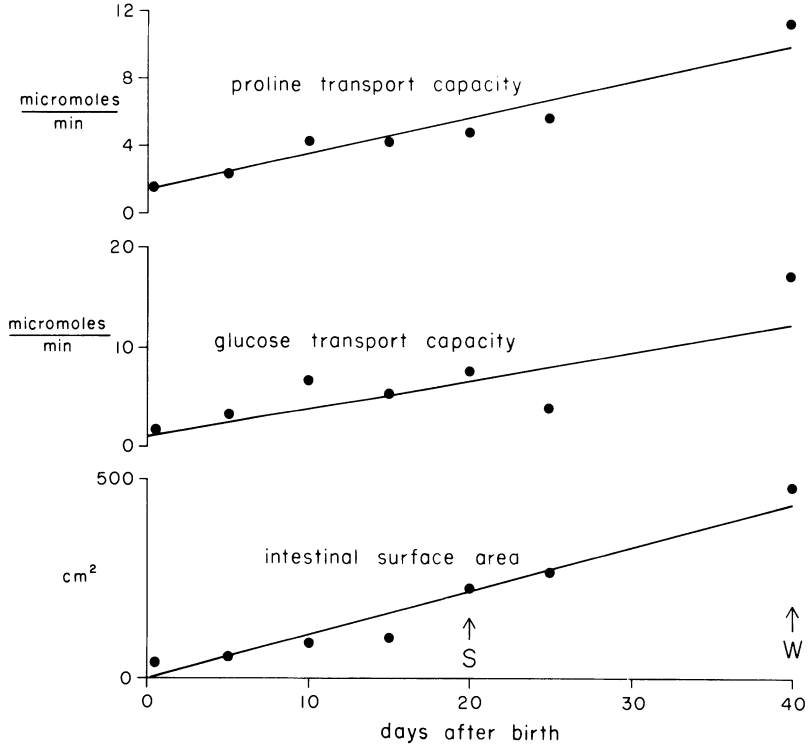


Figure 6. Developmental changes in rabbits, in intestinal surface area and in the whole intestine's uptake capacity for L-proline and D-glucose. S = first day of ingesting solid food, W = day of completed weaning (last day of ingesting milk). Note that intestinal surface area and both uptake capacities increase developmentally, as do weight and metabolic rate. The margin by which the actual uptake capacities exceed those required minimally to meet the animal's metabolic and growth requirements is unknown. (Unpublished data of R. K. Buddington.)

wired timer. To address that question, one could study animals in which weaning had been delayed or else advanced precociously, but it would be preferable if one could study an animal that could be maintained on a diet of unvarying composition as it grew. While this experiment is difficult to perform in mammals, it can readily be performed in some species of fish. Figure 7 illustrates results from a fish called the monkey-faced prickleback, which is a carnivore when young and becomes a herbivore when older. Since a carnivore takes in much less dietary carbohydrate than a herbivore, one might expect small pricklebacks to have lower rates of glucose transport (relative to proline transport) than large pricklebacks, if only because the higher carbohydrate levels in the diet of the large animals would induce the glucose transporter as already illustrated in Figs. 2 and 4. In the experiment depicted in Fig. 7, the immediate effect of diet as an external signal is excluded by maintaining small and large pricklebacks on the same diet. Even then, the ratio of proline transport to glucose transport decreased as the pricklebacks grew larger (Fig.7), showing that the shift in intestinal transporters takes place in the complete absence of dietary changes as an external signal.[34] That is, the development of nutrient transporters in the intestine must be an endogenous event, triggered by an internal timer. In this case we still do not know whether the hard-wired timer is central or local. However, a recent elegant experiment by Yeh and Holt[35] proves the

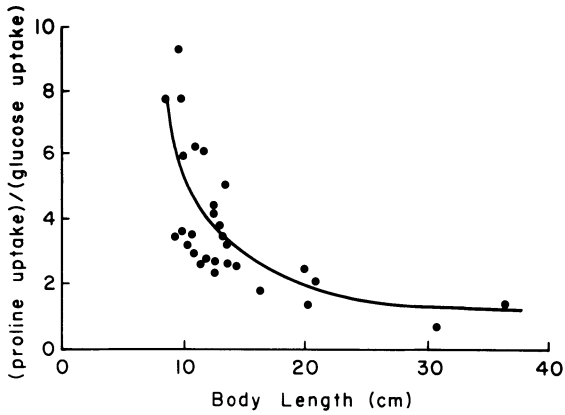


Figure 7. Hard-wired developmental changes in the ratio of the whole intestine's uptake capacity for L-protein to that for D-glucose, in monkey-faced pricklebacks. This fish species switches from carnivorous when small to herbivorous when large. Hence, the normal dietary ratio of protein to carbohydrate decreases with body length, and so does the ratio of intestinal proline to glucose uptake, even though the fish used for these experiments were being maintained in our laboratory on a constant diet. Thus, the developmental change by which the intestine is enabled to adapt to the expected change in the animal's natural diet is independent of the dietary change as an immediate signal. (From Ref. 34.)

existence of a local timer for development of intestinal sucrase. By transplanting isografts of intestine from donor rat pups of one age into host rat pups of a different age, Yeh and Holt showed that the grafted tissue expressed sucrase at an age appropriate to the donor, and many days before expression in the host's intestine. Thus, sucrase expression must depend on a local hard-wired timer within the piece of intestine expressing the sucrase. Similar experiments on development of intestinal nutrient transport await the future.

7. SUMMARY

In this chapter we have seen that intestinal nutrient transporters are subject to regulation. Virtually every transporter examined so far has proved to exhibit specific regulation of its activity by dietary levels of its substrate. A dietary substrate may either induce or repress its transporter, or else transporter activity may go to a minimum or reach an asymptote. These diverse patterns make adaptive sense if one considers the substrate's utility as a caloric source and whether it is essential or toxic. In some cases the most effective signal for regulation is the transported substrate itself; curiously, for other transporters this is not the case. Hormones may serve as the proximate signals for regulation in other instances. In addition to specific regulation of transporters, there is also nonspecific regulation of many transporters in parallel—up-regulation in pregnancy and lactation and other conditions associated with hyperphagia, down-regulation during starvation. Finally, there are ontogenetic changes in transporters in the growing animal, both in regard to the number of transporter copies and in regard to the ratio of activities of different transporters.

REFERENCES

1. Weser E. Nutritional aspects of malabsorption. *Am J Med* 1978;67:1014–1020.
2. Karasov WH, Diamond JM. A simple method for measuring intestinal solute uptake *in vitro*. *J Comp Physiol B* 1983;152:105–116.
3. Karasov WH, Diamond JM. Adaptive regulation of intestinal nutrient transporters for nutrients. *Proc Natl Acad Sci USA* 1987;84:2242–2245.
4. Ferraris RP, Diamond JM. A method for measuring apical glucose transport site density in intact intestinal mucosa by means of phlorizin binding. *J Membr Biol* 1986;94:65–75.
5. Solberg DH, Diamond JM. Comparison of different dietary sugars as inducers of intestinal sugar transport. *Am J Physiol* 1987;252:G574–578.
6. Karasov WH, Solberg D, Carter S et al. Uptake pathways for amino acids in mouse intestine. *Am J Physiol* 1986;251:G501–508.
7. Ferraris RP, Diamond JM, Kwan W. Dietary regulation of intestinal transport of the dipeptide carnosine. *Am J Physiol* 1988;255:G143–150.
8. Rose RC. Water-soluble vitamin absorption in intestine. *Ann Rev Physiol* 1980;42:157–171.

9. Rose RC. Intestinal absorption of water soluble vitamins. In: *Physiology of the Digestive Tract,* 2nd ed. (Johnson LR, ed). New York: Raven Press, pp. 1581–15965, 1987.
10. Wright EM, Harms V, Mircheff AK, van Os CH. Transport properties of intestinal basolateral membranes. *Ann NY Acad Sci* 1981;372:626–636.
11. Stevens BR, Kaunitz JD, Wright EM. Intestinal transport of amino acids and sugars: advances using membrane vesicles. *Ann Rev Physiol* 1984;46:417–433.
12. Karasov WH, Pond RS, Solberg DH, Diamond JM. Regulation of proline and glucose transport in mouse intestine by dietary substrate levels. *Proc Natl Acad Sci USA* 1983;80:7674–7677.
13. Cheeseman CI, Maenz D. Induction of glucose transporter by hyperglycemia in rat enterocyte basolateral membranes. *J Physiol* 1985:369:152P.
14. Diamond JM, Karasov WH. Effect of dietary carbohydrate on monosaccharide uptake by mouse small intestine in vitro. *J Physiol* 1984;349:419–440.
15. Ferraris RP, Diamond JM. Use of phlorizin binding to demonstrate induction of intestinal glucose transporters. *J Membr Biol* 1986;94:77–82.
16. Walker R. *The Molecular Biology of Enzyme Synthesis.* New York: Wiley, 1983.
17. Karasov WH, Diamond JM. Adaptive regulation of sugar and amino acid transport by vertebrate intestine. *Am J Physiol* 1983;245:G443–G462. 1983.
18. Kimber CL, Mukherjee T, Deller DJ. In vitro attachment to the intestinal brush border effect of iron stores and other environmental factors. *Am J Dig Dis* 1973;18:781–791.
19. Morrissey RL, Wasserman RH. Calcium absorption and calcium-binding protein in chicks on differing calcium and phosphorus intakes. *Am J Physiol* 1971;220–1509–1515.
20. Lee DBN, Walling MW, Brautbar N. Intestinal phosphate absorption: Influence of vitamin D and non-vitamin D factors. *Am J Physiol* 1986;250:G369–G373.
21. Rose RC, Nahrwold DL. Intestinal ascorbic acid transport following diets of high or low ascorbic acid content. *Int J Vit Nutr Res* 1978;48:382–386.
22. Patrini C, Cusaro G, Ferrari G, Rindi G. Thiamin transport by rat small intestine in vitro: Influence of endogenous thiamin content of jejunal tissue. *Acta Vitaminol Enzymol* 1981;3 m.s.:17–26.
23. Karasov WH, Solberg DH, Diamond JM. Dependence of intestinal amino acid uptake on dietary protein or amino acid levels. *Am J Physiol* 1987;252:G614–625.
24. Stein ED, Chang, SD, Diamond JM. Comparison of different dietary amino acids as inducers of intestinal amino acid transport. *Am J Physiol* 1987;252:G626–635.
25. Csaky TZ, Fischer E. Intestinal sugar transport in experimental diabetes. *Diabetes* 1981;30:568–574.
26. Debnam ES, Karasov WH, Thompson CS. Hyperglycemia and glucose transport across the rat small intestine. An in vitro and in vivo study. *J Physiol* 1986;376:36P.
27. Debnam ES. Evidence that the presence of sugars in the lower ileum can influence glucose absorption from the upper small intestine. *J Physiol* 1981;324:545P.
28. Lind J, Munck BG, Olsen O. Effects of dietary intake of NaCl on sugar and amino acid transport across isolated hen colon. *J Physiol* 1980;3005:327–336.
29. Weser E, Vandeventer A, Tawil T. Stimulation of small bowel mucosal growth by midgut infusion of different sugars in rats maintained by total parenteral nutrition. *J Ped Gastroenterol Nutr* 1982;1:411–416.
30. Urban E, Weser E. Intestinal adaptation to bowl resection. *Adv Int Med* 1980;26:265–291.
31. Thomson ABR. Uptake of glucose into the intestine of diabetic rats. Effects of variations in the effective resistance of the unstirred water layer. *Diabetes* 1981;30:247–255.
32. Henning SJ. Postnatal development: Coordination of feeding, digestion, and metabolism. *Am J Physiol* 1981;241:G199–G214.

33. Henning SJ. Ontogeny of enzymes in the small intestine. *Ann Ref Physiol* 1985;47:231–245.
34. Buddington RK, Chen J, Diamond JM. Genetic and phenotypic adaptation of intestinal nutrient transport to diet. *J Physiol* 1987;393:261–281.
35. Yeh K-Y, Holt PR. Ontogenic timing mechanism initiates the expression of rat intestinal sucrase activity. *Gastroenterology* 1986;90:520–526.
36. Diamond JM. Hard-wired local triggering of intestinal enzyme expression. *Nature* 1986;324:408.

13

Celiac Disease
Pathogenesis and Clinical Features

Martin F. Kagnoff

1. INTRODUCTION

Celiac disease (gluten-sensitive enteropathy, celiac sprue, or nontropical sprue) is characterized by damage to the small intestine mucosa and the malabsorption of most nutrients. Symptoms most commonly appear during the first 3 years of life after the introduction of cereals into the diet, with a second peak incidence occurring during the third decade.[1] Clinical manifestations predominantly reflect the consequences of malabsorption. Although celiac disease was noted in earlier centuries,[2,3] a striking decrease in the incidence of celiac disease was observed in Holland during the wheat-deprived years of World War II; this suggested an association between celiac disease and the ingestion of wheat-containing products.[4,5]

2. GRAIN PROTEINS

Celiac disease is activated when a genetically susceptible host ingests food products that contain wheat, rye, barley, or oats. It is the gliadin fraction of wheat gluten and similar alcohol-soluble proteins in the other grains (termed *prolamins)* that are associated with the development of intestinal damage.

Martin F. Kagnoff • Laboratory of Mucosal Immunology, Department of Medicine, University of California–San Diego, La Jolla, California 92093. Supported by the National Institutes of Health grant DK 35108.

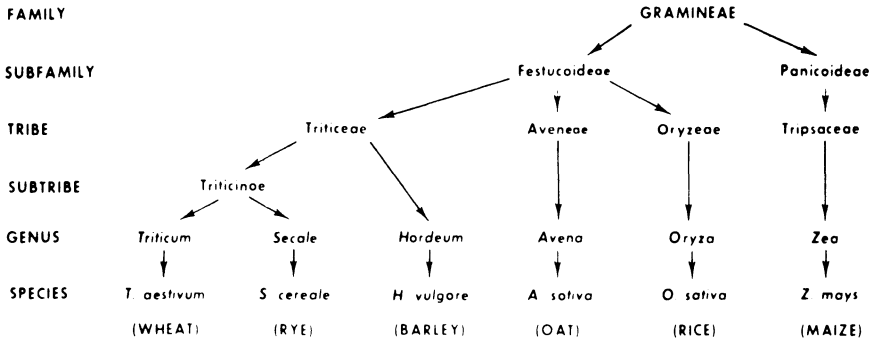


Figure 1. Taxonomic relationships of major cereal grains. (From Kasarda *et al.*, in *Perspectives in Celiac Disease*, B. McNicholl, C. F. McCarthy, and P. F. Fottrell, eds., University Park Press, Baltimore, 1978.)

Cereal grains belong to the grass family (Gramineae). It is interesting that grains other than wheat that activate celiac disease (e.g., rye and barley) bear a close taxonomic relationship to wheat (Fig. 1). Oats, which in large quantities are thought to activate disease, are further removed from wheat, rye, and barley. Grains that do not activate disease (e.g., rice and maize) are still further separated from wheat in terms of their derivation from the primitive grasses.[6–8]

Common bread wheats, developed relatively recently in evolution (i.e., the last 10,000 years), have a hexaploid genome (i.e., genome content AABBCC). Durum wheat, commonly used in the production of pasta, is tetraploid (i.e., genome content AABB), whereas rye and barley are diploid.[6] Gene clusters that code for the gliadins are present on chromosomes of homologous groups 1 and 6.[9] Because gliadins are coded for on more than one chromosome,[10–12] the breeding of wheat varieties devoid of disease-activating properties has not been practical.[13]

Gluten is a major component of the wheat endosperm and serves as a source of nitrogen for the germinating wheat embryo.[6,7] Its elastic properties are important in the production of bread. Gliadins and glutenins (mostly the low molecular weight glutenins) are the major protein components of gluten, but only the gliadins have clearly been demonstrated to activate celiac disease. Minor constituents that contaminate gluten extracts (e.g., wheat albumins, globulins, membrane proteins, lipids, and carbohydrates) also do not appear important in disease activation. In rye, barley, and oats, the alcohol-soluble proteins associated with the activation of disease are termed secalins, hordeins, and avenins, respectively.

Gliadins are single polypeptide chains that range in molecular weight from 30,000 to 75,000. They have a low charge[7] and a high glutamine and proline content (32–56 glutamine and 15–30 proline residues per 100 amino acid res-

idues). Gliadin from a single variety of wheat can be shown to contain 40 or more different but closely related components[7] when examined by two-dimensional gel electrophoresis. By gel electrophoresis, gliadins can be categorized into four major electrophoretic fractions: α-gliadins, β-gliadins, γ-gliadins, and ω-gliadins.[7,14] Each fraction, in turn, contains several subcomponents (e.g., β_1, β_2, β_3-gliadins; γ_1, γ_2, γ_3-gliadins; ω_1, ω_2, ω_3, and ω_4-gliadins).[79] Gliadins of the α, β, and γ fraction share a similar amino acid composition and NH_2-terminal sequence (i.e., α-type sequence).[6,7,15,16] γ_2, γ_3, and ω-Gliadins differ markedly from the α-type sequence in their NH_2-terminal sequence and amino acid composition (i.e., γ-type sequence).[17–20]

When α-gliadins from suitable wheat varieties are ultracentrifuged, a precipitate of aggregatable α-gliadins, termed A-gliadin, is formed.[16,21] This major α-gliadin fraction is known to activate disease.[6,22,23] Recently, the complete primary amino acid sequence of A-gliadin was determined from amino acid sequencing[24]; other α-gliadin sequences have been deduced from sequencing of cDNA clones.[17,24,25] Such information ultimately should allow the resolution of which amino acid sequences in gliadin are responsible for disease activation.

The issue of which wheat gliadin fractions are capable of activating disease is controversial. Earlier reports suggested that only α-gliadins activate celiac disease.[26] Later studies suggested that β- and perhaps γ-, but not ω-gliadins, also might activate disease.[27] Because ω-gliadins have the highest content of glutamine and proline, the high content of those amino acids alone is not thought to be the determining factor in disease activation. However, recent studies suggested that all gluten fractions might activate disease.[28–30] Complete hydrolysis of gliadin destroys its ability to activate celiac disease.

Controversy over which gliadin fractions activate celiac disease may stem in part from (1) heterogeneity in sensitivity to different gliadin fractions among patients, (2) differences in the timing of small intestinal mucosal biopsies after *in vivo* gliadin challenge, (3) the use of different clinical and diagnostic end points to assess mucosal damage after *in vivo* gliadin challenge, (4) using impure gliadin preparations in *in vivo* and *in vitro* challenge studies, or (5) an erroneous assumption that abnormalities in *in vitro* assays equate to disease activation.

3. EPIDEMIOLOGY

Celiac disease is common in Ireland and northern Europe,[32–41] but has been reported in diverse geographic regions (e.g., India, West Pakistan, the Middle East, and Cuba).[42–45] The reported prevalence of celiac disease in specific geographic areas increased after 1960, paralleling the widespread use of the small intestinal mucosal biopsy in diagnosis in general clinical practice.[46,47]

Accurate epidemiological data on celiac disease are difficult to obtain because asymptomatic disease makes ascertainment of true prevalence rates problematic. Nonetheless, the high incidence of Celiac disease in western Ireland appears to have decreased substantially over the past several years.

4. GENETICS

4.1. Families and Twins

The prevalence of celiac disease among asymptomatic first-degree relatives of celiac disease patients may not be as great as the 10–20% initially reported.[48,49] Thus, when all 100 first-degree relatives of 32 Swedish patients were studied, only two individuals had convincing celiac disease.[50] Others have reported celiac disease in less than 5% of family members.[51,52]

At least twenty-four pairs of identical twins with celiac disease have been reported,[53–56] of which 18 (75%) were concordant for disease. Not all had monozygosity unequivocally proven,[56] and some twin pairs have not had sufficient long-term follow-up to be certain that disease would not develop at a later age. Nonetheless, it appears that there are cases of discordance for celiac disease among monozygotic twins. Assuming that twins ingest similar dietary grains, such discordance suggests the possible importance of additional environmental factors in disease expression. Alternatively, because of the genetic mechanisms that generate diversity, monozygotic twins may not be identical in terms of their repertoire of antibodies and T-cell receptors for antigen.

4.2 HLA and Non-HLA Markers

Celiac disease has a striking association with an extended HLA haplotype (Fig. 2), characterized telomerically by the class I HLA B-8 serological market and, more centromerically, the class III SCO1 complotype and the class II HLA-DR3 and HLA-DQw2 serological markers.[57–60] Recent studies using HLA class II gene probes have indicated that this haplotype extends even more centromerically into the HLA class II DP subregion where it is marked by DPα-chain[61] and β-chain genomic restriction fragment length polymorphisms (RFLPs).[62–64] The strongest association of celiac disease is with the HLA class II D-region markers. The class I and class III region gene associations are thought to be secondary to linkage desequilibrium between specific markers (alleles) in those regions and in the HLA class II D region. In several populations,[55,65,66] celiac disease is reported to be associated with a heterozygous phenotype that includes the HLA-DR7 serological marker, or a haplotype containing HLA-DR7, HLA-B44, and the SC 31 complotype[60] on one chromosome, and the DR3/DQw2 haplotype on the other chromosome.

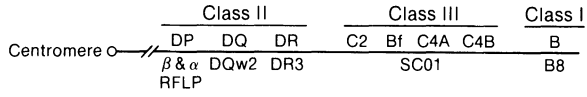


Figure 2. Celiac disease HLA haplotype on chromosome 6.

Among sibling pairs with celiac disease, most (~96%) share one or both HLA haplotypes.[67] However, only 40, 28, and 0–5% of siblings who share two, one, or no HLA haplotypes, respectively, with a propositus develop celiac disease.[46] Differences in the concordance rates for celiac disease between monozygotic twins (i.e., 75%) and HLA-identical sibs suggest that genes outside the HLA locus also confer susceptibility to celiac disease. The observation that celiac disease occurs more frequently in family members who share the HLA susceptibility haplotype than in nonfamily members with the same apparent haplotype further supports this notion.[68] Studies reporting HLA identity in the past depended largely on serological testing for markers at one or more HLA loci and did not document HLA identity across the entire HLA class II D region. It is now recognized that some siblings assumed to be HLA-identical by such serological testing may not be identical according to more detailed studies, which include a direct examination of HLA proteins on the cell surface and HLA genes at the DNA level.

Genes coding for immunoglobulin heavy-chain allotype markers on human IgG heavy chain have been proposed to mark for a second genetic region that may be important in the pathogenesis of celiac disease.[69–71] Such allotype markers, termed Gm markers, reflect inherited differences between individuals in the amino acid composition of the chromosome 14 encoded IgG heavy-chain constant regions. The G2m(n) allotype marker, a marker on IgG2, has been associated with the persistence of antigliadin antibody in celiac disease patients maintained on a gluten-free diet.[71] Other studies have suggested that a particular Gm phenotype, Gm(f;n;b), may be a predisposing factor to celiac disease among some individuals.[72]

5. OTHER ENVIRONMENTAL FACTORS

Breastfeeding has been reported to delay the age of onset of symptoms and to decrease the risk of ultimately developing celiac disease.[73,74] In addition, environmental factors other than dietary grains may be important in the pathogenesis of celiac disease. A-gliadin has a region of amino acid sequence homology with the E1b protein of human adenovirus serotype 12 (Ad12), an adenovirus usually isolated from the human intestinal tract.[75] The region of

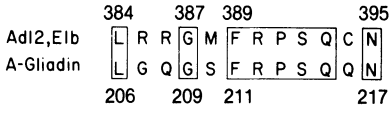


Figure 3. Amino acid sequence of the adenovirus 12 Elb protein (Ad12, Elb) and A-gliadin beginning at amino acid residues 384 and 206, respectively. The region of homology includes 8 of 12 residue identities, including an identical pentapeptide. The single-letter code for amino acids is used. (From Kagnoff *et al., J. Exp. Med.* 1984; 160:1544).

homology spans 12 amino acids and includes 8 of 12 amino acid identities and an identical pentapeptide sequence (Fig. 3). Recent studies reported that 89% (16/18) of untreated celiac disease patients, compared to 17% (16/35) of disease controls, had evidence of prior Ad12 infection (Fig. 4).[76] There was no increased prevalence of infection with another closely related adenovirus (Ad18) or a different enteric virus (Echo 11). The region of A-gliadin that shares amino acid sequence homology with the Ad12 E1b protein appears to be recognized as an antigenic determinant by many celiac disease subjects, but not by controls, in antibody[76] and cell-mediated immune assays.[76,77] These data suggest that a viral protein may play a role in the pathogenesis of celiac disease, perhaps by virtue of immunological cross-reactivity between antigenic determinants shared by the viral protein and α-gliadins.[76]

6. HUMORAL AND CELL-MEDIATED IMMUNITY

6.1. Humoral Immunity

Immune responsiveness is determined by genes that code for three distinct classes of cell surface proteins: HLA antigens encoded by genes of the major histocompatibility complex (MHC), antigen receptors on T cells, and antigen

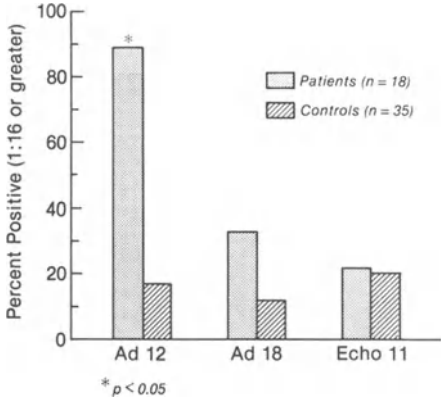


Figure 4. Virus-neutralizing antibody in untreated adult celiac disease patients and controls. n = Number of subjects in each group. Positive subjects defined as titers of 1:16 or greater. * Indicates value significantly different from control group.

receptors (i.e., immunoglobulin molecules) on B cells.[78] Important insights into the regulation of immune responses to wheat gliadins have demonstrated that the antibody response to A-gliadin is governed by genes that map to the murine MHC (termed H-2) and the immunoglobulin heavy-chain region. These studies also indicated that limited regions of the A-gliadin molecules are important in the activation of T-helper cell responses.[70,79,80]

Serum antibodies to whole gliadin and its major electrophoretic fractions can be detected in most celiac disease patients with active[81,82] disease and in many with inactive[83] disease. Furthermore, IgG antigliadin antibody can persist for long periods (up to 20 years) in clinically asymptomatic patients maintained on a gluten-free diet.[71] Such antibody may be directed against one or as many as all four (i.e, α, β, γ, ω) of the major electrophoretic fractions of gliadin.[83]

Antigliadin antibody could play a role in disease pathogenesis by one of several routes. For example, antigliadin antibody complexed with gliadin to form immune complexes could activate tissue-damaging effector mechanisms, including the complement cascade. Alternatively, antigliadin antibody could cause intestinal injury via a cell-mediated cytotoxic reaction in which antigliadin antibody recognizes gliadin peptides bound to mucosal structures and directs a "K"-cell-mediated, antibody-dependent, cytotoxic reaction. In support of the above possibilities, cells producing IgG antibody capable of mediating such reactions are markedly increased in the lamina propria during active disease,[84–86] and organ culture studies have demonstrated increased local production of antigliadin antibody in celiac small intestinal mucosa after gliadin challenge.[85,87,88] Nonetheless, current evidence neither suggests nor refutes a major role for antigliadin antibody in the pathogenesis of celiac disease.

Many celiac disease patients have serum antibodies against other food proteins, including milk, egg, and soya proteins.[89] Antibodies to dietary proteins may simply reflect increased permeability of the intestinal mucosa in this disease. Thus, IgG antigliadin antibodies can also be found in approximately 15% of Crohn's disease patients in whom the small intestinal mucosa is disrupted.[71] However, enhanced absorption may not be a complete explanation, since antigliadin antibodies can also be found in small numbers of apparently normal individuals and in healthy (i.e., biopsy-negative) relatives of celiac disease patients. The significance of other antibodies, including antireticulin antibodies, antiendomysial antibodies, and circulating immune complexes containing a mannose-rich, 90-kDa glycoprotein, in celiac disease is totally unknown at present.[90–92]

6.2. Cell-Mediated Immunity

Cell-mediated immune mechanisms may play a role in the pathogenesis of celiac disease, although, like antibody responses, their precise importance in

this disease requires greater definition. Leukocyte migration assays using gluten components, α-gliadin or a 12 amino acid peptide of A-gliadin, indicate lymphocyte sensitization to gliadin proteins, as well as to other food proteins in celiac disease.[93,94] Other studies have demonstrated that soluble factors active in leukocyte migration inhibition assays are produced when intestinal biopsies from celiacs with active disease are cultured with α-gliadin or gluten fractions.[95–99] Polyclonal stimulation of peripheral blood lymphocytes with the mitogen phytohemagglutinin (PHA) has been reported both as normal or as altered in celiac disease.[100]

Intraepithelial lymphocytes (IEL) in celiac disease patients, as in normal individuals, predominantly have the CD8 (OKT8, Leu-2) surface antigen[101,102] usually associated with class I restricted T cells (i.e., T cells that mediate suppressor or cytotoxic functions); small intestinal biopsies from celiac disease patients that are cultured with gluten develop an increase in CD8-bearing IEL.[101] Such lymphocytes produce an array of lymphokines and could be directly responsible for epithelial damage, although this has not been formally demonstrated. The ratio of CD4 (i.e., helper/inducer phenotype) to CD8 (i.e, suppressor/cytotoxic phenotype) T cells in the circulation and intestinal mucosa has not differed between celiac disease patients and controls,[103,104] although some evidence suggests impaired suppressor cell function in celiac disease.[104,105]

During active celiac disease, there is an increased expression and altered distribution of HLA class II DR molecules on small intestinal epithelial cells.[106–108] Whether this is secondary to increased lymphokine production (i.e., γ-interferon) by mucosal T cells and whether intestinal epithelial cells have an important role in celiac disease by presenting antigen (e.g., gliadin) to mucosal lymphocytes is not known. Finally, it is worth noting the similarity between the small intestinal mucosal villous atrophy and crypt hyperplasia in mice undergoing a graft-versus-host reaction[109] or infection with the parasite *Trichinella spiralis*[110] and celiac disease. Such findings suggest that immune mechanisms in those diseases can lead to a pattern of intestinal tissue injury similar to that observed in celiac disease, although those murine experimental lesions lack the epithelial cell abnormalities characteristic of human celiac disease.

