Contributors

- Johan Bjorksten
- Leonard K. Dunikoski, Jr.
- Masood Haider
- Walter E. Hordynsky
- Louis A. Kazal
- Sheshadri Narayanan
- C. E. Pippenger

Clinical Biochemistry

Contemporary Theories and Techniques

VOLUME 2

Edited by
HERBERT E. SPIEGEL

Clinical Biochemistry Laboratory
Hoffmann-La Roche, Inc.
Nutley, New Jersey

1982

ACADEMIC PRESS
A Subsidiary of Harcourt Brace Jovanovich, Publishers
New York London
Paris San Diego San Francisco São Paulo
Sydney Tokyo Toronto

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data
Main entry under title:

Clinical biochemistry.

Includes bibliographies and indexes.
1. Chemistry, Clinical. 2. Biological chemistry. I. Spiegel, Herbert E. [DNLM: 1. Chemistry, Clinical. QY 90 C6406]
RB40.C56 616.07'56 81-14933
ISBN 0-12-657102-3 (v.2) AACR2

PRINTED IN THE UNITED STATES OF AMERICA

82 83 84 85 9 8 7 6 5 4 3 2 1

This volume is dedicated to the memory of Drs. Alan I. Fleischman and Richard J. Coumbis. The premature death of these distinguished scientists has left a void in the hearts of their many colleagues and friends.

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

Johan Bjorksten (209), Bjorksten Research Foundation, Madison, Wisconsin 53715

Leonard K. Dunikoski, Jr. (45), Department of Laboratories, Perth Amboy General Hospital, Perth Amboy, New Jersey 08861

Masood Haider (145), Departments of Immunochemistry and Research and Development, Roche Clinical Laboratories, Raritan, New Jersey 08869

Walter E. Hordynsky (13), St. Mary's Hospital, Orange, New Jersey 07050

Louis A. Kazal (73), Cardeza Foundation for Hematologic Research, Departments of Medicine and Physiology; Jefferson Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Sheshadri Narayanan (1), Department of Pathology, New York Medical College, Metropolitan Hospital Center, New York, New York 10029

C. E. Pippenger (255), Department of Neurology, Columbia University, College of Physicians and Surgeons, New York, New York 10032

Preface

The purpose of this volume is to again present the ideas and philosophies of scientific leaders in the many collateral fields related to clinical chemistry. The editor has tried to retain the individual style of each contributor rather than to present a highly stylized, regimented volume. It is hoped that this unorthodox method of presentation will still reflect good and scholarly science.

A further intent of this volume is to supplement the existing textbooks of clinical chemistry and biochemistry by providing areas of emphasis such as coagulation chemistry, inborn errors of metabolism, and biochemistry of aging, which are not emphasized in standard clinical chemistry texts. Individual details are de-emphasized in favor of overall concepts and philosophy, where possible. Speculation and interpretation by the contributors has been encouraged, with the idea that these may be stimulating to the reader.

This book is a collection of scientific essays that is designed hopefully to acquaint the reader with the wide spectrum of sciences that is called clinical chemistry.

Herbert E. Spiegel

Acknowledgments

I would like to acknowledge the contributions and professionalism of all those who participated in assembling this volume. My secretary, Mrs. Joan Marks, again has proven to be a valuable and valiant assistant. Finally, to the members of my family, especially my wife Joanne, I express my special thanks for the gifts of patience and concern.

Contents of Volume 1

1

Biochemistry and Clinical Significance of Lipoprotein-X

SHESHADRI NARAYANAN

I. INTRODUCTION

Lipoprotein-X (LP-X) is an unique lipoprotein that has acquired clinical significance in that it is a marker for obstructive jaundice, since it is present in the sera of patients. It is also seen in the plasma of subjects suffering from a rare inborn error of metabolism—the deficiency of lecithin–cholesterol acyltransferase (LCAT).

LP-X has been characterized as a low-density lipoprotein from its behavior in the ultracentrifuge. It is distinct chemically, immunochemically, and electrophoretically from other lipoprotein species. Indeed, because of

CLINICAL BIOCHEMISTRY
Contemporary Theories and Techniques, Vol. 2

ISBN 0-12-657102-3

these distinct characteristics, it was named LP-X to set it apart from other lipoprotein species, such as LP-A and LP-B. The biochemistry of LP-X, in terms of its metabolism and role, can best be understood by gaining an insight into its unique chemical and structural properties and its diagnostic significance.

II. CHEMISTRY OF LP-X

Our knowledge of the chemistry of LP-X is based upon studies conducted with highly purified preparations. Procedures for purification of LP-X from cholestatic plasma entail separation from other lipoprotein fractions. This can be done by exploiting the density differences of the lipoprotein fractions in an ultracentrifuge. Thus, the very low-density lipoprotein fractions (VLDL) can be made to float by applying plasma beneath a layer of buffer with a density of 1.0055 g/ml, and centrifuging at a high speed (105,000 *g*) for 22 hr (*1,2*). The heavier lipoprotein fractions containing LP-X can be isolated by precipitation of the low-density lipoprotein fraction with heparin and manganese. Further manipulative steps include fractionation with ethanol, the adjustment of the density of the supernatant to 1.063 g/ml with sodium chloride, and ultracentrifugation at 105,000 *g* for 44 hr. LP-X bands in the fraction that has a floatation density of 1.006–1.063 g/ml.

Instead of utilizing sequential ultracentrifugation steps, isolation of LP-X can also be accomplished with zonal ultracentrifugation in a sodium bromide gradient, subsequent to a preliminary ultracentrifugation step to remove VLDL (*3*). Zonal ultracentrifugation procedures are less time-consuming than sequential ultracentrifugation. However, the former, because of the introduced effects of dilution, requires a concentration step by ultrafiltration.

The purity of LP-X isolated by zonal ultracentrifugation procedures is apparently dependent on both the duration and speed of centrifugation. While LP-X was isolated as a single species upon high speed zonal ultracentrifugation at 150,000 *g* for 135 min (*3*), lower speeds permitted fractionation of LP-X into two or three species (*4,5*).

Thus, rate zonal ultracentrifugation at 80,000 *g* for 2 hr results in the resolution of LP-X into two fractions, LP-X_1 and LP-X_2 (*4*). Centrifugation at the same speed for approximately 72 hr results in the appearance of another LP-X fraction, LP-X_3 (*5*).

LP-X can also be purified by chromatography on hydroxyapatite since both VLDL and β-lipoprotein are retained on the adsorbent, permitting early elution and separation of LP-X (*3*).

Zonal ultracentrifugation procedures without prior ethanol fractionation yield LP-X preparations that are not completely free from contaminating low-density lipoprotein, apolipoprotein-B(APO-B). Each of the LP-X fractions (LP-X_1, LP-X_2, and LP-X_3) have been purified to homogeneity by ethanol fractionation and zonal ultracentrifugation repeated twice in a linear gradient (*5*).

III. ANALYSIS OF LP-X PREPARATIONS

LP-X purified from human and animal cholestatic plasma are similar in chemical composition (*1,6*). When compared to other lipoprotein fractions (VLDL, LP-B, and high density lipoprotein—HDL), LP-X has a unique chemical composition with the phospholipid content higher than any other lipoprotein fraction. Phospholipid constitutes 66.5% of the LP-X molecule (*1,7*). The next major constituent is unesterified cholesterol, which constitutes 22.4%. Cholesterol esters and triglycerides together constitute approximately 5% of LP-X. Protein content is also low, amounting to approximately 6% of the LP-X molecule. The fact that the esterifying capacity of cholesterol is impaired in patients with obstructive jaundice might explain the extremely low levels of cholesterol esters found in LP-X.

LP-X, when fractionated into three species (LP-X_1, LP-X_2, and LP-X_3), still retains the characteristic phospholipid:unesterified cholesterol ratio seen in the native LP-X molecule (*5*). There are slight differences, however, in the amounts of the minor constituents (VLDL, cholesterol esters, and protein) that make up the three species of LP-X. Lecithin is the major and sphingomyelin is the minor phospholipid found in LP-X. The latter constitutes just one-fifth of the concentration of lecithin (*2*). The lecithin:sphingomyelin ratio of LP-X is almost twice that found in the LDL fraction, but is, however, similar to that found in the HDL fraction.

The 14-carbon fatty acid, myristic acid, is the major fatty acid constituting the cholesterol esters found in the three fractions of LP-X—(LP-X_1, LP-X_2, and LP-X_3 (*5*). It is interesting that the enzyme LCAT prefers a phosphatidyl choline donor of LP-X that has 14-carbon fatty acids (*5,6*). Linoleic acid has also been found as a constituent of cholesterol esters. Linolenic and higher unsaturated fatty acids are esterified to make up the triglycerides in the LP-X molecule. Differences in the fatty acid content of linoleic, linolenic and higher unsaturated fatty acids are seen in the LP-X fractions isolated from different individuals. The significance of this variability in the fatty acid content of LP-X isolated from cholestatic plasma of different individuals is not clear.

The bile acid content of LP-X has been reported to be variable, and ranges from 0.14–3% of the LP-X molecule. The major bile acid found in LP-X is the hepatotoxic lithocholic acid (*8*).

The protein moiety of LP-X is made up of apolipoprotein-X (APO-X), which constitutes 60% of the protein. The remainder of the protein moiety is contributed by albumin, which is tightly associated with the apoprotein. Albumin is implicated in the maintainence of the structural integrity of the LP-X molecule.

The amino acid composition of APO-X is unique in that it differs from that found in apoproteins A and B. APO-X has no cystine or cysteine, and its histidine content is very low (*2*). The major N-terminal amino acids found in APO-X are serine and threonine, whereas the major C-terminal amino acid found was alanine. The composition of APO-X is identical to that of apolipoprotein-C (APO-C) obtained from the partially delipidized VLDL fraction in plasma (*9*). Our current knowledge is that both APO-X and APO-C have three different polypeptide chains made up of N-terminal serine and C-terminal alanine, N-terminal threonine and C-terminal valine, and N-terminal threonine and C-terminal glutamic acid (*2*). Both APO-X and APO-C have similar phospholipid to protein ratios.

Both APO-X and APO-C have a great affinity for phospholipid, as is evidenced by the high phospholipid to protein ratio of 11.5 found in LP-X. The ability to bind phospholipid is related to the presence of helical structures in both APO-X and APO-C. The helical structure with its polar and nonpolar halves presents a conformation that is conducive to the binding of phospholipid. The hydrophobic content of the helix determines the amount of phospholipid bound (*6,10*). Differences are seen in the α-helicity of the three fractions of LP-X (*5*). The secondary structure of APO-X (APO LP-X_1 and APO LP-X_3) is similar to human serum albumin, which is a constituent of LP-X. This secondary structure is relatively stable. LP-X_2 and LP-X_3 contain, in addition to APO-C, the apoproteins APO A-1 and APO-E. APO A-1 is an activator of the enzyme lecithin–cholesterol acyltransferase (LCAT) (*11*). APO-E is rich in the amino acid arginine, and appears in the plasma of patients with a deficiency of LCAT and in type III hyperlipidemia (*12*).

The authenticity of these analytical data obtained with purified LP-X preparations has been documented by immunochemical studies (*1,5,6*). The identity of APO-X and APO-C was established in immunodiffusion experiments using antisera to either APO-C or LP-X (*2*). By performing immunoelectrophoresis of APO-X in 1% agar, the presence of three species of LP-X was demonstrated by the formation of three immunoprecipitin arcs with antibody to LP-X (*2*). The fact that LP-X reacts with antisera to albumin, only upon delipidation, is consistent with the tight association of albumin with APO-X. Our knowledge from immunochemical studies on

the structure of LP-X is that APO C-II and albumin are located within the core of the molecule. The reactivity of the three species of intact LP-X ($LP\text{-}X_1$, $LP\text{-}X_2$, and $LP\text{-}X_3$) with antisera to APO C-I and APO C-III suggests that these two apoproteins are on the surface of the LP-X molecule (*5*). Phospholipids are also located on the surface of LP-X, as demonstrated by the denaturation of LP-X on treatment with phospholipase A_2 (*13*).

Studies using electron microscopy (*13*), X-ray small-angle scattering (*14*), and electron paramagnetic resonance (*5*) have provided valuable information on the structure of LP-X. Our knowledge based on these studies reveals that LP-X is heterogeneous and that the three species ($LP\text{-}X_1$, $LP\text{-}X_2$, and $LP\text{-}X_3$) are very rigid particles. This rigidity is due to its unique chemical composition reflected by its high concentration of free cholesterol and the 14-carbon saturated fatty acid. These rigid particles are spherical vesicles with a lipid bilayer of cholesterol and phospholipids in combination with APO-X. A proposed model for LP-X visualizes it as being a spherical particle which is nearly 400 Å in diameter (*10*). In this model APO-C-I and APO-D are on the outer surface of the particle, with albumin, APO-C-II, and APO-C-III located within the inner core. Albumin exists in the inner core in solution without being bound to the phospholipid-free cholesterol bilayer. This model also depicts a part of the APO-X moiety in combination with the polar head of the phospholipids.

IV. PROPERTIES OF LP-X

LP-X aggregates with membrane-bound enzymes. Complexes of LP-X with alkaline phosphatase have been isolated from cholestatic human sera and characterized. Complexes of LP-X with γ-glutamyltranspeptidase, 5′-nucleotidase, and nucleotide pyrophosphorylase have also been demonstrated (*6,15*). Enlargement of the erythrocyte membrane has been noticed in patients with cholestasis. Electron microscopic studies have ascribed the cause of this enlargement as due to the fusion of LP-X vesicles with the erythrocyte membrane (*16*).

V. METHODOLOGY FOR THE MEASUREMENT OF LP-X

A. Electrophoresis

LP-X has a characteristic cathodal mobility on agar gel which lends itself to both qualitative and quantitative assay. By incorporating a drop of a lipid stain, such as sudan black, at the point of application of the sample

to the agar gel, it is possible to visualize the migration of LP-X. The concentration of agar commonly used is 1%, in barbital buffer, pH 8.6. A typical electrophoretic run lasts 90 min. (*6*). A variety of procedures are available for the quantitation of LP-X subsequent to electrophoresis. It is possible to excise the agar gel corresponding to the migration area of LP-X and quantitate LP-X by performing lipid phosphorus determinations (*17*). Alternatively, LP-X and other LDL fractions can be precipitated by polyanionic compounds, such as heparin, and metal salts, such as magnesium chloride, and the precipitated band of LP-X quantitated by densitometric scanning (*18*). Prior to electrophoresis on agar gel, if the serum is incubated with radiolabeled cholesterol, the label will equilibrate with the various lipoprotein fractions. Subsequent electrophoresis and determination of percent of radioactivity associated with the cathodal area coincident with the migration of LP-X is used to quantitate LP-X (*19*). The sensitivity of these modified electrophoretic procedures vary.

Electrophoresis, with subsequent precipitation by heparin and magnesium chloride, and densitometric scanning can quantitate LP-X in the range of 60 mg–6.3 g/liter. Levels above 6.3 g/liter, seen in cases of severe obstruction, can be quantitated upon appropriate dilution with sera free from LP-X. It must be noted that LP-X levels lower than 60 mg/liter can be encountered in early stages of cholestasis, which might escape detection by the above procedures.

Procedures employing electrophoresis with subsequent lipid phosphorus determination on the excised agar-gel portion containing LP-X are unable to quantitate levels below 500 mg/liter.

B. Immunochemical Techniques

The availability of antisera to highly purified LP-X preparations permits one to measure LP-X. However, procedures using radial immunodiffusion are time-consuming. Since antigenic determinants on LP-X are shared by other lipoproteins, cross-reactivity is also a problem. This has been circumvented by prior incubation of serum with antilipoprotein-B to precipitate lipoprotein-B and subsequent measurement of the supernatant containing LP-X by either radial immunodiffusion or electroimmunodiffusion. The electroimmunodifussion procedure can be performed in 3 hr in contrast to the 72 hr required for the radial immunodiffusion procedure. The former procedure can be shortened by another 75 min by performing the precipitation step with antilipoprotein-B on the agarose gel itself (*20*).

The sensitivity of immunochemical procedures is inferior to that attained by electrophoresis followed by precipitation with polyanionic compounds and densitometry. The radial immunodiffusion procedure (*18*) is

unable to quantitate LP-X levels below 300 mg/liter. The electroimmunodiffusion procedure is only slightly better with the lower limit of detection around 200 mg/liter.

Immunochemical procedures utilizing prior precipitation with antilipoprotein-B may not completely eliminate the cross-reactivity problem because some VLDL subspecies and lipoprotein-C related to HDL are not reactive with antilipoprotein-B. However, the levels of these species in human sera unrecognizable by antilipoprotein-B are negligible, and thus do not present a serious limitation in the quantitation of LP-X.

One must also be wary of pitfalls associated with electrophoretic procedures that are related to the sample and the electrophoretic media. LP-X bands seen in fresh sera are known to disappear upon storage. Batch to batch variability of agar used for preparation of the gel may have an effect upon the migration characteristics of LP-X.

VI. LP-X AND LCAT ACTIVITY

The appearance of LP-X in the plasma of patients with cholestasis is known to depress plasma levels of the LCAT enzyme. This enzyme catalyzes the transfer of the acyl group in the 2-position of phosphatidyl choline to the hydroxyl group of cholesterol. LP-X levels exceeding 2.5 g/liter completely abolish the LCAT activity in plasma (*21*). Correction of biliary obstruction by surgery leads to a decrease in LP-X accompanied by an increase in plasma LCAT activity (*22*).

The rigid structure of LP-X could partly explain why it is a poor substrate for LCAT, although the presence of short-chain fatty acyl donors, such as myristate, associated with the phosphatidyl choline of LP-X, should facilitate acyl group transfer (*5,23*). The presence of bile salts on the surface of LP-X and its characteristic free cholesterol to phospholipid ratio have been implicated in the inhibition of LCAT activity (*21*).

VII. CLINICAL USEFULNESS

The rare familial deficiency of the enzyme LCAT is associated with increased plasma LP-X levels. Our current knowledge is that there are subtle differences between LP-X associated with LCAT deficiency and the species encountered in cholestasis. Thus the apolipoprotein-A-1 (APO-A_1) found in LP-X associated with LCAT deficiency is found only in the LP-X_2 and LP-X_3 fractions of cholestatic LP-X and is missing in the LP-X_1 fraction (*5*). Electron micrographs of LDL fractions obtained from

plasma lacking the LCAT resemble an unique triglyceride-rich lipoprotein referred to as lipoprotein-Y (LP-Y) which has also been seen in plasma of patients with cholestasis (*24*). Incidentally, electrophortic procedures fail to distinguish LP-X from LP-Y. This has to be done either by immunochemical methods or by electron microscopy (*25*).

The diagnostic usefulness of LP-X determinations is primarily aimed at the confirmation of cholestasis. The demonstration of the presence of LP-X in sera even in the absence of clinically significant increases in alkaline phosphatase activity, is indicative of cholestasis (*26*). The specificity of LP-X in cholestasis has been demonstrated by its correlation with histological findings. A limitation, however, is the inability of LP-X to distinguish between intra- and extrahepatic cholestasis (*27*). The presence of extrahepatic cholestasis is invariably accompanied by presence of LP-X in sera. The levels of LP-X encountered in such instances may vary from a median value of 160 mg/dl to greater than 400 mg/dl. In the latter instance increases in alkaline phosphatase activity and bilirubin become evident, and the presence of extraphepatic cholestasis is no longer in doubt (*28*).

In contrast, the presence of intrahepatic cholestasis, in many instances, is not accompanied by the appearance of LP-X in sera (*27*). Although LP-X levels greater than 300 mg/dl have been noted in sera of patients with intrahepatic cholestasis, in general, whenever LP-X is present in sera of patients with intrahepatic cholestasis, the median value is much lower (approximately 25 mg/dl) than that found in extrahepatic cholestasis (approximately 160 mg/dl) (*28*).

While the presence of intrahepatic cholestasis can be missed by relying solely on the absence of LP-X in sera, in instances where LP-X has been demonstrated, histological findings have been consistent with the existence of the above-mentioned disease state (*29*).

The appearance of LP-X in the sera of the newborn is not always diagnostic of cholestasis. Due to immature liver function, a high incidence of false positives, in terms of appearance of LP-X in sera in the absence of cholestasis, has been noted in the first week of the life of the newborn (*30*). The usefulness of serum LP-X as an indicator of cholestasis in infants is applicable only after the baby is over 1 year of age.

The identification of jaundiced infants in whom surgery would be indicated is possible by determining LP-X in sera prior to and after the administration of a bile acid-binding agent, such as cholestyramine. The latter decreases the enterophepatic circulation of bile acid by binding it in the intestinal lumen and excreting it in the feces. Surgical internvetion is unnecessary, if LP-X disappears from circulation after a 2-week oral intake of cholestyramine, thus indicating that cholestasis was of the intrahepatic type (*31*).

Fractionation and determination of levels of the three species of LP-X ($LP\text{-}X_1$, $LP\text{-}X_2$, and $LP\text{-}X_3$) may be of value in assessing the state of the liver disease (*5*). Among the three species of LP-X encountered, $LP\text{-}X_3$ levels are always the lowest. Subsequent to surgery, the denser species ($LP\text{-}X_2$) is cleared faster than $LP\text{-}X_1$ (*4*).

Another useful indicator of extrahepatic cholestasis is the correlation of the α_2-β band of γ-glutamyl transpeptidase in electrophoresis on cellulose acetate gel, in the presence of LP-X (*32*). Although the presence of LP-X in sera has the greatest positive correlation with extrahepatic cholestasis, it has also been reported in patients with other kinds of liver disease, such as viral hepatitis (*33*), particularly in children, and in nonhemolytic jaundice (*34*).

One must be wary of interpreting the disappearance of LP-X in sera of patients receiving heparin. This appearance is not due to the alleviation of extrahepatic cholestasis, but to the activation of the enzyme phospholipase, which converts lecithin to lysolecithin, and thereby degrades LP-X (*35*).

VIII. METABOLISM

Our knowledge of the metabolism of LP-X has been derived from studies in which biliary obstruction was induced in experimental animals. LP-X appears in plasma 5–48 hr after inducing biliary obstruction depending on species experimented upon, such as rats, dogs, or pigs (*7,36*). After reaching peak levels in the plasma, LP-X levels drop and are finally undetectable by either electrophoretic or immunochemical procedures 190–240 hr after inducing biliary obstruction. The characteristic phospholipid: cholesterol ratio of LP-X was, however, maintained even after the assay procedures were unable to detect LP-X. The initial rise and the eventual disappearance of LP-X in bile duct-ligated experimental animals have been attributed to the reflux of lipids from bile into the plasma (*2*). Evidently, a certain concentration of bile salts are needed to maintain the integrity of LP-X, with higher concentrations proving to be inhibitory. Additional evidence has supported the theory that biliary lipoprotein is the precursor of LP-X. In these experiments, the addition of LP-X-negative sera or albumin to native bile resulted in a lipoprotein with electrophoretic mobility on agar, very similar to that of LP-X. When this lipoprotein species was incubated with other isolated lipoproteins, such as HDL and VLDL, a transfer of apoprotein-C (APO-C) from these isolated lipoproteins to the induced LP-X molecule occurred, and the resulting species had an identical cathodal electrophoretic mobility to LP-X on agar-gel electrophoresis (*37*). This characteristic cathodal mobility was, however,

altered in a slightly anodal direction, upon addition of bile salts to this albumin-induced LP-X species.

These results explain the sequence of events that occur upon induction of biliary obstruction in terms of appearance, peaking, and ultimate disappearance of LP-X in the plasma. Apparently, a certain albumin:bile salt ratio is necessary to achieve the characteristic cathodal mobility of LP-X on agar-gel electrophoresis, with an excess of bile salts abolishing this typical cathodal mobility. APO-C appears to stabilize LP-X and preserves its cathodal mobility on agar (*37*). However, in some experiments it was noted that following the initial rise and fall in LP-X levels upon inducing biliary obstruction, an increase in LP-X levels again occurred 2–3 weeks after the original induction of obstruction, the significance of which is not well-understood.

The catabolism of LP-X is apparently mediated through the action of the enzyme phospholipase present in the plasma. Evidence in this direction has been obtained from experiments that resulted in the degradation of LP-X upon incubation with the enzyme phospholipase A_2, and also upon injection of heparin by its activating affect on phospholipase in the plasma (*8,36*). Recent work, however, indicates that phospholipase merely abolishes the cathodal electrophoretic mobility of LP-X on agar gel without completely degrading it (*38*). The mechanism of this effect is, apparently, due to the increase in the lysolecithin content of the phospholipase-treated LP-X, thus conferring on the altered LP-X species a slightly anodal electrophoretic mobility on agar gel. Evidence for this reasoning has been obtained from *in vitro* experiments in which addition of oleic acid induced a slightly anodal electrophoretic mobility on LP-X (*38*). The oleic acid-induced anodal mobility was reversed towards its original characteristic cathodal direction upon addition of albumin, thereby implying that LP-X is not actually degraded by the action of phospholipase (*38*).

IX. CONCLUSION

The appearance of LP-X in extrahepatic cholestasis, together with its unique chemical composition, its aggregating properties, and its inability to incorporate cholesterol ester, even in the face of increased cholesterol ester levels, raises questions about its exact role in biliary obstruction. The readiness with which LP-X incorporates free cholesterol suggests that in biliary obstruction, it, apparently, is able to remove free cholesterol from the circulation (*21*). In this regard, the aggregating properties of LP-X also play a role.

Further characterization of the three species of LP-X (LP-X_1, LP-X_2,

and LP-X_3) and a better understanding of the diagnostic usefulness of LP-X by the introduction of more specific assays for each of the species, coupled with the increased knowledge of the triglyceride-rich lipoprotein, LP-Y, are in the realm of future research into this unique lipoprotein.

REFERENCES

1. Seidel, D., Alaupovic, P., and Furman, R. H., *J. Clin. Invest.* **48,** 1211 (1969).
2. Seidel, D., Alaupovic, P., Furman, R. H., and McConathy, W. J., *J. Clin. Invest.* **49,** 2396 (1970).
3. Danielsson, B., Johansson, B. G., and Petersson, B. G., *Clin. Chim. Acta* **47,** 365 (1973).
4. Patsch, J. R., Patsch, W., Sailer, S., and Braunsteiner, H., *Biochim. Biophys. Acta* **434,** 419 (1976).
5. Patsch, J. R., Aune, K. C., Gotto, A. M., Jr., and Morrisett, J. D., *J. Biol. Chem.* **252,** 2113 (1977).
6. Narayanan, S., Lipoprotein-X. *CRC Crit. Rev. Clin. Lab. Sci.* **11,** 31 (1979).
7. Muller, P., Baggio, G., Barth, H. O., Wieland, H., Anari, M., and Seidel, D., *Clin. Chim. Acta* **80,** 487 (1977).
8. Seidel, D., Agostini, G., and Muller, P., *Biochim. Biophys. Acta* **260,** 147 (1972).
9. Gustafson, A., Alaupovic, P., and Furman, R. H., *Biochemistry* **4,** 596 (1965.).
10. Morrisett, J. D., Jackson, R. L., and Gotto, A. M., Jr., *Biochim. Biophys. Acta* **472,** 93 (1977).
11. Fielding, C. J., Shore, V. G., and Field, P. E., *Biochem. Biophys. Res. Commun.* **46,** 1493 (1972).
12. Utermann, G., Menzel, H. J., and Langer, K. H., *FEBS Lett.* **45,** 29 (1977).
13. Seidel, D., Agostini, B., and Muller, P., *Biochim. Biophys. Acta* **260,** 146 (1972).
14. Laggner, P., Glatter, O., Muller, K., Kratky, O., Kostner, G., and Holasek, A., *Eur. J. Biochem.* **77,** 165 (1977).
15. Brocklehurst, D., Lathe, G. H., and Aparico, S. R., *Clin. Chim. Acta* **67,** 269 (1976).
16. Verkleij, A. P., Nauta, I. L. D., Werre, J. M., Mandersloot, J. G., Reinders, B., Ververgaert, P. H. J. T., and DeGier, J., *Biochim. Biophys. Acta* **436,** 366 (1976).
17. Magnani, H. N., and Alaupovic, P., *Clin. Chim. Acta* **38,** 405 (1972).
18. Neubeck, W., and Seidel, D., *Clin. Chem. (Winston-Salem, N.C.)* **21,** 853 (1975).
19. Ritland, S., *Clin. Chim. Acta* **55,** 359 (1974).
20. Kostner, G. M., Petek, W., and Holazek, A., *Clin. Chem. (Winston-Salem, N.C.)* **20,** 676 (1974).
21. Magnani, H. N., *Biochim. Biophys. Acta* **450,** 390 (1976).
22. Patsch, W., Patsch, J. R., and Sailer, S., *Eur. J. Clin. Invest.* **33,** Suppl. 137, 107 (1974).
23. Patsch, J. R., Soutar, A. K., Morrisett, J. D., Gotto, A. M., Jr., and Smith, L. C., *Eur. J. Clin. Invest.* **7,** 213 (1977).
24. Forte, T., Nichols, A., Glomset, J., and Norum, K., *Scand. J. Clin. Lab. Invest.* **33,** Suppl. 137, 121 (1974).
25. Kostner, G. M., Laggner, P., Prexl, H. J., Holasek, A., Ingolic, E., and Geymayer, W., *Biochem. J.* **157,** 401 (1976).
26. Michael, B., and Ritter, U., *Z. Gastroenterol.* **14,** 556 (1976).
27. Magnani, H. N., and Alaupovic, P., *Gastroenterology* **71,** 87 (1976).
28. Ritland, S., *Scand. J. Gastroenterol.* **10,** 5 (1975).

29. Milewski, B., and Palynyczko, Z., *Pol. Med. Sci. Hist. Bull.* **15,** 551 (1975).
30. Witt, I., and Ober, M., *J. Clin. Chem. Clin. Biochem.* **14,** 197 (1976).
31. Campbell, D. P., and Williams, R. G., *Ann. Surg.* **184,** 89 (1976).
32. Degenaar, C. P., Thijssen, C., Van der Wal, G., and Berends, G. T., *Clin. Chim. Acta* **67,** 79 (1976).
33. Baldi, A., D'Aniello, P., Giuliano, G., Colucci, M., and De Crescenzo, A., *Quad. Sclavo Diagn. Clin. Lab.* **11,** 122 (1975).
34. Samside, G., Lundborg, H., Ivarsson, S., and Gustafson, A., *Acta Chir. Scand.* **142,** 187 (1976).
35. Rubies-prat., J., Ras, M. R., Frison, J. C., Masdeu, S., and Bacardi, R., *Clin. Chim. Acta* **75,** 117 (1977).
36. Seidel, D., Buff, H. U., Fauser, U., and Bleyl, U., *Clin. Chim. Acta* **66,** 195 (1976).
37. Manzato, E., Fellin, R., Baggio, G., Walch, S., Neubeck, W., and Seidel, D., *J. Clin. Invest.* **57,** 1248 (1976).
38. Muckle, T. J., Edwards, J. A., and Auckland, P., *Clin. Chem. (Winston-Salem, N.C.)* **23,** 2302 (1977).

2

Radioimmunoassay

WALTER E. HORDYNSKY

I. INTRODUCTION

Instruction in the fundamentals of immunology and radiochemistry is essential to the understanding of radioimmunoassay (RIA). The purpose of this chapter is to provide some general guidelines in RIA, with basic principles in immunology, radiology, and clinical chemistry. Complete details regarding individual opinions and methods may be found among the references.

CLINICAL BIOCHEMISTRY
Contemporary Theories and Techniques, Vol. 2

ISBN 0-12-657102-3

II. IMMUNOLOGICAL BACKGROUND IN RADIOIMMUNOASSAY

The mechanism by which antigen stimulates antigen-reactive cells and the system regulating development of immune responses is not yet fully elucidated.

Each individual has a unique genetic endowment for his immunity which is governed by the genetic code (*1,2*).

Animals without a functioning system of immunity could not survive. Immunity is generated by two divisions: humoral and cellular. The humoral division is responsible for the formation of antibodies. Antigens are taken up by macrophages where they react with RNA. In metabolized form, they are transferred to the plasma cells, or lymphocytes, which synthesize the antibody to each particular antigen. Thus, the best definition of antigens and antibodies is mutually corresponding. An antigen is a substance with a chemical structure that initiates the synthesis of a characteristic antibody when injected into an animal. An antibody is a plasma protein with characteristic reactive qualities, produced by an animal after exposure to an antigen.

Substances that are not antigenic can form antibodies, if they are covalently linked to proteins or other antigenic macromolecules before immunization (*3*).

The terms antigen and immunogen are not necessarily synonymous. Proteins, polysaccharides, and lipoproteins are usually antigenic as well as immunogenic. Steroid hormones are antigens, but they are not immunogens unless they are coupled to a larger molecule, e.g., albumin. Steroids are incomplete antibodies, or haptens, and must be linked to an immunogen to enhance the antibody reaction.

The immunoglobulins consist of multiple polypeptide chains held together by disulfide bridges. The synthesis of each chain is under separate genetic control. In certain diseases, more than one gene may exert this control (*4*).

The principal immunoglobulin is IgG, accounting for 70–90% of the total immunoglobulins. IgG with a 150,000 MW contains two identical light chains of kappa or lambda type, but never both, and two heavy chains of the gamma type. The chains are held together by disulfide bridges (Fig. 1). It is possible to split the molecule into three components. Two are identical-fragment antigen-binding Fab components. The third does not have the ability to combine with the antigen, but it can be crystallized, and thus is called the *crystallizable fragment*, Fc. The amino acid sequence of the heavy and light chains is constant, with the exception of short variable regions in both chains. This region is fully exposed to sol-

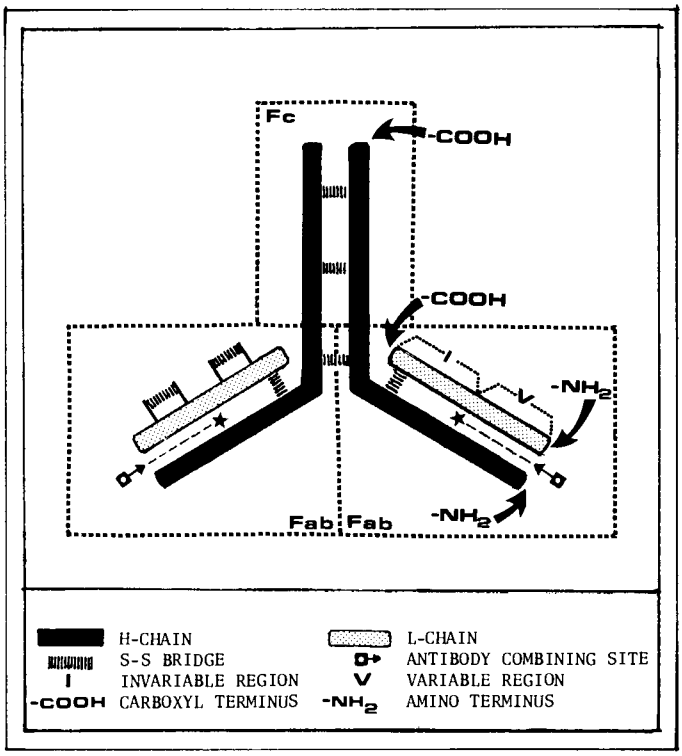


Fig. 1. Digrammatic model of the IgG molecule.

vent and has the characteristic of a combining site (5). The IgG molecule has two antigen-binding sites. There are four subclasses of IgG: IgG_1, IgG_2, IgG_3, and IgG_4 (6).

IgA has an alpha-type heavy chain, and exists in two forms: one in the plasma and the other in seromucous secretions. The plasma form has a 150,000 MW and a sedimentation of 7 S. Secretory IgA represents polymeric aggregates of 7 S units linked at the Fc ends by a J chain, a cysteine-rich polypeptide of 25,000 MW. IgA has two antigen-binding sites.

IgM is composed of five subunits (MW 850,000). Each subunit contains two heavy chains of mu type and two light chains. The subunits are linked by disulfide bridges and by the J chain. IgM has five active antigen-binding sites.

IgD has two heavy chains of delta type and, probably, two antigen-binding sites. The concentration of IgD in human plasma is very low. IgD, together with IgM, are immunoglobulins on the surface of lymphocytes (7).

IgE (MW 200,000) has two heavy chains of epsilon type, with a possibility of two antigen-combining sites. IgE is associated with reaginic activity, which makes this immunoglobin a unique antibody.

Antigen-antibody reactions display the following characteristics:

1. Allergens form antibodies called reagins.
2. Antienzymes are antibodies to enzymes, occurring in abnormal situations in the body, and commonly introduced from without.
3. Antibodies against viruses combine with them and prevent infection by inactivation of the viruses.
4. Antibodies against toxins inactivate them.
5. Opsonins are antibodies against the polysaccharide layer on the surface of the bacterial membrane.
6. Complement is projected against surface constituents of foreign cells in order to lyse them.
7. Agglutinins are antibodies that clump the antigens.
8. Precipitins are multivalent antibodies that precipitate *in vivo* and *in vitro* antigens out of suspension.

The fundamental interactions between antibodies and antigens are utilized *in vitro* as the basis for radioimmunoassay.

III. PREPARATION OF ANTIBODIES

The most important factor in establishing satisfactory RIA is the production of specific antibodies, which often depends on the immunogenic characteristics of the antigen.

Not all antigens are immunogens. Peptide hormones are usually antigens and immunogens. Peptide hormones with a 5000 MW and higher are good immunogens; however, it is possible to develop satisfactory antibodies against a number of unconjugated small peptides with a MW of about 1000, including angiotensin, vasopressin, and oxytocin (*8*).

A number of factors influence the formation of suitable antibodies for use in RIA. Especially for smaller molecules, the application of an adjuvant is important. An adjuvant causes local and general stimulation of the reticuloendothelial system and an increase in macrophages and immunologically responsive lymphoid cells (*9*).

The adjuvant contains light paraffin oil, emulsifiers and killed mycobacteria (*M. tuberculosis, M. butyricum, M. phlei,* and *M. smegmatis*). Equal volumes of antigen and adjuvant are mixed to form an emulsion. Four to

five days after injection of the mixture into the animal, antibodies can be detected in the blood, increasing to a maximal level over 14 days. After that, the antibody level starts to decline. If, however, the animal is given a second injection, the antibody level falls briefly, as circulating antibody binds injected antigen. It then rises rapidly and increases to a much higher level. Usually, on the fifteenth or sixteenth day after the first injection, additional booster dosages are injected, and on the twentieth day the animal is bled. After a month's resting period, the animal is bled again.

The antibody in animal serum is evaluated by determining the binding property of the antigen, the efficiency with which the antibody combines with an antigen, the optimal concentration for an antigen–antibody reaction, and the specificity of the antibody, including cross reactivity.

The antigen–antibody reaction can be affected by several factors such as contamination, pH, and osmolarity, but the most frequent problem is immunological competition for substances other than the particular antigen.

The isolation of immunoglobulins is performed on diethyl–aminoethyl cellulose columns, Sephadex G-200 dextran gel columns, and zinc sulfate precipitation. A practical method of purification of the antibodies is to combine them with the pure antigen to form an antigen–antibody complex. After washing the complex, the antigen is separated, leaving an antibody of high specificity (*10*).

Each antiserum must be characterized according to the following parameters:

1. Dilution of the antibody that will bind approximately 50% of the radioactive antigen, in the absence of unlabeled antigen.
2. Evaluation of the antibody at this optimal dilution with standard amounts of antigen, determining a typical dose–response curve.
3. Definition (under these conditions) of the sensitivity of the antibody, which is the smallest amount of nonradioactive antigen detectable in the RIA system.
4. Analysis of body fluids to determine whether different aliquots of the sample tested reveal a parallel response in the assay system.
5. Cross reactivity of the antibody, to evaluate the capacity of the antibody to bind related substances of a similar, but not identical, structure.

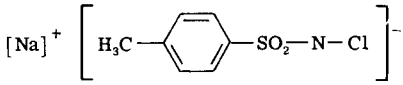
Freeze-dried antibodies retain their activity for long periods of time. Diluted antibody in buffer, stored at −20°C up to 5 months, and thawed only once, does not show any deterioration in titre activity. Repeated thawing and freezing results in denaturation of antibodies.

IV. LABELING OF ANTIGEN

The most widely used radioisotopes for the labeling of antigens are:

1. ^{125}I with a half-life of 57.4 days, with an electron capture* and a gamma mode of decay.
2. Tritium (^{3}H), with a half-life of 12.26 years and beta radiation.
3. Carbon 14(^{14}C), with a half-life of 5770.00 years and a beta mode of decay.

The labeled antigen should preserve the original immunologic activity and should emit the adequate amount of radiation. From the several methods employed for the labeling of an antigen with ^{125}I, lactoperoxidase and the chloramine T systems are the most popular (*11,12*). Chloramine T is the sodium salt of the *n*-monochloride derivative of *p*-toluene sulfonamide.



The mechanism of substitution of iodine, produced by the oxidation of iodide in the tyrosine of protein or polypeptide in the presence of chloramine T, is not fully understood.

Denaturation of the antigen molecule due to chemical, radiation, or reagent damage, or due to the substitution of iodine for hydrogen in the structure of a polypeptide, can influence the immunological completeness of an antigen.

The methods of labeling compounds with ^{3}H can be organized into two main groups: chemical syntheses, which include halogen–tritium replacement, hydrogenation with ^{3}H, introduction of tritiated methyl groups, and reduction with tritiated metal hydrides, and isotope exchange reactions with ^{3}H gas or tritiated solvent and hydrogen-transfer catalyst.

Steroids, fatty acids, and amino acids usually are tritiated by catalytic exchange; after tritiation, the catalyst is removed and nonreacted ^{3}H is treated with hydroxylic solvent. Tritiation of protein molecules causes many problems and is usually unsatisfactory. The complexity of tritiating processes and the high levels of ^{3}H activity in the labeling process require

* Electron capture is a process whereby the nucleus undergoes a transition when the ratio of neutrons to protons is too low, increasing the ratio by capture of an orbital electron. This electron capture produces a vacancy in an orbit and the emission of γ-ray photons and Auger electrons (X rays). An Auger electron is an electron which moves from one orbit to another (closer to the nucleus), sometimes transferring the energy to another electron that is ejected from the atom.

special facilities for safe operation, and are performed only in radiochemical centers.

V. PRINCIPLE OF GAMMA- AND BETA-SCANNING EQUIPMENT*

For the types of labeled material used in RIA, scintillation detection systems are of the greatest use. Scintillation is a process whereby the energy of radioactive emission is converted into a flash of light. Depending upon the emission involved, there are two general types of scintillation systems: liquid scintillation for beta particles and solid-crystal scintillation for γ rays.

Since beta particles and electrons have identical characteristics, beta counters can also be used for measuring conversion of Auger electrons. Likewise, gamma counters can be used for measuring X rays.

When a beta particle passes through an aromatic solvent, its kinetic energy is dissipated, and the beta particles produce excitation and ionization in the solvent. Electrons, in the conjugated carbon–carbon double bonds of these aromatic solvent molecules, are excited to a high energy state by absorbing some energy of the beta particles, and, upon returning to the lower energy ground state, they release their energy in the form of radiant energy. The conversion of kinetic energy to light is referred to as the scintillation process. Since the emitted UV radiation is not suitable for detection in the 200–260-nm wavelength range, fluorescent organic materials are added to the solvent, which are capable of absorbing the UV radiation. These fluors re-emit the energy at longer wavelengths. The fluorescent process is not 100% efficient, but does successfully transfer the energy from the short wavelength to a longer wavelength region, where it is detected with greater sensitivity. Dioxane, despite the fact that it has a lower efficiency than other solvents, is commonly used because of its water solubility. Solutes, such as naphthalene, must be incorporated to transform the beta particle productive of desired effects into light.

Primary and secondary fluors used in the scintillation process are PPO (2,5-diphenyloxazole) and dimethyl POPOP {1,4-bis[2-(4-methyl-5-phenyloxazolyl)]-benzene}. The purpose of the secondary fluor is to further shift the emitted radiation to a still longer wavelength. The new fluors, such as *tert*-butyl PBD [2-(4′-*tert*-butylphenyl)-5-(4″-biphenyl)-1,3,-

* For details the reader is referred to W. F. Ulrich, Ph.D., T. R. 598 Biomedical Report, Beckman Instruments, Inc., Fullerton, California, December, 1975.

4-oxadiazole] and PBBO [2-(4′-biphenyl)-6-phenylbenzoxazole] show still higher performance characteristics.

The choice of fluor used in the scintillation process is dependent upon the sample. In liquid scintillation instrumentation, a sample vial, containing radioactive material in solvent with fluors, is placed between two photomultiplier tubes. Here the photons of light emitted by the scintillation process produce photoelectrons. The process is repeated from anode to anode, increasing the number of electrons. For each initial photoelectron, approximately one-million electrons arrive at the final anode and produce a fairly large electrical pulse in the output circuit of the tube. The pulses are analyzed electronically according to the size in a pulse-height analyzer by discriminators, and stored in bins called scalers (Fig. 2). A computer divides the number of pulses stored in each scaler by the counting time and records the number of counts, or pulses, in the scalers.

One of the problems in liquid scintillation is quenching, which is the reduction of the photon output of the system. Color quenching is the absorption of photons of various wavelengths. Chemical quenching is the reduction of the photon output by flours or solvents. Dilution quenching occurs if the initial concentration of fluor is too low, or if there is too much sample in the system.

The other problem is beta absorption, which occurs when a beta particle must pass through many layers of molecules before it can escape and pass into the solvent. In so doing, the beta particle loses energy.

The following methods are used to calibrate the scintillation counter in order to determine the counting efficiency of each sample:

1. Internal standardization, by repeating the count after addition of a small volume of the same isotope.
2. Channels ratio technique, by the application of the ratio of the count rate in two channels, or portions of pulse height.
3. External standardization, by utilization of the Compton effect in which gamma rays striking a molecule release an orbital electron from the molecule, and the rays proceed at a longer wavelength. These electrons, produced in the scintillation vial, have all of the properties of beta particles and produce scintillation. If scintillation from a source other than the beta particles is produced in the vial, the scintillation is quenched in the same way as the scintillation which originates from the beta particles.

A combination of the channels ratio and the external standard technique eliminates the difficulties and gives the best results.

When gamma or X rays pass through various solid substances, they produce secondary electrons. These, in turn, can produce ionization and excitation, and, therefore, scintillation. Sodium iodide strengthened with

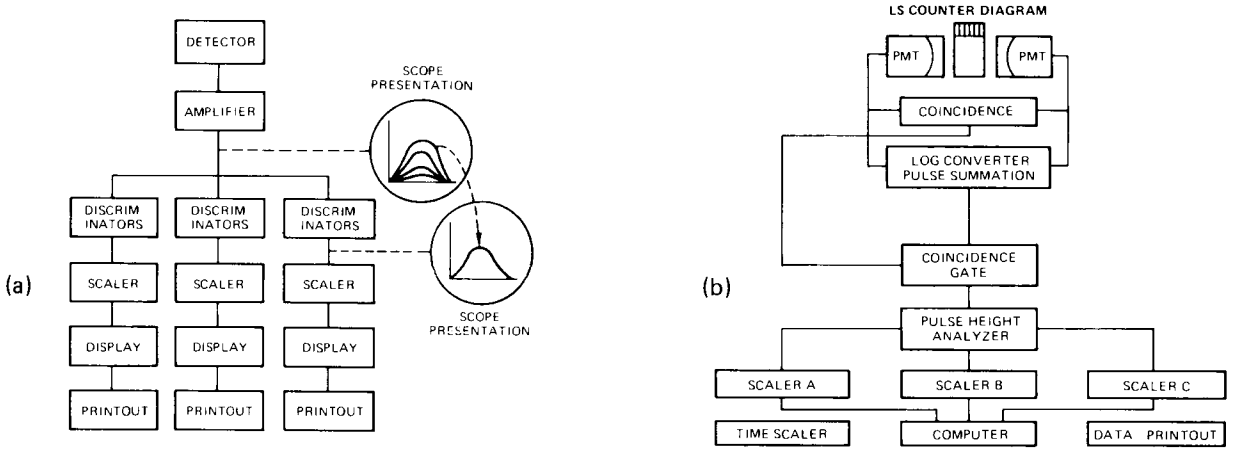


Fig. 2. Schematic diagrams of (a) gamma scintillation and (b) liquid scintillation systems. Reproduced with permission from Beckman Instruments, Inc., Fullerton, California.

Thallium iodide is one such scintillator commonly used in analytical systems.

Regardless of the energy of the rays, i.e., low energy over 400 keV (which produces the Compton effect) or over 1.02 MEV (which converts a gamma photon into an electron–positron pair), gamma rays produce energetic electrons that give rise to scintillation. The light flashes are generated by transitions involving a geometrical arrangement of electrons in the crystal. The number of excitations and resulting photons is dependent upon the energy of incident radiation.

Once radioactive emissions have been converted into light pulses, the instrumental technique used for measurement is essentially the same, whether beta or gamma emissions are involved. The optical signal has a magnitude proportional to the energy of the original radioactive emission.

The output of scalers can be indicated with the use of rate meters (digital displays). Since quantitation is based on the emission rate, it is desirable to have a rate output in counts per minute (cpm) or, using a computer-type system, in true disintegrations per minute (dpm).

VI. RATIONALE OF RADIOIMMUNOASSAY

The basic principle of RIA is the stoichiometric relationship between the binding antibody and the antigen. The amount of antibody-bound antigen is a function of the binding affinity of antibody to the concentration of the antigen. Labeled and unlabeled antigens are competing for available combining sites on antibodies. The unlabeled antigen inhibits the binding of labeled antigen. The amount of labeled antigen bound to antibody is inversely proportional to the amount of unlabeled antigen present in the reaction.

A fixed amount of antibody reacts with a constant amount of radioactive antigen to form an antigen–antibody complex. This reaction is inhibited by unlabeled antigen. Labeled and unlabeled antigens are competing for the reaction with the antibody until equilibrium is attained. It is a reversible process, with two distinct velocities:

$$\text{Antigen} + \text{antibody} \underset{V_2}{\overset{V_1}{\rightleftharpoons}} \text{antigen–antibody complex}$$

where the antigen is a proportional mixture of labeled and unlabeled parts. The interaction between the labeled and unlabeled antigen and antibody is shown schematically in Fig. 3.

The incubation period for the completion of a reaction is specific for each particular antigen. To prevent the incubation damage caused by en-

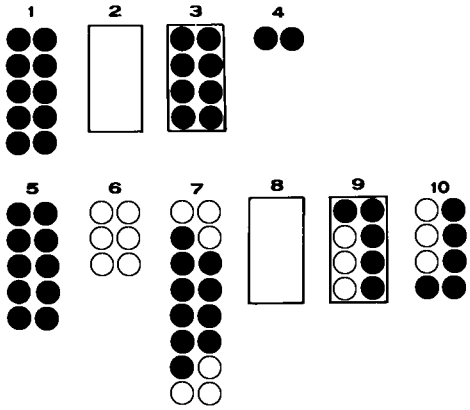


Fig. 3. Scheme of interactions between radioactively labeled or unlabeled antigen and antibody. Key: Radioactive antigen, 1; antibody, 2; antigen–antibody complex, 3; unbound antigen, 4; radioactive antigen, 5; cold antigen, 6; mixture, 7; antibody, 8; antigen–antibody complex, 9; unbound antigen, 10.

zymes, free radicals, or oxidants, it is recommended that the incubation be performed at a low temperature or that proteinase inhibitors be added to the reaction fluids. The unbound antigen, labeled and unlabeled, is separated from the antigen–antibody complex, and the radioactivity is measured. The concentration of antigen in the unknown sample is obtained by comparing the amount of radioactivity in the standard samples, containing known amounts of antigen, to the radioactivity of the unknown sample. Radioimmunoassay allows the quantitation of antigenic substances in body fluids, even in quantities as low as pico (10^{-12}) or femto (10^{-15}) units.

All RIA methods have in common the basic steps of reaction, separation, raw data readout, and data reduction to terminology usable by the physician. In order to perform this sequence with a high degree of sensitivity, specificity, accuracy, and reproducibility, the system used should have the following characteristics:

1. An antibody that is highly specific and demonstrates a strong binding affinity for the antigen.
2. A labeled antigen that has an effective specific activity and strong binding affinity, and which is relatively free of damaged or nonimmunological material.
3. Reagent concentrations that are carefully developed to allow optimum antibody binding of the labeled antigen, in order to provide adequate standard curve slope and amplitude.
4. A method of separating bound from free antigen that is effective and thorough (*13*).

VII. METHODS OF SEPARATING BOUND FROM FREE ANTIGEN

In RIA, the equilibrium between bound and free antigen is time- and temperature-dependent, and can be changed in either direction. The timing and temperature control in separating the antigen-antibody complex from the free antigen are critical.

Many techniques are used for separation, such as electrophoresis, salt precipitation, and gel filtration. However, the most reproducible and economically adaptable are the solid-phase method, adsorption of the free antigen onto charcoal coated with dextran, or double-antibody application.

In the solid-phase antibody procedure, the antibody is firmly fixed to an insoluble material such as Sephadex or Sepharose, or directly to polypropylene or polystyrene plastic tubes. Antibodies are typically covalently linked to the solid phase by a diazotization method (*14*) or by the cyanogen–bromide system (*15*). The antibodies linked to the solid phase are separated from the free antigen by simple centrifugation, leaving the unbound antigen in the supernatant fluid. When antibodies are allowed to stand in polystyrene tubes and then rinsed, a certain amount of antibodies are adsorbed onto the tubes. After addition of antigen and proper incubation, free antigen is decanted, and the radioactivity of the antigen–antibody complex fixed to the tubes is counted (*16*). Measurement of radioactivity in the solid phase enables the determination of the percentage of bound radioactive antigen to antibody against the amount of unlabeled antigen in the standard solutions.

The separation in an optimum charcoal system can almost be complete. The details of preparation of a satisfactory system may be found in "The Separation of Bound Antibody from Free Antigen," by W. M. Hunter and P. C. Ganguli (*17*).

The charcoal technique of separation is based on the principle that if charcoal is coated with a synthetic glucose polymer dextran, it can adsorb only low-molecular-weight antigen. Larger protein molecules, such as antibodies, cannot penetrate the spaces in the charcoal particles, and the antigen–antibody complexes cannot be adsorbed. Centrifugation separates the adsorbed free antigen in the charcoal precipitate, whereas the bound fraction remains in solution. Centrifugation should preferably be performed at a low temperature. Many RIA methods require refrigerated centrifuges.

The double-antibody method is based on the fact that an antibody is simultaneously an antigen. An antibody molecule, developed in an animal of one species, is antigenic to an antibody developed in another species.

The radioactive and unlabeled antigens compete for binding sites on the

antibody. After the reaction achieves equilibrium, a second antibody is added to the antigen–antibody complex; after a period of incubation, the tubes are centrifuged, the supernatant is discarded, and the radioactivity is measured in the precipitate that is the antigen-antibody–antibody complex. After the first antigen–antibody reaction, it is advisable to precipitate calcium, which activates complement present in the solution. Activated complement may cause some interference.

VIII. CALCULATION AND EVALUATION

In evaluating the results, all the rules pertaining to any radioactivity should be strictly followed. Scintillation counting involves the measurement of random occurrences, and it is important to run all tubes in duplicate.

With a restricted amount of antibody and radioactive antigen added to each tube, the radioactivity is inversely related to the ratio of bound and free, unlabeled antigen, according to the law of mass action.

By measuring the radioactivity in the bound phase, it is possible to quantitate the amounts of antigen in the test. The amount of binding is calculated relative to the radioactivity in the blank. The data are plotted on graph paper in relation to a standard, resulting in a curve. The same calculation may be performed by the computer. To produce a sensitive standard curve, the reaction between antibody, radioactive antigen, and standard must be optimized by selecting a proper separation technique having optimal concentrations, temperatures, and timing.

Many procedures have been used to evaluate graphically and mathematically the RIA dose-response data. The standard curve may be constructed graphically, and the transformation of parameters must be established. The choice of parameters depends upon procedures used for analysis.

To avoid individual subjectivity in constructing the curves, it is advisable to estimate the unknowns by using the statistical method of least squares. The line of best fit drawn through the plotted points is defined as the least regression line, or the line for which the sum of the squares of deviations of the observed points from the line is the smallest in comparison to the sum of the squares of deviations from any other line drawn through the scatter of points.

Most RIA data produce a marked curvature. To achieve linearity, the data have to be transformed. However, this has to be done with caution because a new error may be introduced, since data are often forced to a transformation to which they do not fully conform (*18*).

A listing of computer programs, operating instructions, and sample input–sample output data for the calculation of the four parameters of a logistic dose–response curve, as well as the calculation of quality control and statistics for assays using an analysis of variance, is presented by Rodbard *et al* (*19*). The analysis of the logit–log method and the Scatchard plot is documented by Faden and Rodbard (*20*). Automated calculation is discussed by Skelley *et al.* (*21*). The typical dose–response curves

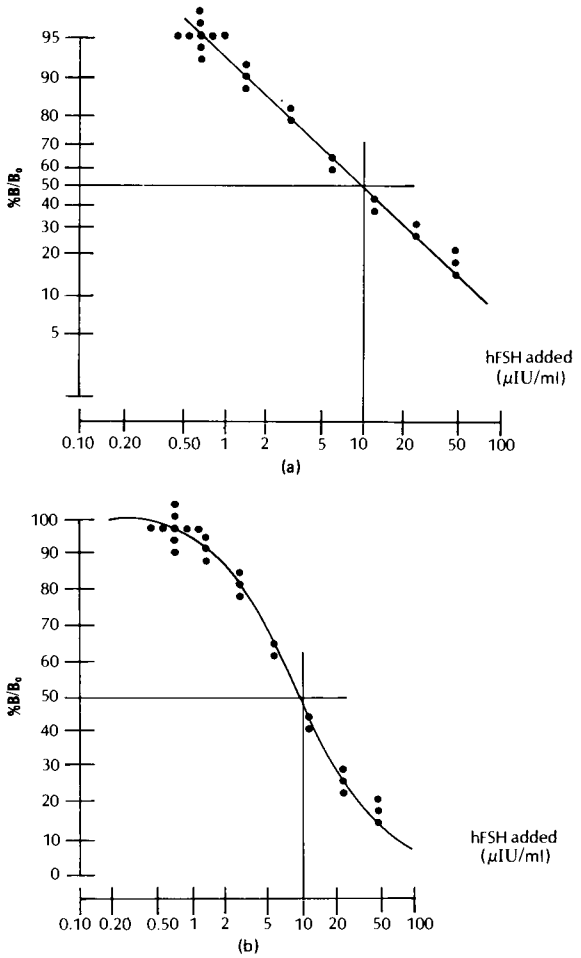


Fig. 4. Representative standard curves of radioimmunoassay for hFSH. Prepared by Serono laboratories; calculated on the Hewlett-Packard computer using Rodbard's program after modification. From Rodbard (*84*) and reproduced with permission from Serono Laboratories, Inc., Braintree, Massachusetts.

are (a) B_0/B versus linear dose; (b) percent-bound versus linear dose; (c) percent-bound versus log dose; and (d) percent-bound logit versus log dose.

The value of B_0/B versus linear dose is determined by dividing the cpm of zero standard (B_0) by the cpm of the standards and samples. B_0 (zero dose) denotes maximum binding tube. A tube contains antibody and labeled antigen but no unlabeled antigen. This specifies how much antigen can bind to a given quantity of antibody under the assay condition. B(bound) denotes that portion of antigens in a tube that is bound to antibodies. B_0/B values are plotted on the y axis against the concentration of the antigen on the x axis of linear graph paper, and the antigen values are determined from the plot.

The percent-bound versus linear dose is determined from calculating the percentage of bound radioactive antigen in each standard, by dividing the cpm of the standards by the cpm of the zero standard, and multiplying the numerical ratio by 100. The percent-bound values are plotted on the y axis and the concentration of antigen on the x axis of linear graph paper.

The percent-bound versus log dose is plotted on semilog paper with the percent-bound values of the standards on the vertical y axis against the concentration of antigen on the horizontal log x axis.

The log–logit is determined by calculating percent-bound values of the standards, and plotting them on the logit y axis against the concentration of standards on the log x axis of log–logit paper. The points are connected by the best-fitting line.

IX. RIA OF POLYPEPTIDE-RELEASING FACTORS

Releasing hormones from the hypothalamus stimulate the liberation of pituitary hormones, whereas release-inhibiting hormones suppress the secretion of the tropic hormones from the pituitary gland. The IUPAC-IUB Commission on Biochemical Nomenclature Recommendations (*22*) has accepted the names *liberin* for releasing hormone and *statin* for release-inhibiting hormone. This nomenclature has been accepted in Europe. The majority of American researchers retained the terms *releasing* and *release-inhibiting* factors or hormones. The RIA method of thyrotropin-releasing factor was developed by Bassiri and Utiger (*23*). The method, however, applies only to tissue assay for in the serum hormone is rapidly inactivated. Improvement of the method by the use of benzamidine to prevent the destruction of thyrotropin-releasing hormone (*24*) has not solved the problem completely.