7. PATHOGENESIS

There are several hypotheses regarding the etiopathogenesis of celiac disease. Based on the observation that gliadin treated with a carbohydrase enzyme did not activate disease in a small number of celiac disease patients, carbohydrate side chains on the gliadins were postulated to be important in disease pathogenesis.[111] However, this does not appear to be the case, as several major α-gliadin components known to activate celiac disease[6] lack carbohydrate side

chains.[112] In addition, little or no evidence supports the hypothesis that purified gluten has lectinlike properties or that lectinlike properties of gluten are important in disease pathogenesis.[113–119] Further, hypotheses proposing abnormal intestinal peptidase activity in the small bowel of celiac disease patients[120] have not been proven. The small intestinal mucosa contains multiple peptide hydrolases with overlapping substrate specificity, and any biochemical abnormalities usually return to normal when patients are treated with a gluten-free diet. Finally, the possibility that a primary defect in intestinal mucosal permeability is responsible for celiac disease[121,122] seems unlikely in that abnormal mucosal permeability appears to normalize on a gluten-free diet.[123]

The importance of genetic factors in the pathogenesis of celiac disease is clear based on family and twin studies and the marked association between this disease and an HLA class II D-region haplotype. The latter also suggests an immunological basis for this disease, as HLA class II genes govern HLA class II restricted T-lymphocyte responses. Although the specific host and environmental factors responsible for celiac disease are not totally known, models for the etiopathogenesis of this disease based on interactions between genetic, immunological, and environmental factors appear most likely at present.

8. PATHOLOGY

The small intestinal lesion of celiac disease is characterized by villous atrophy and crypt hyperplasia. Abnormal mucosal morphology usually is most marked in the duodenum and jejunum, but the extent of abnormality varies between individuals depending on the severity of disease. It can encompass the entire small intestine. Similar histological changes can be seen in other diseases, including tropical sprue, soy and milk protein allergy, diffuse intestinal lymphoma, giardiasis, Zollinger–Ellison syndrome, and viral gastroenteritis. In mild disease, the lesion can be subtle. Caution is needed in interpreting biopsies in children since, in contrast to normal adults, up to 30% of normal children may have mild partial villous atrophy with shorter small intestinal villi and longer crypts than those seen in adults.[124] The most consistent mucosal changes in celiac disease are abnormalities in the villous crypt ratio, increased plasma cell and lymphocyte infiltration of the lamina propria, abnormalities in the epithelial cells (which instead of being columnar become flattened or cuboidal), and increased infiltration of the epithelium with intraepithelial lymphocytes.[125–127]

Following gluten challenge, intraepithelial lymphocytes increase within hours, even in the absence of other overt histological changes.[128] Recent studies also have reported increased numbers of mucosal mast cells and intestinal enterochromaffin cells in the mucosa of celiac disease patients.[129,130] By elec-

tron microscopy, fewer microvillous intramembrane particles and abnormal tight junctions between epithelial cells are seen.[131] Such findings parallel the disaccharidase deficiency and alterations in intestinal permeability in celiac disease. Scanning electron microscopy can demonstrate ultrastructural improvement preceding that evident by light microscopy.[132]

9. CLINICAL ASPECTS

Clinical manifestations are protean and vary markedly with the age of the patient, the duration and extent of disease, and the presence of extraintestinal pathology. A disease spectrum that ranges from minor nutritional deficiencies to more striking weight loss, steatorrhea, and malnutrition can be seen.

Although celiac disease remits clinically during the teenage years, whether or not the disease disappears is not certain. Persistent hematological and morphological abnormalities [133–135] in some patients suggest that symptoms, rather than the disease, remit.

9.1. Associated Diseases

Insulin-dependent diabetes mellitus[102] (the majority are heterozygous for HLA-DR3 and HLA-DR4 haplotypes) and abnormalities in thyroid function[136] are associated with celiac disease. In addition, the majority of patients with dermatitis herpetiformis have a celiac disease-like enteropathy.[137–139] In dermatitis herpetiformis the small intestinal lesion often is patchy rather than diffuse; gluten ingestion may be required to provoke histological abnormalities. Such patients with granular type IgA deposits in the skin usually have an HLA-DR3 haplotype and appear to have an increase in heterozygosity for the HLA-DR2 marker on an HLA-DR2-associated haplotype.[140–142] A specific constellation of HLA class II genes appears to mark for a common link between celiac disease and dermatitis herpetiformis. Case reports of associated disease in other organ systems have been published. In many instances, however, further proof is required to show that the associations with celiac disease are more than simply coincidental.

9.2. Diagnosis

Mucosal biopsy is the cornerstone for diagnosis. Its importance is underscored by the high false-positive rate of diagnosis when it is not done.[143] Endoscopic duodenal biopsy usually yields adequate tissue.[144] However, three issues warrant further comment: (1) the number of small intestinal mucosal biopsies that should be routinely performed, (2) the utility of noninvasive screening tests

in diagnosis, and (3) how to distinguish celiac disease from transient gluten intolerance.

9.2.1. Small Intestinal Biopsy

A biopsy at the onset of illness that demonstrates characteristic mucosal abnormalities, a second biopsy while on a gluten-free diet that demonstrates improvement, and a third biopsy after deliberate gluten challenge help avoid misdiagnosis (i.e., European Society for Pediatric Gastroenterology diagnostic criteria).[82] Whether or not three biopsies and a supervised gluten rechallenge are necessary or advisable in all patients is a matter of debate. In children, particularly those under 2 years of age, such a detailed evaluation seems valid[145,146] as the corollary of diagnosis is commitment to a lifelong gluten-free diet, and in this age group transient gluten tolerance may occur. In adults, one biopsy at the time of initial diagnosis and a second following a gluten-free diet is warranted. However, the potential benefit of gluten rechallenge and re-biopsy does not seem sufficiently high to warrant the additional risk. In practice, patients often conduct their own rechallenge by dietary indiscretion with a return of symptoms. Controlled gluten challenge seems justifiable in adults thought to have celiac disease, who have been placed on a gluten-free diet without a prior biopsy.[147]

9.2.2. Other Screening Tests

A noninvasive screening test for celiac disease would aid in diagnosis, but the tests currently available lack sufficient validation to warrant widespread clinical application. Measurements of serum antibody to purified wheat gliadin fractions have potential use in the follow-up of celiac disease and, in conjunction with a clinical response to a gluten-free diet, might obviate the need for additional biopsies. However, studies of serum antigliadin antibody have yielded a substantial proportion of false-positive results. Similarly, studies of antireticulin antibodies[90] and IgA antiendomysial antibodies[91] have not been sufficiently validated for general clinical use.

Morphological and biochemical studies of small intestinal mucosal biopsies in organ culture using gluten-containing medium[22,147] might obviate the need to obtain sequential biopsies in some patients. However, specialized methodology is required, and the ability to extrapolate from *in vitro* organ culture to *in vivo* disease has not been validated.

Celiac mucosa is impermeable to small polar molecules (i.e., monosaccharides) but not to intermediate-sized polar molecules (i.e., disaccharides). Because of this, sugar permeability tests assessing disaccharide compared to monosaccharide absorption (e.g., cellobiose versus mannitol, or lactulose ver-

sus L-rhamnose[148,149]) have been proposed as more specific for celiac disease than the D-xylose absorption test.[150] In other studies, a [^{51}Cr]-EDTA absorption test was used to show that abnormal permeability of the small intestinal mucosa can persist in patients with celiac disease, despite clinical and apparent histological remission.[151] However, the specificity of abnormal mucosal permeability for celiac disease is unknown and, as with other oral tests, variability of gastric emptying can be problematic. Methods using molecular biological approaches appear to offer the best hope for effective noninvasive screening and diagnostic tests in celiac disease.

9.2.3. Transient Gluten Intolerance

Transient gluten intolerance occurs in early childhood and is difficult to distinguish from celiac disease.[146] Some individuals with apparent celiac disease who become asymptomatic and whose mucosa returns to normal fail to relapse on a normal diet,[152,153] even with follow-up as long as 6 years.[152] Such individuals are said to have transient gluten intolerance. Alternatively, the duration or the intensity of the dietary challenge may not have been sufficient to provoke a recurrence. One view suggests that transient gluten intolerance cannot be diagnosed unless patients are followed for many decades with repeated biopsies. It is known, for example, that patients may be symptom-free despite enteropathy. Conversely, symptoms unrelated to celiac disease (e.g., malabsorption of the carbohydrate in wheat flour[154,155]) may develop after gluten ingestion. In some individuals, changes in jejunal mucosal architecture may be related to a direct toxic effect of gluten[156] or the result of a self-limited viral gastroenteritis.

9.3. Complications

9.3.1. Malignancy

Neoplastic disease occurs in as many as 10% of older celiac disease patients.[157,158] In a study of 259 malignancies in 235 patients with histologically confirmed celiac disease,[159] approximately one-half of the tumors were lymphomas (mainly malignant histiocytosis), most commonly in the small intestine. The malignant cells, which resemble histiocytes by morphology, are reported to have T-cell markers on their surface,[160,161] but definitive proof of their T-cell lineage will require further study.[162,163] Of the nonlymphomatous tumors, one-half arose from the gastrointestinal tract, primarily small intestinal adenocarcinoma and squamous cell carcinoma of the esophagus and pharynx.[159]

The potential value of screening examinations for early detection of malig-

nancy, which exams are appropriate, and what parameters define the celiac population at highest risk are not known. Further, whether or not lifetime adherence to a strict gluten-free diet offers protection from malignancy remains an open question.[164,165] Villous atrophy without overt celiac disease sometimes accompanies small-bowel malignancy. It is not clear whether this reflects silent celiac disease complicated by cancer or a nonceliac carcinomatous enteropathy, although HLA-typing studies have provided some support from the former possibility.[166]

9.3.2. Refractory or Unclassified Sprue

Patients may respond initially to a gluten-free diet and subsequently relapse despite maintaining their diet. Such patients may then be refractory to further dietary therapy. Others are refractory to dietary treatment from its inception and, assuming the validity of their gluten-free diet, may not have celiac disease (i.e., unclassified sprue). Collagenous sprue[167] often is regarded as a separate entity from celiac disease. However, subepithelial collagen has been noted in up to 36% of patients with classic celiac disease and in tropical sprue.[168] Even the presence of large amounts of subepithelial collagen has not precluded a successful response to a gluten-free diet in some patients.[168]

9.3.3. Ulcerative Ileojejunitis

Patients with this uncommon but serious and poorly understood disorder frequently have a past history of celiac disease. It has been suggested that these patients may have developed a superimposed autoimmune reaction to their intestinal epithelial cells.[169]

10. GLUTEN-FREE DIET AND NUTRITION

Treatment consists of a gluten-free diet. After gluten restriction, clinical improvement can occur within days, particularly in children. Morphological abnormalities return toward normal in weeks. However, maintenance of a strict gluten-free diet is not a simple matter, as gliadins and other disease-activating prolamins are ubiquitous in many processed foods. Wheat starch flour, which lacks gliadins, forms the basis for the preparation of gliadin-free bread but is considered unpalatable by some. Durum wheat, a tetraploid variety used in the preparation of pasta, may be less injurious than the common hexaploid bread wheats[73] and has been suggested as an alternative for some patients. However, durum wheat can cause mucosal atrophy[73] and is best avoided.

Significant differences of opinion exist with respect to the inclusion of

other grains in the diet. Studies using D-xylose excretion[170] and intestinal biopsy[171] indicate that barley is harmful in the majority of patients. Nevertheless, some medical centers permit its use[170] unless the patient is refractory to wheat and rye withdrawal.[172] Oats have ~25% of the prolamin content of other cereal grains known to activate celiac disease.[170] Their toxicity may be dose-dependent[170,173]; opinions differ regarding their restriction in a gluten-free diet.[170,173–175]

Corn and rice do not activate celiac disease, but the data regarding several other grains are less certain. Buckwheat does not derive from the grass family and is usually permitted in the diet.[172] Millet and sorghum appear more closely related to corn and rice, and are allowed in most diets.[172] Triticale, a hybrid of wheat and rye, is sold in many health food stores. It may activate disease but is not often included on lists of prohibited foods.[172]

Gluten is not present in distilled spirits. Thus, rye whiskey, Scotch whiskey, and other cereal-derived spirits can be consumed. Similarly, brandy and wine are made from fruit and should cause no problem. Beer and ale are produced from barley. Their approval in the diet differs between medical centers,[172,176,177] although there is a lack of clinical evidence that they activate disease. Malt is made from barley and should be avoided, but the disease-activating proteins are removed in the production of malt extract and malt flavoring. Hydryolyzed vegetable proteins (HVPs) are flavor enhancers in processed foods that can be made from soy, wheat, or other cereal proteins. Their source generally is not provided on food labels and therefore foods containing HVPs are best avoided.

A lifelong gluten-free diet is recommended for all celiac disease patients, regardless of symptoms.[82,178] How strict should the diet be? It should be absolute or tailored according to the level of gluten sensitivity of the patient.[145,152] Because clinical improvement correlates with the strictness of the diet, and because gliadin and related prolamins result in damage to the mucosa, their restriction in theory should be complete for all patients. However, celiac disease diets are often individualized according to symptoms and histology. A recent study reported that a low-gluten diet (i.e, 2.5 g/day) resulted in mild lymphocytic infiltration of the jejunal epithelium, but did not affect gross jejunal morphology or stimulate a significant serum antigluten antibody response.[179] Based on such limited data, these investigators proposed treating celiac disease with a low-gluten rather than a gluten-free diet. However, at present, such a low-gluten diet is not warranted in childhood during normal growth and development, and also is controversial as being ideal in adults.

Multivitamin supplements are advised for all patients, while specific vitamin, mineral, and trace element deficiencies should be corrected. Treatment with iron supplements is warranted for iron deficiency, even in the absence of frank anemia.[180]

11. FAILURE TO RESPOND

Following institution of a gluten-free diet, clinical improvement begins promptly within days to weeks. If this does not occur, three possibilities should be considered: (1) poor dietary compliance or an improperly prescribed diet, (2) presence of an associated disease or complication, or (3) an erroneous diagnosis.

Non-gluten-containing foods also have been suggested to be responsible for persisting symptoms.[181,182] In the enigmatic patient, hospitalization and institution of a supervised dietary plan is appropriate. Should this fail, total parenteral nutrition and reintroduction of one nutrient at a time into the diet can be considered. Refractory disease in four patients was reported to improve when zinc deficiency was corrected.[183]

12. CONCLUSIONS

Research in immunogenetics, protein chemistry, and molecular biology is leading to a greater understanding of the etiopathogenesis of celiac disease, and likely will result in new diagnostic tests and alternative approaches to treatment. Over the next several years, better definition of the disease-activating peptides in wheat, rye, and barley will likely be attained. However, much more information will be needed about the importance of a strict gluten-free diet in preventing the long-term morbidity and complications of this disease. Lessons learned from celiac disease should substantially increase our understanding of the mechanisms that are important in other intestinal and HLA-associated autoimmune diseases.

ACKNOWLEDGMENTS. The authors thank Ms. Debby Sagall and Ms. Sharon Lai for preparation of the manuscript.

REFERENCES

1. Falchuk ZM. Update on gluten-sensitive enteropathy. *Am J Med* 1979;67:1085–1096.
2. Aretaeus, quoted by Major RH. *Classic Descriptions of Disease*, 3rd ed. Springfield, Ill: Charles Thomas, 1945, pp. 600–601.
3. Gee S. On the coeliac affliction. *St Barth Hosp Rep* 1888;24:17.
4. Dicke WK, Weijers HA, van de Damer JH. Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. *Acta Paediatr* 1953;42:34–42.
5. van de Kamer H, Weijers HA, Dicke WK. Coeliac disease. IV. An investigation into the injurious constituents of wheat in connection with the action on patients with coeliac disease. *Acta Paediatr* 1953;42:223–231.

6. Kasarda DD. Toxic proteins and peptides in celiac disease: Relations to cereal genetics. In: *Foods, Nutrition and Evolution* (Walcher D, Kretchmer M, eds). New York: Masson, 1981, pp. 201–216.
7. Kasarda DD, Bernardin JE, Nimmo CC. Wheat proteins. In: *Advances in Cereal Science and Technology,* Vol 1 (Pomerancz Y, ed). St. Paul: American Association of Cereal Chemists, 1976, pp. 158–236.
8. Kasarda DD, Nimmo CC, Bernardin JE. Structural aspects and genetic relationships of gliadins. In: *Proceedings of the Second International Celiac Symposium* (Hekkems WTJM, Pena AS, eds). Leiden: H.E. Stenfert Korese, B.V., 1974, pp. 25–36.
9. Garcia-Olmeda F, Carbonero P, Jones BL. Chromosomal locations of genes that control wheat endosperm proteins. In: *Advances in Cereal Science and Technology* (Pomeranz Y, ed). St. Paul: American Association of Cereal Chemists, 1982, pp. 1–47.
10. Kasarda DD, Autran JC, Lew EJL, Nimmo CC, Shewry PR. N-terminal amino acid sequences of ω-gliadins and ω-secalins: Implications for the evolution of prolamin genes. *Biochem Biophys Acta* 1983;747:138–150.
11. Kasarda DD, Bernardin JE, Qualset CO. Relationship of gliadin protein components to chromosomes in hexploid wheats. *Proc Natl Acad Sci USA* 1976;73:3646–3650.
12. Kasarda DD, Lafiandra D, Morris R, Shewry PR. Genetic relationships of wheat gliadin proteins. *Kulturpflanze* 1984;32:533–552.
13. Ciclitira PJ, Hunter JO, Lennox ES. Clinical testing of bread made from nullisomic 6A wheats in celiac patients. *Lancet* 1980;2:234–236.
14. Woychik JH, Boudy JA, Dimler RJ. Starch gel electrophoresis of wheat gluten proteins with concentrated urea. *Arch Biochem Biophys* 1961;94:477–482.
15. Autran JC, Leue JL, Minno CC, Kasarda DD. N-terminal amino acid sequencing of prolamins from wheat and related species. *Nature* 1979;282:527–529.
16. Kasarda DD. Structure and properties of α-gliadins. *Ann Tech Agric* 1980;29:151–173.
17. Bartels D, Thompson RD. The characterization of cDNA clones coding for wheat storage proteins. *Nuc Acids Res* 1983;11:2961–2977.
18. Bietz JA, Huebna FR, Sanderson JE, Wall JS. Wheat gliadin homology revealed through N-terminal amino acid sequence analysis. *Cereal Chem* 1977;54:1070–1083.
19. Shewry PR, Autran J-C, Nimmo CC, Lew EJ-L, Kasarda DD. N-terminal amino acid sequence homology of storage protein components from barley and a diploid wheat. *Nature* 1980;286:520–522.
20. Shewry PR, Lew EJ-L, Kasarda DD. Structural homology of storage proteins coded by the *Hor-1* locus of barley (*Hordeum vulgare L.*). *Planta* 1981;153:246–253.
21. Bernardin JE, Kasarda DD, Mecham DK. Preparation and characterization of A-gliadin. *J Biol Chem* 1967;242:445–450.
22. Falchuk ZM, Gebhard RL, Sessoms C, Strober W. An in vitro model of gluten sensitive enteropathy. Effect of gliadin on intestinal epithelial cells of patients with gluten-sensitive enteropathy in organ culture. *J Clin Invest* 1974;53:487–500.
23. Hekkens WTJM, Haex AJC, Willighagen RGJ. Some aspects of gliadin fractionation and testing by a histochemical method. In: *Coeliac Disease* (Booth CC, Dowling RH, eds). Edinburgh: Churchill Livingstone, 1970, pp. 11–18.
24. Kasarda DD, Okita TW, Bernardin JE, Baecker PA, Nimmo CC, Lew EJ, Dietler MD, Greene FC. Nucleic acid (cDNA) and amino acid sequences of α-type gliadins from wheat (*Triticum Aestivum L.*). *Proc Natl Acad Sci USA* 1984;81:4712–4716.
25. Rafalski JA, Scheets K, Metzler M, Peterson DM, Hedgcoth C, Soll DG. Developmentally regulated plant genes: The nucleotide sequence of a wheat gliadin genomic clone. *EMBO J* 1984;3:1409–1415.

26. Kendall MJ, Cox PS, Schneider R, Hawkins CF. Gluten subfractions in coeliac disease. *Lancet* 1972;2:1065–1067.
27. Jos J, Charbonnier L, Mougenot JF, Mosse J, Rey J. Isolation and characterization of the toxic fraction of wheat gliadin in celiac disease. In: *Perspectives in Celiac Disease* (McNicholl B, McCarthy CF, Fottrell PF, eds). Baltimore: University Park Press, 1978, pp. 75–90.
28. Ciclitira PJ, Evans DJ, Fagg NLK, Lennox ES, Dowling RH. Clinical testing of gliadin fractions in celiac patients. *Clin Sci* 1984;66:357–364.
29. Howdle PD, Ciclitira PJ, Simpson FG, Losowsky MS. Are all gliadins toxic in celiac disease? An in vitro study of α, β, γ, ω gliadins. *Scand J Gastroenterol* 1984;19:41–47.
30. Jos J, Charbonnier L, Mosse J, Olives JP, de Tand M-F, Rey J. The toxic fraction of gliadin digests in coeliac disease. Isolation by chromatography on Biogel P-10. *Clin Chim Acta* 1982;119:263–274.
31. Kumar PJ, Sinclair TS, Farthing MJG, Ohannesian Ad, Jones D, Emmett S, Waldron N, Clark ML, Dawson AM. Clinical toxicity testing of pure gliadins in coeliac disease. *Gastroenterology* 1984;86:1147 (abstract).
32. Arthur LJH, Langman JJS. Prevalence of coeliac disease in derby. In: *The Genetics of Coeliac Disease* (McConnel RB, ed). Lancaster: MTP, Press Ltd, 1981, pp. 15–16.
33. Hallert C, Gotthard R, Jansson G, Norrby K, Walan A. Similar prevalence of coeliac disease in children and middle-aged adults in district of Sweden. *Gut* 1983;24:389–391.
34. Hallert C, Gotthard R, Norrby K, Walan A. On the prevalence of adult coelic disease in Sweden. *Scand J Gastroenterol* 1981;16:257–261.
35. Mylotte M, Egan-Mitchell B, McCarthy CF, McNicholl B. Incidence of coeliac disease in the west of Ireland. *Br Med J* 1973;1:703–705.
36. Rifkind EA, Logan RFA, Busuttil A, Gilmour H, Ferguson A. Coeliac disease in Edinburgh and the Lothians 1900–1980. *Scot Med J* 1982;27:256.
37. Rossipal E. On the incidence of coeliac disease in Austria: A study comprising a nine-year period. In: *The Genetics of Coeliac Disease* (McConnel RB, ed). Lancaster: MTP, 1981, pp. 23–37.
38. Shmerling DH. Incidence of age distribution of coeliac disease in north-eastern Switzerland. In: *The Genetics of Coeliac Disease* (McConnel RB, ed). Lancaster: MTP, 1981, pp. 19–22.
39. Stevens FM, Egan-Mitchell B, McCarthy CF, McNicholl B. Factors in the epidemiology of coeliac disease in West Ireland. In: *The Genetics of Coeliac Disease* (McConnel RB, ed). Lancaster: MTP, 1981, pp. 7–14.
40. Tobiasen K. Recent Scandinavian data on the epidemiology of coeliac disease. In: *The Genetics of Coeliac Disease* (McConnel RB, ed). Lancaster: MTP, 1981, pp. 47–50.
41. van Stirum J, Baerlocher K, Fanconi A, Gugler E, Tonz O, Shmerling DH. The incidence of coeliac disease in children in Switzerland. *Helv Paediatr Acta* 1982;37:421–430.
42. Al-Hassany M. Coeliac disease in Iraqi children. *J Trop Pediatr Environ Child Health,* 1975,21:178–179.
43. Nelson R, McNeish AS, Anderson CM. Coeliac disease in children of Asian immigrants. *Lancet* 1973;1:348–350.
44. Rabassa EB, Sagaro E, Fragoso T, Castaneda C, Gra B. Coeliac disease in Cuban children. *Arch Dis Child* 1981;56:128–131.
45. Simoons FJ. Celiac disease as a geographic problem. In: *Food Nutrition and Evolution* (Walcher DN, Kretchmer N, eds). New York: Masson, 1981, pp. 178–199.
46. Berg NO, Lindberg T. Incidence of coeliac disease and transient gluten intolerance in children in a Swedish urban community. *Acta Paediatr Scand* 1979;68:397–400.
47. Stenhammer L, Johansson C-G. The incidence of coeliac disease in children in South-East Sweden. *Acta Paediatr Scand* 1981;70:379–381.

48. Ellis A. Coeliac disease: Previous family studies. In: *The Genetics of Coeliac Disease* (McConnel RB, ed). Lancaster: MTP, 1981, pp. 177–200.
49. MacDonald WC, Dobbins WO, Rubin CE. Studies on the familial nature of celiac sprue using biopsy of the small intestine. *N Engl J Med* 1968;272:448–456.
50. Stenhammer L, Brandt A, Wagermark J. A family study of coeliac disease. *Acta Paediatr Scand* 1982;71:625–628.
51. Ellis A, Evans AP, McConnell RB, Woodrow JC. Liverpool coeliac family study. In: *The Genetics of Coeliac Disease* (McConnell RB, ed). Lancaster: MTP, 1981, pp. 265–286.
52. Sagaro E, Jimenez N. Family studies of coeliac disease in Cuba. *Arch Dis Child* 1981;56:132–133.
53. Lee FI, Prior J, Murray SM. Celiac disease in monozygotic twin boys. Asynchronous presentation. *Dig Dis Sci* 1982;27:1137–1140.
54. Lord C, MacGregor GA. Coeliac disease in identical infants. *Postgrad Med J* 1981;57:658–659.
55. Polanco I, Biemond I, van Leeuwen A, Schreuder I, Meera Kahan P, Guerrero J, D'Amaro J, Vazquez C, van Rood JJ, Pena AS. Gluten-sensitive enteropathy in Spain: Genetic and environmental factors. In: *The Genetics of Coeliac Disease* (McConnell RB, ed). Lancaster: MTP, 1981, pp. 211–230.
56. Shale DJ, Johnston DG, Hall R, Robert DF. Coeliac disease in monozygotic twins. *Postgrad Med J* 1982;58:797–798.
57. Cole SG, Kagnoff MF. Celiac disease. In: *Annual Review of Nutrition,* Vol 5 (Olson RE, ed). Palo Alto: Annual Reviews, 1985, pp. 241–266.
58. Tosi R, Vismara D, Tanigaki N, Ferrara GB, Cicimarra F, Buffolano W, Follo D, Auricchio S. Evidence that celiac disease is primarily associated with a DC locus allelic specificity. *Clin Immunol Immunopathol* 1983;28:395–404.
59. Corazza GR, Tabacchi P, Frisoni M, Prati C, Gasbarrini G. DR and non-DR Ia allotypes are associated with susceptibility to coeliac disease. *Gut* 1985;26:1210–1213.
60. Alper CA, Fleischnick E, Awdeh Z, Katz AJ, Unis EJ. Extended major histocompatibility complex haplotypes in patients with gluten-sensitive enteropathy. *J Clin Invest* 1987;79:251–256.
61. Howell MD, Austin RK, Kelleher D, Nepom GT, Kagnoff MF. An HLA-D region restriction fragment length polymorphism associated with celiac disease. *J Exp Med* 1986;164:333–338.
62. Howell MD, Smith JR, Austin RA, Kelleher D, Nepom GT, Kagnoff MF. An HLA-DP allelic specificity associated with celiac disease. *Clin Res* 1987;35:608A.
63. Howell MD, Smith JR, Austin RK, Kelleher D, Nepom GT, Volk B, Kagnoff MF. An extended HLA-D region haplotype associated with celiac disease. *Proc Natl Acad Sci USA* 1987 (in press).
64. Howell MD, Resner J, Austin RK, Kagnoff MF. Rapid identification of hybridization of probes for chromosomal walking. *Gene* 1987;55:41–45.
65. DeMarchi M, Borelli I, Olivetti E, Richiardi P, Wright P, Ansaldi N, Barbera C, Santini B. Two HLA-D and DR alleles are associated with coeliac disease. *Tissue Antigens* 1979;14:309–316.
66. Betuel H, Gebuhrer L, Percebois H, Descos L, Minaire Y, Bertrand J. Adult celiac disease associated with HLA-DRw3 and -DRw7. *Tissue Antigens* 1980;15:231–238.
67. Scholz S, Albert E. HLA and diseases: Involvement of more than one HLA-linked determinant of disease susceptibility. *Immunol Rev* 1983;70:77–88.
68. Mearin ML, Biemond I, Pena As, Polanco I, Vazquez C, Schreuder GTM, DeVries RRP, Van Rood JJ. HLA-DR phenotypes in Spanish coeliac children: Their contribution to the understanding of the genetics of the disease. *Gut* 1983;24:532–537.

69. Kagnoff MF, Weiss JB, Brown RJ, Lee T, Schanfield MS. Immunoglobulin allotype markers in gluten-sensitive enteropathy. *Lancet* 1983;1:952–953.
70. Kangoff MF. Two genetic loci control the murine immune response to A-gliadin, a wheat protein that activates coeliac sprue. *Nature* 1982;296:158–160.
71. Weiser MM, Douglas AP. Cell surface glycosyltransferases of the enterocyte in coeliac disease. In: *Perspective in Celiac Disease* (McNicholl B, McCarthy CF, Fottrell PF, eds). Baltimore: University Park Press, 1978, pp. 451–458.
72. Carbonara AO, DeMarchi M, van Loghem E, Ansaldi N. Gm markers in celiac disease. *Human Immunol* 1983;6:91–95.
73. Auricchio S. Gluten-sensitive enteropathy and infant nutrition. *J Pediatr Gastroenterol Nutr* 1983;2(Suppl.1):S305–S309.
74. Auricchio S, Follo D, de Ritis G, Giunta A, Marzorati D, Prampolini L, Ansaldi N, Levi P, Dall'Olio D, Bossi A, Cortinovis I, Marubini E. Working hypothesis: Does breast feeding protect against the development of clinical symptoms of celiac disease in children? *J Pediatr Gastroenterol Nutr* 1983;2:428–433.
75. Kagnoff MF, Austin RK, Hubert JJ, Kasarda DD. Possible role for a human adenovirus in the pathogenesis of celiac disease. *J Exp Med* 1984;160:1544–1557.
76. Kagnoff MF, Paterson YJ, Kumar PJ, Kasarda DD, Carbone FR, Unsworth DJ, Austin RK. Evidence for the role of a human intestinal adenovirus in the pathogenesis of coeliac disease. *Gut* 1987;28:995–1001.
77. Karagiannis JA, Priddle JD, Jewell DP. Cell-mediated immunity to a synthetic gliadin peptide resembling a sequence from adenovirus 12. *Lancet* 1984;5:884–886.
78. Kagnoff MF. Immunology of the digestive system. In: *Physiology of the Gastrointestinal Tract* (Johnson LR, ed). New York: Raven Press, 1987, pp. 1699–1728.
79. Kagnoff MF, Austin RK, Johnson HCL, Bernardin JE, Dietler MD, Kasarda DD. Celiac sprue: Correlation with murine T-cell responses to wheat gliadin components. *J Immunol* 1982;129:2693–2697.
80. Trefts PE, Kagnoff MF. Gluten-sensitive enteropathy. I. The T-dependent anti-A-gliadin antibody response maps to the murine histocompatibility locus. *J Immunol* 1981;126:2249–2252.
81. Ciclitira PJ, Ellis JH, Evans DJ. A solid phase radioimmunoassay for measurement of circulating antibody titres to wheat gliadin and its subfractions in patients and adult celiac disease. *J Immunol Methods* 1983;62:231–239.
82. Falchuk ZM. Gluten-sensitive enteropathy. *Clin Gastroenterol* 1983;12:475–494.
83. Levenson SD, Austin RK, Dietler MD, Kasarda DD, Kagnoff MF. Specificity of anti-gliadin antibody in celiac disease. *Gastroenterology* 1985;89:1–5.
84. Baklien K, Brandtzaeg P. Immunoglobulins in jejunal mucosa and serum from patients with adult celiac disease. *Scand J Gastroenterol* 1977:12:149–159.
85. Scott H, Ek J, Baklein K, Brandtzaeg P. Immunoglobulin-producing cells in jejunal mucosa of children with coeliac disease on a gluten-free diet and after gluten challenge. *Scand J Gastroenterol* 1980;15:81–88.
86. Scott BB, Goodall A, Stephenson P, Jenkins D. Small intestinal plasma cells in coeliac disease. *Gut* 1984;25:41–46.
87. Falchuk ZM, Strober W. Gluten-sensitive enteropathy: Synthesis of antigliadin antibody in vitro. *Gut* 1974;15:947–952.
88. Ciclitira PJ, Ellis HJ, Wood GM, Howdle PD, Losowsky MS. Secretion of gliadin antibody by coeliac jejunal mucosal biopsies cultured in vitro. *Clin Exp Immunol* 1986;64:119–124.
89. Haeney MR, Goodwin BJF, Barratt MEJ, Mike N, Asquith P. Soya protein antibodies in man: their occurrence and possible relevance in coeliac disease. *J Clin Pathol* 1982;35:319–322.
90. Stern M, Bender SW, Gruttner R, Posselt HG, Strobel S. Serum antibodies against gliadin and reticulin in a family study of coeliac disease. *Eur J Pediatr* 1980;135:31–36.

91. Chorzelski TP, Beutner EJ, Sulej J, Tchorzewska H, Jablonska S, Kumar V, Kapuscinska A. IgA anti-endomysium antibody: A new immunological marker of dermatitis herpetiformis and coeliac disease. *Br J Dermatol* 1984;111:395–402.
92. Maury CPJ, Teppo AM. Demonstration of tissue 90 kD glycoprotein as antigen in circulating IgC immune complexes in dermatitis herpetiformis and coeliac disease. *Lancet* 1984;2:892–894.
93. O'Farrelly C, Feighery C, Greally JF, Weir DG. Cellular response to alpha-gliadin in untreated coeliac disease. *Gut* 1982;23:83–87.
94. Simpson FG, Robertson AF, Howdle PD, Losowsky MS. Cell-mediated immunity to dietary antigens in coeliac disease. *Scand J Gastroenterol* 1982;17:671–676.
95. Ferguson A, Macdonald TT, McClure JP, Holden RJ. Cell-mediated immunity to gliadin within the small-intestinal mucosa in coeliac disease. *Lancet* 1975;1:895–897.
96. Howdle PD, Bullen AW, Losowsky MS. Cell-mediated immunity to gluten within the small intestinal mucosa in coeliac disease. *Gut* 1982;23:115–122.
97. Corazza GR, Rawcliffe PM, Frisoni M, Sarchielli P, Londei M, Campieri M, Lazzari R, Gasbarrini G. Specificity of leukocyte migration inhibition test in coeliac disease: A reassessment using different gluten subfractions. *Clin Exp Immunol* 1985;60:117–122.
98. O'Farrelly C, Hekkens WThJH, Feighery C, Weir DG. The specificity of wheat protein reactivity in coeliac disease. *Scand J Gastroenterol* 1983;18:603–607.
99. Horvath K, Graf L, Walcz E, Bodanszky H, Schuler D. Naloxone antagonises effect of alpha-gliadin on leukocyte migration in patients with coeliac disease. *Lancet* 1985;2:184–185.
100. Ashkenazi A, Levin S, Idar D, Handzel ZT, Altman Y, Or A, Barzilai N. Cellular immunity in children with coeliac disease. *Eur J Pediatr* 1982;138:250–253.
101. Flores AF, Winter HS, Bhan AK. In vitro model to assess immunoregulatory T lymphocyte subpopulations in gluten sensitive enteropathy (GSE). *Gastroenterology* 1982;82:1058 (abstract).
102. Shanahan F, McKenna R, McCarthy CF, Drury MI. Coeliac disease and diabetes mellitus: a study of 24 patients with HLA typing. *Q J Med* 1982;51:329–335.
103. Malizia G, Trejdosiewicz LK, Wood GM, Howdle PD, Janossy G, Losowsky MS. The microenvironment of coeliac disease: T cell phenotypes and expression of the T2 T blast antigen by small bowel lymphocytes. *Clin Exp Immunol* 1985;60:437–446.
104. Pignata C, Troncone R, Monaco G, Ciriaco M, Farris E, Carminati G, Auricchio S. Impaired suppressor activity in children affected by coeliac disease. *Gut* 1985;26:285–290.
105. Corazza GR, Sarchielli P, Londei M, Frisoni M, Gasbarrini G. Gluten specific suppressor T cell dysfunction in coeliac disease. *Gut* 1986;27:392–398.
106. Sarles J, Gorvel JP, Olive D, Maroux S, Mawas C, Giraud F. Subcellular localization of class I (A,B,C) and class II (DR and DQ) MHC antigens in jejunal epithelium of children with coeliac disease. *J Pediatr Gastroenterol Nutr* 1987;6:51–56.
107. Scott H, Brantzaeg P, Solheim BG, Thorsby E. Relation between HLA-DR-like antigens and secretory component (SC) in jejunal epithelium of patients with coeliac disease or dermatitis herpetiformis. *Clin Exp Immunol* 1981;44(2):233–238.
108. Ciclitira PJ, Nelufer JM, Ellis HJ, Evans DJ. The effect of gluten on HLA-DR in the small intestinal epithelium of patients with coeliac disease. *Clin Exp Immunol* 1986;63:101–104.
109. Neild GH. Coeliac disease: A graft-versus-host-like reaction localized to the small bowel wall? *Lancet* 1981;1:811–812.
110. Manson-Smith DF, Bruch RG, Parrott DMV. Villous atrophy and expulsion of intestinal trichinella spiralis are mediated by T cells. *Cell Immunol* 1979;47:285–292.
111. Phelan JJ, Stevens FM, McNicholl B, Fottrell PF, McCarthy CF. Coeliac disease: the abo-

lition of gliadin toxicity by enzymes from *Aspergillus niger*. *Clin Sci Molec Med* 1977;53:35–43.
112. Bernardin JE, Aunders RM, Kasarda DD. Absence of carbohydrate in celiac-toxic A-gliadin. *Cereal Chem* 1976;53:612–614.
113. Douglas AP. The binding of a glycopeptide component of wheat gluten to intestinal mucosa of normal and coeliac human subjects. *Clin Chim Acta* 1976;73:357–361.
114. Torres-Pinedo R. State of the art lectins and the intestine. *J Pediatr Gastroenterol Nutr* 1983;2:588–594.
115. Vasmant D, Feldmann G, Fontaine J-L. Ultra-structural localization of concanavalin A surface receptors on brush-border enterocytes in normal children and during coeliac disease. *Pediatr Res* 1982;16:441–445.
116. Weiss JB, Austin RK, Schanfield MS, Kagnoff MF. Gluten-sensitive enteropathy. Immunoglobulin G heavy-chain (Gm) allotypes and the immune response to wheat gliadins. *J Clin Invest* 1983;72:96–101.
117. Rawcliffe PM, Priddle JD, Jewell DP. Coeliac disease: Possible mannose-specific lectin activity of gluten. *Clin Sci* 1985;69:11p.
118. Colyer J, Farthing MJG, Kumar PJ, Clark ML, Ohannesian AD, Waldron NM. Reappraisal of the "lectin hypothesis" in the aetiopathogenesis of coeliac disease. *Clin Sci* 1986;71:105–110.
119. Kolberg J, Sollid L. Lectin activity of gluten identified as wheat germ agglutinin. *Biochem Biophys Res Commun* 1985;130:867–872.
120. Sterchi EE, Woodley JF. Peptidases of the human intestinal brush border membrane. In: *Perspectives in Celiac Disease* (McNicholl B, McCarthy CF, Fottrell PF, eds). Baltimore: University Park Press, 1978, p. 437–449.
121. Bjarnason I, Peters TJ. In vitro determination of small intestinal permeability: Demonstration of a persistent defect in patients with coeliac disease. *Gut* 1984;25:145–150.
122. Bjarnason I, Marsh MN, Price A, Levi AJ, Peters TJ. Intestinal permeability in patients with coeliac disease and dermatitis herptiformis. *Gut* 1985;26:1214–1219.
123. Hamilton I, Cobden I, Rothwell J, Axon ATR. Intestinal permeability in coeliac disease: The response to gluten withdrawal and single-dose gluten challenge. *Gut* 1982;23:202–210.
124. Penna FJ, Hill ID, Kingston D, Robertson K, Slavin G, Shiner M. Jejunal mucosal morphometry in children with and without gut symptoms and in normal adults. *J Clin Pathol* 1981;34:386–392.
125. Guix M, Skinner JM, Whitehead R. Measuring intraepithelial lymphocytes, surface area, and volume of lamina propria in the jejunal mucosa of coeliac patients. *Gut* 1979;20:275–278.
126. Marsh MN. Studies of intestinal lymphoid tissue. III. Quantitative analyses of epithelial lymphocyte in the small intestine of human control subjects and of patients with celiac sprue. *Gastroenterology* 1980;79:481–492.
127. Rosekrans PCM, Meijer CJLM, Polanco I, Mearin ML, Van der Wal AM, Lindeman J. Long-term morphological and immunohistochemical observations on biopsy specimens of small intestine from children with gluten-sensitive enteropathy. *J Clin Pathol* 1981;34:138–144.
128. Leigh RJ, Marsh MN. Gluten-challenge causes time-related, some dependent lymphocyte responses in celiac sprue. *Gastroenterology* 1984;86:1157 (abstract).
129. Sjolund K, Alumets J, Berg N-O, Hakanson R, Sundler R. Enteropathy of coeliac disease in adults: Increased number of enterochromaffin cells in the duodenal mucosa. *Gut* 1982;23:42048.
130. Strobel S, Busuttil A, Ferguson A. Human intestinal mucosal mast cells: expanded population in untreated coeliac disease. *Gut* 1983;24:222–227.
131. Madara JL, Trier JS. Structural abnormalities of jejunal epithelial cell membranes in celiac sprue. *Lab Invest* 1980;43:254–261.

132. Halter SA, Greene HL, Helinek G. Gluten-sensitive enteropathy: Sequence of villous regrowth as viewed by scanning electron microscopy. *Human Pathol* 1982;13:811–818.
133. McCrae WM, Eastwood MA, Martin MR, Sircus W. Neglected coeliac disease. *Lancet* 1975;1:187–190.
134. Mortimer PE, Stewart JS, Norman AP, Booth CC. Follow-up study of coeliac disease. *Br Med J* 1968;3:7–9.
135. Sheldon W. Prognosis in early adult life of coeliac children treated with a gluten-free diet. *Br Med J* 1969;2:401–404.
136. Farthing MJG, Rees LH, Edwards CRW, Byfield PGH, Himsworth RL. Thyroid hormones and the regulation of thyroid function in men with celiac disease. *Clin Endocrinol* 1982;16:525–535.
137. Gillberg R, Kastrup W, Mobacken H, Stockbrugger R, Ahren C. Endoscopic duodenal biopsy compared with biopsy with the watson capsule from the upper jejunum in patients with dermatitis herpetiformis. *Scand J Immunol* 1982;17:305–308.
138. Katz SI, Hall RP, Lawley TJ, Strober W. Dermatitis herpetiformis: The skin and the gut. *Ann Intern Med* 1980;93:857–874.
139. Lawley TJ, Strober W, Yaoita H, Katz SI. Small intestinal biopsies and HLA types in dermatitis herpetiformis patients with granular and linear IgA skin deposits. *J Invest Dermatol* 1980;74:9–12.
140. Hitman GA, Niven MJ, Festenstein H, Cassell PG, Awad J, Walker-Smith J, Leonard JN, Fry L, Ciclitira P, Kumar P, Sachs JA. HLA class II alpha chain gene polymorphisms in patients with insulin-dependent diabetes mellitus, dermatitis herpetiformis, and celiac disease. *J Clin Invest* 1987;79:609–615.
141. Sachs JA, Awad J, McCloskey D, Navarrete C, Festenstein H, Elliot E, Walker-Smith JA, Griffiths CEM, Leonard JN, Fry L. Different HLA-associated gene combinations contribute to susceptibility for coeliac disease and dermatitis herpetitformis. *Gut* 1986;27:515–520.
142. Nepom GT, Hansen JA, Nepom BS. The molecular basis for HLA class II associations with rheumatoid arthritis. *J Clin Immunol* 1987;7:1–7.
143. Stenhammar L. Transient gastrointestinal disorders during infancy and early childhood. *Acta Paediatr Scand* 1981;70:383–387.
144. Mee AS, Burke M, Newman J, Cotton PB. Comparison of capsule and duodenoscopic biopsy specimens in the diagnosis of small intestinal disease. *Gut* 1980;20:A913 (abstract).
145. Kumar PJ, O'Donoghue DP, Stenson K, Dawson AM. Reintroduction of gluten in adults and children with treated coeliac disease. *Gut* 1979;20:743–749.
146. McNeish AS, Harms HK, Rey J, Shmerling DH, Visakorpi JK, Walker-Smith JA. The diagnosis of coeliac disease. A commentary on the current practices of members of the European Society for Paediatric Gastroenterology and Nutrition (ESPGAN). *Arch Dis Child* 1979;54:783–786.
147. Katz AJ, Falchuck ZM. Definitive diagnosis of gluten-sensitive enteropathy. Use of an in vitro organ culture model. *Gastroenterology* 1978;75:695–700.
148. Cobden I, Rothwell J, Axon ATR. Intestinal permeability and screening tests for coeliac disease. *Gut* 1980;21:512–518.
149. Menzies IS, Laker MF, Pounder R, Bull J, Heyer S, Wheeler PG, Creamer B. Abnormal intestinal permeability to sugars in villous atrophy. *Lancet* 1979;2:1107–1109.
150. Hill R, Cutz E, Cherian G, Gall DB, Hamilton JR. An evaluation of d-xylose absorption measurements in children suspected of having small intestinal disease. *J Pediatr* 1981;99:245–247.
151. Bjarnasin I, Peters TJ, Veall N. A persistent defect in intestinal permeability in coeliac disease demonstrated by a ^{51}Cr-labelled EDTA absorption test. *Lancet* 1983;1:323–325.

152. McNicholl B, Egan-Mitchell B, Fottrell PF. Variability of gluten intolerance in treated childhood coeliac disease. *Gut* 1979;20:126–132.
153. Packer SM, Charlton V, Keeling JW, Risdon RA, Oguilvie D, Rowlatt RJ, Larcher VF, Harries JT. Gluten challenge in treated coeliac disease. *Arch Dis Child* 1978;53:449–455.
154. Anderson IH, Levine AS, Levitt MD. Incomplete absorption of the carbohydrate in all-purpose wheat flour. *N Engl J Med* 1981;304:891–892.
155. Cooper BT, Holmes GKT, Ferguson R, Thompson RA, Allan RN. Gluten-sensitive diarrhea without evidence of celiac disease. *Gastroenterology* 79:1980;801–806.
156. Doherty M, Barry RE. Gluten-induced mucosal changes in subjects without overt small-bowel disease. *Lancet* 1981;1:517–520.
157. Cooper BT, Holmes GKT, Cooke WT. Lymphoma risk in coeliac disease of later life. *Digestion* 1982;23:89–92.
158. Selby WS, Gallagher ND. Malignancy in a 19-year experience of adult celiac disease. *Dig Dis Sci* 1979;24:684–688.
159. Swinson CM, Slavin G, Coles EC, Booth CC. Coeliac disease and malignancy. *Lancet* 1983;1:111–115.
160. Isaacson PG, O'Connor NTJ, Spencer J, Bevan DH, Connolly CE, Kirkham N, Pollack DJ, Wainscoat JS, Stein H, Mason DY. Malignant histiocytosis of the intestine: A T-cell lymphoma. *Lancet* 1985;2:688–691.
161. Salter DM, Krajewski AS, Dwar AE. Immunophenotype analysis of malignant histiocytosis of the intestine. *J Clin Pathol* 1986;39:8–15.
162. Kelleher D, Kagnoff MF. Malignant histiocytosis of T-cell lymphoma? *Gastroenterology* 1986;91:777–778.
163. Kagnoff MF. Histogenesis of malignant histiocytosis of the intestine. *Gastroenterology* 1987;92:2050–2054.
164. Cooper BT, Holmes GKT, Ferguson R, Cook WT. Celiac disease and malignancy. *Medicine* 1980;59:249–261.
165. Holmes GKT, Stokes PL, Sorahan TM, Prior P, Waterhouse JAH, Cooke WT. Coeliac disease, gluten-free diet, and malignancy. *Gut* 1976;17:612–619.
166. O'Driscoll BRC, Stevens FM, O'Gorman TA, Finnegan P, McWeeney JJ, Little MPG, Connolly CE, McCarthy CF. HLA type of patients with coeliac disease and malignancy in the west of Ireland. *Gut* 1982;23:662–665.
167. Trier JS, Falchuk M, Carey MC, Schrieber DS. Celiac sprue and refractory sprue. *Gastroenterology* 1978;75:307–316.
168. Bossart R, Henry K, Booth CC, Doe WF. Subepithelial collagen in intestinal malabsorption. *Gut* 1975;16:18–22.
169. Strober W, Falchuk ZM, Rogentine GN, Nelson DL, Klaeveman HL. The pathogenesis of gluten-sensitive enteropathy. *Ann Intern Med* 1975;83:242–256.
170. Baker PG, Read AE. Oats and barley toxicity in coeliac patients. *Postgrad Med J* 1976;52:264–268.
171. Anand BS, Piris J, Truelove SC. The role of various cereals in coeliac disease. *Q J Med* 1978;47:101–110.
172. Campbell JA. Foods for patients with celiac disease. *CMA J* 1982;127:963–965.
173. Baker PG. Oats and coeliac disease. *Br Med J* 1974;4:588–589.
174. Barry RE, Henry C, Read AE. The patient's view of a gluten-free diet. In: *Perspectives in Celiac Disease* (McNicholl B, McCarthy CF, Fottrell PF, eds). Baltimore: University Park Press, 1978, pp. 487–493.
175. Dissanayake AS, Truelove SC, Whitehead R. Lack of harmful effect of oats on small-intestinal mucosa in coeliac disease. *Br Med J* 1974;4:189–191.

176. Bell L, Hoffer M, Hamilton JR. Recommendations for foods of questionable acceptance for patients with celiac disease. *J Can Diet Assoc* 1981;42:143–158.
177. Wood MN. *Gourmet Food on a Wheat Free Diet.* Springfield, Ill.: Charles C Thomas, 1972.
178. McNeish AS. Coeliac disease: Duration of gluten-free diet. *Arch Dis Child* 1980;55:110–111.
179. Montgomry AMP, Goka AKJ, Kumar PJ, Farthing MJG, Clark ML. Jejunal morphology and humoral responses to gluten in adult celiac disease: Gluten free diet vs low gluten diet. *Gastroenterology* 1987;92:1539 (abstract).
180. Kosnai I, Kuitunen P, Silmes MA. Iron deficiency in children with coeliac disease on treatment with gluten-free diet. *Arch Dis Child* 1979;54:375–378.
181. Baker AL, Rosenberg IH. Refractory sprue: recovery after removal of nongluten dietary proteins. *Ann Intern Med* 1978;89:505–508.
182. Pock-Steen OCH. The role of gluten, milk, and other dietary proteins in chronic or intermittent dyspepsia. *Clin Allergy* 1973;3:373–383.
183. Love AHF, Elmes M, Golden MK, McMaster D. Zinc deficiency and celiac disease. In: *Perspectives in Celiac Disease* (McNicholl B, McCarthy CF, Fottrell PF, eds). Baltimore: University Park Press, 1978, pp. 335–342.

14

Pathophysiology of Diarrhea
Mechanisms of Action of Antidiarrheal Agents

N. W. Read

1. INTRODUCTION

It has been estimated that approximately 9 liters of fluid enters the gut each day: 1–1.5 liters from the diet; the rest as digestive secretions from the liver, pancreas, stomach, and small intestine. The majority of this fluid is absorbed during fairly rapid passage through the small intestine. About 1–1.5 liters passes through the ileocecal valve into the colon, and 90% of this is absorbed in the colon leaving the subject to pass a solid stool of approximately 100 g.

Diarrhea is often regarded as the failure of the gut to cope with the influx of fluid and, according to clinical investigators, a state of diarrhea exists when the patient passes more than 200 g of stool per day. With more people ingesting a high-fiber diet, it is likely that the criterion for diarrhea should increase to more than 400 g of stool per day. Whatever the line of demarcation, the definition implies either that the patient is ingesting an abnormally large amount of fluid or that an abnormally large volume of fluid is secreted into the gut or that fluid absorption is impaired. The role of intestinal motility and transit in the pathogenesis of diarrhea is incompletely understood, but certain types of contractile activity can impair absorption or enhance secretion.

Not all patients who complain of diarrhea pass abnormally large volumes of stool. Patients with irritable bowel syndrome or ulcerative colitis may pass small amounts of stool or mucus very frequently and often complain of urgency

N. W. Read • Sub-Department of Human Gastrointestinal Physiology and Nutrition, Royal Hallamshire Hospital, Sheffield, United Kingdom S10 2JF.

and incontinence. Appropriate treatment for diarrhea depends on knowing the nature of the patient's complaint and having accurate insights into the pathophysiological process involved.

2. EXCESSIVE FLUID INGESTION

Diarrhea caused by excessive fluid ingestion is rare but has been described in compulsive water drinkers and heavy beer drinkers. Drinking large volumes of saline solutions is an excellent means of cleansing the colon prior to colonoscopy.

3. EXCESSIVE FLUID SECRETION

Excessive fluid secretion from the salivary glands, pancreas, or biliary tree sufficient to cause diarrhea has not been described, though it is likely that excessive pancreatic secretion may contribute to the fluid entering the gut in patients with vasoactive intestinal polypeptidomas. The absorptive capacity of the small intestine is normally sufficient to cope with excess fluid secretion from these sources. Abnormally high levels of gastric secretion occurs in the Zollinger–Ellison syndrome, but the diarrhea observed in this condition is related as much to the noxious effect of acid in the small intestine affecting digestion and fluid transport as to the increased gastric secretion.[1] Thus, excessive secretion sufficient to result in diarrhea almost always comes from the small or large intestine.

Excessive small intestine secretion may be produced by tumors secreting polypeptide hormones, such as vasoactive intestinal polypeptide, by infection with enterotoxigenic organisms, such as *Escherichia coli* and *Vibrio cholerae,* and by ingestion of certain laxatives. Excessive colonic secretion may be caused by ingestion of irritant laxatives and by the malabsorption of either long-chain fatty acids or bile acids in the small intestine. These agents are converted by colonic bacteria into potent secretagogues.

Despite the large number of mechanisms that may result in intestinal secretion, diarrhea caused by abnormal intestinal secretion is for the most part a transient, self-limiting phenomenon. Chronic secretory diarrhea is rare and suggests the possibility of endocrine tumors or laxative abuse.

3.1. Mechanisms of Intestinal Secretion

The consensus is that secretion occurs from the crypt epithelium whereas absorption of fluid takes place on the small intestinal villi or from the surface epithelial cells of the colon.

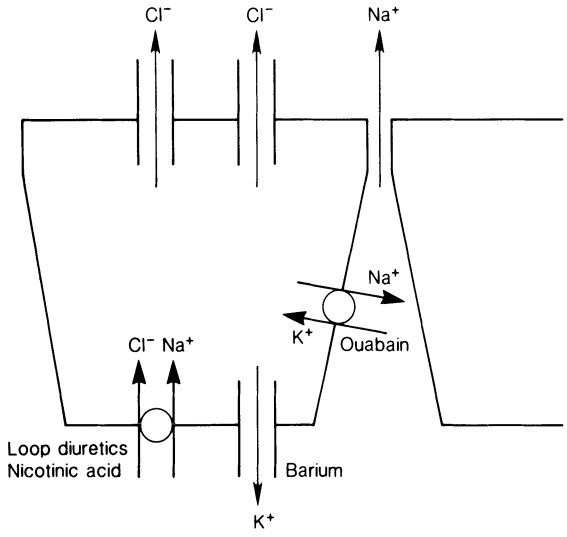


Figure 1. Alterations in transport of ions across the mucosal and basolateral membranes of the enterocyte that are implicated in the secretory process. Also indicated are the agents that could be used to block secretion by action at each of these sites.

The changes in membrane ion transport associated with intestinal secretion are shown in Fig. 1. The key mechanism facilitating secretion is an alteration in the permeability of the microvillous membrane of the enterocyte to chloride.[2] This depolarizes the membrane and alters the transmucosal potential difference, the lumen becoming more negative. The change in membrane potential encourages sodium to enter the lumen, probably from the lateral intercellular spaces. The passage of these ions across the epithelium then facilitates the flow of water into the intestinal lumen by osmotic mechanisms.

In order for chloride to continue to flow across the apical membrane, an equivalent amount of chloride has to enter the cell from the serosal membrane. Recent data[3,4] suggest that chloride enters the cell from the basolateral border by interacting with a carrier mechanism that also transports sodium and potassium into the cell. In order to maintain the flow of chloride into the cell, the sodium has to be pumped out of the cell and this takes place via the sodium- and potassium-dependent ATPase (Na^+, K^+-ATPase) also situated at the basolateral cell membrane. This causes more potassium to enter the cell. If potassium levels are allowed to increase too much, they will eventually inhibit the chloride entry process. Thus excess potassium is allowed to leak into the extracellular fluid via conductance channels situated at the basolateral border.[4] Inhibition of these potassium channels with barium inhibits the secretory process.[4] Secretion can be inhibited by drugs that act at any of these sites (Fig. 1), yielding several possibilities for development of new antidiarrheal agents.

The problem is that agents that influence the transport of ions across the membrane of the enterocyte will almost certainly alter ion transport across other cell membranes. Thus, unless the drug can be given in such a way as to confine its action to the gut, its use may be complicated by unwanted effects. Nicotinic acid, which acts on the basolateral membrane to inhibit sodium- and chloride-coupled entry,[5] has been shown to reduce diarrhea when administered to patients with cholera.

3.2. Intracellular Transduction of the Secretory Stimulus (Fig. 2)

It is generally agreed that the combination of a secretory transmitter with receptors situated on the membrane of the enterocyte initiates secretion either by increasing membrane permeability to calcium or by increasing the intracellular concentrations of either cyclic AMP or cyclic GMP. Intracellular calcium and the cyclic nucleotides serve as second messengers and induce secretion by activating appropriate protein kinases, which then phosphorylate specific membrane proteins[6–8] to alter the permeability of cell membrane to ions. There is some evidence that cAMP may work in conjunction with calcium. The entry of calcium into the cell activates adenylate cyclase in the rat colon[9] and agents that decrease cytosolic calcium also decrease the activity of cAMP in the rat small intestine.[10] An increase in cytosolic calcium may mediate intestinal secretion by combining with calmodulin to activate a specific calmodulin-dependent protein kinase,[7] though some studies have implied that calmodulin is not directly involved in active electrolyte secretion in the intestine.[12] More recent work suggested that calcium may induce secretion by activating protein kinase C (C-kinase)[11] via a mechanism that involves the turnover of phosphatidylinositol (PI) in the membrane and the release of inositol triphosphate (IP_3) and diacylglycerol (DG).

The increased understanding of the biochemical transduction of intestinal secretion yields several opportunities for the development of novel antisecretory drugs (Fig. 2). The problem is that cells in many organs share similar intracellular machinery, creating risks from unwanted effects if such agents are given. Phenothiazines have been shown to reduce the diarrhea of cholera, probably by inhibiting adenylate cyclase,[13–16] but these drugs also have potent effects on the central nervous system.