RIA of lutenizing hormone-releasing factor in serum has been per-

formed by the sensitive methods of Nett *et al.* (*25*) and Arimura (*26*), with the average normal level ranging from 250 fg/ml to 10 pg/ml of serum. Somatotropin release-inhibiting factor was performed by the method of Brownstein *et al.* in the rat brain (*27*).

Many research programs are involved in the systematic study of the synthesis of polypeptide factors from the hypothalamus, and this field is rapidly developing. However, despite the obvious progress in RIA methods of these peptide hormones, releasing and release-inhibiting factors are still restricted to investigational use.

X. POLYPEPTIDE HORMONES FROM THE NEUROHYPOPHYSIS

The hormones from the posterior pituitary are actually manufactured in the hypothalamus and stored in the pituitary gland. The RIA of vasopressin, an antidiuretic hormone, is successfully performed with the double-antibody separation method. Its average normal level in the blood by RIA is 1.0–3.5 pg/ml (*28*).

The RIA of oxytocin, a hormone responsible for uterine contractions, release of milk, and reabsorption of water from the kidneys into the blood stream, is performed by the double-antibody separation of the bound portion. The average normal serum level of oxytocin is up to 1.5 pg/ml (*29*). The cord blood and amniotic fluid have a very high content of oxytocin; however, the normal levels in the body fluids of pregnant women have not yet been established.

XI. POLYPEPTIDE HORMONES FROM THE ADENOHYPOPHYSIS

Growth hormone, somatotrophic hormone, and somatotropin are synonyms. This hormone is a polypeptide secreted by the acidophil cells of the anterior pituitary gland. RIA of growth hormone can be performed on serum or heparinized plasma. The blood must be obtained from a fasting subject at rest, after an overnight sleep. Clinical observation indicates that psychic or physical stimulation releases the growth hormone-releasing factor from the hypothalamus and raises the growth hormone level in the blood. A number of procedures for RIA of human growth hormone (HGH) have been evolved, differing mainly in the system used to separate bound from free hormone. In modern assay the majority of laboratories are using the method of Hunter (*30,31*). Special care is required to opti-

mize the assay conditions. The normal serum level of HGH in man is in the range of 1–50 ng/ml, but the clinical range is up to about 500 ng/ml, and serial dilutions are required to fit the standard curve. Rapid changes occur in serum concentration after exercise, or after the intake of proteins, steroid hormones (*32*), glucose, olive oil, or terpene oils (*33*). The interpretation of results has to take into account the existence of at least four forms of immunoreactive HGH with a MW of 20,000, 40,000, 80,000, and 150,000. The nature of larger molecules is unknown.

The ACTH radioimmunoassay is performed directly on plasma. The polyethylene glycol separation technique used by Orth (*34*) permits the measurement of low levels of ACTH with high accuracy. Because plasma contains proteolytic enzymes which decompose ACTH, blood has to be collected into chilled, heparinized tubes and separated in refrigerated centrifuges at 4°C. Plasma must be transferred to plastic tubes and frozen immediately. Added trypsin inhibitors in the plasma only partially repress the action of the proteolytic enzymes. The average normal level of plasma ACTH drawn between 8–10 A.M. is 20–80 pg/ml. Antibodies to follicle stimulating hormone (FSH) and leuteinizing hormone (LH), and to their alpha and beta subunits, have been developed, and each can be measured by RIA in the presence of the other. Human chorionic gonadotropins react also with LH, so that during pregnancy this fact must be taken into consideration. There is no general agreement about the structure of FSH and LH, our standards being based on the Second International Reference Preparation produced by the National Institutes of Health or by the Pituitary Agency of the United States. The Midgley method (*35*), modified by Serono Laboratories (*36*), is based on a double-antibody system and may by performed on serum, EDTA plasma, or urine. Serum or plasma are stable up to 24 hr at 2°–8°C. Stability of 1–3 months can be achieved when stored at −20° to −40°C. A 24-hour urine sample should be collected and immediately kept under refrigeration without any antifermentation agent. The test must be performed within 48 hr. Urine must be filtered on Whatman's Paper No. 4 and dialyzed against distilled water at 4°C for 2–4 hr. The ratio of the volume before and after dialysis is used for the calculation of the results. The assay is performed on 0.1 ml of sample or standard. After addition of 0.1 ml of antibodies to human FSH, the tubes are vortexed and incubated at room temperature for 6–20 hr. After completion of the first incubation, 0.1 ml of labeled antigen is added, the tubes are vortexed, and again incubated for 20–24 hr. After completion of the second incubation, 0.1 ml of the second antibody is added, and the tubes are vortexed and incubated at room temperature for 18–24 hr. At the end of the third incubation, 2 ml of cold, deionized water is added, the tubes are centrifuged, decanted, and the radioactivity is measured.

As in the RIA of FSH, Serono Laboratories modified the method of Midgley (*37*) for human luteinizing hormone (hLH), using an almost identical procedure (*38*).

Human prolactin is a single-chain polypeptide with a 21,000 MW. The antiserum developed in rabbits is specific, and the cross-reactivity with growth hormone or placental lactogen is minimal. The RIA method for prolactin with double antibodies was developed by Hwang *et al.* (*39*). The method is sensitive up to 2 ng/ml; the average normal level in the male is 5 ng/ml and in the female, 10 ng/ml. The Second International Standard prepared by the National Institutes of Health of the United States has an activity of 1 IU/45 μg. However, prolactin prepared by different sources has a different reactivity, and, because the unified standards are not available, the material from the National Institute of Arthritis, Metabolism and Digestive Diseases (Bethesda, Maryland) serves as a criterion for comparison.

Thyroid stimulating hormone (TSH) is a glycoprotein with a 28,000 MW, which is composed of alpha and beta subunits. The RIA procedures for hTSH follow the basic principle of competition by labeled and unlabeled antigen for a fixed amount of antibody. The separation of bound and free antigen can be accomplished by several methods; Odell *et al.* (*40*) apply double-antibody. The Pharmacia Company (Piscataway, N.J.) modified the method of Wide and Porath (*41*) by utilizing a solid-phase technique with TSH antibodies bound to special Sephadex particles. Corning Medical Diagnostics (Medfield, Massachusetts) modified the methods of Sluiter *et al.* (*42*) and Miles and Hales (*43*) and employed more stable radioactive antibodies instead of radioactive antigen. TSH is quantitated in terms of ^{125}I-antibodies bound to the hormone. The bound radioactive antibodies are separated from the free antibodies with a second TSH antibody, which is covalently coupled to 1-μm glass particles. A test of this kind is known as a "sandwich assay," or a two-site immunoradiometric assay (*44*). The reaction time is 2.5 hr, with a rapid and quantitative separation after centrifugation. The observed average range in euthyroid individuals by this method is 1.8–10.1 μIU/ml. The sensitivity of the method is about 0.7 μIU. Euthyroid and hyperthyroid levels do overlap. The test does not require special preparation of the patient. After collection the blood is centrifuged and separated, the serum being stable for one week at 4°C.

XII. GASTROINTESTINAL POLYPEPTIDE HORMONES

For all gastrointestinal-tract hormone studies, the blood has to be collected into a prechilled tubes. It should then be immediately separated in a

refrigerated centrifuge at 4°C. The serum or plasma must be transfered to plastic tubes and frozen at −20°C until assayed.

Cholecystokinin or pancreozymin is a single substance now referred to only as *cholecystokinin*. It consists of 33 amino acids, with a variant form having 39 amino acids. As with other gastrointestinal hormones, the preparation of the patient is essential. All medication must be discontinued for at least 48 hr, and the blood must be drawn in a fasting state. Postprandial specimens are also required for special diagnostic purposes. The tests performed basically by the method of Englert (*45*) reveal the average fasting level of cholecystokinin in a normal individual to be from 10 to 110 pg/ml.

The main factors contributing to the problems connected with cholecystokinin RIA are difficulty of iodination, the low immunogenic potency of cholecystokinin, the substitution of porcine for human hormone in the assay, and a lack of a standard reference preparation (*46*).

Gastrin exists in two major forms: Big Gastrin G34, composed of 34 amino acids, and Little Gastrin G17, composed of 17 amino acids. The RIA method of Yalow and Berson (*47*) was modified by Becton Dickinson Immunodiagnostics in such a way that unbound labeled gastrin is specifically adsorbed onto an ion–exchange resin, and separated from antibody-bound labeled gastrin by centrifugation. A standard curve is plotted, which relates inhibition of binding to a known added amount of unlabeled G17. Concentrations of gastrin in unknown samples are then determined from this curve. The average normal range of the gastrin level in serum is less than 150 pg/ml.

Certain substances, such as heparin and high concentrations of salt, may interfere with the separation system. Enzymes present in the serum gradually degrade gastrin. Therefore, serum should be separated within 1 hr of the collection, and assayed or stored frozen.

Secretin is a 27-amino acid peptide with a 3055 MW that lacks tyrosine. Antibodies are developed in rabbits or guinea pigs by synthetic secretin coupled to bovine serum albumin. Despite the lack of tyrosine, the iodination of secretin is possible (*48*). Straus *et al.* (*49*) successfully iodinated the histidine residue of secretin. The RIA of secretin is performed by the double-antibody system (*50*). The time required to accomplish the test is about 1 week. The average normal fasting level of secretin is 15–150 pg/ml.

Glucagon is a peptide consisting of 29 amino acids, with a 3485 MW. However, there are four molecular forms of glucagon with molecular weights of 40,000, 9000, 3484, or 2900. Specific antibodies that can differentiate between different types of glucagon have been developed in animals (*51*). RIA methods apply the extraction of glucagon. The bound and

free glucagon are separated by a double antibody. Valverde *et al.* (*52*) used gel filtration prior to RIA and did find "big plasma glucagon," with a range of 45–293 pg/ml, "large glucagon," with a range of 0–50 pg/ml, "true glucagon," with a range of 0–98 pg/ml, and "small glucagon," with a range of 0.56 pg/ml. The time required to perform RIA is about 3 days.

A molecule of insulin consists of 51 amino acids and is composed of A- and B-peptide chains. Yalow and Berson (*53*) first described the RIA method, presenting the RIA of insulin. Today, we have the following methods to separate bound from free insulin: dextran-coated charcoal, membrane filters, Sephadex coupled with insulin antibodies, polyethylene glycol precipitation, antibodies coated to the walls of plastic tubes, as well as antibodies bound to dextran beads. Corning Medical Diagnostics modified the method of Ceska *et al.* (*54*) and chemically immobilized the antibody on a porous glass particle with a large surface area and high relative density. The time required to complete the test is 3 hr. The average fasting insulin level in serum is 10–24 μIU/ml (1 IU = 41.67 μg).

Gastric inhibitory polypeptide (GIP) is a 43-amino acid peptide with a 5100 MW. The antibodies to GIP were successfully developed. The RIA method of Kuzio *et al.* (*55*) found the average fasting serum level to be 75–500 pg/ml.

A series of other gastrointestinal hormones has been investigated and their RIA procedures elaborated. After evaluation, the new RIA methods will be introduced to clinical laboratories.

XIII. CALCITROPHIC POLYPEPTIDES

Parathyroid hormone (PTH) is a polypeptide of 84 amino acids with a 9500 MW. Several methods for RIA were developed (*56–58*), but many problems associated with the assay are still not resolved. The main difficulty is the unavailability of pure human PTH as a standard and as an immunogen in animals. Replacement of bovine PTH or the synthetic, active portion of human hormone with the amino-terminal fragment does not yet solve the problem. Still, many issues relating to the physiology and metabolism of PTH must be elucidated. Current assays report the average normal level of PTH as being up to 500 pg/ml.

Calcitonin is a polypeptide of ~3400 MW. The antibodies to calcitonin are developed in animals by the injection of synthetic human calcitonin. The same synthetic substance serves as a standard as well as the labeled antigen. RIA performed according to Deftos *et al.* (*59*) resulted in the average normal serum level of calcitonin up to 400 pg/ml.

XIV. PLACENTAL POLYPEPTIDES

Human chorionic gonadotropin is a glycoprotein with a 56,000 MW of its main peptide portion. The carbohydrate part is 40% and, additionally, 9% forms sialic acid. A molecule of chorionic gonadotropin consists of alpha and beta subunits. Antibodies produced against HCG cross-react with LH. The double-antibody RIA method was developed by Morgan and Lazarow (*60*). However, Vaitukaitis *et al.* (*61*) raised antibodies separately for alpha and beta subunits. A beta subunit of HCG reacts insignificantly with a beta subunit of LH.

The average normal pregnancy values (in mIU/ml of serum or plasma) are:

First week	10–30
Second week	30–100
Third week	100–1000
Fourth week	1000–10,000
Second to third month	10,000–100,000
Second trimester	10,000–30,000
Third trimester	5,000–15,000

According to the Second International Standard, 1 IU of chorionic gonadotropin is contained in 1.279 μg. Human placental lactogen (HPL) and human chorionic somatomammotrophin hormone (HCS) are synonymous. HPL is a single-chain polypeptide of 190 amino acids with a MW of ~20,000. As in other polypeptide RIAs, the methods differ mainly in the system of separating the bound from the free antigen. The method of Letchworth *et al.* (*62*) applies ethanol precipitation, which makes the assay simple and accurate. The average normal level of HPL during pregnancy at 24 weeks of gestation should be about 3 μg/ml of serum, which should increase steadily to about 7 μg/ml of serum at 42 weeks of gestation.

The discussion of the thyroid hormones has been provided in this volume in the chapter titled "Thyroid Function" by Leonard K. Dunikoski. The clarity and completeness of the subject material presented by Dr. Dunikoski obviates the need for further discussion within this chapter.

XV. STEROID HORMONES

Steroids as haptens must be coupled to a carrier, usually bovine albumin, to make them antigenic. The specificity of antibodies varies accord-

ing to the position of conjugation to the carrier. The antisera to steroids coupled via carbons 1, 6, and 7 have good sensitivity and specificity. Antisera to steroids coupled via carbons 3, 11, 17, or 20 have low specificity (*63*). If antibodies are not sufficiently specific, it is necessary to extract and eliminate the interfering substances from the samples of the body fluids. Each extraction requires estimates of recovery, which are performed by tritiated steroids. In RIA the most widely used label for steroids is ^{3}H. Despite a series of methods using iodinated steroids for RIA, there is a curious lack of published information concerning well documented, clinically useful assays (*64*).

An almost complete spectrum of clinically applicable antibodies to human steroids has been developed. Still, some problems pertaining to sensitivity and specificity have to be resolved. A typical procedure of estimation of steroids is the estriol method of Kunzig and Geiger (*64*) adapted by Wien Laboratories (Succasunna, New Jersey):

1. Glass-stoppered centrifuge tubes (50 ml) for each unknown and control serum are labeled.
2. A sample (200 μl) of unknown serum or heparinized plasma and control serum is transferred to its respective tube.
3. [^{3}H]Estriol (5 μl) is added to each tube. At the same time, 5 μl of radioactive estriol is placed in a liquid scintillation vial for use in the recovery determination.
4. Distilled water (5 ml) is added to each tube.
5. For the blank, 6.0 ml of distilled water is transferred to a 50-ml glass-stoppered centrifuge tube.
6. To each tube, 20 ml of purified methylene chloride is added, and the tubes are shaken vigorously for 60 sec.
7. Then the tubes are centrifuged for 7 min at 2000 g, and the upper layer is aspirated and discarded.
8. Three 5-ml aliquots of each centrifuge tube are removed and transferred to three labeled 16 $\times$ 125 mm glass test tubes. This is equivalent to one-fourth of the original amount of serum or plasma.
9. The tubes are evaporated to dryness in a warm water bath (40°–50°C), with the aid of a stream of air or nitrogen.
10. The 16 $\times$ 125 mm glass test tubes are labeled in duplicate as 0, 50, 100, 250, 500, and 1000.
11. Estriol standard of 0, 5, 10, 25, or 100 μl is placed into each tube (in duplicate) respectively, and the standard is evaporaated to dryness along with the extracted samples.
12. To each tube, 0.8 ml of phosphate buffer (pH 7.4) is added, and the tubes are mixed.

13. Using a pipetting device, 50 μl of [^{3}H]estriol is added to each tube, and the tubes are mixed.

14. While mixing, 50 μl of estriol antibody solution is added to each tube.

15. All tubes are incubated in an ice bath for 60 min. While the tubes are in the ice bath, 0.5 ml of a cold, dextran-coated charcoal suspension is added to each tube. The tubes are mixed, and allowed to stand for 5 min.

16. The tubes are then centrifuged at high speed for 7 min.

17. The clear supernatant solution from each tube is decanted into liquid scintillation vials, and 10 ml of scintillation fluid is added to each vial. The tubes are closed, mixed well by shaking vigorously, and allowed to stand for 30 min before counting in the dark, at room temperature.

18. Each vial is counted for at least 2 min in a beta scintillation counter, with the window set for counting ^{3}H. Counts per minute (cpm) for each tube are recorded.

The recovery factor is calculated by the following formula:

$$\frac{\text{cpm of unknown or control} \times 4}{\text{cpm of recovery vial}}$$

The ratio for each blank, standard control, and unknown counted is calculated as

$$\text{Ratio} = \frac{\text{cpm for zero standard tube}}{\text{cpm for blank, or control, or unknown}}$$

A standard curve is constructed, and the concentration of estriol is determined where

$$\text{Estriol (ng/ml)} = \frac{\text{ng/ml}}{\text{recovery factor}}$$

RIA of urinary steroids requires, before processing, an acid hydrolysis of the urine; it also often requires chromatographic separation of samples from interfering steroids.

Urinary steroids are analyzed in 24-hr samples; the total volume is recorded. A 1-ml aliquot of the sample urine is placed in a 16 × 125 mm tube; 0.1 ml of 3.2 *N* HCl is added to convert the urine to pH 1. The sample is allowed to stand in the dark, at room temperature, for 24 hr. It is neutralized by the addition of 0.9 ml. of 0.35 *N* NaOH. After addition of 8 ml of borate-saline buffer, the sample is mixed, and is ready for RIA. If further purification is necessary, it is accomplished on silica, alumina, or Sephadex columns (*65*).

Today, it is generally accepted that for estimation of plasma or serum steroid hormones, the procedures of choice are radioimmunoassays.

XVI. CANCER MARKERS

Most human malignancies have tumor-associated antigens that are released into the circulation and which can be measured by RIA. In addition, various malignant lesions may produce enzymes, which may be used as tumor markers to identify the presence of the lesion. The carcinoembryonic antigen (CEA) is a glycoprotein with a 150,000–200,000 MW. Radioimmunoassay of CEA is performed by the method of Hansen *et al.* (*66*) adapted by Roche Diagnostics (Nutley, New Jersey):

To a mixture of 0.5 ml plasma sample or control and 2 ml of 0.85% NaCl, 2.5 ml of 1.2 *M* perchloric acid is added. After centrifuging, the supernatant is dialyzed against deionized water and, finally, against 0.01 *M* ammonium acetate. The standards are prepared in 5 ml EDTA buffer, and 25 μl of CEA antiserum is added to each tube. The tubes are incubated in 45°C-water bath for 30 min. Then 25 μl of ^{125}I-labeled CEA is added, and the tubes reincubated for 30 min. The tubes are then removed from the bath, and 2.5 ml of zirconyl phosphate gel is added. After centrifugation and washing of the precipitate with 0.1 *M* ammonium acetate, the contents of the tubes are counted in the gamma scintillator.

The method is sensitive to the ionic strength and the pH of the buffers. Precise timing and temperature are important factors. Dialysis may be replaced by desalting in preequilibrated Sephadex or high-resolution, spherical polyacrylamide gel columns. In the future this step will probably be eliminated by the direct method with radioactive monoclonal antibodies.

Diagnostically, much better results are achieved when CEA is performed simultaneously in plasma and in body fluids from a location closer to the cancer (*67*). The average normal level of CEA is 0–2.5 ng/ml.

α-1-Fetoprotein is a glycoprotein with a ~70,000 MW. A double-antibody RIA (*68,69*) has the sensitivity of 0.5 ng/ml. The average normal concentrations have been reported to be between 2–20 ng/ml.

In normal pregnancy, the following values have been reported:

First trimester	20–120 ng/ml
Second trimester	160–550 ng/ml
Third trimester	100–400 ng/ml

In hepatoma, the level of α-1-fetoprotein may rise to milligram quantities per milliliter.

Elevation of serum acid phosphatase levels has long been regarded as evidence for the dissemination of prostatic cancer. However, the results have not been conclusive, and a more specific RIA for prostatic acid phosphatase has been developed. The double-antibody methods of Choe *et al.* (*70*) and Cooper and Foti (*71*), performed on normal male individuals, averaged 0.0–2.0 ng/ml. In prostatic carcinoma, the levels were above 2.0 ng/ml. Preferably, serum is used for the assay. Certain sera give false high readings. In such cases, the sera should be reassayed in diluted samples.

XVII. VARIOUS PROTEINS

Ferritin is a protein with an approximate 450,000 MW, which may contain 24% iron. The RIA method of Miles *et al.* (*72*) is based on a two-stage reaction. In stage one, the human serum ferritin is bound to a solid-phase antihuman ferritin. In stage two, radiolabeled antihuman ferritin is bound to the insoluble antihuman ferritin complex from stage one. The solid phase is washed and counted. The average level in normal sera is 130–140 ng/ml. It is important to note that RIA results are affected by isoferritin and a subunit of liver ferritin.

IgE is a glycoprotein with a 200,000 MW. Several RIA methods have been developed, but the most utilized is the solid-phase method of Wide and Porath (*41*). The normal level is 6–780 ng/ml.

Myoglobin is similar to hemoglobin, but it contains only one iron and one heme group bound to a single polypeptide chain with an ~17,200 MW. The double-antibody method is sensitive to 0.5 ng of myoglobin. The average normal values are 6–85 ng/ml of serum. Rosano and Kenny discuss the method's development (*73*).

XVIII. MISCELLANEOUS

Renin, a proteolytic enzyme with a molecular weight of ~40,000, is released from the juxta-glomerular cells of the kidney. The enzyme which is present in plasma acts in the general circulation to cleave its substrate, an α_2-globulin synthesized by the liver, to produce the decapeptide, angiotensin I. Angiotensin I is rapidly cleaved by the activity of a converting enzyme to the biologically active octapeptide, angiotensin II.

Angiotensin II is rapidly inactivated by angiotensinases. In the RIA of renin by the solid-phase system, the plasma is incubated to generate An-

giotensin I, thereby allowing for its quantitation. Plasma renin activity is expressed as ng/ml/hr (*74*).

Prostaglandin studies are performed by the RIA methods of Jaffe *et al.* (*75*), Caldwell *et al.* (*76*), and Hickler (*77*). To eliminate the interference of proteins, a protein denaturation and extraction step is performed on the sample prior to assaying. The prostaglandins are labeled by ^{3}H, and separation of bound from free prostaglandins is achieved by precipitating the bound portion with a second antibody-binding reaction. Attention must be paid to the cross-reactivity of the particular prostaglandin.

The extraction of cAMP and cGMP must occur prior to serum RIA. Urine may be used directly. All reagents must be refrigerated, and cAMP and cGMP must be frozen at −20°C. The cAMP method of Gilman (*78*) and the cGMP method of Murad (*79*) apply the double-antibody system with tritiated antigen.

Radioimmunoassays of therapeutic and toxic drugs such as digoxin, digitoxin, gentamicin, and barbiturates belong to routine procedures in clinical laboratories, and the list of new assays is growing steadily.

XIX. NEW CONCEPTS IN RIA

At the present level of experience with RIA, the solutions to several problems pertaining to specimen collection and handling, patient preparation, and dietary and drug interference have been elucidated. Application of statistical methods and new basic concepts to clinical problems for evaluation of test results increased the precision, sensitivity, and accuracy of RIA. The real major advance in the RIA technique has been the development of automated, fully computerized equipment.

The automated RIA system (ARIA II, Becton Dickinson immunodiagnostics) applies reusable antibody chambers. Antibodies are covalently bound to glass microbeads packed in a tiny chamber, which can be used again for up to 3000 consecutive assays. The system has both beta- and gamma-counting capabilities, and automatically performs up to 120 tests. A changeover from one type of test to another takes 5 min, the time from sample aspiration to the result being about 3 min. The ARIA II system uses automated computation and control of clinical results.

A new technology in the physical separation of bound antibody from free antigen has been developed, using *Staphylococcus aureus* as a bacterial immunoadsorbent. Protein A, present on the surface of *Staphylococcus aureus,* binds IgG. Boehringer Mannheim Biochemicals are commercially preparing staphylococcal immunoadsorbent Pansorbin, which can be adapted to the RIA of viral and subviral antigen–antibody systems.

Soergel *et al.* (*80*) has described the method of assay of antibodies to caliciviruses by radioimmune precipitation, using staphylococcal protein A as an IgG adsorbent.

Although methods for radioactive iodination and tritiation are available, the accessibility and relatively low cost of commercial preparations make them most practical.

XX. MONOCLONAL ANTIBODIES

Genetic engineering on recombinant DNA and production of monoclonal, monospecific antibodies by hybridomas are about to make fundamental changes in radioimmunoassay technology.

A single clone of antibody-producing cells yields only one type of homogenous antibody molecule with high specificity and reproducibility. Multiple myeloma, a single antibody-forming cell, implies the best possible exemplification for production of such monoclonal antibody.

Through hybridoma technology, the monospecific antibodies permit the development of new diagnostic procedures and improvement of the reliability of radioimmunoassay.

Specific antibodies to a known antigen produced by hybridoma clone are used to detect early changes on the surfaces of cells. It may be that this is the direction in which the future application of immunology and radioimmunoassay to cancer diagnosis lies (*81–83*).

XXI. CONCLUSION

The performance of RIA is not a simple procedure. It requires, besides the knowledge of immunology, radiology, and clinical chemistry, an understanding of the physiological and pathological processes of each particular clinical case. Endocrinological interrelationships, immunological cross reactions, the relationship of radioactive decay to the stability of labeled antigen, dietary interference, and the effects of metabolic by-products are only a few problems pertaining to the simplest RIA procedure. Automation may eliminate the human error, but it does not solve the main problems which make the test valid.

The complexities of the problems involved mandate that only a well-trained chemist with a broad clinical experience can properly perform RIA, one who must also share in the responsibility of the interpretation of the results with the medical staff.

TABLE I

RIA Problem-Solving Guide[a]

Apparent problems	Possible causes	Certified causes	Suggested solutions
No (maximum) binding	1. No antibody added 2. Wrong tracer or wrong antibody added 3. Separation procedure failed	1. All tubes had same counts as NSB 2. Inventory of some reagents do not reflect assays run 3. No pellets in tubes or all counts in charcoal	1. Devise pipetting procedure to insure addition of all reagents 2. Code reagents of each assay to avoid mismatching 3. Check for pellets before decanting
Low (maximum) binding	1. Old tracer used 2. Incorrect dilution (too much) tracer used 3. Incorrect dilution (too little) of antibody used	1. Check expiration date of tracer 2. Calculate cpm that should be in the assay 3. Check dilution of antibody	1. Don't use tracer close to expiration date or tracer that has been too long diluted or stored improperly 2. Count 100 μl of tracer to verify cpm before adding to assay 3. When diluting make sure final volume gives right number of tubes
Flat curve with normal NSB and B_0	1. No standards added 2. Standards not diluted 3. Wrong standards added	1. All standard tubes had same counts as B_0 2. Binding in standard curve very low ($B/B_0 < 20\%$) 3. All standard tubes had about the same counts as B_0	1. Devise pipetting procedure to insure addition of all reagents 2. Do standard serial dilution just before the RIA 3. Code reagents of each assay to avoid mismatching
High NSB	1. Antibody added to NSB tubes 2. Old tracer used 3. Damaged or improperly stored tracer	1. cpm in NSB = B_0 and no tubes have greater cpm, whereas most have fewer cpm 2. Check expiration date of tracer 3. Binding should be low throughout; assay out of control	1. Devise method to isolate NSB tubes from antibody addition 2. Don't use tracer close to expiration date or tracer that has been too long diluted or stored improperly 3. Don't use tracer close to expiration date or tracer that has been too long diluted or stored improperly
Low sensitivity	1. Standards improperly diluted (overdiluted) 2. Too much tracer used per tube 3. Too much antibody used 4. Old tracer used	1. Check procedure for serial dilution of standard 2. Calculate cpm that should be in the assay (low B_0) 3. Check dilution of the antibody 4. Check expiration date of tracer	1. Do standard serial dilution just before the RIA 2. Count 100 μl of tracer to verify cpm before adding to assay 3. When diluting make sure final volume gives right number of tubes 4. Don't use tracer close to expiration date
Too much sensitivity	1. Standards improperly diluted (underdiluted) 2. Too little tracer used per tube 3. Too little antibody added 4. Old tracer used	1. Check procedure for serial dilution of standard 2. Calculate cpm that should be in the assay (high B_0) 3. Check dilution of the antibody (low B_0) 4. Check expiration date of tracer	1. Do standard serial dilution just before the RIA 2. Count 100 μl of tracer to verify cpm before adding to assay 3. When diluting make sure final volume gives right number of tubes 4. Don't use tracer close to expiration date

High control values	1. Standards improperly diluted (overdiluted) 2. High NSB due to old tracer or improper storage 3. Old or improperly stored controls used	1. Curve will be flatter than normal in the beginning; all values will be high 2. Abnormally high NSB (2 or 3 × normal) 3. Check expiration of control	1. Do standard serial dilution just before the RIA 2. Do not use tracer close to expiration date 3. Store reagents according to protocol and don't use after expiration date
Low control values	1. Standards improperly diluted (underdiluted) 2. High NSB due to old tracer or improper storage 3. Old or improperly stored control used	1. Curve will drop more steeply than normal in the beginning; all values will be low 2. Abnormally high NSB (2 or 3 × normal) 3. Check expiration of control	1. Do standard serial dilution just before the RIA 2. Do not use tracer close to the expiration date 3. Store reagents according to protocol and don't use after expiration date
No pellets	1. Normal serum not added 2. Second (precipitating) antibody not added 3. Normal serum and second antibody mismatched 4. Normal serum and/or second antibody overdiluted	1. Inventory of NS does not reflect assay run 2. Inventory or Ab_2 does not reflect assay run 3. Inventory of Ab_2, NS do not correspond 4. Check dilution of the NS and Ab	1. Devise pipetting procedure to insure addition of all reagents 2. Devise pipetting procedure to insure addition of all reagents 3. Code reagents of each assay to avoid mismatching 4. When diluting make sure final volume gives right number of tubes
Poor duplication	1. Poor technique 2. Poor pipettors 3. Old or damaged reagents used	1. No pattern in duplication, problem in all assays that particular technician performs 2. Errors are occasional; problems occur in all assays 3. Check expiration date and storage of all reagents	1. Technician training and qualifications programs should be developed 2. Pipette ^{125}I solution into 10 tubes and make sure CV is acceptable 3. Store reagents according to protocol and don't use after expiration date
Poor reproducibility	1. Poor dispensors 2. Old or damaged reagents used 3. Differences in technicians' techniques	1. Volumes of various aliquots are unequal 2. Check expiration date and storage of all reagents 3. Technicians get different results on all assays they do	1. Pipette ^{125}I solution into 10 tubes; make sure CV is acceptable 2. Store reagents according to protocol and don't use after expiration date 3. Establish QC information on each technician
Low activity	1. Tracer diluted improperly 2. Insufficient activity in tracer stock solution 3. Old tracer used	1. Check tracer dilution; calculate cpm that should be in the assay (B_0 high) 2. Count sample of stock solution 3. Check expiration date of tracer (B_0 low)	1. Count 100 μl of tracer to verify cpm before adding to the assay 2. Check tracer vial to insure removal of all activity for the assays 3. Don't use tracer close to expiration date
No activity	1. No activity in tracer stock solution 2. Tracer not added	1. Count sample of stock solution 2. Check to see if any volume is present in TC tubes	1. Count 100 μl of tracer to verify cpm before adding to assay 2. Devise pipetting procedure to insure addition of all reagents

[a] Abbreviations: Ab, antibody; Ab_2, second antibody; **B**, bound portion of antigens complexed to antibodies; B_0, maximum binding tube; CV, coefficient of variation; NS, normal serum; NSB, nonspecific binding; QC, quality control; TC, total count. Reproduced with permission from the Cambridge Nuclear Radiopharmaceutical Corporation, Billerica, Massachusetts.

ACKNOWLEDGMENTS

We are grateful to Beckman Instruments, Inc., Becton Dickinson Immunodiagnostics, Cambridge Nuclear Radiopharmaceutical Corporation, Radioassay Systems Laboratories, Inc., RIA Products, Inc., Serono Laboratories, Inc., the Upjohn Company, and Wien Laboratories, Inc. for permission to use the product literature, graphics, and schematics.

REFERENCES

1. Watson, J. D., "Molecular Biology of the Gene." Benjamin, New York, 1965.
2. Ackroyd, J. F., "Immunological Methods." Davis, Philadelphia, Pennsylvania, 1964.
3. Landsteiner, K., "The Specificity of Serological Reactions." Harvard Univ. Press, Cambridge, Massachusetts, 1945.
4. Smith G. P., Hood, L., and Fitch, W., *Annu. Rev. Biochem.* **40,** 469 (1971).
5. Poljak, R. J., *Nature (London)* **256,** 373 (1975).
6. Terry, W. D., and Fahey, J. L., *Science* **146,** 400 (1964).
7. Rowe, D. S., Hug, K., Forni, L., and Pernis, B., *J. Exp. Med.* **138,** 965 (1963).
8. Hurn, B. A. L., and Landon, J., *In* "Radioimmunoassay Methods" (K. E. Kirkham and W. M. Hunter, eds.), p. 123. Churchill-Livingstone, Edinburgh and London, 1971.
9. Freund, J., *Am. J. Clin. Pathol.* **21,** 645 (1951).
10. Walker, S., and Lukas, Z., *J. Immunol.* **113,** 813 (1974).
11. Thorell, J. I., and Larson, I., *J. Immunochem.* **11,** 203 (1974).
12. Hunter, W. M. *in* "Radioimmunoassay Methods" (K. E. Kirkham and W. M. Hunter, eds.), p. 3. Churchill-Livingstone, Edinburgh and London, 1971.
13. "T_4 RIA Test System 1004." Company Report, Corning Medical Diagnostics Medfield, Massachusetts, 1977.
14. Gurvich, A. E., Kuzovleva, O. B., and Tumanova, A. E., *Biokhimiya* **26,** 803 (1961).
15. Wide, L., *Acta Endocrinol. (Copenhagen)* **63,** Suppl. 142, 219 (1969).
16. Catt, K. J., and Tregear, G. W., *Science* **158,** 1570 (1967).
17. Hunter, W. M., and Ganguli, P. C., *in* "Radioimmunoassay Methods" (K. E. Kirkham and W. M. Hunter, eds.), p. 243. Churchill-Livingstone, Edinburgh and London, 1971.
18. Hunter, W. M., *in* "Handbook of Experimental Immunology" (D. M. Weir, ed.), pp. 14.35–14.36. Blackwell, Oxford, 1973.
19. Rodbard, D., Faden, V. B., Hutt, D. M., and Knisley, S., *U.S. NTIS, PB Rep.* **PB-246224** (1975).
20. Faden, V. B., and Rodbard, D., *U.S. NTIS, PB Rep.* **PB-246223** (1975).
21. Skelley, D. S., Brown, L. P., and Besch, P. K., *Clin. Chem. (Winston-Salem, N.C.)* **19,** 2, 146 (1973).
22. IUPAC–IUB Commission of Biochemical Nomenclature, *J. Biol. Chem.* **250,** 3215 (1975).
23. Bassiri, R. M., and Utiger, R. D., *Endocrinology* **90,** 722 (1972).
24. Jeffcoate, S. L., and White, N., *J. Clin. Endocrinol. Metab.* **38,** 155 (1974).
25. Nett, T. M., Akbar, A. M., and Niswender, G. D., *J. Clin. Endocrinol. Metab.* **36,** 880 (1973).
26. Arimura, A., Kastin, A. J., and Schally, A. V., *J. Clin. Endocrinol. Metab.* **38,** 510 (1974).
27. Brownstein, M., Arimura, A., and Sata, H., *Endocrinology* **96,** 1456 (1975).
28. Beardell, C. G., *J. Clin. Endocrinol. Metab.* **33,** 254 (1971).

29. Chard, T., Boyd, N. R. H., Forsling, M. L., McNeilly, A. S., and Landon, J., *J. Endocrinol.* **48,** 223 (1970).
30. Hunter, W. M., *in* "Handbook of Experimental Immunology" (D. M. Weir, ed.), p. 17. Blackwell, Oxford, 1973.
31. Hunter W. M., *in* "Methoden der Hormonbestimmung" (H. Breuer, D. Hamel, and H. L. Krüskemper, eds.), p. 1. Thieme, Stuttgart, 1975.
32. Hunter, W. H., *in* "Growth Hormone in Hormone Assays and Their Clinical Application" (J. A. Loraine and E. T. Bell, eds.), p. 221. Churchill-Livingstone, Edinburgh and London, 1976.
33. Hordynsky, W., *Arch. Arzneither.* **2,** 151 (1977).
34. Orth, D. N., *in* "Methods of Hormone Radioimmunoassay" (B. M. Jaffe and H. R. Behrman, eds.), p. 125. Academic Press, New York, 1974.
35. Midgley, A. R., Jr., *J. Clin. Endocrinol. Metab.* **27,** 295 (1967).
36. "FSH Protocol," Company Report, Serono Laboratories, Inc., Braintree, Massachusetts, 1977.
37. Midgley, A. R., Jr., *Endocrinology* **79,** 10 (1966).
38. "hLH Protocol," Company Report, Serono Laboratories, Inc., Braintree, Massachusetts. 1978.
39. Hwang, P. Guyda, H., and Friesen, H. G., *Proc. Natl. Acad. of Sci. U.S.A.* **68,** 1902 (1971).
40. Odell, W. D., Wilber, J. F., and Paul, W. E., *J. Clin. Endocrinol. Metab.* **25,** 1179 (1965).
41. Wide, L., and Porath, J., *Biochim. Biophys. Acta* **130,** 257 (1966).
42. Sluiter, W. J., Van Kersen, F., VanZanten, A. K., Beekhuis, H., and Doorenbos, H., *Clin. Chim. Acta* **42,** 255 (1972).
43. Miles, L. E. M., and Hales, C. N., *Lancet* **2,** 492 (1968).
44. "Corning TSH ^{125}I-RIA Test System." Company Report Corning Medical Diagnostics, Medfield, Mass., 1978.
45. Englert, E., Jr., *Clin. Res.* **21,** 207 (1973).
46. Go. V. L. W., and Reilly, W. M., *in* "Gastrointestinal Hormones, A Symposium" (J. C. Thompson, ed.), p. 295. Univ. of Texas Press, Galveston, 1975.
47. Yalow, R. S., and Berson, S. A., *Gastroenterology* **58,** 1 (1970).
48. Holohan, K. N., Murphy, R. F., Flanagan, R. W. J., Buchanan, K. D., and Elmore, D. T. *Biochim. Biophys. Acta* **322,** 178 (1973).
49. Straus, E., Urbach, H. J., and Yalow, R. S., *Biochem. Biophys. Res. Commun.* **64,** 1036 (1975).
50. Bloom, S. R., *in* "Gastrointestinal Hormones, A Symposium" (J. C. Thompson, ed.), p. 257. Univ. of Texas Press, Galveston, 1975.
51. Valverde, I., Rigopoulou, D., and Marco, J., *J. Diabetes* **19,** 614 (1970).
52. Valverde, I., Lemon, H. M., Kessinger, A., and Under, R. H., *J. Clin. Endocrinol. Metab.* **42,** 804 (1976).
53. Yalow, R. S., and Berson, S. A., *J. Clin. Invest.* **39,** 1157 (1960).
54. Ceska, M., Grossmuller, F., and Lundkvist, U., *Acta Endocrinol. (Copenhagen)* **64,** 111 (1970).
55. Kuzio, M., Dryburgh, J. R., Malloy, K. M., and Brown, J. C., *Gastroenterology* **66,** 357 (1974).
56. Arnaud, C. D., Tsao, H. S., and Littledike, E. T., *J. Clin. Invest.* **50,** 21 (1971).
57. Buckle, R. M., *Br. Med. J.* **2,** 789 (1969).
58. Segré, G. V., Habener, J. F., Powell, D., Tregear, G. W., and Potts, J. T., Jr., *J. Clin. Invest.* **51,** 3163 (1972).

59. Deftos, L. J., Bury, A. E., Habener, J. F., Singer, F. R., and Potts, J. T., *Metab., Clin. Exp.* **20,** 1129 (1971).
60. Morgan, C. R., and Lazarow, A., *Proc. Soc. Exp. Biol. Med.* **110,** 29 (1962).
61. Vaitukaitis, J. L., Braunstein, G. D., and Ross, G. T., *Am. J. Obstet. Gynecol.* **113,** 751 (1972).
62. Letchworth, A. T., Boardman, R. J., Bristow, C., Landon, J., and Chard, T., *J. Obstet. Gynaecol. Br. Commonw.* **78,** 542 (1971).
63. Pratt, J. J., *Clin. Chem. (Winston-Salem, N.C.)* **24,** 1869 (1978).
64. Kunzig, J., and Geiger, W., *Arch. Gynecol.* **216,** 387 (1974).
65. Abraham, G. E., *Acta Endocrinol. (Copenhagen)* **75,** Suppl. 183, 7 (1974).
66. Hansen, H. J., Lance, K. P., and Krupey, J., *Clin. Res.* **19,** 143 (1971).
67. Pschibul, F., and Hordynsky, W., *J. Med. Soc. N.J.* **74,** 535 (1977).
68. Kirkpatrick, A., Wepsic, T. H., and Nakamura, R. M., *Clin. Chem. (Winston-Salem, N.C.)* **23,** 50 (1977).
69. Brock, D. J. H., and Sutcliffe, R. G., *Lancet* **2,** 197 (1972).
70. Choe, B. K., Pontes, E. J., McDonald, I., and Rose, N. R., *Cancer Treat. Rep.* **61,** 201 (1977).
71. Cooper, J. F., and Foti, A., *Invest. Urol.* **12,** 98 (1974).
72. Miles, L. E. M., Lipschitz, D. A., Bieber, C. P., and Cook, J. D., *Anal. Biochem.* **61,** 209 (1974).
73. Rosano, T. G., and Kenny, M. A., *Clin. Chem. (Winston-Salem, N.C.)* **23,** 69 (1977).
74. "Clinical Assays, Gamma Coat, Plasma Renin Activity, Company Report, Clinical Assays Inc., Division of Travenol Laboratories Inc., Cambridge, Mass., 1976.
75. Jaffe, B. M., Smith, J. W., Newton, W. T., and Parker, C. W., *Science* **171,** 494 (1971).
76. Caldwell, B. V., Burnstein, S., Brock, W. A., and Speroff, L., *J. Clin. Endocrinol. Metab.* **33,** 171 (1971).
77. Hickler, R. B., "Prostaglandin Symposium of the Worcester Foundation for Experimental Biology," p. 279. Wiley (Interscience), New York, 1978.
78. Gilman, A. G., *Proc. Natl. Acad. Sci. U.S.A.* **67,** 305 (1970).
79. Murad, F., *Proc. Natl. Acad. Sci. U.S.A.* **68,** 736 (1971).
80. Soergel, M. E., Schaffer, F. L., Sawyer, J. C., and Prato, C. M., *Arch. Virol.* **57,** 271 (1978).
81. Kohler, G., and Milstein, C., *Nature (London)* **256,** 495 (1975).
82. Olson, L., and Kaplan, H. S., *Proc. Natl. Acad. Sci. U.S.A.* **77,** 5429 (1980).
83. Kwan, S. P., Yelton, E. S., and Scharff, M. D., "Genetic Engineering," pp. 31–46. Plenum, New York, 1980.
84. Rodbard, D., *Clin. Chem. (Winston-Salem, N.C.)* **20,** 1255–1277 (1974).

3

Thyroid Function

LEONARD K. DUNIKOSKI, JR.

I. INTRODUCTION

Tests of thyroid function represent the single largest type of endocrine testing performed by most clinical laboratories. This reflects the relatively high incidence of thyroid disease as well as the wide availability of reliable laboratory tests which can aid in confirming diagnosis or monitoring therapy. Study of the thyroid gland can therefore serve the clinical chemist as both an introduction to endocrinology as well as a source of important continuing interaction between the clinical chemist and the endocrinologist, internist, or other physician.

The metabolic effects of thyroid hormones are wide-spread, affecting the metabolism of carbohydrates, lipids, proteins, and vitamins, and ex-

CLINICAL BIOCHEMISTRY
Contemporary Theories and Techniques Vol. 2

ISBN 0-12-657102-3

erting systemic effects on many organs. In turn, the thyroid gland is influenced by several other endocrine glands, and is regulated by hypothalamic, pituitary, and, possibly, other feedback mechanisms. It is not surprising, then, that the presenting signs and symptoms of thyroid disease are often nonspecific and confusing to the physician. Reliable and inexpensive laboratory procedures can be of great help in clarifying the presence and type of thyroid disease for these patients as well as for those with the classic symptoms of thyroid disorders.

This chapter will not convert a clinical laboratory scientist into a clinician. Rather, it will attempt to clarify some of the common problem areas in the laboratory diagnosis of thyroid disease. An informed laboratory scientist who has a good working relationship with physicians will undoubtedly be interested in more detailed clinical and laboratory knowledge. To this end, several suggestions for further reading are included at the end of the chapter.

II. THYROID PHYSIOLOGY

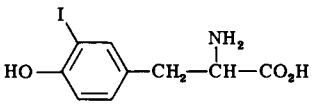
A. Synthesis

The thyroid is a bilobed or butterfly-shaped gland that weighs 20–30 g in an adult, making it the largest human endocrine gland. It is located just beneath the anterior strap muscles of the neck, and is extremely vascular in nature, receiving almost twice the blood supply (per gram of tissue) compared to the kidney. The gland consists of spheroids called follicles that vary in size but average 200–300 μm in diameter. The follicles are the functional units of the thyroid gland. Each follicle is lined by a single layer of epithelial cells, the base of which rests on a basement membrane oriented towards the capillary endothelium, the apex pointing to the colloid-filled follicular lumen. These follicular cells synthesize thyroglobulin, a protein of 670,000 MW and a sedimentation of 19 S. In addition, they are the site of synthesis of the thyroid hormones thyroxine (T_4) and triiodothyronine (T_3).

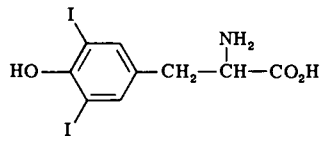
In addition to the follicular cells, special cells called C cells (parafollicular cells) constitute about 0.1% of the epithelial cells of the thyroid gland. These are located intrafollicularly (between the basement membrane and the lumen, but not reaching the lumen) as well as parafollicularly (Wolfe *et al.*, 1974). They are responsible for the synthesis and storage of the hormone calcitonin, a 32-amino acid peptide that plays a part in the regulation of serum calcium homeostasis. The C cells have been implicated in medullary carcinoma of the thyroid, a calcitonin-producing tumor.

Within the follicular cells, synthesis and release of T_3 and T_4 are normally considered in three distinct steps: iodide trapping, iodination and coupling of tyrosines, and proteolysis with release of iodothyronines. Iodide trapping (uptake) occurs by active transport of plasma iodide against chemical and electrical gradients across the basement membrane into the follicular cell. Plasma iodide levels reflect dietary and other intake of iodine-containing substances. In the United States, iodized salt and iodine in baked bread provide the bulk of ingested iodine; however, radiographic diagnostic agents, medications, and food supplements can increase iodine intake dramatically and consequently raise plasma iodide concentrations. Plasma inorganic iodide levels normally remain below 1.0 mEq/dl, with iodide clearance occurring in both the thyroid and the kidney. The total thyroid gland iodide content normally averages 6–7 mg, accounting for about 99% of total body iodine and demonstrating the magnitude of the active transport mechanism required for iodide trapping. Although the exact mechanism of active transport has not been clarified, uptake is stimulated by thyroid stimulating hormone (TSH) and is related to the amount of iodide present within the thyroid. Iodide transport will be increased when thyroidal iodide stores are low, but decreased when glandular iodide stores are increased. This can markedly affect some diagnostic testing results, as will be described later.

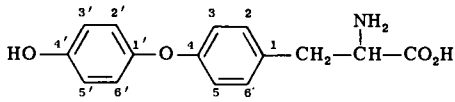
The second step in synthesis, iodination of tyrosines, begins with oxidation of trapped iodide to form iodine, a process catalyzed by a thyroid peroxidase. The overall mechanism of oxidation and iodination has not been clarified, but peroxidase activity is known to be present mainly in the apical portion of the thyroid follicular cells, a location which would minimize iodination of nonthyroglobulin intracellular protein (Taurog, 1978). Thyroglobulin synthesized by the follicular cell serves as the substrate for both iodination and storage. While iodination occurs within the follicular cells, storage occurs mainly in the colloid-filled follicular lumen, an extracellular location. Within the follicular cell, tyrosine is probably incorporated into thyroglobulin and then iodinated at the cell–colloid interface, forming either 3-monoidotyrosine (MIT) or 3,5-diiodotyrosine (DIT). The chemical structures for these and other related compounds are given in Fig. 1. Iodination *in vivo* does not necessarily occur in stepwise fashion, with MIT always being a precursor of DIT. Rather, the relative amounts of MIT and DIT formed seem to relate to iodide availability and hormonal synthesis, with MIT normally accounting for only 5% of total tyrosyl residues (Van Zyl and Edelhoch, 1967), but MIT being preferentially formed with iodide deficiency and reduced serum thyroxine (T_4) levels. The MIT and DIT formed are biologically inactive, remain bound to thyroglobulin stored in the follicular lumen, and represent about two-



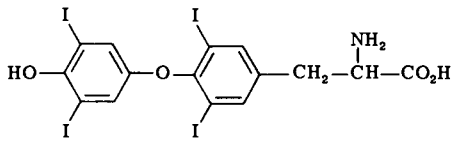
3-Monoiodotyrosine (MIT)



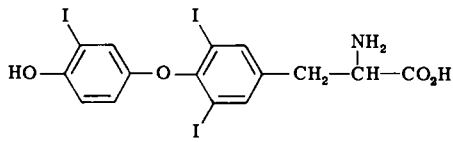
3,5-Diiodotyrosine (DIT)



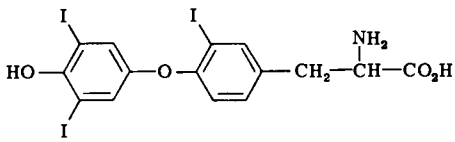
Thyronine nucleus



Thyroxine (L-3,5,3′,5′-tetraiodothyronine, T_4)



L-3,5,3′-Triiodothyronine (liothyronine, T_3)



L-3,3′,5′-Triiodothyronine (reverse T_3, rT_3)

Fig. 1. Thyroid hormones and related compounds.

thirds of the total iodine stored in the gland, equal to a 20-day supply for synthesis of iodothyronines (Robbins *et al.*, 1974).

Coupling of the peptide-linked mono- and diiodotyrosines produces the iodothyronines, with one MIT and one DIT yielding L-3,5,3′-triiodothyronine (T_3), and two molecules of DIT producing thyroxine (T_4). Although intermolecular *in vitro* coupling of DIT to T_4 has been documented, most workers feel that enzymatic intramolecular coupling represents the predominant *in vivo* mechanism of thyronine formation, and that thyroglobulin creates special steric effects necessary for maximum reaction efficiency. Enough T_3 and T_4 are stored in the thyroid bound to thyroglobulin to permit adequate availability for 1–2 months without additional synthesis (Robbins *et al.*, 1974). The ratio of T_3 to T_4 synthesized and stored depends on iodine availability and probably on TSH stimulation. Normally, 10–20 times more T_4 than T_3 is stored in thyroglobulin. In animal studies, restricted iodine intake produced an increased proportion of T_3 as compared to T_4 (Taurog, 1974), reflecting preferential MIT formation. In addition to T_3 and T_4, a small amount of L-3,3′,5′-triiodothyronine (reverse T_3, rT_3) is also formed by thyroidal iodotyronsine coupling. Reverse T_3 has very little or no calorigenic activity, but serum and/or amniotic fluid rT_3 levels may be clinically significant in various disease states (Chopra, 1978). Coupling of iodotyrosines is inhibited by many compounds, among them the antithyroid drugs methimazole and propylthiouracil. Both of these drugs also inhibit iodination of tyrosines as well.

Secretion of T_3 and T_4 occurs via TSH-stimulated proteolysis of thyroglobulin, releasing the free iodothyronines into the blood stream. The mechanism of proteolysis involves intracellular lysosomal enzymes which degrade the peptide-linked hormones into free amino acids. These enzymes normally are not secreted with the thyroid hormones. The iodotyrosines can also be released during the proteolysis of thyroglobulin, are normally deiodinated within the thyroid by a iodotyrosine deiodinase, and are at least partially reutilized for hormone synthesis. Inhibition of TSH-stimulated thyroid hormone release can be caused by a number of compounds, including high doses of iodide or lithium.

B. Hormonal Transport and Metabolism

In the circulation, thyroid hormones T_3 and T_4 are almost entirely bound by serum proteins. For both T_3 and T_4, thyroxine-binding globulin (TBG) is the principle binding protein; thyroxine-binding pre-albumin (TBPA) also binds T_4, but not T_3. Finally, albumin binds a small portion of both T_3 and T_4. The protein-bound ("bound") fraction represents 99.97% of total T_4 and 99.7% of total T_3. This large, physiologically inac-

tive pool of protein-bound T_3 and T_4 serves as a reservoir for extra-thyroidal hormone as well as a damper of the effects of transient increases in hormonal secretion. The remaining unbound ("free") fraction constitutes the physiologically active form of thyroid hormones. Although assay of the free hormone levels in the serum should logically correlate best with physiologic activity and/or disease states, these assays have been more difficult technically and no better clinically. Most clinical laboratories measure total serum levels, i.e., the sum of the bound plus the free fractions.

The normal serum concentration of total T_4 is 50–100 times that of total T_3 (5–12 μg/dl versus 70–200 ng/dl), making it the most abundant thryoid hormone. All of the T_4 in the body is synthesized in the thyroid. T_3, although present in lower concentration than T_4, is biologically the more active hormone. About 80% of serum T_3 results from peripheral partial deiodination of T_4, and only 20% from thyroidal secretion in the euthyroid patient. Substantial differences exist in the clearance and distribution of T_3 and T_4. Partially as a result of weaker binding to serum proteins, T_3 is cleared more rapidly than T_4 ($T_{1/2}$ for T_3 = 1 day; $T_{1/2}$ for T_4 = 7 days). The time for labeled doses of intravenously administered hormones to equilibrate with target organs is 3 days for T_4 and 22 hr for T_3 (Cavalieri and Ingbar, 1975), again reflecting the differences in protein binding. In addition, the distribution rates for T_3 and T_4 are not uniform for all organs. For T_4, two different kinetic "compartments" have been identified (Irvine, 1975). The first, consisting of the kidney, liver, and lungs, demonstrates rapid equilibration of T_4 to the tissues, and the second, consisting of the brain, fat, muscle and skin, demonstrates slow equilibration of T_4. The differences between these two compartments are presumably due to variations in the capillary barrier to passage of serum proteins, as well as to differences in intracellular binding sites for T_4. Most of the extrathyroid T_4 is located in the extracellular fluid and the rapidly equilibrating compartment. For T_3, the distribution is not as well characterized, but most T_3 appears to be located in the slowly equilibrating compartment, again demonstrating the differences in protein binding affinity between T_3 and T_4.

Changes in serum proteins, mainly TBG, will affect the level of total circulatory hormones. The binding of T_4 to TBG is a dynamic relationship:

$$\underset{\text{free}}{T_4 + TBG} \rightleftharpoons \underset{\text{bound}}{T_4 - TBG}$$

$$\text{Total} = \text{free} + \text{bound}$$

From this equation, a T_4 association constant can be calculated. In normal

TABLE I

Factors Associated with Changes in Serum TBG Levels

Increased TBG	Decreased TBG
Estrogens	Androgens
Pregnancy	Genetic
Newborn	Glucocorticoids
Estrogen-secreting tumors	Nephrotic syndrome
Oral contraceptives	Stress
Genetic	
Infectious hepatitis	

thyroid function, the law of mass action dictates that increases (or decreases) in TBG will increase (or decrease) the total T_4 concentration, but the free fraction will remain within normal limits. For example, an increase in TBG will bind more T_4, lowering the free T_4 concentration and increasing the bound T_4 concentration. However, at equilibrium (assuming normal thyroid function) the free T_4 concentration will return to normal and the total T_4 will be increased. Factors which alter TBG concentration are listed in Table I.

Thyroid hormone disposal consists of excretion and metabolism. About 20% of the daily T_3 and T_4 disposal occurs by biliary excretion of glucuronide and sulfate conjugates. The remaining portion is disposed of by metabolism via deiodination. For T_4, deiodination leads to formation of T_3(3,5,3′) and reverse T_3(3,3′,5′). For T_3 and rT_3, *in vivo* deiodination appears to occur sequentially, producing a variety of diiodothyronines, monoiodothyronines, and their glucuro- and sulfoconjugates (Sakurada *et al.*, 1978). Sensitive radioimmunoassays have demonstrated significant serum levels of MIT and DIT (Nelson *et al.*, 1974, 1975), and the clearance and production of several metabolites are now being extensively studied.

C. Regulation

Control of the amount of free thyroid hormones in the circulation and in the peripheral tissues is achieved by a complex series of mechanisms, three of which have already been mentioned. During normal synthesis, thyroglobulin serves as a storage reservoir for iodinated intermediates and hormones, with hydrolysis required to release T_3 and T_4 into the circulation. Once secreted into the circulation, the thyroid-binding proteins also limit the amount of physiologically active available to the tissues. Finally, by mainly secreting T_4, the physiologically less active hormone,

peripheral deiodination to T_3 provides a source of more active hormone. In addition to these three mechanisms, regulation of thyroid hormone occurs via hypothalamic, pituitary, and autoregulatory mechanisms.

Thyroid-stimulating or thyrotropic hormone (TSH) is a well defined glycoprotein which stimulates iodine uptake, organification, coupling, and secretion of thyroid hormones. TSH is considered the major regulator of thyroid function. TSH is secreted by the anterior pituitary gland, with secretion controlled by two regulatory mechanisms: feedback control by circulating free thyroid hormone levels and neural control by the hypothalamus. The feedback control mechanism, also known as the *thyroid–pituitary axis,* represents a classic example of a negative feedback system. The amount of free thyroid hormone in the circulation has an inverse effect on the secretion of TSH, so that low levels of free T_3 or T_4 cause increased TSH secretion, and high levels of free T_3 or T_4 suppress TSH secretion. The monitoring of unbound T_3 and T_4 levels occurs by the pituitary itself, separately from any control mechanisms influenced by the hypothalamus. Secretion of TSH is regulated over the complete range of thyroid hormone levels (Reichlin and Utiger, 1967) to maintain proper circulating levels of active T_3 and T_4. Regulation appears to involve binding of T_3 and T_4 to specific receptor sites in the pituitary; this binding may be more sensitive to T_3 than T_4 (Schadlow *et al.,* 1972).

Neural control of TSH secretion involves the tripeptide TRH (thyrotropin-releasing hormone) which is formed in the hypothalamus and in other parts of the brain. Hypothalamic TRH is one of the hypophysiotropic hormones (also called hypothalamic hormones or releasing factors) which will be described in greater detail in the next chapter (non-thyroid endocrinology). Secretion of TRH stimulates TSH as well as prolactin, and has several direct actions on the brain. The mechanism of regulation of TRH secretion has not been completely clarified, and may or may not involve circulating thyroid hormones. TRH secretion is definitely dependent on biogenic amines, and is modified by exposure to cold and stress (Reichlin, 1978). Hypothalamic lesions in animals cause a reduction in baseline TSH levels, but maintanence of normal pituitary response to low plasma hormone levels (Martin and Reichlin, 1970). However, even massive lesions in rat hypothalamus do not eliminate TRH secretion (Jackson and Reichlin, 1978), indicating that overall TSH secretion is a result of both feedback control and neural control. Figure 2 gives an overview of the pituitary–hypothalamic–thyroid relationship.