4. IMPAIRMENT OF INTESTINAL ABSORPTION

Most fluid absorption occurs in the upper small intestine secondary to the absorption of sodium. Sodium entry across the microvillous membrane of the jejunum occurs by cotransport with a number of organic solutes, such as hexose

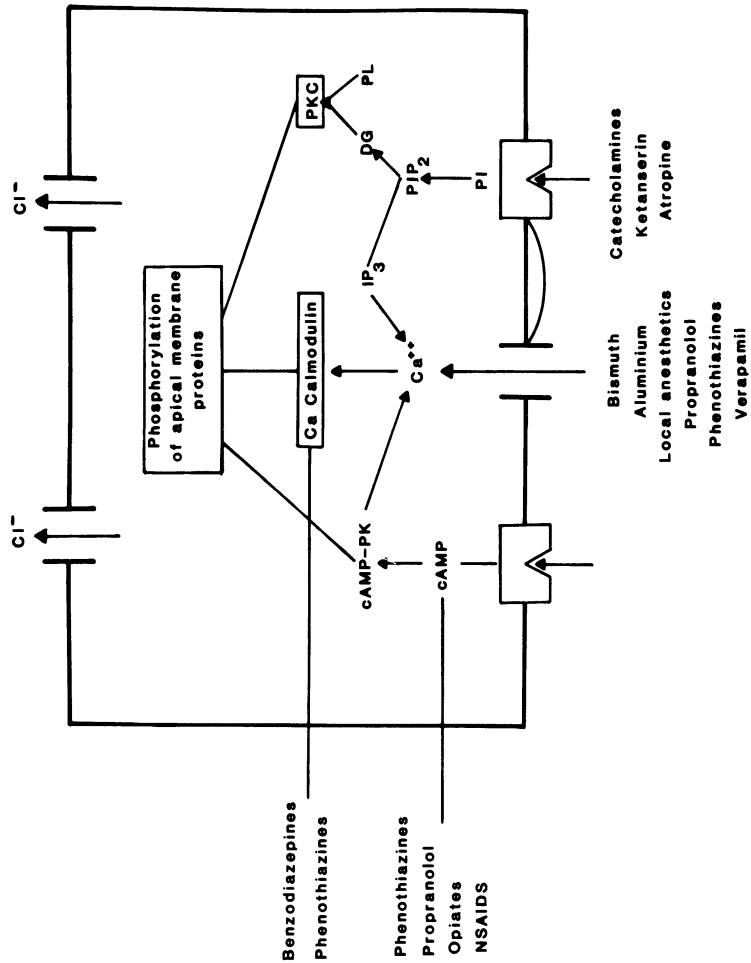


Figure 2. Putative intracellular mechanisms thought to transduce the action of a secretory neurotransmitter on the enterocyte and the agents that block these mechanisms.

sugars, amino acids or peptides, phosphate, and sulfate. Thus, any condition that impairs solute absorption by the small intestine also impairs fluid and electrolyte absorption. Such conditions include celiac disease, which causes atrophy of the intestinal epithelium, and chronic pancreatitis, which reduces digestion of solute. In both of these conditions, the unabsorbed solute retains fluid in the intestinal lumen by virtue of its osmotic activity, and may be voided as diarrhea.

Some absorption of salt and water can take place independently of net solute transfer: sodium can enter the cell across the microvillous membrane in exchange for hydrogen ion in both the jejunum and the ileum; in the ileum this mechanism is functionally coupled to a parallel entry of chloride in exchange for bicarbonate. The diarrhea of congenital chloridorrhea[17] is thought to be due to the absence of the chloride/bicarbonate exchange mechanisms. An absence of sodium/hydrogen exchange has also been implicated in a form of chronic diarrhea in neonates and infants.[18]

Perhaps the greatest impact on the treatment of infectious diarrhea in infants and in the Third World has come from the use of glucose and electrolyte solutions. These stimulate the absorption of fluid from the bowel and thus act to restore fluid balance. They are not, strictly speaking, antidiarrheal agents; the increased oral intake of fluid and solute may actually make the diarrhea worse, but it saves lives by restoring fluid balance.

5. ROLE OF MOTILITY

An increased flow of feces always involves the interaction of motor activity and epithelial transport. In some conditions, disturbances in absorption or secretion may occur secondary to disturbances in motor activity; in others, the alteration of epithelial transport may influence motor patterns. It is well established that an increase in the rate of perfusion of fluid into the gut will enhance the rate of transit.[19,20] This is not simply a passive phenomenon; fluid will not flow down the gut unless propelled by contractions. Luminal distension with fluid encourages peristalsis in isolated segments of gut[21] and generates propagative motor complexes in the intact small intestine,[22] clearing the fluid downstream.

The intracellular mediation of smooth muscle contraction is similar in some ways to the intracellular mediation of secretion. Calcium plays a pivotal role in both; the interaction of a transmitter with a membrane receptor either causes calcium to enter the cell or releases calcium from the sarcoplasmic reticulum. The increase in cytosolic calcium then activates calmodulin, which is necessary for the activation of myosin light-chain kinase, the enzyme that phosphorylates

mysoin and initiates contraction. The similarity in stimulus–secretion and stimulus–contraction coupling indicates that agents that act on the intracellular machinery to initiate intestinal secretion will also initiate contraction of intestinal smooth muscle and vice versa.

Increases in motor activity may cause diarrhea either by accelerating transit, which would reduce contact time, inducing secretion, or reducing contact area. Decreased motor activity may in theory impair absorption by reducing mixing or convection, though the concomitant increase in contact time may compensate for this effect. Intestinal stasis, however, may result in diarrhea by encouraging the overgrowth of anaerobic bacteria in the small intestine. These bacteria deconjugate and dehydroxylate bile acids, impairing fat absorption and causing a greater delivery of altered bile acids into the colon, where fatty acids and unabsorbed bile acids stimulate secretion[23] and propagative motor activity.[24]

Reduction in Contact Time

Gastrointestinal motor activity determines the rate at which material passes along the different parts of the gut and hence the degree of contact between gut contents and both the digestive enzymes and the absorptive epithelium. The degree of absorption is determined in part by the degree of contact between luminal contents and the absorptive epithelium and in part by the efficiency of the absorptive mechanisms.

Under normal conditions, it appears that the rate at which food moves through different parts of the gut closely matches the rate of digestion and absorption. The delivery of food into the small intestine is regulated by a duodenal feedback mechanism, and is compatible with the rate at which it can be digested by enzymes secreted into the duodenum.[25] Recent evidence suggests that the rate of transit through the small intestine and the degree of absorption may be regulated by the interaction of nutrients, particularly lipid, with receptors situated in the ileum.[26,27]

Reductions in the time for markers to pass from mouth to anus are associated with large stool volumes and diarrhea, whereas increases in whole-gut transit time are associated with low stool weights and constipation.[28] Patients with the irritable bowel syndrome who complain of diarrhea have more rapid transit through the small intestine and the colon than either patients who complain of constipation or patients who have pain and distension with no alteration in bowel habit.[29] Psychological stress, which is thought to induce diarrhea in some people, accelerates small-bowel transit.[30] Acceleration of small-bowel transit is also seen in patients with thyrotoxicosis, who may also present with diarrhea,[38] and in patients who have undergone partial gastrectomy and vagot-

omy,[32–34] which are surgical procedures commonly complicated by diarrhea. These observations indicate an association between impairment of net fluid absorption and rapid transit but do not tell us which was the primary event.

Similar relationships between contact time and absorption have been reported in studies investigating the absorption of meals in the small intestine. Johansson and her colleagues[35] fed a liquid test meal containing a nonabsorbable marker to normal volunteers, sampling the meal in the stomach and 65 cm down the small intestine. By analysis of the samples, they were able to calculate the absorption of the food constituents. During the same experiment they infused a continuous series of nonabsorbable markers into the duodenum and calculated the mean transit time in the proximal segment of small intestine. Their results demonstrated a direct relationship between transit time and the degree of absorption. We have used similar techniques to investigate the degree of absorption of a liquid carbohydrate meal in the proximal 210 cm of small intestine of normal volunteers.[27,36] Slowing small intestinal transit by infusing lipid into the ileum increased the degree of absorption, whereas the use of metoclopramide, which accelerated transit, was associated with reduced absorption.

Reversed electrical pacing slows small intestinal transit and enhances absorption.[37] Caution is needed in interpreting these results, however, because forward pacing enhances absorption as well[38] and recent data suggest that electrical pacing may increase epithelial transport by activating sympathetic nerves.[39]

To determine if changes in transit time can influence the degree of absorption down the whole small intestine, we studied the effect of accelerating small-bowel transit time on the absorption of a radiolabeled solid meal from the small intestine of 14 patients equipped with terminal ileostomies (less than 6 inches of ileal resection).[40] The transit time to the terminal ileum was monitored by plotting the cumulative delivery of the radioactive marker to the stoma, while chemical analysis of the meal and the ileostomy effluent provided an index of the degree of absorption. Administration of either lactulose, magnesium sulfate, or metoclopramide reduced small intestinal transit time to approximately the same degree. The reduction in transit time was accompanied in each case by an increase in wet and dry weight of ileostomy effluent and an increase in fat output. Administration of magnesium sulfate and lactulose also caused an increase in the output of protein and carbohydrate. In an earlier study using perfusion techniques in children, the addition of mannitol to a test drink accelerated transit through 50-cm segments of proximal small intestine and reduced the absorption of xylose, arginine, and palmitic acid.[41] Although these data supported the hypothesis that acceleration of transit time through the small intestine could reduce the degree of absorption, the extent to which absorption was reduced depended on the agent used to accelerate transit time. Mannitol, lactulose, and perhaps also magnesium sulfate accelerate small-bowel transit by

restricting fluid absorption osmotically. It is possible that the greater reduction in absorption observed with these "osmotic" agents was caused in part by luminal distension, dilution of luminal contents, and a less intimate association between the nutrients and both the absorptive epithelium and digestive enzymes. Thus, although a number of experiments support the logical assumption that factors that accelerate transit through the bowel reduce the degree of absorption, the hypothesis is difficult to prove because it is difficult to rule out other factors, such as dilution of luminal contents or the reduction in available absorptive surface plus the possibility that substances which act on smooth muscle may have an independent action on epithelial transport.

It could be said that most, if not all, of the effective antidiarrheal drugs act by increasing intestinal contact time. Opiates, α_2-adrenoreceptor agonists, and somatostatin all prolong transit through the intestine and inhibit contractile activity. It is possible, however, that the effects on transit and contractile activity in the intact intestine could be secondary to a reduction in luminal volume caused by a primary action on the epithelium.

To investigate whether antidiarrheal agents could reduce stool volumes by a primary action on motor activity or on transport, Fordtran's group perfused the whole gastrointestinal tract with a balanced electolyte solution containing polyethylene glycol (PEG) as a nonabsorbable marker at a steady rate of 30 ml/min for a period of 8 hr.[42,43] After about 60–90 min, fluid issued from the anus and continued to flow at a steady rate until the end of the experiment. Administration of codeine phosphate or loperamide did not alter the rate of elimination of fluid or the concentration of PEG, indicating that under the conditions of this experiment, these drugs did not affect epithelial transport. In contrast, administration of the α_2-adrenoreceptor agonist clonidine[44] or the addition of glucose to the perfusate[43] reduced the rate of anal effluent under steady-state conditions and increased the concentration of PEG, confirming that these substances can stimulate intestinal absorption of fluid. Both loperamide and codeine phosphate, however, delayed the appearance at the anus of a gastric infusion of 2700 ml of the same solution over 90 min and reduced the volume of effluent. Further studies showed that the drugs prolonged transit of fluid through segments of small intestine in normal volunteers. These findings suggested that in normal subjects opiate-like antidiarrheal agents act primarily on motor activity, increasing the time over which fluid remains in contact with the absorptive epithelium.

6. ENTERIC REFLEXES

Diarrhea, like vomiting, should perhaps be regarded as the response of the gut to potentially toxic or injurious solutions, acting to clear the gut of its

contents. If true, then the mechanism of diarrhea involves the interaction of the toxic or damaging substance with a mucosal receptor, mediation by certain chemical transmitters, and an effector system that increases excretion to flush out the damaging substance and stimulates propulsive motor activity to expel the fluid containing the offending substances from the body. The generation of a secretory response without a motor response or vice versa is unlikely to result in efficient clearance of luminal contents. Figure 3 illustrates the possible mechanisms for these responses.

Recent studies have shown that organisms such as *Vibrio cholerae*[45,46] and toxigenic *E. coli*[47,48] can evoke within a few hours both intestinal secretion and a highly propagative motor pattern known as the migrating action potential complex (MAPC).[45] Both secretory and motor responses to these organisms and/or their toxins can be inhibited by tetrodotoxin, which blocks fast sodium channels in nerves, by ganglion blockers and by local anesthetics applied to the serosal or mucosal aspects of the intestine. This suggests that both secretory and motor responses are mediated by nervous reflexes.[49–51] Since both secretion and MAPCs can be generated in isolated loops of small intestine,[48] the reflexes are mediated by the enteric nervous system, although they may be modulated by extrinsic nerves.

Both cholera-induced secretion and MAPCs can be reduced or abolished by indomethacin but restored by prostaglandin. The generation of cholera-induced secretion probably also involves the release of serotonin since cytofluorometric studies have shown a reduction in the serotonin content of enterochromaffin cells after exposure to cholera toxin,[52] and cholera secretion can be abolished by serotonin blockers or serotonin tachyphylaxis.[53] Exposure of the mucosa to cholera toxin also causes the release of vasoactive intestinal polypeptide (VIP),[54] a potent intestinal secretagogue found only in nerves in the gut. To date there are no published data to implicate the involvement of 5-hydroxytryptophan (5-HT) or VIP in the generation of MAPCs, though the latter are partially inhibited by muscarinic antagonists.

There is also evidence to suggest that laxatives and bile acids may generate secretion[48] and propagative motor complexes[55–57] via enteric reflexes.

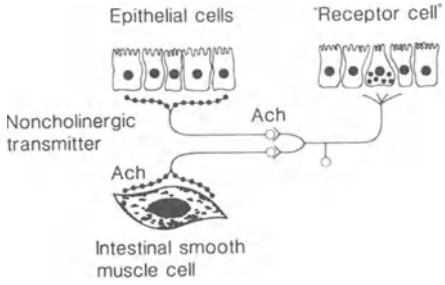


Figure 3. A simplified enteric reflex that may induce contraction and secretion in response to a luminal stimulus.

Although most of the work on MAPCs has been carried out using a preparation of rabbit ileum, reports of similar myoelectrical phenomena or of analogous propagated manometric phenomena have been described in other species including man.[58–59]

Distension of the intestinal lumen induces secretion[60] and a rise in transepithelial potential differences (PD)[61,62] indicate a change in intestinal ion transport. Both the secretion and the electrical change can be blocked by hexamethomium, suggesting mediation by a nervous reflex triggered by stimulation of intraluminal tension receptors. Luminal distension also releases local transmitters such as acetylcholine,[63,64] serotonin,[65] and prostaglandins,[66,67] all of which induce intestinal secretion and elevate the PD. Similarly, changes in fluid transport and PD can also be obtained by stimulation of the vagus nerve in experimental animals.[68]

Recordings from the human small intestine have shown that spontaneous and repetitive bursts of contractions occurring at intervals of about a minute were associated with fluctuations in PD that peaked about 45 sec after the pressure event.[62] Similar changes could be induced in the anaesthetized ferret by prolonged low-frequency stimulation of the distal cut end of the cervical vagus.[69] Administration of atropine blocked spontaneous pressure and PD changes in the ferret[69] but did not block the PD and secretory changes evoked by vagal stimulation.[68] These observations suggest that spontaneous increases in PD and presumably secretion occurred secondary to the pressure event.

Our results are compatible with the hypothesis that fluctuations in transepithelial PD (and presumably secretion) are triggered by stimulation of in-series tension receptors by intestinal contraction, though they could also occur by stimulation of rapidly adapting mucosal mechanoreceptors by food or the opposing mucosal surface. Rubbing the mucosal surface gently with the smooth end of a glass rod or a small balloon has been done for many years to induce secretion in the intestine of experimental animals.[67]

The implication from the work is that several factors may induce diarrhea via nervous reflexes mediated by the enteric nervous system. If this is the case, then the most effective antidiarrheal agents will be those that interrupt these nervous loops without affecting the basal motor and transport function of the gut.

Opiatelike agents act on the enteric nervous system and are the most widely used and effective antidiarrheal agents acting to modify intestinal motor activity and inhibiting epithelial secretion. The modification of intestinal motility prolongs the transit through the small intestine and the colon by reducing propagative motor activity[70–73] and/or by increasing nonpropulsive or poorly propulsive motor patterns.[74–81] Opiates act on the fasting small intestine, increasing the frequency of phase III of the migratory myoelectric complex but reducing motor activity during phase II. Although it is often assumed that opiates exert these effects through an action on the myenteric plexus,[70,71] the effects could be modified by a direct effect on the muscle or via the central

nervous system. Synthetic opiate analogs such as loperamide, which do not supposedly cross the blood–brain barrier, have similar effects on the bowel as opiates that penetrate the central nervous system.

The effect on phase III might be expected to accelerate small-bowel transit, except that the most rapid propulsion in the intact small intestine usually occurs during phase II or after feeding. Ruppin showed that the delay in small-bowel transit induced by loperamide was associated with an abolition of propagative contractile patterns during phase II.[82] No study has investigated the effect of opiates or opiate derivatives on the specific patterns of the small or large intestinal motor activity associated with diarrhea in man and other species. In experimental animals loperamide can inhibit the intestinal fluid secretion induced by cholera toxin or prostaglandins.[83] Perfusion studies carried out in human volunteers have indicated a similar phenomenon, though less convincingly.[84]

Since opiate receptors have not been demonstrated on enterocytes and loperamide does not cross the blood–brain barrier, it is assumed that opiate and opiate-like agents influence intestinal fluid transport by an action on the enteric nerve plexi. This would explain why an opiate-like antidiarrheal failed to inhibit intestinal secretion induced by VIP,[42,43] which acts on enterocyte receptors but could inhibit secretion[85] and diarrhea induced by bile acids,[85] dioctylsulfonylsuccinate,[86] a laxative detergent.

Sympathomimetic agents likely also act at the level of the enteric plexus to enhance intestinal absorption, inhibit secretion,[87–90] and depress motor activity. Although α_2-adrenoreceptors have been demonstrated in enterocytes,[91] Mathias recently showed that pretreatment of the rat with 6-hydroxydopamine to selectively destroy the varicosities containing norepinephrine or postganglionic adrenergic nerves[92] caused the appearance of MAPCs, prior to phase III of the MMC.

7. COLONIC SALVAGE

The capacity of the colon to absorb material entering from the ileum is much greater than is usually required. Under normal circumstances, between 500 and 1500 ml of unabsorbed material containing at least 20 g of carbohydrate passes into the colon each day. The colon can, however, adapt to absorb between 4 and 7 liters/day of plasmalike solution, when this is infused into the cecum at a steady rate.[92] The colon can also accommodate the entry of at least 25 g unabsorbable carbohydrate, taken in a single meal, before the appearance of carbohydrate in the stool and onset of diarrhea.[94] The carbohydrate is converted by anaerobic bacteria to short-chain fatty acids, mainly acetate, propionate, and butyrate. These substances are absorbed rapidly by the colonic epithelium, and stimulate the absorption of water.[95,96] The efficiency of carbohydrate salvage may help to explain why many patients with lactase deficiency or pa-

tients with celiac disease can drink milk and eat very large amounts of carbohydrate without getting osmotic diarrhea.

Transit time through the colon is normally about 10 times as long as it is through the small intestine.[28] Prolonged transit is necessary to allow sufficient time for the extraction of salt and water from the colonic contents against high osmotic and concentration gradients. The relative stagnation of colonic contents also encourages the growth of the colonic bacteria that are responsible for the conversion of unabsorbed carbohydrate to volatile fatty acids. The fact that the colon is the most distal part of the intestine and can adapt its absorption in response to a variable delivery from the ileum means that it plays a critical role in determining whether a person has diarrhea or not.

One useful way to look at diarrhea is to regard it as the failure of colonic salvage.[97] Diarrhea occurs only when the absorptive capacity of the colon is inadequate to cope with the amount of ileal effluent. This situation arises either because the amount or composition of ileal effluent overwhelms the normal capacity for colonic salvage or because abnormal colonic function or colonic disease prevents salvage of a normal ileal effluent.

Several factors could impair the colon's ability to salvage and absorb even a normal amount of ileal effluent. Recent studies from our laboratory indicated that impairment of the ability of the colon of patients with total ulcerative colitis to salvage carbohydrate is related to a reduction in absorption of short-chain fatty acids.[98]

Impaired fermentation caused by reduction in the numbers of colonic bacteria could explain the diarrhea frequently observed after administration of broad-spectrum antibiotics. An abnormal pattern of colonic motility could cause diarrhea by accelerating transit through the colon. Changes in colonic motility can impair colonic salvage by propelling material rapidly to the anus leaving insufficient time for fermentation of carbohydrate and absorption of fatty acids, salt, and water to take place. The slow rate of epithelial transport in the colon compared with the rate in the small intestine implies that changes in contact time have a relatively greater influence on colonic absorption. Comparisons of stool weight with colonic transit time suggests that stool weights in the diarrhea range would occur if colonic transit is less than about 30 hr.[28] Impaired absorption of fat or bile acids in the small intestine may also inhibit colonic salvage because bile acid and fat are converted by colonic bacteria to potent laxative agents that increase colonic secretion and propulsion.[23,24,98]

8. MOTILITY OF THE ANORECTUM

Many patients who complain to their doctors of diarrhea do not pass large volumes of feces. Instead, they tend to pass small volumes of stool frequently or they suffer from urgency or fecal incontinence. These patients have a distur-

bance of anorectal function causing either a hypersensitivity of the rectum or a weakness of the anal sphincter or both. In patients with the irritable bowel syndrome, distension of the rectum with a balloon causes discomfort at low volumes[99,100] and may also induce abnormally large amplitude rectal contractions, associated with precipitous anal relaxations.[100]

Drug therapy of anorectal disorders has not been developed to the extent of drug therapy of bladder disturbances. Nevertheless, agents that reduce smooth muscle contractility such as nifedipine[101,102] may reduce discomfort and frequency of defecation in the irritable bowel syndrome. When loperamide was administered to 26 patients with chronic diarrhea and incontinence, it reduced the frequency of episodes of fecal incontinence and severe urgency.[103] The improvement in these symptoms was probably explained by the action of this drug in increasing sphincter tone and enhancing continence as measured by a standard volume of saline infused rectally. Thus it would appear that loperamide may have a specific action on the anal sphincter that may improve continence in patients with diarrhea.[103] This action does not appear to be shared by diphenoxylate,[104] which had no specific effect on saline continence and sphincter manometry in incontinent patients and was much less effective than loperamide and codeine phosphate for alleviating fecal incontinence or suppressing urgency.[105]

9. CONCLUSION

Diarrhea cannot be regarded solely as the net result of a disorder in epithelial transport, nor can it be regarded solely as the result of a disorder of motility. It seems likely that most factors that induce diarrhea act via the enteric nervous system to stimulate both secretion and propulsive motor patterns. In any case, motor and transport phenomena are so closely related that a disorder of one must involve an alteration of the other. Distension of the intestine by fluid secretion can generate propulsive motor patterns leading to rapid transport and reduced solute and water absorption, thus creating a vicious spiral. The onus of drug treatment of diarrhea is to break the vicious spiral by slowing motility and/or reversing net secretion.

However, the effective and safe use of an antidiarrheal agent requires an accurate knowledge of the mechanisms of the disease. If diarrhea occurs as the response to an injurious agent, such as pathogenic bacteria, then limiting its expulsion from the body could increase the morbidity.[106]

Ulcerative colitis is associated with increased irritability of the distal colon but slowed transit through the proximal colon[107] irrespective of the extent of disease. Thus, administration of agents that slow transit could cause prolonged stasis and distension in the proximal colon. This may account for the case

reports of toxic megacolon in patients treated with opiate-like agents or anticholinergics.[108]

Agents that delay small-bowel transit may exacerbate the diarrhea associated with bacterial overgrowth in the small intestine because they increase stasis and encourage overgrowth.

Finally, opiatelike agents contract the sphincter of Oddi[109] and could exacerbate the steatorrhea of pancreatic disease.

Because of these potential hazards of antimotility agents, several drug companies have sought to develop agents that inhibit secretion without influencing motility. So far the search has not identified a satisfactory agent. Most antisecretory substances also inhibit motility and those that have little effect on motility are not very potent antisecretory agents. In any case, a pure antisecretory agent would have a somewhat limited application; excessive secretion is a relatively uncommon mechanism for diarrhea and, where implicated, it usually causes a brief self-limited illness like traveller's diarrhea, which can be successfully managed with loperamide. Fluid balance in severe life-threatening secretory diarrheas such as cholera and VIPomas is best treated by intravenous infusion or oral glucose–electrolyte solutions, and though it may be more convenient to use an agent that blocked the secretion, this idea has not been accepted with enthusiasm.

REFERENCES

1. Rambaud JC, Modigliani R, Emorts P et al. Fluid secretion in the duodenum and intestinal handling of water and electrolytes in Zollinger-Ellison syndrome. *Am J Dig Dis* 1978;23:1089–1097.
2. Field M. Secretion of electrolytes and water by mammalian small intestine. In: *Physiology of the Gastrointestinal Tract* (Johnson LR, ed). New York: Raven Press, 1981, pp. 963–982.
3. Heintze K, Stewart CP, Frizzell RA. Sodium-dependent chloride secretion across rabbit descending colon. *Am J Physiol* (*Gastrointest Liver Physiol* 7) 1983;244:G357–G365.
4. Dharmsathaphorn KA, Weymer KA, McRoberts JA. Chloride secretion induced by prostaglandin E_1 (PGE_1): Participation of Na, K, Cl cotransport, Cl channels and K channels (abstract). *Gastroenterology* 1985;88:1364.
5. Turjman N, Gotterer GS, Hendrix TR. Prevention and reversal of cholera enterotoxin effects on rabbit jejunum by nicotinic acid. *J Clin Invest* 1978;61:1155–1160.
6. Cohen ME, Donowitz M, Gudiurch R, Sharp GWG. cAMP-induced phosphorylation of rabbit ileal microvillus membrane and cytosol: Evidence of interaction (abstract). *Fed Proc* 1983;42:1288.
7. Rao MC, Nash N, Palfrey HC. Ca-calmodulin and cyclic neucleotide-dependent phosphorylation in epithelial cells (abstract). *J Cell Biol* 1982;95:254A.
8. Shlatz LJ, Kimberg DV, Cattieu KA. Phosphorylation of specific rat intestinal microvillus and basolateral membrane proteins by cyclic nucleotides. *Gastroenterology* 1979;76:293–298.
9. Craven PA, DeRubertis FR. Patterns of prostaglandin synthesis and degradation in isolated

superficial and proliferative colonic epithelial cells compared to residual colon. *Prostaglandins* 1983;26:583–604.
10. Gunther RD, Wright EM. Na, Li, and Cl transport by brush border membranes from rabbit jejunum. *J Membr Biol* 1983;74:85–94.
11. Fondacaro JD. Intestinal ion transport and diarrheal disease. *Am J Physiol* (*Gastrointest Liver Physiol* 13) 1986;250:G1–G8.
12. Donowitz M, Wicks J, Madara JL, Sharp GWG. Studies on role of calmodulin in Ca^{2+} regulation of rabbit ileal Na and Cl transport. *Am J Physiol* (*Gastrointest Liver Physiol* 11) 1985;248:G726–G720.
13. Holmgren J, Lange S, Lonnroth I. Reversal of cyclic AMP-mediated intestinal secretion in mice by chlorpromazine. *Gastroenterology* 1978;75:1103–1108.
14. Lonnroth I, Holmgren J, Lange S. Chlorpromazine inhibits cholera toxin-induced intestinal hypersecretion. *Med Biol* 1977;55:126–129.
15. Rabbani GH, Holmgren J, Greenough WB III, Lonnroth I. Chlorpromazine reduces fluid loss in cholera. *Lancet* 1979;1:410–412.
16. Holmgren J, Lange S, Lonnroth I. Reversal of cyclic AMP-mediated intestinal secretion in mice by chlorpromazine. *Gastroenterology* 1978;75:1103–1108.
17. Bieberdorf FA, Gorden P, Fordtran JS. Pathogenesis of congenital alkalosis and diarrhea. Implications for the physiology of normal ileal absorption and secretion. *J Clin Invest* 1972;51;1958–1964.
18. Booth IW, Strange G, Murer H, Fenton TR, Milla PJ. Defective jejunal brush border Na^+/H^+ exchange. A cause of secretory diarrhea. *Lancet* 1985;2:1066.
19. Dillard RL, Eastman H, Fordtran JS. Volume flow relationship during transport of fluid through the human small intestine. *Gastroenterology* 1965;49:58–66.
20. Williams NS, Meyer JH, Jehn D, Miller J. Canine intestinal transit and digestion of radiolabelled liver particles. *Gastroenterology* 1984;86:1451–1459.
21. Trendelenberg P. Physiologische and pharmakologische verusche uber die Dunndarmperistaltik. *Arch Exp Path Pharmak* 1917;81:55–129.
22. Fleckenstein P, Bueno L, Fioramonti J. Minute rhythm of electrical spike bursts of the small intestine in different species. *Am J Physiol* 1982;842:G654–G659.
23. Binder HJ. Pathophysiology of bile acid and fatty acid induced diarrhea. In: *Secretory Diarrhea* (Field M, Fordtran JS, Schultz SG, eds). American Physiological Society, 1980, pp. 159–178.
24. Snape WJ, Shiff S, Cohen S. Effect of deoxycholic acid on colonic motility in the rabbit. *Am J Physiol* 1980;238:G321–G325.
25. Cooke AR. The control of gastric emptying and motility. *Gastroenterology* 1975;68:804–816.
26. Read NW, Krejs GJ, Read MG, Santa Ana CA, Morawski SG, Fordtran JS. Chronic diarrhea of unknown origin. *Gastroenterology* 1980;78:264–271.
27. Holgate AM, Read NW. The effect of ileal infusion of intralipid on gastrointestinal transit, ileal flow rate and carbohydrate absorption in humans after a liquid meal. *Gastroenterology* 1985;88:1005–1011.
28. Read NW, Miles CA, Fisher D, Holgate AM, Kyme ND, Mitchell MA, Roche TB, Walker M. Transit of a meal through the stomach, small intestine and colon in normal subjects and its role in the pathogenesis of diarrhea. *Gastroenterology* 1980;79:1276–1282.
29. Cann PA, Read NW, Brown C, Hobson N, Holdsworth CD. The irritable bowel syndrome (IBS) relationship of disorders in the transit of a single solid meal to symptom patterns. *Gut* 1983;24:405–411.
30. Cann PA, Read NW, Cammack I, Childs H, Holden S, Kashman R, Longmore I, Nix S, Simms N, Swallow K, Weller J. Psychological stress and the passage of a standard meal through the stomach and small intestine in man. *Gut* 1983;24:236–240.