In addition to TSH and TRH, the thyroid responds to variations in the availability of iodide. Since the responses to iodide are dependent on normal thyroid function, they are termed *autoregulatory,* and are aimed at maintaining the glandular content of thyroid hormone. For example,

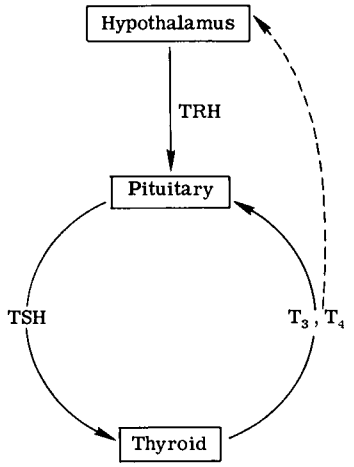


Fig. 2. Overview of thyroid regulation.

when plasma iodide concentrations are increased dramatically, the active transport of iodide is decreased. Animal studies show the thyroid gland has a finite capacity for organification and coupling, and that increasing levels of iodide cause a plateau in the rate of organic iodide synthesis (Nagataki and Ingbar, 1964). Still higher levels of iodide cause a decrease in organification, and a tendency to produce a larger quantity of MIT as compared to DIT, and a larger quantity of iodotyrosine as compared to iodothyronine. Increased iodide levels also inhibit release of T_4 (and T_3) from thyroglobulin. The overall effect of greatly increased iodide can, therefore, be a reduction in circulating T_3 and T_4 levels known as the Wolff–Chaikoff effect (after the investigators who first studied this effect, Wolff and Chaikoff, 1948).

III. LABORATORY APPROACH TO THYROID DISEASE

A. Overview

Diagnosis of thyroid disease is based on a clinical assessment of the patient's history and physical examination; laboratory tests are ordered to confirm or clarify the clinical diagnosis and to monitor therapy. Whereas thyroid disfunction is often complex or nonspecific, two types of patients are routinely seen by endocrinologists or internists: first, the patient with a thyroid goiter or nodule, and second, the patient with classic symptoms of hyper- or hypothyroidism. Diagnosis of the first type of patient relies

heavily on tests of radioactive iodine uptake normally not performed in the clinical laboratory but often performed in conjunction with tests of serum hormone levels. Workup of the second type of patient, i.e., the clinically hypo- or hyperthyroid patient, relies heavily on clinical laboratory tests, which will be stressed in the following section.

The clinical laboratory scientist, while often unfamiliar with specific types of thyroid disfunction, should be aware of the classic symptoms of the hyper- or hypothyroid patient, as well as of some commonly used nomenclature. Many hyperthyroid patients present nonspecific symptoms that usually result from accelerated functioning of one or more organ systems. The signs and symptoms best associated with hyperthyroidism (thyrotoxicosis) include fatigue, heat intolerance, palpitations, nervousness, excessive sweating, loss of weight despite good appetite, thyroid enlargement, and tachycardia or atrial fibrillation (Kaplan and Utiger, 1978). In the elderly patient these findings often are less obvious, and may be sometimes confused by other nonthyroidal illnesses. This less obvious hyperthyroidism of the elderly patient is termed apathetic or "masked" hyperthyroidism. Of the various disease states classed as hyperthyroidism, Graves' disease (Basedow's disease) is the most common, and involves autonomous secretion of thyroid hormones. While the cause of Graves' disease is still obscure, an autoimmune mechanism is suspected, a strong familial tendency is recognized, and a female to male ratio of 4:1 to 10:1 is reported. The disease is most common between 20 and 40 years of age, but patients may present symptoms at any age, including the newborn. In addition to the usual symptoms of hyperthyroidism, patients may have opthalmopathy, goiter, dermopathy, and loosening or separation of fingernails from their beds. This last phenomenon, called Plummer's nails, allows dirt to accumulate deeply under the nail, which is a distinct clinical diagnostic clue. Toxic nodular goiter (Plummer's disease) is a second type of hyperthyroidism that involves one or more autonomously functioning adenomas. These adenomas usually grow slowly and secrete increasing amounts of thyroid hormone, producing an insidious, slow onset of hyperthyroidism, usually appearing in older (over age 40) patients. There is no evidence of an autoimmune-type mechanism. A third type of hyperthyroidism, called factitious thyrotoxicosis, is a result of covert self-administration of excess thyroid hormones to which hospital and medical personnel seem to be particularly susceptible. In most forms of hyperthyroidism, the nonspecific accelerated functioning of organ systems also results in nonspecific changes in several clinical laboratory tests, including decreased cholesterol, decreased creatine kinase, and increased calcium levels in the serum. These changes are neither specific nor sensitive enough to be used in confirming diagnosis or monitoring

therapy. As a result of the continuing extra burden on the stimulated organ systems, several common complications (congestive heart failure, arrhythmias, heart block, and malabsorption) may further complicate these other laboratory tests.

In hypothyroidism, slower functioning of organ systems often produces signs and symptoms which are directly opposite to those observed in hyperthyroidism. The classic manifestations of hypothyroidism include fatigue, cold intolerance, weight gain, dryness of skin and hair, and hoarseness of the voice (Hall and Scanlon, 1979). Hypothyroidism may be divided into three classifications: Primary hypothyroidism (the most common type) refers to intrinsic glandular hyposecretion; secondary hypothyroidism refers to the inability of the pituitary to secrete adequate amounts of TSH; tertiary hypothyroidism refers to inability of the hypothalamus to secrete adequate amounts of TRH. The great majority of hypothyroid patients are women, most of whom are between 30 and 60 years old. Effects of hypothyroidism on biochemical tests are the opposite of those seen in hyperthyroidism, i.e., the serum cholesterol is increased. These changes likewise are nonspecific and nonsensitive. The laboratory can play a vital role in the assessment of the patient with thyroid disfunction. The following sections describe the common laboratory tests of thyroid function, dividing them into the general areas of tests of iodide clearance and tests of serum or plasma.

B. Tests of Iodide Clearance

1. Radioactive Iodine Uptake

The uptake of radioactive iodine (RAI) by the thyroid gland is really an estimate of plasma clearance of iodide by the thyroid. As such, it only provides indirect evidence of thyroid activity, and is no longer considered a primary test of thyroid function. However, its continued use alone or together with stimulation or suppression tests mandates that the laboratory scientists have a fundamental understanding of this type of testing. Since RAI testing is normally not performed in the clinical laboratory, a brief explanation of the test procedure is appropriate.

RAI Procedure. A known quantity of a tracer dose of radioactive iodine is administered to the patient either intravenously (if early uptake values are sought) or orally (for intermediate or late uptake values). Various isotopes may be used, depending on the patient's age and initial diagnosis. At fixed time intervals following the dose, radioactivity trapped in the thyroid is determined with suitable instrumentation and compared to a

dose standard; in some procedures, correction factors for extrathyroid radioactivity are used. In the euthyroid patient, early radioactive iodine uptake (20–30 min) reflects thyroidal trapping of iodide, whereas intermediate uptake (2–6 hr) reflects both trapping and organification, and late uptake (24 hr or longer) begins to reflect loss of iodide from the gland in addition to trapping and organification. In the euthyroid patient, maximum thyroid radioactivity is found after 24 hr, which is the most commonly employed counting time. Since iodide trapping is decreased in the hypothyroid patient, RAI uptakes are also decreased, with 24-hr uptakes less than that of about 10% commonly found. In the hyperthyroid patient, iodide trapping is accelerated and 24-hr RAI uptakes greater than 30% are commonly seen. Some hyperthyroid patients can trap, organify, and lose radioactive iodine so quickly that by 24 hr the thyroid radioactivity may be decreasing rapidly. For this reason, many radiologists include a 2-hr or 4-hr thyroid count in the evaluation of a possibly hyperthyroid individual. In hypothyroidism, the radioactivity sequestered in the gland is too low for acceptable precision, so the 2-hr or 4-hr result is not used. For most untreated patients, the 24-hr RAI uptakes correlate well with clinical and laboratory findings.

There are several reasons why the RAI uptake is not used as a primary test of thyroid function. First, "normal ranges" for the 24-hr RAI uptake procedure have been decreasing steadily in the United States, from 15–40% uptake in the early 1960s to about 10–30% uptake in the late 1970s. Presumably, this reflects the American's higher dietary iodide intake as a result of increased iodide in baked bread, as well as increased iodide ingestion in the form of drugs and dietary supplements. This increased intake expands the iodine pool in the body and dilutes the radioactive tracer in the plasma. A second problem with the RAI uptake arises from geographical variations in iodine intake due to diet, soil, or water. Since RAI uptake is inversely proportional to iodide intake, high or low plasma inorganic iodide concentrations of whatever cause will produce shifts in RAI uptake. Most importantly, since RAI values chiefly estimate plasma clearance of iodide rather than hormonal secretion, high or low RAI values may or may not indicate hyper- or hypothyroidism.

The problems are compounded with patients receiving therapy. Antithyroid drugs (propylthiouracil, methimazole, carbimazole) and other drugs which block organification and coupling of tyrosines cause a decrease in the 24-hr RAI uptake. Likewise, clinically euthyroid patients taking proper doses of replacement hormone will demonstrate decreased 24-hr RAI uptake due to TSH suppression. In such patients, a low RAI value should be expected, since a "normal" result would actually be interpreted as being abnormally high. Furthermore, after withdrawal of the

thyroid hormone or antithyroid drugs, a rebound effect may yield elevated RAI values. Although all of these possibilities can be properly anticipated and interpreted by the experienced clinician, they do demonstrate some of the pitfalls in using the RAI uptake technique. The laboratory scientist should be especially careful of situations in which RAI uptake and plasma hormone results apparently do not coincide. Often a case of "you're wrong" or "we're wrong" is the result of insufficient information or incomplete patient history.

The clinical usefulness of routine RAI is increased by the performance of two modifications—the T_3 suppression test and the TSH stimulation test. When the initial RAI value is elevated, T_3 can be administrated daily for 4–7 days, and the RAI uptake repeated. If the RAI uptake falls to less than half the baseline (i.e., initial) value, the initial elevation was probably due to the iodine rebound effect, dyshormonogenesis, etc. Hyperthyroid patients will not demonstrate T_3 suppression of the RAI uptake.

When the initial RAI uptake value is low, a TSH stimulation test may be performed to distinguish primary (thyroidal) hypothyroidism from secondary (pituitary) hypothyroidism. One or more doses of bovine TSH are given and the RAI uptake is repeated. In primary hypothyroidism, the damaged thyroid will not respond to TSH stimulation, whereas in secondary hypothyroidism (far less common than primary hypothyroidism) the repeat RAI uptake is greater than the initial procedure. Because of the common availability of reliable serum TSH determinations, the TSH-stimulation procedure has become seldom used, eliminating the need for a foreign protein injection (bovine TSH is used for stimulation due to inadequate availability of human TSH).

2. *Thyroid Imaging*

Whereas RAI uptake measures only the total uptake of radioactivity by the thyroid, thyroid imaging provides a visual display of the radioactive material incorporated in the gland. Several instruments may be used to produce this visual display. The rectilinear scanner first used for this procedure moves back and forth over the patient's neck, the image being produced by such an instrument called a *scan*. The newer gamma camera remains stationary and provides a high resolution image that is still sometimes (improperly) referred to as a scan. A third procedure, fluorescent imaging, does not even involve administration of radioisotope to the patient. Instead, a photon beam from a conventional X-ray tube or radioisotope source excites the iodide stores already present in the gland, generating characteristic transient X rays from the excited iodide. The image is recorded as with the rectilinear scanner. Thyroid imaging, by providing a visual display of the functioning thyroid, provides valuable information

on the size and shape of thyroid tissue as well as the uniformity of thyroid functional activity. It is widely used to evaluate patients with goiters or thyroid nodules since it can differentiate hyperfunctioning ("hot") and hypofunctioning ("cold") areas very reliably.

C. Analysis of Serum or Plasma

1. *Thyroxine (T_4)*

The analysis of total T_4 alone or together with other tests continues to be the most widely performed clinical laboratory test of thyroid function. Since total T_4 far exceeds that of total T_3 (70:1 by weight), it has become simpler to make measurements rapidly and reproducibly. Furthermore, since T_4 is more tightly bound to serum proteins than is T_3, most of the body pool of T_4 is circulating in the serum, whereas T_3 is more widely distributed in the tissues. Therefore, serum T_4 measurement may give a better indication of the total hormonal pool. In any case, procedures based on competetive protein binding (CPB), radioimmunoassay (RIA), or enzyme immunoassay permit routine total T_4 measurements in most clinical laboratories. Normal ranges for total T_4 average about 5–12 μg/dl (65 to 155 nmole/liter) except in neonates, in which they are higher.

The Committee on Nomenclature of the American Thyroid Association has recommended a uniform, unambiguous nomenclature for tests of serum thyroid hormones (Solomon *et al.*, 1976). Adoption of these recommendations by clinical laboratories and physicians would minimize the confusion in terminology which now exists. An abbreviated version of the nomenclature of the American Thyroid Association is given in Table II; further references to laboratory tests will use this nomenclature. For total thyroxine the abbreviation T_4(RIA) should be used if radioimmunoassay methodology is used, or T_4(D), if displacement assays are used (Murphy–Pattee, competitive protein binding, saturation analysis). Thyroxine determined by column chromatography should be designed as T_4(C), and reported as thyroxine rather than iodine. Free thyroxine, if actually measured and not calculated, should be reported as FT_4, and should not be confused with the free thyroxine index [FT_4 index (see Section III)]. Since total T_4 values can be used to confirm (or rule out) hyper- or hypothyroidism, several authors have advocated initial performance of total T_4 alone in an effort to contain the cost of laboratory testing. Should the total T_4 be normal, no further testing would be performed; an abnormal or borderline result would be followed by further tests to indicate the status of plasma-binding proteins. If employed conscientiously, this approach could indeed result in cost savings to the laboratory and to the patient,

TABLE II

Recommended Nomenclature for Tests of Thyroid Hormones in Serum

Abbreviation	Method
T_4 (D)	Total thyroxine assayed by displacement methodology (Murphy–Pattee, competitive protein binding, saturation analysis, etc.) (μg/dl).
T_4 (RIA)	Total thyroxine assayed by radioimmunoassay methodology (μg/dl).
T_4 (C)	Total thyroxine assayed by column chromatography (μg thyroxine (not iodine)/dl).
T_3U Ratio	Triiodothyronine uptake ratio (ratio, not percentage).
FT_4 Index	Free thyroxine index [calculated from (total T_4) × (T_3U ratio); units should be omitted to avoid confusion with T_4 or FT_4].
T_3 (RIA)	Total triiodothyronine assayed by radioimmunoassay methodology (ng/dl).
FT_3 Index	Free triiodothyronine index; calculated from [T_3 (RIA)] × (T_3U ratio); units should be omitted to avoid confusion with T_3 (RIA) or FT_3.
FT_4	Free thyroxine, assayed by equilibrium dialysis or special techniques (ng/dl).
FT_3	Free triiodothyronine assayed by equilibrium dialysis or special techniques (ng/dl).

especially in situations in which a relatively small percentage of patients would require further testing or if obvious clinical symptoms were to make the diagnosis straightforward. Such an approach requires a commitment by the laboratory director to insure maintenance of the highest standards for personnel, the quality of reagents, and type of equipment. More sophisticated clinicians may desire concommitant information about binding protein levels and may be reluctant to accept a total T_4 result alone, especially in the light of recent reports showing that T_3(RIA) is a more sensitive test for hyperthyroidism. Significant variations in binding proteins do occur in the general population, so that the use of T_4 alone certainly would not be appropriate for inclusion in a general screening profile. The interpretation of total T_4 is discussed under the free thyroxine index.

2. *Triiodothyronine Uptake Ratio (T_3U Ratio)*

The T_3 uptake (T_3U) test probably causes more confusion among relatively unsophisticated physicians than any other laboratory test of thyroid function. Although widely performed together with a total T_4, the nomenclature confusion between "T_3" and "T_3(RIA)," together with interlaboratory variation in normal ranges, can present real problems.

All T_3U tests are based on a similar procedure: Radioactive T_3 is added to an aliquot of serum and allowed to react with the unsaturated binding sites on the serum TBG. A nonspecific binder is then added to take-up the remaining unbound T_3, and the resulting radioactivity in the binder is then counted. The T_3U is an estimate of unsaturated TBG binding sites, and is not useful by itself. Its use lies in providing the ability to interpret a total T_4 (or total T_3) result. When serum TBG levels are increased, the T_3U will be decreased (due to less uptake by the resin or other nonspecific binder) and conversely, when serum TBG levels are decreased, the T_3U will be increased (due to greater uptake). The T_3U results are therefore *inversely* related to serum TBG levels or *directly* related to saturation of TBG binding sites. The T_3U is decreased in hypothyroidism since the TBG binding sites are relatively unsaturated, and conversely, it is increased in hyperthyroidism. However, any change which affects the levels of plasma TBG will affect the T_3U test. Table I lists some of the factors which can alter plasma TBG levels and, therefore, affect the T_3U.

In the past, the T_3U was reported either as a percentage or as "percent uptake," or sometimes as an inverted fraction. In an effort to standardize normal values for T_3U testing, laboratories should be encouraged to report T_3U results not as percentages, but as a T_3U ratio, using a two- or three-decimal place number. The ratio is attained by taking the mean of the percentage range for normal reference sera, and dividing it into the observed percentage result on any given patient specimen. For example,

Patient A

T_3U result: 32.5% uptake

Mean T_3U for all euthyroid patients: 30.0% uptake

$$T_3U \text{ ratio: } \frac{32.5\%}{30.0\%} = 1.083$$

In this way, a T_3U result from one laboratory may be compared to that from any other. The normal range, although centered on 1.00, will not necessarily be identical for each method. However, with T_4 procedures being performed by RIA procedures in most laboratories, the total T_4 results are becoming fairly well standardized, with only a few commercial kits displaying significant bias. If T_3U ratio results can be similarly standardized, a significant improvement in test reporting will have been achieved.

3. Free Thyroxine Index (FT_4 Index)

The free thyroxine index (FT_4 index) is a calculated value obtained by multiplying the total T_4 result [T_4(D) or T_4(RIA)] by the T_3U ratio. For example, a normal T_4 might be 6.5 μg/dl, and a normal T_3U ratio might be

1.033, resulting in a calculated FT_4 index of $6.5 \times 1.033 = 6.715$. Note: No units are used for the FT_4 index to avoid confusion with FT_4. Although the total T_4 or T_3U ratio alone are dependent on plasma-binding protein concentrations, the FT_4 index corrects for changes in binding protein and provides a mathematically derived value which correlates well with free thyroxine (FT_4) concentration. Patients with an increased T_4 due to estrogens, for example, would have a decreased T_3U ratio, but a normal FT_4 index. The FT_4 index is widely viewed as the best single screen for thyroid disorders. An elevated FT_4 index would be consistent with the diagnosis of hyperthyroidism, whereas a decreased FT_4 index would be consistent with hypothyroidism. Unfortunately, a normal FT_4 index does not necessarily rule out either of these diagnoses, although further investigation might not be indicated without clear clinical symptoms.

4. *Triiodothyronine* [T_3(RIA)]

Although T_3 thyrotoxicosis has been recognized for some time, a widespread interest in total serum T_3 determinations has not developed until fairly recently. The upsurge in serum T_3 determinations is partially the result of the recent commercial availability of reliable radioimmunoassay kits, and partially the realization that compensatory changes in T_3 can occur in several disease states (Cavalieri and Rapoport, 1977). Since total T_3 determinations are almost exclusively performed by radioimmunoassay, the test is designated T_3(RIA).

T_3(RIA) now appears to be the most sensitive laboratory test for hyperthyroidism, with T_3(RIA) levels often elevated earlier than T_4 levels (Shalet *et al.*, 1975). In the patient with suspected hyperthyroidism and a normal FT_4 index, the T_3(RIA) often will be elevated and clarify the diagnosis. The only examples of hyperthyroidism without an elevated T_3(RIA) seem to occur in hyperthyroid patients with an associated severe non-thyroidal illness, in which T_4(RIA) may be elevated, but T_3(RIA) may actually be decreased (Turner *et al.*, 1974; Engler *et al.*, 1978). Even though T_3 is more widely distributed in the tissues and less firmly bound to TBG than is T_4, alterations in plasma-binding proteins affect the total T_3 level. The FT_3 index (similar to the FT_4 index) should correct for these changes, and its use should be encouraged. The FT_3 index is calculated from the T_3(RIA) multiplied by the T_3U ratio, giving a single unitless number. The normal range for T_3(RIA) in most laboratories is 70–200 ng/dl (1.1 to 3.1 nmole/liter) with elevated values seen in almost all patients having clinical hyperthyroidism. Although extremely useful in hyperthyroidism, relative imprecision in assaying low levels makes T_3(RIA) currently unsuitable in the diagnosis of hypothyroidism.

The ratio of total serum T_4 to total T_3 in normal and abnormal patients

shows some interesting variations and illustrates the flurry of interest in T_3(RIA). In euthyroid adults, the serum T_4/T_3 ratio (by weight) averages about 67 (Abuid and Larsen, 1974). In untreated hyperthyroidism, the serum T_4/T_3 ratio falls to about 43, reflecting a greater thyroidal contribution to the serum T_3 level, and possibly explaining the sensitivity of T_3(RIA) in confirming the diagnosis of hyperthyroidism. The source of the elevated serum T_3 changes in thyroid disease. About half of the circulating T_3 derives from the thyroid in hyperthyroid patients, compared to only 20% in euthyroid patients (Abuid and Larsen, 1974); the remainder in both cases results from peripheral deiodination of T_4. Interestingly, in hypothyroidism the ratio of T_4/T_3 is also about 43. Possibly, T_4 secretion in hypothyroidism is decreased to a greater extent than is T_3 secretion, and peripheral deiodination of T_4 is increased in an attempt to compensate for reduced T_4 levels by maintaining a higher concentration of physiologically more active T_3. This type of physiological adaptation is well documented in animal models as well as in humans, where iodine deficiency results in decreased serum T_4 concentrations, but normal or increased serum T_3(RIA) levels. This further complicates the poor correlation between clinical hypothyroidism and serum T_3(RIA) levels. Several procedures for free triiodothyronine (FT_3) are now available, but additional experience is required before a firm evaluation of their usefullness can be made.

5. *Reverse T_3(rT_3)*

In some situations, the deiodination of T_4 to T_3 is reduced and often monodeiodination to produce reverse T_3(rT_3) seems to be enhanced. The rT_3 produced has little or no calorigenic activity, but can be measured in serum using sensitive radioimmunoassay procedures (Chopra, 1978). When formation of rT_3 is enhanced, the laboratory may report a normal T_4(RIA), but a decreased T_3(RIA), as observed in severe illness, cirrhosis, dexamethasone therapy, and starvation. Even an elevated serum T_4(RIA) may accompany a decreased serum T_3(RIA), as in some geriatric patients who clinically appear euthyroid (Burrows *et al.*, 1975). Studies by Chopra *et al.* (1975b) show that serum rT_3 concentrations are increased in a variety of systemic illnesses in which T_3(RIA) levels are decreased. Further, fasting alone causes serum T_3(RIA) levels to decrease and rT_3 levels to increase (Spaulding *et al.*, 1976). These data suggest that a change in deiodination of T_4 to rT_3 (instead of T_3) might serve as a protective mechanism to minimize excess hormonal stimulation during illness or decreased caloric intake (Cavalieri and Ingbar, 1975). Such a change in deiodination is influenced by both carbohydrate and noncarbohydrate diets by a mechanism which is not well understood (Davidson and Chopra, 1979). These

and other data strongly imply a selective (rather than random) deiodination mechanism for T_4 (Chopra, 1978).

Levels of rT_3 also have been studied in amniotic fluid and in serum from the newborn. Compared to adults, amniotic fluid and newborn serum levels of rT_3 are usually increased, whereas T_3(RIA) levels are usually decreased (Chopra *et al.*, 1975a; Chopra and Crandall, 1975). Decreased fetal deiodination of T_4 to T_3 and increased conversion to rT_3 has been postulated to explain these observations. In normal pregnancy, amniotic rT_3 levels decrease with advancing gestational age, while in patients with Rh isoimmune disease higher amniotic fluid rT_3 levels are associated with less favorable outcomes (Osathanondh *et al.*, 1978). In congenital hypothyroidism, decreased cord blood rT_3 levels are seen, but at levels which overlap the normal population (Klein *et al.*, 1978).

While extensive studies on T_3 and rT_3 are expected to continue, our knowledge of the interrelationship between T_4, T_3, and rT_3 has already indicated that interpretation of serum levels must be made with caution where laboratory data and the clinical picture do not correlate. For example, as a practical result of T_3(RIA) and rT_3 levels, one might imply that decreased serum T_3(RIA) in systemic illness does not indicate a need for thyroid replacement, and that an elevated FT_4 index accompanied by a low T_3(RIA) does not indicate hyperthyroidism (Merimee, 1978). At the author's institution, we are considering a breakdown of the traditional "thyroid panel" [T_4(RIA) + T_3U ratio] into two separate panels—a hyperthyroid panel which would include T_3(RIA) and a hypothyroid panel which would include TSH.

6. *Thyroid-Stimulating Hormone (TSH)*

Thyroid-stimulating hormone (TSH) is a glycoprotein hormone consisting of noncovalently bound, dissimilar α and β subunits. The α subunits are immunologically identical to other glycoprotein hormones in man (LH, FSH, HCG), whereas the β subunits have a distinct biological and immunological specificity (Pierce, 1971). Serum TSH levels are easily determined by radioimmunoassay. Mean serum levels of 1–4 μU/ml (reference range 0–10 μU/ml) are found in euthyroid adults and children other than neonates. In the newborn, serum TSH levels rise rapidly from 5 to 15 μU/ml at birth to about 50 μU/ml within the first hour after birth, and then return to adult levels about 3 days after delivery. With TSH secretion inhibited by elevated serum T_3 and/or T_4 levels, TSH levels are low or non-detectable in hyperthyroid patients. Since precision in TSH assays is less than optimal at low levels, a serum TSH is usually not helpful in supporting the diagnosis of hyperthyroidism.

Serum TSH levels are extremely important in establishing the diagnosis of primary (thyroidal) hypothyroidism where greatly elevated TSH levels can be obtained. In patients with a low or borderline T_4 or FT_4 index, a serum TSH is clearly indicated. Slight to marked elevations of TSH characterize primary hypothyroidism, while secondary (pituitary) or tertiary (hypothalamic-extremely rare) hypothyroid patients have low or normal TSH concentrations. While a greatly elevated TSH level usually correlates well with clinically evident hypothyroidism, patients with borderline TSH elevations may not display the classic clinical symptoms of hypothyroidism. These patients should be investigated further with the TRH stimulation test.

In some patients, a euthyroid state may be preserved by TSH stimulation of a functionally-impaired thyroid to produce additional T_3 and T_4. For example, in Hashimoto's disease, in which elevated TSH levels may produce enlargement of the thyroid to increase hormonal synthesis, the serum T_4 levels may be normal. While this may also be observed in other cases of thyroid hypertrophy, such as chronic thyroiditis or endemic goiter, most euthyroid patients with nontoxic nodular goiter have normal TSH levels (Krugman and Hershman, 1978). In extremely rare hyperthyroid patients, serum TSH levels may be elevated.

7. *Thyrotropin-Releasing Hormone (TRH) Stimulation Test*

Thyrotropin-releasing hormone (TRH) is a small peptide (pyroglutamyl–histidyl–prolineamide) which is normally produced in the hypothalamus, but which is also synthesized synthetically and is available commercially. In euthyroid patients, administration of TRH causes a rapid increase in serum TSH levels, reaching a peak at 30 min and gradually returning to baseline levels after several hours. TRH also stimulates release of prolactin (Jacobs *et al.*, 1973) and growth hormone (Faglia *et al.*, 1973), and TSH response to TRH stimulation is significantly reduced by increased serum levels of T_3 or T_4. In the last few years, the TRH stimulation test has moved out of the research laboratory and into greater clinical use whenever serum hormone assays are inconclusive.

Performance of the TRH test involves a baseline TSH and specimens drawn at 15, 30, 45 and 60 min. Sometimes a single post-dose serum TSH level is used in addition to the baseline, although the multiple-sampling technique has been employed more extensively in academic institutions. Figure 3 shows idealized TSH levels encountered in the TRH test. In hyperthyroid patients, a blunted TSH response is observed due to inhibition of the pituitary response by increased plasma T_3 and/or T_4. A reduced response is also seen in patients taking thyroid hormones, as well as in

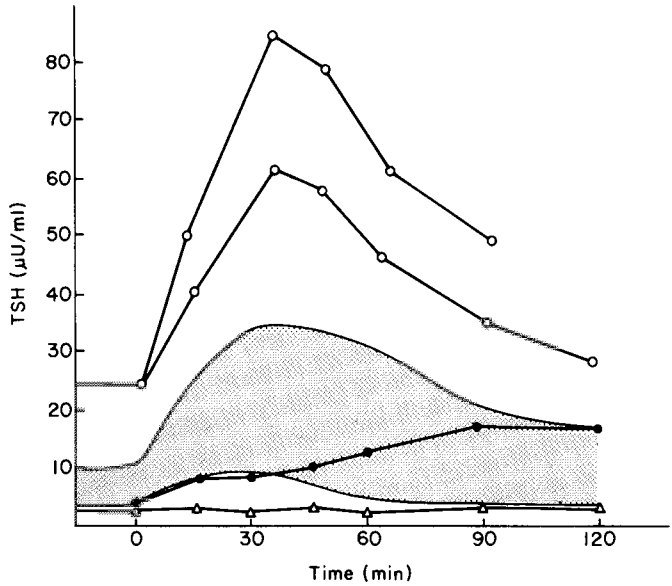


Fig. 3. Serum TSH response to TRH stimulation. The shaded area represents the euthyroid range, whereas the curves represent typical responses seen in primary hypothyroidism (○) (two different patients), tertiary hypothyroidism (●), and secondary hypothyroidism (△). The response in hyperthyroidism is similar to that found in secondary hypothyroidism.

euthyroid patients taking glucocortoids (Otsuki *et al.*, 1973). A blunted TSH response to TRH is, therefore, not conclusive evidence of hyperthyroidism, whereas a normal response almost certainly excludes the diagnosis.

In patients with suspected hypothyroidism, the TRH stimulation test need not be performed if a markedly elevated baseline TSH level supports the diagnosis of primary hypothyroidism. In patients with borderline or normal serum TSH levels, the TRH test can differentiate primary from non-thyroidal hypothyroidism, and help assess pituitary TSH reserve. A large increase in TSH (20 μU/ml or greater) is seen in primary hypothyroidism, making the TRH stimulation test even more sensitive than serum TSH in indicating mild thyroid impairment (Utiger, 1978). In secondary (pituitary) hypothyroidism, the pituitary fails to respond to TRH stimulation, generating a blunted response curve. In the extremely rare case of tertiary (hypothalamic) hypothyroidism, a delayed response curve is sometimes obtained. The test is not affected in normal pregnancy and does not change with increasing gestational age (Ylikorkala *et al.*, 1979). To a large extent, the TRH test has replaced the T_3 suppression test/RAI uptake and the TSH stimulation/RAI uptake test (Hamburger, 1978),

since the newer test can be performed in a single brief visit without the risks involved in administering T_3 or bovine TSH (a foreign protein). Especially in elderly patients with compromised cardiac function, the risks of administering T_3 can easily be avoided. Laboratories performing TSH assays as part of a TRH test must be extremely exact, since precision becomes very important in a test of this type which emphasizes small changes of TSH. Unfortunately, surveys and proficiency testing studies involving TSH and other endocrine assays have demonstrated a great problem in interlaboratory variation. Confidence in interpretation of the TRH response depends on the validity of TSH results and the ability to reproduce those results; thus, the clinical laboratory scientist is well advised to select a TSH method carefully. When evaluating a commercial kit for possible use, the ability to reproduce controls and standards in the low range should be considered an important criterion for acceptability. After a choice has been made, careful monitoring with blind samples or split samples should be a part of routine quality control, as should close monitoring of pipetting devices, temperature, and all other variables that can affect the data.

IV. SPECIAL CONCERNS FOR THE CLINICAL CHEMIST

A. Monitoring Replacement Therapy

Treatment of hypothyroidism involves replacement of T_3 and T_4 to provide adequate reserves and to counter the quantity used in normal metabolism. Replacement therapy can be in the form of desiccated thyroid, L-thyroxine, triiodo-L-thyronine, or mixtures of L-thyroxine and triiodo-L-thyronine. In recent years, L-thyroxine (levothyroxine) has been advocated as the treatment of choice (Jackson and Cobb, 1978; Bunner, 1978) because of its purity, low cost, and predictable response. Since T_4 is converted peripherially to T_3, additional T_3 need not be given. Use of desiccated thyroid, once the mainstay for replacement therapy, has been widely criticized because of its nonuniformity and inappropriate excess content of T_3 over T_4. In many patients, thyrotoxic symptoms due to elevated serum T_3(RIA) can be seen after desiccated thyroid is given, even if the serum T_4 or FT_4 is low or normal. Use of triiodo-L-thyronine (Cytomel) has also fallen into disfavor because it produces a rapid increase in serum T_3 that is transient due to the relatively short half-life. Treatment with T_3 is both expensive and difficult to monitor, since the FT_4 will not be affected by T_3. Likewise, mixtures of T_3 and T_4 are more expensive than T_4 alone, and are also more difficult to monitor by methods other than physical examination for the same reasons.

When levothyroxine is used for replacement therapy, a single oral dose can correct all the symptoms of hypothyroidism (Werner, 1978). Normally, minimal doses are used in the beginning and the serum TSH is monitored to assess adequacy of the dose in primary hypothyroidism; the patient is not euthyroid until the TSH has returned to normal. Monitoring the serum T_4 alone instead of TSH may produce erroneous results. In one study (Maeda *et al.*, 1976), 73 to 76 patients receiving minimal doses of T_4 had normal serum T_4 values, but 45 of 76 still had elevated TSH levels, accompanied by decreased T_3(RIA) values. After a euthyroid state is achieved by adequate doses of T_4, periodic assays for elevated T_4(RIA) and T_3(RIA) can be used to check for overdosage, and elevated TSH used to check for underdosage. Many patients, especially older ones or those with concomitant heart disease, are extremely sensitive to any administration of replacement thyroid hormone and, thus, must be approached slowly with repeated clinical and laboratory assessments. Final adequacy of minimum replacement therapy should be assessed with the TSH (Hamburger, 1978).

The TRH test has also been used as an index of adequate levothyroxine therapy in patients with euthyroid goiter (Krugman and Hershman, 1978). The purpose of therapy in these patients is suppression of goiter-producing TSH secretion. Increased TSH secretion is presumed to be responsible for the goiters in these patients, despite the fact that serum TSH levels are often normal before treatment. Whereas treatment previously was monitored by measuring the suppression of the 24-hr RAI uptake to less than the lower limit of normal (Astwood *et al.*, 1960), it also can be followed by suppression of the TSH response to TRH below the lower limit for normals. Again, the caveat about TSH method selection applies.

B. Antithyroid Drug Therapy

Five types of therapeutic drugs can be used in the treatment of hyperthyroidism; iodide, which inhibits organification (Wolff–Chaikoff effect); complex anions such as perchlorate, which competitively inhibit active transport of iodide; lithium salts, which inhibit hormonal release; β-adrenergic blockers such as propranolol; and thionamides, which inhibit hormonal synthesis. Of these five types, the thionamides are most widely used.

Propylthiouracil, methimazole, and carbimazole are commonly used thionamide drugs which inhibit synthesis of T_3 and T_4 by blocking iodide organification and/or coupling reactions in the thyroid. They are used in long-term treatment, preparation for radioactive iodine therapy or thyroidectomy, hyperthyroidism in pregnancy, and for the treatment of thyroid storm. It is important to remember that release of hormones already syn-

thesized in the gland is not inhibited by thionamides, so that utilization of glandular stores as well as circulating hormones must occur before the effects of the drugs can be seen. If propylthiouracil is given, an additional factor must be considered, since it also inhibits peripheral deiodination of T_4 to T_3. In attempting to assess the adequacy of antithyroid drug therapy, a reduction of the FT_4 index and T_3(RIA) can be helpful when used together with a careful evaluation of the patient's clinical response. As with much drug therapy, toxic reactions and patient noncompliance pose two types of problems, both of which can be monitored with the FT_4 index and the T_3(RIA) if desired.

C. Congenital Hypothyroidism

Maternal TSH, T_3, and T_4 possess only a limited capacity to cross the placenta. Since adequate fetal levels of thyroid hormones are essential for normal growth and development of the central nervous system, congenital hypothyroidism of any cause may result in mental retardation. Primary congenital hypothyroidism (formerly called cretinism) in the United States has an incidence twice that of hyperphenylalaninemia (1 in 5,000 compared to 1 in 10,000), and accounts for 1–27% of admissions to mental institutions. With early diagnosis and treatment soon after birth (less than 3 months of age), the mental retardation can be minimized and possibly reversed (Klein *et al.*, 1972). Although characteristic clinical features of congenital hypothyroidism can be observed during the first few weeks of life, symptoms are sometimes nonspecific or delayed (Smith *et al.*, 1957). Laboratory diagnosis of neonatal hypothyroidism is now simple and reliable enough to be included in every newborn workup.

In New Jersey and several other states, mandatory screening for hypothyroidism is performed on every newborn. Dried blood spots collected on filter paper from infants 3–4 days of age can be used in a joint screening for PKU. Such mass screening programs can be extremely effective (Dussault *et al.*, 1975), especially by a central laboratory performing large volumes of assays that can keep cost to a minimum. Initial screening on filter paper blood spots can involve T_4 or TSH assay or both. Alternately, studies on cord blood T_4 or TSH (Klein *et al.*, 1974) have been used for screening purposes, as have T_4 or TSH measurements performed on serum collected by heelstick. Newer RIA procedures employing very small volumes of serum make these and other RIA procedures easier to perform on newborns. However, the ease of speciment collection using dried blood samples makes this a very attractive procedure for many hospitals and clinics. Variations in filter paper absorption of blood are widely recognized, and caution should be taken to minimize compromising the

quality of the specimens used in the study. Both the committee of the newborns of the American Thyroid Association and the committee on genetics of the American Academy of Pediatrics (Scriver *et al.*, 1977) have recommendations which should be consulted when establishing screening programs.

D. Other Tests

Several other laboratory tests may be important in the evaluation of specific thyroid diseases. Among these are antimitochondrial and antithyroglobulin antibodies, calcitonin, LATS (long-acting thyroid-stimulating hormone), thyroid-stimulating immunoglobulins (TSI), and serum thyroglobulin. This last test, serum thyroglobulin (Tg), is of interest because of reports that indicate elevated serum thyroglobulin levels in all hyperthyroid disease, regardless of cause (Torrigiani *et al.*, 1969; Van Herle *et al.*, 1973) and in euthyroid patients with differentiated thyroid carcinoma (Van Herle and Uller, 1975). It is speculated that elevated serum thyroglobulin levels reflect an overstimulated thyroid gland (Uller and Van Herle, 1978) and that Tg levels may be useful in monitoring therapy. In a recent study (Uller and Van Herle, 1978) of patients on long-term antithyroid therapy, high levels of Tg were associated with an exacerbation of hyperthyroidism, whereas low levels of Tg were associated with remission. This test promises to play an increasingly important role in the laboratory assessment of thyroid disease. A complete discussion of thyroid function is beyond the scope of this chapter, and the reader is referred to the excellent references suggested for additional reading, if detailed descriptions of laboratory testing are required.

REFERENCES

Abuid, J., and Larsen, P. R. (1974). *J. Clin. Invest.* **54,** 201–208.

Astwood, E. B., Cassidy, C. E., and Aurbach, G. D. (1960). *JAMA J. Am. Med. Assoc.* **174,** 459–464.

Bunner, D. L. (1978). *Arch. Intern. Med.* **138,** 978–979.

Burrows, A. W., Shakespear, R. A., Hesch, R. D., Cooper, E., Aickin, C. M., and Burke, C. W. (1975). *Br. Med. J.* **4,** 437–439.

Cavalieri, R. R., and Ingbar, S. H. (1975). *In* "Methods in Enzymology" (B. W. O'Malley and J. G. Hardman, eds.), Vol. 36, pp. 126–132. Academic Press, New York.

Cavalieri, R. P., and Rapoport, B. (1977). *Annu. Rev. Med.* **28,** 57–65.

Chopra, I. J. (1978). *In* "The Thyroid" (S. C. Werner and S. H. Ingbar, eds.), pp. 100–114. Harper and Row, Hagerstown, Maryland.

Chopra, I. J., and Crandall, B. F. (1975). *N. Engl. J. Med.* **293,** 740–743.

Chopra, I. J., Sack, J., and Fisher, D. A. (1975a). *J. Clin. Invest.* **55,** 1137–1141.

Chopra, I. J., Chopra, U., Smith, S. R., Reza, M., and Soloman, D. H. (1975b). *J. Clin. Endocrinol. Metab.* **41,** 1043–1049.

Davidson, M. B., and Chopra, I. J. (1979). *J. Clin. Endocrinol. Metab.* **48,** 577–581.

Dussault, J. H., Coulombe, P., Laberge, C., Letarte, J., Guyda, H., and Khoury, K. K. (1975). *J. Pediatr.* **86,** 670–674.

Engler, D., Donaldson, E. B., Stockigt, J. R., and Taft, P. (1978). *J. Clin. Endocrinol. Metab.* **46,** 77–82.

Faglia, G., Beck-Peccoz, P., Ferrari, C., Travaglini, P., Ambrosi, B., and Spada, A. (1973). *J. Clin. Endocrinol. Metab.* **36,** 1259–62.

Hall, R., and Scanlon, M. F. (1979). *Clin. Endocrinol. Metab.* **8,** 29–38.

Hamburger, J. I. (1978). "Clinical Exercises in Internal Medicine," Vol. 1. Saunders, Philadelphia, Pennsylvania.

Irvine, C. H. G. (1975). *In* "Thyroid Hormone Metabolism" (W. A. Harland and J. S. Orr, eds.), pp. 163–188. Academic Press, New York.

Jackson, I. M. D., and Cobb, W. E. (1978). *Am. J. Med.* **64,** 284–288.

Jackson, I. M. D., and Reichlin, S. (1979). *Endocrinology* **104,** 1814–1821.

Jacobs, L. S., Snyder, P. J., Utiger, R. D., and Daughaday, W. H. (1973). *J. Clin. Endocrinol. Metab.* **36,** 1069–1073.

Kaplan, M. M., and Utiger, R. D. (1978). *Clin. Endocrinol. Metab.* **7,** 97–113.

Klein, A. H., Meltzer, S., and Kenny, F. (1972). *J. Pediatr.* **81,** 912–915.

Klein, A. H., Augustin, A. V., and Foley, T. P. (1974). *Lancet* **2,** 77–79.

Klein, A. H., Foley, T. P., Bernard, B., Ho, R. S., and Fisher, D. A. (1978). *J. Clin. Endocrinol. Metab.* **46,** 336–338.

Krugman, L. G., and Hershman, J. M. (1978). *J. Clin. Endocrinol. Metab.* **47,** 78–83.

Maeda, M., Kuzuya, N., Masuyama, Y., Imai, Y., Ikeda, H., Uchimura, H., Matsuzaki, F., Kumagai, L. F., and Nagataki, S. (1976). *J. Clin. Endocrinol. Metab.* **43,** 10–17.

Martin, J. B., and Reichlin, S. (1970). *Endocrinology* **87,** 1032–1040.

Merimee, T. J. (1978). *Postgrad. Med.* **63,** No 6, 113–117.

Nagataki, S., and Ingbar, S. H. (1964). *Endocrinology* **74,** 731–736.

Nelson, J. C., Weiss, R. M., and Lewis, J. E. (1974). *J. Clin. Invest.* **53,** 416–422.

Nelson, J. C., Weiss, R. M., Palmer, F. J. Lewis, J. E., and Wilcox, R. B. (1975). *J. Clin. Endocrinol. Metab.* **41,** 1118–1124.

Osathanondh, R., Chopra, I. J., Tulchinsky, D., and Umansky, I. (1978). *J. Clin. Endocrinol. Metab.* **46,** 365–368.

Otsuki, M., Dakoda, M., and Baba, S. (1973). *J. Clin. Endocrinol. Metab.* **36,** 95–102.

Pierce, J. G. (1971). *Endocrinology.* **89,** 1331–1344.

Reichlin, S. (1978). *In* "The Thyroid" (S. C. Werner and S. H. Ingbar, eds.), pp. 151–173. Harper and Row, Hagerstown, Maryland.

Reichlin, S., and Utiger, R. D. (1967). *J. Clin. Endocrinol. Metab.* **27,** 251–255.

Robbins, J., Rall, J. E., and Gorden, P. (1974). *In* "Duncan's Diseases of Metabolism" (P. K. Bondy and L. E. Rosenberg, eds.), 7th ed., pp. 1009–1104. Saunders, Philadelphia, Pennsylvania.

Sakurada, T., Rudolph, M., Fang, S. L., Vagenakis, A. G., Braverman, L. E., and Ingbar, S. H. (1978). *J. Clin. Endocrinol. Metab.* **46,** 916–922.

Schadlow, A. R., Surks, M. I., Schwartz, H. I., and Oppenheimer, J. H. (1972). *Science* **176,** 1252–1254.

Scriver, C. R., Feingold, M., Mamunes, P., and Nadler, H. L. (1977). *Pediatrics* **60,** 389–404.

Shalet, S. M., Beardwell, C. G., Lamb, A. M., and Gowland, E. (1975). *Lancet* **2,** 1008–1010.

Smith, D. W., Blizzard, R. M., and Wilkins, L. (1957). *Pediatrics* **19,** 1011–1022.

Solomon, D. H., Benotti, J., De Groot, L. J., Greer, M. A., Oppenheimer, J. H., Pileggi, V. J., Robbins, J., Selenkow, H. A., Sterling, K., and Volpe, R. (1976). *J. Clin. Endocrinol. Metab.* **42,** 595–598.

Spaulding, S. W., Chopra, I. J., Sherwin, R. S., and Lyall, S. S. (1976). *J. Clin. Endocrinol. Metab.* **42,** 197–200.

Taurog, A. (1974). *In* "Handbook of Physiology" (R. O. Greep and E. B. Astwood, eds.) Sect. 7, Vol. III, p. 101. Am. Physiol. Soc., Washington, D. C.

Taurog, A. (1978). *In* "The Thyroid" (S. C. Werner and S. H. Ingbar, eds.), pp. 31–61. Harper and Row, Hagerstown, Maryland.

Torrigiani, G., Doniach, D., and Roitt, I. M. (1969). *J. Clin. Endocrinol. Metab.* **29,** 305–314.

Turner, J. G., Brownlie, B. E. W., and Sadler, W. A. (1974). *Lancet* **1,** 407.

Uller, R. P. and Van Herle, A. J. (1978). *J. Clin. Endocrinol. Metab.* **46,** 747–755.

Utiger, R. D. (1978). *In* "The Thyroid" (S. C. Werner and S. H. Ingbar, eds.), pp. 367–374. Harper and Row, Hagerstown, Maryland.

Van Herle, A. J., and Uller, R. P. (1975). *J. Clin. Invest.* **56,** 272–277.

Van Herle, A. J., Uller, R. P., Matthews, N. L., and Brown, J. (1973). *J. Clin. Invest.* **52,** 1320–1327.

Van Zyl, A., and Edelhoch, J. (1967). *J. Biol. Chem.* **242,** 2423–2427.

Werner, S. C. (1978). *In* "The Thyroid" (S. C. Werner and S. H. Ingbar, eds.), pp. 965–970. Harper and Row, Hagerstown, Maryland.

Wolfe, H. J., Voelkel, E. F., and Tashjian, A. H. (1974). *J. Clin. Endocrinol. Metab.* **38,** 688–693.

Wolff, J., and Chaikoff, I. L. (1948). *J. Biol. Chem.* **174,** 555–564.

Ylikorkala, O., Kivinen, S., and Reinila, M. (1979). *J. Clin. Endocrinol. Metab.* **48,** 288–292.

ADDITIONAL READING

The following texts offer the interested reader additional insight into the theoretical, research, laboratory, and clinical aspects of thyroid function (several references are classic endocrinology texts, whereas others are smaller paperback working references):

Bondy, P. K., and Rosenberg, L. E., eds. (1980). "Metabolic Control and Disease," 8th ed. Saunders, Philadelphia, Pennsylvania. (Classic text).

Evered, D., and Hall, R., eds. (1979). "Clinics in Endocrinology and Metabolism," Vol. 8, No. 1. Saunders, Philadelphia, Pennsylvania. (Recent update).

Hershman, J. M., ed. (1977). "Endocrine Pathophysiology: A Patient Oriented Approach." Lea & Febinger, Philadelphia, Pennsylvania. (Paperback; stresses clinical aspects).

Hamburger, J. E. (1978). "Clinical Exercises in Internal Medicine," Vol. 1. Saunders, Philadelphia, Pennsylvania. (Paperback; stresses clinical aspects).

Watts, N. B., and Keffer, J. H., eds. (1978). "Practical Endocrine Diagnosis." Lea & Febiger, Philadelphia, Pennsylvania. (Paperback; excellent ready reference).

Werner, S. C., and Ingbar, S. H., eds. (1978). "The Thyroid." Harper and Row, Hagerstown, Maryland. (Extensive in depth approach).

Williams, R. H., ed. (1974). "Textbook of Endocrinology," 5th ed. Saunders, Philadelphia, Pennsylvania. (Classic text).

4

Coagulation Chemistry

LOUIS A. KAZAL

CLINICAL BIOCHEMISTRY
Contemporary Theories and Techniques, Vol. 2

ISBN 0-12-657102-3

I. INTRODUCTION

The last four decades have been witness to enormous advances in our knowledge of blood coagulation and hemostasis.* The combined efforts of many scientific disciplines, especially of clinical medicine, physiology, and biochemistry, have produced a massive accumulation of experimental data in support of current concepts and theories. The data are not always complementary and are sometimes even conflicting; nevertheless, sound principles and new directions have emerged that have prepared a foundation for a promising and challenging future. More than a dozen clotting factors have been discovered in the blood, and an intricate dual mechanism of blood clotting has been proposed to account for their interactions and for the ultimate transformation of fibrinogen to fibrin in the easily observable blood clot, which has bewitched investigators from the time of Malpighi, who first described the fibrillar nature of the blood clot in 1666 (Cohn, 1953). Recent advances in our knowledge of the blood platelet have provided new insights into hemostasis. The platelet appears to have the unique capability of interacting not only with certain components of the coagulation mechanism, but also with itself and with the injured vessel wall in order to produce a hemostatic plug. The coagulation system, itself, can no longer be considered a simple straightforward mechanism, solely concerned with the formation of a blood clot, for its activation becomes the focus for the activation of other equally complex physiologic mechanisms leading to the formation of complement, fibrinolysin, kinins, and tissue permeability factors. If we add to all these systems the specific inhibitors and activators that provide the regulatory physiologic controls, then we have a galactic, physiologic ensemble of interrelated mechanisms that somehow maintain hemostasis in a normal individual. When these systems become unbalanced, hemorrhagic and thrombotic events occur.

These new observations and theories undoubtedly go beyond the expectations and dreams of scientists and physicians of the nineteenth and early twentieth centuries, many of whom had broad and basic concepts of hemostasis that were not unlike those of our day. If present-day knowledge of hemostasis and blood coagulation had been available to them, one can imagine the delight of Giulio Bizzozero (1873–1901), who first distinguished the platelet from other blood cells; of Olof Hammarsten (1841–1883), who purified fibrinogen and studied its coagulability by thrombin;

* It is impossible in the present explosion of literature to cite all pertinent publications on this subject. Hopefully, the reader will refer to the cited review papers for other important publications that otherwise deserve specific citation.

of Andrew Buchanan (1798–1828), who, by his pioneering studies, laid the foundation for an enzymic concept of blood clotting*; of Alexander Schmidt (1831–1894), the "father of blood coagulation," who, by his work with tissue thromboplastin and fibrin ferment [named "thrombin" by Rudolf Virchow (1821–1902)], firmly established the enzymic nature of the clotting enzyme; of Maurice Arthus (1862–1945), who demonstrated the requirements for calcium ions in blood coagulation; and of Paul Morawitz (1879–1936), who demonstrated the absence of thrombin or its hypothetical precursor, prothrombin, in tissue juices. It was Morawitz who, in 1904, welded together the observations of his predecessors into the Classical Theory of Blood Coagulation, which is the basis of our current theory and the beginning of modern concepts of blood coagulation (Morawitz, 1905). This four-factor theory remained basically unchanged for close to four decades.

What follows historically is too extensive and intricate to relate here. However, with the development of the prothrombin time test by Armand Quick (1894–1978) in the 1930s and 1940s, the biochemical studies of prothrombin by Walter Seegers, and the discovery of proaccelerin (a fifth clotting factor) by Paul Owren, research in blood coagulation advanced exponentially at an unprecedented pace, albeit often with confusing results. For an historical treatise, the reader is referred to two excellent works that provide a detailed and interesting retrospect of developments in hemostasis and blood coagulation (Owen *et al.*, 1969; Tullis, 1976). Chemistry always has played an important role in the development of mechanisms in blood coagulation. Today, especially under the influence of modern biochemical technology, it is unfolding the basic molecular events and interactions involved in the blood clotting process and its related systems. It is this aspect of coagulation to which this review is directed.

II. HEMOSTASIS

Before considering the chemical aspects of coagulation, a few simple generalizations about hemostasis seem appropriate in order to place the coagulation mechanism in its proper perspective. Hemostasis usually is defined from its Greek roots as the "arrest of bleeding" (hemo: blood; statikos: arrest or stop). This is a physiologic mechanism that prevents

* An interesting historical account of contemporary comments and experiments on the pioneering work of Buchanan, and its relation to that of Schmidt and his "fibrin ferment," is given by Gamgee (1879).

the loss of blood through a "break" in the vessel wall arising from physical or pathologic injury. In the normal individual, hemorrhage ceases with the formation of a hemostatic plug, except in larger arterial vessels, and the flow of blood through the "break" stops. The hemostatic plug requires functional platelet and blood coagulation mechanisms, with the former mechanism most likely being the primary event. According to the foregoing concept hemostasis is defined solely in terms of events leading to hemorrhage. However, hemostasis must also explain thrombotic phenomena. In clinical thrombosis and intravascular clot formation, blood also ceases to flow, in this case through the intact vessel itself, as a result of the formation of a thrombus, a plug of platelets or fibrin, or both attached to the inner surface of the vessel wall. This has been aptly referred to as "hemostasis in the wrong place" (Macfarlane, 1977). Broadly speaking, both the thrombus and the hemostatic plug invoke essentially similar mechanisms, although the inciting causes are different, and both involve the blood clotting mechanism to provide the fibrin that cements the platelet aggregates into a firm plug. It follows, then, that normal hemostasis must provide the essentials required for the formation of a hemostatic plug in order to prevent bleeding, and, at the same time, it must maintain the fluidity of the blood (Tocantins, 1952), i.e., it must not permit intravascular clot or platelet aggregate formation (thrombus formation) to occur within the vasculature in order to maintain a proper circulation of the blood. This kind of control calls for a vastly complicated mechanism involving not only platelet function and blood coagulation, but also the action of related inhibitors and activators of various types, a subject beyond the scope of this review, and treated elsewhere (e.g., Reich *et al.*, 1975; Shapiro and Hultin, 1975).

A. Formed Elements and the Role of Platelets

The concept of a hemostatic plug is that of a platelet aggregate reinforced with fibrin strands or of a fibrin clot attached to a platelet base on the vascular endothelium. Platelets react to many foreign surfaces (see Mason *et al.*, 1976), especially to the damaged endothelium of vessel walls, by first undergoing a shape change from smooth disk to sphere, after which adhesiveness develops. The platelets then adhere to the vessel wall and to each other to form aggregates. In the absence of a participating blood coagulation process and of fibrin formation, platelet aggregation, when it occurs, is reversible. With the formation of fibrin, irreversible platelet aggregation is established, and a firm hemostatic plug develops. The essential event, apparently, is the activation of the platelet, an extremely rapid process that occurs in fractions of a second or less

(Alfors *et al.*, 1976), and must precede fibrin formation, since blood coagulation is a relatively slower process.

The role of erythrocytes and leukocytes in hemostasis has only recently been studied in depth. Heretofore, erythrocytes were assigned a more or less passive role; their presence in a clot was considered as providing only rigidity and strength. Leukocytes have been considered only as contributory to coagulation processes in pathologic states.

More recently, however, hemodynamic evidence has suggested that erythrocytes enhance the adhesiveness of platelets by supplying adenosine diphosphate (ADP) under conditions of shear stress developed by fluid mechanical forces on the erythrocyte at the point of turbulence created by the injured endothelial surface of the vessel wall (Born, 1977). Recent research on leukocytes focuses on their protease tissue-factor activity that is capable of activating the clotting mechanism (Niemetz, 1972; Lerner *et al.*, 1977). Polymorphonuclear leukocytes also are attracted to platelet aggregates by chemotactic agents elaborated by platelets from platelet arachidonate (Turner *et al.*, 1975) or a factor that interacts with the fifth component of complement (Weksler and Coupal, 1973). In these more subtle ways, therefore, both erythrocytes and leukocytes may contribute to hemostasis, a field of study not yet fully elucidated, but exciting and challenging.

Of the formed elements, the platelets, without question, provide the keynote to hemostasis. These smallest of blood cells contain a host of active protoplasmic and membranous agents. In addition to releasing ADP upon degranulation, platelets elaborate other potent aggregating agents, among them prostaglandin endoperoxides and thromboxane A_2 (Samuelson *et al.*, 1976), if they are exposed to subendothelial structures of the vessel wall. Collagen, like thrombin, induces release of platelet granules that contain ADP, ATP, and vasoactive serotonin (5-hydroxytryptamine). Phospholipase A_2 in platelets also is activated by contact with collagen (or thrombin) and is responsible for the release of arachidonate from platelet phospholipids. The platelet enzyme, cyclooxygenase, then converts arachidonate to thromboxane A_2, a potent aggregating agent. Platelets also release platelet factor 3, a membrane phospholipoprotein, which is an essential component of the clotting mechanism by virtue of the fact that it provides a phospholipid surface for the activation of factor X to Xa, and of prothrombin to thrombin. Thrombin, itself, activates platelets and also causes the deposition of fibrin on their surfaces, thus participating in the reactions necessary for the formation of the irreversible platelet aggregate that constitutes the physiologic hemostatic plug. Recent surveys of the role of platelets in normal and abnormal hemostasis document these many interrelationships and the complicated chemical machinery of the platelet

(Mustard and Packham, 1977; Born, 1977; Crawford and Taylor, 1977; Hardisty, 1977; Miletich, 1977). Exciting as the chemistry of the platelet may be, we can only consider the biochemical properties of platelet factor 3, which functions in the intrinsic clotting mechanism.

B. Coagulation Factors of Blood Plasma

In a hemostatic plug the fibrin strands are enmeshed in the platelet aggregates. The fibrin is derived from the blood plasma by a mechanism requiring the activation of a number of clotting factors present in the plasma. The mechanism that is activated solely by factors within the blood is called the "intrinsic clotting mechanism"; that which is activated by tissue factor (factor III), "the extrinsic clotting mechanism." The intrinsic system participates in the formation of a hemostatic plug whereas the extrinsic system participates in the formation of a fibrin clot. Such a clear-cut and simple separation of the two mechanisms, however, is theoretical, since there is evidence to suggest that the extrinsic system also may operate intravascularly by virtue of the presence of tissue factor III in endothelial cells (Zeldis *et al.* 1972). Furthermore, normal hemostasis requires both systems, since hereditary deficiencies of certain plasma clotting factors in either of these systems produce bleeding diatheses. In spite of such reasoning, nevertheless, much experimental evidence supports the intrinsic system as the most important facet in the formation of a hemostatic plug (Mustard and Packham, 1977).

Some of the coagulation factors present in blood plasma function in both clotting mechanisms whereas others are more specific:

1. Participating in the intrinsic pathway are clotting factors VIII, IX, XI, XII, prekallikrein, and high-molecular-weight kininogen (platelet factor 3, which also functions in this mechanism, is not a plasma constituent).
2. Participating in the extrinsic pathway are factors III and VII (factor III in this system is tissue-derived).
3. Factors common to both pathways are factors V, X, XIII, prothrombin, fibrinogen, and calcium ions (factor IV).

The three groups of factors are shown in Table I. The properties and mechanism of interactions of these factors is the subject of this review. The discovery of so many factors in human or bovine plasma as necessary components in the formation of a fibrin clot clearly underscores the complicated nature of the biological process that has been devised by nature and modified by evolutionary development into a sequence of reactions in order to protect the organism against the loss of blood and the maintenance of its fluidity (Tocantins, 1950).

TABLE I

Blood Coagulation Factors in Intrinsic and Extrinsic Mechanisms[a]

Intrinsic		
Factor XII	Prekallikrein	
	High-molecular-weight kininogen	
Factor XI		*Common Pathway*
[Factor IX	Phospholipid][b]	[Factor X Phospholipid][b]
[Factor VIII	Ca^{2+}]	[Factor V Ca^{2+}]
		Prothrombin
		Fibrinogen
Extrinsic		Factor XIII
Factor III		
Factor VII		
Ca^{2+}		

[a] Factors are grouped in sequential order of activation.
[b] Brackets indicate factors involved in phospholipid complex formation.