31. Evans DF, Ballantyne KC, Pegg CA, Hardcastle JD. Abnormal motility patterns in thryrotoxicosis. *Dig Dis Sci* 1985;30:768.
32. Lundh G. Intestinal digestion and absorption after partial gastrectomy. *Acta Chir Scand* 1958; Suppl 231:1–79.
33. Bond JH, Levitt MD. Use of breath hydrogen to quantitate small bowel transit following partial gastrectomy. *J Lab Clin Med* 1977;90:30–36.
34. Madsen P, Peterson G. Postvagotomy diarrhea examined by means of a nutritional contrast medium. *Scand J Gastroenterol* 1968;3:545–552.
35. Johansson C. Studies of gastrointestinal interactions. VII. Characteristics of the absorption pattern of sugar, fat and protein from composite meals in man. A quantitative study. *Scand J Gastroenterol* 10:33–42.
36. Holgate AM, Read NW. Effect of metoclopramide on gastrointestinal transit, ileal flow rate and carbohydrate absorption in humans following a liquid meal. *Br J Clin Pharmacol* 1985;19:67–74.
37. Collin J, Kelly KA, Phillips SF. Increased canine jejunal absorption of water, glucose and sodium with intestinal pacing. *Am J Dig Dis* 23:1121–1124.
38. Collin J, Kelly KA, Phillips SF. Absorption from the jejunum is increased by forward and backward pacing. *Br J Surg* 1979;66:489–492.
39. Bjork S, Phillips SF, Kelly KA. Mechanism of enhanced intestinal absorption with electrical pacing. *Gastroenterology* 1984;86:1029.
40. Holgate AM, Read NW. The relationship between small bowel transit time and absorption of a solid meal: influence of metoclopramide, magnesium sulfate and lactulose. *Dig Dis Sci* 1983;28:812–819.
41. Launiala K. The effect of unabsorbed sucrose or mannitol induced accelerated transit of absorption in the human small intestine. *Scand J Gastroenterol* 1969;4:25–32.
42. Schiller LR, Davis GR, Santa Ana CA, Morawski SG, Fordtran JS. Studies of the mechanism of antidiarrheal effect of codeine. *J Clin Invest* 1982;70:999–1008.
43. Schiller LR, Santa Ana CA, Morawski SG, Fordtran JS. Mechanism of the antidiarrheal effect of loperamide. *Gastroenterology* 1984;86:1475–1480.
44. Schiller LR, Santa Ana CA, Morawski SG, Fordtran JS. Studies on the antidiarrheal order of clonidine. Effects on motility and intestinal absorption. *Gastroenterology* 1986;89:982–988.
45. Mathias JR, Carlson GM, DiMarino AJ et al. Intestinal myoelectrical activity in response to live *Vibrio cholerae* and cholera enterotoxin. *J Clin Invest* 1976;58:91–96.
46. Cassuto J, Jodal M, Tuttle R, Lundgren O. On the role of transmural nerves in the pathogenesis of cholera toxin-induced intestinal secretion. *Scand J Gastroenterol* 1982;16:337.
47. Burns TW, Mathias JR, Carlson GM, Martin JL, Shields RP. Effect of toxigenic *Escherichia coli* on myoelectric activity on small intestine. *Am J Physiol* 1978;235:E311.
48. Jodal M, Lundgren O. Effects on enterotoxins from *Vibrio cholerae* and *Escherichia coli* on intestinal fluid and electrolyte transport. In: *The Relationships between Intestinal Motor Activity and Epithelial Transport* (Read NW, ed). Janssen Research Foundation, Beerse, Belgium, 1987.
49. Mathias JR, Carlson GM, Bertiger G, Cohan S. The effect of cholera toxin on ileal myoelectric activity: A neural-hormonal mechanism. In: *Proceedings of the Fifth Internal Motility Symposium* (Vantrappen G, ed). Harentals, Belgium: Typoff, 1976, pp. 219–266.
50. Cassuto J, Jodal M, Lundgren O. The effect of nicotinic and muscarinic receptor blockade on cholera toxin induced intestinal secretion in rats and cats. *Acta Physiol Scand* 1982;114:573.
51. Cassuto J, Siewert A, Jodal M, Lundgren O. The involvement of intramural nerves in cholera toxin induced intestinal secretion. *Acta Physiol Scand* 1983;117:195–201.
52. Nilsson O, Cassuto J, Larsson P-A, Jodal M, Lidberg P, Ahlman H, Dahlstrom A, Lundgren

O. 5-Hydroxytryptamine and cholera secretion: A histochemical and physiological study in cats. *Gut* 1983;24:542–548.
53. Cassuto J, Jodal M, Tuttle R, Lundgren O. 5-Hydroxytryptamine and cholera secretion: Physiological and pharmacological studies in cats and rats. *Scand J Gastroenterol* 1982;17:695.
54. Cassuto J, Fahrenkrug J, Jodal M, Tuttle R, Lundgren O. Release of vasoactive intestinal polypeptide from the cat small intestine exposed to cholera toxin. *Gut* 1981;22:958.
55. Mathias JR, Martin JL, Burns TW, Carlson GM, Shields RP. Ricinoleic acid effect on the electrical activity of the small intestine in rabbits. *J Clin Invest* 1978;61:640.
56. Atchison WD, Stewart JJ, Bass P. A unique distribution of laxative-induced spike potentials from the small intestine of the dog. *Am J Dig Dis* 1978;23:513.
57. Mathias JR, Sninsky CA, Martin JL, Fernandez A. The migrating action potential complex in a human subject: A case of surreptitious laxative abuse diagnosed by recording probe. *Clin Res* 1983;31:655A.
58. Coremans G, Chaussade S, Janssens J, Van Trappen G. Migrating action potential complexes, a motility pattern associated with diarrhea in man. *Dig Dis Sci* 1985b;30:765.
59. Mathias JR, Svinsky CA, Martin JL, Fernandez A. MAPC in a human subject, a case of surreptitious laxative abuse, diagnosed by recording probe. *Clin Res* 1983;31:655a.
60. Caren JF, Meyer JH, Grossman MI. Canine intestinal secretion during and after rapid distension of the small bowel. *Am J Physiol* 1974;227:183–188.
61. Fuller J, Hardcastle J, Hardcastle PT. Effect of intraluminal pressure on electrical activity of rat ileal mucosa. *J Physiol* 1980;302:12.
62. Read NW, Smallwood RH, Levin RJ et al. Relationship between changes in intraluminal pressure and transmural potential difference in the human and canine jejunum in vivo. *Gut* 1977;18:141–151.
63. Ruppin H, Kachel G, Soergel KH. The association of motor events with propulsion. In: *Intestinal Absorption and Secretion* (Skandhauge E, Heintze K, eds). Lancaster: MTP, 1983, pp. 141–151.
64. Kazic T, Varagic VM. Effect of increased intraluminal pressure on the release of actycholine from the isolated guinea pig ileum. *Br J Pharmacol Chemother* 1968;32:185–192.
65. Burke TF, Lang JP. 5-Hydroxytryptamine release into dog intestinal vasculature. *Am J Physiol* 1966;211:619–625.
66. Yagasaki O, Susuki H, Sohji J. Effects of loperamide on acetylcholine and prostaglandin release from isolated guinea pig ileum. *Jap J Pharmacol* 1978;28:873–883.
67. Beubler E, Juan H. PGE-released blood flow and transmucosal water movement after mechanical stimulation of the rat jejunal mucosa. *Naun-Schmeide Arch Pharmacol* 1978;305:91–95.
68. Greenwood B, Read NW. The effect of vagal stimulation on jejunal fluid transport, transmural potential difference and intraluminal pressure in the anesthetized ferret. *Am J Physiol* 1986;12:G651–G654.
69. Greenwood B, Read NW. The neural control of jejunal and ileal motility and transmural potential difference in the ferret. *Can J Physiol Pharmacol* 1986;64:180–187.
70. Paton WDM. The action of morphine and related substances on contraction and on acetylcholine output of co-axially stimulated guinea pig ileum. *Br J Pharmacol* 1957;12:119–127.
71. Van Nueten JM, Janssen PAJ, Fontaine J. Loperamide, a novel type of antidiarrheal agent. *Arzneim-Forshng Drug Res* 1976;24:1641–1645.
72. Schultz R, Wuster M, Herz A. Centrally and periferally mediated inhibition of intestinal motility by opioids. *Naum-Schmeide Arch Pharmacol* 1979;308:255–260.
73. Weisbrodt NW, Sussman SE, Stewart JJ, Burks TF. Effect of morphine sulfate on intestinal transit and myoelectric activity of the small intestine of the rat. *J Pharmacol Exp Ther* 1980;214:333–338.

74. Pruitt DB, Grubb MN, Jaquette DC, Burns TF. Intestinal effects of 5-hydroxytryptamine and morphine in guinea pigs, dogs, cats and monkeys. *Eur J Pharmacol* 1974;26:298–305.
75. Bass P, Wiley JN. Effect of ligation and morphine on electrical and motor activity of dog duodenum. *Am J Physiol* 1965;208:408–413.
76. Gillan MGC, Pollock D. Acute effects of morphine and opioid peptides on the motility and response of rat colon to electrical stimulation. *Br J Pharmacol* 1980;68:381–392.
77. Burks TF, Long JP. Release of intestinal 5-hydroxytryptamine by morphine and related agents. *J Pharmacol Exp Ther* 1967;156:267–276.
78. Kreuger H. The action of morphine on the digestive tract. *Physiol Rev* 1931;19:618–645.
79. Adler HF, Atkinson AJ, Ivy AC. Effect of morphine and dilaudid on the ileum and of morphine dilaudid and atrophine on the colon of man. *Arch Intern Med* 1942;69:974–985.
80. Painter NS, Truelove SC. The intraluminal pressure patterns in diverticulosis of the colon. The effect of morphine. *Gut* 1964;5:207–213.
81. Garrett JM, Sauer WG, Moertel CG. Colonic motility in ulcerative colitis after opiate administration. *Gastroenterology* 1967;53:93–100.
82. Ruppin H, Kachel G, Soergel KH. The association of motor events with propulsion. In: *Intestinal Absorption and Secretion* (Skadhauge E, Heintze K, eds). Lancaster: MTP, 1983, p. 141.
83. Sandhu B, Tripp JH, Candy DCA, Harries JT. Loperamide inhibits cholera toxin induced small intestinal secretion. *Lancet* 1979;1:689–690.
84. Hughes S, Higgs NB, Turnberg LA. Loperamide has antisecretory activity in the human jejunum in vivo. *Gut* 1984;25:931–935.
85. Farack VM, Laeshke K. Inhibition by loperamide of deoxycholic acid induced intestinal secretion. *Naun-Schmeid Arch Pharmacol* 1984;325:286–289.
86. Moriarty KJ, Rolston DDK, Kelly MH, Shield M, Clark M. Nufenoxole, a new antidiarrheal agent, inhibits fluid secretion in the human jejunum. *Gut* 1985;26:75–80.
87. Chang EB, Field M, Miller RJ. Alpha$_2$ adrenergic receptor regulation of ion transport in rabbit ileum. *Am J Physiol* (*Gastrointest Liver Physiol*) 1982;242:G237–G242.
88. Field M, McColl I. Ion transport in rabbit ileal mucosa, III. Effects of catecholamines. *Am J Physiol* 1973;225:852–857.
89. Nakaki T, Nakadate T, Yamazmoto S, Kato K. Alpha$_2$ adrenergic inhibition of intestinal secretion induced by prostaglandin E_1, vasoactive intestinal peptide and dibutyryl cyclic AMP in rat jejunum. *J Pharmacol Exp Ther* 1982;220:637–641.
90. Edwards CA, Read NW. The effect of lidamidine, a proposed alpha$_2$ receptor agonist, on salt and water transport in the human jejunum. *Dig Dis Sci* 1986;31:817–821.
91. Chang EB, Field M, Miller RJ. Enterocyte alpha$_2$ adrenergic receptors: Yohimbine and p-aminoclonidine binding relative to ion transport. *Am J Physiol* (*Gastrointest Liver Physiol* 7) 1983;244:G76–G82.
92. Mathias JR, Clench MM, Davis RH, Sninsky CA, Pineiro-Carrero VM. Migrating action potential complex, unmasked by 6-hydroxydopamine. *Am J Physiol* 1985;12:G416–G420.
93. Debongnie JC, Phillips SF. Capacity of the human colon to absorb fluid. *Gastroenterology* 1978;74:698–703.
94. Sanders DR, Wiggins HS. Conservation of mannitol, lactulose and raffinose by the human colon. *Am J Physiol* 1981; G397–G402.
95. McNeil NI, Cummings JA, James WPT. Short chain fatty acid absorption by the human large intestine. *Gut* 1978;19:819–822.
96. Ruppin H, Bar-Meir S, Soergel KH, Wood CM, Schmitt MG. Absorption of short chain fatty acids by the colon. *Gastroenterology* 1980;78:1500–1507.
97. Read NW. Diarrhea: Failure of colonic salvage. *Lancet* 1982;2:481–483.

98. Spiller RC, Brown ML, Phillips SF. Segmental colonic function in experimental steatorrhoea. Decreased capacities of the proximal colon. *Gut* 1985;26:A1136–A1137.
99. Ritchie J. In: The GI tract in stress and psychological disorder (Almy TP, Fielding JF, eds). *Clin Gastroenterol* 1977;6:622–631.
100. Whitehead WE, Engel BT, Schuster MM. Irritable bowel syndrome: physiological and psychological differences between diarrhea predominant and constipation predominant patients. *Dig Dis Sci* 1980;404–412.
101. Schuster MM. Anorectal disorder in the irritable bowel syndrome. In: *Irritable Bowel Syndrome* (Read NW, ed). London: Grune and Stratton, 1985, pp. 191–200.
102. Narducci R, Bassatti G, Gaburni M, Farroni F, Marelli A. Nifedipine reduces the colonic motor response to eating in patients with the irritable colon syndrome. *Am J Gastroenterol* 1985;80:317–319.
103. Read MG, Read NW, Barber DC, Duthie HL. Effects of loperamide on anal sphincter function in patients complaining of chronic diarrhea with fecal incontinence and urgency. *Dig Dis Sci* 1982;29:239–247.
104. Harford WV, Kregs GJ, Santa Ana C, Fordtran JS. Acute effect of diphenoxylate with atropine (lomotil) on patients with chronic diarrhea and fecal incontinence. *Gastroenterology* 1980;78:440–443.
105. Palmer KR, Corbett CL, Holdsworth CD. Double blind cross over study comparing loperamide, codeine and kiphenoxylate in the treatment of chronic diarrhea. *Gastroenterology* 1980;79:1272–1275.
106. Dupont HL, Hornick RS. Adverse effect of lomotil therapy in Shigellosis. *J Am Med Assoc* 1973;226:1525–1528.
107. Rao SSC, Read NW, Bruce C, Brown C, Holdsworth CD. Abnormalities in bowel transit in patients with ulcerative colitis. *Gastroenterology* 1987;93:1013–1019.
108. Bank S, Saunders SJ, Marks IN, Novis BM, Barbezat GO. Gastrointestinal and hepatic disease. In: *Drug Treatment Principles and Practice of Clinical Pharmacology and Therapeutics* (Avery GS, ed). Sydney: ADIS Press; Edinburgh: Churchill Livingstone, 1976, pp. 520–539.
109. Economou G, Ward-McQuiad JN. A crossover comparison of the effects of morphine, pathidine, pentazocine and phenazocine on biliary pressure. *Gut* 1971;12:218–221.

15

Sexually Transmitted Intestinal Diseases

Linda Rabeneck

1. INTRODUCTION

Sexual transmission of enteric pathogens was first suggested by Most in 1968, when he described several cases of enteric infection occurring in homosexual men in New York City in a paper entitled "Manhattan: A tropic isle?"[1] Subsequently, numerous studies have documented the increased prevalence of both enteric parasites and bacteria in homosexual men.[2–7]

In 1981, the Centers for Disease Control alerted the medical community to the unprecedented occurrence of Kaposi's sarcoma and *Pneumocystis carinii* pneumonia in homosexual men.[8] Subsequently, human immunodeficiency virus (HIV) was established as the agent that causes this acquired immunodeficiency syndrome (AIDS) and that it is transmitted sexually. It has been suggested that synergism exists between HIV and intestinal parasites, the latter promoting the development of AIDS in HIV-infected individuals.[9] There is no evidence, however, that these agents act as cofactors in the development of AIDS. Rather, like anal receptive intercourse, they are an exposure factor for HIV infection.[10,11]

The gastrointestinal tract may be involved in acute (primary) HIV infection[12] and is a major target organ in AIDS. The gastrointestinal manifestations of AIDS is presented in a subsequent paper.[13] This chapter will focus on the common sexually transmitted intestinal infections, excluding HIV infections.

Linda Rabeneck • Department of Medicine, University of British Columbia, and St. Paul's Hospital, Vancouver, British Columbia, Canada V6Z 1Y6; *present address:* Robert Wood Johnson Clinical Scholars Program, Yale University School of Medicine, New Haven, Connecticut 06510.

2. RISK FACTORS

The increased risk of acquiring sexually transmitted diseases in homosexual men is due to a number of factors. These include multiple sexual partners, anonymous partners, and specific sexual practices such as direct oral-anal contact and rectal intercourse. Oral-anal contact may be indirect as well, e.g., fellatio with a partner whose penis has been contaminated by previous anal intercourse.[14] In addition, many of the infectious agents affecting homosexual men may be asymptomatic and therefore serve as reservoirs of infection while escaping diagnosis and therapy.

In order to diagnose these infections, the physician must be aware of (1) the high prevalence of such infections in homosexual and bisexual men, (2) the patient's sexual preferences and, to a lesser extent, sexual practices, (3) the spectrum of infections, and (4) an approach to their diagnosis and treatment.[14] The more common sexually transmitted intestinal infections in homosexual men can be grouped into three categories: proctitis, proctocolitis, and enteric infection (also called enteritis).[15]

3. ANORECTAL INFECTION OR PROCTITIS

Symptoms associated with proctitis include tenesmus, mucopurulent or bloody rectal discharge, anorectal pain, and constipation. At sigmoidoscopy the abnormality is confined to the distal 15 cm of the bowel (i.e., rectum). The gross lesion may include increased mucus, cotton swab friability, or mucosal ulcerations. The commonest causes of proctitis are *Neisseria gonorrhoeae,* herpes simplex virus, *Chlamydia trachomatis,* and *Treponema pallidum.*

3.1. Neisseria gonorrhoeae

Anorectal gonorrhea may be symptomatic or asymptomatic. If symptoms are present they often consist of mild anorectal pain, itching, and mucopurulent discharge. At sigmoidoscopy the mucosa may be normal or show gross features of proctitis (described above). Gram stain of rectal exudate may yield the diagnosis, but its sensitivity is less than 50%.[16] Culture is the diagnostic test.

The treatment of gonococcal proctitis is aqueous penicillin G 4.8 million units I.M. plus 1 g of probenicid by mouth. Penicillinase-producing *N. gonorrhoeae* (PPNG) has been described. The incidence varies depending on geographic location. Physicians must be aware of the prevalence of these strains in their own communities. The therapy of PPNG proctitis is spectinomycin 2 g I.M.

All known sexual contacts of patient should have rectal cultures. The pa-

tient should be advised to abstain from sexual activity until cure is documented by a negative culture.

3.2. Herpes Simplex Virus (HSV)

HSV (usually type 2) is the most common cause of nongonococcal proctitis in male homosexuals,[16] causing symptoms more commonly than other rectal infections. It has, however, been cultured from the rectum in asymptomatic homosexual men.[17] While the symptoms of HSV proctitis may be the same as in other infectious causes of acute proctitis, the severity of rectal pain may be greater.[17] Systemic symptoms may be present, including malaise and fever. In addition, neurological symptoms in the distribution of the sacral nerve roots may occur. These include difficulty in initiating micturition, impotence, and posterior thigh pain or paresthesiae of the buttock or perineal region.[17] The pathogenesis of these neurological symptoms is not fully understood. Inguinal lymphadenopathy may occur.[8] In addition to the sigmoidoscopic findings of proctitis, HSV proctitis may be associated with external perianal or anal ulcerative or vesicular lesions.[17] A mucosal lesion extending to 20 cm (i.e., distal sigmoid colon) has been described in one patient.[17] Cytological scrapings taken from the base of a vesicle or ulcer may reveal intranuclear inclusion bodies or multinucleated giant cells. Rectal biopsies may reveal multinucleated cells, intranuclear inclusions, and lymphocytic infiltration around submucosal vessels.[17] Viral isolation by culture of the rectal exudate establishes the diagnosis.

The mean duration of symptoms in HSV proctitis is 21 days.[17] The natural history of recurrences of HSV proctitis is unclear. Oral acyclovir in a dose of 200 mg by mouth five times daily for 10 days has been demonstrated to reduce virus shedding, new lesion formation, and the duration of lesions in both men and women with first episodes of genital HSV infection.[18] Acyclovir is therefore being used to the treat herpectic proctitis.

3.3. Chlamydia trachomatis

Chlamydia trachomatis has two subtypes: lymphogranuloma venereum (LGV) and non-LGV agents. The subtypes are further separated into several serotypes. The severity of disease depends in part on the serotype of the infecting strain. Infections with non-LGV strains produce mild symptoms of proctitis or no symptoms at all. Sigmoidoscopy may be normal or may reveal changes consistent with proctitis.[19] LGV infections produce more severe inflammation which can resemble Crohn's disease, clinically and histologically.[20,21] The lesion can extend proximally into the descending colon. Fistulae, perirectal abscess, and strictures have been reported in these patients.

The diagnosis can be confirmed by culture for *C. trachomatis*. Microim-

munofluorescence tests are also available. Rectal biopsies may demonstrate a diffuse inflammatory infiltrate with plasma cells, round cells, and polymorphonuclear leukocytes. Granulomata and giant cells may be present.[16] Because of the similarity of some of these histological findings to those of Crohn's disease, caution should be used in establishing a diagnosis of inflammatory bowel disease in homosexual or bisexual men.[22]

Infection with *C. trachomatis* responds to tetracycline or doxycycline. Non-LGV infections may be treated with tetracycline 500 mg by mouth four times a day for 1 week, whereas LGV infections should be treated with the same dosage for at least 2 weeks.

3.4. *Treponema pallidum*

Primary anorectal syphilis can be asymptomatic or may cause symptoms of proctitis. Anorectal chancres usually occur 2–6 weeks after exposure to the agent. Secondary syphilis may cause erythematous nodular or indurated rectal mucosal lesions.[15] Dark-field examination of fluid from chancres will reveal the presence of the organism. Serology also will assist in the diagnosis. The organism can be found in biopsy tissue by appropriate silver stains. Immunofluorescence using specific antibody against *T. pallidum*, however, is a more reliable method.

The therapy of early syphilis is penicillin G benzathine 2.4 million units I.M. Tetracycline 500 mg by mouth four times a day for 15 days may be used to treat penicillin-allergic patients. All sexual contacts should be examined for evidence of the disease and have serology performed.

4. ANOTHER CAUSE OF ANORECTAL INFECTION: CONDYLOMATA ACUMINATA

Condylomata acuminata or venereal warts are caused by human papilloma virus and are common in homosexual men.[23] The lesions, either single or confluent, occur in the perianal area, the anal canal, or at both sites. They are usually painless.

The diagnosis is established on the basis of the gross appearance of the lesions. Topical application of podophyllin may be used to treat these lesions, although recurrence is common.

5. PROCTOCOLITIS

The symptoms of proctocolitis may include some of the symptoms of proctitis, i.e., anorectal pain, mucopurulent or bloody rectal discharge, and

tenesmus, plus diarrhea. In addition, proctocolitis may be associated with abdominal cramps. Sigmoidoscopy reveals the mucosal lesion to extend above 15 cm (i.e., above the rectum). The mucosal lesion may include increased mucus, cotton swab friability, or mucosal ulceration. The commonest causes of proctocolitis include *Campylobacter, Shigella, C. trachomatis* (LGV-type), and *Entamoeba histolytica.*[15] Nontyphoidal *Salmonella* infections have not been reported to occur commonly in otherwise healthy homosexual men,[15] but typhoid fever can be sexually transmitted.[24]

5.1. Shigella

Shigellosis usually presents with an abrupt onset of fever, diarrhea, and lower abdominal cramps. The stools may be watery or contain mucus or blood. Sigmoidoscopic examination reveals a mucosal lesion consistent with proctocolitis. Stool cultures establish the diagnosis.

The illness is usually self-limited. Therapy is supportive but antibiotics should be considered if the illness is particularly severe. Trimethoprim-sulfamethoxazole 160/800 mg (double strength) taken twice a day by mouth or ampicillin 500 mg by mouth four times daily for 7 days may be prescribed.

5.2. Campylobacter

Campylobacter jejuni is the commonest species to cause proctocolitis in homosexual men.[25] Infection with *Campylobacter*-like organisms has also been described.[26] *Campylobacter* infections may be asymptomatic or may present with symptoms of proctocolitis.

Sigmoidoscopic examination is usually abnormal (as described above). The diagnosis is established culturing the organism on selective media in microaerophilic atmosphere.

Campylobacter infection is usually self-limited. Thus the need for antimicrobial therapy is not established, although erythromycin 500 mg by mouth four times daily for 1 week is effective.[27]

5.3. Entamoeba histolytica

In the first controlled study of the prevalence of parasitic infections in homosexual men, *Entamoeba histolytica* was identified in the stools of 27% of homosexual men.[28] In that study the presence of symptoms did not correlate with infection. Other studies have confirmed this high prevalence and have also failed to find an association between the symptoms and the presence of the parasite in the stool.[29] Proctocolitis may be caused by *E. histolytica,* however.[15,30,31]

Recent evidence indicates that although this agent is prevalent in the stools of homosexual men, it rarely causes invasive disease.[32,33] With the use of isoenzyme electrophoresis, *E. histolytica* isolates from homosexual man have been shown to be nonpathogenic in type,[33] probably accounting for the lack of disease. It is unclear at present if, in the absence of mucosal disease, homosexual men with *E. histolytica* in their stools should receive therapy.

5.4. Nonpathogenic Protozoa

Endolimax nana, Entamoeba hartmanni, Entamoeba coli, Iodamoeba buetschlii, and *Dientamoeba fragilis* have also been isolated from the stools of homosexual men. In one study they were found in 30% of homosexual men.[28] There is no evidence to date that these agents cause disease. Further studies are necessary to determine if therapy is indicated for these agents.

6. ENTERIC INFECTION

The symptoms of enteric (i.e., small-bowel) infection may include diarrhea, abdominal pain, bloating, and nausea. Tenesmus and bloody or mucopurulent rectal discharge are usually absent. Sigmoidoscopic examination reveals a normal rectum.

Giardia lamblia is the most common agent associated with small-bowel infection[15]; it is found in the stools of 13% of homosexual men.[28] The presence of trophozoites or cysts in stool samples, duodenal aspirates, or duodenal biopsy specimens establishes the diagnosis. Metronidazole 250 mg by mouth three times a day for 10 days is the recommended therapy.[34] It is important to also screen the stools of any regular sexual partners for infestation with this agent.

7. OTHER INFECTIOUS AGENTS

The prevalence of several other infections is increased in homosexual men. Although infection with these agents does not usually present with any of the clinical syndromes discussed previously, they deserve mention at this point.

Both hepatitis A virus and hepatitis B virus are sexually transmitted in homosexual men.[35,36] The initial prevalence of cytomegalovirus seropositivity in homosexual men was 87% in one study.[37] The subsequent attack rate among susceptible men was 71% in 9 months, but no clinical illness was identified with cytomegalovirus seroconversion.

Spirochetosis has been demonstrated in rectal biopsy specimens in 30% of homosexual men.[38] The organism is a spirochete, the classification of which is not clear at present. The significance of this finding is uncertain as no associa-

tion of spirochetosis with clinical or histological findings has been demonstrated.

8. SUMMARY

A number of microbial agents are sexually transmitted in homosexual man and cause a variety of illnesses. Physician awareness—of both the sexual preference of the patients and the spectrum of illness caused by these agents—is the first and crucial step toward the diagnosis of these infections. Next comes the history, physical examination, and sigmoidoscopic exam. These steps are usually sufficient to determine whether the patient has proctitis, proctocolitis, or an enteric infection. Knowledge of the likely causative organisms in each of these cases leads to the appropriate laboratory studies that can successfully establish the diagnosis.

REFERENCES

1. Most H. Manhattan: "A tropic isle?" *Am J Trop Med Hygiene* 1968;17:333–354.
2. Kean BH. Venereal amebiasis. *NY State J Med* 1976;76:930–931.
3. Drusin LM, Genvert G, Topf-Olstein B, Levy-Zombek E. Shigellosis. Another sexually transmitted disease? *Br J Vener Dis* 1976;52:348–350.
4. Bader M, Pedersen AHB, Williams R, Spearman J, Anderson H. Venereal transmission of shigellosis in Seattle-King County. *Sex Transm Dis* 1977;4:89–91.
5. William DC, Felman YM, Marr JS, Shookhoff HB. Sexually transmitted enteric pathogens in a male homosexual population. *NY State J Med* 1977;77:2050–2052.
6. Schmerin MJ, Jones TC, Klein H. Giardiasis: Association with homosexuality. *Ann Intern Med* 1978;88:801–803.
7. Markel EK, Havens RF, Kuritsubo RA, Wingerd J. Intestinal protozoa in homosexual men of the San Francisco Bay area: prevalence and correlates of infection. *AM J Trop Med* 1984;32:239–245.
8. Centers for Disease Control. Kaposi's sarcoma and *Pneumocystis* pneumonia among homosexual men: New York City and California. *Morbid Mortal Week Rep* 1981;25:305–308.
9. Pearce RB, Abrams DI. *N Engl J Med* 1987;316:690–691 (letter).
10. Jeffries E, Willoughby B, Boyko WJ et al. The Vancouver Lymphadenopathy–AIDS Study. 2. Seroepidemiology of HTLV-III antibody. *Can Med Assoc J* 1985;132:1373–1377.
11. Boyko WJ, Schechter MT, Craib KJP et al. The Vancouver Lymphadenopathy–AIDS Study. 5. Antecedent behavioural, clinical and laboratory findings in patients with AIDS and HIV-serpositive controls. *Can Med Assoc J* 1986;135:881–887.
12. Rabeneck L, Popovic M, McLean DM, McLeod WA, Read E, Wong KK, Boyko WJ. Esophageal ulcers in acute HIV infection. Presented at the Canadian Society for Clinical Investigation Annual Meeting, September 15, 1987, Winnipeg, Manitoba.
13. Rabeneck L. Gastrointestinal manifestations of the acquired immune deficiency syndrome. In: *Modern Concepts in Gastroenterology,* Vol 3. (Shaffer EA, Thomson ABR, eds.) New York: Plenum, 1989 (in press).
14. Rompalo AM, Stamm WE. Anorectal and enteric infections in homosexual men. *West J Med* 1985;142:647–652.

15. Quinn TC, Stamm WE, Goodell SE et al. The polymicrobial origin of intestinal infections in homosexual men. *N Engl J Med* 1983;309:576–582.
16. Quinn TC, Corey L, Chaffee RG, Schuffler MD, Brancato FP, Holmes KK. The etiology of anorectal infections in homosexual men. *Am J Med* 1981;71:395–406.
17. Goodell SE, Quinn TC, Mkrtichian E, Schuffler MD, Holmes KK, Corey L. Herpes simplex virus proctitis in homosexual men. *N Engl J Med* 1983;308:868–871.
18. Bryson YJ, Dillon M, Lovett M et al. Treatment of first episodes of genital herpes simplex virus infection with oral acyclovir. *N Engl J Med* 1983;308:916–921.
19. Quinn TC, Goodell SE, Mkrtichian E et al. *Chlamydia trachomatis* proctitis. *N Engl J Med* 1981;305:195–200.
20. Klotz SA, Drutz DJ, Tam MR, Reed KH. Hemorrhagic proctitis due to lymphogranuloma venereum serogroup L2. *N Engl J Med* 1983;308:1563–1565.
21. Levine JS, Smith PD, Brugge WR. Chronic proctitis in male homosexuals due to lymphogranuloma venereum. *Gastroenterology* 1980;79:563–565.
22. Surawicz CM, Goodell SE, Quinn TC et al. Spectrum of rectal biopsy abnormalities in homosexual men with intestinal symptoms. *Gastenterology* 1986;91:651–659.
23. Sohn N, Robilotti JG. The gay bowel syndrome. *Am J Gastroenterol* 1977;67:478–484.
24. Dritz SK, Braff EH. Sexually transmitted typhoid fever. *N Engl J Med* 1977;296:1359–1360.
25. Quinn TC, Goodell SE, Fennell C et al. Infections with *Campylobacter jejuni* and *Campylobacter*-like organisms in homosexual men. *Ann Intern Med* 1984;101:187–192.
26. Totten PA, Fennell CL, Tenover FC et al. *Campylobacter cinaedi* (sp. nov.) and *Campylobacter fennelliae* (sp. nov.): Two new *Campylobacter* species associated with enteric disease in homosexual men. *J Infect Dis* 1985;151:131–139.
27. Anders BJ, Lauer BA, Paisley JW, Reller LB. Double-blind placebo controlled trial or erythromycin for treatment of *Campylobacter* enteritis. *Lancet* 1982;1:131–132.
28. Keystone JS, Keystone DL, Proctor EM. Intestinal parasitic infections in homosexual men: Prevalence, symptoms and factors in transmission. *Can Med Assoc J* 1980;123:512–514.
29. Phillips SC, Mildvan D, William DC, Gelb AM, White MC. Sexual transmission of enteric protozoa and helminths in a venereal-disease clinic population. *N Engl J Med* 1981;305:603–606.
30. Saltzberg DM, Hall-Craggs M. Fulminant amebic colitis in a homosexual man. *Am J Gastroenterol* 1986;81:209–212.
31. Burnham WR, Reeve RS, Finch Rg. Entamoeba histolytica infection in male homosexuals. *Gut* 1980;21:1097–1099.
32. Allason-Jones E, Mindel A, Sargeaunt P, Williams P. Entamoeba histolytica as a commensal intestinal parasite in homosexual men. *N Engl J Med* 1986:315:353–356.
33. Goldmeier D, Sargeaunt PG, Price AB et al. Is *Entamoeba histolytica* in homosexual men a pathogen? *Lancet* 1986;1:641–644.
34. Drugs for parasitic infections. *Med Lett* 1986;28:9–18.
35. Corey L, Holmes KK. Sexual transmission of hepatitis A in homosexual men. *N Engl J Med* 1980;302:435–438.
36. Dietzman DE, Harnisch JP, Rag CG, Alexander ER, Holmes KK. Hepatitis B surface antigen (HBsAg) and antibody to HBsAg: Prevalence in homosexual and heterosexual men. *J Am Med Assoc* 1977;238:2625–2626.
37. Mintz L, Drew WL, Miner RC, Brafaf EH. Cytomegalovirus infections in homosexual men. *Ann Intern Med* 1983;99:326–329.
38. Surawicz CM, Roberts PL, Rompalo A, Quinn TC, Holmes KK, Stamm WE. Intestinal spirochetosis in homosexual men. *Am J Med* 1987;82:587–592.

16

Intestinal Gas

Michael D. Levitt

1. INTRODUCTION

Excessive intestinal gas is thought to cause a variety of abdominal symptoms. Such gas problems clearly cannot be attributed to recent alterations in diet or environment since the first known treatise on abdominal gas, *The Winds*, was authored by Hippocrates over 2000 years ago. While the intervening years have not resulted in great strides in our therapy of gas problems, our understanding of the etiology of these complaints has been enhanced by studies carried out over the past 20 years. This chapter reviews recent advances in our understanding of the origin of gaseous complaints and the present, unsatisfactory state of treatment of this problem.

2. PHYSIOLOGY OF GASTROINTESTINAL GAS

Gas can enter the gut via three routes: (1) air swallowing, (2) diffusion from the blood, and (3) production within the lumen. The relative contribution of each of these three mechanisms has not been clearly defined.

Studies in which the stomach is constantly aspirated via a nasogastric tube would suggest that enormous quantities of air are swallowed each day.

For example, every swallow is said to deposit several milliliters of air in the stomach, and in the process of drinking liquids, a volume of air equivalent to that of the liquid can be swallowed.[1] In such studies, a minimum of several

Michael D. Levitt • Research Service, Veterans Administration Medical Center, and Department of Medicine, University of Minnesota, Minneapolis, Minnesota 55417.

thousand milliliters of gas is aspirated from the stomach over a 24-hr period. As will be discussed, virtually all swallowed N_2 that enters the small bowel should subsequently be passed via the rectum. In fact, N_2 is probably added to the lumen as gas passes through the intestine. Measurements of N_2 passed in flatus are usually in the range of 400–600 ml/day[2] and seldom reaches the level of several thousand milliliters per day, as predicted by gastric aspiration studies. Thus, it seems likely that the presence of the nasogastric tube, a requirement in gastric aspiration studies, enhances the tendency to swallow air or that much of the air that is normally swallowed is subsequently eructated and never enters the intestine.

Various maneuvers can drastically increase the volume of gas that is swallowed as well as the fraction of this gas that enters the duodenum. Large volumes of air can be "swallowed" if the subject relaxes the upper esophageal sphincter simultaneous with the production of a negative intrathoracic pressure. Such a maneuver sucks air into the esophagus and some of this air may subsequently be propelled by peristalsis into the stomach. Laryngectomized patients consciously aspirate air in this fashion in the process of producing esophageal speech. Surprisingly, a sizable fraction of healthy individuals commonly perform this maneuver, often as a nervous habit in an attempt to stimulate an eructation. A prodigious volume of gas can be swallowed in this fashion. I studied a patient who had had a Nissen repair of a hiatal hernia several years prior to undergoing a laryngectomy for carcinoma of the larynx. The patient complained of massive abdominal distension following short periods of esophageal speech. Measurement of the rate at which gas passed through the duodenum (assessed via an intestinal tube) during esophogeal speech showed that up to 200 ml/min of swallowed air was entering the intestinal tract. Presumably most swallowed air is subsequently eructated, but the Nissen procedure in this patient prevented such eructation.

A factor postulated to influence the efficiency of eructation is posture. The esophagus enters the dorsosuperior aspect of the stomach. When a subject reclines in the supine position, liquid in the stomach overlaps the gastroesophageal junction forming a liquid "trap" for stomach gas.[3] The common habit of reclining in the supine posture after meals may enhance the tendency of gas to enter the small bowel. Such a trap also has been postulated to account for the high prevalence of infant colic in the Western world where infants frequently are bottle-fed while in the supine position.

Failure to appreciate the rate at which air can be swallowed and the rapidity with which air can be propelled through the gut (stomach to cecum transit time of several minutes) has led to erroneous conclusions regarding the source of the bowel gas that rapidly accumulates during various medical procedures. For example, it is not uncommon in patients undergoing urological procedures for the small bowel to fill with gas over a period of minutes. This gas has been

assumed to result from diffusion from the blood. However, such an origin seems physiologically impossible (as will be discussed in a subsequent section). Far more likely is a scenario in which these anxious patients swallow a large volume of air that cannot be readily eructated because of their supine position. This air is then rapidly propelled through the gut giving the appearance of filling from diffusion from the blood.

3. PRODUCTION AND CONSUMPTION OF GAS IN THE INTESTINAL TRACT

The two recognized mechanisms that liberate luminal gas are the interaction of bicarbonate and acid, and bacterial metabolism. The interaction of 1 meq of bicarbonate with 1 meq of acid releases about 22.4 ml of CO_2. The quantities of acid and base that may react in the duodenum suggest that there is an enormous potential for CO_2 production via such a reaction. For example, after a meal normal subjects often secrete 30 meq of HCl.[4] The interaction of this acid with bicarbonate in the duodenum could release 600 ml of CO_2. Another source of duodenal acid is the digestion of triglycerides with the release of fatty acids. After a hearty meal, up to 100 meq of acid could be formed by this mechanism yielding an additional 2200 ml of CO_2. In actuality, such gas volumes never accumulate in the duodenum because of the high solubility of CO_2 in water and the rapid absorption of this gas from the gut. However, measurements of the pCO_2 of duodenal contents after a meal has revealed values in the range of 500 mm Hg.[5] Thus, CO_2 represents about two-thirds (500/760) of the gas equilibrated with duodenal fluid and CO_2 should be the predominant gas of the upper intestinal tract after a meal.

In the ileum, Na^+ and Cl^- are thought to be absorbed by a double-exchange mechanism, H^+ for Na^+, HCO_3^- for Cl^-.[6] This secretion is followed by the interaction of H^+ and HCO^- to form CO_2 and H_2O. Thus, the pCO_2 of the ileum also tends to be elevated above blood levels because of the reaction of bicarbonate and acid.

Because the stomach and small intestine of normal subjects have low bacterial counts, the bacterial production of gases in the gut normally is limited largely to the colon.[7] However, patients with gastric outlet destruction, achlorhydria, or bacterial overgrowth of the small bowel may have appreciable bacterial production of gas in the upper gut. Multiple gases are produced by bacteria; however, only three are quantitatively important: CO_2, H_2, and CH_4. Carbon dioxide may be a direct metabolic product of oxidation by the bacterial flora or may be an indirect result of fermentation reactions that release organic acids such as acetic, propionic, and butyric acids. These acids react with bicarbonate to liberate CO_2.

Fermentation reactions of colonic bacteria also release H_2 and CH_4. Bacteria appears to be the sole source of these gases in rats, since germ-free rats excrete neither gas.[8] Excretion of these two gases is readily detected within several days of contamination of germ-free rats with feces from conventional rats. Neither H_2 nor CH_4 is excreted by the germ-free human newborns.

Studies of H_2 production in the normal human intestine have shown that virtually all H_2 is produced in the colon.[7] This colonic H_2 production is very low when subjects have fasted for long periods and increases dramatically when fermentable substrate (largely carbohydrate) is supplied to the colonic bacteria. Thus, endogenous substrates are a poor source of fermentable substrates for H_2 production in the colon. Copious H_2 production is dependent on the ingestion of nonabsorbable carbohydrates that pass through the small bowel to supply substrate for H_2-liberating reactions. Flatus CO_2 concentration usually correlates closely with flatus H_2. Thus, it appears that most colonic CO_2 similarly results from bacterial fermentation reactions in the colon.

The bacterial production of CH_4 seemingly is less dependent on ingested substrates than is the case with H_2. For example, prolonged fasting seldom reduces CH_4 production by more than 50%.[9] Thus unknown endogenous substrates serve as the source of CH_4 production in the colon.

There are huge ($\sim 10^6$) individual variations in the excretion rate of CH_4.[10] Although not intensively studied, this variability presumably reflects differences in the fecal concentration of the CH_4-producing bacteria. When CH_4 production is assessed via breath measurements, about one-third of the normal adult population are found to excrete CH_4 in concentrations well above the atmospheric concentration of about 1 ppm, while the other two-thirds excrete very little or no CH_4 above the atmospheric level. High CH_4 production appears to be a familial trait.[10]

Bacterial metabolic reactions release a number of other gases in low quantity. Of particular importance are the odoriferous gases that are readily detected in trace quantities because of the extreme sensitivity of the nose. For example, the human nose can detect H_2S in a concentration of 1 ppb. While the odoriferous gut gases have received an enormous amount of attention from the lay public, scientific attention to these gases has been limited. Recently, using relatively sophisticated technology, Moore and co-workers[11] reported that most of the offensive odor of feces represents sulfur-containing gases such as methanethiol and dimethyl disulfide. The long-standing concept that aromatic amino acid metabolites such as indole and skatole are the source of fecal odor appears to be incorrect.

It is seldom appreciated that bacteria can consume as well as produce gas and the volume excreted represents the net of these two reactions.[12] For example, when carbohydrate is added to a human fecal homogenate in a syringe, initially H_2 is produced in large quantity. However, the high pH_2 in the syringe drives H_2 consumption by other bacteria. As production slows (and H_2 con-

sumption increases), the volume of H_2 in the syringe falls dramatically. We have also carried out studies with carbon monoxide (CO) demonstrating that this gas is produced in feces when incubated aerobically while CO is consumed by feces when incubated anaerobically.[13]

4. DIFFUSION OF GASES

The diffusion of gases between the blood and the intestinal lumen is somewhat more complex than the situation in nonintestinal tissue. In most tissues, exchange of gases consists simply of O_2 diffusing from blood to tissue, and CO_2 diffusing from tissue to blood. Nitrogen tension is equal in blood and tissue and there is no net movement of N_2. Diffusion of gas in the gut is complicated by the production of large quantities of CO_2, H_2, and CH_4 in the lumen.

In the gut, as in all other tissue, the net movement of gas via diffusion depends on the partial pressure gradient. However, the gradients in the gut vary along the linear extent of the gut. For example, swallowed air has a higher pO_2 and lower pCO_2 than does blood and O_2 will diffuse into the stomach mucosa while CO_2 diffuses into the lumen. However, as this stomach gas enters the duodenum, the pCO_2 increases dramatically due to the interaction of acid with the pancreatic and intestinal bicarbonate. The direction of CO_2 diffusion now shifts and the net movement is from lumen to blood. The high duodenal CO_2 concentration also dilutes luminal N_2 and O_2, and these gases begin to diffuse from blood to lumen. The high aqueous solubility of CO_2 allows the gas to diffuse many times faster than any other gut gas and by the time gas has reached the ileum, the pCO_2 may have nearly equilibrated with the blood. However, the metabolism of the colonic bacteria results in the liberation of large quantities of CO_2, H_2, and CH_4. These gases will diffuse from lumen to blood. In addition, these gases dilute luminal N_2 and O_2 causing these gases to diffuse from blood into the lumen.

The net result of diffusion is that this process tends to reduce the volume of CO_2, H_2, and CH_4 in the gut and probably increases the volume of N_2. The pO_2 of luminal contents is consistently well below that of blood and O_2 is constantly diffusing from blood to lumen. However, O_2 consumption by epithelial cells and the intestinal bacteria render gut contents nearly anaerobic.

5. OVERVIEW OF THE PHYSIOLOGY OF INTESTINAL GAS

Flatus passed per rectum represents the net result of the three processes described above: air swallowing, intraluminal gas production and consumption, and diffusion. The rate at which flatus is passed is extremely variable but gen-

erally ranges from 500 to 2000 ml/day.[2,14] The frequency with which boluses of gas are passed per rectum was found to average about 13.5 ± 4.5[15] in a small number of healthy, young, male adults. The volume of an average bolus or rectal gas measured by a single patient was about 75 ml.

Literature values suggest that the composition of rectal gas is highly variable. While some of this variability probably reflects atmospheric contamination of the samples, much of the variability presumably reflects differences in the metabolic activity of colonic bacteria.

Five gases—N_2, O_2, CO_2, H_2, and CH_4—are the only quantitatively important gases in flatus.[16] When flatus is being passed in low volumes, N_2 is the predominant gas and pN_2 values above that of blood (about 580 mm Hg) may be observed. When large volumes of flatus are being passed, H_2 and CO_2 usually are present in high concentrations and the sum of these two gases may exceed that of N_2.[16] Thus, intraluminal production clearly may be the predominant source of rectal gas. The pO_2 of flatus is always from negligible to 20% depending on the CH_4-producing status of the individual.

The importance of swallowed air as a source of flatus cannot be determined from available data. The maximal contribution would be the N_2 plus minimal quantities of O_2 passed in flatus. However, it is not clear to what extent flatus N_2 represents swallowed N_2 versus N_2 that diffused into the gut lumen down a gradient created by the production of gas (H_2, CO_2, CH_4) by the colonic bacteria. That such diffusion may be important is supported by studies of flatus excretion after ingestion of beans, a food that markedly enhances bacterial H_2 and CO_2 production. After bean ingestion the excretion of N_2, as well as that of CO_2 and H_2, was markedly increased.[14] Since N_2 is not produced in the colon, presumably this enhanced N_2 excretion reflects increased diffusion of this gas from blood to lumen, an indirect result of the bacterial liberation of H_2 and CO_2.

The excretion rate of gas per rectum should not be considered a reflection of the volume of gas in the gut. The rapid production and excretion of gas in the colon might result in a minimal accumulation of intestinal gas while a similar volume of swallowed air slowly passing through the bowel would have a much greater effect on intestinal gas volume.

The volume of gas in the human intestinal tract has been measured by perfusing the gut rapidly with argon via a tube at the ligament of Treitz.[17] All gas washed out at the rectum was quantitatively collected and analyzed for the gases native to the gut (N_2, O_2, H_2, CH_4, and O_2). These studies indicated that the normal human small and large bowel contain a mean of 200 ml of gas, the bulk of which is usually N_2. Measurements obtained with the body plethysmograph supported the finding that the human gut contains only about 100–200 ml of gas. While such a volume, intuitively, seems low, abdominal roentgenograms showed that infusion of 500 ml of gas into the gut produced a gas

pattern that appeared to contain several fold more gas than was present prior to the infusion.[18]

6. CLINICAL GAS SYNDROME

The patient with a gaseous problem usually complains to the physician using rather nonspecific terms, i.e., "I have too much gas." In reality, this complaint usually means that the patient has abdominal bloating and distension that the patient attributes to too much gas in the intestine. Less commonly, the patient is referring to the excessive passage of gas per rectum. Least commonly, the patient has excessive eructation. Since these problems have different origins and different treatments, the physician must first determine the nature of the gaseous complaint.

7. EXCESSIVE ERUCTATION

Analysis of gas obtained from the gastric air bubble shows that N_2 and O_2 are the predominant gases. Thus excessive eructation is caused by air swallowing or, more accurately, air aspiration.

For unclear reasons, air swallowing is a very common habit. Possibly the habit derives from the mistaken belief that all sorts of thoracic and abdominal discomforts are caused by excessive gas and the remedy should be eructation. In an attempt to eructate, patients aspirate air into the esophagus and then immediately regurgitate most of this air, mistakenly believing that they have produced a true eructation that removes gas from the stomach. In other subjects air swallowing clearly represents a response to anxiety. Lastly, there are some subjects who claim to get transitory relief of abdominal or chest discomfort via air swallowing. It is possible that this maneuver could improve symptoms of reflux esophagitis by stimulating an esophageal peristaltic wave that clears the esophagus of refluxed acid and by delivering salivary bicarbonate to the lower esophagus.

In my experience, excessive eructation is usually part of a vicious cycle in which the patient senses minor abdominal discomfort, attempts to eructate, but actually swallows air. The resulting abdominal air may cause further distress with more air swallowing. If an abdominal problem (esophagitis, peptic ulcer) is present, it should of course be treated. However, most patients have no demonstrable abdominal pathology and the problem appears to be functional. The only treatment I have found to be effective is educational therapy, in which the origin of eructation is explained to the patient. The voluntary nature of eructation is best demonstrated if the physician is capable of swallow-

ing air and belching several times. It must be explained to the patient that air swallowing is aggravating, not alleviating, the problem and the urge to eructate should be resisted. In the receptive patient, an insight into the cause of the problem interrupts the cycle and the problem is alleviated. However, in many patients the air-swallowing habit is so deeply ingrained that no amount of education is effective. I know of no treatment that is beneficial to these subjects. Theoretically, and in practice, drugs that influence gastric function (H_2 blockers, antacids) are of no benefit.

Abdominal bloating and discomfort that the patient attributes to bowel gas is an extremely common complaint. Surprisingly, measurement of the volume of intestinal gas using the argon washout method showed perfectly normal volumes of bowel gas—about 200 ml—in bloating patients.[19] This quantitative result is supported by subjective estimates of intestinal gas present on abdominal roentgenogram. Most patients complaining of bloating have normal volumes of bowel gas on X-ray exam. Doctor Walter Alvarez actually attempted to predict how much gas would be present on a roentgenogram of his abdomen based on how bloated he felt at the moment. Over a several week period, he found no correlation between his sensation of bloating and the quantity of gas he observed on roentgenograms. Thus, patients are unable to predict the state of gaseous distension of their gut.

We did find in our argon perfusion studies that patients with bloating and distension poorly tolerated the rapid infusion of gas into their intestine.[19] The majority of bloating patients perceived such severe pain from the gas perfused into the gut that they demanded that the study be discontinued. In contrast, the majority of controls had no appreciable symptoms from the infusion. This response of bloating patients was not entirely psychogenic. Each infusion study was initiated by a sham infusion at which time the patient was told that the study was starting, the infusion pump was turned on, but a stopcock was adjusted so that no argon entered the gut. No appreciable pain developed during the sham infusion. The patient began to complain within minutes of argon entering the gut. We concluded from these studies that the primary problem of bloating patients is an irritability of the gut that causes the patient to perceive pain when the gut is distended by volumes of gas that were well tolerated by most healthy controls.

If bloating patients have normal volumes of bowel gas, what accounts for the distended abdomen observed in many of these subjects? As first demonstrated by Alvarez, bloating subjects can voluntarily reduce the girth of their abdomens. Patients who truly have an excess of abdominal contents (i.e., ascites, obesity) cannot so reduce their girth. Thus, it appears that the bloating patient subconsciously lowers his diaphragm and relaxes his abdominal musculature, thus creating a distended abdomen. The explanation of this phenom-

enon seems to be that the irritable loops of gut of bloating patients are more comfortable when protruding than when crammed up under the diaphragm.

Therapy for bloating and distension presumably should be directed toward correcting the underlying dysmotility or irritability rather than reducing bowel gas. Anticholinergics, which decrease motility, are commonly used for this purpose. However, controlled trials (and my own anecdotal observation) indicate that this type of drug is not very effective in bloating patients. In fact, patients often complain of increased bloating and distension while on anticholinergics.

There is some evidence that drugs that enhance gastrointestinal motility are useful in bloating subjects. Johnson *et al.*[20] found that metoclopramide improved the symptoms of "flatulent dyspepsia" better than did placebo in double-blind studies. At present, there are insufficient data to determine the possible value of cisapride, a new prokinetic drug, on symptoms of bloating.

There is little evidence that dietary manipulation, increased fiber intake, antacids, charcoal, enzyme supplements, or surface tension-reducing agents (simethicone) are helpful to bloating patients. Thus, the major therapeutic tool of the physician, as is the case with most forms of irritable bowel, is reassurance and understanding. Many of these patients are worried that their symptoms reflect some severe abdominal problem, usually cancer. Exclusion of this diagnosis and a detailed explanation of the cause of the irritable gut (motor dysfunction) often leads to a dramatic reduction in symptoms. In my experience, the patient who does not respond to reassurance seldom benefits from drug therapy.

The final clinical problem associated with gas is the passage of excessively voluminous or odoriferous gas. Unlike eructation or bloating, this complaint may actually reflect a problem with excessive gas.

Despite the lay public's enormous interest in excessive rectal gas, only one patient with this problem has received appreciable study.[15,21] This patient carefully counted the number of times he passed gas. His daily average was 35 passages per day, well above the upper limit of normal. Ingestion of several quarts of milk increased his productivity to a record-breaking 141 passages per day. His flatus contained high concentrations of H_2 and CO_2 indicating that bacterial fermentation reactions were the source of the gas. The patient was lactase-deficient and removal of lactose from his diet produced some improvement. Employing a carefully evaluated elimination diet, the patient was able to develop a diet on which he passed gas only 13 times per day.[21] Ingestion of milk markedly increased his productivity demonstrating that his colonic flora still had the potential to produce appreciable gas. Thus, it appears the patient developed a diet that contained carbohydrates that were nearly completely absorbed in the small bowel, thus depriving the colonic bacteria of substrates that

could be used for gas-producing reactions. This diet does not benefit all flatulent patients, suggesting that there are individual differences in the ability to absorb carbohydrates.

My approach to the patient complaining of excessive flatus is to first have the subject count how many times gas is passed each day. Not infrequently, the patient passes gas less than 13 times per day (the normal mean). Thus, there is a good deal of misunderstanding with regard to how frequently normal individuals pass gas. Obviously, such patients only require assurance.

Patients who are passing excessive gas almost always have high concentrations of the bacterial gases H_2 and CO_2 in flatus. Such excessive production reflects some combination of malabsorption of abnormal quantities of carbohydrates and/or a bacterial flora that is particularly adept at producing gas during fermentation reactions. Flatulent patients who are lactase-deficient invariably benefit from lactose restriction. Many are not cured, however. On occasion, non-lactase-deficient patients also seemingly benefit from lactose restriction.

Although not intensively studied, our preliminary investigations suggested that the fecal flora of flatulent subjects produces more gas from a given quantity of carbohydrate than do feces of control subjects. A carbohydrate-free diet usually results in a marked reduction in flatus volume in such patients, but it is usually impossible to pinpoint the carbohydrates that are causing the problem. A low-carbohydrate diet is not very appealing and most subjects prefer their flatulence. Antibiotics usually are of little benefit in flatulent subjects and on some occasions aggravate the problem. This result apparently reflects the ability of antibiotics to reduce the gas-consuming flora thereby increasing the volume or malodorous quality of flatus. An extensive gastrointestinal evaluation seldom, if ever, turns up a disease that could account for flatulence.

REFERENCES

1. Maddock WG, Bell JL, Tremain MJ. Gastro-intestinal gas. Observation on belching during anaesthesia, operations and pyelography; and rapid passage of gas. *Ann Surg* 1949;130:512–535.
2. Kirk E. The quantity and composition of human colonic flatus. *Gastroenterology* 1949;12:782–794.
3. Hood JH. Effect of posture on the amount and distribution of gas in the intestinal tract of infants. *Lancet* 1964;2:107–110.
4. Fordtran JS, Walsh JH. Gastric acid secretion rate and buffer content of the stomach after eating. Results in normal subjects and in patients with duodenal ulcer. *J Clin Invest* 1973;52:645.
5. Rune SJ. Acid-base parameters of duodenal contents in man. *Gastroenterology* 1972;62:533.
6. Turnberg LA, Fordtran JS, Carter NW, Rector FC. Mechanism of bicarbonate absorption and its relationship to sodium transport in the human jejunum. *J Clin Invest* 1970;49:548.
7. Levitt MD. Production and excretion of hydrogen gas in man. *N Engl J Med* 1969;281:122–127.

8. Levitt MD, Engel RR. Intestinal gas. In: *Advances in Internal Medicine,* Vol 20 (Strollerman E, ed). Chicago: Year Book Medical Publishers, Inc., 1975, p. 151.
9. Bond JH, Levy M, Levitt MD. Explosion of hydrogen gas in the colon during proctosigmoidscopy. *Gastrointe Endosc* 1976;23:41–42.
10. Bond JH, Engel RR, Levitt MD. Factors influencing pulmonary methane excretion in man. *J Exp Med* 1971;133:572.
11. Moore JG, Jessop LD, Osborne DN. A gas-chromatographic and mass spectrometric analysis of the odor of human feces. *Gastroenterology* (in press).
12. Levitt MD, Berggren T, Hasting J, Bond JH. Hydrogen (H_2) catabolism in the colon of the rat. *J Lab Clin Med* 1974;84:163–167.
13. Levine AS, Bond JH, Prentiss R, Levitt MD. Metabolism of carbon monoxide by the colonic flora of man. *Gastroenterology* 1982;83:633–637.
14. Andersen K, Ringsted A. Clinical and experimental investigations; ileus with particular reference to the genesis of intestinal gas. *Acta Chir Scand* 1943;88:475–502.
15. Levitt MD, Lasser R, Schwartz J, Bond JH. Studies of a flatulent patient. *N Engl J Med* 1976;295:260–262.
16. Levitt MD, Ingelfinger FJ. The volume, composition, and rate of accumulation of human intestinal gas. *Gastroenterology* 1968;54:1296.
17. Levitt MD. Volume and composition of human intestinal gas. *N Engl J Med* 1971;284:1394.
18. Wittneberg J, Levitt MD. Correlation of radiologic appearance and measured volume of intestinal gas. *Invest Radiol* 1970;5:244.
19. Lasser RH, Bond JH, Levitt MD. The role of intestinal gas in functional abdominal pain. *N Engl J Med* 1975;293:524–526.
20. Johnson AG. Controlled trial of metoclopramide in the treatment of flatulent dyspepsia. *Br Med J* 1971;2:25–26.
21. Levitt MD, Sutalf LO. Follow-up of a flatulent patient. *Dig Dis Sci* 1979;24:652–654.