C. Role of Vasculature and Endothelial Cells

The endothelium of blood vessels has long been visualized as perhaps the most important biologic structure that contributes to normalcy in hemostasis, in that its endothelial cells provide a nonactivating surface for blood. As long as this inner layer of cells remains intact, neither platelets nor the coagulation mechanism are activated. When this surface layer of cells is injured and its subendothelial structures become exposed, both platelet and coagulation systems become activated (Smith, 1977). Among the activators observed in endothelium are the following: collagen, which can activate platelets (Hovig, 1963), and factor XII (Niewiarowski, 1966; Walsh, 1972); thromboplastin in the endothelial cell surface (Nemerson and Pitlick, 1972); and plasminogen activator (Todd, 1964; Nilsson and Pandolfi, 1970). The nonactivating character of intact endothelial cells, however, does not appear to be the result of a simple passive effect but may be due to an active elaboration of a platelet-derived prostaglandin inhibitor (prostaglandin X, PGI_2, or more commonly, prostacylin) which is capable of inhibiting platelet aggregation (Moncada *et al.*, 1976). Arterial vessel wall contains a microsomal enzyme that transforms platelet endoperoxides to prostacylin. This new observation supports the unique antithrombotic property of intact vascular endothelium, although the issue is clouded by the observation that the microsomal enzyme may really be present in the nonendothelial part of the blood vessel (Hornstra *et al.*, 1978). If the function of prostacylin, the most potent antiaggregating agent of the prostaglandin type, can be established, and the evidence for it is highly convincing, then a unique system exists for maintaining normal hemostasis in terms of an inhibitor of one of the early reactions in clotting, namely, platelet aggregation.

D. Contributions of the Extravascular System

The extravascular tissue that surrounds blood vessels has long been known to contain substances, tissue thromboplastins, which accelerate the coagulation of blood (Schmidt, 1875: zymoplastin). Almost all tissue extracts elicit some degree of thromboplastic activity in the prothrombin time test (Quick, 1942; 1946) with brain, lung, and placenta exhibiting the greatest activity. Tissue thromboplastin, which is not normally found in blood or body fluids, functions in the extrinsic clotting mechanism via a specific interaction with blood clotting factor VII. From the viewpoint of hemostasis, thromboplastin, commonly referred to as factor III or, more recently, as "tissue factor" (Nemerson and Pitlick, 1970), theoretically functions as a rapidly acting clotting agent when blood escapes from the vascular system into the tissue space. However, the rapidity with which such extravasated blood would clot must be dependent upon the type of tissue involved, since some tissue extracts, e.g., saline muscle extract, are considerably less active; however, even a slowly acting thromboplastin should contribute in some measure to the hemostatic process. Undoubtedly, the contribution of tissue factor to hemostasis must be quite significant, although very difficult to evaluate, since a genetic deficiency of factor III has not been recorded.

Extravascular tissue also can contribute to hemostasis physically by the pressure effect of tissue on extravasated blood, resulting in the localization of hemorrhage, and it may also contribute in other ways, e.g., through the elaboration of hormones and other agents contributed by organ tissue (Brozović and Path, 1977); however, since only the chemistry of factor III is appropriate in this review, these aspects of hemostasis cannot be considered here.

III. OVERVIEW OF THE MECHANISM OF BLOOD COAGULATION

The *autoprothrombin-derivative* (Seegers, 1962), the *waterfall* (Davie and Ratnoff, 1964) and the *cascade* (Macfarlane, 1964) mechanisms of blood clotting laid the groundwork for current concepts in the chemistry of blood coagulation. The last two schemes feature what is currently referred to as the intrinsic mechanism, based on the activation of a series of zymogen clotting factors. The *autoprothrombin-derivative theory* expounds a scheme in which prothrombin is the central and most important reactant. Comparative reviews of these mechanisms have been published earlier (Kazal, 1971; Macfarlane, 1976). The clotting factors operative in

these mechanisms are shown in Table I. Our knowledge of these factors is mostly derived from studies on human and bovine blood. Although differences in properties exist experimentally between human and bovine factors, the overall mechanism of coagulation accommodates factors isolated from either source, as well as factors isolated from other mammalian species.

The basic postulates of the *waterfall* and *cascade* mechanisms were (1) the existence in plasma of proenzyme clotting factors, (2) the activation of these proenzymes to active serine proteases, and (3) the sequential activation of individual clotting factors in a stepwise manner, such that starting with the activation of Hageman factor (XII) and ending with the conversion of prothrombin to thrombin, fibrinogen would be transformed to fibrin. Both mechanisms emphasized the intrinsic pathway, but their authors also recognized the intervention of an extrinsic tissue factor in these reactions. Both of the proposed mechanisms indicated the importance of ionic calcium and of an external source of phospholipid for the activation of certain factors.

Subsequent experimental evidence modified the proposed mechanisms. Not all of the clotting factors (factors V and VIII) proved to be convertible to active enzymes; although these two were capable of indirect enzymatic activation by thrombin, resulting in an increased state of activity, they were not proenzymatic components, but rather they participated as phospholipid micellar complexes with the active forms of these two clotting factors. Specifically, factor VIII, activated factor IX, and calcium ions were found to form a complex with phospholipid, which activates factor X. Activated factor X, factor V, and calcium then form a complex with phospholipid, which activates prothrombin to thrombin.

Two divergent activation mechanisms of coagulation eventually were proposed to account for (a) the slower activation process of fibrin formation by the intrinsic mechanism, which is wholly dependent on clotting factors residing in blood itself, and (b) the considerably more rapid activation process of the extrinsic system, which is dependent upon the participation of an extravascular factor, i.e., tissue factor or tissue thromboplastin (factor III). Factor III alone does not account for the major difference between intrinsic and extrinsic clotting, for it must interact with a plasma component, factor VII, a specific plasmatic component of the extrinsic system. Interestingly, recent evidence suggests that factor VII also may become involved in the intrinsic mechanism under certain conditions. In any case, both mechanisms subsequently arrive at a common pathway at the level of factor X activation, prothrombin conversion, fibrinogen transformation to soluble fibrin, and its transamidation by factor XIII to insoluble fibrin.

Strong support for the existence of coagulation factors, aside from those involving factors III and IV, has been found in patients with clinically manifested or with hereditary factor deficiencies. Biochemical studies have contributed to identification and isolation of clotting factors, most of which proved to be plasma proteins, many having the characteristics of proteolytic enzymes. The initiating event in the intrinsic system is clearly not enzymatic and appears to be a physical one, involving surface contact. However, the first factor to be involved, the Hageman factor, acquires enzymatic activity upon contact with foreign surfaces, an activation first demonstrated *in vitro* with glass or diatomaceous earths and, subsequently, by biologic and subcellular structures such as platelets, collagen, and basement membranes of the vessel wall. The function of Hageman factor itself is now envisioned as a complicated series of activations involving a feedback mechanism requiring the activation of prekallikrein and a high-molecular-weight kininogen. Both clinical and biochemical observations support this mechanism. The proteolytic fragments of factor XII also participate in other physiologic mechanisms, namely, fibrinolysis, complement activation, kinin formation, and tissue permeability. These complicated interrelationships in the contact system of the activation of factor XII (and of platelets) are presently under intensive investigation and hold the key to our understanding of hemostasis.

The biochemistry of blood coagulation, however, was not solely established by the study of *cascade* and *waterfall* sequences. Since 1940, the biochemistry of prothrombin and its accessory factors (Seegers, 1940) has been investigated extensively, in fact, providing the foundation and much of the stimulus for our present state of knowledge of many areas of blood coagulation. Seegers' concept of the coagulation of blood embodied in his *autoprothrombin derivative theory* places emphasis on prothrombin as the central molecule from which well-defined biochemical fragments, prethrombin and autoprothrombin, are derived (Seegers, 1967; 1976). With accessory factors (phospholipids, calcium, factors III, V, and VIII), prothrombin subunits are believed to be converted to thrombin, mainly by the enzyme autoprothrombin C. According to this hypothesis, prothrombin is a single molecule containing the activities of factors II, VII, IX and X, all of which have similar biochemical properties and which are dependent on vitamin K for their biosynthesis. The enzymatic nature of this mechanism has long been stressed, actually since the advent of prothrombin activation by citrate ions (Seegers *et al.*, 1950). Furthermore, the derivatives of prothrombin correct clotting abnormalities of plasma deficient in factors VII, IX, or X. Although the terminology of this theory has not received widespread adoption, its contribution to our knowledge of the biochemistry of blood coagulation has been very significant, especially for

the chemistry of prothrombin and thrombin. The concept of a single molecular entity comprising individual plasma proteins is not new in biochemistry and was supported by the work of Block and Keller (1960), who considered isolated plasma proteins as experimentally derived fragments of a larger molecule called orosin. Interactions leading to an association of individual plasma proteins were demonstrated by the work of Cohn and Edsall and their collaborators, as early as 1950 (for references and brief account, see Kazal, 1962).

A. Nomenclature of Blood Clotting Factors

The earlier confusion of nomenclature characterized by numerous (up to 13) common names for a single clotting activity was clarified by the adoption of a Roman numeral system for clotting factors by the International Committee for Nomenclature of Blood Clotting Factors (Wright, 1962). All clotting factors were assigned a specific number, but the use of well-known common names (prothrombin, fibrinogen, calcium, thromboplastin, or tissue factor) has persisted, although the Roman numerals are frequently appended parenthetically. The committee numbered the nonactivated factors from I to XII, omitting VI, which in reality was an activated state of factor V. It suggested the use of symbols and common names, some of which are listed in Table II. The action of the committee brought order out of chaos to the terminology of clotting factors.

As indicated before, all of these factors are proteins, except for calcium, and all are zymogens except factors I, III, IV, V, and VIII. All are present in blood plasma except factor III. General reviews of coagulation proteins have been published; some are brief (e.g., Aronson, 1977; Green, 1978), others are more extensive (e.g., Doolittle, 1975; Davie and Hanahan, 1978). The activated enzymatic states of these factors are indicated usually by appending "a" or prefixing "act" to the Roman numeral. The thrombin-activated states of factors V and VIII (V′ and VIII,′ respectively) do not possess enzymatic activity. Frequently, the word "factor" is omitted in the literature in referring to specific clotting factors, and numerals are used alone.

Subsequent research has disclosed the probable existence of a number of clotting factors based on studies of the blood of patients with specific clotting abnormalities or clinical deficiencies. Most of these have not been accredited the status of a coagulation factor; the distinguishing characteristics of about 20 such factors have been described (Owen *et al.*, 1969). Two of these, however, have gained acceptance as clotting factors functioning in the contact phase of clotting as part of the feedback loop in the activation of factor XII. These are listed in Table I as prekallikrein and

TABLE II

Nomenclature of Blood Clotting Factors

Factor no.	Common name	Symbol	Activated state
I	Fibrinogen	∅	
II	Prothrombin		IIa (Thrombin)
III	Tissue factor		
	Tissue thromboplastin		
IV	Calcium	Ca^{2+}	
V	Accelerator globulin	AcG	V′
VII	Proconvertin		VIIa
	Autoprothrombin-1	Auto-I	
VIII	Antihemophilic factor	AHF	VIII′
	Antihemophilic globulin	AHG	
IX	Plasma thromboplastin component	PTC	IXa
	Autoprothrombin II		
X	Stuart factor	SF	Xa
	Autoprothrombin III	Auto-III	Autoprothrombin C
XI	Plasma thromboplastin antecedent	PTA	XIa
XII	Hageman factor	HF	XIIa, HFa
XIII	Fibrin stabilizing factor	FSF	XIIIa
	Profibrinoligase		Fibrinoligase
	Plasma protransglutaminase		Transglutaminase
	Prekallikrein (Fletcher factor)	PK	Kallikrein
	High-molecular-weight kininogen (Flaujeac, Fitzgerald, or Williams factor)	HMWK	

high-molecular-weight kininogen. Numeral designations have not been assigned, since these are well-known biochemical entities.

Finally, the recently suggested terminology for factor VIII, the antihemophilic factor, is not listed in Table II. The complex chemistry of this factor is still not completely elucidated. Factor VIII is now described by its coagulant and antigenic properties, The nomenclature used in this review will be factor VIII: C for coagulant and factor VIII: RAg for VIII-related antigen. Activities related to factor VIII in von Willebrand's disease associated with abnormal aggregation of platelets and bleeding time will be referred to as factor VIII: RWF (factor VIII-related von Willebrand factor, Bloom and Peake, 1977).

B. The Intrinsic Clotting System

A simplified scheme illustrating current concepts in the sequence of reactions in intrinsic clotting is shown in Fig. 1. It is based on the *waterfall* and *cascade* hypotheses, modified to include the phospholipid micel-

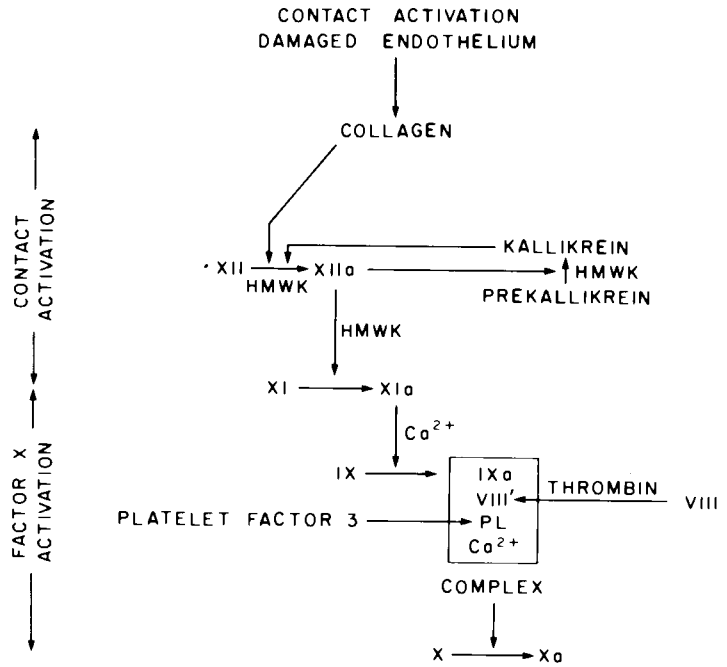


Fig. 1. Intrinsic mechanism of blood coagulation emphasizing collagen as the contact activator on exposure of subendothelium resulting from vascular injury and the sequence of activations leading to factor Xa. The feedback loop involving kallikrein and high-molecular-weight kininogen, and the phospholipid micellar complex of IXa and VIII′ are illustrated. PL, equals phospholipid. Thrombin-activated VIII is assumed to be a functional part of the complex.

lar complex of activated factor IX, and the recent advances in the contact activation stage of coagulation. Present evidence suggests that the activation of factor XII becomes amplified by the protease, kallikrein, which is formed from prekallikrein by proteolytic action of XIIa released in the initial stages of clotting. High-molecular-weight kininogen acts as a cofactor and accelerates the formation of kallikrein by XIIa, of XII to XIIa by kallikrein, and of XI to XIa by XIIa. Factor XIIa continues the clotting sequence by converting XI to XIa, which enzymatically activates with ionic calcium factor IX to IXa. Enzymatically active factor IX, factor VIII, phospholipid, and calcium ions form a complex, presumably on the surface of phospholipid micelles. It is assumed that factor VIII initially participates in complex formation in a nonactive form, but experimental evidence indicates that thrombin-activated VIII, i.e., VIII′, is considerably more active and, therefore, it must be the primary participant. The phospholipid micellar complex containing IXa and VIII′ activates factor X to

Xa, bringing this mechanism to the point where it participates in the common pathway.

C. The Extrinsic Clotting System

As indicated previously, recalcified blood and plasma coagulate rapidly when tissue extracts of brain, lung, or placenta are added. The rapid clot formation (10–15 sec, depending upon the care with which preparations are made) is the result of the reaction of tissue factor III and plasma factor VII. In the absence of factor III, intrinsic coagulation is much slower, requiring minutes instead of seconds for fibrin formation. The increased extrinsic rate is due to the direct activation of factor X and the bypassing of reactions that occur in intrinsic clotting. The phenomenon is the basis of the Quick prothrombin time test.

Contact of blood with tissue is necessary for the activation of the extrinsic system from a physiologic point of view. Recent demonstration of the presence of high a concentration of factor III in endothelial cells (Zeldis *et al.*, 1972) lends much support for the participation of the extrinsic mechanism *in vivo* following injury to blood vessels without extensive damage of surrounding tissue. Figure 2 illustrates two pathways that have evolved from studies with purified factors. Unfortunately, no claims can be made that these reactions actually represent physiologic coagulation.

According to one group of investigators, factor III has peptidase and endopeptidase activity, and bovine factor VII is a protease with a serine

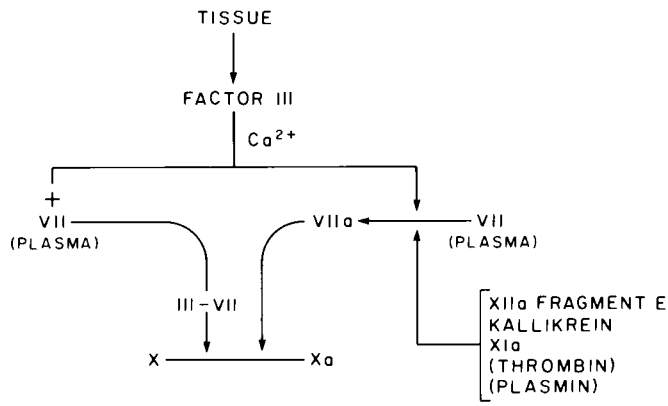


Fig. 2. Extrinsic mechanisms of blood coagulation dependent on factors III and VII. Two possible pathways are supported by *in vitro* experiments. One is through a complex of factors III and VII, in which tissue factor causes an unfolding of an already existent, active serine site, and by which a complex of III + VII activates factor X. The other pathway involves the activation of factor VII by factor III to VIIa, which, by itself, activates factor X. Other agents also can convert VII to VIIa.

active site. Factor VII, however, will not react with its substrate, factor X, unless it forms a complex with factor III, which induces a conformational change in factor VII, thereby allowing a substrate–enzyme interaction. The complex activates factor X (Jesty and Nemerson, 1974; Jesty *et al.*, 1975).

Other investigators hold the belief that factor VII has no enzymatic activity until factor III converts it to factor VIIa, which by itself and in the absence of factor III can convert factor X to Xa (Østerud *et al.*, 1972). A number of enzymes also can convert factor VII to VIIa, as shown in Fig. 2 (Radcliffe and Nemerson, 1976).

D. Common Pathway to Thrombin Formation and the Fibrin Clot

Figure 3 illustrates the final reactions in the formation of the fibrin clot. The extrinsic or intrinsic activators, which are enzymatically active, react with factor X to produce the same end product, Xa. The same activation is produced by such nonphysiologic agents as trypsin and Russell's viper venom (Jesty *et al.*, 1974).

Activated Xa is capable of converting prothrombin to thrombin at a low rate; however, in the presence of phospholipid, ionic calcium, and factor V, the reaction rate is greatly increased and is even greater if thrombin-activated factor V is involved in the formation of the phospholipidized complex. The product of this interaction, which has protease activity, is referred to as prothrombinase, a term used to describe the prothrombin activator originally derived from the reaction of brain thromboplastin; factors V, VII, and X; and calcium ions (Owren *et al.*, 1954). The function of factor V in this complex will be discussed later.

Prothrombinase converts prothrombin to thrombin by cleavage of the prothrombin molecule. Thrombin is a highly specific enzyme for the rapid and effective transformation of fibrinogen, although it will split bonds in some other proteins. The end product of thrombin activity is fibrin. When formed in plasma, fibrin is a tough, insoluble, and fibrous protein equivalent to fibrin in the hemostatic clot. However, in a system of purified reactants, a fibrin is formed (fibrin S) which is soluble in monochloracetic acid, 5 *M* urea, etc. The tough, insoluble fibrin in the hemostatic clot (fibrin I) is the result of the stabilizing action of another proenzyme, factor XIII, which is converted by thrombin to factor XIIIa, which is a transamidase. Factor XIIIa produces the cross-linking of fibrin fibers that stabilize the clot. Ionic calcium is essential for the action of factor XIIIa; however, it is not an absolute requirement for the action of thrombin, although it will increase the rate of fibrin formation.

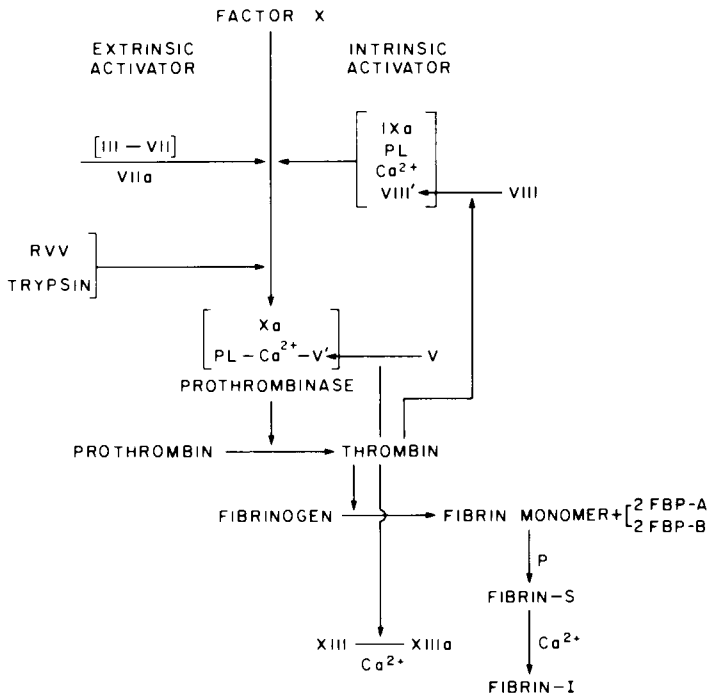


Fig. 3. The common pathway to thrombin formation and the fibrin clot. The final activators of intrinsic and extrinsic mechanisms that convert factor X to Xa (from Figs. 1 and 2) are included, in order to emphasize the multiactivating effect of thrombin and to illustrate the action of other agents on factor X, such as Russel's viper venom (RVV) and trypsin, which activate factor X in the same way as intrinsic and extrinsic activators. Factor Xa, ionic calcium, and thrombin-activated factor V form a phospholipid–micellar complex (prothrombinase), which converts prothrombin to thrombin. Thrombin then converts fibrinogen to soluble fibrin (fibrin S). Thrombin-activated factor XIII (XIIIa) stabilizes fibrin by cross-linking the fibrin polymers in fibrin S to form insoluble fibrin I, which is the form of fibrin in the hemostatic clot. Thrombin has other functions, in addition to its transformation of fibrinogen; it activates factors V, VIII, and XIII. Ionic calcium is required for prothrombinase activity, the activation of factor XIII, and the stabilization of fibrin. FBP-A and FBP-B are fibrinopeptides A and B, respectively, released by the action of thrombin on fibrinogen.

IV. BIOCHEMISTRY OF THE INTRINSIC BLOOD COAGULATION SYSTEM

Significant advances have been made in recent years in elucidating the mechanism of blood coagulation by the application of biochemical techniques to every phase of the mechanism. New factors uncovered by the clinician in patients with bleeding disorders, a new amino acid γ-carboxy-

glutamic acid that further explains the function of vitamin K-dependent proteins, the detailing of proteolytic cleavage of clotting factors during their activation and interactions, and the elucidation of the protein composition and structure of clotting factors all have provided new insights into blood coagulation on a molecular level. In the intrinsic system only factors V and VIII lag behind other factors in the evolution of their composition and function, primarily because they apparently exist as complex high-molecular-weight entities in the blood and are more difficult to purify.

A. The Contact System and Surface Activation Reactions

Exposure of citrated plasma to glass, kaolin, celite, collagen, and other foreign surfaces initiates a series of reactions known as the contact phase of coagulation in which factor XII is activated (Ratnoff, 1977a). Although new factors have been discovered that operate in the contact system, factor XII still holds a central position in that it appears to be the factor that is affected first when blood contacts a foreign surface. Lack of Hageman factor (XII), first discovered in a patient by that name, caused his blood to clot slower than normal when exposed to glass (Ratnoff and Colopy, 1955), even though the patient had no bleeding problems. Hemostasis, thus, is functionally normal in these individuals, but the defect becomes evident *in vitro* in the prolonged, activated partial thromboplastin and whole blood clotting time tests. These observations cast some doubt on the physiologic importance of factor XII in coagulation and have raised the question of other routes and factors involved in contact activation (Waaler, 1959; Schiffman and Lee, 1974, 1975).

Subsequently, the following intrinsic defects were discovered in subjects having normal factor XII levels, but abnormal clotting tests: Fletcher factor (Hathaway *et al.*, 1965), Flaujeac factor (Lacombe *et al.*, 1975), Williams factor (Colman *et al.*, 1975), and Fitzgerald factor (Saito *et al.*, 1975). The Fletcher factor proved to be prekallikrein (Wuepper, 1972); the other three were shown to be high-molecular-weight kininogens. Like factor XII, the new factors apparently are not essential for hemostasis, and these individuals are subject to thrombotic disease. The discovery, hereditary aspects, implications for hemostasis, and the interrelationship of all three contact factors to coagulation, fibrinolysis, kinin formation, and vascular permeability have been reviewed, and their implication in human disease states evaluated (Colman and Wong, 1977; Bennett, 1977, Ogston and Bennett, 1978).

The activation of contact factors in intrinsic clotting can result, also, in

the activation of factor VII by factor XII fragments, which are cleaved by kallikrein or plasmin, and thus may provide a concomitant *in vivo* activation of the extrinsic pathway (Radcliffe *et al.*, 1977) as well as a rapid source of incipient thrombin for the intrinsic activation of factors V and VIII (V′ and VIII′, respectively.) Such alternate pathways may provide the normal hemostasis observed in subjects deficient in one or the other of the contact factors. Thus, factor XII, prekallikrein and high-molecular-weight kininogen in the contact phase of coagulation have broad implicative actions, encompassing not only intrinsic clotting, but also other physiologic mechanisms.

In intrinsic clotting, it is factor XIIa that proceeds to activate the subsequent sequence of reactions by activating its substrate, factor XI. The first hemorrhagic deficiency of factor XI was reported in a patient by Rosenthal *et al.* (1953), and the activation of factor IX by XIa was demonstrated much later by Ratnoff and Davie (1962). Factor XI is transmitted as an autosomal recessive gene. The mechanism of the activation of factor XI to XIa has been the subject of numerous studies (see Davie and Hanahan, 1978; Davie and Fujikawa, 1975; Esnouf, 1977).

1. Hageman Factor (Factor XII)

The more recent purification procedures developed for factor XII have made it possible to define chemical properties more precisely, since the preparations are free or nearly free of the factor XII-activated molecule. Many methods of isolation have been devised in the past two decades with variable successes in purification and recovery, usually in yields of 1–25% (Schiffman *et al.*, 1960; Ratnoff and Davie, 1962; Schoenmaker *et al.*, 1965; Speer *et al.*, 1965; Kaplan and Austen, 1970; Komiya *et al.*, 1972; Movat and Ozge-Anwar, 1974; Revak *et al.*, 1974; Cochrane and Wuepper, 1971; Griffen and Cochrane, 1976; Fujikawa *et al.*, 1977a). Chromatography, gel filtration, adsorption, and precipitation techniques have been applied to human, bovine, and rabbit plasmas, and procedures, properties, and molecular mechanisms have been reviewed (Davie and Hanahan, 1978; Davie and Fujikawa, 1975). Spontaneous activation during isolation has been the principal problem which, in turn, has complicated evaluation of the contact mechanism.

Human and bovine factor XII are single polypeptide-chain glycoproteins with internal disulfide bridges. Molecular weights have been reported in the range of 80,000–147,000, but recently purified preparations appear to be 74,000 (bovine: Fujikawa *et al.*, 1977a) and 80,000 (human: Revak *et al.*, 1977). The carbohydrate content of 13.5–15% includes hexose, hexosamine, and neuraminic acid. Rabbit factor XII apparently has three polypeptide chains of 30,000 daltons held together by disulfide

bonds (Cochrane and Wuepper, 1971). The amino acid compositions of factor XII have been determined for human (Revak *et al.*, 1974) and bovine (Fujikawa *et al.*, 1977a) preparations and are quite, but not exactly, similar.

Factor XII isolated from human serum has a molecular weight of only 30,000–40,000, and possesses esterase and clotting activity (Kaplan and Austen, 1970); apparently it is a fragment of a larger molecule.

Bovine factor XII shows no esterase activity toward arginine esters, if it is isolated from plasma and solutions containing heparin, soybean trypsin inhibitor, or polybrene, which prevent spontaneous activation. The amino-acid sequencing of bovine nonactivated factor XII discloses an active site in the carboxyl-terminal region, homologous to that present in thrombin, factors XIa and Xa, trypsin, and plasmin. In the amino-terminal region, homology to serine protease inhibitors has been found. Thus, the inhibitor sequence may mask the esterase active site in nonactivated factor XII. It has been suggested that an unfolding of the amino terminal and of factor XII uncovers the active site in the carboxyl-terminal end when activated by adsorption to surfaces such as kaolin (Fujikawa *et al.*, 1977a).

Activated bovine factor XII also has been separated from nonactivated XII by benzamidine–Sepharose chromatography; the nonactivated molecule was shown to have a single chain with one internal disulfide, whereas the activated molecule was composed of a heavy and light chain held together by a disulfide bond. The amino terminal of the activated molecule was threonine and the carboxyl terminal was arginine (Fig. 4). Amino acid

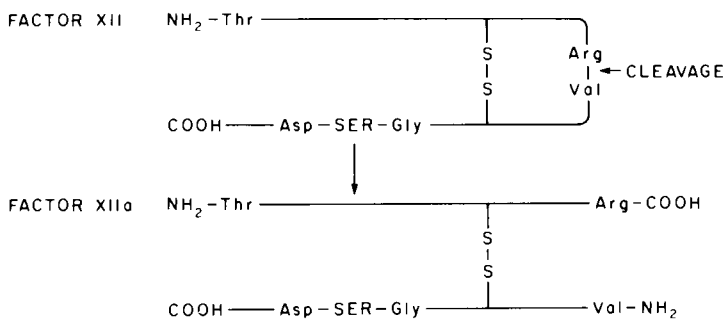


Fig. 4. Schematic depiction of the activation of factor XII by serine proteases, which cleave an Arg–Val bond, leaving a heavy chain (NH_2–Thr---Arg, MW 46,000) attached to a light chain (COOH---Val, MW 28,000) by a disulfide bond. The serine active site (SER) is shown in the light chain. Val in the light chain may form an ion pair with COOH of Asp at the active-site SER in a manner analogous to the activation of trypsinogen to trypsin. Adapted with permission from Fujikawa *et al.* (1977b), *Biochemistry* **16**, 4182. Copyright (1977) American Chemical Society.

sequencing indicated that an original valine bond is cleaved during activation on kaolin. The molecular weight was unchanged. The esterase and clotting activities of XIIa were inhibited by DFP and antithrombin III (Fujikawa *et al.*, 1977b). These observations have significant meaning for understanding contact activation reactions.

Equally significant in this respect have been the investigations of human factor XII (Revak and Cochrane, 1976; Revak, *et al.*, 1977), which was activated by exposure to glass or kaolin. Using immunologic and radiolabeling techniques, the single polypeptide chain of 80,000 daltons was cleaved enzymatically by kallikrein, plasmin, or factor Xla to fragments of 40,000, 12,000, and 28,000 MW and labeled c, d, and e, respectively. A 52,000-MW fragment, cd is cleaved initially. Fragmentation studies also have been reported by others, (see Bennett, 1977; Davie and Hanahan, 1978). The c fragment (40,000) and the cd (52,000) fragment bind to kaolin; the e fragment (28,000) does not bind, and actually is released from kaolin during activation, then bound, and inhibited by a plasma protein of 110,000 daltons (C-1 esterase inhibitor). The 28,000 XIIe fragment is probably what has been isolated as factor XII from human serum by Kaplan and Austen (1970). Three cleavage sites have been proposed (Ulevitch *et al.*, 1975; Revak *et al.*, 1977) to explain the formation of factor XII fragments, based on the presence of a disulfide bond linking cd and e fragments:

Cleavage site	Composition of fragments	Molecular weight	Cleavage rate
1	cd-s-s-e	80,000	Rapid
2	cd	52,000	Rapid
	e	28,000	
3	c	40,000	Slow
	d	12,000	
	e	28,000	

Thus, activation of factor XII by adsorption onto glass induces first a cleavage at site 1, without release of the fragment e with hydrolase activity; cleavage at site 2, corresponding to experimental conditions under which fragment e diffuses away from the adsorbent (and is inactivated by C-1 esterase inhibitor); cleavage at site 3, which occurs only after prolonged exposure to glass. These cleavages presumably occur as a result of the action of an unknown enzyme, or by plasmin, kallikrein, or factor Xla as secondary enzyme agents.

2. Prekallikrein (Fletcher Factor)

The amplification of the contact activation reactions of factor XII requires two additional factors: kallikrein and high-molecular-weight kininogen (HMWK).

Kallikrein is present in plasma as the proenzyme, prekallikrein, and was first recognized by its accelerating action in blood clotting tests (Wuepper and Cochrane, 1972). This serine protease was well known to biochemists and physiologists; its substrate is an α_2-globulin from which it releases bradykinin, a vasoactive peptide (Wuepper, 1972). As a proteolytic activator of factor XII, kallikrein appears to be 5–10 times more active than plasma or even factor XIIa. Clinically, it has been identified with Fletcher factor, a deficiency characterized by abnormal clotting, fibrinolysis, and kinin generation (Revak *et al.*, 1977). Such plasmas have prolonged partial thromboplastin times and poor kaolin-activated fibrinolytic activity.

Prekallikrein has been purified extensively from human and other plasma by a variety of chromatographic procedures (Nagasawa, 1968; Colman *et al.*, 1969; Takahashi *et al.*, 1972; McConnell and Mason, 1970; Wuepper and Cochrane, 1972; Mandle and Kaplan, 1977).

Prekallikrein is classified as a γ-globulin in complex formation with HMWK. It exists in human plasma in the molecular forms of 85,000 and 88,000 daltons. Activated factor XII (HF_f) cleaves both forms to kallikreins I and II by limited proteolysis at a single bond, forming two disulfide-linked chains: a heavy chain of 52,000 and a light chain of 33,000 or 36,000 daltons (Mandle and Kaplan, 1977). The serine active site is present in the light chains of kallikreins I and II. The kallikreins activate plasminogen to plasmin; their relationship to plasminogen proactivator and plasminogen activator has been discussed recently (Mandle and Kaplan, 1977).

Prekallikrein in bovine or rabbit plasma has a different pattern of cleavage, i.e., single bond cleavage, but with no change in molecular weight, producing a molecule with two disulfide-linked chains (Takahashi *et al.*, 1972; Johnston *et al.*, 1976). The substrates of these two enzymes are the same as that of human prekallikrein.

3. High-Molecular-Weight Kininogen (HMWK)

The third participant in contact-activated blood coagulation, and also in fibrinolysis, kinin formation, and vascular permeability is HMWK (bovine, 76,000 MW; human, 160,000 MW). Low-molecular-weight kininogen (LMWK) (bovine, 48,000 MW) does not participate in contact-me-

diated reactions; however, LMWK is a substrate of kallikrein for the release of vasokinins, and it is lacking in some HMWK-deficient plasmas (see Bennett, 1977).

Purified preparations of HMWK have been obtained from bovine (Komiya *et al.*, 1974) and human plasma (Griffin and Cochrane, 1976; Saito *et al.*, 1974), and as contact activation cofactor, CAC (Schiffman and Lee, 1975; Schiffman *et al.*, 1977). Isolation procedures typically involve ion-exchange chromatography and gel filtration. Bovine LMWK can be separated from HMWK by chromatography on CM cellulose (Komiya *et al.*, 1974).

Bovine HMWK is a glycoprotein with 12.6% carbohydrate (hexose 4.6%, hexosamine 3.7%, sialic acid 4.4%). Its amino acid composition has also been determined. It is a single polypeptide chain, having a carboxy-terminal leucine and a blocked amino-terminal residue: The vasoactive peptide, bradykinin, is in the intrachain disulfide loop. Plasma kallikrein cleaves HMWK into fragments 1 and 2; fragment 2 inhibits factor XII activation (Komiya *et al.*, 1974). The amino acid sequence of fragment 2 indicates that 41 residues are present; in fragment 1, which is a glycopeptide, there are 69 residues. HMWK is a histidine-rich polypeptide (Han *et al.*, 1976).

4. Plasma Thromboplastin Antecedent (Factor XI)

Factor XI is a plasma proenzyme with endopeptidase and esterase activity, developing after its activation by factor XIIa in a system with purified components, or by other proteases (Kingdon *et al.*, 1964; Fujikawa *et al.*, 1974a; Kurachi *et al.*, 1980).

Factor XI has been purified from human and bovine plasma after adsorbtion with barium sulfate, aluminum hydroxide, or tricalcium phosphate to remove prothrombin complex factors, and subsequently subjected to ion-exchange chromatography and ammonium sulfate precipitation and/or gel filtration (Ratnoff *et al.*, 1961; Schiffman *et al.*, 1963; Saito *et al.*, 1973; Heck and Kaplan, 1974; Koide *et al.*, 1977a). It also has been purified from cryoprecipitated plasma (Schiffman and Lee, 1974).

Human and bovine factor XI preparations are dimeric glycoproteins with reported molecular weights of 185,000, 160,000 and 124,000, containing similar polypeptide chains of 80,000 MW each, being held together by disulfide bonds. In the recent studies with a bovine preparation of 124,000 MW, the 11.1% carbohydrate was distributed as 5.4% hexose, 4.1% *N*-acetylhexosamine, and 1.0% *N*-acetylneuraminic acid in a product purified 28,000-fold, and obtained in 30% yield (Koide *et al.*, 1977a). Each polypeptide chain of the dimer was of 55,000 MW. The preparation was

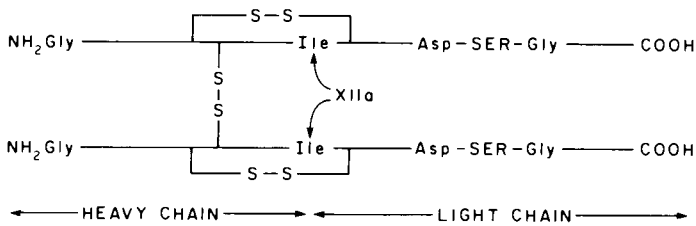


Fig. 5. Proposed structure of human factor XI. Two identical chains are linked by a disulfide bridge. After activation by XIIa, four chains are present in disulfide linkage; the exact number of disulfide bonds is unknown. Adapted with permission from Kurachi and Davie (1977), *Biochemistry* **16**, 5831. Copyright (1977) American Chemical Society.

unactivated; spontaneous activation had been depressed by the use of protease inhibitors and anticoagulants during its preparation. The specific activity was 494 units/mg protein; the plasma concentration was calculated to be approximately 2 μg/ml of plasma. It appears that the amino-terminal residues in the peptide chain are blocked. Amino acid sequence studies (Koide *et al.*, 1977b) indicate the presence of an active-site serine at position 195 and aspartic acid at 194, just as in factor XIIa, with an amino-terminal isoleucine residue. There appear to be two identical chains held together by disulfide bonds. After cleavage by factor XIIa, the factor XIa molecule is composed of four chains, bonded by an undetermined number of disulfide linkages (Davie *et al.*, 1979). A schematic presentation is shown in Fig. 5.

5. Activation of Factors XI and XII in The Contact System

Once plasma is exposed to a negatively charged surface, the contact reactions shown in Fig. 1 move through an intricate pattern of molecular interactions, involving surface adsorption and complex formation, release of active sites on proenzymes, and proteolytic cleavages. Unravelling these events requires the use of purified reagents in various combinations and the measurement of specific enzyme activities (by esterase activity, with immunologic techniques or by clotting assays). The contact reactions proceed in the absence of calcium ions.

There is now agreement that rapid proteolytic cleavage of factor XII takes place on the negative surface (Fujikawa *et al.*, 1977b; Revak *et al.*, 1977; Griffin and Cochrane, 1976), and that fragmentation of factor XII to a 28,000-MW derivative occurs. The 28,000-MW fragment is active only when attached to the surface, either as part of a larger fragment or perhaps as a free fragment. Once it diffuses away from the surface, it is rapidly inactivated by C'-1 esterase (Revak *et al.*, 1977). Significantly, factor XII also can be activated to a serine protease without change in molecular

TABLE III

Specific Sites of Action in Factor XII Reactions on the Surface of Kaolin[a]

Substrate	Activation reactants	Product	Action on:
Factor XI	XIIa/kaolin + HMWK	XIa	IX → IXa
PK	XIIa/kaolin + HMWK	Kallikrein	XII/kaolin
XII/kaolin	Kallikrein + HMWK	XIIa/kaolin	XI → XIa PK → kallikrein

[a] Based on reactions demonstrated by Griffen and Cochrane, 1976.

size (Fujikawa *et al.*, 1977b); cleavage to HF fragments must be due to the action of kallikrein in a secondary reaction, which is part of the feedback loop (Fujikawa *et al.*, 1980). Neither of these reactions takes place in the absence of HMWK. Evidence for the formation of a complex of activated factor XII and the histidine-rich, basic HMWK has been presented (Griffen and Cochrane, 1976). These investigators propose the sequence of reactions shown in Table III, which emphasizes the important role of the nonenzymatic HMWK cofactor. The reactions illustrate the catalytic effect of activated factor XII and kallikrein in the activation of factor XI to XIa, but beg the question of how the initial amount of XIIa becomes available for reaction 1. What actually happens on negatively charged surfaces is not clear, but it is evident that some kind of conformational change in factor XII takes place to provide incipient amounts of XIIa, presumably in the absence of any known enzymes and without any cleavage of peptide bonds.

B. The Factor IX and Factor VIII Complex

The next set of reactions initiated by factor XIa involve the conversion of factor IX to a serine protease (IXa) and the formation of a phospholipid micellar complex with calcium ions and factor VIII; this multimolecular complex enzymatically converts factor X to Xa. Although the biochemistry of factor IX has been explored extensively with much success, an equally extensive investigation of factor VIII has not unequivocally clarified the action of the antihemophilic factor.

1. Activation of Factor IX—The Antihemophilic B Factor

Factor IX is a clotting factor clinically identified as absent or nonfunctional in hemophilia B, which is an X-linked recessive trait characterized by spontaneous bleeding and second in severity only to classical hemo-

philia A. Named "Christmas factor" after study of a British patient by that name (Biggs *et al.*, 1952) or "plasma thromboplastin component," PTC (Aggeler *et al.*, 1952, after study of an American patient), factor IX contributed to the differentiation of hemophilia A and B, the blood of each having a prolonged clotting time, but a mixture of both producing a normal clotting time (Pavlovsky, 1947). Concentrates of human factor IX are available for treatment of Christmas disease (Aronson, 1977).

Factor IX is a member of the prothrombin complex group, which requires vitamin K for its biosynthesis by the liver, and which is readily adsorbed by aluminum hydroxide, barium sulfate, and tricalcium phosphate from serum or plasma. (Along with other vitamin K-dependent proteins, prothrombin and factors VII and X usually are contaminants in partially purified preparations). The plasma level of factor IX is about 510 μg/ml; the p*I* is 4–4.6, and the carbohydrate content is 20% (Aronson, 1977).

The initial step in purification is adsorption by and elution from these adsorbents, first accomplished with barium sulfate (Aggeler *et al.*, 1952). Numerous procedures developed since 1961 employing ion-exchange chromatography, gel filtration, and electrophoresis, have been summarized recently (Davie and Hanahan, 1978). Earlier preparations were of low yield and of purity to about 3000-fold. More recently, factor IX has been purified 22,300-fold from bovine plasma by DEAE-Sephadex and heparin–agarose column chromatography with 75% yield (Fujikawa *et al.*, 1973), and from human plasma, 10,000-fold with 38% yield by preparative polyacrylamide-gel electrophoresis and affinity chromatography on Sepharose-4B, having factor X and albumin antibody ligands (Østerud and Flengsrud, 1975); It has also been purified on heparin–Sepharose (DiScipio *et al.*, 1977; Anderson *et al.*, 1975; Vennerød *et al.*, 1977).

A comparison of the properties of bovine and human factor IX show significant differences [Davie and Hanahan, 1978 (bovine); Østerud and Flengsrud, 1975 (human)]. The molecular weight of human preparations is 70,000–72,000, and of bovine preparations, 55,400. For human preparations, the total carbohydrate content is 20.3% (hexose 11.7%, hexosamine 1.6%, sialic acid 3.1%, and galactose 3.9%). The corresponding value for a bovine preparation is 25.8% (10.6%, 6.5%, 8.7%, respectively). The p*I* values are 4.3–4.4 (human) or 3.7 (bovine), and the N-terminal NH_2 groups are glycine (Østerud and Flengsrud, 1975) or tyrosine (Andersson *et al.*, 1975; DiScipio *et al.*, 1977) for human preparations, and tyrosine for bovine preparations.

The amino acid composition of bovine factor IX has been determined (Davie and Hanahan, 1978). The NH_2-terminal region of bovine factor IX is homologous with prothrombin and the light chain of factor X (Davie and Fujikawa, 1975; Davie *et al.*, 1975a); six of the first 14 amino acids are

the same in all three proteins. Carboxyglutamic acid (Gla) is present in factor IX as in other prothrombin complex factors, and plays a role in the binding of calcium during coagulation (Stenflo and Suttie, 1977).

The proteolytic conversion of factor IX to IXa by factor XIa is a two-step process, requiring divalent ions (Davie and Fujikawa, 1975; DiScipio *et al.*, 1978; Lindquist *et al.*, 1978). Calcium is the physiologic ion in this reaction; strontium is less effective. The single polypeptide chain is first cleaved at a single bond, forming an inactive molecule with the same molecular weight (55,400) as factor IX, but consisting of a heavy and light chain held together by disulfide bonds. The N-terminal amino acid (tyro-

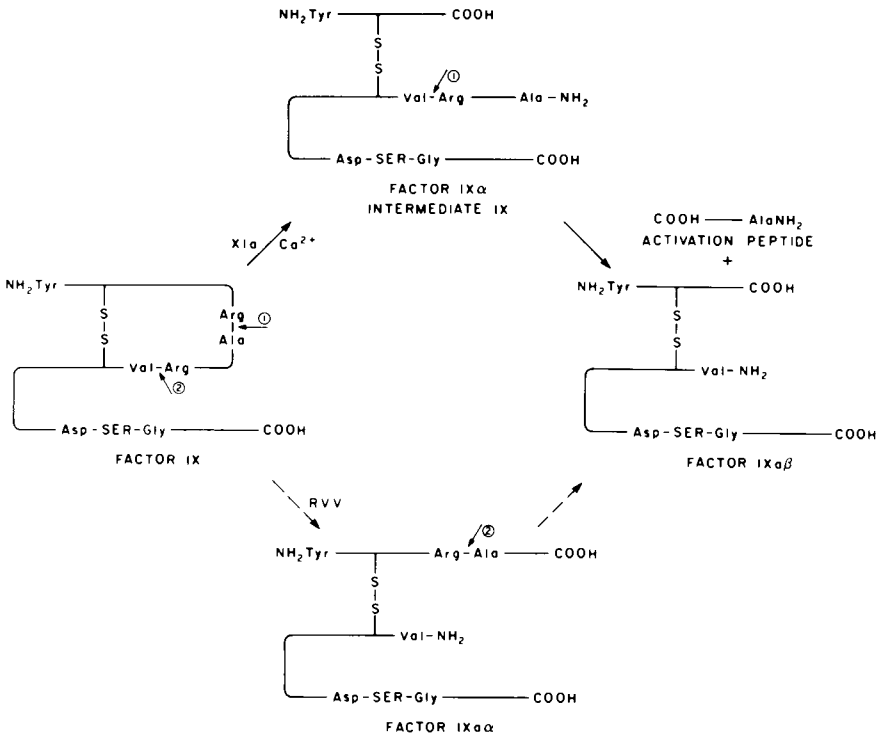


Fig. 6. Schematic depiction of the activation of factor IX to IXa through a two-step cleavage induced by factor XIa. ① An Ala–Arg bond is cleaved first to provide a two-chain polypeptide factor IXα or Intermediate IX, (MW 55,400) with an attached light chain (16,000 MW) and a heavy chain (38,800 MW). A second cleavage removes an activation peptide (19,000 MW), leaving an active-site serine (SER) in the heavy chain, now with a MW of 27,300. The intermediate IX is not an active serine protease; its site must be masked at this stage. This reaction sequence is indicated by solid arrows. ② The cleavage of factor IX by Russel's viper venom protease is indicated by dash-arrows in the pathway below. Adapted with permission from Lindquist *et al.* (1978).

sine) of factor IX is retained by the light chain, and a new one, alanine, appears on the heavy chain of this factor-IX intermediate (IXα). In the second step, an activation peptide of 9000 MW with 50% of the carbohydrate is cleaved from the amino-terminal end of the heavy chain at a Val-Arg bond to yield factor IXa (IXaβ) with a molecular weight of 46,500. Factor IXa, thus, is a two-chain disulfide-bonded molecule, having a light chain of 16,000 MW with an N-terminal tyrosine, and a heavy chain of 27,300 MW with valine as the N-terminal amino acid. Factor XIa, which is responsible for this cleavage, acts as an endopeptidase. Factors IX and XI are serine proteases, in as much as both have the amino acid 194-195-196 sequence of -Asp-SER-Gly- (Koide *et al.*, 1977b). Figure 6 illustrates the molecular change. Factor IX also can be activated by kallikrein and trypsin in the absence of calcium ions, but not by tissue factor and factor VII or by Russell's viper venom (Østerud *et al.*, 1975); however, other investigators find that factor IX is activated by a protease from Russel's viper venom and by factor Xa, but not by kallikrein, thrombin, or factor XIIa (Lindquist *et al.*, 1978).

2. *The Complexity of Factor VIII—The Antihemophilic A Factor (AHF)*

The fact that classic hemophilia, a sex-linked, severe bleeding disorder, is associated with a deficiency of a protein, antihemophilic globulin, was known four decades ago (Patek and Taylor, 1937). Since then, the development of knowledge regarding the structure and function of this protein has been slow, and only in recent years has the complexity of this factor been recognized. The slow progress is understandable in the light of the following observations: (1) the molecular weight of factor VIII is in excess of one million; (2) the molecule can be reduced to subunits or dissociated into high and low molecular components; (3) a high molecular component may be a second protein, the von Willebrand factor (VWF), which is an essential plasma cofactor for platelet aggregation and the absence of which in von Willebrand's disease subjects is the cause of the illness and of the prolonged bleeding time. Factor VIII coagulant activity (VIII:C) is measured by correction of the prolonged clotting time of hemophilic A plasma; VWF is measured by correction of bleeding time or, more readily, as a cofactor for ristocetin-induced platelet aggregation. Factor VIII also is detectable immunologically with antibodies to factor VIII, but further complications arise from the fact that AHF antigen and VWF both have the property of high molecular weight, and may well be the same protein. Furthermore, patients with hemophilia A synthesize a nonfunctional factor VIII molecule that reacts like AHF antigen immunologically (Poon and Ratnoff, 1977), whereas patients with von Willebrand's disease

are deficient in VWF and partly (50%) in factor VIII. The immunologically responsive, but factor VIII-inactive protein has been named "factor VIII-related antigen" (VIII:RAg or VIII:RAG), and is probably identical to VWF. The carbohydrate moiety of VIII:RAg may constitute the abnormality in von Willebrand's disease (Sodetz *et al.*, 1977). Excellent general reviews of this subject are available (Vermylen, 1975; Herschgold, 1975; Barrow and Graham, 1974; Bloom and Peake, 1977; Ratnoff, 1977a; Aledorf, 1975; Davie and Hanahan, 1978).

Factor VIII precipitates from plasma with fibrinogen in a number of procedures utilizing alcohol, salt, ether, or freezing temperatures (Pennell, 1960; Wickerhauser, 1971; Pool, 1964), polyethylene glycol, and glycine (Polson and Ruiz-Bravo, 1972). Differential polyethylene glycol precipitation with PEG-4000 and -6000 gives a high order of purification (Johnson *et al.*, 1971). Snake venom has been used to purify factor VIII (Green, 1971). Unfortunately, the complexity of the factor VIII molecule in terms of VIII:C, VIII:RAg, and VIII:VWF activities casts doubts on the purity of these preparations, and various chromatographic procedures have been employed to effect separation. The question as to whether factor VIII is one molecule, two molecules, or a molecular complex is a problem, which has not been resolved to everyone's satisfaction, as will be evident from the discussion below.

It is evident from one line of investigation that human and bovine factor VIII is a glycoprotein with a molecular weight greater than 1.1 million. There are differences in carbohydrate composition (human :6% carbohydrate with 2.2% hexose, 2.7% hexosamine, 0.9% neuraminic acid; bovine: 9% carbohydrate with 3.8%, 4.2% and 0.6% of the foregoing, respectively). Reducing agents (2-mercaptoethanol or dithiothreitol) produce similar subunits of molecular weight, approximately 200,000 all of which are held together by disulfide bonds. These preparations have VIII:C and VIII:RAg activities, which are interpreted from their biochemical behavior to be activities of a single large molecule (McKee *et al.*, 1975; Legaz *et al.*, 1975) since the factor VIII preparations were not dissociated in 1 *M* NaCl or 0.25 *M* $CaCl_2$. Activation of bovine factor VIII by thrombin did not change the molecular weight, although it did lead to enhancement and then subsequent instability of VIII:C activity. Proteolytic enzymes (plasmin, trypsin) enhance activity and cause degradation, the latter process curiously continuing even in the presence of serine protease inhibitors.

Other lines of study demonstrated that purified factor VIII can be dissociated at high ionic strength (1 *M* NaCl or 0.25 *M* $CaCl_2$) into high- and low-molecular-weight components. The low-molecular-weight components retained VIII:C activity; the high-molecular-weight components had VIII:RAg activity. Regeneration of the larger molecule occurred in

solution at low ionic strength. The disocciation and regeneration were effected with a cryoprecipitate obtained by Sepharose-4B or -6B gel filtration (Gordon and Shulman, 1975; Rick and Hoyer, 1978; Austen, 1974). On solid-phase ethylene maleic anhydride polyelectrolytes, factor VIII separated into individual molecules of 200,000 MW, which aggregated *in vitro* on cooling (Newman *et al.*, 1976). More recently, factor VIII:C with a molecular weight of 250,000–300,000 was obtained from bovine plasma in low yield by glycine precipitation, DEAE-Sephadex A 50 chromatography, gel filtration on Sephadex-SO_4 and G200, and elution from a Factor X–Sepharose column equilibrated with $CaCl_2$. The highly purified preparation had high VIII:C but no VIII:RAg activity in plasma, and was activated by limited proteolysis with factor Xa or activated protein C. On polyacrylamide gel electrophoresis, a triplet was observed, suggesting that subunits of 93,000, 88,000, and 85,000 MW exist in an associated state in the isolated preparation, with the 93,000 polypeptide having VIII:C activity (Vehar and Davie, 1980).

Still another line of investigation has demonstrated the separation of factor VIII into low- and high-molecular-weight components with VIII:C and VIII:RAg activities, respectively, with components aggregated together rather than separately (Sussman *et al.*, 1976), leading to the concept of a molecular complex in which VIII:RAg is a carrier protein for VIII:C. Other investigators have separated AHF and VWF activities from fresh human plasma as activities of separate molecules of 240,000 and 280,000 daltons, which aggregate at independent rates into larger molecules (Sodetz *et al.*, 1977). Significant for physiologic interpretation is the observation that 0.25 *M* $CaCl_2$ dissociated a low-molecular-weight factor VIII of 340,000 daltons, which does not aggregate in 0.002 *M* $CaCl_2$, and this might perhaps be the *in vivo* molecular size of factor VIII (Sussman and Weiss, 1976).

That factors VIII:C and VWF exist in plasma as separate but attached molecules in dynamic equilibrium is supported by affinity chromatography studies with an antibody to VWF (Koutts *et al.*, 1976). Even more interesting for the dissociation hypothesis is the observation that a certain type of collagen absorbs factor VIII:RAg and increases factor VIII:C activity. This dissociation may be a prerequisite for coagulant activity (Nyman, 1977).

The overall results of these and other investigations of factor VIII are still controversial. The complex relationships between VIII:C and VIII:RAg and the substructure of factor VIII need much elucidation on biochemical, chemical, and genetic levels (Ratnoff, 1977b; 1978). In most simple terms, factor VIII is a molecule(s) comprising two distinct properties of procoagulant activity and of von Willebrand factor activity, which

traverse two hemostatic mechanisms, namely blood coagulation and platelet agglutination, and therein resides the great importance of this protein in normal hemostasis. The uncertainties of chemical structure add to the difficulties of the interpretation of this complex physiologic system.

3. Role of Phospholipids and Calcium in Complex Formation

The interaction of factor IXa, calcium, phospholipid, and factor VIII is construed to be a reaction in which phospholipid molecules and calcium ions provide a chemically reactive micellar surface on which the above factors proteolytically cleave factor X to Xa.

The basic unit of the complex is a multimolecular bilayer or sphere, consisting primarily of phospholipid molecules oriented with hydrophobic tails inwardly, and hydrophylic, charged groups outwardly, the latter providing the negatively charged surface at the aqueous interface of biologic fluids. The proteolipids or lipoproteins of tissue cell membranes provide a similar activating surface, as in the structures of platelet cell surface membranes and internal, subcellular lipid granules and organelles, which contain phospholipids. Pure phospholipids such as phosphatidylethanolamine and/or phosphatidylserine function well in the complex of factors under consideration, but the true *in vivo* coagulation factor is platelet factor 3, a platelet membrane lipoprotein containing these phospholipids. Such membranes provide the surface for the activation of factor X and of prothrombin (Crawford and Taylor, 1977), located probably in or on the membrane (Joist *et al.*, 1974). The lipoprotein structure of these membranes provides a more intricate and effective system for these activation reactions.

Much less research has been conducted on the isolation of platelet factor 3 than on tissue factor III. Its purification has proven to be more difficult, and its biochemical behavior has not paralleled that of tissue factor III, which can be separated into lipid and protein moieties and recombined with retention of thromboplastic activity (see Section IIIA). Recent isolation studies demonstrate the lipoprotein nature of this factor and support the concept that protein(s) is essential for the full expression of coagulant activity of the lipid moiety. Platelet factor 3 activity is very labile, and its protein(s), once separated from the phospholipids, no longer produces coagulant activity upon recombination with lipid (Wu and McCoy, 1977).

4. The Factor X Intrinsic Activator Complex

The four reactants involved in the formation of the factor Xa intrinsic activator are factors IX and VIII, phospholipid, and calcium ions; all four are required for the rapid conversion of factor X to Xa (Davie *et al.*,

1975b; Varadi *et al.*, 1976). Of these, factor IX is converted by XIa to an intermediate that is enzymatically inactive; the intermediate, after the loss of its activation peptide, becomes the active serine protease, factor IXa (Fig. 6). Factor IXa alone can activate factor X to Xa, but the reaction is very slow and probably nonphysiological, since the conversion of the intermediate is a rate-limiting reaction. As indicated before (Section IV, B), the active site is in the heavy chain, whereas the light chain contains the vitamin K-dependent, calcium ion-binding sites (Gla residues), which are essential for complex formation with phospholipid.

The slow rate of cleavage of factor X by IX alone is increased in the presence of calcium ions and phospholipid; hence, factor IXa is the catalytic agent in this reaction. Factor VIII and VIII′ (thrombin-activated) have no enzymatic activity, but VIII′ will increase the rate of conversion of factor X several thousandfold. The function of factor VIII is considered to be that of a regulatory protein, a concept proposed for the one-molecule structural hypothesis (Legaz *et al.*, 1975). How thrombin or other proteolytic enzymes enter into physiologic clotting reactions to enhance factor VIII activity, and whether any feedback between factor IXa and factor VIII occurs, is not known. Neither is it clear whether the low-molecular-weight VIII:C activity or the intact molecule participate in complex formation. A recent report suggests that a serine protease is produced by the action of thrombin on bovine factor VIII, which is inactivated by diisopropylfluorophosphate (Vehar and Davie, 1977).

It is only possible to speculate that factor X undergoes conformational changes on the surface of the phospholipid micelle, which permits cleavage of a glycopeptide from its terminal end, and that in this reaction, factor Xa itself may be the enzyme that activates factor VIII, which, in turn, can function efficiently as a regulatory protein for the enzymic action of factor IXa (Davie *et al.*, 1975b). In this complex, positive divalent calcium ions must bind the reactants to the phospholipid surface by means of the reactants' negatively charged ionic groups. The final product of these reactions is the intrinsic factor-Xa activator.

V. BIOCHEMISTRY OF THE EXTRINSIC BLOOD COAGULATION SYSTEM

The extrinsic system requires a plasma factor (VII), calcium ions, and a tissue factor (III) for the activation of factor X to Xa. Trace quantities of factor III will increase markedly the rate of clotting in normal plasma, but not in factor VII-deficient plasma. It appears that factor VII circulates *in vivo* as a serine protease that does not attack its substrate, except in the presence of tissue factor. Tissue factor is not found in blood plasma.

A. Tissue Factor—Tissue Thromboplastin (Factor III)

Tissue factor is present in many tissues, but lung, brain, and placenta provide the most active product; there are other sources also (see Section III,C). Tissue factor has been shown to be a lipoprotein containing about 38–45% phospholipid; the protein and lipid moieties have been separated by several investigators (see Davie and Hanahan, 1978). The first reported separation of lipid and protein in human brain preparations (Deutsch *et al.*, 1964) demonstrated that the protein is insoluble and that the organic-solvent soluble fraction is a phospholipid. Recombining the fractions reproduced thromboplastic activity. Separated from its protein carrier, the phospholipid acts in test clotting systems only as a partial thromboplastin, and is much less active than tissue factor; the protein moiety is inactive. The protein moiety of tissue factor from lung has been extracted with deoxycholate, and is precipitated with 30–60% ammonium sulfate, treated with DEAE-Sephadex and gel filtered on agarose. Two major water-soluble proteins of 220,000 and 330,000 MW have been obtained (Nemerson and Pitlick, 1970; Pitlick and Nemerson, 1970). Upon recombination with phosphatidylcholine and phosphatidylethanolamine in the presence of deoxycholate, followed by dialysis, an unstable, but active complex is formed (Pitlick *et al.*, 1971). An endopeptidase activity is associated with the tissue factor apoprotein, but it is unrelated to coagulant activity (Nemerson and Pitlick, 1972). Tissue factor apoprotein, in its present state of purity, contains 18% carbohydrate comprising seven components, including sialic acid; these carbohydrates are responsible for adsorption of the protein to concanavalin-A Sepharose in the presence of deoxycholate (Pitlick, 1976).

B. Proconvertin (Factor VII)

The plasma concentration of factor VII is approximately 2 μg/ml, the lowest in concentration of the vitamin K-dependent factors (Aronson, 1977). Factor VII is a glycoprotein synthesized by the liver; the carbohydrate content for human preparations is 9% and for bovine, 13%. Its low concentration calls for special chromatographic procedures for effective purification.

The crude prothrombin complex is first separated from plasma or serum by adsorption to $BaSO_4$, $Al_2(OH)_3$, or barium citrate, and is desorbed by citrate or oxalate ions. Further purification is accomplished usually with ion-exchange chromatography or gel filtration (see Davie and Hanahan, 1978). More recently, affinity chromatography on benzamidine–Sepharose enabled purification of 148,000-fold to be achieved in 10% yield

from bovine plasma. The molecular weight found was 55,000 (Jesty and Nemerson, 1974). A 200,000-fold purification of bovine factor VII was obtained when benzamidine was used as a protease inhibitor during isolation (Radcliffe and Nemerson, 1975). The molecular weight was reduced to 37,000 by 2-mercaptoethanol. A 500,000-fold purification of bovine factor VII (Kisiel and Davie, 1975) resulted in a single-chain polypeptide with an amino-terminal alanine residue. The 13% carbohydrate consisted of 14.1% hexose, 6.2% *N*-acetylglucosamine, and 7.9% *N*-acetylneuraminic acid. Human factor VII from serum or plasma, purified 60,000-fold, has a 60,000 MW (Laake and Ellingsen, 1974). The enzymatic site in factor VII is not blocked, and is readily inhibited by DFP (diisopropylfluorophosphate). Factor VII circulates in plasma as a serine protease (Jesty and Nemerson, 1974).

C. The Extrinsic Factor X Activator

Factors III and VII interact to form an extrinsic factor-X activator. Experimental evidence has been presented, however, in support of two different ways by which factor VII becomes involved. In one, a complex of factors III and VII is claimed (Nemerson and Pitlick, 1970); according to the other, factor VII is activated directly (Østerud *et al.*, 1972).

In the first hypothesis, factor VII is unable to activate factor X until it forms a complex with tissue factor and calcium ions. Complex formation between III and VIII is theoretically dependent upon interaction of calcium ions and Gla residues. Although Gla residues have not been reported in factor VII, the amino-acid residue homology with the vitamin K-dependent proteins (II, IX, and X) is highly suggestive of this (Stenflo and Suttie, 1977; Jesty *et al.*, 1975). When complex formation does occur, the serine active site in bovine factor VII becomes accessible. It has been shown that the single-chain factor-VII protein can be converted to a two-chain, disulfide-bonded protein by cleavage of an Arg-Lys bond in the presence of phospholipid, calcium, and catalytic amounts of factor Xa (Fig. 7); other enzymes also can activate VII, e.g., XIIE fragment, kallikrein, IXa, and plasmin (Radcliffe and Nemerson, 1975, 1976; Kisiel *et al.*, 1977).

The direct activation route involves the activation of factor VII to VIIa by tissue factor III, and the conversion of factor X to Xa by factor VIIa (Østerud *et al.*, 1972). After reaction with factor III, the activated form of factor VII has been isolated and shown to activate factor X in the absence of factor III. Whether various other enzymes, which are capable of a similar activation of factor VII, function physiologically in this manner is not clear.

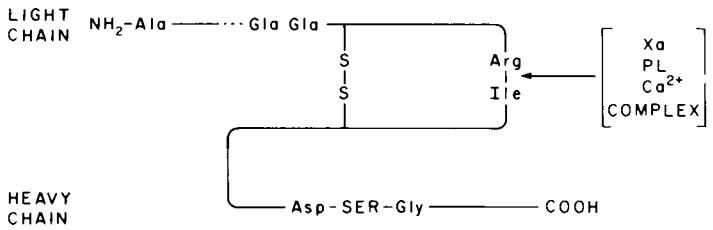


Fig. 7. Schematic illustration of activation of factor VII to VIIa by phospholipid, ionic calcium, and catalytic amounts of factor Xa. Location of Gla residues is arbitrary, but found in the light chain. Adapted with permission from Kisiel *et al.* (1977) *Biochemistry* **16,** 4189. Copyright (1977) American Chemical Society.

Obviously, a clear choice of mechanisms cannot yet be made, and the reactions illustrated in Fig. 2, depicting these choices, need further study.

VI. THE COMMON PATHWAY OF BLOOD COAGULATION

The common pathway begins with the activation of factor X by the factor-X activators of the intrinsic or extrinsic systems. Although different in composition and source, their action on factor X is, broadly speaking, the same.

A. Biochemistry of Factor X and Its Cleavage by Intrinsic and Extrinsic Activators

Factor X is a precursor primarily responsible for the enzymic conversion of prothrombin to thrombin. The plasma level is 20 μg/ml (Aronson, 1977). Two forms are known: factors X_1 and X_2. A number of investigators have prepared factor X (Duckert, 1964; Esnouf and Williams, 1962; Papahadjopoulus *et al.*, 1964; Jackson and Hanahan, 1968; Fujikawa *et al.*, 1972a; Kisiel and Hanahan, 1973; Yue and Gertler, 1977; Aronson *et al.*, 1969; DiScipio *et al.*, 1977). Its preparation and properties have been reviewed by Davie and Hanahan (1978). A glycoprotein with 10% carbohydrate (2.9% hexose, 3.6% hexosamine, 3.8% neuraminic acid), it has a molecular weight of 55,100, of which a heavy chain accounts for 35,000 and a light chain for 17,000; both chains are joined by disulfide bond(s). The heavy chain is the zymogenic moiety, for it contains the serine-active enzymatic site and the Arg-51–Ile-52 peptide bond, which is split when factor X is activated. Also, in the intact heavy chain, an amino-terminal residue has been identified as tryptophan, and almost all of the carbohydrate is attached to this glycopeptide (Fujikawa *et al.*, 1972b; Jackson,

1972). No difference was found between factors X_1 and X_2, while the earlier preparation showed a small difference in carbohydrate content. The light chain contains essentially no carbohydrate, its total amino acid sequence has been determined, and it has an alanine N-terminal residue (Enfield *et al.*, 1975). The function of the light chain may be to define substrate specificity (Davie and Hanhan, 1978).

The activation of bovine factor X by extrinsic or intrinsic activators, by proteolytic enzymes of Russel's viper venom, or by trypsin involve the cleavage of an Arg-51–Ile-52 peptide bond in the heavy chain (Radcliffe and Barton, 1973; Morris *et al.*, 1978; Lindhout *et al.*, 1978; Fujikawa *et al.*, 1974b; Davie and Hanahan, 1978) (Fig. 8). An activation glycopeptide of 10,500 daltons is released from the amino-terminal end of the heavy chain by these activators; the remainder of the molecule is factor Xaα or α-Xa (Fujikawa *et al.*, 1972b, 1975; Jesty *et al.*, 1974). It now has a molecular weight of 45,300, and Ile-52 amino-terminal and Leu-307 carboxy-terminal residues. The molecule is folded into two halves, such that residues His-93 and Asp-138 form one loop, and Asp-232 and Ser-233

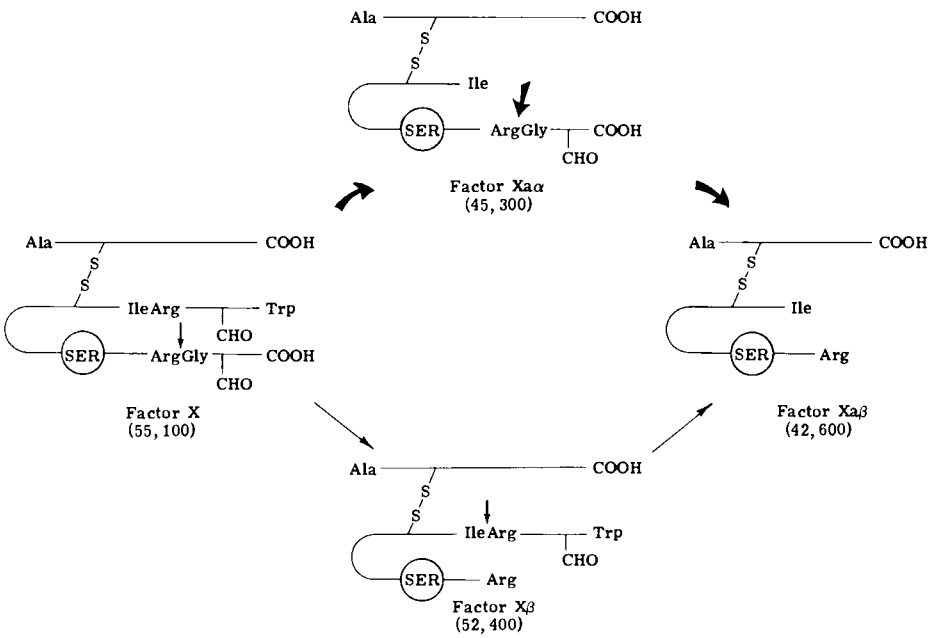


Fig. 8. Mechanism of activation of bovine factor X. The Arg–Ile bond is cleaved first to form factor Xaα, followed by a split at the Arg–Gly bond to give factor Xaβ. Both peptides that are released contain the carbohydrate residues. The Arg–Ile bond is cleaved in physiologic clotting; cleavage of the Arg–Gly bond is a degradative reaction. Reproduced with permission from Davie and Hanahan, 1978.

form the second loop. These residues form a charge–relay network characteristic of pancreatic enzymes. A diagrammatic representation of the amino acid sequence of the heavy chain of factor Xaα(α-Xa) illustrating these properties is shown in Fig. 9, according to Davie and Hanahan (1978).

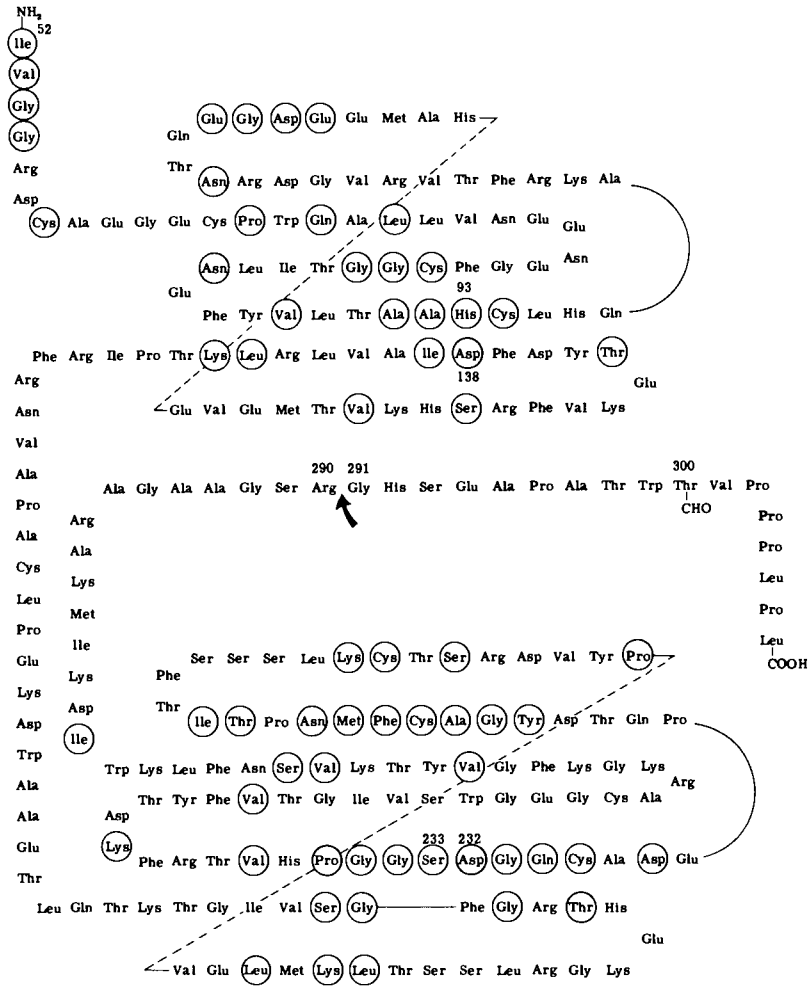


Fig. 9. The amino acid sequence of the bovine heavy chain of factor Xaα. The circled residues represent those found in trypsin. The degradative peptide is cleaved at Arg 290–Gly 291. The histidine–aspartic (93–138) loop is shown in the top half of the sequence; the serine loop (232–233), in the bottom half. The entire molecule is folded into two halves. Reproduced with permission from Davie and Hanahan, 1978.

A second glycopeptide also is released from factor Xaα(α-Xa) as a degradation product (Fig. 8); the degradation peptide, factor Xaβ, arises from the bond split between Arg-290 and Gly-291 in the carboxy terminal of the heavy chain. Factor Xaβ(β-Xa) is carbohydrate-free, and its coagulant activity is the same as that of α-Xa. The carbohydrate in the degradation peptide is attached to Thr-300 (Davie and Hanahan, 1978).

Bovine factor X also can be degraded to an enzymatically inactive protein, factor Xβ (β–X) by cleavage of factor X only at the Arg-290–Gly-291 bond (Fig. 8). Cleavage of the activation peptide from the amino-terminal end of the heavy chain results in the formation of factor Xaβ (β–Xa) (Davie and Hanahan, 1978). The degradation glycopeptides are feedback hydrolysis products of the action of factor Xa in the presence of phospholipid and calcium, which complicate the activation reaction (Jesty *et al.*, 1974). The chemistry and enzymology of bovine factor X have been reviewed by Henriksen and Jackson (1975) and Jackson (1977).

Further elucidation of the mechanism of activation of factor X by factors VII and III in the extrinsic system has been provided by kinetic studies with a new assay, which measures the release of a radiolabeled activation peptide resulting from the hydrolysis of the Arg-51–Ile-52 bond in the presence of the inhibitor, benzamidine. Cleavage of factor VII by Xa at the Arg-Ile bond yields α-VIIa, with an increase in factor VII coagulant activity, which, in turn, causes the hydrolysis of more factor X to Xa. The individual reactions are envisioned as follows:

1. Factor X to Xa by factors VII and III
2. Factor VII to α-VIIa by factor Xa and phospholipid
3. Factor X to Xa by α-VIIa and phospholipid
4. Factor α-VIIa to β-VIIa (inactive) by factor Xa and phospholipid

The exact order is not clearly established, but the observations support the concept of a significant self-regulating feedback mechanism. These important interrelated reactions between factors III, VII, and X in the presence of phospholipid have been described by Silverberg, Nemerson, and Zur (1977), and may have some bearing on the physiologic regulation of factor Xa formation.

B. Biochemistry of Factor V

Factor V is a high-molecular-weight plasma protein that functions as a cofactor for factor Xa, along with calcium ions and phospholipid, the latter providing a matrix for a micellar complex which converts prothrombin to thrombin. The role of factor V is much like that of factor VIII, in that it is part of a micellar complex that produces an enzymatic surface for

activating a clotting factor, yet it, itself, is nonenzymatic, but does acquire enhanced activity after exposure to thrombin (factor V′ or Va). Most investigations have been made with factor V isolated from bovine plasma because of the great instability of human factor V. Excellent reviews of factor V are available (Davie and Hanahan, 1978; Colman, 1976).

The partial purification of factor V was reported in 1947, then still referred to as "accelerator globulin," AcG (Ware and Seegers, 1948; Owren, 1948). Numerous procedures since then have produced inhomogeneous and unstable preparations (see Davie and Fujikawa, 1975; Colman, 1976; Davie and Hanahan, 1978; Nesheim *et al.*, 1979). Procedures utilizing highly charged reagents alter factor V.

More recent preparations, in which such treatment was avoided, produced high purity, good yield, and stability of bovine factor V (Esmon, 1979). Adsorption on $BaSO_4$ from DFP-treated plasma, extraction with QAE-cellulose, polyethylene glycol (PEG) precipitation, and chromatography on desulfated Sepharose-6B gave a preparation (1000–2000-fold purified) with a molecular weight of 800,000 to 1,000,000, stable at 4°C for one week and at −20°C for several months. To prevent protease degradation, DFP was necessary. Factor V is a glycoprotein with 8.7–21% carbohydrate; purified bovine factor V has been found to have 18% carbohydrate and to contain galactose, mannose, glucosamine, and sialic acid. Factor V is considerably more active (7- to 10-fold) after treatment with thrombin, factor Xa, Russel's viper venom, or chymotrypsin, an enhancement observed by many investigators and clearly demonstrated with the highly purified factor. The activated form of factor V has a lower molecular weight (240,000), and probably does not exist in this form in the plasma, but becomes active on the phospholipid–calcium–factor Xa micelle (Nesheim and Mann, 1979; Bartlett *et al.*, 1980).

C. The Factor Xa–Factor V Catalytic Complex (Prothrombinase)

Factors Xa and Va alone do not form a complex (Smith and Hanahan, 1976); phospholipid and calcium ions are necessary to form the micellar complex with enzymatic properties. The light chain of factor Xa contains Gla residues (Nelsestuen *et al.*, 1977) that bind to phospholipid through divalent calcium ion interactions (Henriksen and Jackson, 1975) on the micellar surface, based on evidence of positive cooperativity of calcium-binding isotherms. Thus, the negatively charged Gla residues of the light chain play a critical role in holding factor Xa close to the negatively charged surface of the phospholipid micelle by the positive electrostatic forces of calcium ions.

The rate of conversion of prothrombin to thrombin by factor Xa, calcium, and phospholipid is slow, but is markedly increased by factor Va. A mixture of these four reagents provides a powerful enzymatic complex for the rapid conversion of prothrombin, and has been called "prothrombinase" (Owren *et al.*, 1954). Factor Xa functions as the serine protease responsible for cleavage of specific peptide bonds in prothrombin, but the total complex is required for rapid conversion. While the role of factor V is this conversion is that of a cofactor or regulatory protein, which enhances prothrombinase activity considerably, its mode of action when bound to the micellar surface is not clearly understood. Presumably, it provides a more optimal orientation of factor Xa and prothrombin (Davie and Fujikawa, 1975), and this may be effected by hydrophobic binding of factor Va to phospholipid and/or to prothrombin (Hemker *et al.*, 1977). Finally, factor V binds to fragment 2 of prothrombin (Esmon and Jackson, 1974), and may protect fragment 2 from inhibition by fragment 1 of prothrombin, which is released by thrombin during the incubation of prothrombin with factor Xa and calcium, thereby accelerating the rate of thrombin formation (Prowse *et al.*, 1976). It is believed that the complex, phospholipid–calcium–factor Xa–factor V (prothrombinase), is the physiologic prothrombin converter.

D. Chemistry of Prothrombin

The concept of a precursor of thrombin (prothrombin) dates back to the work of Buchanan and of Schmidt (see Section I). Numerous preparations of prothrombin have been made since then. Only recently, however, has the chemistry of this protein molecule been explored in depth. The in-depth studies of prothrombin really originated with Seegers about 1940, when prothrombin chemistry was attacked on a purely biochemical basis, and from which the autoprothrombin-derivative theory of blood clotting evolved (see Section III). These developments have been reviewed recently (Seegers *et al.*, 1975). Although the terminology for prothrombin, its fragments, and related factors that evolved biochemically differs in part from that evolved on clinical grounds (factor terminology), both types of investigations substantially paralleled each other in the final outcome. In the last decade, the chemistry and physiology of prothrombin have undergone explosive growth; these events have been thoroughly reviewed quite recently (Shapiro and McCord, 1978). Also a number of reviews on various aspects of prothrombin and its activation have appeared (Davie and Hanahan, 1978; Magnusson *et al.*, 1975c; David and Fujikawa, 1975; Mann, 1976; Suttie and Jackson, 1977; Stenflo, 1978; Seegers, 1967). A critical analysis of structure–function relationships in prothrom-

bin activation, in which prothrombin and factor X are compared, has been made (Jackson, 1978).

Prothrombin has been prepared from the blood of various mammalian species, but the most-studied preparations are of human and bovine origin. Generally, the initial step in isolation is adsorption on insoluble salt [$BaSO_4$, $Al(OH)_3$, barium citrate], followed by elution with sodium citrate and removal of barium ions, e.g., by dialysis against ethylenediaminetetraacetic acid. Stabilization against spontaneous activation to thrombin is essential with DFP, phenylmethylsulfonyl chloride, or protease inhibitors. Various chromatographic procedures have been applied for the purification of bovine and human prothrombin (Seegers, 1940; Cox and Hanahan, 1970; Owen *et al.*, 1974; Wallin and Prydz, 1975; Bajaj and Mann, 1973; Aronson *et al.*, 1969; Kisiel and Hanahan, 1973). It is essential to remove traces of factors VII, IX, and X, if kinetic or activation studies are to be meaningful. A chromatographic procedure for separating human factors II, IX, and X from each other in high yield has been developed recently (DiScipio *et al.*, 1977). The plasma level of this vitamin K-dependent protein in human blood is 100 μg/ml (Aronson, 1977).

Bovine and human proteins are glycoproteins containing 7–8% carbohydrate (mannose, galactose, glucosamine, galactosamine, and sialic acid constituents). The carbohydrates are present as three individual chains, each linked to an asparagine residue. Sialic acid is not functional since desialated prothrombin retains its clotting activity and calcium-binding properties. Calcium binding mediates the interaction of prothrombin with phospholipid; the binding of calcium by prothrombin exhibits cooperativity (about 10 moles of calcium are bound per mole of protein). The binding of calcium is quite specific, being limited to one region of the molecule, namely, fragment 1, which is nonthrombogenic, and is in the amino-terminal end of the molecule. Calcium binding is a function of Gla residues on fragment 1, not of sialic acid. The molecular weight of prothrombin is 68,000–75,000 (bovine) and 72,000 (human). The molecule is a single, asymetric polypeptide chain with little or no helical content. The amino-acid composition of bovine and human prothrombin has been determined. The amino-terminal residue in both bovine and human prothrombin is alanine; the carboxy terminal in human prothrombin is glutamic acid, and in bovine prothrombin it is serine (for references to original observations, see Davie and Hanahan, 1978; Seegers *et al.*, 1975; Shapiro and McCord, 1978).

The complete amino acid sequence has been established for bovine prothrombin (Magnusson *et al.*, 1975c; Reuterby *et al.*, 1974; Hewett-Emmett *et al.*, 1975) and for human prothrombin (Butkowski *et al.*, 1977; Walz *et al.*, 1977a, b). The amino acid sequences of both prothrombins

are quite similar. The sequence homology with factor X and related factors shows a close structural relationship in certain important portions of the molecules (DiScipio *et al.*, 1977).

1. Proteolytic Cleavage Pathway to Thrombin

The structure of prothrombin becomes evident from its activation to thrombin, during which various fragments of the molecule are cleaved, the sequence and number of fragments being dependent upon the proteolytic agent, factor Xa or thrombin, and the conditions of proteolysis (see Seegers *et al.*, 1975; Davie and Hanahan, 1978; Shapiro and McCord, 1978; Jackson, 1978). The various products and bond cleavages of human prothrombin are illustrated in Fig. 10 (according to Shapiro and McCord, 1978), in which the terminology is that adopted by the International Committee on Thrombosis and Hemostasis (Myrmel *et al.*, 1976). Since other fragmentation terminologies are found in the literature, Table IV compares the trivial names of two other systems. The terminology applies to both human and bovine prothrombin; however, slight differences are to be noted. The numbering of the amino acids in bovine prothrombin is not exactly identical to human prothrombin, due to the presence of an extra amino acid in the N-terminal region of the bovine protein. Thus, bovine prothrombin cleavage points are Arg-155–Ser-156 and Arg-323–Ile-324 with factor Xa and Arg-274–Thr-275 with thrombin. The Arg-286–Thr-287 bond, which is cleaved by thrombin in human prothrombin, is not

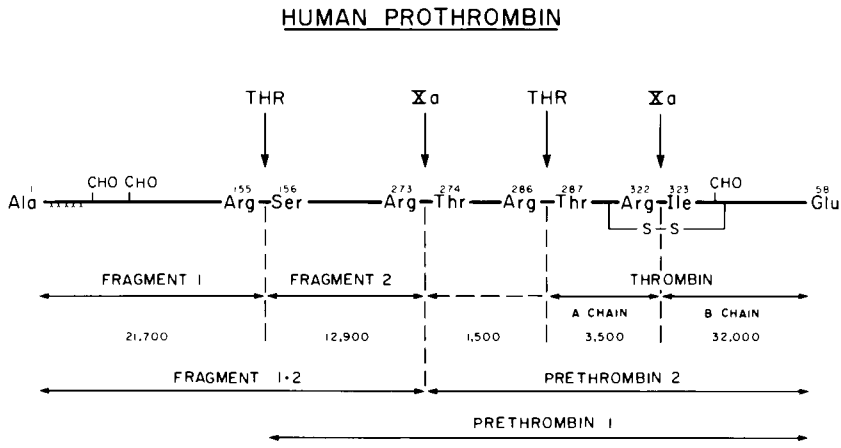


Fig. 10. Cleavage points and products in human prothrombin illustrating terminology adopted by the International Committee on Thrombosis and Hemostasis. Reproduced with permission from Shapiro and McCord (1978). *Prog. Hemostasis Thromb.* **4**, 177, Grune and Stratton.

TABLE IV

Proteolysis Products of Prothrombin: Terminology

International[a]	Derivative theory[b]	Intermediate[c]
Prethrombin 1	Prethrombin	Intermediate 1
Prethrombin 2	Prethrombin E	Intermediate 2
Fragment 1	PR fragment	Fragment 1[d]
Fragment 2	0 fragment	Fragment 2[e]
Fragment 1 · 2	PRO chain, piece	Fragment 1 · 2
Alpha thrombin	Thrombin	Alpha thrombin
Thrombin-A chain	Thrombin-A chain	
Thrombin-B chain	Thrombin-B chain	

[a] Myrmel *et al.*, 1976; [b] Seegers *et al.*, 1975; [c] Davie and Hanahan, 1978; [d] intermediate 3; [e] intermediate 4.

cleaved in the bovine protein. These cleavage points demarcate the various domains of the molecule (Mann, 1976). Another perspective of the bovine prothrombin molecule is shown in Fig. 11 (according to Jackson, 1978), in which the structures of bovine prothrombin and factor X are compared, and the essential chemical features of the molecules are presented. Three primary domains are shown for bovine prothrombin: fragment 1, fragment 2, and prethrombin 2, which becomes thrombin after cleavage.

Fragment 1, the first activation product, contains carbohydrate residues, internal disulfides which define "kringles," and 10 Gla residues. Fragment 1 is bound to phospholipid by divalent calcium ions that react with its Gla residues and the negatively charged ionic surface of phospholipid, which is part of the prothrombinase complex. In this manner, fragment 1 is anchored to the micelle and performs an essential function in the activation of prothrombin. There are 10 to 12 calcium-binding sites involved. Of these, the first two to four calcium ions bound induce a conformational change in fragment 1, which leads to phospholipid binding, involving the remainder of calcium-binding sites. The phospholipid–prothrombin complex contains 8 to 10 phospholipid residues per molecule of prothrombin if the micelle is prepared from pure phosphatidylserine. The platelet lipid surface membrane or lipid granule provides a similar but, undoubtedly, a more complicated interface for reaction with fragment 1. These calcium-binding sites are in the first 33 amino acids of fragment 1 (Magnusson *et al.*, 1974; Stenflo and Suttie, 1977). The remainder of fragment 1 contains disulfide bonds which stricture the polypeptide chain into loops or kringles, and which are assumed to aid in the binding of two factor Xa molecules to one prothrombin molecule (Magnusson *et al.*, 1975c).

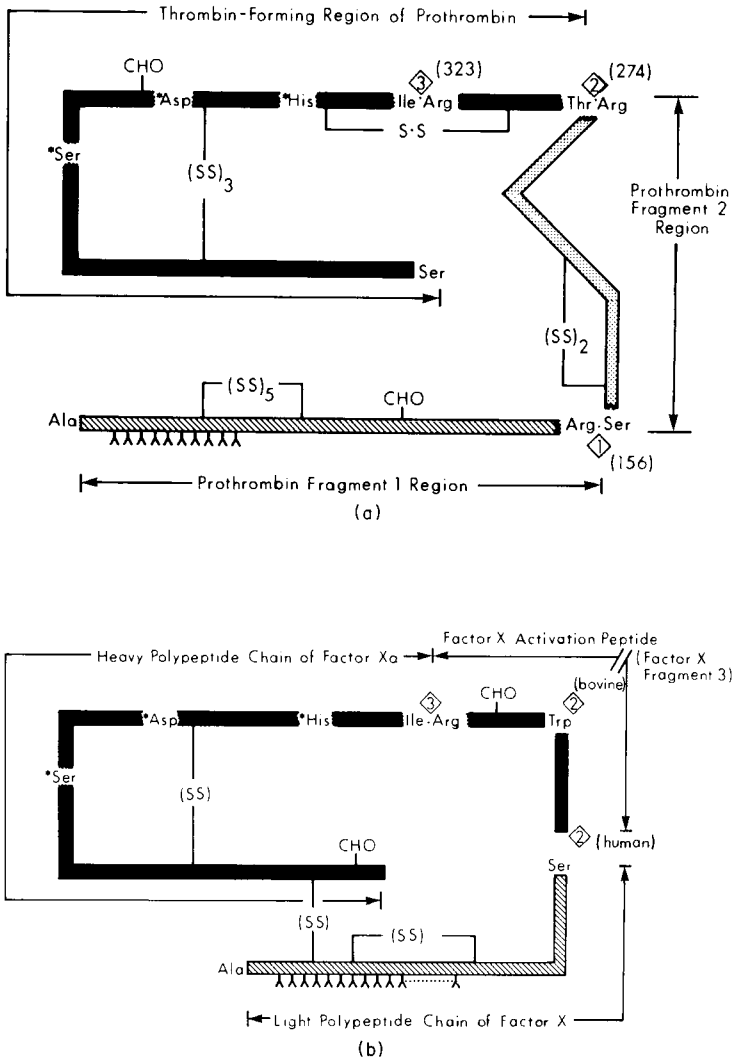


Fig. 11. Proposed structure of the bovine prothrombin molecule in (a) and of the factor X molecule in (b). Structure (a) illustrates the three prothrombin fragments: prothrombin fragment 1 (hatched bar), prothrombin fragment 2 (stippled), and thrombin-forming fragment (solid). Numbers in ( ) identify the amino acid residue. The active serine center is designated by * Ser. The inverted Ys (ʌ) represent Gla residues present in profragment 1; CHO represents carbohydrate residues. Structure (b) illustrates the positions of the light and heavy chains and of the activation peptide. The light chain (hatched) is similar to prothrombin fragment 1, but only in the presence of Gla residues. The heavy chain (solid) contains the active serine (*Ser) and is similar in structure to thrombin. Factor Xa is derived by cleavage of the activation peptide, which in bovine X is shorter (Arg ③ to bovine ②) than that in human factor X (Arg ③ to human ②). Reproduced with permission from Jackson, 1978.

Fragment 2 has the capacity to bind factor V (Esmon and Jackson, 1974). Nonactivated factor V participates in prothrombin activation at a slower rate than thrombin-activated factor V, which has a 300-fold accelerating effect on the rate of reaction. Factor Va binds to prothrombin and prethrombin 1, which contains fragment 2, but not to prethrombin 2, the immediate percursor of thrombin. The binding requires four to five calcium ions, but these weak binding sites in fragment 2 are not Gla residues. Thus, when factor V is activated by thrombin (or other enzymes), it acquires the capability of binding to the fragment 2 moiety of prothrombin. In this capacity, it is a receptor for prothrombin (and for factor X) in the formation of the prothrombinase complex. *In vivo,* it must circulate as a nonfunctional factor until it is activated by thrombin. The source of the incipient thrombin formation essential for this reaction must be factor Xa-derived. Factor Va, apparently, is responsible for orienting factors II and Xa, which are held by calcium ion bonds to the phospholipid-micelle surfaces, in such a manner that factor Xa can induce effective proteolysis of prothrombin. Since factor V binds to platelets, this is a potential mechanism for physiologic coagulation reactions. A bonding mechanism visualized for these interactions has been proposed (Jackson, 1978).

Prethrombin 2 is converted to α-thrombin by factor Xa–calcium with the single cleavage of bond Arg-322–Ile-323, leaving a thrombin with a short A chain attached to a long B chain by a disulfide bond. This reaction is not accelerated by factor V or phospholipid. Bovine fragment 2 binds to bovine prethrombin 2, restoring the factor V acceleration of thrombin formation; the physiologic significance of these reactions is not clear since the human prothrombin counterparts do not function in this manner. The carbohydrate present in fragment 1 and in prethrombin 2 (β chain of thrombin) apparently has no functional capacity as indicated above (Shaprio and McCord, 1978).

In summary, the chemistry of prothrombin and its activation has been gleaned from *in vitro* experimental systems made possible by modern biochemical techniques. The sequence of probable events may be construed to be as follows: Factor Xa, prothrombin, factor V, and calcium ions form a complex in which factor V organizes the enzyme and substrate on a phospholipid–platelet micellar surface on which calcium ions are binding agents for the enzyme and substrate. Binding via calcium occurs through the light chain of factor Xa and fragment 1 of prothrombin. Then, one molecule of factor Xa cleaves the Arg-273–Thr-274 peptide bond to release the nonthrombogenic fragments 1 and 2, and in prethrombin 2 another molecule of Xa cleaves the Arg-322–Ile-323 bond to form a two-chain polypeptide, α-thrombin. In a feedback mechanism thrombin can attack prothrombin itself at Arg-155–Ser-156 and at Arg-286–Thr-287,

the latter reaction shortening the A chain by 13 residues in the human system. The thrombin formed by factor Xa activates factor V to Va, and induces autocatalytic formation of larger amounts of thrombin. The bulk of the thrombin now diffuses away from the surface and a new prothrombin molecule takes its place, being subject to rapid peptide cleavages. The basic reactions involved can be expressed as follows:

$$\text{Prothrombin} \longrightarrow [\text{prethrombin 2–fragment 1} \cdot 2 \text{ complex}] \longrightarrow \text{thrombin} + \text{fragment 1} \cdot 2$$

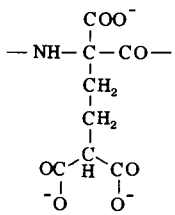
Whether this entire sequence of reactions occurs physiologically with platelets participating as phospholipid donors remains to be established, especially in respect to the thrombin feedback cleavage of the Arg-155–Ser-156 bond. Fragment 1 · 2 but not fragment 1 can be detected in serum, suggesting that thrombin cleavage of this bond is of secondary importance. In the physiologic context, much of the thrombin diffusing away from the activation site is neutralized by antithrombin (Jackson, 1978; Shapiro and McCord, 1978; Magnusson *et al.*, 1975a; Jackson *et al.*, 1975).

2. *Vitamin K and Prothrombin Complex Proteins*

Vitamin K is essential for the biosynthesis of normal prothrombin and factors VII, IX, and X, known as vitamin K-dependent proteins. In vitamin K deficiency or in subjects treated with vitamin K antagonists (e.g., coumarins), an abnormal protein is synthesized by the liver. Such proteins are called PIVKA proteins: proteins induced by vitamin K absence or antagonists. The older as well as the more recent literature has been reviewed, especially that which is concerned with the carboxylation of glutamic acid in proteins and, specifically, in the prothrombin complex (Stenflo and Suttie, 1977).

In the synthesis of prothrombin, vitamin K functions as a cofactor for a carboxylation enzyme in the posttranslational step, whereby a carboxyl group is attached to the terminal carbon of glutamic acid to form Gla residues in a polypeptide chain. Gla residues have been identified in fragment 1 of prothrombin, in the light chain of factor X (Figs. 10 and 11), in factor IX, and possibly in factor VII, as well as in some noncoagulation proteins, protein C and protein S, (DiScipio *et al.*, 1977; Stenflo and Suttie, 1977). The importance of Gla residues in these proteins becomes evident from the discovery that the abnormal proteins, lacking Gla residues, do not bind calcium and fail to be adsorbed by the usual inorganic prothrombin adsorbents; they also fail to bind to phospholipid via calcium-ion bonding. Thus, abnormal prothrombin fails to be activated for lack of binding to and orientation on the phospholipid micelle, even though the abnormal

prothrombin contains prethrombin 2, which by other means (trypsin, venom) can be converted to thrombin. The affinity of Gla residues, for calcium ions is self-evident from its structure in peptide linkage, as shown here;



The double carboxyl grouping imparts unusual binding capacity to a glutamate residue, since two calcium ions may react locally with phospholipid. These observations have contributed significantly to our understanding of part of the mechanism of prothrombin activation.

E. The Clotting Enzyme—Thrombin

Thrombin (α-thrombin) is a serine protease, the primary substrate of which is fibrinogen. It is derived from prethrombin 2 after activation of prothrombin by various activators, as discussed in VI, D, 1. Recent, as well as earlier, separation and purification procedures for bovine and human thrombins are based on ion-exchange chromatography. Various ion exchangers have been used, such as IRC 50 (Fenton *et al.*, 1977), sulfoethyl or sulfopropyl Sephadex (Lundblad *et al.*, 1976; Workman and Lundblad, 1978) and CM Sephadex (Gorman *et al.*, 1976). Specific activities of thrombin preparations ranged from approximately 2000–3000 NIH units/mg. Stable solutions of thrombin can be made in sodium acetate buffer at pH 5.0; however, fibrinogen clotting is lost on aging while the esterase activity is retained. Preservation of activity in 25% trisodium citrate (Seegers, 1967) or in the frozen state at −80°C has been used.

The molecular weight of α-thrombin is 39,000, but smaller degradative species of 26,000, 28,000, 25,000, and 23,000 result from autolytic cleavage (Gorman *et al.*, 1976). Commercial thrombins, now marketed for many years, contain several species of thrombin, in addition to other protein impurities, and can be used as a source material for isolation. The first thrombin preparations (e.g., Mellanby, 1933; Seegers *et al.*, 1940) were crude but active; nevertheless, much of the earlier knowledge about coagulation in general, and especially about fibrinogen, was gained through their use. The autolytic cleavage products of degradation do not clot fibrinogen but do retain esterase activity.

The esterase activity of thrombin is directed toward a number of ester substrates (e.g., arginine, lysine, and *p*-nitrophenylesters). As a protease, however, its specificity is limited to arginyl and lysyl bonds. Some thrombin substrates, in addition to fibrinogen, are factors II, V, VII, IX, X, and possibly VII; in addition, thrombin induces proteolysis of several enzymes and some hormones, and aggregates platelets (see Shapiro and McCord, 1978; Seegers *et al.*, 1975). Thrombin also is a potent platelet aggregator for which an active site is necessary (see Section II,A).

Bovine thrombin is a two-chain disulfide-linked polypeptide. The A chain contains 49 amino acids, and the B chain, 259; complete chain sequences have been reported (Magnusson *et al.*, 1975b). Human thrombin has a similar two-chain structure; the A chain contains 36 residues, and corresponds to residues 14–49 of bovine A chain; hence, 13 residues (Thr-274–Arg-286) at Arg-286 (Fig. 10) have been cleaved by thrombin. In bovine thrombin, Lys-286 replaces Arg-286 of human thrombin, the lysine bond being less sensitive to cleavage by thrombin; accordingly, the 13-residue chain is retained in bovine thrombin. There is no difference, however, in the specific activity of both enzymes (Thompson *et al.*, 1977). The B chain of human thrombin has 264 residues. The catalytic charge–relay system on both B chains is similar; the active serine (SER) is at residue 527, and the histidine at 365 in human prothrombin (see Table III; Shapiro and McCord, 1978). The 13 residues deleted in human thrombin appear not to have a significant function.

VII. BIOCHEMISTRY OF THE TRANSFORMATION OF FIBRINOGEN TO FIBRIN

The final phase of blood coagulation and the ultimate consummation of all the chemical reactions preceding thrombin formation is the transformation of a soluble protein, fibrinogen, to an insoluble protein, fibrin, which forms a pseudocrystalline meshwork of fibers that constitute the gel or fibrin clot. This represents the first enzymatic event in this final coagulation sequel. The second enzymatic event is a transamidation reaction, which is catalyzed by factor XIII, whereby the fibrin fibers are crosslinked to form the stable, hemostatic fibrin clot. Although fibrinogen and fibrin are the oldest clotting factors known, the total chemistry of these proteins is not yet understood, due to its high molecular weight, its complicated primary structure, and the insoluble nature of its polymeric state. Over the years, a vast literature has accumulated with the advance of knowledge in this field, and it is possible only to skin the surface in this review, which will be limited to human and bovine fibrinogen, although

the fibrinogen of many vertebrate and invertebrate species has been studied (Doolittle, 1973, 1975).

Many excellent reviews have appeared over the years, which touch on the various aspects of the chemistry, physiology, and therapeutic application of plasma fibrinogen, among which may be cited the following: Scheraga and Laskowski (1957), Blombäck (1967), Blombäck *et al.* (1976a), Laki (1968), Blombäck and Blombäck (1972), Murano (1974), Doolittle (1973, 1975), Mosesson and Finlayson (1976), and Gaffney (1977a,b; Blombäck, 1979). In this review, we are concerned only with the chemistry of plasma fibrinogen. A molecular species of fibrinogen does exist in platelet α granules, but this so-called platelet fibrinogen has undergone only limited biochemical characterization. It, apparently, is different from platelet-associated fibrinogen, which presumably is simply fibrinogen adsorbed from the plasma. Developments in this field have been recently reviewed (James *et al.*, 1977).

A. Chemistry of Fibrinogen

Fibrinogen is a hepatic plasma protein with a relatively high molecular weight of 340,000 (Scheraga and Laskowski, 1957). Normal human plasma levels range from 200 to 400 mg/dl; thus, fibrinogen is the highest in concentration among clotting factors, followed by prothrombin at 10 mg/dl. All other clotting factors are trace components of plasma. Fibrinogen is a euglobulin; it precipitates when plasma is diluted to a low ionic strength with water; however, in this form fibrinogen is unstable and is subject to spontaneous fibrinolysis.

Highly clottable fibrinogen can be prepared by a variety of methods. Clottability is expressed as percent of total protein converted to fibrin by thrombin; however, high clottability is not necessarily synonymous with purity since a few impurities remain in trace amounts in highly clottable preparations.

Since fibrinogen is the least soluble of plasma proteins at room temperature, it becomes the first of the plasma proteins to separate by precipitation methods. In 1879, Olof Hammarsten prepared the first fibrinogen preparation by repeated precipitation with half-saturated NaCl. Since then, numerous precipitation methods have been developed with neutral salts (ammonium sulfate and potassium phosphate); with solvents (the Cohn cold ethanol and the Keckwick cold ether procedures), and with nonorganic solvent methods utilizing 20 m*M* zinc acetate for precipitation. These and other, older methods have been reviewed many times (for references, see Pennell, 1960; Kazal, 1962). More recently, precipitation of fibrinogen from $BaSO_4$-adsorbed plasma has been described for amino

acids (glycine: Kazal *et al.*, 1963; Walker and Catlin, 1971; β-alanine: Jacobsen and Kieruff, 1973) and for heavy metal coordination complexes (potassium tetrathiocyanate-(*S*)-mercurate: Brown and Rothstein, 1967). The freeze–thaw technique of cryoprecipitation of plasma fibrinogen (Ware *et al.*, 1947; Pool *et al.*, 1964) separate a crude but important preparation of fibrinogen, suitable for therapeutic application or as an initial step for further purification. The glycine precipitation method (Kazal *et al.*, 1963) is a simple procedure for obtaining small or large amounts of highly clottable fibrinogen, and is particularly adaptable to clinical studies (Martinez *et al.*, 1974, 1977). The Cohn Cold Ethanol Procedure is a standard for laboratory or commercial preparations (Cohn *et al.*, 1946), and further purification of Cohn Fraction I can be achieved easily with ammonium sulfate precipitation (Laki, 1951). The Cohn Cold Ethanol–Glycine Procedure provides a fibrinogen of high purity, which has been used extensively for structural studies (Blombäck and Blombäck, 1956). A cold ethanol procedure as well has been devised for a high purity preparation for these studies (Doolittle *et al.*, 1967). Also, polyethylene–glycol (5%) precipitates plasma fibrinogen (Polson and Ruiz-Bravo, 1972).

Generally, additional steps are necessary to achieve ultimate purity, and chromatography on DEAE-cellulose (Finlayson, 1968, 1972) or precipitation in the presence of lysine (Bergström and Wallén, 1961) are especially valuable for the removal of biologic impurities, particularly plasminogen, factor XIII, and traces of factor VIII. Cold-insoluble globulin is another contaminant if fibrinogen is prepared at low temperatures. For example, it is found in Cohn Fraction I, which separates at 8% ethanol, pH 7.2, ionic strength 0.14, and −3°C (Cohn *et al.*, 1946). It also is present in the 2.2 *M* glycine precipitate, if precipitation is conducted at +5°C. With glycine at 20°C, however, coprecipitation of cold-insoluble globulin is eliminated and 98% clottable fibrinogen is possible after a few precipitations (Kazal *et al.*, 1963). Actually, fibrinogen cannot be regarded as a single species of protein for its exhibits heterogeneity (Finlayson and Mosesson, 1963) in solubility and coagulability, and such subfractions have been separated chromatographically from 2.1 *M* glycine precipitates obtained at +5°C (Mosesson and Sherry, 1966; Mosesson *et al.*, 1967; Mosesson and Umfleet, 1970). Fibrinogen of low, intermediate, or high solubility also has been demonstrated in ethanol–water solvent systems. The most extensive structure studies have been conducted with fibrinogen of intermediate solubility (fraction I-4 of Blombäck and Blombäck, 1972). Finally, as an aside, some subfractions of cold insoluble globulin itself are partly clottable and have assumed importance as a circulating cell surface protein (Mosesson, 1977).

Physiochemical studies are not in complete agreement on the shape of

the fibrinogen molecule. Its hydrodynamic behavior leaves much room for the interpolation of data in regard to shape, although from its sedimentation coefficient (S_{20}, ω = 7.9 S), the molecular weight can be clearly placed at 340,000 ± 20,000. Its frictional ratio (f/f_0) of 2.34 supports an elongated molecule rather than a sphere; however, under minimum hydrated conditions, the data support an ellipsoid shape with an axial ratio of 30, whereas a hydrated molecule could be a sphere of 200-Å diameter. The dimensions of the molecule more likely are in the neighborhood of 90 × 450 Å (Shulman, 1953; Scheraga and Laskowski 1957; Koppel, 1967). Electron microscopic studies support either an elongated trimodular shape (Hall and Slayter, 1959) or a polygonal shape (Koppel, 1967), but whether any of these represent the true physiological shape is mere speculation. More recent investigators of subunit structure have produced other concepts in structure (see Gaffney, 1977a), and the electron microscopic determination of shape from studies of microcrystals of partially degraded fibrinogen suggest's either a compact or a swollen structure (Tooney and Cohen, 1972).

Human and bovine fibrinogens are glycoproteins containing about 4.5% carbohydrate; variable amounts of hexose, hexosamine, and neuraminic acid have been reported (Murano, 1974). Fibrinogen is a dimer; each monomer is composed of three chains: Aα, Bβ, and γ chains having molecular weights of 63,500, 56,000, and 47,000, respectively. The chemical formula for vertebrate fibrinogens may be expressed as $(A\alpha B\beta\gamma)_2$ or $A\alpha_2B\beta_2\gamma_2$, to indicate the existence of three pairs of similar but apparently nonidentical chains. The molecular weight of the monomer is 170,000, and that of the dimer, 340,000. The amino terminal residue of the Aα chain is alanine, and that of the γ chain is tyrosine in human fibrinogen; however, the Bβ chain has a terminal pyrolidine carboxylic acid (pyroglutamic acid) and, therefore, its terminal is masked. The carboxy-terminal residues in the α, β, and γ chains consist of valine in human fibrinogen. N- and C-terminals vary in different mammalian fibrinogens (Doolittle, 1975).

The amino acid compositions of mammalian fibrinogens are similar. Of interest is the presence of substituted amino acids, such as tyrosine-*O*-sulfate in mammalian fibrinogen and phosphoserine in higher primates and man. Of importance are the 58 to 60 cysteine residues. Since fibrinogen has no free sulfydryl groups, these residues form disulfide bridges (29 to 30), some of which interconnect the individual chains, whereas others are clustered at the amino terminus of the molecule (Blombäck *et al.*, 1976a). The cluster of chains has been named the "N-terminal disulfide knot" (N-DSK) by Blombäck (1972), and it represents an important structural feature of the fibrinogen molecule. A diagrammatic sketch of the N-DSK part of fibrinogen is shown in Fig. 12, which actually illustrates the present

general concept of the overall structure of fibrinogen (Blombäck *et al.*, 1976b); however, only about 15% of the sequence is defined (Gaffney, 1977a).

The illustration (Fig. 12) suggests the existence of symmetry in the molecule of fibrinogen. Heterogeneity, however, appears to exist among subunit chains of fibrinogen. Physical, chemical, physiologic, and clinical evidence all support the concept of heterogeneity with, perhaps, 36 distinct molecular forms having been suggested for bovine fibrinogen (Gaffney, 1977a; Blombäck *et al.*, 1976b). Whether these are artefacts of isolation, effects of enzymic degradation, or results of variations in biosynthesis is not clear; nevertheless, they add to the complexity of the compositions of fibrinogen and the elaboration of its structure.

The N-terminal sequences of the Aα and Bβ chains contain the fibrinopeptides A and B, which are located in the N-DSK domain of fibrinogen. It is this region of the molecule that is attacked by thrombin, which first releases two peptide A chains from the α chains and, subsequently, two peptide B chains from the β chains. The fibrinopeptidic amino acid sequences were the first to be determined in the search for the structure of fibrinogen. The complete sequences have been determined by Blombäck *et al.* (1966), Doolittle (1970), and others for many mammalian species (see Gladner, 1968). The four human fibrinopeptides were isolated by column chromatography on Dowex 50 × 2, with an eluting buffer of 0.2 *M* ammonium acetate, or by extraction with 0.05 *M* pyridine of a fibrin clot supernatant, followed by gel filtration on Sephadex G-25 (Blombäck *et al.*, 1966). These had the following properties:

Peptide	Molecular weight	NH_2 terminal	COOH terminal	No. of amino acids
A	1536	Ala	Arg	16
B	1552	PCA[a]	Arg	14
AP	1616	Ala	Arg	16
Y	1465	Asp	Arg	15

[a] PCA, pyrollidine carboxylic acid.

Peptides A, AP, and Y possessed identical sequences, except for a phosphoserine in fibrinopeptide AP in position 14 and one amino acid residue less in fibrinopeptide Y.

The primary structure of fibrinogen is partially known from two approaches: cyanogen bromide fragmentation (CNBr) and enzymic digestion with plasmin, recently reviewed by Gaffney (1977a).

CNBr cleaves methionyl bonds in fibrinogen and produces several large molecular fragments. Fibrinogen contains many susceptible bonds among

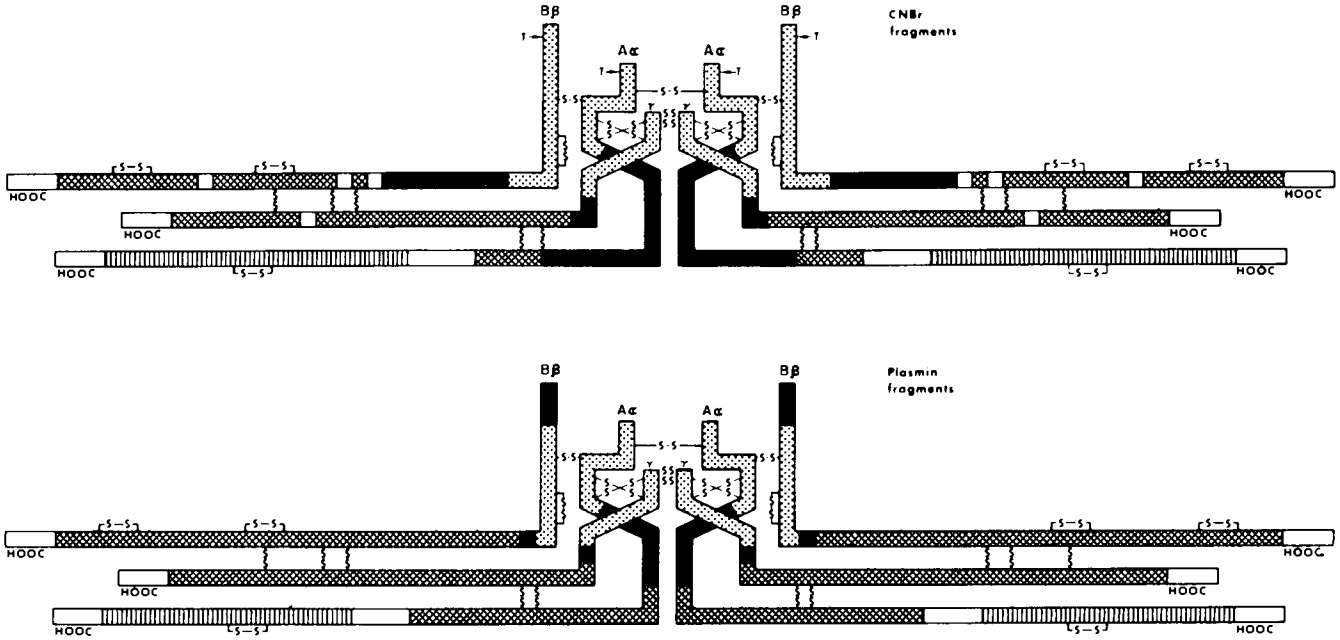


Fig. 12. Structure of the fibrinogen molecule based on present knowledge of CNBr and plasmic digestion fragments isolated from human fibrinogen. The top half shows arrangement of CNBr fragments in the Aα, Bβ, and γ chains. The dotted area is the N-DSK domain, showing disulfide linkages and the points of cleavage of fibrinopeptides A and B by thrombin (T). The black area represents CNBr-1 and CNBr-3 fragments; the cross-hatched area, CNBr fragments Ho1- and Ho2-DSK; the striped area, Hi2-DSK. Unknown structures are denoted by the white areas. The lower half shows the plasmic fragments arranged with this structure. Dotted area is fragment E; the cross-hatched area, fragment D; the striped area, PL-1 and PL-2; the white area, unknown. The black areas show the overlaps of structure between CNBr and plasmic fragments. Reproduced with permission from Blombäck *et al.* (1976b), *Thromb. Res.* **8,** 329, Pergamon Press, Ltd.

TABLE V

CNBr and Plasmic Fragments of Human Fibrinogen

Fragment	Molecular weight	Disulfides per mole	Chain origin
CNBr			
N-DSK	58,000	11	$A\alpha,B\beta,\gamma$
Hi2-DSK	28,000	2	α
Ho1-DSK	42,500	12	α,β,γ
Ho2-DSK	7,000	2	β
CNBr-1	6,000	0	α
CNBr-2	10,000	0	β
CNBr-3	14,000	0	α
Plasmin			
D	85,000	15	α,β,γ
E	50,000	11	$A\alpha,B\beta,\gamma$
PL-1	50,000	2	α
PL-2	20,000	0	α

its 60 methionine residues, so that the resultant peptides are either chains or disulfide-bonded chains. The disulfide-bonded chains, which are called disulfide knots (DSK), are derived from parts of adjacent polypeptide chains of fibrinogen. Some are hydrophilic (N-DSK and HL2-DSK); others are hydrophobic (Ho1-DSK and Ho2-DSK). The N-DSK, (Fig. 12, Table V) contains 11 of 28 disulfide bridges of fibrinogen; the A and B fibrinopeptides are part of the N-DSK. The remaining DSK regions are distributed away from the N-DSK "head" in the C-terminal direction, as shown in Fig. 13, in which they are aligned with plasmin-derived fragments (Blombäck *et al.*, 1976a,b).

The enzymatic digestion of fibrinogen by plasmin produces several fragments representing different segments of the α, β, and γ chains. Two fractions, D and E (Nussenzweig *et al.*, 1961), following exhaustive digestion, make up 70% of the fibrinogen molecule, whereas earlier degradation products (X and Y) represent larger fragments. Intermediate fragments representing an early phase of enzymatic cleavage also are recognized. These disappear with extensive digestion. A nomenclature for these fragments has been proposed (Marder and Budzynski, 1972). The classification of plasmin-derived fragments and their coagulant, anticoagulant, and molecular size is shown in Table VI.

The properties of the degradation products of fibrinogen have been reviewed (Marder and Budzynski, 1975b). The fragments can be distinguished immunologically since they have different antigenic determinants and have been separated by polyacrylamide gel electrophoresis (5% gel, 2 *M* urea, pH 2.7), in which fragment E has the greatest electrophoretic

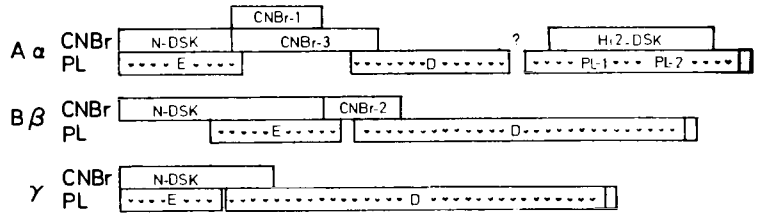


Fig. 13. Areas of the overlapping fragments of the CNBr and plasmic fragments in human fibrinogen are illustrated. The CNBr fragments are shown in the upper arrangement and the plasmic fragments (PL) in the lower in each of the three chains of fibrinogen. Reproduced with permission from Blombäck *et al.* (1976b), *Thromb. Res.* **8,** 329, Pergamon Press, Ltd.

TABLE VI

Classification of Plasmin-Derived Fragments of Human Fibrinogen[a]

Fragment	Molecular weight	Chain identity	Chain molecular weight	Antigenic determinant	Properties
(Fibrinogen)	340,000	Aα	68,000	F,X-Y,D,E	Normal polymerization and clottability
		Bβ	58,000		
		γ	49,000		
Fragment X	250,000	Aα	25,000	X-Y,D,E	Slow polymerization; inhibition of polymerization
		Bβ	52,000		
		γ	47,000		
Fragment Y	150,000	*Light*		X-Y,D,E	Nonclottable
		Aα	13,000		Inhibition of polymerization
		Bβ	9,000		
		γ	10,000		
		Heavy			
		Aα	25,000		
		Bβ	51,000		
		γ	40,000–48,000		
Fragment D	80,000–100,000	Aα	13,000		Nonclottable
		Bβ	42,000		Inhibition of polymerization
		γ	26,000–39,000		
Fragment E	45,000	Aα	10,000		Nonclottable
		Bβ	7,000		Inhibits thrombin
		γ	9,000		Inhibits thromboplastin generation; antiplatelet

[a] Marder and Budzynski, 1975a.

mobility. They also have differing biologic properties, as shown. Their relationship to CNBr fragments is illustrated in Fig. 13, from which it becomes obvious that the two methods of fragmentation produce overlapping peptides (Blombäck and Blombäck, 1972; Blombäck *et al.*, 1976a, 1976b; Murano, 1974; Doolittle *et al.*, 1977a; Collen *et al.*, 1975).