17

Immunology of Inflammatory Bowel Disease

Fergus Shanahan and Stephan Targan

1. INTRODUCTION

Crohn's disease and ulcerative colitis are chronic inflammatory diseases of the intestine, the causes of which are not known. Although genetic factors, biochemical abnormalities, and infectious or other exogenous agents might contribute to the etiopathogenesis or trigger the onset of disease, the immune system probably mediates the tissue damage.[1–3] The role of the immune system in the pathogenesis of inflammatory bowel disease was first suspected on clinical grounds. These included an association with other putative immunologically mediated disorders, a responsiveness to immunosuppressive agents and the failure to consistently detect an infectious culprit. Subsequent investigations revealed multiple and varied immunological alterations that have been difficult to interpret or to fit to a coherent concept of the pathogenetic sequence. A major problem has been the difficulty in distinguishing primary immunological abnormalities from those that are secondary to the disease process itself. In addition, research progress has been impeded by the absence of a suitable animal model of inflammatory bowel disease; most of the available models exhibit an acute mucosal inflammation and do not mimic the chronic relapsing condition in hu-

Fergus Shanahan and Stephan Targan • Division of Gastroenterology, University of California–Los Angeles, School of Medicine, Center for the Health Sciences, Los Angeles, California 90024. The authors are supported by the UCLA/Harbor Inflammatory Bowel Disease Center (USPHS grant AM 36200) and by the National Foundation for Ileitis and Colitis, and by NIH USPHS grant DK 40057.

mans.[4] However, several advances have been made, and the purpose of this chapter is to review current theories and experimental approaches to the immunopathology of these disorders. Recent developments will be emphasized; other sources provide comprehensive accounts of earlier work.[5–11]

2. POSSIBLE TRIGGERS AND POTENTIAL IMMUNOPATHOGENETIC SEQUENCES

A working model of possible sequences of events in the pathogenesis of inflammatory bowel disease may be formulated based on available information and on studies of other putative immune-mediated disorders such as chronic active hepatitis or insulin-dependent diabetes. A schematic overview of the elements that might predispose to and trigger immune-mediated intestinal injury in inflammatory bowel disease is shown in Fig. 1. It is implicit in this overview that Crohn's disease and ulcerative colitis, in addition to being individually distinct, might each be heterogeneous disorders in which there are different underlying immune defects and possibly distinct trigger factors. The clinical expression of diseases may be similar because of common final pathways of tissue injury and because any tissue has a limited repertoire of responses to a given insult.

Two main pathways of immune-mediated mucosal injury have been

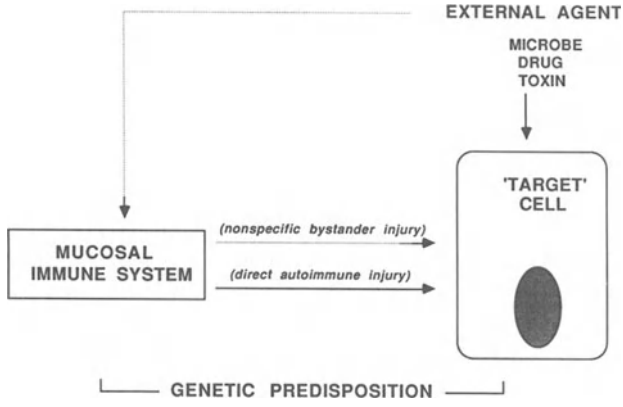


Figure 1. Summary of elements that might predispose to and trigger the onset of inflammatory bowel disease. Tissue damage might be due to a direct attack by the mucosal immune system on a specific target such as the surface or glandular epithelial cell, or might be a nonspecific outcome of disordered mucosal immune regulation with uncontrolled overreactivity to environmental antigens. Genetic predisposing factors and exogenous triggers might operate at the level of the "target" cell or at the level of the mucosal immune system.

hypothesized[1,3] (Figs. 2 and 3). First, there might be an immune assault directed against a specific intestinal target cell such as the epithelial cell. This could develop because the target cell is altered in some way, perhaps by an environmental trigger agent (drug, toxin, virus, or other microbial factor), and this stimulates an immune response that mediates the tissue damage. Chronicity occurs because the response of the immune system is inappropriate or fails to clear the inciting agent. Alternatively, in another subset of patients, the defect may lie at the level of the immune system in which a breakdown of immune regulation occurs and there is an autoimmune attack on otherwise normal target cells. Second, immune-mediated tissue injury may not be directed against a specific target cell; rather, the tissue damage might be a nonspecific bystander effect of uncontrolled or inappropriate reactivity to multiple dietary or intestinal bacterial antigens due to a primary disorder of mucosal immunoregulation.

3. AUTOIMMUNITY IN INFLAMMATORY BOWEL DISEASE

Unlike organ-specific autoimmune diseases such as pernicious anemia, autoimmune chronic active hepatitis, or thyroiditis, a specific cellular target for the immune reaction has not been identified conclusively in either ulcerative colitis or Crohn's disease. Histological studies of "early lesions" have not been conclusive and are difficult to interpret.[12–14] Despite this, several investi-

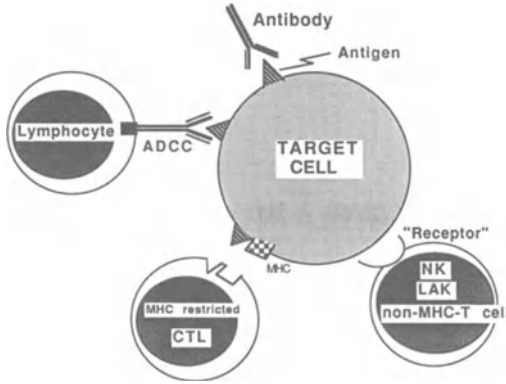


Figure 2. Mucosal immune effector mechanisms that have been proposed as possible mediators of a direct specific attack against the putative target cell in inflammatory bowel disease. These might be generated either in response to an alteration in the target cell, perhaps triggered by an external agent, or might arise from a breakdown in mucosal immune tolerance to normal host epithelial cells.

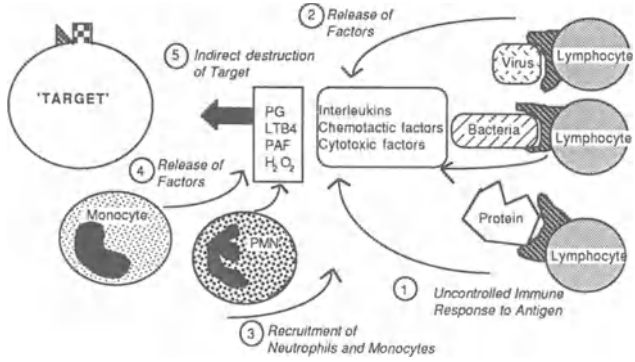


Figure 3. Tissue damage as a nonspecific indirect or "innocent bystander" attack. A disorder of mucosal immune regulation leads to uncontrolled overreactivity to a variety of exogenous antigens. The increased proliferation of lymphocytes with release of cytokines results in recruitment of additional inflammatory cells and auxiliary effector cells such as neutrophils. The release of inflammatory mediators such as prostaglandins (PG) and leukotrienes (LTB4) into the mucosa amplifies the inflammatory response and leads to local tissue damage.

gators have considered the surface or glandular epithelial cells to be the most likely targets of immune-mediated injury in these diseases. This is based to a large extent on evidence for antibody and lymphocyte reactivity to epithelial cells.[15–17]

Circulating antibodies to colonic epithelial cells were reported in patients with ulcerative colitis and Crohn's disease over 25 years ago.[18,19] These antibodies also reacted with extracts of certain strains of *E. coli;* this led to the speculation that inflammatory bowel disease might result from an autoimmune cross-reaction with shared antigens on intestinal epithelium and intestinal bacteria. Later, the significance of such antibodies was questioned because they could not be correlated with disease activity, extent, or duration, and were found in other diseases and in low titer in normal subjects.[1,15–17] In addition, they were not pathogenic for colon epithelial cells *in vitro*[20] or in an animal model of colitis.[21]

However, autoantibodies are heterogeneous and the possibility that certain antiepithelial antibodies might mediate antibody-dependent cellular cytotoxicity has been raised by several studies. Serum antibodies participating in antibody-dependent cellular cytotoxicity against human colon carcinoma cells[22] and rat colonic epithelial cells[23] have been reported in ulcerative colitis. In one study, serum antibody-dependent cellular cytotoxicity against a human colon carcinoma cell line was correlated with disease activity.[24] In addition to circulating antibodies, a colonic tissue-bound IgG antibody has been identified in patients

with ulcerative colitis but not Crohn's disease.[25,26] The antigen recognized by this antibody has been isolated and partially characterized, although its function is not known.[27] It occurs on normal colonic epithelium but is apparently absent from the small bowel. Evidence using monoclonal antibodies has confirmed the organ specificity and suggests that the same or a cross-reactive antigen may also be present in the skin and biliary tract.[28,29] Whether this distribution might explain some of the extraintestinal disorders associated with ulcerative colitis is an intriguing possibility.

Lymphocyte-mediated cytotoxicity against epithelial cells has been reported in both ulcerative colitis and Crohn's disease by several investigators.[30–40] This phenomenon occurs with peripheral blood lymphocytes and intestinal mononuclear cells. The mechanism is thought to be a form of antibody-dependent cellular cytotoxicity in which lymphocytes are armed with an antiepithelial antibody.[32–37,40] However, the cytotoxic reaction occurs only with colonic and not with small-bowel epithelial cells—even in patients with Crohn's disease limited to the small intestine. Therefore, it cannot explain all of the lesions of Crohn's disease. In addition, some potential technical limitations of these experiments may complicate their interpretation.[3] Chief among these is the lability and unreliability of freshly isolated colonic epithelial cells as targets in cytotoxicity assays. This is likely to be a particular problem when epithelial cells are isolated from inflamed tissue such as in inflammatory bowel disease, and therefore comparisons with noninflamed tissues may be inherently biased.

Another approach to the study of epithelial autosensitization has been the use of erythrocytes coated with epithelial antigens as targets instead of intact epithelial cells.[41,42] Soluble epithelial cell antigen complexes isolated from rat and murine colon and intestine have been shown to induce an organ-specific inflammation when injected with adjuvant.[43] When such antigens were bound to chicken erythrocytes as indicator cells, peripheral blood[41] and colonic lamina propria lymphocytes[42] from patients with both Crohn's disease and ulcerative colitis, but not from controls, were found to be cytotoxic to the indicator cells. The cytotoxic reaction was specific for gut-associated epithelial antigens because erythrocytes labeled with kidney epithelial antigens were resistant to lysis. Although the mechanism of lysis was antibody-dependent cellular cytotoxicity when peripheral blood lymphocytes were the effector cells,[41] direct T-cell cytotoxicity appeared to be involved when intestinal lamina propria lymphocytes were tested.[42] The significance of these findings is uncertain because of the low levels of lysis observed and because classical cytotoxic T-cell killing requires that the target antigen be presented to the effector cell in association with self-antigens of the major histocompatibility complex[44,45]; reactions against epithelial antigens on the surface of chicken erythrocytes cannot be major histocompatibility complex restricted.

In summary, several different studies have provided evidence for autoimmune reactivity against intestinal epithelial cells in inflammatory bowel disease. However, the autoreactivity may not be a primary pathogenic event and might simply be an intriguing epiphenomenon secondary to tissue damage. This coupled with the lack of any histological evidence for a specific target of immune-mediated tissue injury led some investigators to examine the possibility that tissue damage occurs nonspecifically as a result of an abnormality of mucosal immune regulation.

4. IMMUNOREGULATION IN INFLAMMATORY BOWEL DISEASE

Because an inciting antigen or target for the immune response in inflammatory bowel disease has not been determined, evaluations of the immune system have necessarily been limited to nonspecific variables of immune competence. No primary deficiency of systemic or mucosal immunity has been identified in either Crohn's disease or ulcerative colitis. However, numerous immunological alterations in these patients have been reported and are detailed elsewhere.[1,5–10] Some of the diverse immunological disturbances which have been observed might reflect an underlying defect in immunoregulation; others are probably secondary to the debilitating effects of the chronic inflammatory process or reflect an appropriate response to the antigenic load associated with increased gut permeability in these disorders. Additional variables that complicate the assessment of immune regulation in patients with inflammatory bowel diseases include nutritional status[46] and the influence of treatments such as antibiotics[47] and sulfasalazine or steroids (Section 6) on the immune system. The possible heterogeneity of both Crohn's disease and ulcerative colitis might also account for inconsistencies between different patients and different studies.

Regulation of the immune system occurs on several levels. One that is amenable for study and has received much attention in inflammatory bowel disease is the influence of suppressor T cells. Unfortunately, reports on suppressor cell activity both in the peripheral blood[48–52] and within the intestinal mucosa[53,54] have been contradictory. Some of the discrepancy is probably due to the heterogeneity of suppressor cells[55,56]; the various assays used have examined different types of suppressor cells. Non-antigen-specific suppressor cell function appears to be particularly sensitive to the effects of disease activity. In an important study of patients with Crohn's disease, with several assay systems for peripheral blood suppressor cells, defects were found only in patients with active disease.[56] When patients were in remission, no suppressor cell abnormality was found.

Using a different indicator system, one group of investigators has de-

scribed a "covert" suppressor cell disorder in Crohn's disease.[52] This suppressor activity was unmasked (covert) when peripheral blood lymphocytes were fractionated into T- and B-cell subsets *in vitro,* and was associated with a numerically small subset of cells having the CD8+, Leu 7+ phenotype.[57] In some patients the suppressor activity appeared to become clinically overt leading to hypogammaglobulinemia.[58] Because patients with ulcerative colitis and other disease-control subjects were not included in these studies, the specificity of the finding is not known. The suppressor activity does not occur in all patients with Crohn's disease and is not present at the mucosal level[59,60]; it probably represents an expansion of a normal regulatory mechanism in response to mucosal inflammation.

Immune regulation also operates on the afferent limb of the immune response, even at the level of antigen presentation. There is now clear evidence that the columnar epithelial cells in the normal intestine are capable of presenting antigens to the cells of the mucosal immune system.[61,62] It has been known for several years that epithelial cells express Ia or class II major histocompatibility complex antigens[63]; this expression is increased in the presence of inflammation[64,65] and may even be modulated by intraepithelial lymphocytes.[66] The significance of this is not yet clear but does suggest the possibility that abnormal antigen processing or enhanced antigen presentation by the epithelium in inflammatory bowel disease might lead to exaggerated tissue-damaging immune responses to a variety of antigens that would normally be excluded or ignored. Whether gut epithelial cells provide other signals such as interleukins, which act on the mucosal immune system, is not clear but seems likely. Recently, allogeneic mixed lymphocyte cultures with Ia^+ gut epithelial cells from control subjects but not from patients with Crohn's disease or ulcerative colitis have been shown to lead to the selective induction of non-antigen-specific suppressor T cells.[67] This defect in induction of suppressor T cells by epithelial cells in inflammatory bowel disease, if confirmed, might be responsible for the initiation or perpetuation of an inflammatory response to perhaps a myriad of luminal antigens.

5. MUCOSAL EFFECTOR CELL FUNCTION IN INFLAMMATORY BOWEL DISEASE

Our understanding of the mucosal immune system has advanced in recent years. The cellular and humoral components of the local mucosal immune system are uniquely adapted for host defense at the interface with the environment.[68,69] They are largely independent of the systemic immune response and are functionally and phenotypically distinct from their counterparts within the

systemic immune system. This implies that studies of peripheral blood do not necessarily reflect events at the intestinal mucosa. In addition, animal studies of mucosal immunity cannot reliably be extrapolated to the human.[69] Therefore, a precise delineation of the effector cell populations within the human intestinal mucosa in control subjects and patients with inflammatory bowel disease is important.

The lesions of Crohn's disease and ulcerative colitis contain a two- to fourfold increase in inflammatory cells. The infiltrate consists of an expanded T-cell population and, to a slightly greater extent, the B-cell population, with a variable increase in auxiliary effector cells such as neutrophils, mast cells, and macrophages. Subclassification of the T-cell population into distinct regulatory subsets such as helpers and suppressors has not shown any deviation from the normal proportions.[70,71] Studies of mucosal cytotoxic effector cells in inflammatory bowel disease have been limited by the lack of an appropriate target cell such as viable autologous epithelial cells.[69,72] In contrast, examination of the B-cell population shows marked differences from the normal mucosa and differences between Crohn's disease and ulcerative colitis.[73,74]

5.1. B-Cell Function

Immunohistological studies of the intestinal lesions in both Crohn's disease and ulcerative colitis show that the increased population of mucosal B cells, including plasma cells, is associated with a proportionately greater increase in the numbers of IgG- and, to a lesser extent, IgM-bearing and secreting cells. Up to a 30-fold increase in IgG-bearing cells, in addition to increased IgA (twofold) and IgM (five- to sixfold) cells, has been observed in both diseases.[75–78] Further immunohistological analysis of B-cell subsets, based on the subclass of surface IgG, indicates differences between ulcerative colitis and Crohn's disease.[73] The proportion of IgG_1-producing cells is significantly higher in ulcerative colitis than in Crohn's colitis, whereas the proportion of IgG_2-producing cells is significantly higher in Crohn's colitis than in ulcerative colitis.

The alteration in the profile of IgG subclass-bearing cells has been confirmed by studies of spontaneous immunoglobulin secretion by intestinal cells isolated from patients with inflammatory bowel disease.[74,79,80] Whether these changes reflect an underlying disorder of immune regulation or are an appropriate response to the tissue injury and breakdown in mucosal integrity is not known. Whatever the mechanism, the relative increase in IgG secretion might contribute to the perpetuation of tissue injury because of its potential for complement activation, opsonization, and antibody-dependent cellular cytotoxicity.[74,81] The different patterns of IgG subclass secretion in Crohn's disease and

ulcerative colitis might reflect immune response to different classes of antigens. For example, IgG_1 and IgG_3 are the predominant subclasses produced in response to protein and T-cell-dependent antigens, whereas IgG_2 is the predominant subclass response to carbohydrate antigens.[74,81] Studies of IgA production and secretion in inflammatory bowel disease have shown that the transepithelial transport process for IgA is normal.[82,83] Total IgA secretion by isolated intestinal cells from patients with both diseases is slightly reduced compared to control intestinal cells, and there is an increase in the proportion of monomeric IgA and IgA_1 subclass secretion.[84] This shift in IgA subclass is probably due to the influx of peripheral blood B cell into the inflamed mucosa.

5.2. Cytotoxic Lymphocytes

As discussed earlier, lymphocyte-mediated cytotoxicity against epithelial cells has been proposed as a possible mechanism of tissue injury in inflammatory bowel disease. Antibody-dependent cellular cytotoxicity (or K-cell killing) has been highlighted by several studies.[37,40,41] However, whether isolated intestinal mononuclear cells can mediate antibody-dependent cellular cytotoxicity has been controversial.[85–89] Differences in results might be technical and related to the influence of the enzymatic isolation procedure on lymphocyte[90–92] and Fc receptor function.[93] We have been able to consistently find significant antibody-dependent cellular cytotoxicity activity against tumor cell targets using freshly isolated mucosal cell preparations, but this appears to be macrophage- and not lymphocyte-mediated (F. Shanahan and S. Targan, unpublished observations). This is consistent with the finding that lymphocytes bearing the CD16 antigen (a marker of Fc receptors for IgG) are absent from the human colon.[94,95]

Natural killer (NK) cells are thought to have a role in surveillance against certain tumor and virally infected cells and also appear to have immunoregulatory effects. Natural killer cell function has been reported to be decreased in the peripheral blood of patients with Crohn's disease[96–98] and ulcerative colitis.[98] This is probably secondary to the inflammatory process; it is known that plasma from the mesenteric vein of patients with intestinal inflammation is inhibitory to NK cells.[99] Studies of mucosal NK cells in patients with inflammatory bowel and other intestinal diseases are difficult to interpret because of their low numbers and low levels of activity in unfractionated cell preparations from normal or inflamed intestines.[88,89,91,94,95,99–102] Direct NK cytolytic effects are therefore not thought to be important in intestinal disease.[72] However, dismissal of NK cells on a numerical basis may be inappropriate because the number of cells required for effective tumor and viral infection surveillance and immunoregulation *in vivo* is not known.[104]

Lymphokine-activated killer cells are induced when human mucosal lymphocytes and peripheral blood lymphocytes are cultured in the presence of interleukin-2 (IL-2).[94,95,105] These cells have lytic activity against both NK-sensitive and NK-resistant target cells, and are currently being intensively investigated as a form of treatment for metastatic cancer, including colorectal cancer.[106] Unlike the peripheral blood, where the majority of lymphokine-activated killer cells are generated from NK cells bearing the CD16 surface marker,[107,108] the precursors of the mucosal lymphokine-activated killer phenomenon are $CD16^-$.[94,95] In inflammatory bowel disease, the generation of mucosal lymphokine-activated killer cell activity *in vitro* in response to exogenous IL-2 is normal.[94] However, the production of IL-2 by mucosal mononuclear cells from these patients is significantly diminished under certain conditions *in vitro*.[109] Although the role of lymphokine-activated killer as a defense mechanism *in vivo* is not known, it is possible that defective generation of mucosal lymphokine-activated killer cells due to reduced production of IL-2 locally might contribute to the increased frequency of intestinal cancer in patients with longstanding inflammatory bowel disease.

In contrast to antibody-dependent cellular cytotoxicity, NK and lymphokine-activated killer, little is known of cytotoxic T-lymphocyte function in either the normal human mucosa or in inflammatory bowel disease. Mitogen-induced cellular cytotoxicity in culture has been used as a measure of T-cell cytotoxicity[88,100,110] and has reported to be decreased in the mucosa and blood of patients with inflammatory bowel disease.[110] However, the nonspecific cytotoxicity induced by mitogens is probably distinct from that of specifically *in vivo*-activated cytotoxic T lymphocytes. In addition, this technique does not permit a precise distinction between killing due to T cells and that due to culture-activated or lymphokine-activated killer cells.[95,111] In inflammatory bowel disease, it is possible that a subset of *in vivo*-activated cytotoxic T lymphocytes (possibly reactive to self antigens) might coexist with depressed (probably secondary to the disease) nonspecific T-cytolytic function.

Even though the antigens responsible for the immune reactivity in inflammatory bowel disease are not known, it may still be possible to indirectly examine cytotoxic T cells that are reactive to them. The human T-cell antigen receptor consists of two functional components: one responsible for antigen recognition and the other (CD3) responsible for signal transduction.[45,112] If a cytotoxic T cell has already been primed by antigen *in vivo*, subsequent stimulation *in vitro* with monoclonal antibodies to the CD3 component of its receptor will trigger the cytolytic activity of the cell. This technique might therefore be a useful indirect method for measuring and identifying *in vivo*-primed cytotoxic T lymphocytes when the antigen to which they are reacting is not known.[113] The lytic function of freshly isolated cytotoxic T lymphocytes from noninflamed intestinal mucosa has recently been demonstrated using this method.[114]

In inflammatory bowel disease, anti-CD3-triggered peripheral blood cytotoxic T lymphocytes have been shown to have significantly greater activity than those of normal subjects.[115] Whether the function or numbers of such cells is altered in the inflamed mucosa is not yet clear.

5.3. Macrophages, Mast Cells, and Polymorphonuclear Leukocytes

An increase in intestinal macrophages is seen in inflammatory bowel disease, particularly in Crohn's disease.[116–118] This is not surprising because the pathological hallmark of Crohn's disease is granuloma formation where macrophage/epithelioid cells are the most prominent cell type. Indeed, it has even been postulated that a basic defect in macrophage function might explain the clinicopathological differences between Crohn's disease and ulcerative colitis.[120] However, no major monocyte-macrophage defect has been found in either condition. Although there is evidence from several studies for macrophage activation in inflammatory bowel disease,[117] this is probably a nonspecific response to the intestinal inflammatory process.

Increased intestinal mast cell numbers in both Crohn's disease and ulcerative colitis have been reported by several investigators.[121–124] This led to the notion that the antiallergic agent cromolyn sodium might have a therapeutic benefit.[125] Subsequent clinical trials, however, failed to confirm any therapeutic benefit. Experimental studies with isolated cells have indicated that intestinal mucosal mast cells differ from mast cells at other sites in being unresponsive to this agent.[69,126,127] There is additional extensive evidence for the heterogeneity of mast cells. It is now clear that mast cells within the mucosa are morphologically, biochemically, and functionally distinct from connective tissue mast cells at nonmucosal sites and in the deeper layers of the intestinal wall.[69,127] Whether one or both types of mast cell subpopulation undergoes hyperplasia in inflammatory bowel disease is not known.

Neutrophils are prominent components of active lesions in both ulcerative colitis and Crohn's disease. They provide a major contribution to the tissue damage and amplify the inflammatory response once it has been initiated. Tracer studies using radiolabeled white cells have demonstrated the passage of neutrophils and monocytes out of the circulation into active lesions where some of them are destroyed in the inflammatory process and others continue to migrate into the crypt and then the intestinal lumen.[128,129] The chemotactic stimuli directing the neutrophil to the active lesions are probably multiple, including lymphocyte-derived factors. Recent studies have indicated the importance of arachidonate metabolites, particularly the leukotrienes, as mediators of the inflammatory response in inflammatory bowel disease.[130–132] Leukotriene B_4 (LTB_4) is responsible for most of the chemotactic activity in the inflamed gut mucosa[133] and neutrophils themselves appear to be the major source of the LTB_4. This is

therefore an important mechanism for the amplification of inflammation and is a primary target for the action of many of the drugs used today in the treatment of inflammatory bowel disease.

6. IMMUNOPHARMACOLOGY OF INFLAMMATORY BOWEL DISEASE

Theories of the immunopathogenesis of inflammatory bowel disease have important implications for the design of new approaches and treatments. Differences between the mucosal and systemic immune systems suggest that it might be possible to design drugs for selective modulation of mucosal immunity. The notion of immune-mediated tissue injury in inflammatory bowel disease also suggests that the mechanism of action of drugs that are currently used and of proven efficacy in these diseases might exert their mechanisms of action on the cellular components or secretory products of the immune system.

The drugs used to inhibit the inflammatory response in ulcerative colitis and Crohn's disease include sulfasalazine, corticosteroids, and 6-mercaptopurine. These agents appear to have multiple mechanisms of action on different components of the inflammatory response. Sulfasalazine inhibits both the cyclooxygenase and the lipoxygenase pathways of arachidonate metabolism and therefore reduces levels of prostaglandins and leukotrienes in the inflamed tissue.[2,134] Additional effects of sulfasalazine that are independent of arachidonate metabolism include the inhibition of binding of the bacterial chemotactic peptide f-Met-Leu-Phe to its receptor on neutrophils,[134] inhibition of neutrophil chemotaxis,[135] inhibition of lymphocyte-mediated cytotoxicity,[136] and a blocking effect on the binding of killer lymphocyte factor(s) to their target cell.[137] A possible effect on humoral immunity is suggested by reports of sulfasalazine-induced deficiencies of IgA and IgG_2 subclasses.[138]

Corticosteroids are nonspecific inhibitors of several components of the immune response. Of probable significance in inflammatory bowel disease is their inhibitory effect on leukotriene and prostaglandin production within the intestinal mucosa; corticosteroids promote the synthesis of lipomodulin, which in turn blocks the release of arachidonic acid from phospholipids by inhibiting phospholipase A_2.[2]

The use of the purine analog azathioprine and its active metabolite 6-mercaptopurine in inflammatory bowel disease has been controversial.[139] However, there is now a reasonable body of evidence to support its use in selected patients.[139–141] The primary action of these drugs is probably the inhibition of nucleic acid metabolism, and this alters the proliferation and clonal expansion of immunocytes that follows antigen stimulation. Unlike sulfasalazine and corticosteroids, 6-mercaptopurine does not appear to have a major direct action on

acute inflammatory cells such as neutrophils or their soluble mediators. Therefore 6-mercaptopurine is not useful for acute situations but rather its therapeutic benefit only becomes evident after an average period of 3 months of therapy.[141] A number of studies have emphasized the inhibitory effect of these agents on the cytolytic and immunoregulatory activities of NK cells.[142–144] Inhibition of T-cell cytotoxicity has also been reported.[145] Some of these drug-induced alterations in immune function might provide clues to the immunopathogenesis or can be used to monitor drug therapy and predict those patients who are likely to have a therapeutic response.

7. FUTURE DIRECTIONS

Future advances in the immunopathogenesis of inflammatory bowel disease will require a better understanding of the regulation of the normal human mucosal immune system. This should include investigations of interactions between the mucosal immune system and the local neuroendocrine system, and also the other components of the intestinal mucosa, including the epithelium.[146] The role of the epithelial cell both as a possible target of the immune response and, by its antigen-presenting and processing functions, as a participant in the immune response should be clarified with the development of better methods for the analysis of these cells in culture.

The possibility that Crohn's disease and ulcerative colitis might each be heterogeneous disorders should be considered in the design and monitoring of future immunomodulatory drugs. Finally, an animal model of both ulcerative colitis and Crohn's disease would greatly facilitate studies of the immunopathogenesis and immunotherapy of these diseases. The colitis that occurs in the cotton top tamarin *(Sanguinus oedipus)* might provide important clues with respect to ulcerative colitis[147] whereas chlamydial infections in primates have already been used by some investigators to mimic the granulomatous inflammation of Crohn's disease.[148]

REFERENCES

1. Strober W, James S. The immunologic basis of inflammatory bowel disease. *J Clin Immunol* 1986;6:415–432.
2. MacDermott RP, Stenson WF. The immunology of inflammatory bowel disease. *Hosp Practice* 1986;21:97–116.
3. Shanahan F. Inflammatory bowel disease, pp 862–866. In: Immunology of Intestinal Diseases (Targan SR, moderator). *Ann Intern Med* 1987;106:853–870.
4. Strober W. Animal models of inflammatory bowel disease: An overview. *Dig Dis Sci* 1985;30:3S–10S.