Studies of proteolytic degradation of human and bovine fibrinogen stimulated the development of new models of protein structure. A trinodular structure was derived from observations on the plasmic digestion of fibrinogen (Marder and Budzinski, 1975a). A kinetic model, based on the rates of proteolytic cleavage of fibrinogen with plasmin and trypsin into fragments X, Y, D, E, and other smaller fragments, suggested the existence of three pairs of chains that connect each of the two D-fragment domains to the central E-fragment domain, and a seventh chain that connects the two outer D fragments (Schrager *et al.*, 1976). Both of these models were tailored after the trinodular concept of structure derived from electron microscopic studies (Hall and Slayter, 1957). In another model an elongated trinodular structure was proposed, in which the $A\alpha$, $B\beta$, and γ chains were arranged in hexagonal configuration held together by interchain disulfide bridges with the N-DSK region fixed by disulfide bonds and the C-terminal region portrayed in a random, non-ordered conformation (Mosesson and Finlayson, 1976). Still another model, structured around a 518-amino-acid sequence, depicts an amino-terminal central domain consisting of six chains, held together by disulfide bonds in such fashion that the fibrinopeptides protrude beyond the N-DSK and the C-terminal regions extended in the opposite direction, its long flexible chains having regions of α-helices, with carbohydrate residues located in the γ-chains (Doolittle, 1975).

Most of these models are based on the unique N-terminal disulfide knot discovered by Blombäck and co-workers. Cyanagen bromide degradation of fibrinogen yielded many fractions with from approximately 2000 to 6000 MW. The 58,000-MW fraction represented 16% of the entire molecule and contained one-half of its disulfide bridges. By sulfitolysis, disulfide bridges are broken, and fibrinogen yields three polypeptide chains, $A\alpha$, $B\beta$, and γ, which are present as chain pairs in the intact molecule. There is strong experimental evidence for a dimeric structure. Furthermore, the amino-acid sequencing of the CNBr degradation fragments provided the necessary insight into the primary structure of fibrinogen. The N-terminal disulfide knot has been shown to consist of an intricate dimeric arrangement of chains, linked by disulfide bridges. At its head are the fibrinopeptides. Internal disulfide bonds, which can take part in disulfide exchange, produce polypeptide loops. The unraveling of the N-DSK

obviously is the key to an understanding of the structure of fibrinogen (Blombäck, 1979).

It has proven to be only a matter of time before the complete primary structure of the three chains of fibrinogen would be determined. This now has been achieved by contributions from several laboratories. The α chain has been shown to contain 610 amino acids, starting with Ala-1 and ending with Val-610 (Doolittle *et al.*, 1977b; 1979). The β chain has 461 residues with an N-terminal Gln-1 and a C-terminal Gln-461 (Blombäck *et al.*, 1976b, 1979; Henschen and Lottspeich, 1977b; Watt *et al.*, 1979). The γ chain has 411 residues with a Tyr-1 and a Val-411 (Henschen and Lottspeich, 1977a). Essentially the entire amino acid sequence of the N-DSK was determined by Blombäck. The remainder of the molecule was determined in the laboratories of Henschen and Lottspeich and of Doolittle.

The calculated molecular weight for the 2964 amino acids in the three chains was 329,840, with a molecular weight of 10,000 for the four carbohydrate groupings, one on each of the two β and two γ chains, which gave a final molecular weight of 340,000, in excellent agreement with physicochemical measurements.* These observations provide an excellent base for further evaluation of the structure of fibrinogen.

The role of calcium in the structure of fibrinogen has been probed (Marguerie, 1977). Three high-affinity calcium-binding sites and some nonspecific ones have been established. Calcium protects against denaturation by acid and offers resistance to plasmic digestion. The binding sites are located in the terminal part of the Aα chains; calcium appears to stabilize a more compact structure of the fibrinogen molecule.

B. Transformation of Fibrinogen to Soluble Fibrin

The first reaction in this transformation is initiated by thrombin, which releases four acidic fibrinopeptides (FBP) from the amino-terminal ends of two α and two β chains by cleavage of specific arginyl–glycine bonds, according to the reaction:

$$\text{Fibrinogen} \xrightarrow{\text{thrombin}} \text{fibrin monomer} + 2\ \text{FBP-A} + 2\ \text{FBP-B}$$

The A peptides are rapidly released first and the B peptides are released only after a lag phase; calcium ions are not required. The A and B peptides represent a little less than 2% of the fibrinogen molecule, leaving a

* I am deeply indebted to Dr. Russell F. Doolittle of the University of California, San Diego and LaJolla, California, for kindly providing a prepublication copy of his review "Fibrinogen and Fibrin," in which these calculations and estimates of molecular weights were presented.

huge fibrin monomer that can polymerize to fibrin. The function of the fibrinopeptides is inferred to be a nonspecific blocking of those chemical groups of amino acid residues of the fibrin monomer that are involved in the polymerization reaction. In support of this is the observation that they bear a strong negative net charge, and thus may maintain fibrinogen molecules in an unreactive state by mutual electrostatic repulsions. The γ chains of fibrinogen are not attacked by thrombin (Doolittle, 1975).

The polymerization of fibrin monomer follows spontaneously after the release of either A peptides alone, or after both A and B peptides are released. The reaction is as follows:

$$n \text{ (fibrin monomer)} \xrightarrow[\text{(neutral pH)}]{\text{spontaneous}} \text{(fibrin monomer)}_n$$

or

$$\underset{\text{fibrin monomer}}{(\alpha_2\beta_2\gamma_2)} \xrightarrow{Ca^{2+}} \underset{\text{fibrin}}{(\alpha_2\beta_2\gamma_2)_n}$$

where n represents an indeterminate number of fibrin monomer molecules that constitute the fibrin fibers of the clot, and where $\alpha_2\beta_2\gamma_2$ is the fibrinogen dimer denuded of its fibrinopeptides. The product of this reaction is a large-molecular-weight insoluble protein $(\alpha_2\beta_2\gamma_2)_n$, which constitutes the fibrin clot. The fibrin in experimental clots formed from factor XIII-free fibrinogen is fibrin S (Fig. 3), a nonstabilized, molecular structure consisting of long chains of fibrin monomers that are presumed to be electrostatically attracted by end-to-end and lateral interactions of oppositely charged groups on the fibrin monomer (Ferry *et al.*, 1954). These interactions lead to the formation of dimers, trimers, etc., and to intermediate polymers, which associate laterally to form fibrils. Fibrils eventually branch to form a network of fibers (Fig. 14) that constitute the gel (Doolittle, 1975).

The nature of the fibrous gel is determined in experimental systems by the kind of clotting agent used. Thrombin-induced clotting leads to end-to-end and lateral interactions; however, Reptilase, a *Bothrops jacaraca* snake venom enzyme, which releases only fibrinopeptide A, induces polymer formation by only end-to-end interactions of the fibrin monomer. In this case, the structure of fibrin would be $(\alpha_2B\beta_2\gamma_2)_n$, the physical properties of the clot differing somewhat from a thrombin-induced clot (Blombäck, 1967).* Ancrod, a purified fraction of venom from the Malayan pit viper, also releases only fibrinopeptide A (Esnouf and Tunnah,

* These molecular subunit structures also are expressed by some investigators with the subscript n replaced by P, for instance,

$$\underset{\text{monomer}}{(\alpha\beta\gamma)_2} \longrightarrow \underset{\text{fibrin polymer}}{(\alpha\beta\gamma)_{2P}}$$

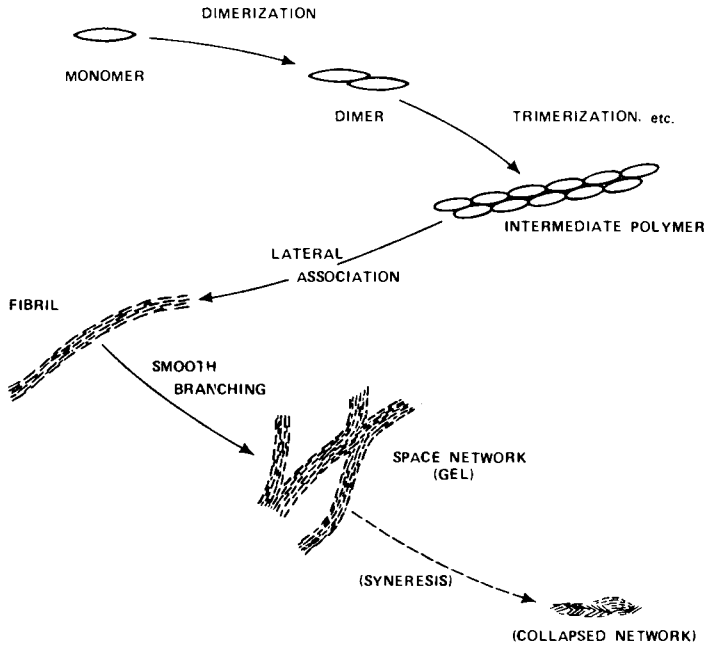


Fig. 14. Polymerization of fibrin monomers followed by clot retraction (syneresis). Reproduced with permission from Doolittle, 1975.

1967). Since all four fibrinopeptides are located in the N-DSK domain of fibrinogen, the release of only the A peptides, rather than A and B peptides, should alter the nature of the interaction products. Reptilase and Ancrod-induced fibrin is more susceptible to plasmic digestion than thrombin-induced fibrin. Fibrinopeptide B can be selectively removed by another venom-derived enzyme from a copperhead snake without induction of fibrin formation (Herzig *et al.*, 1970), attesting to the importance of the polymerization site in the fibrin monomer, which is uncovered by the release of fibrinopeptide A.

The polymerization or binding site covered by fibrinopeptide A is in the N-DSK or fragment-E domain, for which there is much convincing evidence (Blombäck *et al.*, 1976a; Murano, 1974). Other binding sites become operative when fibrinopeptide B is released by thrombin (Blombäck and Blombäck, 1972), which apparently occurs only after a conformational change takes place in fibrinogen upon the release of peptide A. This probably accounts for the lag phase in release of peptide B by thrombin. All three chain (α, β, and γ) remnants in the N-DSK or fragment-E domain are important for binding, including three critical tyrosine residues, which if iodinated cause a loss of binding capacity. Polymerization also is inhibited by photooxidation of fibrinogen, which results in the oxidation

of histidine and/or tryptophan residues. Polymerization sites uncovered by release of fibrinopeptide A also may be affected (Inada *et al.*, 1978). A second set of binding sites may be present in fragment D in the C-terminal end of its chains, a postulate based on the interaction of the fibrin monomer–Sepharose binding of fragment D, and on studies of the abnormal, nonclottable fibrinogen "Detroit," which has a mutation (Aα19-Arg → Ser) in the N-DSK region. Since these sites apparently are located in N- and C-terminal domains, the new chemical evidence strongly supports end-to-end alignment and side-to-side attraction of the aligned polymers (Blombäck *et al.*, 1976a), an interaction scheme similar to that of the electrostatic–hydrodynamic model suggested earlier (Ferry *et al.*, 1954).

The above events have been elaborated for a purified fibrinogen and for reactions taking place in the absence of calcium ions and fibrin-stabilizing factor, factor XIIIa. Fibrin monomers formed under these conditions undergo only reversible aggregation to produce a clot (fibrin S) that is insoluble in saline, but which can be solubilized in solutions of 5 *M* urea, 1% monochloracetic acid, or 1 *M* NaBr (Lorand, 1962). Physiologic clotting, however, occurs in the presence of the cofactors calcium and factor XIII. The latter is activated to XIIIa by thrombin, in the presence of calcium ions. The transpepditation reaction induced by enzymic factor XIIIa results in the cross-linking of fibrin polymers to produce a stabilized fibrin clot, which is insoluble in the above solvents (Lorand, 1975). The stabilized clot is the physiologic, hemostatic blood clot. The polymerization reaction is essential for normal wound healing. The overall transformation of fibrinogen to fibrin through polymerization reactions is illustrated schematically in Fig. 14.

C. Factor XIII-Induced Transamidation and Insoluble Fibrin Formation

Human factor XIII is present in plasma at levels approximating 2 mg/dl, and is present also in many tissues. In plasma, it is a proenzyme that can be activated to its enzymic state by thrombin in the presence of calcium ions. It is synthesized by the liver. Megakaryocytes also synthesize a similar enzyme, platelet factor XIII. Common names frequently used for factor XIII are fibrin-stabilizing factor (Lorand, 1950; 1976), protransglutaminase (Chung, 1972), fibrinase, now in disuse (Loewy *et al.*, 1961), and plasma transglutaminase (Loewy, 1968). Fibrinoligase refers to the activated enzyme, XIIIa (Lorand, 1972). Excellent reviews of the study of factor XIII are available (Loewy, 1970, 1972; Lorand, 1972; Folk and Chung, 1973, 1975; Finlayson, 1974; Doolittle, 1973, 1975; Bohn, 1978).

Factor XIII has been prepared from heated (56°C) plasma by ammo-

nium sulfate precipitation and DEAE-cellulose chromatography (Loewy *et al.*, 1961). Heat treatment is used to remove the bulk of the plasma protein and especially fibrinogen, since factor XIII tends to be separated with fibrinogen by coprecipitation in procedures not utilizing heat treatment. Precipitation of factor XIII can be accomplished with ethanol and with polyethylene glycol. Ion-exchange chromatography provides final purification. The adsorption of prothrombin factors from the plasma before fractionation is desirable in order to prevent the activation of factor Xa and especially thrombin, which would activate factor XIII. Purification in the range of 2000–4000-fold with 20% recovery of activity has been obtained (Lorand and Gotoh, 1970; McDonagh *et al.*, 1969; Schwartz *et al.*, 1973; Takagi and Konishi, 1972; Kazame *et al.*, 1976).

Plasma factor XIII is a glycoprotein with an electrophoretic mobility of a β_2-globulin having approximately 5% carbohydrate (Bohn, 1978). The molecule is a tetramer in which two pairs of nonidentical chains are held together by noncovalent bonds, so that its subunit molecular structure has been represented as a_2b_2 (Schwartz *et al.*, 1971, 1973; Chung *et al.*, 1974) or A_2S (Bohn, 1978), where subunit dimer a_2 or A_2 contains the catalytic sites of the enzyme and the dimer b_2 or S is a nonenzymatic component. All of the carbohydrate (8.5%) is found in the b_2 or S chains (Loewy *et al.*, 1961; Schwartz *et al.*, 1973; Bohn, 1978). It has been suggested that b chains may stabilize the a chains or aid in secretion at the site of synthesis, and it has been demonstrated that they influence the rate of activation by thrombin (Chung *et al.*, 1974); however, the biologic function and physiologic significance of the b chains are unknown. The amino terminal of the a chain is masked by an *N*-acetyl serine residue; that of the b chains is glutamic acid or glutamine. Six sulfhydryl groups are part of the a chain, but there are none on the b chain; there are 16 or 17 intramolecular disulfide groups in the b chain (Takagi and Doolittle, 1974; Chung *et al.*, 1974; Bohn, 1978). The molecular weight of the plasma factor XIII tetramer has been reported as 300,000 to 320,000 (Loewy *et al.*, 1961; Schwartz *et al.*, 1973; Chung *et al.*, 1973; Chung *et al.*, 1974), and has been revised up to 340,000 recently (Bohn, 1978).

The physical and chemical properties of plasmic, placental, and platelet fibrin-stabilizing factors have been compared (Folk and Chung, 1973; Bohn, 1978). Platelet and placental factor XIII also have β_2-globulin electrophoretic mobility, and have been prepared in crystalline form. Unlike the plasma factor, however, they are dimers with the composition of an a_2 or A_2 subunit structure, and molecular weights of 160,000. The a and b chains of plasma factor XIII have been reported as 75,000 and 80,000, respectively, and that of placental and platelet factor-XIII a chains as 80,000 daltons. The dimers of the latter enzymes (a_2 or A_2) have only 1.5% carbo-

hydrate. The a-chain subunits of these three factors are quite similar and essentially indistinguishable in amino acid composition, electrophoretic mobility, molecular weight, and activation by thrombin.

Thrombin and other proteases activate factor XIII by cleavage of an arginyl-glycine bond in each a chain with the liberation of two activation peptides (4,000 daltons) from the amino-terminal portion of the a chains, thereby reducing the molecular weight by 8,000 daltons according to the following reactions:

$$\underset{\substack{\text{(factor XIII,}\\ \text{320,000 MW)}}}{a_2b_2} \xrightarrow{\text{thrombin}} \underset{\text{(intermediate)}}{a_2'b_2} + \underset{(2 \times 4{,}000)}{\text{2 peptides}}$$

$$a_2'b_2 \xrightarrow{Ca^{2+}} \underset{\substack{\text{(factor XIIIa,}\\ \text{140,000 MW)}}}{a_2'} + b_2$$

In the absence of calcium ions, the tetrameric molecular structure is retained, but the molecule is inactive. Calcium ions are required for the dissociation of the tetramer to the dimeric and enzymically active form; this dissociation is reversible, with the interesting observation that platelet a_2 dimers will combine with plasma b_2 dimers to reproduce an active hybrid tetramer. Factor XIIIa (transamidase or fibrinoligase) is an enzyme with a cysteine-SH active site as the functional group (Lorand, 1975), and is the only clotting factor enzyme of this type, since other activated factors have serine active centers. The cysteine-SH center in only one of the chains of the dimer (a_2') can be inhibited by alkylation to produce an enzyme with "half-of-the-site" reactivity; the amino acid sequence in the active center region has been shown to be Gly-Gln-Cys-(SH)-Trp. Platelet factor XIII is activated more rapidly by thrombin than plasma factor XIII; accordingly, the b_2 dimer in the latter may exert some control over the rate of activation (Chung *et al.*, 1974).

In the clotting of fibrinogen, factor XIIIa functions as a unique cross-linking agent, building a multitude of chemical bridges between adjacent polypeptide chains of fibrin polymers (fibrin S), to produce fibrin I. Factor XIIIa is a transglutaminase, which, in the presence of calcium ions, forms covalent intermolecular bonds between lysine and glutamyl residues on adjacent polypeptide chains of fibrin S, which results in the stabilization of fibrin. Cross-links are rapidly formed between antiparallel γ chains to give γ dimers. Alpha chains react more slowly to produce α polymers, while β chains are not cross-linked at all. The condensation reaction between lysyl donor groups (ϵ-NH_2) and the acyl (γ-$CONH_2$) groups of the glutamyl acceptor results in the release of ammonia and the formation of an ϵ-lysyl–γ-glutamyl bond (Fig. 15) that imparts insolubility to fibrin S

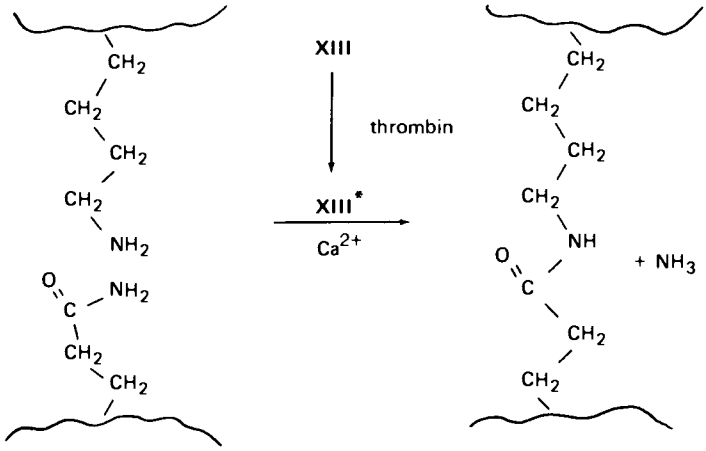


Fig. 15. The cross-linking of polymer chains of fibrin S by the action of factor XIIIa, fibrinoligase. This transamidase unites specific donor lysine residues in one polymer chain with acceptor glutamine residues in another chain to produce an $\epsilon(\gamma$-glutamyl) lysine cross-link (usually 6 bonds per mole of fibrin). The resulting polymer is fibrin I. Reproduced with permission from Doolittle, 1975.

and greater resistance to plasmic digestion; two γ–γ linkages are formed per mole of fibrin. Platelet transglutaminase functions in the same manner.

In the γ chains the cross-linkages occur within the fifteenth-residue region of the C terminus between Lys-15 of one chain and Glu-7 of the adjacent chain in both human and bovine fibrin (Chen and Doolittle, 1971). In the α chains, the acceptor–donor sites are in the middle 20% of the chain, based on plasmic digestion and CNBr cleavage of acceptor sites of fibrin, labeled with fluorescent monodansylcadaverine donor. It has been suggested that the physiologically important cross-linking acceptor sites for α-chain polymerization are located in residue 310 or 385 of the A chain of fibrinogen (Fretto *et al.*, 1978).

The γ–γ linkage also has been demonstrated in plasmic fragment D (Pizzo, 1973). The release of plasmic fragments D and E from fibrin S results from the straightforward cleavage of polypeptides that are not cross-linked; however, in fibrin I, the noncovalent bonds holding the polypeptide chains together alter the cleavage pattern with the plasmin release of a different product, a D-dimer–E-dimer complex, as well as other complexes. The major product of cross-linked fibrin digestion by plasmin (the D-dimer–E-complex) may offer a means of monitoring fibrin clot lysis *in vivo* by means of specific antisera to this complex. The complex mechanistic pattern involved is under intense investigation and must yet be resolved (Gaffney, 1977a).

VIII. SIGNIFICANCE OF COAGULATION BIOCHEMISTRY

It is obvious from the foregoing account of the mechanisms of blood coagulation that our knowledge of blood coagulation has made and is making tremendous advances on many fronts, as a result of a wide and concerted application of biochemical techniques and expertise. In contrast to a few decades ago, we now have a phenomenal and exciting base on which to proceed. This is quite evident from the rapidly and vastly expanding literature in the biochemical, clinical, physiological, and pharmaceutical disciplines in which the more recent knowledge of blood coagulation mechanisms is being applied to develop a better understanding of physiologic function and to devise new therapeutic modes of treatment and analysis. Most importantly, this has been and will be dependent upon achieving a high degree of purification of the isolated coagulation factors and the elaboration of their biochemical properties and reactions. The end is not yet in sight in this vast and complicated system, originally studied only by the appearance and timing of an apparently simple blood clot.

ACKNOWLEDGMENTS

I am deeply indebted to Joan Marks of the Department of Clinical Laboratory Research Division at Hoffman-La Roche, Inc. for typing the manuscript of this chapter. Permission to reproduce and/or adapt illustrations from their publications was provided kindly by the following: Professor Birger Blombäck of the Karolinska Institute (Stockholm, Sweden); Professor Earl W. Davie of the University of Washington (Seattle, Washington); Professor Russell F. Doolittle of the University of California San Diego and La Jolla California; Professor Craig M. Jackson of Washington University School of Medicine (St. Louis, Missouri); and Professor Sandor S. Shapiro of Jefferson Medical College, Thomas Jefferson University (Philadelphia, Pennsylvania). To each I extend my sincerest thanks for their kind and helpful cooperation. I appreciate greatly the help of Drew Likens, photographer at the Cardeza Foundation of Jefferson Medical College, for preparing excellent drawings and providing photographic prints of most of the figures in this chapter; and the assistance of Alice Mackov, Readers' Service Librarian at Thomas Jefferson University, with literature searches. Finally, I am most grateful to Dr. Herbert E. Spiegel, the editor of this volume, for the invitation to write this essay on blood coagulation, and for his help and encouragement along the way.

REFERENCES

Aggeler, P. M., White, S. G., Glendening, M. B., Page, E. W., Leake, T. B., and Bates, G., *Proc. Soc. Exp. Biol. Med.* **79**, 692–697 (1952).

Aledort, L. M., *Ann. N.Y. Acad. Sci.* **240**, 1–149, 349–426 (1975).

Alfors, K. E., Cockburn, J. S., and Gross, J. F., *Microvasc. Res.* **11**, 79–87 (1976).

Andersson, L.-O., Borg, H., and Miller-Andersson, M., *Thromb. Res.* **7**, 451–459 (1975).

Aronson, D. L., *in* "Proceedings of the International Workshop on Technology for Protein

Separation" (H. E. Sandberg, ed.), DHEW Publ. No. NIH-78-1422, pp. 345–357. USDHEW, Reston, Virginia, 1977.
Aronson, D. L., Mustafe, A. J., and Mushinski, J. T., *Biochim. Biophys. Acta* **188,** 25–30 (1969).
Austen, D. E. G., *Br. J. Haematol.* **27,** 89–100 (1974).
Bajaj, S. P., and Mann, K. G., *J. Biol. Chem.* **248,** 7729–7741 (1973).
Barrow, E. M., and Graham, J. B., *Physiol. Rev.* **54,** 23–74 (1974).
Bartlett, S., Latsom, P., and Hanahan, D. J., *Biochemistry* **19,** 273–277 (1980).
Bennett, B., *Semin. Hematol.* **14,** 301–317 (1977).
Bergström, K., and Wallén, P., *Ark. Kemi* **17,** 505–514 (1961).
Biggs, R., Douglas, A. S., Macfarlane, R. G., Dacie, J. V., Pitney, W. R., Mersky, C., and O'Brien, J. R., *Br. Med. J.* **2,** 1378–1388 (1952).
Block, R. J., and Keller, S., *Contrib. Boyce Thompson Inst.* **20,** 385–394, 1962.
Blombäck, B., *in* "Blood Clotting Enzymology" (W. H. Seegers, ed.) pp. 143–215. Academic Press, New York, 1967.
Blombäck, B., *In* "Plasma Proteins" (B. Blombäck and L. A. Hanson, eds.) pp. 223–253, Wiley, New York, 1979.
Blombäck, B., and Blombäck, M., *Ark. Kemi* **10,** 415–443 (1956).
Blombäck, R., and Blombäck M., *Ann. N.Y. Acad. Sci.* **202,** 77–97 (1972).
Blombäck, B., Blombäck, M., Edman, P., and Hessel, B., *Biochim. Biophys. Acta* **115,** 371–396 (1966).
Blombäck, B., Hessel, B., and Hogg, D. H., *Thromb. Res.* **8,** 639–658 (1976a).
Blombäck, B., Hogg, D. H., Gärlund, B., Hessel, B., and Kudryk, B., *Thromb. Res.* **8,** Suppl. II 329–346 (1976b).
Bloom, A. L., and Peake, I. R., *Br. Med. Bull.* **33,** 219–223 (1977).
Bohn, H., *Mol. Cell. Biochem.* **20,** 67–75 (1978).
Born, G. V. R., *Br. Med. Bull.* **33,** 193–198 (1977).
Brown, M. E., and Rothstein, F., *Science* **155,** 1017–1019 (1967).
Brozović, M., and Path, M. R. C., *Br. Med. Bull.* **33,** 231–237 (1977).
Butkowski, R. J., Elion, J., Downing, M. R., and Mann, K. G., *J. Biol. Chem.* **252,** 4942–4957 (1977).
Chen, R., and Doolittle, R. F., *Biochemistry* **10,** 4486–4491 (1971).
Chung, S. L., *Ann. N.Y. Acad. Sci.* **202,** 240–255 (1972).
Chung, S. L., Lewis, M. S., and Folk, J. E., *J. Biol. Chem.* **249,** 940–950 (1974).
Cochrane, C. G., and Wuepper, K. D., *J. Exp. Med.* **134,** 986–1004 (1971).
Cohn, E. J., *in* "Blood Cells and Plasma Proteins" (J. L. Tullis, ed.), pp. 3–17. Academic Press, New York, 1953.
Cohn, E. J., Strong, L. E., Hughes, W. L., Mulford, D. J., Ashworth, J. N., Melin, M., and Taylor, H. L., *J. Am. Chem. Soc.* **68,** 459–475 (1946).
Collen, D., Kudryk, B., Hessel, B., and Blombäck, B., *J. Biol. Chem.* **250,** 5808–5817 (1975).
Colman, R. W., *Prog. Hemostasis Thromb* **3,** 109–144 (1976).
Colman, R. W., and Wong, P. Y., *Thromb. Haemostasis* **38,** 751–775 (1977).
Colman, R. W., Mattler, L., Sherry, S., J. *Clin. Invest.* **48,** 11–22, 1969.
Colman, R. W., Bagdasarian, A., Talamo, R. C., Scott, C. F., Seavy, M., Guimaraes, J. A., Pierce, J. W., and Kaplan, A. P., *J. Clin. Invest.* **56,** 1650–1662 (1975).
Cox, A. C., and Hanahan, D. J., *Biochim. Biophys. Acta* **207,** 49–64 (1970).
Crawford, N., and Taylor, D. G., *Br. Med. Bull.* **33,** 199–206 (1977).
Davie, E. W., and Fujikawa, K., *Annu. Rev. Biochem.* **44,** 799–829 (1975).
Davie, E. W., and Hanahan, D. J., *in* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 3, pp. 421–597. Academic Press, New York, 1978.

Davie, E. W., and Ratnoff, O. D., *Science* **145,** 1310 (1964).
Davie, E. W., Fujikawa, K., Kata, H., and Legaz, M. E., *Ann. N.Y. Acad. Sci.* **240,** 34–42 (1975a).
Davie, E. W., Fujikawa, K., Legaz, M. E., and Kato, H., **2,** 65–77 *Cold Spring Harbor Conf. Cell Proliferation* (1975b).
Davie, E. W., Fujikawa, K., Kurachi, K., and Kisiel, W., *Adv. Enzymol.* **48,** 277–318 (1979).
Deutsch, E., Irsigler, K., and Lomoschitz, H. *Thromb. Diath. Haemorrh.* **12,** 12 (1964).
DiScipio, R. G., Hermodson, M. A., Yates, S. G., and Davie, E. W., *Biochemistry* **16,** 698–706 (1977).
DiScipio, R. G., Kurachi, K., and Davie, E. W., *J. Clin. Invest.* **61,** 1528–1538 (1978).
Doolittle, R. F., *Thromb. Diath. Haemorrh.* **39,** Suppl., 25–35 (1970).
Doolittle, R. F., *Adv. Protein Chem.* **27,** 1–109 (1973).
Doolittle, R. F., *in* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 2, pp. 110–161. Academic Press, New York, 1975.
Doolittle, R. F. Schubert, D., and Schwartz, S. A., *Arch. Biochem. Biophys.* **118,** 456–467 (1967).
Doolittle, R. F., Cassman, K. G., Cottrell, B. A., Friesner, S. J., Hucko, J. T., and Takagi, T., *Biochemistry* **16,** 1703–1709 (1977a).
Doolittle, R. F., Cassman, K. G., Cottrell, B. A., Friezner, S. J., and Takagi, T., *Biochemistry* **16,** 1710–1714 (1977b).
Doolittle, R. F., Watt, K. W. K., Cottrell, B. A., Strong, D. D., and Riley, M., *Nature* **280,** 464–468, (1979).
Duckert, F., *in* "Blood Coagulation, Hemorrhage and Thrombosis" (L. M. Tocantins and L. A. Kazal, eds.), pp. 133–137. Grune & Stratton, New York, 1964.
Enfield, D. L., Ericsson, L. H., Walsh, K. A., Neurath, H., and Titani, K., *Proc. Natl. Acad. Sci. U.S.A.* **72,** 16–19 (1975).
Esmon, C. T., *J. Biol. Chem.* **254,** 964–973 (1979).
Esmon, C. T., and Jackson, C. M., *J. Biol. Chem.* **249,** 7791–7797 (1974).
Esnouf, M. P., *Brit. Med. Bull.* **33,** 213–218 (1977).
Esnouf, M. P., and Tunnah, G. W., *Br. J. Haematol.* **13,** 581–590 (1967).
Esnouf, M. P., and Williams, W. J., *Biochem. J.* **84,** 62–72 (1962).
Fenton, J. W., Fasco, M. J., Stackrow, A. B., Aronson, D. L., Young, A. M., and Finlayson, J. S., *J. Biol. Chem.* **352,** 3587–3598 (1977).
Ferry, J. D., Katz, S., and Tinoco, I., Jr., *J. Polym. Sci.* **12,** 509–516 (1954).
Finlayson, J. S., *in* "Fibrinogen" (K. Laki, ed.), pp. 39–59. Dekker, New York, 1968.
Finlayson, J. S., *Ann. N.Y. Acad. Sci.* **202,** 149–154 (1972).
Finlayson, J. S., *Semin. Thromb. Hemostasis* **1,** 33–62 (1974).
Finlayson, J. S., and Mosesson, M. W., *Biochemistry* **2,** 42–52 (1963).
Folk, J. E., and Chung, S. I., *Adv. Enzymol.* **38,** 109–191 (1973).
Folk, J. E., and Chung, S. I., *Cold Spring Harbor Conf. Cell Proliferation* **2,** 157–170 (1975).
Fretto, L. J., Ferguson, E. W., Steinman, H. M., and McKee, P. A., *J. Biol. Chem.* **253,** 2184–2195 (1978).
Fujikawa, K., Legaz, M. E., and Davie, E. W., *Biochemistry* **11,** 4882–4891, (1972a).
Fujikawa, K., Legaz, M. E., and Davie, E. W., *Biochemistry* **11,** 4892–4899 (1972b).
Fujikawa, K., Thompson, A. R., Legaz, M. E., Meyer, R. G., and Davie, E. W., *Biochemistry* **12,** 4938–4945 (1973).
Fujikawa, K., Legaz, M. E. Kato, H., and Davie, E. W., *Biochemistry* **13,** 4508–4516 (1974a).

Fujikawa, K., Coan, M. H., Legaz, M. E., and Davie, E. W., *Biochemistry* **13,** 5290–5299 (1974b).
Fujikawa, K., Titani, K., and Davie, E. W., *Proc. Natl. Acad. Sci. U.S.A.* **72,** 3359–3363 (1975).
Fujikawa, K., Walsh, K. A., and Davie, E. W., *Biochemistry* **16,** 2270–2278 (1977a).
Fujikawa, K., Kurachi, K., and Davie, E. W., *Biochemistry* **16,** 4182–4188 (1977b).
Fujikawa, K., Heimark R. L., Kurachi, K., and Davie, E. W., *Biochemistry* **19,** 1322–1330 (1980).
Gaffney, P. J., *Br. Med. J.* **33,** 245–252 (1977a).
Gaffney, P. J., *in* "Haemostasis: Biochemistry, Physiology and Pathology," (Ogston, D. and Bennett, B., eds.). pp. 105–168. Wiley, New York, 1977b.
Gamgee, A., *J. Physiol. (London)* **2,** 145–157 (1879).
Gladner, J. A. *In* "Fibrinogen" (K. Laki, ed.) p. 87–115, Dekker, New York, 1968.
Gordon, N. R., and Shulman, N. R., *Ann. N.Y. Acad. Sci.* **240,** 79–83 (1975).
Gorman, J. J., Castaldi, P. A., and Shaw, D. C., *Biochim, Biophys. Acta* **439,** 1–16 (1976).
Green, D., *J. Lab. Clin. Med.* **77,** 153–158 (1971).
Green, D., *Ann. Clin. Lab. Sci.* **8,** 95–105 (1978).
Griffin, J. H., and Cochrane, C. G. *in* "Methods in Enzymology" (L. Lorand, ed.), Vol. 45, Part B, pp. 56–65. Academic Press, New York, 1976.
Hall, C. E., and Slayter, H. S., *J. Biophys. Biochem. Cytol.* **5,** 11–16 (1959).
Han, Y. N., Kato, H., Iwanage, S., and Suzuki, T., *FEBS Lett.* **63,** 197–200 (1976).
Hardisty, R. M., *Br. Med. Bull.* **33,** 207–212 (1977).
Hathaway, W. E., Belhasen, L. P., and Hathaway, H. S., *Blood* **26,** 521–532 (1965).
Heck, L. W., and Kaplan, A. P., *J. Exp. Med.* **140,** 1615–1630 (1974).
Hemker, H. C., Lindhout, M. J., and Vermeer, C., *Ann. N.Y. Acad. Sci.* **283,** 104–113 (1977).
Henschen, A. and Lottspeich, F., *Hoppe-Seyler's Z. Physiol. Chem.* **358,** 935–938 (1977a).
Henschen, A., and Lottspeich, F., *Hoppe-Seyler's Z. Physiol. Chem.* **358,** 1643–1646 (1977b).
Henriksen, R. A., and Jackson, C. M., *Semin. Thromb. Hemostasis* **1,** 284–309 (1975).
Hershgold, E. J., *Ann. N.Y. Acad. Sci.* **240,** 70–73 (1975).
Herzig, R. H., Ratnoff, O. D., and Shainoff, J. R., *J. Lab. Clin. Med.* **76,** 451–465 (1970).
Hewett-Emmett, D., Walz, D. A., Reuterby, J., McCoy, L. E., and Seegers, W. H., *Thromb. Res.* **7,** 227–234 (1975).
Hornstra, G., Haddeman, E., and Don, J. A., *Thromb. Res.* **12,** 367–374 (1978).
Hovig, T., *Thromb. Diath. Haemorr.* **9,** 264–269 (1963).
Inada, Y., Hessel, B., and Blombäck, B., *Biochim. Biophys. Acta* **532,** 161–170 (1978).
Jackson, C. M., *Biochemistry* **11,** 4873–4881 (1972).
Jackson, C. M., *Prog. Hematol.* **10,** 333–358 (1977).
Jackson, C. M., *Br. J. Haematol.* **39,** 1–8 (1978).
Jackson, C. M., and Hanahan, D. J., *Biochemistry* **7,** 4506–4517 (1968).
Jackson, C. M., Esmon, C. T., and Owen, W. G., *Cold Spring Harber Conf. Cell Proliferation* **2,** 95–109 (1975).
Jacobsen, E., and Kieruff, P., *Thromb Res.* **3,** 145–149 (1973).
James, H. L., Ganguly, P., and Jackson, C. W., *Thromb. Haemostasis* **38,** 939–954 (1977).
Jesty, J., and Nemerson, Y., *J. Biol. Chem.* **248,** 509–515 (1974).
Jesty, J., Spencer, A. K., and Nemerson, Y., *J. Biol. Chem.* **249,** 5614–5622 (1974).
Jesty, J., Maynard, J. R., Radcliffe, R. D., Silverberg, S. A., Pitlick, F. A., and Nemerson, Y., *Cold Spring Harbor Conf. Cell Proliferation* **2,** 171–190 (1975).
Johnson, A. J., Karpatkin, M. H., and Newman, J., *Br. J. Haematol.* **21,** 21–41 (1971).

Johnston, A. R., Ulevitch, R. J., and Ryan, K., *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **35,** 693 (1976).
Joist, J. H., Dolezel, G., Lloyd, J. V., Kinlough-Rathbone, R. L., and Mustard, J. T., *J. Lab. Clin. Med.* **84,** 474–482 (1974).
Kaplan, A. P., and Austen, K. F., *J. Immunol.* **105,** 802–811 (1970).
Kazal, L. A., *Prog. Hematol.* **3,** 293–341 (see p. 297) (1962).
Kazal, L. A., *Ann. Clin. Lab. Sci.* **1,** 139–154 (1971).
Kazal, L. A., Amsel, S., Miller, O. P., and Tocantins, L. M., *Proc. Soc. Exp. Biol. Med.* **113,** 989–994 (1963).
Kazame, M., McDonagh, J., Wagner, R. H., Langdell, R. O., and McDonagh, R. P., *Haemostasis* **5,** 329–340 (1976).
Kingdon, H. S., Davie, E. W., and Ratnoff, O. D., *Biochemistry* **3,** 166–173 (1964).
Kisiel, W., and Hanahan, D. J., *Biochim. Biophys. Acta* **304,** 103–113 (1973).
Kisiel, W., and Davie, E. W., *Biochemistry* **14,** 4928–4934 (1975).
Kisiel, W., Fujikawa, K., and Davie, E. W., *Biochemistry* **16,** 4189–4199 (1977).
Koid, T., Kato, H., and Davie, E. W., *Biochemistry* **16,** 2279–2286 (1977a).
Koid, T., Hermodson, M. A., and Davie, E. W., *Nature (London)* **266,** 729–730 (1977b).
Komiya, M., Nagasawa, S., and Suzuki, T., *J. Biochem. (Tokyo)* **72,** 1205–1218 (1972).
Komiya, M., Kato, H., and Suzuki, T., *J. Biochem. (Tokyo)* **76,** 811–822, 823–832, 833–845 (1974).
Koppel, G., *Z. Zellforsch. Mikrosk. Anat.* **77,** 443–517 (1967).
Koutts, J., Gude, N., and Firkin, B. G., *Thromb. Res.* **8,** 531–541 (1976).
Kurachi, K., and Davie, E. W., *Biochemistry* **16,** 5831–5839 (1977).
Kurachi, K., Fujikawa, K., and Davie, E. W., *Biochemistry* **19,** 1330–1338 (1980).
Laake, K., and Ellingsen, R., *Thromb. Res.* **5,** 539–545 (1974).
Lacombe, M. J., Varet, B., and Levy, J. P., *Blood* **46,** 761–770 (1975).
Laki, K., *Arch. Biochem. Biophys.* **32,** 317–324 (1951).
Laki, K., ed., "Fibrinogen," pp. 1–398. Dekker, New York, 1968.
Legaz, M. E., Weinstein, M. J., Heldebrant, C. M., and Davie, E. W., *Ann. N.Y. Acad. Sci.* **240,** 43–61 (1975).
Lerner, R. G., Goldstein, R., and Cummings, G., *Thromb. Res.* **11,** 253–261 (1977).
Lindhout, M. J., Kop-Klaasen, B. H., and Hemker, H. C., *Biochim. Biophys. Acta* **533,** 327–341, 342–354 (1978).
Lindquist, P. A., Fujikawa, K., and Davie, E. W., *J. Biol. Chem.* **253,** 1902–1909 (1978).
Loewy, A. G., *in* "Fibrinogen" (K. Laki, ed.) Dekker, New York, pp. 185–220,1968.
Loewy, A. G., *Thromb. Diath. Haemorrh. Suppl.* **39,** 103 (1970).
Loewy, A. G., *Ann. N.Y. Acad. Sci.* **202,** 41–76 (1972).
Loewy, A. G., Dunathan, K., Kriel, R., and Wolfinger, H. L., *J. Biol. Chem.* **236,** 2625–2633 (1961).
Lorand, L., *Nature (London)* **166,** 694 (1950).
Lorand, L., *Thromb. Diath. Haemorrh.* **3,** Suppl. 1, 238, (1962).
Lorand, L., *Ann. N.Y. Acad. Sci.* **202,** 6–40 (1972).
Lorand, L., *Cold Spring Harbor Conf. Cell Proliferation* **2,** 79–84 (1975).
Lorand, L., *in* "Methods in Enzymology" (L. Lorand, ed.), Vol. 45, Part B, pp. 31–303. Academic Press, New York, 1976.
Lorand, L., and Gotoh, T., *in* "Methods in Enzymology" C. G. E. Perlmann and L. Lorand, eds.), Vol. 19, pp. 770–774. Academic Press, New York, 1970.
Lundblad, R. L., Kingdon, H. S., and Mann, K. G., *Methods in Enzymology* **45,** 156 (1976).
McConnell, D. J., and Mason, B. *Br. J. Pharmacol.* **38,** 490–502 (1970).

McDonagh, J., McDonagh, R. P., Delage, J.-M., and Wagner, R. H., *J. Clin. Invest.* **48,** 940–946 (1969).
Macfarlane, R. G., *Nature (London)* **202,** 498–499 (1964).
Macfarlane, R. G., *in* "Human Blood Coagulation, Haemostasis and Thrombosis" (R. Biggs, ed.), pp. 1–31. Blackwell, Oxford, 1976.
Macfarlane, R. G., *Br. Med. Bull.* **33,** 183–185 (1977)
McKee, P. A., Andersen, J. C., and Switzer, M. E., *Ann. N.Y. Acad. Sci.* **240,** 8–33 (1975).
Magnusson, S., Sottrup-Jensen, L., Peterson, T. E., Morris, H. R., and Dell, A., *FEBS Lett.* **44,** 189–193 (1974).
Magnusson, S., Gottrup-Jensen, L., Petersen, T. E., and Claeys, H., *Boerhaave Ser. Postgrad. Med. Educ.* **10,** 25 (1975a).
Magnusson, S., Petersen, T. E. Gottrup-Jensen, L., and Claeys, H., *Cold Spring Harbor Conf. Cell Proliferation* **2,** 1281–1310 (1975b).
Magnusson, S., Petersen, T. E., Gottrup-Jensen, L., and Claeys, H., *Cold Spring Harbor Conf. Cell Proliferation* **2,** 123–149 (1975c).
Mandle, R., Jr., and Kaplan, A. P., *J. Biol. Chem.* **252,** 6097–6104 (1977).
Mann, K. G., *in* "Methods in Enzymology" (L. Lorand, ed.), Vol. 45, Part B, pp. 123–156. Academic Press, New York, 1976.
Marder, V. J., and Budzynski, A. Z., *Thromb. Diath. Haemorrh., Suppl.* **51,** 267–275 (1972).
Marder, V. J., and Budzynski, A. Z., *Thromb. Diath. Haemorrh.* **33,** 199–207 (1975a).
Marder, V. J., and Budzynski, A. Z., *Prog. Hemostasis Thromb.* **2,** 141–174 (1975b).
Marguerie, G., *Biochim. Biophys. Acta* **494,** 172–181 (1977).
Martinez, J., Holburn, R. R., Shapiro, S. S., and Erslev, A. J., *J. Clin. Invest.* **53,** 600–611 (1974).
Martinez, J., Palascak, J., and Peters, C., *J. Lab. Clin. Med.* **89,** 367–377 (1977).
Mason, R. G., Mohammad, S. F., Chuang, H. Y. K., and Richardson, P. D., *Semin. Thromb. Hemostasis* **3,** 98–116 (1976).
Mellanby, J., *Proc. Royal Soc. London B,* **113,** 93 (1933).
Miletich, J. P., Jackson, C. M., and Majerus, P. W., *Proc. Natl. Acad. Sci. U.S.A.* **74,** 4033–4036 (1977).
Moncada, S., Gryglewski, R., Bunting, S., and Vane, J. R., *Nature (London)* **263,** 663–665 (1976).
Morawitz, P., *Ergeb. Physiol.* **4,** 307 (1905) (translation by R. Hartman and P. F. Gunther, Thomas, Springfield, Illinois, 1958).
Morris, S., Robey, F. A., and Kosow, D. P., *J. Biol. Chem.* **253,** 4604–4608 (1978).
Mosesson, M. W., *Thromb. Haemostasis* **38,** 742–750 (1977).
Mosesson, M. W., and Finlayson, J. S., *Prog. Hemostasis Thromb.* **3,** 61–108 (1976).
Mosesson, M. W., and Sherry, S., *Biochemistry* **5,** 2829–2835 (1966).
Mosesson, M. W., and Unfleet, R. A., *J. Biol. Chem.* **245,** 5728–5736 (1970).
Mosesson, M. W., Alkjaersig, N., Sweet, B., and Sherry, S., *Biochemistry* **6,** 3279–3287 (1967).
Movat, H. Z., and Ozge-Anwar, A. H., *J. Lab. Clin. Med.* **84,** 861–869 (1974).
Murano, G., *Semin. Thromb. Hemostasis* **1,** 1–32 (1974).
Murano, G., *Am. J. Hematol.* **4,** 409–417 (1978).
Mustard, J. F., and Packham, M. A., *Br. Med. Bull.* **33,** 187–192 (1977).
Myrmel, K. H., Lundblad, R. L., and Mann, K. G., *Biochemistry* **15,** 1767–1773 (1976).
Nagasawa, S., Takahashi, H., Koide, M., and Suzuki, M., *Biochem. Biophys. Res. Commun.* **32,** 644–649 (1968).
Nelsestuen, G. L., *in* "Chemistry and Biology of Thrombin" (R. L. Lundblad *et al.*, eds.), pp. 291–296. Ann Arbor Sci. Publ., Ann Arbor, Michigan, 1977.

Nemerson, Y., and Pitlick, F. A., *Biochemistry* **9,** 5100–5105 (1970).
Nemerson, Y., and Pitlick, F. A., *Prog. Hemostasis Thromb.* **1,** 1–37 (1972).
Nesheim, M. E., and Mann, K. G., *J. Biol. Chem.* **254,** 1326–1334 (1979).
Nesheim, M. E., Myrmel, K. H., Hibbard, L., and Mann, K. G., *J. Biol. Chem.* **254,** 508–517 (1979).
Newman, J., Harris, R. B., and Johnson, A. J., *Nature (London)* **263,** 612–613 (1976).
Niemitz, J., *J. Clin. Invest.* **51,** 307–313 (1972).
Niewiarowski, S., Stuart, R. K., and Thomas, D. P., *Proc. Soc. Exp. Biol. Med.* **123,** 196–200 (1966).
Nilsson, I. M., and Pandolfi, M., *Thromb., Diath. Haemorrh., Suppl.* **40,** 231–242 (1970).
Nussenzweig, V., Seligmann, M., Pelmont, J., and Grabar, P., *Ann. Inst. Pasteur, Paris* **100,** 377–389 (1961).
Nyman, D., *Thromb. Res.* **11,** 433–438 (1977).
Ogston, D., and Bennett, B., *Br. Med. Bull.* **34,** 107–112 (1978).
Østerud, B., and Flengsrud, R., *Biochem. J.* **145,** 469–474 (1975).
Østerud, B., Berre, Å., Otnaess, A. B., Bjorklid, E., and Prydz, H., *Biochemistry* **11,** 2853–2857 (1972).
Østerud, B., Laake, K., and Prydz, H., *Thromb. Diath. Haemorrh.* **33,** 553–563 (1975).
Owen, C. A., Jr., Bowie, E. J. W., Didisheim, P., and Thompson, J. H., "The Diagnosis of Bleeding Disorders", Little, Brown, Boston, Massachusetts, 1969.
Owen, W. G., Esmon, C. T., and Jackson, C. M., *J. Biol. Chem.* **249,** 594–605, (1974).
Owren, P. A., *Biochem. J.* **43,** 136–146 (1948).
Owren, P. A., Rapaport, S. I., Hjort, P., and Aas, K., *Sang* **25,** 752–757 (1954).
Papahadjopolous, D., and Hahahan, D., *Biochim. Biophys. Acta* **90,** 436–442 (1964).
Patek, A. J., Jr., and Taylor, F. H. L., *J. Clin. Invest.* **16,** 113–124 (1937).
Pavlovsky, A., *Blood* **2,** 185–195 (1947).
Pennell, R. B., *in* "The Plasma Proteins" (F. W. Putnam, ed.), Vol. 1, pp. 9–42. Academic Press, New York, 1960.
Pitlick, A. F., *Biochim. Biophys. Acta* **428,** 27–34 (1976).
Pitlick, A. F., and Nemerson, Y., *Biochemistry* **9,** 5105–5113 (1970).
Pitlick, A. F., Nemerson, Y., Gottlieb, A. J., Gordon, R. G., and Williams, W. J., *Biochemistry* **10,** 2650–2657 (1971).
Pizzo, S. V., Schwartz, M. L., Hill, R. L., and McKee, P. A., *J. Biol. Chem.* **248,** 4574–4583 (1973).
Polson, A., and Ruiz-Bravo, C., *Vox Sang.* **23,** 107–118 (1972).
Pool, J. G., Hershgold, E. J., and Pappenhagen, A. R., *Nature (London)* **203,** 312 (1964).
Poon, M.-C., and Ratnoff, O. D., *Blood* **50,** 367–376 (1977).
Prowse, C. V., Mattock, P., Esnouf, M. P., and Russell, A. M., *Biochim. Biophys. Acta* **434,** 265–279 (1976).
Quick, A. J., "Hemorrhagic Diseases and the Physiology of Hemostasis", Thomas, Springfield, Illinois, 1942.
Quick, A. J., "Hemorrhagic Diseases and Thrombosis." Lea & Febiger, Philadelphia, Pennsylvania, 1946.
Radcliffe, R. D., and Barton, P. G., *J. Biol. Chem.* **248,** 6788–6795 (1973).
Radcliffe, R. D., and Nemerson, Y., *J. Biol. Chem.* **250,** 388–395 (1975).
Radcliffe, R. D., and Nemerson, Y., *J. Biol. Chem.* **251,** 4797–4802 (1976).
Radcliffe, R. D., Bagdasarian, A., Colman, R., and Nemerson, Y., *Blood* **50,** 611–617 (1977).
Ratnoff, O. D., *in* "Hemostasis: Biochemistry, Physiology and Pathology" (D. Ogston and B. Bennett, eds.) p. 25. Wiley, New York, 1977a.
Ratnoff, O. D., *J. Lab. Clin. Med.* **89,** 1159–1162 (1977b).

Ratnoff, O. D., *Ann. Int. Med.* **88,** 403–409 (1978).
Ratnoff, O. D., and Colopy, J. E., *J. Clin. Invest.* **34,** 602–613 (1955).
Ratnoff, O. D., and Davie, E. W., *Biochemistry* **1,** 677–687 (1962).
Ratnoff, O. D., Davie, E. W., and Mallet, D. L., *J. Clin. Invest.* **40,** 803–813 (1961).
Reich, E., Rifkin, D. B., and Shaw, E., eds., "Proteases and Biological Control," Vol. 2. Cold Spring Harbor Lab., Cold Spring Harbor, New York, 1975.
Reuterby, J., Walz, D. A., McCoy, L. E., and Seegers, W. H., *Thromb. Res.* **4,** 885–890 (1974).
Revak, S. D., and Cochrane, C. G., *J. Clin. Invest.* **57,** 852–860 (1976).
Revak, S. D., Cochrane, C. G., Johnson, A. R., and Hugli, T. E., *J. Clin. Invest.* **54,** 619–627 (1974).
Revak, S. D., Cochrane, C. G., and Griffen, J. H., *J. Clin. Invest.* **59,** 1167–1175 (1977).
Rick, M. E., and Hoyer, L. W., *Br. J. Haematol.* **38,** 107–119 (1978).
Rosenthal, R. L., Dreskin, O. H., and Rosenthal, N., *Proc. Soc. Exp. Biol. Med.* **82,** 171–175 (1953).
Saito, H., Ratnoff, O. D., Marshall, J. S., and Pensky, J., *J. Clin. Invest.* **52,** 850–861 (1973).
Saito, H., Ratnoff, O. D., and Donaldson, V. H., *J. Lab. Clin. Med.* **84,** 62–72 (1974).
Saito, H., Ratnoff, O. D., Waldman, R., and Abraham, J. P., *J. Clin. Invest.* **55,** 1082–1089 (1975).
Samuelsson, B., Hamberg, M., Malsten, C., and Svensson, J., *Adv. Prostaglandin Thromboxane Res.* **2,** 737–746 (1976).
Scheraga, H. A., and Laskowski, M., Jr., *Adv. Protein Chem.* **12,** 1–131 (1957).
Schiffman, S., and Lee, P., *Br. J. Haematol.* **27,** 101–105 (1974).
Schiffman, S., and Lee, P., *J. Clin. Invest.* **56,** 1082–1092 (1975).
Schiffman, S., Rapaport, S. I., Ware, A. G., and Mehl, J. W., *Proc. Soc. Exp. Biol. Med.* **105,** 453–458 (1960).
Schiffman, S., Rapaport, S. I., and Patch, M. J., *Blood* **22,** 733 (1963).
Schiffman, S., Pecci, R., and Lee, P., *Thromb. Res.* **10,** 319–323 (1977).
Schmidt, A., *Pfluegers Arch. Ges. Physiol.* **11,** 515 (1875).
Schoenmaker, J. G. G., Matze, R., Haanen, C., and Zilliken, F., *Biochim. Biophys. Acta* **101,** 166–176 (1965).
Schwartz, M. L., Pizzo, S. V., Hill, R. L., and McKee, P. A., *J. Biol. Chem.* **246,** 5851–5854 (1971).
Schwartz, M. L., Pizzo, S. V., Hill, R. L., and McKee, P. A., *J. Biol. Chem.* **248,** 1395–1407 (1973).
Seegers, W. H., *J. Biol. Chem.* **136,** 103–111 (1940).
Seegers, W. H., "Prothrombin." Harvard Univ. Press, Cambridge, Massachusetts, pp 1–728, 1962.
Seegers, W. H., *in* "Blood Clotting Enzymology" (W. H. Seegers, ed.), pp. 1–21. Academic Press, New York, 1967.
Seegers, W. H., and Murano, G., *Pol. Arch. Med. Wewn.* **55,** 1–15 (1976).
Seegers, W. H., McClaughry, R. I., and Fahey, J. L., *Blood* **5,** 421–433 (1950).
Seegers, W. H., Hassouna, H. I., Hewett-Emmett, D., Walz, D. A., and Andary, T. J., *Semin. Thromb. Hemostasis* **1,** 211–283 (1975).
Shapiro, S. S., and McCord, S., *Prog. Hemostasis Thromb.* **4,** 177–209 (1978).
Shaprio, S. S., and Hultin, M., *Semin. Thromb. Hemostasis* **1,** 336–385 (1975).
Shulman, S., *J. Am. Chem. Soc.* **75,** 5846–5857 (1953).
Shrager, R. I., Mihalyi, E., and Towne, D. W., *Biochemistry* **15,** 5382–5386 (1976).
Silverberg, S. A., Nemerson, Y., and Zur, M., *J. Biol. Chem.* **252,** 8481–8488 (1977).
Smith, C. M., and Hanahan, D. J., *Biochemistry* **15,** 1830–1838 (1976).

Smith, E. B., Alexander, K. M., and Massie, I. B., *Adv. Exp. Med. Biol.* **82,** 872–877 (1977).
Sodetz, J. M., Pizzo, S. V., and McKee, P. A., *J. Biol. Chem.* **252,** 5538–5546 (1977).
Speer, R. J., Ridgeway, H., and Hill, J. M., *Thromb. Diath. Haemorrh.* **14,** 1–11 (1965).
Stenflo, J., *Adv. Enzymol.* **46,** 1–31 (1978).
Stenflo, J., and Suttie, J. W., *Annu. Rev. Biochem.* **46,** 157–172 (1977).
Sussman, I. I., and Weiss, H. J., *Thromb. Res.* **9,** 267–276 (1976).
Sussman, I. I., Rosner, W., and Weiss, H. J., *Am. J. Physiol.* **230,** 434–440 (1976).
Suttie, J. W., and Jackson, C. M., *Physiol. Rev.* **57,** 1–70 (1977).
Takagi, T., and Doolittle, R. F., *Biochemistry* **13,** 750–756 (1974).
Takagi, T., and Konishi, K., *Biochim. Biophys. Acta* **271,** 363–370 (1972).
Takahashi, H., Nagasawa, S., and Suzuki, T., *FEBS Lett.* **24,** 98–100 (1972).
Thompson, A. R., Enfield, D. L., Ericsson, L. H., Legaz, M. E., and Fenton, J. W., II, *Arch. Biochem, Biophys.* **178,** 356–367 (1977).
Tocantins, L. M., *Surgery* **27,** 465–466 (1950).
Tocantins, L. M., *Adv. Med. Surg.* **27,** 205 (1952).
Todd, A. S., *Br. Med. Bull.* **20,** 210–212 (1964).
Tooney, N. M., and Cohen, C., *Nature (London)* **237,** 23 (1972).
Tullis, J. L., "Clot". Thomas, Springfield, Illinois, 1976.
Turner, S. R., Tainter, J. A., and Lynn, W. S., *Nature (London)* **257,** 680–681 (1975).
Ulevitch, R. J., Cochrane, C. G., Revak, S. D., Morrison, D. C., and Johnston, A. R., *Cold Spring Harbor Conf. Cell Proliferation* **2,** 85–94 (1975).
Varadi, K., and Hemker, H. C., *Thromb. Res.* **8,** 303–317 (1976).
Vehar, G. A., and Davie, E. W., *Science* **197,** 374–376 (1977).
Vehar, G. A., and Davie, E. W., *Biochemistry* **19,** 401–410 (1980).
Vennerød, A. M., Ørstavik, K. H., Laake, K., Fagerhol, M., and Ly, B., *Thromb. Res.* **11,** 663–672 (1977).
Vermylen, J., *Pathol. Biol.* **23,** 5–10 (1975).
Waaler, B. A., *Scand. J. Clin. Lab. Invest., Suppl.* **37,** 1 (1959).
Walker, L., and Catlin, A., *Thromb. Diath. Haemorrh.* **26,** 99–102 (1971).
Wallin, R., and Prydz, H., *Biochem. Biophys. Res. Commun.* **62,** 398–406 (1975).
Walsh, P. N., *Br. J. Haematol,* **22,** 393–405 (1972).
Walz, D. A., Hewett-Emmett, D., and Seegers, W. H., *Proc. Natl. Acad. Sci. U.S.A.* **74,** 1969 (1977a).
Walz, D. A., Hewett-Emmett, D., and Seegers, W. H., *Life Sci.* **20,** 79–85 (1977b).
Ware, A. G., and Seegers, W. H., *J. Biol. Chem.* **172,** 669 (1948).
Ware, A. G., Guest, M. M., and Seegers, W. H., *Arch. Biochem. Biophys.* **13,** 231–236 (1947).
Watt, K. W. K., Takagi, T., Doolittle, R. F., *Biochemistry* **18,** 68–76 (1979).
Weksler, B. B., and Coupal, C. E., *J. Exp. Med.* **137,** 1419–1430 (1973).
Wickerhauser, M., *Thromb. Diath. Haemorrh., Suppl.* **43,** 165–175 (1971).
Workman Jr., E. F., and Lundblad, R. L., *Thromb. Haemostasis* **39,** 193–200 (1978).
Wright, I. S., *Thromb. Diath. Haemorrh.* **7,** 381–388 (1962).
Wu, V. Y., and McCoy, L. E., *Thromb. Res.* **11,** 581–593 (1977).
Wuepper, K. D., *in* "Inflammation, Mechanisms and Control" (I. N. Lepow and P. A. Ward, eds.)., pp. 93–117. Academic Press, New York, 1972.
Wuepper, K. D., and Cochrane, C. G., *Proc. Soc. Exp. Biol. Med.* **141,** 271–276 (1972).
Yue, R. H., and Gertler, M. M., *Biochim. Biophys. Acta* **490,** 350–362 (1977).
Zeldis, S. M., Nemerson, Y., Pitlick, F. A., Lentz, T. L., *Science* **175,** 766–768 (1972).