5. Strickland RG, Sacher DB. The immunology of inflammatory bowel disease. *Progr Gastroenterol* 1977;3:821–838.
6. Sacher DB, Auslander MO, Walfish JS. Aetiological theories of inflammatory bowel disease. *Clin Gastroenterol* 1980;9:231–247.
7. Strickland RG, Jewell DP. Immunoregulatory mechanisms in nonspecific inflammatory bowel disease. *Ann Rev Med* 1983;34:195–204.
8. Kirsner JB, Shorter RG. Recent developments in nonspecific inflammatory bowel disease. *N Engl J Med* 1982;306:837–848.
9. MacDermott RP. Cell mediated immunity in gastrointestinal disease. *Hum Pathol* 1986;17:219–233.
10. Fiocchi C. Etiology of mucosal ulcerative colitis. In: *Mucosal Ulcerative Colitis* (Jagleman DG, ed). Mount Kisco, NY: Futura, 1986, pp. 1–35.
11. Elson CO, Kagnoff MF, Fiocchi C, Befus AD, Targan S. Intestinal immunity and inflammation: recent progress. *Gastroenterology* 1986;91:746–768.
12. Rickbert RR, Carter HW. The "early" ulcerative lesion of Crohn's disease: Correlative light- and scanning electron-microscopy studies. *J Clin Gastroenterol* 1980;2:11–19.
13. Schmitz-Moormann P, Becker H. Histologic studies on the formal pathogenesis of the epithelial cell granuloma in Crohn's disease. In: *Recent Advances in Crohn's Disease* (Pena AS, Weterman IT, Booth CC, Strober W, eds). The Hague: Martinus Nijhoff, 1981, pp. 76–79.
14. Korelitz BI, Sommers SC. Rectal biopsy in patients with Crohn's disease: normal mucosa on sigmoidoscopic examination. *J Am Med Assoc* 1977;237:2742–2744.
15. Bartnik W, Shorter RG. Inflammatory bowel disease: Immunological developments. In: *Developments in Digestive Diseases* (Berk JE, ed). Philadelphia: Lea & Febiger, 1980, pp. 109–127.
16. Fiocchi C, Farmer RG. Autoimmunity in inflammatory bowel disease. *Clin Aspects of Autoimmun.* 1987;1:12–19.
17. Wright R, Truelove SC. Autoimmune reactions in ulcerative colitis. *Gut* 1966;7:32–40.
18. Broberger O, Perlmann P. Autoantibodies in human ulcerative colitis. *J Exp Med* 1959;110:657–674.
19. Lagercrantz R, Hammarstrom S, Perlmann P, Gutaffson BE. Immunological studies in ulcerative colitis. III. Incidence of antibodies to colon-antigen in ulcerative colitis and other gastrointestinal diseases. *Clin Exp Immunol* 1966;1:263–276.
20. Broberger O, Perlmann P. In vitro studies of ulcerative colitis. I. Reaction of patients' serum with human fetal colon cells in tissue culture. *J Exp Med* 1963;117:705–715.
21. Rabin BS, Rogers SJ. Nonpathogenicity of anti-intestinal antibody in the rabbit. *Am J Pathol* 1976;83:269–278.
22. Nagai T, Das K. Demonstration of an assay for specific cytolytic antibody in sera from patients with ulcerative colitis. *Gastroenterology* 1981;80:1507–1512.
23. Hibi T, Aiso S, Yoshida T, Watanabe M, Asakura H, Tsuru S, Tsuchiya M. Anti-colon antibody and lymphocytophilic antibody in ulcerative colitis. *Clin Exp Immunol* 1982;49:75–80.
24. Das KM, Kadona Y, Fleischner GM. Antibody-dependent cell-mediated cytotoxicity in serum samples from patients with ulcerative colitis. *Am J Med* 1984;77:791–796.
25. Das KM, Dubin R, Nagai T. Isolation and characterization of colon tissue-bound antibodies from patients with idiopathic ulcerative colitis. *Proc Natl Acad Sci USA* 1978;75:4528–4532.
26. Nagai T, Das KM. Detection of colonic antigen(s) in tissues from ulcerative colitis using purified colitis colon tissue-bound IgG (CCA-IgG). *Gastroenterology* 1981;81:463–470.
27. Takahashi F, Das KM. Isolation and characterization of a colonic autoantigen specifically recognized by colon tissue-bound immunoglobulin G from idiopathic ulcerative colitis. *J Clin Invest* 1985;76:311–318.

28. Vecchi M, Sakamaki S, Diamond B, Novikoff AB, Novikoff PM, Das KM. Development of a monoclonal antibody specifically reactive to gastrointestinal goblet cells. *Proc Natl Acad Sci USA* 1987;84:3425–3429.
29. Das KM, Sakamaki S, Vecchi M, Diamond B. The production and characterization of monoclonal antibodies to a human colonic antigen associated with ulcerative colitis: Cellular localization of the antigen by using the monoclonal antibody. *J Immunol* 1987;139:77–84.
30. Perlmann P, Broberger O. In vitro studies of ulcerative colitis. II. Cytotoxic action of white blood cells from patients on human fetal colon. *J Exp Med* 1963;117:717–733.
31. Watson DW, Quigley JA, Bolt RJ. Effects of lymphocytes from patients with ulcerative colitis on human adult colon epithelial cells. *Gastroenterology* 1966;51:985–993.
32. Shorter RG, Cardoza M, Huizinga KA, ReMine G, Spencer RJ. Further studies of in vitro cytotoxicity of lymphocytes for colonic epithelial cells. *Gastroenterology* 1969;57:30–35.
33. Shorter RG, Cardoza M, Spencer RJ, Huizinga KA. Further studies of in vitro cytotoxicity of lymphocytes from patients with ulcerative and granulomatous colitis for allogeneic colonic epithelial cells, including the effect of colectomy. *Gastroenterology* 1969;56:304–309.
34. Shorter RG, Cardoza MR, ReMine SG, Spencer RJ, Huizinga KA. Modifications of in vitro cytotoxicity of lymphocytes from patients with chronic ulcerative colitis or chronic granulomatous colitis for allogenic colonic epithelial cells. *Gastroenterology* 1970;58:692–698.
35. Shorter RG, Huizinga KA, Spencer RJ, Aas J, Guy SK. Inflammatory bowel disease. Cytophilic antibody and the cytotoxicity of lymphocytes for colonic cells in vitro. *Dig Dis Sci* 1971;16:673–680.
36. Shorter RG, Huizinga KA, Spencer RJ. A working hypothesis for the etiology and pathogenesis of nonspecific inflammatory bowel disease. *Dig Dis Sci* 1972;17:1024–1032.
37. Stobo JD, Tomasi TB, Huizinga KA, Spencer RJ, Shorter RG. In vitro studies of inflammatory bowel disease: surface receptors of the mononuclear cells required to lyse allogeneic colonic epithelial cells. *Gastroenterology* 1976;70:171–176.
38. Watson DW, Bolt RJ. Immune mechanisms in ulcerative colitis. In: *Progress in Gastroenterology,* Vol 1 (Glass GBJ, ed). New York: Grune & Stratton, 1968, pp. 391–411.
39. Kemler BJ, Alpert JE. Inflammatory bowel disease: a study of cell-mediated cytotoxicity for isolated human colonic epithelial cells. *Gut* 1980;21:353–359.
40. Shorter RG, McGill DB, Bahn RC. Cytotoxicity of mononuclear cells for autologous colonic epithelial cells in colonic diseases. *Gastroenterology* 1984;86:13–22.
41. Aronson RA, Cook SL, Roche JK. Sensitization to epithelial antigens in chronic mucosal inflammatory disease. I. Purification characterization and immune reactivity of murine epithelial cell-associated components (ECAC). *J Immunol* 1983;131:2796–2804.
42. Roche JK, Fiocchi C, Youngman K. Sensitization to epithelial antigens in chronic mucosal inflammatory disease. Characterization of human intestinal mucosa-derived mononuclear cells reactive with purified epithelial cell-associated components in vitro. *J Clin Invest* 1985;75:522–530.
43. Roche JK, Cook SR, Day ED. Cellular cytotoxicity and gastrointestinal inflammation in inbred rats: Induction with gut organ-specific antigens. *Immunology* 1981;44:489–497.
44. Zinkernagel RM, Doherty PC. H-2 compatibility requirement for T cell mediated lysis of target cells infected with lymphocytic choriomeningitis virus. Different cytotoxic T cell specificities are associated with structures coded in H-2K or H-2D. *J Exp Med* 1975;141:1427–1436.
45. Acuto O, Reinherz EL. The human T-cell receptor. Structure and function. *N Engl J Med* 1985;312:1100–1111.
46. Chandra RK. Nutrition and immune responses. *Can J Physiol Pharmacol* 1983;61:290–294.
47. Hauser WE, Remington JS. Effect of antibiotics on the immune response. *Am J Med* 1982;72:711–716.

48. Hodgson HJF, Wands JR, Isselbacher KJ. Decreased suppressor cell activity in inflammatory bowel disease. *Clin Exp Immunol* 1978;32:451–458.
49. Kemler B, Alpert E. Immune regulation in inflammatory bowel disease: Absence of a serum inhibitor of suppressor cell function. *Clin Exp Immunol* 1980;42:280–284.
50. Ginsburg CH, Falchuk ZM. Defective autologous mixed lymphocyte reaction and suppressor cell generation in patients with inflammatory bowel disease. *Gastroenterology* 1982;83:1–9.
51. Holdstock G, Chastnay BF, Krawitt EL. Functional suppressor T cell activity in Crohn's disease and the effects of sulfasalazine. *Clin Exp Immunol* 1982;48:619–624.
52. Elson CO, Graeff AS, James SP, Strober W. Covert suppressor T cells in Crohn's disease. *Gastroenterology* 1981;80:1513–1521.
53. Fiocchi C, Youngman KR, Farmer RG. Immunoregulatory function of human intestinal mucosa lymphoid cells: evidence for enhanced suppressor cell activity in inflammatory bowel disease. *Gut* 1983;24:692–701.
54. Goodacre RL, Bienenstock J. Reduced suppressor cell activity in intestinal lymphocytes from patients with Crohn's disease. *Gastroenterology* 1982;82:653–658.
55. Lobo PI, Spencer CE. Inhibition of humoral and cell-mediated immune responses in man by distinct suppressor cell systems. *J Clin Invest* 1979;63:1157–1163.
56. Auer IO, Roder A, Frohlich J. Immune status in Crohn's disease. IV. Immunoregulation evaluated by multiple, distinct T-suppressor cell assays of lymphocyte proliferation, and by enumeration of immunoregulatory T-lymphocyte subsets. *Gastroenterology* 1984;86:1531–1543.
57. James SP, Neckers LM, Graeff AS, Cossman J, Balch CM, Strober W. Suppression of immunoglobulin synthesis by lymphocyte subpopulations in patients with Crohn's disease. *Gastroenterology* 1984;86:1510–1518.
58. Elson CO, James SP, Graeff AS, Berensdon RA, Strober W. Hypogammaglobulinemia due to abnormal suppressor T cell activity in Crohn's disease. *Gastroenterology* 1984;86:569–576.
59. James SP, Fiocchi C, Graeff AS, Strober W. Immunoregulatory function of lamina propria T cells in Crohn's disease. *Gastroenterology* 1985;88:1143–1150.
60. Elson CO, Machelski E, Weiserbs DB. T cell-B cell regulation in the intestinal lamina propria in Crohn's disease. *Gastroenterology* 1985;89:321–327. 1985.
61. Bland PW, Warren LG. Antigen presentation by epithelial cells of the rat small intestine. I. Kinetics, antigen specificity and blocking by anti-Ia antisera. *Immunology* 1986;58:1–7.
62. Bland PW, Warren LG. Antigen presentation by epithelial cells of the rat small intestine. II. Selective induction of suppressor T cells. *Immunology* 1986;58:9–14.
63. Scott H, Solheim BG, Brandtzaeg P, Thorsby E. HLA-DR-like antigens in the epithelium of the human small intestine. *Scand J Immunol* 1980;12:77–82.
64. Selby WS, Janossy G, Mason DYU, Jewell DP. Expression of HLA-DR antigens by colonic epithelium in inflammatory bowel disease. *Clin Exp Immunol* 1983;53:614–618.
65. Hirata I, Austin LL, Blackwell WH, Weber JR, Dobbins WO. Immunoelectron microscopic localization of HLA-DR antigen in control small intestine and colon and in inflammatory bowel disease. *Dig Dis Sci* 1986;31:1317–1330.
66. Cerf-Bensussan N, Quarnoni A, Kurnick JT, Bhan AK. Intraepithelial lymphocytes modulate Ia expression by intestinal epithelial cells. *J Immunol* 1984;132:2244–2252.
67. Mayer L, Eisenhardt D. Lack of induction of suppressor T cells by gut epithelial cells from patients with inflammatory bowel disease. The primary defeat? *Gastroenterology* 1987 (abstract); 92(2):1524.
68. Targan SR. The intestine as an immunologic organ. In: Immunologic mechanisms in intestinal diseases (Targan R, moderator). *Ann Intern Med* 1987;106:853–870.
69. Shanahan F, Targan S. Role of natural effector cells in human gastrointestinal disease. In: *Functions of the Natural Immune System* (Reynolds CW, Wiltrout RH, eds). New York: Plenum (in press).
70. Selby WS, Janossy G, Bonfil M, Jewell DP. Intestinal lymphocyte subpopulations in inflam-

matory bowel disease: An analysis by immunohistological and cell isolation techniques. *Gut* 1984;25:32–40.
71. Hirata I, Berrebi G, Austin LL, Keren DR, Dobbins WO. Immunohistologic characterization of intraepithelial and lamina propria lymphocytes in control ileum and colon and in inflammatory bowel disease. *Dig Dis Sci* 1986;31:593–603.
72. James SP, Strober W. Cytotoxic lymphocytes and intestinal disease. *Gastroenterology* 1986;90:235–240.
73. Kett K, Rognum TO, Brandtzaeg P. Mucosal subclass distribution of immunoglobulin G-producing cells is different in ulcerative colitis and Crohn's disease of the colon. *Gastroenterology* 1987;93:919–924.
74. MacDermott RP, Nahm MH. Expression of human immunoglobulin G subclasses in inflammatory bowel disease. *Gastroenterology* 1987;93:1127–1134.
75. Baklien K, Brandtzaeg P. Comparative mapping of the local distribution of immunoglobulin-containing cells in ulcerative colitis and Crohn's disease of the colon. *Clin Exp Immunol* 1975;22:197–209.
76. Rosekrans PCM, Meijer CJLM, van der Wal AM, Cornelisse CJ, Lindeman J. Immunoglubulin containing cells in inflammatory bowel disease of the colon: a morphometric and immunohistochemical study. *Gut* 1980;21:941–947.
77. Scott BB, Goodall A, Stephenson P, Jenkins D. Rectal mucosal plasma cells in inflammatory bowel disease. *Gut* 1983;24:519–524.
78. Keren DF, Appelman HD, Dobbins WO, Wells JJ, Whienant B, Foley BS, Dieterle BS, Geisinger K. Correlation of histopathologic evidence of disease activity with the presence of immunoglobulin-containing cells in the colons of patients with inflammatory bowel disease. *Hum Pathol* 1984;15:757–763.
79. MacDermott RP, Nash GS, Bertovich MJ, Seiden MV, Bragden MJ, Beale MG. Alterations of IgM, IgG, and IgA synthesis and secretion by peripheral blood and intestinal mononuclear cells from patients with ulcerative colitis and Crohn's disease. *Gastroenterology* 1981;81:844–852.
80. Scott MG, Nahm MH, Macke K, Nash GS, Bertovich MJ, MacDermott RP. Spontaneous secretion of IgG subclasses by intestinal mononuclear cells: differences between ulcerative colitis, Crohn's disease and controls. *Clin Exp Immunol* 1986;66:209–215.
81. Ochs HD, Wedgewood RJ. IgG subclass deficiencies. *Ann Rev Med* 1987;38:325.
82. Brandtzaeg P, Baklien K, Fausa O, Hoel PS. Immunohistochemical characterization of local immunoglobulin formation in ulcerative colitis. *Gastroenterol* 1974;66:1123–1136.
83. Gebbers J-O, Otto HF. Immunohistochemical and electronmicroscopic obervations in Crohn's disease. In: *Recent Advances/In Crohn's Disease* (Pena AS, Weterman IT, Booth CC, Strober W, eds). The Hague: Martinus Nijhoff, 1981, pp. 136–142.
84. MacDermott RP, Nash GS, Bertovich MJ, Mohrman RF, Kodner IJ, Delacroix DL, Vaerman J-P. Altered patterns of secretion of monomeric IgA and IgA subclass 1 by intestinal mononuclear cells in inflammatory bowel disease. *Gastroenterology* 1986;91:379–385.
85. Clancy R, Pucci A. Absence of K cells in the human gut mucosa. *Gut* 1978;19:273–276.
86. Fiocchi C, Battito JR, Farmer RG. Gut mucosal lymphocytes in inflammatory bowel disease. Isolation and preliminary characterization. *Dig Dis Sci* 1979;24:705–717.
87. Chiba M, Shorter RG, Thayer WR, Bartnik W, ReMine S. K-cell activity in lamina proprial lymphocytes from the human colon. *Dig Dis Sci* 1979;24:817–822.
88. MacDermott RP, Franklin GO, Jenkins KM, Kodner IJ, Nash GS, Weinrieb IJ. Human intestinal mononuclear cells. 1. Investigation of antibody-dependent, lectin-induced, and spontaneous cell-mediated cytotoxic capabilities. *Gastroenterology* 1980;78:47–56.
89. Bland PW, Britton DC, Richens ER, Pledger JV. Peripheral, mucosal and tumor-infiltrating components of cellular immunity in cancer of the large bowel. *Gut* 1981;22:744–751.
90. Bland PW, Richens ER, Britton DC, Lloyd JV. Isolation and purification of human large

bowel mucosal lymphoid cells: Effect of separation technique on functional characteristics. *Gut* 1979;20:1037–1046.
91. Chiba M, Bartnik W, ReMine SG, Thayer WR, Shorter RG. Human colonic intraepithelial and lamina proprial lymphocytes: Cytotoxicity in vitro and the potential effects of the isolation method on their functional properties. *Gut* 1981;22:177–186.
92. Gibson PR, Hermanowicz A, Verhaar HJJ, Ferguson DJP, Lopez Bernal A, Jewell DP. Isolation of intestinal mononuclear cells: Factors released which may affect lymphocyte viability and function. *Gut* 1985;26:60–68.
93. Persussia B, Trinchieri G, Cerottini JC. Functional studies of Fc receptor-bearing human lymphocytes: Effects of treatment with proteolytic enzymes. *J Immunol* 1979;123:681–687.
94. Fiocchi C, Tubbs RR, Youngman KR. Human intestinal mucosal mononuclear cells exhibit lymphokine-activated killer cell activity. *Gastroenterology* 1985;88:625–637.
95. Shanahan F, Brogan M, Targan S. Human mucosal cytotoxic effector cells. *Gastroenterology* 1987;92:1951–1957.
96. Auer IO, Zeimer E, Sommer H. Immune status in Crohn's disease. V. Decreased in vitro natural killer cell activity in peripheral blood. *Clin Exp Immunol* 1980;42:41–49.
97. Beeken WL, MacPherson BR, Gundel RM, St. Andre-Ukena S, Wood SG, Sylvestor DL. Depressed spontaneous cell-mediated cytotoxicity in Crohn's disease. *Clin Exp Immunol* 1983;51:351–358.
98. Ginsburg CH, Dambrauskas JT, Ault KA, Falchuk ZM. Impaired natural killer cell activity in patients with inflammatory bowel disease: Evidence for a qualitative defect. *Gastroenterology* 1983;85:846–851.
99. Gibson PR, Verhaar HJJ, Selby WS, Jewell DP. The mononuclear cells of human mesenteric blood, intestinal mucosa and mesenteric lymph nodes: Compartmentalisation of NK cells. *Clin Exp Immunol* 1984;56:445–452.
100. Falchuk ZM, Barnhard E, Machado I. Human colonic mononuclear cells: Studies of cytotoxic function. *Gut* 1981;22:290–294.
101. Targan S, Britvan L, Kendall R, Vimadalal S, Soll A. Isolation of spontaneous and interferon inducible natural killer like cells from human colonic mucosa: Lysis of lymphoid and autologous epithelial target cells. *Clin Exp Immunol* 1983;54:14–22.
102. Beeken WL, Gundel M, St. Andre-Ukena S, McAuliffe T. In vitro cellular cytotoxicity for a human colon cancer cell line by mucosal mononuclear cells of patients with colon cancer and other disorders. *Cancer* 1985;55:1024–1029.
103. Gibson PR, Jewell DP. Local immune mechanisms in inflammatory bowel disease and colorectal carcinoma. Natural killer cells and their activity. *Gastroenterology* 1986;90:12–19.
104. Gibson PR. NK cells, IBD and cancer. *Gastroenterology* 1986;90:1314–1315.
105. Hogan PG, Hapel AJ, Doe WF. Lymphokine-activated and natural killer cell activity in human intestinal mucosa. *J Immunol* 1985;135:1731–1738.
106. Rosenberg SA, Lotze MT, Muul LM, Chang AE, Avis FP, Leitman S, Linehan WM, Robertson CN, Lee RE, Rubin JT, Seipp CA, Simpson CG, White DE. A progress report on the treatment of 157 patients with advanced cancer using lymphokine-activated killer cells and interleukin-2 of high dose alone. *N Engl J Med* 1987;316:889–897.
107. Itoh K, Tilden AB, Kumagai K, Balch CM. Leu 11^+ lymphocytes with natural killer (NK) activity are precursors of recombinant interleukin-2 (rIL-2)-induced activated killer (AK) cells. *J Immunol* 1985;134:802–807.
108. Phillips JH, Lanier LL. Dissection of the lymphokine activated killer phenomenon. Relative contribution of peripheral blood natural killer cells and T lymphocytes to cytolysis. *J Exp Med* 1986;164:814–825.
109. Fiocchi C, Hilfiker ML, Youngman KR, Doerder NC, Finke JH. Interleukin-2 activity of human intestinal mucosal mononuclear cells. Decreased levels in inflammatory bowel disease. *Gastroenterology* 1984;86:734–742.

110. MacDermott RP, Bragdon MJ, Kodner IJ, Bertovich MJ. Deficient cell-mediated cytotoxicity and hyporesponsiveness to interferon and mitogenic lectin activation by inflammatory bowel disease peripheral blood and intestinal mononuclear cells. *Gastroenterology* 1986;90:60–111.
111. Shanahan F, Brogan M, Newman W, Targan S. K562 killing by K and KL-2-responsive NK and T cells involves different post binding trigger mechanisms. *J Immunol* 1986;137:723–726.
112. Weiss A, Imboden J, Hardy K, Manger B, Terhost C, Stobo J. The role of the T3/antigen receptor complex in T-cell activation. *Ann Rev Immunol* 1986;4:593–619.
113. Phillips JH, Lanier LL. Lectin-dependent and anti-CD3 induced cytotoxicity are preferentially mediated by peripheral blood cytotoxic T lymphocytes expressing Leu-7 antigen. *J Immunol* 1986;136:1579–1585.
114. Shanahan F, Deem R, Nayersina R, Leman B, Targan S. Human mucosal T-cell cytotoxicity. *Gastroenterology* 1988;94:960–967.
115. Shanahan F, Leman B, Deem R, Niederlehner A, Brogan M, Targan S. Enhanced peripheral blood T cell cytotoxicity in inflammatory bowel disease. J Clin Immunol (in press).
116. LeFevre ME, Hammer R, Joel DD. Macrophages of the mammalian small intestine: a review. *J Reticuloendothelial Soc* 1979;26:553–573.
117. Tanner AR, Arthur MJP, Wright R. Macrophage activation, chronic inflammation, and gastrointestinal disease. *Gut* 1984;25:760–783.
118. Thyberg J, Graf W, Klingenstrom P. Intestinal fine structure in Crohn's disease. Lysosomal inclusions in epithelial cells and macrophages. *Virchows Arch (Pathol Anat)* 1981;39:141–152.
119. Meurat G, Bitzi A, Hammer B. Macrophage turnover in Crohn's disease and ulcerative colitis. *Gastroenterology* 1978; 74:501–503.
120. Ward M. The pathogenesis of Crohn's disease. *Lancet* 1977;2:903–905.
121. Zweiman B. Mast cells in human disease. *Clin Rev Allergy* 1983;1:417–426.
122. Lloyd G, Green FHT, Fox H, Mani V, Turnberg LA. Mast cells and immunoglobulin E in inflammatory bowel disease. *Gut* 1975;16:861–866.
123. Ranlov P, Nielson MH, Wanstrup J. Ultrastructure of the ileum in Crohn's disease: Immune lesions and mastocytosis. *Scand J Gastroenterol* 1972;7:471–476.
124. Rao SN. Mast cells as a component of the granuloma in Crohn's disease. *J Pathol* 1973;109:79–82.
125. Heatley RV, Calcraft BJ, Rhodes J, Owen E, Evans BK. Disodium cromoglycate in the treatment of chronic proctitis. *Gut* 1975;16:559–563.
126. Shanahan F, Denburg JA, Bienenstock J, Befus AD. Mast cell heterogeneity. *Can J Physiol Pharmacol* 1984;62:734–737.
127. Befus AS, Bienenstock J, Denburg JA (eds). *Mast Cell Differentiation and Heterogeneity.* New York: Raven Press, 1986.
128. Saverymuttu SH, Peters AM, Lavender JP, Pepys MB, Hodgson HJF, Chadwick VS. Quantitative fecal indium 111-labelled leukocyte excretion in the assessment of disease in Crohn's disease. *Gastroenterology* 1983;85:1333–1339.
129. Savermuttu SH, Chadwick VS, Hodgson HJ. Granulocyte migration in ulcerative colitis. *Eur J Clin Invest* 1985;15:60–68.
130. Sharon P, Stenson WF. Enhanced synthesis of leukotriene B4 by colonic mucosa in inflammatory bowel disease. *Gastroenterology* 1984;86:453–460.
131. Rampton DS, Hawkey CF. Prostaglandins and ulcerative colitis. *Gut* 1985;25:1399–1413.
132. Donowitz M. Arachidonic acid metabolites and their role in inflammatory bowel disease. *Gastroenterology* 1985; 88:580–587.
133. Lobos EA, Sharon P, Stenson WF. Chemotactic activity in inflammatory bowel disease. Role of leukotriene B4. *Dig Dis Sci* 1987;32:1380–1388.
134. Stenson WF. Pharmacology of sulfasalazine. *Viewpts Dig Dis* 1984;16:13–16.

135. Rhodes JM, Bartholomew TC, Jewell DP. Inhibition of leucocyte motility by drugs used in ulcerative colitis. *Gut* 1981;22:642–647.
136. MacDermott RP, Kane MG, Steele LL, Stenson WF. Inhibition of cytotoxicity by sulfasalazine. I. Sulfasalazine inhibits spontaneous cell-mediated cytotoxicity by peripheral blood and intestinal mononuclear cells from control and inflammatory bowel disease patients. *Immunopharmacology* 1986;11:101–109.
137. Shanahan F, Niederlehner A, MacDermott RP, Stenson WF, Kane MG, Targan S. Inhibition of cytotoxicity by sulfasalazine. II. Sulfasalazine and sulfapyridine inhibit different stages of the NK and NKCF lytic process. *Immunopharmacol* 1986;11:111–118.
138. Leickly FE, Buckley RH. Development of IgA and IgG2 subclass deficiency after sulfasalazine therapy. *J Pediatr* 1986;108:481–482.
139. Goldstein F. Immunosuppressant therapy of inflammatory bowel disease. *J Clin Gastroenterol* 1987;9:654–658.
140. Donoghue DP, Dawson AM, Powell-Tuck J, Brown RL, Lennard-Jones JE. Double-blind withdrawal trial of azathioprine as maintenance treatment for Crohn's disease. *Lancet* 1978;2:955–957.
141. Present DH, Korelitz BI, Wisch N, Glass JL, Sacher DB, Pasternack BS. Treatment of Crohn's disease with 6-mercaptopurine. A long-term, randomized, double-blind study. *N Engl J Med* 1980;302:981–987.
142. Campbell AC, Skinner JM, Maclennan ICM, Hersey P, Waller CA, Wood J, Jewll DP, Truelove SC. Immunosuppression in the treatment of inflammatory bowel disease. II. The effects of azathioprine on lymphoid cell populations in a double blind trial in ulcerative colitis. *Clin Exp Immunol* 1976;24:249–258.
143. Brogan M, Hieserodt J, Oliver M, Stevens R, Korelitz B, Targan S. The effect of 6-mercaptopurine on natural killer-cell activities in Crohn's disease. *J Clin Immunol* 1985;5:204–211.
144. Fahey JL, Sarna G, Gale RP, Seeger R. Immune interventions in disease. *Ann Intern Med* 1987;106:257–274.
145. Leman B, Shanahan F, Targan S. Serial study of immunologic alterations in patients on long-term treatment with 6-MP: effect on CTL (abstract). *Gastroenterology* (in press).
146. Shanahan F, Anton P. Neuroendocrine modulation of the immune system. Possible implications for inflammatory bowel disease. *Dig Dis Sci* 1988;33:415–495.
147. Madara JL, Podolsky DK, King NW, Sehgal P, Moore R, Winter HS. Characterization of spontaneous colitis in cotton-top tamarins *(Sanguinus oedipu)* and its response to sulfasalazine. *Gastroenterology* 1985;88:13–19.
148. James SP, Strober W, Quinn TC, Danovitch SH. Crohn's disease. New concepts of pathogenesis and current approaches to treatment. *Dig Dis Sci* 1987;32:1297–1310.

Index