5

Inborn Errors of Metabolism

MASOOD HAIDER

CLINICAL BIOCHEMISTRY
Contemporary Theories and Techniques, Vol. 2

ISBN 0-12-657102-3

I. INTRODUCTION AND HISTORICAL BACKGROUND

Inborn errors of metabolism (IEM) comprise a large number of relatively rare biochemical disorders due to defects in the structure or function of protein molecules caused by mutant genes. During the last two decades, a great deal of attention has been paid to the elucidation of the molecular mechanisms underlying these disorders and to the prenatal diagnosis, genetic screening, and clinical management of these diseases. Although most of IEM are rare to uncommon, taken as a group, they may affect approximately 1.0% of newborn infants, causing significant morbidity and mortality in about one-third of the affected newborns. Moreover, aside from the clinical impact of understanding the molecular basis of an IEM in terms of providing a rational basis for treatment, the study of IEM is of great value since it gives us useful information about defining normal biochemical pathways and their genetic control. For example, the study of hemoglobinopathies has contributed greatly to our understanding of the hemoglobin structure. The identification of enzymatic defects in disorders of amino acid metabolism, such as phenylketonuria and alcaptonuria, of carbohydrate metabolism, such as galactosemia and pentosuria, and of ganglioside metabolism, and also blood clotting disorders (to give just a few more examples), have led to the crucial details of the respective normal pathways. Important genetic concepts have also been derived from the study of IEM, such as the recognition that these disorders represent only a small portion of the larger number of mutations that constantly take place in the genetic material, but do not cause any clinical effects. The discovery and understanding of IEM, which has been increasing at a rapid rate and will continue to do so in the future, has been possible due to progress in the multiple disciplines of biochemistry, molecular biology, genetics, and clinical chemistry. In order to appreciate the impact of these developments on the various aspects of IEM, a brief historical background is worthwhile.

The concept of IEM was put forward by A. E. Garrod in 1908 in Croonian lectures (*1*) and in his monograph on the subject in 1909. It was based on his studies of albinism, cystinuria, pentosuria, and, most significantly, on his work with alcaptonuria (*2*). Garrod found that patients with alcaptonuria excreted homogentisic acid in their urine due to a congenital failure of the body to metabolize aromatic structures. He also noted that the

condition was found in one or more siblings of the patients, and nearly 50% of the patients with the disease were the offspring of first-cousin marriages. The disorder was explicable as an inheritable recessive condition by the recently rediscovered Mendelian laws of heredity. Garrod clearly recognized that the disorders were caused by metabolic block as a result of enzyme deficiencies. The enzymatic deficiency in alcaptonuria was elucidated in 1958 by LaDu and co-workers (*3*), who found virtually nonexistent activity of homogentisic acid oxidase in the liver and kidney of patients with alcaptonuria. The enzymatic deficiency in pentosuria was established in 1970 (*4*). Only in cystinuria a different mechanism for the disorder than the one postulated by Garrod has been found (*5*).

The formulation of the concept of IEM established the implicit direct relationship between genes and enzymes. However, it was the "one gene–one enzyme" hypothesis enunciated by Beadle (*6*) and Tatum (*7*), based on their studies on *Drosophila* and *Neurospora*, which established the genetic control of biochemical processes and the relationship between gene and enzyme. The one gene–one enzyme hypothesis has been revised in recent years to include proteins that are not enzymes and complex proteins made up of nonidentical polypeptide subunits. Also, the original hypothesis is now called the "one cistron–one polypeptide" hypothesis and the term *cistron* refers to the linear sequence of nucleotides in DNA controlling the structure of a single polypeptide chain.

The discovery in 1949 of an electrophoretic difference between normal hemoglobin and hemoglobin from sickle cell anemia patients (*8*) by Pauling and co-workers laid the groundwork for the understanding of the inherited metabolic diseases at a molecular level. This landmark work elucidated the fact that diseases may result from mutations affecting structural as well as enzymatic protein. The term *molecular disease* was coined by Pauling and associates to describe a disease in which a structural alteration in a macromolecule leads to a special functional change, causing the diseased state. Ingram (*9*) demonstrated that the difference in electrophoretic behavior is the result of a single amino acid substitution in the β chain of hemoglobin, thus providing the first example of a point mutation. The extensive work on hemoglobinopathies has contributed greatly to our present understanding of biochemical genetics in man.

In 1952, the biochemical basis of an IEM was reported for the first time by Cori and Cori (*10*), who demonstrated the deficiency of glucose-6-phosphatase in the liver of patients with glycogen storage disease. This was followed by the report of Jervis (*11*) showing the absence of phenylalanine hydroxylase in the liver of patients with phenylketonuria. The variant form of an enzyme deficiency was documented by Marks (*12*) for glucose-6-phosphate dehydrogenase. Since then, there has been very rapid progress in the field. McKusick (*13*) has catalogued over 2800 variants of

conditions which are genetically determined and, by this time, over 500 enzyme deficiencies including the variant forms have been shown. This number will certainly grow, both for enzyme defects to be identified as well as for variant forms, as minor clinical symptoms and biochemical abnormalities in the pattern of circulating and excreted metabolites are more easily recognized. The variant forms may be clinically similar and metabolically related disorders, involving deficiencies of different enzymes. These variants may be caused by mutations in different genes known as *genocopies* or by alterations in the same enzyme due to different mutations at a single locus or *allelic series*.

The literature on IEM is voluminous and increasing at a rapid pace. A comprehensive treatment of various inherited metabolic disorders is available in reference texts [e.g., by Stanbury *et al.* (*14*) and by Bondy and Rosenberg (*15*)]. Several excellent reviews and books on the subject have recently become available (*16–19*). The literature on selected topics in the area of inherited metabolic disorders until 1978 has been reviewed by Öckerman and Lundblad (*20*). For the purpose of this chapter, emphasis will be placed on advances in screening, prenatal diagnosis and therapy, and laboratory methods used in the study of IEM. A few of the conditions, especially those which are now considered to be treatable, will be considered in detail.

II. MENDELIAN INHERITANCE: MODES OF GENETIC TRANSMISSION

A detailed treatment of the principles of Mendelian inheritance and the current concepts of the principles of biochemical genetics is beyond the scope of this chapter. The topics are covered comprehensively in textbooks on metabolic disorders (*14,15*) and molecular biology (*21*). However, in order to properly understand the subject of IEM, it is necessary to briefly discuss our present understanding of the basic principles of these two areas.

Stated at the simplest level, the genotype (or genetic constitution of an organism) determines its appearance, or the phenotype. In case of IEM, the gene dosage and chromosomal location of the mutant gene will decide the characteristic genetic pattern of the inheritable disorders and traits. In humans, the somatic cell contains 22 pairs of autosomes and one pair of sex chromosomes, and the location of a mutant gene may be on either type of chromosome.

The effect of one member of a pair of genes on the phenotype may be referred to as dominant when it is greater than that of the other. The gene

may be fully dominant, if the effect is expressed irrespective of the other member of the gene pair, or incompletely dominant, in which case it promotes an effect that it cannot fully achieve. Genes determining recessive effects do not express themselves at the level chosen for ascertainment. The terms *dominant* and *recessive* are arbitrary and refer to phenotypic expression rather than to the gene itself. The assignment of dominance will depend on the phenotype measured; the same condition may be called dominant or recessive, depending on the criteria used. For example, sickle cell anemia may be recessive, as judged from the clinical picture, but dominant when judged by a positive sickling test. It has been suggested (*22*) that the terms *homozygous* or *heterozygous* and *single dose* or *double dose* are preferable to dominant and recessive when referring to mutations that result in specific chemical pathology. In the heterozygous state, only a single mutant gene is present, whereas if the mutant gene is present in double dose, thus detectable phenotypically in only the homozygous state, it is most often recessive. Males in whom a recessive gene on a single X chromosome produces a given trait or disorder are referred to as hemizygous. The term *codominant* is used to describe the traits that are jointly expressed in the heterozygotes. Two other terms commonly used are *expressivity* and *penetrance*, which refer to the degree or frequency at which the effects are shown in an individual or a population, respectively.

In contrast to the use of the terms dominant and recessive, when referring to mutants that result in specific chemical pathology, the terminology has a very definite meaning when it pertains to patterns of inheritance of specific traits in humans. A knowledge of patterns of inheritance is of great value in understanding the mechanism of inheritance of IEM. Such information is an important step for the proper genetic counseling of families with these disorders.

There are four main Mendelian patterns of inheritance in man; *autosomal dominant, autosomal recessive, X-linked dominant,* and *X-linked recessive*. These are discussed below.

A. Autosomal Dominant Inheritance

The effect of a dominant gene will be recognized in both the homozygous or heterozygous state, but a patient with a disease caused by such a gene will usually be heterozygous, having received the gene from one affected parent. There is a 50% chance that the child of an affected heterozygote and normal parent will be affected by the disease. The characteristics of recognition of autosomal dominant inheritance are the following: (1) one of the parents of the diseased individual must have the disorder,

transmitting it directly to the child, yielding a vertical pattern of inheritance, unless the abnormal gene is caused by mutation; (2) the disease is transmitted by the affected individual to approximately half of his or her offspring, regardless of sex; (3) normal parents do not transmit the abnormality to their children. Another feature of dominant traits is wide variability in the expression of traits and symptoms (*expressivity*), which may be so low as to make it impossible to detect the gene by available means (*nonpenetrance*).

Patients with dominant disease develop clinical signs later in life, and show a less severe prognosis than with a recessive disorder. Very few autosomal dominant diseases occur in the homozygous state, which requires that both parents be affected with the disease. Inborn errors of metabolism showing this mode of inheritance are congenital spherocytosis, acute intermittent porphyria, and hereditary angioneurotic edema (*22*).

B. Autosomal Recessive Inheritance

Autosomal recessive inheritance requires the presence of the mutant gene in a double dose or homozygous state, and the affected individual inheriting one mutant gene from each parent. In recessive disorders, both parents are heterozygotes, carrying the gene for the trait, but generally show no clinical abnormality.

The chance of transmiting the mutant gene is 50% for each of the heterozygous parents. The offspring of two heterozygous individuals will have a 1:4 chance of inheriting the disorder and a 1:2 chance of being heterozygotes. Autosomal recessive disorders follow a horizontal pattern of inheritance and, in general, occur with equal frequency in males and females. Cystinuria, galactosemia, and cystic fibrosis are some of the diseases which follow this mode of inheritance.

C. X-Linked Recessive Inheritance

A disorder inherited as an X-linked recessive is expressed by all males who carry the single mutant gene, since males are hemizygous for all X-linked traits, and by females who are homozygous, i.e., carry the mutation on both X chromosomes. Thus, a distinctive feature of this mode of inheritance is that the incidence of the trait will be much higher in males than in females. An affected male will transmit the trait to all of his daughters, but to none of his sons. Females who are homozygous for the trait will transfer the trait to all of their sons and all of their daughters, but only the sons will show clinical signs. The disorder is never transferred directly

from father to son, thus causing an oblique pattern of transmission. The distinctive features are that (1) male-to-male transmission does not occur, (2) all daughters of an affected father will be affected, and (3) there is an excess of affected females.

D. X-Linked Dominant Inheritance

In this unique mode of inheritance, affected females who are heterozygous transmit the disorder to one-half of their sons and one-half of their daughters, thus resembling the pattern of autosomal dominant inheritance. The affected male, on the other hand, transmits the disorder to all of his daughters and none of his sons. Thus, X-linked dominant inheritance cannot be distinguished from autosomal dominant inheritance by progeny of affected females, but rather only by the progeny of affected males.

Most of the IEM are inherited as autosomal recessive traits. This means that clinically detectable disorders occur only when the mutant gene is present in double dose. The reason why no significant clinical abnormality is present in heterozygotes is that most enzymes are very likely to be present far in excess of that required. Thus, the effect of reduction of enzyme activity to one-half due to heterozygosity is compensated for by the enzyme's presence in excess amounts. For those IEM which are inherited as autosomal dominant traits, the enzymes responsible for the disorder may be synthesized in such amounts that reduction of the activity to one-half may affect the rate of enzyme reaction, and may lead to a block in a metabolic pathway.

In regard to the inheritance pattern of the X-linked traits, the activity of some enzymes, such as glucose-6-phosphate dehydrogenase and hypoxanthine–guanine phosphoribosyltransferase (the genes of which reside on the X chromosome) was found to be equal in men and women (*23,24*), in which case the expected activity should be twice for the females, for they have two X chromosomes. These observations can be explained by the *Lyon hypothesis* or *principle,* enunciated by Mary Lyon (*25*), as a result of her studies of X-linked color mutation in mice. This principle postulates that in the female one of the X chromosomes is inactivated in early embryonic life, and the selection of a survivor is a random process. The Lyon principle explains the presence of the *Barr body:* a dense clump of chromatin in the nucleus as a degenerate X chromosome. As a result of lyonization, in the female carrier of an X-linked disorder there should be two kinds of cells—those with a completely normal phenotype and those with a mutant phenotype. Two populations of cells have in fact been found in carriers of many X-linked metabolic diseases (*26–28*).

III. GENE MAPPING

In recent years, a number of techniques have contributed to the rapid progress in assigning specific genes to individual chromosomes. For X-linked traits, the characteristic pattern of inheritance has made it possible to assign more than 100 genes to the X chromosome (*13*), and to map it using proximity of certain genes to others on this chromosome. Using the technique of interspecific, human–animal somatic-cell hybridization in tissue culture has resulted in rapid advances in the definition of autosomal linkages in man, and the assignment of more than 100 genes to specific human autosomes. This approach should enable us to determine the chromosomal localization of a number of inborn errors that are expressed in cell culture. Use of human cells containing chromosomes marked by specific rearrangements is making it possible even to place loci on a specific portion of the autosome. A list of the gene loci established to be on specific chromosomes is provided by Bondy and Rosenberg (*15*) and in Stanbury *et al.* (*14*).

IV. BIOCHEMICAL GENETICS

The current knowledge of the mechanism of genetic control of protein synthesis has been derived mainly from studies with microbial systems. Despite the important differences in gene structure and the cellular regulatory aspects of protein synthesis in microbial and animal systems, much of our understanding about the genetic mechanisms for IEM is also based on microbial systems.

In the simplest terms, the sequence of amino acids in the protein is specified by the triplet sequence of nucleotides in the structural gene, which constitutes the genetic code and comprises a large fraction of the total DNA of the cell. The biosynthetic steps involved in the conversion of genotype of phenotype appear to be relatively simple. Under the action of the enzyme RNA polymerase, using a strand of DNA as a template, a replica RNA or mRNA is synthesized, which passes from the nucleus to the cytoplasm, is bound to the ribosomes, and serves in the protein synthetis machinery. In eukaryotic cells, a large-molecular-weight RNA (H_nRNA) is first produced, most of which remains in the nucleus and is degraded, while a small fraction of the long sequence, reduced in length, emerges into the cytoplasm as mRNA. The conversion of the genetic message from the structural gene in the nucleus into a utilizable cytoplasmic form (mRNA) is referred to as *transcription,* whereas the conversion of this information in protein synthesis is called *translation.* In the cyto-

plasm, activated amino acids are attached to their specific tRNAs, which bear the triplet of nucleotides complimentary to mRNA or anticodon, and are assembled on the ribosomes, which have the mRNA bound to them to synthesize proteins. Initiator and terminator codons in mRNA, as well as initiator and release factors, are involved in signaling the beginning and end of the polypeptide chain synthesis.

The conceptual basis of the regulatory mechanisms underlying the genetic control of protein synthesis was provided by the model proposed by Jacob and Monod in 1961. In this model, the synthesis of mRNA by the structural gene is regulated by two other types of genes—the *operator* and *regulator genes*. Regulator genes produce a repressor that interacts with DNA at the operator gene locus and shuts off the synthesis of mRNA by the adjacent structural gene. A group of contiguous operator genes and the structural gene is referred to as an *operon*.

The mechanism of the genetic control of protein synthesis in the eukaryotic cell is largely unknown and is, perhaps, a very complex process, involving controls, both at the transcriptional as well as on the translational levels. Attempts have been made to apply the regulatory concepts of Jacob and Monod to humans, for example, in explaining the normal postnatal change of fetal hemoglobin synthesis to adult hemoglobin synthesis (*29*). Originally, it was suggested (*22*) that structural gene mutations lead to an abnormality in the amino acid sequence of a protein, whereas those diseases in which the function of an enzyme is absent are caused as a result of controller gene mutations. However, it is, at the present time, difficult to justify the existence of a metabolic disease as a consequence of a controller gene defect, since the evidence in support of such a mechanism is not convincing.

A. Point Mutations

Most of the IEM are caused by point mutations that take place as a result of changes affecting a single nucleotide. Mutations in which a single base in DNA has been replaced by another are referred to as *replacement mutations*. The substitution of a purine base for a purine, or a pyrimidine base for a pyrimidine, is called *transition,* whereas a purine–pyrimidine or pyrimidine–purine substitution is called a *transversion*.

Point mutations may be "mis-sense" or "no-sense," depending upon the kind of change in the structural gene. A large number of structurally altered enzymes that characterize many inborn errors are caused by mis-sense mutations, resulting in the substitution of a different amino acid at a particular site in the affected protein. Nonfunctional polypeptides generally result as a consequence of no-sense mutation produced in a number

of ways. Deletion or insertion of a single base in the nucleotide sequence will lead to the shifting of the triplet codon reading frame, causing a "frame shift" no-sense point mutation. Insertions can also lead to the generation of a no-sense codon, causing premature chain termination and, thus, an incomplete peptide chain. In general, mis-sense mutations are cross-reactive material (CRM)-positive, whereas no-sense mutations are CRM-negative.

It is also possible for a mutation to change the rate of synthesis of a polypeptide rather than its primary structure. Some of the amino acids may be coded by two or more triplets (referred to as *degeneracy*). Substitution of one triplet for the other as a result of mutation will cause no effect if tRNAs complimentary to these codons are present in excess. However, the rate of synthesis of the polypeptide would be affected if the tRNA to the new codon varied in amount from the tRNA in the wild-type gene.

A single-step base substitution in a structural gene may also result in an increased synthesis of a variant protein. For example, one human variant of glucose-6-phosphate dehydrogenase (G6PD Hektoen), which differs from normal glucose-6-phosphate dehydrogenase in a single amino acid substitution, is associated with a fourfold increase in the rate of enzyme synthesis.

B. Gross Chromosomal Changes

The genetic abnormalities as a result of gross chromosomal errors constitute a significant percentage of all births. Approximately 5% of all fetuses may carry the chromosome errors due to faulty meiotic division, and the number of birth defects caused in this way may be far greater than the genetic abnormalities due to point mutations. The abnormalities in this group may be due to errors in chromosome number resulting from faulty chromosome segregation or to defects in chromosome structure due to chromosome breakage. Gross structural change leads to most chromosomal abnormalities such as Down's syndrome, Klinefelter's syndrome, or Turner's syndrome characterized by phenotypic abnormalities. However, no useful information for the study of IEM has been derived from such cytogenetic disorders.

V. CONSEQUENCES OF INBORN ERRORS OF METABOLISM

The normal biological role of a protein will determine the effect of a genetic alteration in the quality or quantity of that protein. Although inborn errors have been described for all types of proteins, the majority of disor-

ders cause an enzyme defect or derangement in the control of reaction rates. Abnormalities of cell membrane function, transport processes, or deficiency of nuclear receptors represent other types of metabolic disorders. For an exhaustive list of the disorders belonging to each such category, the reader is referred to Rosenberg (*22*).

In those disorders in which a metabolic block is present, the quantity of the normal product of the reaction concerned and its metabolites will be affected, and an accumulation of the precursors of the reaction may also result. The accumulated precursors may show concentration in body fluids or increased excretion in urine if the precursor is water-soluble, or it may be stored, if water-insoluble. Accumulation of precursors or metabolites is the cause of storage diseases, such as disorders in lipid, mucopolysaccharide, and glycogen storage. In lysosomal storage diseases, as defined by Hers (*30*) to be characterized by deficiency of a single lysosomal enzyme, abnormal deposits of utilized precursors of a metabolic reaction lie within the lysosomes.

Disorders affecting transport mechanisms generally involve the small intestine and kidneys and impair the transport of nutrients such as sugars and amino acids, due to a defective membrane carrier protein. If the transport mechanism for the entry of the precursors in the cell is defective, such small amounts of the precursor may be present intracellularly that the enzyme will not be saturated with the substrate, and the entire reaction sequence may be affected. In Hartnup's disease, defective intestinal transport of tryptophan causes low levels of nicotinamide, which will produce clinical symptoms unless corrected by dietary niacin supplements. Cystinuria, iminoglycinuria, and renal glycosuria are other examples of inborn errors due to transport defects. Inherited metabolic disorders may also involve a defect in the binding of hormones to their respective receptors due to a deficiency (as in familial hypercholesterolemia). Still other IEM alter levels of circulating protein such as albumin, γ-globulin, thyroxin-binding globulin, and β-lipoprotein.

The clinical consequences of inborn errors show a wide spectrum; some of the disorders may result in a virtually complete cessation of normal function; others may show no effect at all. The loss of normal enzymatic function may result either due to defective synthesis or production of a mutant enzyme with altered biochemical properties. In the socalled K_m mutants, partial loss of enzyme activity may result because of altered affinity of the enzyme for its substrate or cofactor. When the abnormal enzyme activity is due to altered affinity of the enzyme for a vitamin coenzyme, the decreased activity may be enhanced by administration of large amounts of the vitamin concerned.

The clinical symptoms and signs of inborn errors may also vary greatly. Some disorders such as pentosuria, fructosuria, and β-aminoisobutyric

aciduria are innocuous; others such as phenylketonuria, galactosemia, and cretinism produce severe clinical effects in early childhood. Some disorders may show symptoms only when the individual is exposed to certain environmental influences, such as G6PD deficiency, which may manifest itself only after exposure of the individual to certain drugs. Several disorders, although producing symptoms, are not life-threatening, and are compatible with long life. Patient's sex is an important determinant for some disorders, which are inherited as X-linked recessive or dominant traits. The incidence of an X-linked dominant trait, such as familial hypophosphatemia rickets in women, is twice that of men. On the other hand, X-linked recessive disorders, such as glucose-6-phosphate dehydrogenase deficiency, manifest themselves clinically in women only rarely. The clinical considerations of disease frequency and the effect of factors such as age and sex are dependent on the methods of diagnosis for IEM.

VI. EXPERIMENTAL APPROACHES AND METHODS OF DIAGNOSIS FOR INBORN ERRORS OF METABOLISM

The experimental approach to IEM can be considered from two aspects: how to diagnose a patient suspected of having the disease and what methods to use in understanding the mechanism of the genetic defect.

The physician would usually suspect the disease either from clinical symptoms or on the basis of the results of screening tests for certain IEM. In IEM having clear-cut clinical symptoms and which feature characteristics such as red macular spots in some sphingolipidoses, it is easy to perform a definite diagnostic test quickly. Other metabolic disorders may be associated with unusual odors or coloration of urine. The presence of accumulated metabolites typical of the disease may be identified by means of laboratory methods, such as qualitative color tests, formation of derivatives, and chromatographic techniques. Diagnosis may be made by enzyme assays on blood, tissues, leukocytes, or red blood cells, as seems appropriate. However, it has to be borne in mind that enzyme of interest may exhibit tissue specificity, which is of importance in making the correct diagnosis of an IEM. Shih (*31*) has summarized the main clinical and biochemical features of various IEM, and gives details of commonly available methodologies for the laboratory diagnosis of IEM.

Improved and newer methods and techniques for diagnosis of individual IEM, as well as those with general applicability for a group of IEM, are continuously being investigated. Anderson and Anderson (*32,33*), used two-dimensional electrophoresis with isoelectric focusing in one direction and electrophoresis in sodium dodecylsulphate in the second, to

obtain a high-resolution separation of serum, tissue proteins, and protein subunits. This method could be of great usefulness in the future for characterizing abnormalities in protein composition in IEM.

By use of gas chromatography/mass spectrometry and high performance liquid chromatography (HPLC), new areas of IEM, such as oligosaccharidoses, have become amenable to investigation, and methods have become available for the quantitative determination of oligosaccharides in the urine (*34,35*).

The ultimate aim of the experimental approach to IEM is the elucidation of the basic genetic defect. Since it is not yet generally feasible to determine this defect at the DNA level, efforts are directed at identification of the molecular abnormality at the protein level. Before arriving at this stage of understanding, which has been achieved for only a relatively small percentage of genetic disorders, a number of indirect and direct approaches may be undertaken in the understanding of the IEM. These have been thoroughly discussed in an excellent review by Brock (*17*).

The elucidation of a number of inborn errors, especially storage diseases and aminoacidopathies, has been helped by the delineation and structural characterization of accumulated metabolites, and by identification of depleted metabolites in other cases (*17*). In many other conditions the clue to the nature of the genetic defect has been provided by the highly differentiated nature of the tissue or organ involved, e.g., in hemolytic anemias. The metabolic blocks in a number of aminoacidopathies have been located by oral and intravenous administration of an amino acid and measuring the rate of product formation (*36,37*). Some indication of IEM involving intestinal and kidney transport deficiencies may be obtained by measurement of plasma and urinary metabolites, after an oral load of the compound of interest.

The extent to which information leading to the elucidation of the mechanism or underlying cause of an IEM can be obtained from human subjects or patients is rather limited because of ethical and logistic considerations. A better understanding of the molecular basis as well as pathology of the metabolic disorders is of great importance in order to place the treatment of such conditions on a sound scientific basis. Studies with true genetic animal models of disease as well as chemically induced models and tissue culture models have been very valuable in understanding the nature of IEM and in evaluating therapeutic approaches such as enzyme replacement therapy.

A list of true genetic models summarizing IEM with a genetic counterpart in animals has been provided by Hommes (*38*). The number of such models is small at present and does not answer the question whether the model sufficiently resembles the human disease. Chemically induced

models suffer from lack of specificity of the inhibitor used and may give more useful information if synthesis of enzyme is inhibited specifically.

The rapid progress in cell and tissue culture technology has been very valuable for study of the genetics of metabolic disorders. It is now possible to perform biochemical studies of varying degree of complexity on cultured cells. The enzymatic defect responsible for the IEM is expressed in fibroblast cultures for at least 35 IEM (*14*). It has been shown that fibroblasts cultured from patients with various forms of mucopolysaccharidoses and sphingolipidoses exhibit metachromasia or uptake of toluidine blue dye to give a reddish-pink complex, due to the presence of abnormal amounts of polyanionic high-molecular-weight compounds. An improved technique that uses the affinity of another dye (Alcian Blue) for polyanionic molecules may be more specific (*39*) for disorders in which mucopolysaccharide accumulation takes place. There are, however, two disadvantages in the use of cultured fibroblasts. First, some of the IEM may not be expressed in fibroblasts; for example, phenylalanine hydroxylase activity is confined to hepatocytes, making it impossible to diagnose the disease (PKU) with fibroblast culture. Secondly, in the fibroblast-cultured cell, there is no phenotypic specificity for the tissue where the consequences of the disease are observed.

Finally, Pena and Wrogemann (*40*) have proposed that the application of labeling technique for detection of protein defects in genetic diseases on the basis of structural rather than functional alterations may be a valuable new tool in biochemical genetics. This technique, which has been used in the past by other workers (*41,42*) in the investigation of the molecular pathology of inherited metabolic disorders, is based on the double labeling of normal fibroblast proteins, followed by extensive fractionation and analysis. It is important, however, that the mutant protein be distinguishable from the wild type, either by structural changes or decreased amounts.

VII. GENETIC SCREENING

Genetic screening has been defined as the search in a population for persons possessing certain genotypes that are known to be associated with or to predispose to disease in the individuals or their descendents (*43*). The consequences of treatable genetic diseases can be prevented by identifying the disease in the newborn at an early stage before irreversible damage takes place. For some diseases it may be possible to identify the parents with a high risk of transmitting a disease by methods for heterozygote detection, and to prevent the birth of an affected fetus either by abor-

tion after prenatal diagnosis of the disorder or by providing to the parents the odds of having an affected fetus through genetic counseling.

Screening of newborns for IEM has been conducted in a number of countries for several years. The discovery in 1953 by Bickel *et al.* (*44*) that in phenylketonuric children fed a phenylalanine-free diet, mental retardation associated with the disease could be prevented, although not corrected, underlined the need for early detection of the disorder. The earlier test for detection of PKU was based on the reaction of ferric chloride with phenylpyruvic acid in urine, as a dipstick (Phenistix) method. The use of this method in England was abandoned after it was found that over one third of phenylketonuric children gave a false negative result. In 1963, Guthrie and Susi (*45*) developed a microbiological screen for PKU in which the filter paper containing the blood specimen is placed on an agar plate containing a phenylalanine-requiring bacterium. In the presence of phenylalanine, a circle of growth takes place the size of which is proportional to the amount of phenylalanine. In the United States, approximately 90% of all newborn babies are screened for PKU, using the Guthrie assay. A positive test by the screening procedure is confirmed by a quantitative test. Screening for PKU has made it possible to identify the vast majority of infants with the disease, and to start therapy in time to prevent irreversible damage. A further refined test developed by Guthrie (*46*) allows the screening of other disorders on a single sample of dried blood. Techniques for detection of several amino acid disorders by a chromatographic method have been reported which have been very effective (*47,48*). An exhaustive review by Levy (*49*) lists the disorders for which screening tests are available. Some of the inborn errors for which screening is possible are listed in Table I. The subject of screening of newborns for detection of IEM has recently been reviewed (*50,51*).

For a screening test to be useful, a number of criteria have to be met, which accounts for the small number of the IEM for which such tests are available. The procedure should be simple and cheap. There must be a suitable treatment for the disease which should be cost-effective; delays in diagnosis would result in increased risk to the patient. The objectives of a screening program should include (1) prevention of irreversible damage, (2) prevention of delays in diagnosis, (3) reduction of exposure to harmful drugs, and (4) education of patients on recurrence of the disease (*50*).

Early institution of therapy can prevent irreversible damage in conditions such as galactosemia, branched-chain ketoaciduria, and homocystinuria. In diseases like histidinemia, in which neither the frequency of the disorder nor the effectiveness of dietary restricion is known, the usefulness of a screening program is doubtful. In other cases such as tyrosinemia, in which transient elevations of tyrosine take place in about 1% of all

TABLE I

Conditions for Which Newborn Screening Is Possible (*14,50*)

Adenosine deaminase deficiency
α_1-Antitrypsin deficiency
Branched-chain ketoaciduria
Cystic fibrosis
Cystinuria
Duchenne muscular dystrophy
Familial hypercholesterolemia
Galactosemia
Glucose-6-phosphate dehydrogenase deficiency
Homocystinuria
Hartnup's disease
Histidinemia
Infantile hypothyroidism
Phenylketonuria
Sickle cell anemia
Tyrosinemia

newborn infants, screening for the disorder may not be cost-effective. Most cases of sickle cell anemia, thalassemia, and orotic aciduria can be detected at 6 months to 1 year of age by hematocrit determinations. For diseases such as glucose-6-phosphate dehydrogenase deficiency or α_1-antitrypsin deficiency, which make infants susceptible to certain drugs or environments in the future, screening may be helpful to identify the individuals. It has been suggested by Holtzman (*50*) that screening of cord blood samples of the first pregnancy may be helpful in identifying mothers with PKU, who can be counseled about the extremely high probability that all of their offspring will be retarded.

There are numerous difficulties in accomplishing a broad screening program. The mere fact that a screening test is available does not guarantee the success of a screening program. Several additional conditions should be satisfied and the entire newborn population should be screened. The specificity of the screening test must be very stringent, and both false positives and false negatives should be exceedingly low. A number of factors may account for the false negative result. Thus, for example, the protein defect may not be apparent in certain individual children at the age when screening for that disorder is done, or if the test substance may not accumulate in a large enough amount due to inadequate dietary intake. In addition, poor reliability of the laboratory performing the test may account for increased negative results. It is very important that the physician evaluat-

ing the screening test be aware of the possibility of a false negative test; the clinical signs of any IEM, which may present themselves in the face of a negative screening test, should be given due consideration.

Despite the high specificity of screening tests, the ratio of false positives to true positives may still be high due to the low incidence of the disease. For example, 1 in 1000 normal infants will give a false positive test for PKU. As the incidence of PKU is 1 in 15,000, the true positive to false positive ratio will be 15:1. It is necessary to follow up all positives before a definite disorder of the particular inborn error can be made, and a very high false positive to true positive ratio may make it difficult to follow up the large number of infants with positive results. The whole objective of screening to prevent irreparable damage is lost if treatment cannot be started in time. Holtzman (*50*) describes that in PKU screening programs in the United States more than a month passed by for 23% of infants with positive PKU tests before follow-up was obtained. Levy *et al.* (*52*) found that 9 out of 35 infants who had positive tests for galactosemia died as a result of sepsis in the neonatal period. Results of a survey of 156 physicians by Holtzman (*53*) suggested that a high percentage of the participants failed to recognize and manage high-risk situations for inherited metabolic disorders and were deficient in their response to screening test results.

Thus, despite the availability of screening tests for several disorders, it is not clear that newborn screening will accomplish the object or objectives most frequently claimed. Screening for only PKU and, more recently, for congenital hypothyroidism have found wide application. Recently, Paul and Guthrie (*54*) have reported a urine screening test for metabolic diseases in newborn infants, in which several bacterial inhibition assays can be conducted on urine-impregnated paper for disorders of amino acid, purine, and pyrimidine metabolism. This method may find wide application due to ease, specificity, and low cost. In all screening programs, the maximum benefit can only be obtained when all the implications and requirements of such programs are well understood by all concerned, especially the physician involved.

Heterozygote Detection: Screening of High-Risk Populations

At the present time, no treatment is available for the vast majority of genetic diseases, and for these disorders efforts can be directed only towards the prevention of the birth of an affected child. The most effective way to meet this objective would be the identification of heterozygote carriers, who can then be given genetic counselling. For this approach to

be feasible, reliable methods for carrier detection that can be applied to large populations in a cost-effective way are needed. Moreover, the incidence of heterozygous carriers for the disease concerned must be high enough to justify screening.

Detection of heterozygotes has been very effective for those proteins which can be studied by physical methods, such as hemoglobins or plasma proteins. Since the amount of protein produced by a cell is proportional to the number of structural genes controlling its synthesis, in heterozygous carriers there is generally a 50% reduction in activity of the enzyme relevant to the disorder. Measurement of enzyme activity or of a specific metabolite in the heterozygote carrier in a suitable body fluid would, in some instances, reveal lower heterozygous values as compared to normal values. Enzymatic assays using appropriate body fluid have been helpful in identification of heterozygotes for several diseases such as acatalasemia (*55*), pseudocholinesterase deficiency (*56*), branched-chain ketoaciduria (*57*), and galactosemia (*58*). In some disorders, blood or urine analysis for relevant enzyme activity or metabolite concentration fails to distinguish the heterozygote carriers from normals, but distinction between them can be made after administration of a load of the normal substrate of the defective enzyme. A decreased tolerance to the substance administered as shown by elevated blood concentration or excretion of unusual metabolites indicates the heterozygosity of the individual. However, neither of these two methods provides the complete separation of homozygotes from heterozygotes with certainty. Raine (*16*) suggests the designation of heterozygosity in terms of probability, using the methods described by Westwood and Raine (*59*).

The criteria for heterozygote carrier screening in large populations referred to above have been most closely met only in the case of screening for carriers of Tay–Sachs disease among Ashkenazi Jews. The heterozygote carrier frequency of this lethal disorder, which is caused by deficiency of β-*N*-acetylhexosaminidase-A leading to GM_2-ganglioside storage, is 1 in 30 to 1 in 60 in the American Jewish population, as compared to 1 in 300 among the general population. Heterozygotes for this disorder can be simply, accurately, and relatively inexpensively determined by assay of serum hexosaminidase-A, using a highly accurate automated technique (*60*). The limitation of Tay–Sachs disease to a defined population, the availability of a satisfactory and cost-effective assay, and the alternative of terminating the pregnancy during a safe period, following prenatal diagnosis of the disease, make this disorder ideally suited for population screening. By the end of 1976 (*61*), more than 150,000 individuals had been screened for Tay–Sachs disease, from which 6700 were

found to be carriers and 124 were couples at risk. By proper genetic counseling and prenatal monitoring, the birth of 12 affected children was prevented. By the end of 1979, over 250,000 individuals had been screened and 210 couples were identified as being at risk (*62*).

Screening for sickle cell anemia and sickle cell trait by electrophoresis of blood specimens has been conducted regularly in many places. Both homozygotes and heterozygotes for hemoglobin S have been identified (*63,64*), but the procedures for prenatal diagnosis of an affected fetus are not as well developed as for Tay–Sachs and some other diseases. Heterozygote-carrier screening programs are not feasible at this time for other inherited metabolic disorders, either because the metabolic defect is not understood, that there is no suitable assay available, or the disease is so uncommon that the screening program may be very cost-ineffective (*61*).

VIII. PRENATAL DIAGNOSIS OF INHERITED METABOLIC DISORDERS

In recent years, due to the development of accurate and safe intrauterine diagnostic procedures, such as ultrasound, amniocentesis, and fetoscopy, it is now possible to detect a number of hereditary disorders in early fetal life. The most useful of these techniques is amniocentesis, which is capable of detecting virtually all cytogenetic abnormalities, as well as over 70 inherited metabolic disorders. Ultrasonography and fetoscopy that allow a direct visualization of the fetus are rapidly becoming invaluable in the diagnosis of certain major structural defects. Fetal blood sampling made possible by the technique of fetoscopy allows detection of other serious conditions, such as Duchenne muscular dystrophy and β-thalassemia.

Although it does not seem likely that treatment for genetic diseases in fetal life will be possible soon, the availability of the option to terminate an affected pregnancy, currently possible as a result of prenatal diagnosis, represents one aspect of the control of serious metabolic disorders. The advances in the field of prenatal diagnosis, coupled with the Supreme Court's decision allowing therapeutic abortions, have thus added a new dimension to the prevention of IEM and other hereditary diseases. This mode of control of certain inherited disorders, however, has profound medico–ethical and legal implications. These problems have been discussed by Raine (*65*). The subject of prenatal diagnosis has been reviewed by several authors (*66–68*).

Amniocentesis

Amniocentesis is the most important tool in the prenatal diagnosis of genetic abnormalities. In the 1930s, third trimester amniocentesis was used as an obstetric tool, and in the 1950s, it was used to diagnose fetal sex prenatally using the Barr-body technique. Identification of fetal sex is also of relevance when the disease causing concern is X-linked. If the fetal sex is female, there will be no risk, whereas in males, the risk will be 50%.

The discovery that second-trimester amniotic fluid cells of fetal origin could be grown in tissue culture led to the demonstration by Nadler that enzyme deficiencies could be effectively determined from the amniotic fluid cells (*69*).

Amniocentesis is generally performed at 16 weeks of gestation when the amniotic fluid volume is 200 ml, thus enabling the obstetrician to obtain the fluid easily. The gestational age of 16 weeks is optimal for amniocentesis because it allows sufficient time for chromosomal and metabolic studies and repeat taps if necessary, so that if the termination of pregnancy were indicated, it could be done before 24 weeks of gestation. Gestational age is determined by menstrual history and confirmed by ultrasonic scanning. Approximately 10–15 ml of amniotic fluid is withdrawn and separated into liquid and cellular proteins by centrifugation. The supernatant is used for α-feto protein assay, and has also been useful in the detection of the adrenogenital syndrome and methylmalonicaciduria (*22*), whereas the sedimented material is used for cell culture. At 16 weeks, approximately 20% of these cells are viable, and most cells are of fetal origin. Sometimes contamination with maternal cells does take place and is a source of serious problems; as many as 0.8% of amniotic cell cultures could be derived from maternal rather than fetal cells. This problem is avoided by setting up separate cultures from each syringe of amniotic fluid and then subculturing clones from single cells, so that multiple cultures are established, grown, and analyzed independently (*70*).

The procedure for cell culture has been described in detail by Kaback and co-workers (*70*). The point can hardly be overemphasized that a population of actively dividing cells is needed for both cytogenetic and biochemical studies; performance of such tests on nonviable cells can be dangerous (*66*). Recently developed microbiochemical techniques may in the future make it possible to circumvent the time-consuming process of cell culture, which is now necessary for accurate *in utero* enzyme analysis.

In order to diagnose an IEM *in utero,* certain prerequisites have to be met. The enzymatic reaction of interest must be expressed in amniotic fluid cells. Certain enzymes such as phenylalanine hydroxylase, which is deficient in phenylketonurics, are not expressed in amniotic fluid, and

therefore such conditions cannot be diagnosed by amniocentesis. Normal levels for the enzyme levels have to be established in a large number of pregnancies at various stages of gestation before a level can be considered to be abnormal in an affected fetus.

A large number of metabolic diseases which include disorders of carbohydrate, lipid, amino acid, mucopolysaccharide, and nucleic acid metabolism, as well as defects in other cellular reactions, are now amenable to prenatal diagnosis, and the number is increasing. Prenatal diagnosis can be considered to be feasible in about 75 different inherited metabolic diseases (*61*); practical experience has been obtained with the diagnosis of more than 35 such disorders (*61,71,72*). Table II gives a partial list of some important IEM for which prenatal diagnosis is available. The largest group of diseases detected by prenatal analyses are mucopolysaccharide or lipid storage disorders.

It has been shown conclusively in various studies that amniocentesis in the 15th–16th week of pregnancy carries very little risk for both the mother and fetus. The National Institute of Child Health and Development prospective collaborative study (*73*), involving nine United States institutions, to assess the risks of mid-trimester amniocentesis and the Canadian collaborative amniocentesis study, for example, show amniocentesis to carry a combined maternal and fetal risk of no higher than 1 in 200. The diagnostic accuracy has been determined to be 99.4%. Golbus *et al.* (*74*) found that abortions following amniocentesis showed only a slight increase when compared to the rate of spontaneous abortions. Ethical guidelines for prenatal diagnosis have recently been suggested by Powledge and Fletcher (*75*).

With the use of conventional analytical techniques for the detection of an enzyme deficiency in cultured amniotic fluid cells, 4–6 weeks of culturing is required to give enough cellular proteins for analysis. Ultramicrochemical techniques have been developed (*61*) that use (sub)microliter incubation volumes and extinction and fluorescence measurements that use microscope spectrofluorometer techniques, and allow for analyses of small numbers of cultured cells, thereby reducing the cultivation period. It is possible to establish a prenatal diagnosis within 7–14 days after amniocentesis (*76*), using these ultramicro techniques.

Inborn errors have been diagnosed *in utero* without analysis of cultured amniotic fluid cells by the use of other techniques such as fetoscopy. Examples of disorders which have been diagnosed by fetoscopy or placental puncture include mainly hemoglobinopathies such as sickle cell disease, β-thalassemia (*77*), and Duchenne's muscular dystrophy (*78*). When and if the technique of fetoscopy becomes routine with established safety for mother and fetus, it could be useful as a general technique for the prenatal

TABLE II

Inborn Errors of Metabolism Diagnosed Prenatally (*14,15*)

Acid phosphatase deficiency (lysosomal)
Adenosine deaminase deficiency
Adrenogenital syndrome
Arginosuccinic aciduria
Citrullinemia
Congenital adrenal hyperplasia
Cystinosis
Congenital erythropoietic porphyria
Duchenne muscular dystrophy
Fabry's disease
Galactosemia
Gaucher's disease
GM_1 Gangliosidosis, type I
GM_2 Gangliosidosis, type I (Tay–Sachs)
GM_2 Gangliosidosis, type II (Sandhoff's)
Glycogen storage disease, type II
Hemoglobin αG-Philadelphia
Hunter's syndrome
Hurler's syndrome
Hypophosphatasia
Homocystinuria
I Cell disease
Krabbe's disease
Lesch–Nyhan disease
Metachromatic leukodystrophy
Mucolipidosis IV
Methylmalonic acidemia, type I & II
Menke's disease
Niemann–Pick disease
Propionic acidemia
α-Thalassemia
β-Thalassemia
Xeroderma pigmentosum

diagnosis of metabolic disorders. It has been shown by Galjaard (*79*) that genetic enzyme deficiences, which are expressed in leukocytes, are also demonstrable in fetal white blood cells as early as the 16th to the 18th week of pregnancy, and that by ultramicrochemical assay of fetal leukocytes, it could be possible to establish a prenatal diagnosis in 1 or 2 days, without the need for amniocentesis.

A new and novel approach to prenatal diagnosis was reported in 1978 by Kan and Dozy (*80*), who used recombinant DNA techniques to diagnose a fetus for sickle cell anemia. The method which can be carried out

on amniotic fluid cells is safe and sensitive, and has the promise of a wider application to the prenatal diagnosis of any disease for which a gene-specific probe can be identified. Sampling of fetal blood, which can be a hazardous procedure with the present techniques, resulting in a 5–10% mortality, was the method used for the prenatal diagnosis of such diseases. The procedure developed by Kan and co-workers does not require fetal blood, and involves the digestion of fetal DNA by a restriction enzyme designated as Hpa I. Restriction enzymes exhibit specificity for a certain nucleotide sequence in the DNA, and each enzyme produces a characteristic fragment. By the use of a gene probe, it is possible then to identify a specific gene or gene portion in the fragments. Kan and co-workers found that for people of white, black, or Asian origin without any hematological disorders, the β-globin gene is located on a DNA fragment 7.6 kilobases long, produced by the action of Hpa I restriction enzyme. By contrast, blacks with sickle cell anemia have lost the 3′-Hpa I recognition site, and the gene is located on a 13-kilobase fragment about 87% of the time. In order to diagnose the disease in the fetus, the fragment pattern produced by the parental DNA must also be determined. Diagnosis can be made with 15 ml of amniotic fluid. There is no need for culturing the amniotic fluid cells; the result can be obtained by the 16th or 17th week of pregnancy. The work of Kan and co-workers has been reviewed briefly by Williamson (*81*) and Marx (*82*).

Restriction mapping has been used for prenatal diagnosis of a rare form of thalassemia (*82*) caused by gene deletion, and may be used for common thalassemias also, once the causative gene defects have been delineated. Prenatal diagnosis using restriction endonucleases has been used for diseases with defects in globin genes because the appropriate probes are readily available. It is theoretically possible to use this technique for any genetic disease, provided the right restriction enzymes and probes are available. The work on cloned human genes may prove useful for preparation of such probes, and restriction mapping may be used for the prenatal diagnosis of a wide variety of genetic defects. This technique would be especially useful for those diseases in which the metabolic defect is not expressed in cultured amniotic fluid cells.

IX. THERAPY OF INBORN ERRORS OF METABOLISM

As indicated before, IEM are caused fundamentally by a change in one specific protein, which may be produced in smaller amounts, may not be produced at all, or may be present in an altered state. Although there is as yet no effective therapy for many of the inherited metabolic diseases, some disorders, particularly those in which the biochemical abnormalities

have been well characterized, are amenable to effective therapeutic management. A therapeutic approach to IEM can theoretically be directed at either altering the genotype of the affected individual or to lessening or eliminating the harmful effect of the mutant phenotype. Ideally, one would like to correct the error at the level of the gene; thus, research in transgenosis and genetic engineering is directed towards this end, using recombinant DNA research techniques that involve introduction of a foreign DNA into the host cell. This approach may prove to be very promising in the future, especially in terms of the eventual understanding of gene control, but at the present time, its therapeutic usefulness is negligible. There are major ethical (*83*) and legalistic concerns about gene therapy by means of viral transduction, such as the possibility of unwanted consequences as neoplasia, which have led to the establishment to strict guidelines for DNA recombinant research (*84*). The possibility of using genetic engineering for therapy of inborn errors must, therefore, only remain a promise for the future for a long time, and successful therapy of inborn errors has, so far, been achieved only at the level of the phenotype. The nutritional aspects of the therapy for inborn errors of metabolism are discussed by O'Brien (*85,86*). Raine (*87*) describes the practical precautions in the laboratory monitoring of therapy.

A. Replacement Therapy

Replacement of a missing or abnormal protein constitutes one of the approaches to the treatment of IEM. It has been suggested by Boyer and co-workers (*88*) that this approach is more likely to be effective and safe in CRM-positive patients. Protein replacement therapy has been successful with plasma protein disorders, such as treatment of a gammaglobulinemia with γ-globulin, hemophilia with antihemophilic globulin, and analbuminemia with albumin. However, there are several limitations to this mode of treatment: (1) large amounts of protein or enzyme of interest need to be available; (2) the protein in question must be given parenterally and may not reach the exact target tissue; (3) therapy must be continuous or repeated at frequent intervals; (4) the protein must be delivered across the blood–brain barrier without any serious toxic reactions; (5) repeated injections of the protein may invoke an immunological response and produce antibodies.

In contrast to the experience with plasma protein disorders, success has, as a whole, been rather limited with enzyme replacement therapy. In general, enzyme replacement therapy is more likely to be successful with disorders involving lysosomal enzyme deficiencies than other deficiency states, in which the enzyme is not localized to lysosome. Brady and co-

workers (*89*) have reported the first successful attempt at treatment of a genetic disease by infusion of a purified enzyme. A decreased glucocerebroside content of erythrocytes and liver was found following administration of purified glucocerebrosidase to two patients with Gaucher's disease. However, the reduction of hepatic glucocerebroside was not paralleled by a reduction in the organ size. Immobilization of an enzyme by attachment to microparticles of inert material that are taken up by lysosomes may be helpful in directing the enzyme to a site where they are needed, and alleviate this difficulty with the use of enzymes in therapy. The problems associated with presence of free enzymes in the circulation could also be obviated by entrapping enzymes in liposomes. It has been know (*90,91*) that liposomes will prevent or diminish both the contact of circulating immunoglobulins with intraliposomal antigens and, in certain cases, premature enzyme–substrate interactions (*92*). Work with model lysosomal storage conditions (*93,94*) supports the view that liposomes may be utilized as enzyme carriers in treatment of lysosomal storage diseases. Braidman and Gregoriadis (*95*) have purified glucocerebroside β-glucosidase in large amounts by the use of a rapid affinity chromatography technique, and have found that 60% of the enzyme can be associated with liposomes. However, despite the fact that liver is the principal site for localization of carrier associated enzyme, no appreciable reduction in liver size has been noted, although some clinical improvement of the patient has been seen. Liposome-entrapped enzymes are unlikely to be of use in patients with neural tissue (e.g., Tay–Sachs disease) and cardiac or skeletal muscle involvement (e.g., Pompe's disease) because of the low endocytic activity in such tissues. With neural tissues, it has been proposed to link the enzyme of interest with the patients' own leukocytes, which may delay enzyme degradation sufficiently to infiltrate the brain. Another approach in which small substrate molecules penetrate the lipid layer of liposome to reach their respective enzymes may be feasible, and may be applicable to treatment of conditions such as phenylketonuria. The subject of the carrier potential of the liposome has been reviewed by Gregoriadis (*96*). The prospects for the usefulness of enzyme replacement therapy has recently been enhanced by methodological advances that have allowed preparation of bulk quantities of human enzyme, and by the discovery that many enzymes possess recognition markers which allow the rapid uptake by the cell (*97*). For example, there is a phosphomannosyl-recognition marker on several lysosomal hydrolases, facilitating their uptake by fibroblasts, and a fucosyl marker on glycoproteins that are recognized by liver cells. It may be possible to prevent the uptake of an enzyme by a particular tissue in favor of other tissues by substances which compete for similar binding sites, allowing the enzyme to circulate longer.

Accessibility of enzymes to the brain may be achieved by increasing the permeability of the blood–brain barrier by exposure of animals to hyperbaric oxygen or by hypotonic shock (*98*).

B. Enzyme Induction

For a very limited number of disorders, induction of a deficient mutant enzyme in the patient may be used to correct enzyme defects. Administration of phenobarbital and certain other drugs stimulates the increased production of enzymes associated with smooth endoplasmic reticulum, some of which are involved in detoxication mechanisms. In patients with Crigler–Najjar syndrome (*99*) or with Gilbert's syndrome (*100*), who exhibit hyperbilirubinemia, a decrease in the level of bilirubin may result following phenobarbital administration, probably due to the induction of hepatic glucuronyltransferase. Moses and co-workers (*101*) reported that in a patient with Type III glycogen storage disease, adrenal steroid therapy resulted in the reappearance of glucose-6-phosphatase activity. Although the mechanism of this effect is not clearly established, it is likely that the hormone induced the formation of the enzyme.

C. Product Supplementation

Another approach toward therapy that has been rarely used is that of supplementation with the end product, when a metabolic block prevents its formation in sufficient amounts. A good example of this approach is the addition of uridine to the diet of patients with macrocytic anemia of orotic aciduria. The treatment of the inborn error of tyrosyl iodination at an early stage by administration of thyroxine or of cortisone in treatment of adrenogenital syndrome are two conditions in which the absence of the synthesized metabolite is the primary abnormality.

D. Supplementation with Vitamin Cofactors

In conditions in which the enzyme may have a structural defect of the cofactor site rather than the substrate site, or in which there may be a defect in the active form of the cofactor, supplementation with the cofactor is a mode of therapy. More than 20 metabolic disorders are known to show a clinical response to supplementary amounts of specific vitamins (*102*). Most of these disorders are a result of primary enzyme disturbances that cause impaired affinity for a cofactor. Others are caused by abnor-

malities in the synthetic pathway of coenzymes from vitamin precursors. This approach to therapy has not been extensively used, and clinical improvement does not always take place, even in those cases in which the biochemical abnormality is correctable.

E. Substrate Limitation

The most important and usual approach to therapy, which has been applied primarily to the inborn errors of amino acid and carbohydrate metabolism, is dependent on withholding from the diet a constituent that results in an abnormal toxic product or an unmetabolizable substrate. Phenylketonuria is the most well-known example of this approach. Other disorders which have been effectively treated in this way include branched-chain ketoaciduria, homocystinuria, galactosemia, hereditary fructose intolerance, and Refsum's disease. In disorders of amino acids that can be synthesized in the body, such as hyperprolinemia or citrullinemia, dietary restriction is of no benefit. Although the dietary treatment appears simple in theory, there are several pitfalls and possible hazards of such treatment that should be carefully evaluated. The restricted compound should be sufficient to meet the metabolic requirement, yet not accumulate in the body at a toxic level. Semisynthetic diets containing low amounts of the implicated substance need to be supplemented with other nutrients. This is especially true in early infancy during the most active form of brain myelination.

In order to properly manage the patient with a treatable IEM, it is essential to carefully monitor treatment. In the case of amino acidopathies, the concentration of the amino acid of interest should be monitored to ascertain that it is maintained at the optimal level. It has to be borne in mind that levels of the restricted amino acid in the blood may rise not only as a result of excess intake, but also from body protein catabolism when intake is low. The laboratory monitoring of IEM requires such a degree of sophistication that the American Academy of Pediatrics recommends management of such patients in specially equipped centers.

F. Avoidance of Certain Drugs

In certain inherited metabolic disorders, the patient does not show any symptoms until after coming in contact with certain specific stressful stimuli. In such conditions, the clinical consequences of the disorders can be prevented by avoiding exposure to the particular environmental influences. For example in patients with glucose-6-phosphate dehydrogenase

deficiency, hemolytic crises can be avoided by preventing exposure to primaquine or other precipitating agents.

G. Toxicity Limitation

In several metabolic diseases, toxic reactions may be caused by deposition of a specific substance in tissues. Depletion of the affected tissue by increasing the excretion of the stored substance or preventing its deposition will help alleviate the symptoms of the disease. For example, in Wilson's disease, copper (which is the incriminating agent) can be removed by chelation with British anti-lewisite (BAL) or penicillamine, and thus excreted more rapidly. In hemochromatosis, the excretion of iron is increased by frequent phlebotomy or by the administration of desferrioxamine-B. Drugs such as sulfinpyrazone, which increase the excretion of uric acid, are of benefit in treatment of gout, as are drugs such as allopurinal, which inhibit its biosynthesis.

H. Organ Transplantation

Restoration of a missing enzyme has also been attempted by organ transplantation. This approach has produced clinical and biochemical changes in a few instances. In patients with childhood cystinosis hyperoxaluria or Fabry's disease, renal transplantation has shown encouraging results in some cases (*22*). Bone marrow transplantation has been reported to yield beneficial results in certain immune deficiency states, and may be the only mode of treatment for patients with such disorders. Transplants of HLA-identical fibroblasts were found to increase the levels of the relevant enzymes in patients with Hurler's or Sanfilippo-type A diseases, without evidence of graft rejections for several years. Although this method may find greater applicability in the future, at the present time, it can only be justified if the organ concerned suffers progressive and ultimately lethal injury, the disease entity being associated with severe mortality and morbidity.

X. BIOCHEMICAL AND GENETIC HETEROGENEITY

The structure of each gene present on the chromosome is subject to variation due to mutation, resulting in a change either in the primary structure on in the rate of synthesis of a specific protein. Extensive qualitative and quantitative variation exists among healthy individuals, as de-

termined by the structure and properties of many proteins. Harris and Hopkins (*103,104*) searched for electrophoretic variants of several enzymes from a large number of normal subjects and estimated that individual heterozygosity may be present at 20% of structural loci. Many variations are inconsequential and do not effect the individual clinically or biologically. When the variant protein is altered to the extent that the fitness of the individual is affected, an inherited metabolic disease may occur.

In the study of inborn errors of metabolism after a disorder has been identified and described, closer analysis generally reveals a collection of disorders with similar clinical signs and symptoms. The identification of variant forms of established inborn errors of metabolism has increased considerably with the availability of cell and tissue culture techniques and advancements in microchemical analysis.

Genetic heterogeneity may either involve mutation at a single locus (allelic) or at separate loci (nonallelic). The allelic mutations may be further subdivided into homo-allelic or hetero-allelic, depending upon whether the amino acid substitutions occur at the same or at different points in the polypeptide chain.

As it is not possible, at the present time, to understand genetic variation at the level of the genome, observations about differences in qualities or quantities of protein may give information about gene differences. The demonstration of protein variants is most commonly achieved by differences in electrophoretic patterns or use of chemical and biochemical assays, but other methods such as immunological methods are also used. Variant hemoglobins have been identified by the use of amino acid sequence determination. At least 240 variants have been discovered, most of which are single amino acid substitutions.

Sequence data on human mutant enzymes are limited, but Yoshida has demonstrated two glucose-6-phosphate dehydrogenase variants, in which a single amino acid substitution occurs (*105*). Similarly, a mutant human-erythrocyte carbonic anhydrase $I_{c\ guam}$ differs from the normal enzyme by one amino acid (*106*). Stanbury *et al.* (*14*) list disorders in which a deficient activity of a specific enzyme has been demonstrated in humans.

For an increasing number of diseases, it has been found that different clinical features may be presented by a similar deficiency of an enzyme. Most often this clinical heterogeneity is not associated with differences in the biochemical characteristics. Whether or not different gene mutations are involved in different clinical forms of the same enzyme deficiency can be determined by the technique of complementation analysis after somatic cell hybridization. Hybridization of cultured skin fibroblasts from

two different patients, both with a similar enzyme deficiency, is done using Sandai virus or polyethylene glycol (*22,61*). Enzyme assays are carried out after a few days of *in vitro* cultivation to determine if restoration of enzyme activity has taken place in the heterokaryons, which would indicate that the clinical heterogeneity is due to a different gene mutation.

Such studies have been of value in understanding the genetic background of several inherited metabolic disorders (*107–109*). Different gene mutations have been shown to be involved in clinical variants of several inborn errors of metabolism such as GM_1 and GM_2 gangliosidoses, Niemann–Pick disease, xeroderma pigmentosa, methylmalonic aciduria, and maple-syrup urine disease.

Genetic methods also have been of use in demonstrating heterogeneity. Evidence for separate genes is provided by family studies that show the existence of different modes of inheritance for identical disorders. For example, Hunter's disease is an X-linked trait, whereas Hurler's disease is inherited as an autosomal recessive. Classic linkage analysis also has revealed heterogeneity.

XI. CLASSIFICATION OF INBORN ERRORS OF METABOLISM

Inborn errors of metabolism can be classified biochemically on the basis of the protein defect in more than one way. The disorders can be separated into groups according to the class of compounds whose metabolism is affected, such as disorders of amino acid, carbohydrate, lipid, or mucopolysaccharide metabolism. They can also be classified as defects due to (1) enzyme deficiencies, such as amino acidopathies; (2) structural protein changes, e.g., hemoglobinopathies; (3) membrane transport alterations, e.g., cystinuria; (4) other conditions associated with cell membrane change, e.g., Duchenne's muscular dystrophy; and (5) storage diseases, such as mucopolysaccharidoses. With this approach to classification, some disorders may be classifiable in more than one category of biochemical alterations. Lists of disorders belonging to various classes are given in textbooks such as Bondy and Rosenberg (*22*). Raine lists (*14*) disorders along with their clinical features, biochemical diagnosis, treatability, and whether or not prenatal diagnosis and heterozygote detection is available for them. It is impossible to discuss in detail even a small percentage of the more common disorders within the scope of this chapter. The following account concentrates on a few inborn errors of metabolism that emphasize some aspect of these disorders such as newborn screen-

ing, prenatal diagnosis, and dietary treatment. Therapeutic success has most often been achieved with aminoacidopathies.

XII. EXAMPLES OF INBORN ERRORS OF METABOLISM

A. Phenylketonuria

Phenylketonuria is one of the most familiar and typical examples of inborn errors of metabolism, and one which has provided the most extensive experience with newborn screening. Considerable progress has been made in recent years in our understanding of the various aspects of hyperphenylalaninemic states.

Phenylketonuria (PKU) was first reported by Folling in 1934 (*110*), who identified phenylpyruvic acid in excess in the urine of several siblings with mental retardation. Penrose and Quastel (*111*) in 1937 suggested the name phenylketonuria for this condition, which has been adopted ever since. Jervis showed that this trait is transmitted as an autosomal recessive disorder (*11*) and demonstrated the inability of the liver of a patient with this order to convert phenylalanine to tyrosine. From Jervis' studies a rational dietary approach to the treatment of the disease became possible and was first reported by Bickel *et al.* (*44*). Phenylketonuria has remained a model for the study of mental retardation caused by inborn errors of metabolism, and for its prevention by appropriate and timely dietary intervention. The development of simple screening methods for determination of blood phenylalanine concentration (*45*) ushered in a new era in regard to the detection and treatment of PKU patients. For newborn screening, PKU has served as a prototype and the experience obtained has helped in the assessment of the implications of screening programs (*43*). As a result of the data obtained with mass screening programs, various degrees and kinds of hyperphenylalaninemias have been found, due to the varying extent of the deficiency of phenylalanine hydroxylase and other enzymes; the trait is recognized to be genetically heterogenous. In order to properly manage the patients with this disorder, it is necessary to accurately characterize the abnormality, since hyperphenylalaninemia and PKU are not the same (*112,113*). Classic PKU has the incidence of about 1 in 12,000 in the general population, although the incidence varies greatly from country to country. The detection of heterozygotes may be of value for genetic counseling, but all procedures involved have not been reliable. Detection of heterozygotes may be made using phenylalanine loading test, but it is unreliable in identifying some cases (*115*). Jagenburg and Rodjer (*115*)

recently used a method of heterozygote detection for PKU by constant intravenous infusion of L-phenylalanine. The heterozygotes can be discriminated rather accurately by determining postprandial phenylalanine/tyrosine ratios, as proposed by Perry *et al.* (*116*) and Rosenblatt and Scriver (*117*).

1. Metabolic Pathway of Phenylalanine

Phenylalanine is an essential amino acid for protein synthesis in man. Approximately 50% of the normal dietary intake of phenylalanine can be used for protein synthesis during early growth, and this proportion increases as the growth rate declines. The fraction of phenylalanine not used for protein synthesis is converted irreversibly primarily to tyrosine by hydroxylation, by the action of phenylalanine hydroxylase in the liver. Tyrosine is also an essential amino acid and serves as a precursor for dopamine, norepinephrine, thyroxine, and melanin.

The hydroxylating system has been resolved into two components by Mitoma (*118*): a labile fraction(phenylalanine hydroxylase) that is found only in the liver and a stable fraction (dihydropteridine reductase) that is widely distributed in mammalian tissues. Kaufman (*119,120*) has shown that an oxidized pteridine (dihydrobiopterin) is the natural cofactor for the hydroxylation reaction after it has been reduced to tetrahydrobiopterin by dihydrofolate reductase. During hydroxylation the tetrahydro form of the cofactor is oxidized to the quinonoid and is replenished by reduction with the action of the enzyme dihydropteridine reductase.

In addition to the major pathway of phenylalanine producing tyrosine by hydroxylation, there are some minor pathways as well for phenylalanine which yield other products. Transamination of phenylalanine by a specific aminotransferase in the presence of pyruvate as cosubstrate and pyridoxal phosphate as coenzyme yields phenylpyruvic acid. Synthesis of phenylpyruvic acid is impaired by delayed postnatal augmentation; this should be taken into account in the interpretation of a negative ferric chloride test in the newborn. Phenylpyruvic acid can be converted to phenylacetic and phenyllactic acids, and can also be orthohydroxylated to *o*-hydroxylphenylpyruvic acid, which is an important constituent of phenylketonuric urine. Phenylalanine can also be decarboxylated to give phenylethylamine, which in turn can be converted to phenylacetic acid by the action of an amine oxidase.

2. Animal Models of Phenylketonuria

Studies on the biochemical basis of PKU are possible only to a very limited degree on human patients. Animal models of the disease, therefore, have the objective of creating the biochemical and behavioral ana-

logue of human phenylketonuria. Only a few natural mutations occur in animals, which lead to diseases resembling the inborn errors of metabolism found in man. For PKU, there are no genetic animal models with partial or complete absence of hepatic phenylalanine hydroxylase.

The homozygous dilute lethal (d^l/d^l) phenotype in the mouse originally described by Searle in 1952 (*121*), and initially thought to be a genetic model for PKU, has been found not to be that after subsequent investigations by several workers (*122,123*). In order to use the rodent animal model for PKU, the hyperphenylanemic state must be induced in the first 3 or 4 weeks of postnatal life, which is a critical period for the differentiation of the brain. The biochemical and genetic criteria for PKU in humans, i.e., absence of liver phenylalanine hydroxylase activity, persistently elevated plasma and tissue phenylalanine levels, and decreased level of the product of phenylalanine hydroxylation (tyrosine) must also be met.

Due to the scarcity of the genetic animal models, chemical models for PKU have been sought. Phenylalanine administration (*124*) by dietary route or repeated intraperitoneal or intravenous injections leads to wide fluctuations of tissue and plasma phenylalanine levels as well as tyrosine levels, since the hydroxylase is present in normal amounts, and elevations are transitory. Lipton *et al.* (*125*) introduced the use of the phenylalanine hydroxylase inhibitor, *p*-chlorophenylalanine; this inhibitor, when used with L-phenylalanine orally (*126*), increased the phenylalanine levels without elevations in tyrosine. The models using frequent simultaneous injections of phenylalanine and *p*-chlorophenylalanine with supplemental phenylalanine have satisfied the biochemical criteria of PKU, and have been used in the study of mechanisms by which phenylalanine causes the brain dysfunction characteristic of PKU. However, in developing rats, *p*-chlorophenylalanine causes a severe inhibition of growth, and other toxic reactions, such as a high incidence of cataracts. Greengard *et al.* (*127*) proposed the use of α-methylphenylalanine as inhibitor of phenylalanine hydroxylase, which shows less side effects. They were able to induce sustained elevations of plasma and cerebral phenylalanine concentration, comparable to those in phenylketonuric children by treatment of animals with α-methylphenylalanine plus phenylalanine, and found hyperphenylalaninemia to affect cerebral glycine metabolism (*128*). Lane *et al.* (*129*), using this model, found long-term behavioral impairment and changes in biochemical parameters in the animals to be close to the clinical finding in PKU. As pointed out by Delvalle *et al.* (*130*), if chemically induced models are used, it is possible with comparative studies to identify effects that are unique to the inhibitor, and thus not attributable to hyperphenylalaninemia itself.

In all of the models referred to above, it is not possible to carry out dietetic studies because of the addition of phenylalanine and inhibitors of phenylalanine hydroxylase to the diet. Recently, Dhondt *et al.* (*131*) used another model for PKU, and induced hyperphenylalaninemia by a combination of *p*-chlorophenylalanine and cotrimoxazole injections into weaning and adult rats. Since a high phenylalanine load was not needed to obtain the hyperphenylalaninemic state, the model could be used for dietetic studies.

3. Clinical Symptoms and Biochemical Features: Variant Forms

As a result of mass screening programs for PKU detection, it has become clear that phenylketonuria is not the only abnormality of phenylalanine metabolism. Several abnormalities of phenylalanine metabolism in man have existed due to different acquired or inherited alterations in the hydroxylation reaction. In short, phenylketonurics generally have hyperphenylalaninemia, but all hyperalaninemic patients do not have phenylketonuria. There are several variant forms of hyperphenylalaninemia (*112*) which differ in the nature of the presumed defect as well as in the clinical aspect. The major variants fall in three distinct groups involving components of the phenylalanine hydroxylase system: those due to a defect in phenylalanine hydroxylase, dihydropteridine reductase, or biopterin synthesis. The symptoms of the classic phenylketonuric phenotype or type-I hyperphenylalaninemia, reported in the literature, are seldom seen due to the wide use of neonatal screening, which prevents the development of the full range of symptoms in the untreated patient. In general, infants with PKU are normal at birth. Some patients may have a musty odor due to the presence of phenylacetic acid in their excretions. An untreated patient will develop severe mental retardation and an IQ < 30 by the end of 2 years, and a number of neurological deficits may be present (*112,113*). Pigment dilution of skin, hair, and eyes occurs in the untreated phenylketonuric, who is lighter than other siblings in the family.

Mental retardation caused as a result of untreated PKU is almost always irreversible, but the other symptoms disappear when dietary intervention takes place at any age, in order to bring the biochemical abnormalities under control.

In classic phenylketonuria, abnormal accumulation of phenylalanine takes place in body fluids, the plasma concentration generally being greater than 1 m*M* or 16.5 mg%, whereas the concentration of other amino acids is decreased. Many of the metabolites which are products of minor pathways of phenylalanine metabolism, such as phenylacetic acid and phenyllactic acid are excreted in increased amounts in the urine. The

increased excretion of phenylethylamine may be of importance in the pathogenesis of mental deficiency. A number of secondary biochemical abnormalities result, which are a consequence of the accumulation of phenylalanine and its metabolites (*112,113*). Tyrosine metabolism as well as melanin synthesis are impaired, the latter deficit being responsible for pigment dilution. Glutamine deficiency in body fluids and brain may be the cause of mental retardation (*132*), but the exact mechanism of mental retardation in PKU has yet not been clearly resolved. Using an animal model of PKU induced by *p*-chlorophenylalanine and L-phenylalanine, Loo *et al.* (*133*) reported that the toxic effects of PKU were reproduced in rats treated with phenylacetate in amounts approximating those likely to be produced in phenylketonuria. They attributed the retarded growth of the body and brain of the young animal treated with phenylacetate to the formation of phenylacetyl coenzyme A in the tissues. They concluded that growth retardation caused by phenylacetate during the period of rapid development of the brain is responsible for the mental retardation in PKU.

In the variant forms of hyperphenylalaninemia, referred to as type II and III, the patients have a relaxed tolerance to dietary phenylalanine. They do not need a special diet and there is no risk of mental retardation. The infants show no symptoms at birth, which are detectable only as a result of mass screening. The mechanism for increased tolerance to dietary phenylalanine in this type of hyperphenylalaninemia is not known; the presence of small amounts of residual phenylalanine hydroxylase is one of several possibilities.

In the variant form due to transaminase deficiency or type IV hyperphenylalaninemia, which has been described by several workers (*135,136*), the patients do not show an abnormal phenylalanine level until they are exposed to a high protein diet. This condition is of special interest because it needs to be recognized at birth and treated with a low phenylalanine diet to avoid mental retardation. Also, the patient requires close monitoring to prevent a phenylalanine deficiency if a low phenylalanine diet is continued, after a change in phenylalanine tolerance occurs.

a. Hyperphenylalaninemia with Normal Phenylalanine Apohydroxylase Activity: Dihydropteridine Reductase Deficiency Type V. In this variant, which was recognized a few years ago (*136,137*), typical signs of PKU were present in the newborn period and, despite early detection and proper management, the neurological effects continued. Assays of enzymes in several of the patients revealed a normal hydroxylase activity, but marked deficiency of dihydropteridine reductase. Kaufman *et al.* (*138*) demonstrated the deficiency of the enzyme in brain and in

cultured skin fibroblasts. Differential diagnosis of hyperphenylalaninemia can be made by measurement of enzyme activity in cultured skin fibroblasts or in liver biopsies, but neither method is ideal. Milstein *et al.* (*139*) presented evidence that normal and classic PKU children excrete mainly tetrahydrobiopterin in their urine, whereas children with dihydropteridine reductase deficiency excreted only the oxidized forms of the biopterin. They described a rapid HPLC method for the measurement of various forms of biopterin in urine, which can be used as a screen for suspected cases of dihydropteridine reductase deficiency. Dihydropteridine reductase is also required for the hydroxylation of tryptophan and tyrosine to give 5-hydroxytryptophan (5-HT) and DOPA, respectively, which are precursors of the important neurotransmitters serotonin and norepinephrine. As a decreased level of these neurotransmitters may be the cause of the neurological deficits, Kaufman *et al.* (*138*) suggest providing substrates such as 5-HT and DOPA along with a low phenylalanine diet as a therapeutic approach.

Recently Kaufman (*140*) identified another type of defect that is due to marked tissue deficiency of all reduced derivatives of biopterin as a result of defective synthesis. The level of this cofactor was found to be only 20% of normal in the liver. The suggested treatment is the same as for dihydropteridine reductase deficiency.

Curtius *et al.* (*141*) have described the use of synthetic tetra-hydrobiopterin to discriminate between patients who have a hydroxylase deficiency and those with the socalled malignant hyperphenylalaninemia or dihydropteridine deficiency. A rapid decrease in serum levels of phenylalanine following tetrahydrobiopterin administration occurs only in persons with defects in cofactor generation. The test is helpful in identifying persons who require further study and who may need preemptive care.

The type VI variant is similar to type V in that the hyperphenylalaninemia does not respond to dietary manipulation, but both the hydroxylase and reductase activities are normal. Bartholome and Byrd (*142*) treated a patient with this variant hyperphenylalaninemia, who had an undefined block in dihydropteridine reductase with DOPA, 5-HT, and carbidopa, to compensate for the deficient synthesis of neurotransmitters, and noted improvement after 9 months of treatment. It must be noted that there are serious risks involved in the use of L-dopa and 5-HT, especially in persons in whom brain development is occurring. It may be acceptable to take these risks in persons who are destined to suffer from devastating neurological disease, but not in those who are expected to be normal or nearly normal in development.

b. Hyperphenylalaninemia without PKU. In patients with hyperphenylalaninemia without PKU, who comprise about one-quarter of the hyper-

phenylalaninemics, even with unrestricted dietary intake, the plasma phenylalanine levels are below 1 m*M*, and significant amounts of phenylpyruvate and its derivatives are not formed. In this variant form, normal development takes place in the absence of treatment (*113*). Clearance of phenylalanine following a loading test is found to be slower than normal in the probands, but faster than in homozygotes with classic PKU. Residual hydroxylase activity (10–20%) was found in material from liver biopsy and could be increased at least threefold by the addition of synthetic cofactor. Kaufman (*143*) has suggested that phenylalanine causes inhibition of a mutant form of hydroxylase in the presence of the natural coenzyme *in vivo*, and this inhibition could perhaps be overcome by the use of a synthetic cofactor.

c. Transient Hyperphenylalaninemia with Hypertyrosinemia. In a small percentage of births, especially in premature births, transient hyperphenylalaninemia accompanied by transient tyrosenemia is a normal common occurrence. This condition is the result of the functional immaturity of phenylalanine-hydroxylating and tyrosine-oxidizing enzymes. Both transient hyperphenylalaninemia and transient tyrosinemia are benign, but followup investigation of patients is needed to distinguish this condition from other hyperphenylalaninemias that may be harmful.

4. Diagnosis

In the early months of life, there are no symptoms in patients with hyperphenylalaninemia, but the trait can be diagnosed by tests that detect increased levels of phenylalanine or its metabolites in blood or urine. The $FeCl_3$ test for detection of phenylpyruvic acid in urine is not an accurate or reliable test, and early diagnosis of hyperphenylalaninemia can only be made by determining the blood phenylalanine levels, which are abnormal after the infant has been fed protein. From 8 to 10% of PKU patients are likely to be missed if the PKU testing is done before the fourth day of birth, a time when adequate levels of phenylalanine may not have yet accumulated. Dietary treatment for the disease should be started before 1 month in order to be effective.

The requirements and criteria for a screening program have been discussed under the topic of genetic screening in this chapter. A positive test in the screening program should be confirmed by several criteria needed to establish the diagnosis of PKU or to distinguish any variant forms. These criteria (*113*) include (1) sustained hyperphenylalaninemia (>1 m*M*) on a normal dietary intake, (2) no increase in plasma tyrosine levels after challenge with phenylalanine, (3) formation of phenylpyruvic acid and its derivatives, and (4) decrease in plasma phenylalanine concentra-

tion to near-normal levels, when dietary phenylalanine intake is restricted to 250–500 mg/day.

5. *Treatment*

Patients with PKU are normal at birth, but in the first few months, retardation begins and progresses to a severe state. Dietary restriction of phenylalanine is the only practical mode of therapy of PKU at this time. It has been well documented that a regulated diet low in phenylalanine is effective in treatment of PKU, if initiated soon after birth, and that a delay in the start of therapy can impair later cognitive development (*114,144*). It has been shown that the average IQ for patients treated early approaches the normal range, and tends to be higher than the average for untreated infants or infants treated late. The infant is given a low phenylalanine diet as the primary source of protein and supplemented with other foods. The monitoring of blood phenylalanine level is necessary for adequate treatment and is maintained between 3 and 10 mg/dl. Too low of a phenylalanine intake may also be harmful for the proper growth and development of the infant. There is no consensus as yet on the age at which dietary restriction can be stopped. Horner *et al.* (*145*), Vanderman (*146*), and Solomons and co-workers (*147*), among others, recommend an early termination. Cabalska (*148*), in a study of 22 children with PKU, who were placed early on a diet therapy that was terminated at age 5, found a fall in IQ of up to 31 points 4–6 years later. On the other hand, important improvements in personality traits were reported when the diet was terminated at age 8 (*149*). Smith *et al.* (*150*) recently reported the comparison of IQ between PKU patients treated at two medical centers, where they were put on a normal diet between the ages of 5 and 15 in one center or placed on a relaxed low phenylalanine diet in the other. Their results suggested that complete withdrawal of the low phenylalanine diet during childhood leads to a fall in intellectual progress in many patients.

B. Cystinuria

Cystinuria is a rather common inherited disorder of the membrane transport of amino acis that affects the epithelial cells of renal tubules and the intestines. This disorder has captured the attention of biochemists, physiologists, geneticists, and clinicians for years because of features of interest for all these specialities. This disorder is inherited as an autosomal recessive trait and is characterized by formation of urinary tract calculi caused primarily by precipitation of cystine, which is also excreted in the urine. Cystinuria was one of the four diseases studied by Garrod in his postulation of the concept of inborn errors of metabolism, and was

considered by him to be due to a defect in the metabolism of cystine. Our present understanding of cystinuria resulted after the development of techniques such as paper chromatography and polarographic and microbiological assays (*151*), leading to the recognition that lysinuria, argininuria (*152*), and ornithinuria (*153*) accompanied cystinuria. Dent and Rose (*5*) observed the structural similarities between cystine and the dibasic amino acids, and postulated that they share a single renal transport mechanism, which is defective or absent in cystinuria. Milne and co-workers (*154*) demonstrated that, as in Hartnup's disease, reduced intestinal absorption of dibasic amino acids is present in cystinuria patients.

1. Incidence

The overall incidence of cystinuria has been indicated to be 1 in 7000 by Levy (*155*), making it one of the most common errors of metabolism. Although the incidence of the disease is as low as 1 in 100,000 in some populations such as Sweden, in other populations such as the Jews of Libyan origin, homozygous cystinuria is a common inherited disorder with a frequency of 1 in 2500. The incidence is equal in males and females, but males are more severly affected and have a higher mortality rate as is true of all types of renal lithiases.

2. Transport of Cystine and Dibasic Amino Acids

The transport of small molecules such as amino acids and monosaccharides across membranes is an active process mediated by mechanisms involving participation of carrier molecules that serve a catalytic function. There are five main transport systems involved with reabsorption of amino acids in the renal tubules, which are concerned with neutral amino acids, dibasic amino acids and cystine, dicarboxylic acids, imino acids and glycine, and β-amino acids (*16*). Transport of cystine and dibasic amino acids in intestines, kidney, leukocytes, and cultured fibroblasts shows important differences. In the intestine, cystine and dibasic amino acids are transported by a common mechanism as determined by studies of intestinal absorption in cystinurics (*154*), which show increased amounts of the decarboxylation product of the unabsorbed amino acids and the diamines putrescine, agmatine, and cadaverine in the stool and urine. It has been found from feeding experiments that homozygous cystinurics can be divided into three groups(types I–III). In the majority of patients, accumulation of neither cystine nor dibasic amino acids takes place against a gradient. In some patients active accumulation of cystine, but not of the dibasic amino acids can be detected, and in others accumulation of both takes place, but not to the full extent (*151*). It, therefore,

appears that there are at least three different mutations that lead to a common clinical phenotype in cystinuria, accounting for the observed gentic heterogeneity. There is no defect in cystine accumulation in any of these conditions.

The renal transport system for these amino acids is more complex and shows important differences from the intestinal transport mechanism. The common transport mechanism is not the only mode of renal transport of dibasic amino acids, as suggested by a number of observations (*113*), and at least two other transport mechanisms exist in the kidney, one for the reabsorption and secretion of cystine and the other for reabsorption of dibasic amino acids, but not of cystine. The renal transport of cystine, as observed in kidney slices from cystinuric patients, is normal in contrast to the defective transport in the intestinal mucosa (*151*). Also, the active transport of lysine is absent in the mucosa of cystinurics, but only partially inhibited in kidney slices. The transport of either dibasic acids or cystine does not show any defect in leukocytes, erythrocytes, or skin fibroblasts in culture. These cells accumulate the dibasic amino acids by a transport process common to them, but not shared by cystine.

3. Clinical Features and Symptoms

The major clinical features of cystinuria are cystalluria and calculus formation, due to the low solubility of cystine. Cystine calculi have been detected in the first or second year of life, but the symptoms appear most commonly during the third and fourth decade of life. The dibasic amino acids (lysine, arginine, and ornithine) and cysteine–homocysteine mixed disulfides are also present in excess in the urine. Colic is a common presenting clinical symptom, and can be associated with obstruction of the urinary tract which leads to infection and eventual loss of renal function.

4. Diagnosis

The diagnosis of cystinuria is relatively easy to make and would be suspected in a patient who presents with the clinical symptoms of calculi or symptoms suggestive of calculi. Microscopic examination of first-morning concentrated urine may show hexagonal flat crystals. The cyanide–nitroprusside test is a simple diagnostic test for this disorder that has been widely used a a screening procedure (*156*). Cystine in an alkalinized urine sample is reduced to cysteine by sodium cyanide. The cysteine thus formed gives a magneta red complex on the addition of sodium nitroprusside. The reaction allows easy detection of homozygotes who excrete 600–1300 mg cystine/24 hr, but is weakly positive in only those heterozygotes who excrete a modest or moderate excess of cystine. A positive test

is also given by acetone, certain drugs, and homocystinurics; however, it is possible to easily differentiate them from cystinuria. Confirmation of a positive screening test should be made by demonstration of a characteristic excretion of cystine, lysine, arginine, and ornithine by paper chromatography and quantitative amino acid determination by column chromatography. Recently, Kinoshita *et al.* (*157*) have reported a new screening test for cystinuria that utilizes the dark-brown coloration given by a neutral aqueous solution of cystine on addition of nickel ion and sodium hydrosulfite. Cystinuria can easily be distinguished from the far more serious disorder, cystinosis, which is characterized by accumulation of cystine in the tissues and an aminoaciduria distinct from that of cystinuria marked by the absence of dibasic amino acids.

5. Treatment

Treatment of cystinuria is directed at reducing excretion in the urine by dietary restriction and increasing the solubility of cystine by physical means or by conversion to a more soluble compound. The dietary approach, using diets low in methionine, has yielded variable and conflicting results, and is not considered to be the treatment of choice. The most effective means of treatment is retaining a high urine volume by high fluid intake, which should be continued throughout the 24-hr period to prevent formation of a concentrated urine during the night. Patient compliance with this mode of therapy, however, is not easy. The use of penicillamine (β-β-dimethyl cysteine) was introduced in 1963 by Crawhall and co-workers (*158*). It can produce the more soluble cystine–penicillamine by the disulfide exchange reaction and thus minimize stone formation. However, penicillamine does have some side effects, e.g., allergic reaction, development of a nephrotic syndrome, vitamin B_6 deficiency (*159*), and an increase in the ratio of soluble to insoluble collagen (*160*). The potential for serious side effects makes it necessary to watch the patients on penicillamine therapy closely for any signs of toxic reactions. Rosenberg and Scriver (*113*) suggest the use of penicillamine only for high-risk patients.

6. Animal Models

In order to comprehend the underlying abnormalities of membrane transport in cystinuria and in other inborn errors of transport, investigation of the basic molecular abnormalities of the isolated renal membranes is needed, which can be best done with suitable animal models. Three animal models have been suggested: the cystinuric dog, the Kenya genet, and the mink. Only one mink with cystine stones has been reported and the cystinuric state in this animal has been largely uninvestigated (*151*). The cystinuric dog is the only animal model for study. Canine cystinuria is

inherited as an X-linked trait and is quite variable in expression, some dogs showing the full human cystinuric pattern, some excreting cystine and lysine, and others showing isolated cystinuria. Cystinuric dogs resemble the type III human cystinuria (*161*) with normal intestinal cystine transport. Bovee and co-workers (*162*) studied the defect in renal reabsorption in cystinuric dogs by the renal clearance techniques and found a slight defect in reabsorption of dibasic amino acids and a wide spectrum of cystine reabsorption. Segal and Bovee (*163*) have described the characteristics of canine cystinuria and the Fanconi syndrome in Basenji dogs, and found a resemblance to the human counterpart in many ways, offering the possibility of using these animals in studies on the nature of the basic defects in cystinuria. A chemical model of cystinuria in humans and rats has been described (*164*) that utilizes feeding of the nonmetabolizable amino acid, cycloleucine, to produce urinary excretion of several amino acids, resembling the pattern seen in cystinuria.

C. Galactosemia

Galactosemia is a congenital disorder transmitted as an autosomal recessive trait that is characterized by a defect in the metabolism of galactose. Mason and Turner (*165*) first described this syndrome in 1935, and the disease entity was clearly established after numerous other cases followed. This inborn error of metabolism particularly exemplifies the suitability and usefulness of routine newborn screening because any clinical complication that may develop in early infancy can be completely prevented or reversed by early therapeutic intervention with a relatively simple diet (*166*).

1. Metabolic Pathway of Galactose

Galactose in the humans is mainly derived by the hydrolysis of milk disaccharide lactose by lactase in the intestines. The main pathway of galactose utilization involves the conversion of galactose, under the influence of galactokinase, to galactose-1-phosphate which reacts with UDPglucose to give UDPgalactose and glucose-1-phosphate in a reaction catalyzed by galactose-1-phosphate uridyltransferase (*167,168*). The UDPgalactose undergoes epimerization of the hydroxyl group at carbon-4 to form UDP glucose catalyzed by an epimerase. Under conditions in which high levels of tissue galactose may be present, the sugar is reduced to galactitol by aldose reductase by an alternative pathway. Studies in patients with transferase deficiency galactosemia indicate that another pathway in which galactose is oxidized to galactonic acid may be of some importance in such patients. In older galactosemic patients, the transferase step may be by-

passed by the formation of UDPgalactose by the interaction of galactose-1-phosphate and UTP.

Three separate inborn errors of galactose metabolism are now known to occur in man depending upon which enzyme deficiency is present. Transferase deficiency was described the earliest and is the most commonly recognized disorder usually referred to as galactosemia. The other two forms of galactosemia are due to deficiency of galctokinase or epimerase. Estimates of the incidence of various forms of galactosemia vary from country to country, ranging from 1 in 16,000 to 1 in 180,000 (*167*). The overall incidence of galactosemia based on routine screening in several countries has been found to be 1 in 62,000 (*169*).

2. Transferase Deficiency Galactosemia

In this form of galactosemia, accumulation of galactose takes place in the blood and urine of the patient, galactose-1-phosphate is found in the erythrocytes, and galactitol is excreted in the urine. Galactitol is also present in ocular lenses and other body tissues (*167,168*). Clinically, this syndrome is marked by failure to thrive, vomiting, jaundice, liver disease, and cataracts. In untreated patients there may be cirrhosis, mental retardation, and a high frequency of sepsis due to *E. coli,* which leads to death (*52*). Some homozygous individuals may be capable of metabolizing galactose, and are found to be devoid of any symptoms while ingesting milk. Albuminuria and generalized aminoaciduria as manifestations of renal toxicity syndrome may be present.

3. Diagnosis, Screening Tests, and Management

The diagnosis of galactosemia is suggested by the presence of reducing substances in the urine that are negative for glucose by a specific test such as glucose oxidase. Newborn screening for galactosemia is not as common or as simple as for PKU, but the therapeutic outcome can be greatly improved by the institution of such programs. In order to be effective, screening for galactosemia should be performed on a blood specimen no later than the third or fourth day after birth. Levy and Hammersen (*169*) have recently reviewed the subject of newborn screening for galactosemia.

An inhibition assay similar to the Guthrie PKU assay detects increased levels of galactose in the blood, utilizing the dried filter paper blood specimen obtained for PKU. The main disadvantage of the auxotrophic assay is the gradual loss of sensitivity of the bacteria to galactose.

The most reliable and effective test for detection of galactosemia is the assay developed by Paigen and Pacholec (*170*) that utilizes an epimerase-deficient strain of *E. coli.* If galactose is present in the blood specimen on

the dried filter-paper disc, the strain of *E. coli* will resist destruction by *C21* bacteriophage, and growth proportional to galactose concentration will take place. The assay can be used for other forms of galactosemia as well and detects galactose-1-phosphate, thus enabling the identification of the infant who has ingested little or no milk. Detection of galactose-1-phosphate can also be made on umbilical cord blood. Since this test will not distinguish between the different enzyme defects, identification of the type of galactosemia has to be made by another test.

There are a number of chemical tests for the detection of galactose, including a dipstick method. Beutler and Baluda (*171*) in 1966 introduced a spot test for galactosemia based on the detection of transferase activity by fluorescence of NADPH. This method is subject to some errors due to loss of transferase activity after exposure to heat and humidity (*172*). A positive test in a neonatal screening program needs to be confirmed by other suitable tests, so that a definitive diagnosis and course of treatment can be established. Confirmation of diagnosis is made by demonstrating the absence or deficiency of the transferase in the erythrocytes, leukocytes, or fibroblasts cultured from the skin biopsies (*173*), using tests such as the UDPglucose consumption assay (*174*). Transferase activity in erythrocytes should be measured soon after collection of blood, as rapid deterioration in the activity may take place on storage (*175*). The consumption assay is based on the determination of UDPglucose before and after incubation with galactose-1-phosphate and red cell hemolysate, using the formation of NAD by the action of UDPglucose dehydrogenase as an endpoint. Using this assay, homozygotes have been found to be devoid of erythrocyte transferase activity; heterozygous carriers have intermediate activity (*176*). Several reports of prenatal diagnoses of galactosemia have appeared (*177–179*) from observations of deficiency in transferase in the amniotic cells. Schapira *et al.* (*180*) monitored three pregnancies at risk by deficiency in the transferase, and reported the first prenatal diagnosis of a "double heterozygote." For those cases in which pregnancy is continued despite a prenatal diagnoses of galactosemia, the diagnosis permits a better monitoring of pregnancy and a precocious treatment of the newborn.

Management of the galactosemic patient is based on the elimination of galactose from the diet as completely as possible for at least the first two years of life (*181*). Fishler *et al.* (*182*) have recently reported on the developmental status of 60 galactosemic infants, including their subsequent intellectual level, school status, visual perceptual skills, and EEG results. The results were consistent with earlier observations that better progress is found in those individuals whose dietary control is instituted at the youngest age. Failure to institute a dietary regimen will usually lead to

progressive hepatic failure and eventual death. In view of the observations that galactosemic syndrome is present at birth (*183*) and that severe damage to the fetus may occur if the mother is mildly galactosemic, galactose intake by the mother should be avoided during such pregnancies. Schwarz (*184*) has advocated assays of erythrocyte galactose-1-phosphate for monitoring adequate dietary control. In patients who are placed on the restricted diet at an early stage, most of the symptoms of the syndrome are reversed, but the deficit in intelligence is not totally restored.

4. Variants of Transferase Deficiency Galactosemia

Several variant forms of the transferase deficiency have been identified that show abnormalities in the electrophoretic mobility of the enzyme and varied ranges of clinical findings. In the socalled Negro variant, there is no detectable transferase activity in the erythrocytes (*185*), but about 10% of normal activity is present in liver and intestines. Black patients with this syndrome are still able to metabolize galactose, and may show a milder clinical manifestation of the disease than whites with galactosemia. In the Duarte variant described by Beutler *et al.* (*186*), there is a diminished red cell transferase activity, but no clinical symptoms. The variant enzyme is indistinguishable from the normal, except that it has a faster electrophoretic mobility on starch gel electrophoresis (*187*). The Rennes (*188*) and Indiana (*189*) variants, both characterized by decreased transferase activity and a slower electrophoretic mobility of the enzyme, show almost all the signs of galactosemia and are usually caused by a double heterozygous state. On the other hand, in the variant termed Los Angeles (*190*), the transferase activity is elevated in both homozygotes and heterozygotes and an electrophoretic pattern similar to that observed with the enzyme from the Duarte variant. No clinical abnormalities are associated with this variant.

5. Mechanism of Biochemical Defect

The abnormalities observed in galactosemia are entirely a consequence of the deranged metabolism of galactose leading to an accumulation of galactose-1-phosphate and galactitol, and have been mainly studied in experimental animals fed high galactose diets (*191,192*). The nature of the toxicity and the underlying biochemical cause for it may differ from organ to organ, depending on the metabolic pattern of the organ. Although the cause of the entire range of toxic clinical manifestations of galactosemia is not known, it seems conclusive that at least the cataract formation is a consequence of galactitol formation in the lens. Several metabolic alterations, such as the effects on protein synthesis (*193*), amino acid transport

(*194*), and glutathione levels (*195*), are known to take place in the lens of galactosemic patients along with the cataract formation. The specific toxic effects of galactosemia on nervous tissue may also be due to galactitol. Patients with transferase-deficiency galactosemia have galactitol in their brains (*196*), this polyol is also present in high concentrations in the brains of experimental animals. It has been suggested that decreased levels of serotonin in the brain is the underlying cause of mental disorders (*197*) in various toxic states, and this condition may be the result of decreased serotonin receptors in galactosemia (*198*). Deranged galactose metabolism may also interfere with the formation of galactose-containing cerebral lipids.

Important earlier work has been done to delineate the specific molecular basis of the genetic abnormality in galactosemia. It was reported by Tedesco and Mellman (*199*) that a protein from transferase-deficient erythrocytes is immunologically identical to the active enzyme, but devoid of catalytic activity. All the available data indicate that inactive transferase results from a structural gene mutation involving an amino acid near the active site.

6. *Galactosemia Due to Galactokinase and Epimerase Deficiencies*

Galactokinase-deficiency galactosemia was described by Gitzelmann (*200*). It is a syndrome inherited as an autosomal recessive trait, marked by galactose accumulation in the blood and urine and galactitol accumulation in the urine and ocular lenses. In this disorder, in contrast to transferase galactosemia, galactose-1-phosphate does not accumulate, and the only clinical feature is the development of cataracts appearing early in life, possibly as a toxic consequence of galactose or galactitol accumulation. However, recently Segal *et al.* (*201*) studied two brothers with galactokinase deficiency, who, in addition to cataracts, had severe mental retardation. The disease can be detected in infancy by screening tests that measure increased levels of galactose in urine or blood. The diagnosis is confirmed by the finding of normal activity of the transferase and no detectable activity of the galactokinase in erythrocytes. The estimate of the incidence of this disease is 1 in 40,000 (*202*). Early diagnosis of the disease is desirable by timely initiation of therapy in order to prevent the progression of cataract formation to blindness. Heterozygotes may also be at risk for developing cataracts (*203*).

The differences in the toxic effects observed in the two kinds of galactosemia have contributed to our understanding of the basic biochemical abnormality responsible for the toxicity in each syndrome. As no symptoms involving the kidney, liver, or brain are observed in galactokinase-defi-

ciency galactosemia, which shows no accumulation of galactose-1-phosphate and since galactitol is present in the ocular lenses in both types, it can be inferred that galactose-1-phosphate is responsible for the toxic symptoms in organs other than the lens.

In galactosemia, due to the epimerase deficiency described recently (*204*), the epimerase activity was found to be low or absent in erythrocytes and leukocytes without any presenting clinical symptoms, whereas it was normal in the liver and the cultured skin fibroblasts. Normal to slightly increased concentrations of galactose and greatly elevated levels of galactose-1-phosphate were present in blood, without any red cell abnormality. Epimerase-deficiency galactosemia is thus a benign disorder.

D. GM_2 Gangliosidosis (Tay–Sachs Disease)

GM_2 gangliosidosis, commonly known as Tay–Sachs disease (TSD), is inherited as an autosomal recessive trait and is a progressive, untreatable neuromuscular disorder, which invariably proves fatal. This disorder provides a classic example of the elucidation of the underlying molecular basis of an inborn error of metabolism and its potential control as a result of contributions and approaches from multiple disciplines. Tay–Sachs disease is a prototype of the neuronal sphingolipidoses and, more importantly, of the prenatal diagnosis, prevention, and control of serious metabolic disorders. It is the most frequently diagnosed prenatal inborn error of metabolism. Although TSD has been reported in many non-Jewish groups, it occurs predominantly in a defined population, being 100 times more common in infants of Ashkenazi Jewish ancestry, with an incidence of 1 in 3600. The recessive gene is estimated to be carried by 1 in 30 Jewish individuals in the United States compared to 1 in 300 persons from the general population (*61*).

The clinical findings of the cherry-red macular degeneration in the fundus of an affected infant suffering from progressive neuromuscular degeneration was first described by Tay, a British opthalmologist, in 1881 (*205*). Bernard Sachs in 1887 (*206*) gave a detailed description of the cortical pathology in the disease and later concluded it to be a familial disorder that was usually fatal. The excessive accumulation of lipids in the brain of children with TSD was found to belong to a class of neuraminic acid-containing sphingolipids, referred to as gangliosides by Klenk (*207*). Svennerholm (*208*) identified the structure of GM_2-ganglioside in 1962, known as ceramide-glucose-galactose-(*N*-acetylneuraminic acid)-*N*-acetyl galactosamine; further characterization of the lipid was conducted by Ledeen and Salsman (*209*). Following the concept of lysosomal storage diseases by Hers (*30*) and the demonstration of deficiency of specific lysosomal hy-

drolases in the first two neuronal storage diseases, Gaucher's disease and Niemann–Pick disease, it was speculated that a deficiency of lysosomal hexosaminidase necessary to cleave the terminal *N*-acetyl galactosamine from GM_2-ganglioside may be the defect in TSD, accounting for the intralysosomal glycolipid accumulation.

The presence of at least two acidic hexosaminidase isozymes in human spleen was shown by Robinson and Stirling in 1968 (*210*). In the same year, Sandhoff *et al.* (*211*) reported their clinical findings in a child with apparent TSD, who showed evidence of visceral storage of gangliosides and was completely lacking in lysosomal hexosaminidase, in contrast to the normal hexosaminidase activity in children with TSD. In 1969, the very important observation of the underlying metabolic defect of TSD was made by Okada and O'Brien (*212*), which showed that in tissues from children with this disorder, the activity of one of the isozymes of hexosaminidase, called hexosaminidase A or "hex A," was nearly absent. The deficient activity of the isozyme was also demonstrable in leukocytes and serum of affected children and, most importantly, the isozyme activity was partially deficient in the serum and leukocytes of heterozygotes (*213*). The isozyme deficiency was shown by using *p*-nitrophenyl-4-methylumbelliferyl and naphthyl ASBI derivatives of both β-D-*N*-acetylglucosamine and β-D-*N*-acetylgalactosamine as synthetic substrates (*214*). Brain activity of the other isozyme hexosaminidase B or "hex B" is elevated severalfold, probably as a result of lysosomal stimulation following ganglioside storage.

Shortly following the discovery of hex A deficiency in tissues as the molecular basis of TSD, the same defect was demonstrable in cultured skin fibroblasts from affected infants (*215*). The demonstration of hex A activity in cultured amniotic fluid from normal pregnancy led to the development of a uniformly successful method of prenatal diagnosis of the disease by several groups of workers (*216–218*). With the use of a fully automated hexosaminidase assay based on heat inactivation of the more heat-labile hex A (*60*), it became feasible to screen a large number of child-bearing-age Ashkenazi Jews, who comprise a higher-risk group for detection of heterozygotes. Both hex A and hex B are found in all normal human tissues (except red blood cells) and in body fluids such as leukocytes (*212*), serum (*213*), urine (*219*), and tears (*220*).

Although the more commonly used assay for heterozygote detection is based on the heat lability of hex A and a heat inactivation step to distinguish hex B from total hexosaminidase activity, some laboratories have utilized an automated pH inactivation method (*221*) or electophoretic (*222*) method to separate the major hexosaminidases. The assay results in the heat inactivation assay can be affected by relatively small changes in

pH, ionic strength, and temperature; minor changes in reaction conditions may cause sera from noncarriers to resemble heterozygotes, and vice versa. The major drawback of the serum assay is the problem of false positives. When the total hexosaminidase activity is abnormally high, as is the case in certain chronic diseases, in patients on oral contraceptives, or in pregnancy, the ratio of hex A to hex B is no longer a reliable indicator of the carrier state. Nonetheless, serum hexoasaminidase assay in the heterozygotes can accurately assign the correct genotype to 96% of tested individuals, and the false positive rate falls to less than 1% when the patients in the inconclusive range are retested by the more accurate leukocyte (*214*) assay. It has been suggested that tears (which require no further preparation) might be a more convenient and accurate source of specimen for the hexosaminidase assay. Nakagawa *et al.* (*223*) have reported a method for heterozygote detection during pregnancy using platelet extracts for hexosaminidase isozyme assay, which is claimed to be better than the leukocyte assay. The probability that a heterozygote would be missed by the available methods has been estimated to be 1 in 30,000 (*214*). With appropriate precautions and use of follow-up testing, both false positives and false negatives would be extremely rare.

As is clear from the above discussion, it is relatively easy to identify the heterozygotes for TSD at risk because of the availability of easy and inexpensive hexosaminidase assays, and also because the high-frequency gene is present in a defined and accessible population (Ashkenazi Jews). Although no therapeutic strategy exists for TSD at the present time, high-risk couples identified through screening could prevent the birth of a child with TSD by prenatal diagnosis and elective abortion of an affected foetus. The pitfalls of prenatal diagnosis are discussed by O'Brien (*214*), and should be carefully guarded against.

By the end of 1976, screening programs were available in 60 cities in the United States and in several countries around the world (*224*). The results obtained with such programs have already been referred to under genetic screening in this chapter. Various aspects of TSD and the screening experiences in five countries have been reviewed in the monograph edited by Kaback *et al.* (*224*).

1. Clinical Symptoms

Children with TSD do not exhibit any symptoms during the first few months after birth and appear normal. One of the earliest signs is the "startle" response to sudden sharp noises. There may be feeding difficulties, hypotonia, and spasticity. Between the ages of 3 and 6 months, signs of motor weakness appear, and by 7–8 months of age, lack of coordinated motor activity is easily noticeable. The child becomes progressively less

attentive and may suffer convulsions. Blindness may result in some children. A cherry-red spot can be seen in the macular region of the eye within the first year of life. There is an increase in the head size at about 2 years of age, the cranium becoming 50% greater than normal. The child does not generally survive 3 or 4 years of life. At the time of death no enlargement in the size of organs is noticeable. Accumulation of GM_2 ganglioside and Asialo Tay–Sachs ganglioside (GA_2) takes place primarily in the central nervous system.

2. *Variants of TSD*

Several variants of TSD, or more correctly of *N*-acetyl β-hexosaminidase deficiency, have been found that have very similar clinical manifestations of the disease, but which show varied activities of the hexosaminidase isozymes in tissues and body fluids. As already mentioned, in contrast to TSD that is marked by deficient activity of hex A, in the major variant, Sandhoff's disease, activities of both hex A and hex B are deficient. Both traits show 100–300 times elevated levels of GM_2 ganglioside and increased levels of Asialo Tay–Sachs ganglioside (GA_2) as well. Sandhoff's disease is not confined to Ashkenazi Jews, for patients of non-Jewish extractions have been reported. In juvenile GM_2 gangliosidosis, the onset of the disease is late and the course of the disease is more protracted than in the infantile form. The milder accumulation of GM_2 ganglioside is consistent with the presence of only a partial deficiency of hex A in the tissues of patients, varying from as low as 10% to greater than 50%, having no relationship between residual enzyme level and the clinical course of the disease (*225*). Assays of hex A in patients with this disorder have, however, revealed much greater deficiency when GM_2 is used as a substrate (*214*). A single case of variant AB is reported (*226*) with TSD phenotype in which the tissue activities of hex A and hex B towards synthetic artificial substrates were either normal or elevated, but deficient activity towards the natural substrate GM_2 was present. The puzzling feature of this variant is the increased storage of GA_2 in the presence of of hex A and hex B with normal GA_2-degrading activity. A chronic GM_2 gangliosidosis was reported in adults by Rapin (*227*), with slowly progressive symptoms, indicative of spinocerebellar degeneration and biochemically marked by greatly decreased hex A in serum and leukocytes. Significant deficiency of hexosaminidase, as assayed with synthetic substrates, has been reported in normal healthy adults (*228,229*). It has been suggested that these clinically normal individuals with a deficiency of hex A have allelic compounds (double heterozygotes) (*214*). Tallman *et al.* (*230*) have discussed the impossibility of distinguishing prospective parents of healthy, hex A-deficient children from TSD carriers,

based on the use of artificial substrates in screening programs and prenatal diagnosis. However, due to the rarity of this variant, the practical implication of this is insignificant. Deficiency of both hex A and hex B in a normal adult has also been reported (*231*).

3. Molecular Genetics

Both hex A and hex B have a molecular weight of approximately 100,000. Antibodies raised against hex A or hex B show immunological cross-reaction with the other isozymes, suggesting that the two isozymes share at least one commón antigenic determinant (*232*). Kinetic studies (*233,234*) indicate that hex A cleaves GM_2 ganglioside, but hex B does not. The observation that hex A has a specific antigen not found in hex B led Srivastava and Beutler (*235*) to propose that hex A might be a heteropolymer composed of two different types of subunits α and β, whereas hex B is a homopolymer made up of identical β subunits. Hybridization studies indicate that the locus involved in the expression of hex A is assigned to chromosome 15, and that of hex B to chromosome 5 (*214*). According to O'Brien (*214*), the genetic defect in TSD appears to be a structural mutation of the gene at chromosome 15, which codes for an α subunit, whereas in Sandhoff's disease the defect seems to be of a gene on chromosome 5, coding for the β subunit.

4. Therapy

At the present time, there is no therapy available for TSD, the control of the disease being directed towards preventing the birth of an affected child. The problems associated with enzyme replacement therapy (discussed in Section IX,A) are particularly difficult in TSD because of several considerations. The administered enzyme needs to be taken not only into the brain tissue, but specifically into the lysosomes, where the GM_2 ganglioside accumulates (*236*). Even if this could be accomplished, the replacement therapy would have to be initiated early *in utero* because pathological accumulation of the ganglioside starts in fetal life. Because of the complexities in evaluating the potential of therapeutic modalities for TSD in humans, preliminary studies should be undertaken with appropriate animal models. The domestic cat has been used as an animal model for Sandhoff's disease, and is well-suited for the development of enzyme-replacement therapeutic approaches (*237*). Inhibition of the rapid disappearance of administered hexosaminidase, due to hepatic uptake, appears useful in inducing permeability of the blood–brain barrier by exposure of animals to hyperbaric oxygenation (*98*). Brooks *et al.* (*238*), working with TSD brain cells in culture, found that concanavalin A, when bound to the cell membrane, functions as an artificial enzyme and mediates the uptake

of significant amounts of exogenous hex A into TSD glial cells. Ultrastructural and biochemical evidence of GM_2 ganglioside breakdown has been presented. Although a few theoretical approaches to either replacement of hex A or depletion of accumulated gangliosides are possible (*236*), the outlook for a succeesful therapeutic approach to TSD appears bleak. TSD, however, remains a prime example of prevention and control of a lethal disorder by means of genetic screening and prenatal diagnosis.

E. Congenital Hypothyroidism

Congenital hypothyroidism, or the more inclusive infantile hypothyroidism (*239*), is one of the most common causes of mental retardation, with an approximate incidence of 1 in 4000 (*240*). However, mental retardation is preventable and the prognosis for developing a normal IQ is good, if the infant is treated before the age of 3 months (*241*). Approximately 80% of the infants with hypothyroidism, who were diagnosed and treated before 3 months of age, had IQs greater than 90, whereas leass than 50% of those diagnosed and treated at a later time achieved this level of IQ (*241,242*). An early diagnosis of infantile hypothyroidism is difficult to make clinically because of the nonspecificity of symptoms, such as feeding difficulties, constipation, and hypothermia, which can be easily missed. It has been estimated that only about 10% of hypothyroid infants are diagnosed clinically before 2 months of age (*243*). The need for an early detection of this condition and the difficulties associated with making a diagnosis on the basis of the clinical manifestations of the disease underscore the benefits of a screening program, which is now recommended by the American Thyroid Association (*244*). The progress in the development of methodologies for the measurement of thyroid hormones by radioimmunoassay has made it possible to accurately measure thyroxine and TSH, allowing analysis on less than 20 μl of blood, thus making possible the establishment of screening procedures. Before discussing the current status of screening for infantile hypothyroidism, it is important to briefly present an account of the development of thyroid function in the human fetus as well as to outline the steps involved in thyroid hormone synthesis, in order to understand the various testing procedures.

The thyroid follicular cell synthesizes the distinctive protein thryoglobulin, and also accumulates iodide from the extremely small amounts present in the blood in a highly efficient manner, followed by iodination of tyrosyl residues in the thyroglobulin by the action of thyroid peroxidases, giving rise to mono- and diiodotyrosines. The synthesis of the thyroid hormones triiodothyronine and thyroxine takes place principally within the thyroglobulin molecule, by the coupling of one mono- and one diiodotyrosine residue or two diiodotyrosine residues, respectively, without the in-

volvement of a specific coupling enzyme. The hormones are split off from the peptide linkage in the thyroglobulin by deiodinases. Thyroxine circulates in the blood, bound principally (75%) to thyroxine-binding globulin and, to a lesser extent, to prealbumin and albumin. The major route of thyroxine and triiodothyronine metabolism is deiodination. All reverse T_3 is produced by deiodination of thyroxine. Different aspects of thyroid function and metabolism are comprehensively covered by Robbins and Gorden (*245*), by Stanbury (*246*), and by standard texts of endocrinology.

The developments in the field of fetal and newborn endocrinology during the last several years have affected the approach to diagnosis and treatment of the hypothyroid infant. It has been demonstrated by the work of many investigators that the human placenta is almost impermeable to thyroid hormones, and fetal serum concentrations of T_4, TSH, and free T_4 are very low until midgestation (*247*). Fetal development is thus entirely dependent on endogenous production of thyroid hormones. In contrast to T_3 levels, T_4 and TSH levels start to rise after midgestation, and, as has been recently recognized, preferential formation of metabolically inactive reverse T_3 from T_4 also takes place in the fetus. Immediately following birth, circulating levels of TSH, T_3, and T_4 are increased to a level considerably higher than in normal adults (*248*), but fall toward normal within 1 week of birth. Due to rapid changes in thyroid hormone levels within a few days of birth, it is important for the evaluation of test results to use the normal range for the post-natal age at which the specimen was taken. Since the TSH regulation of T_4 secretion in the human fetus is well-established in the third trimester of pregnancy, hypothyroidism at birth is marked by elevated TSH and low T_4 levels (*249*).

1. Inheritable Disorders of Thyroid Metabolism

Infantile hypothyroidism includes all children with hypothyroidism in whom brain damage may result as a consequence of the thyroid function abnormality and can be divided into at least three groups: (1) children who are born with hypothyroidism as a result of transplacental poisoning with drugs such as iodides, thiourea derivatives given to the mother, (iatrogenic group); (2) those infants without any demonstrable ectopic glands or very small glands in a normal location; and (3) infants with inheritable disorders of thyroid metabolism, comprising about 1/5 to 1/3 of all infantile hypothyroidism patients. In general, patients with inborn errors of thyroid metabolism are all hypothyroid, and have a goiter due to increased levels of TSH. Information about the biomedical basis of inheritable disorders began in 1950 with the report by Stanbury and Hedge (*250*) of a goitrous cretin with a defect in the organification of iodine in the thyroid. Stanbury and co-workers have made major and extensive contributions to this field.

The inheritable disorders of thyroid hormone metabolism can be generally due to (1) a defect in iodide concentration mechanism, (2) an iodination defect, (3) a defect in iodotyrosine deiodinase, or (4) a defect in hormone transport (*245*). A congenital decrease or elevation in thyroxine-binding globulin is inherited as a sex-linked dominant trait, with a decreased or increased synthesis of a normal protein, and generally is not associated with a pathology. Salvatore *et al.* (*251*) have recently reviewed the subject of inherited disorders of thyroid metabolism.

2. *Newborn Screening*

The earliest large-scale screening programs to test newborns for hypothryoidism were begun in Quebec and in Pittsburgh in 1973. Since then, such programs have increased in number and size, and a number of states have passed laws facilitating newborn screening, so that it is likely that all children born in the United States and Canada will be screened for hypothyroidism in the future.

Screening tests utilize filter paper impregnated with cord blood or capillary blood from an infant in the first week of life, which is frequently obtained for PKU screening. An RIA procedure is used to determine T_4 or TSH. Screening with TSH results in a smaller number of false positive, but will miss the relatively rare cases of hypothyroidism secondary to lack of TSH stimulation, which are, however, probably of no clinical significance. When T_4 measurement is used for screening, the approximately 2.5% false-positive patients require testing for TSH on the same specimen. Although TSH measurement is used for screening in some United States, Canadian, and European programs, all large-scale programs in the United States utilize T_4 testing, followed by confirmatory testing with a repeat T_4 and a TSH measurement (*239*). Neonatal T_4 and TSH analyses are now also routinely performed by hospital and commercial clinical laboratories due to easy availability of the radioimmunoassay reagents for such testing. The specimens for hypothyroid screening should preferably be obtained at least 5 days after birth to allow the elevated TSH levels following birth to subside. In the case of a positive result in screening, therapy should be initiated without waiting for the confirmatory results. Neonatal screening for hypothyroidism has been reviewed by Klein (*239*) and by Fisher (*247*).

Fisher *et al.* (*240*) have recently reported on the results of screening 1 million North American infants for hypothyroidism. As of August 1978, the incidence of primary hypothyroidism was 1 in 4254 (239 cases detected, 7 undetected) and that of secondary or tertiary (hypothalamic) hypothyroidism was 1 in 68,200 births. The incidence of goitrous hypothy-

roidism, as a consequence of an inborn error in thyroid hormonogenesis, was estimated to be 1 in 30,824 infants. The comprehensive program for mass screening in North America, at the present time, would use T_4 testing of filter-paper soaked blood-spot samples, with confirmatory follow-up testing by TSH measurement of the patients with the lowest 3% T_4 results (*240*). It is suggested that TSH may be more cost-effective and practical, although technically somewhat difficult for screening purposes.

The data regarding the mental development of infants, treated within 1 month of the detection of hypothyroidism, as a result of the screening programs is preliminary only at this time, but suggests that the infants develop normally. Klein (*239*) has critically reviewed the available data on the early treatment of infantile hypothyroidism and IQ.

3. Clinical Symptoms

Despite the availability of sensitive laboratory indicators of thyroid function, clinical examination continues to be an important part of diagnosis and follow-up. Nonspecific complaints such as skin mottling, persistence of mild jaundice, feeding difficulties, hoarse cry, and constipation or diarrhea may suggest hypothyroidism (*252*), and should prompt a physical examination and appropriate laboratory tests. Physical signs such as lethargy, hypothermia, dry skin, thick tongue, and umbilical hernia may be present in about 1/3 of the patients. Presence of goiter is usually diagnostic of hypothyroidism. A diagnostic workup is needed for those infants who present with suggestive physical or historical findings. A normal neonatal T_4 value does not necessarily rule out infantile hypothyroidism, as, for example, infants with an ectopic thyroid gland may have adequate or borderline thyroid function for months or years.

4. Treatment

Prompt treatment of infantile hypothyroidism is necessary for improving the prognosis of adequate mental development. Whether hypothyroidism is suspected on clinical grounds or as a result of positive screening results, therapy should be initiated without waiting for a definite diagnosis, since replacement therapy carries little risk. A number of replacement agents are available. The committee on drugs of the American Pediatric Association (*252*) recommends the use of synthetic levothyroxine as the drug of choice. Maintenance of T_4 levels (appropriate for the age) rather than suppression of TSH levels, should be the main criteria for dose adjustment. Criteria for response to therapy, dose adjustment, and drug interactions have been discussed in the report by the committee on drugs (*252*).

REFERENCES

1. Garrod, A. E., *Lancet*, **2,** 1,73,142,214 (1908).
2. Garrod, A. E., Lancet **2,** 1616 (1902).
3. La Du, B. N., Zannoni, V. G., Laster, L., and Seegmiller, J. E., *J. Biol. Chem.* **230,** 251 (1958).
4. Wang, Y. M., and Van Eys, J., *N. Engl. J. Med.* **282,** 892 (1970).
5. Dent, D. E., and Rose, G. A., *Quart. J. Med.* **20,** 205 (1951).
6. Beadle, G. W., *Science* **129,** 1715 (1959).
7. Tatum, E. L., *Science* **129,** 1711 (1959).
8. Pauling, L., Itano, H. A., Singer, S. J., and Wells, I. C., *Science* **110,** 543 (1949).
9. Ingram, V. M., *Nature (London)* **180,** 326 (1957).
10. Cori, G. T., and Cori, C. F., *J. Biol. Chem.* **199,** 661 (1952).
11. Jervis, G. A., *Proc. Exp. Biol. Med.* **82,** 514 (1953).
12. Marks, P. A., *in* the "Red Blood Cell", (C. Bishop and D. M. Surgenor, eds.) p. 225. Academic Press, New York, 1964.
13. McKusick, V. A., "Mendelian Inheritance in Man: Catalog of Autosomal, Dominant, Recessive and X-Linked Phenotypes." 5th ed., Johns Hopkins Press, Baltimore, Maryland, 1978.
14. Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S.," The Metabolic Basis of Inherited Disease," 4th ed., McGraw-Hill, New York, 1978.
15. Bondy, P. K., and Rosenberg, L. E., "Metabolic Control and Disease" Saunders, Philadelphia, Pennsylvania, 1980.
16. Raine, N. D., *in* "Chemical Diagnosis of Disease," (S. S. Brown, F. L. Mitchell, and D. S. Young, ed.) p. 927. Elsevier North-Holland Biomedical Press, New York, 1979.
17. Brock, D. J. H., *in* "Biochemical Genetics of Man," (D. J. H. Brock, and O. Mayo, ed.) p. 469. Academic Press, London, 2nd ed., 1978.
18. Goodman, R. M., and Motulsky, A. G., ed. "Genetic Disease Among Ashkenazi Jews," Raven, New York, 1979.
19. Goodman, R. M., "Genetic Disorders Among the Jewish People," Johns Hopkins Press, Baltimore, Maryland 1979.
20. Öckerman, P. A. and Lundblad, A., *Ann. Rev. Clin. Biochem.*, **1,** 303 (1980).
21. Watson, J. D., "The Molecular Biology of the Gene," Benjamin Press, New York, 1976.
22. Rosenberg, L. E., *in* "Metabolic Control and Disease," (P. K. Bondy and L. E. Rosenberg, ed.) p. 73, Saunders, Philadelphia, Pennsylvania, 1980.
23. Grumbach, M. M., Marks, P. A., and Morishima, A., *Lancet* **1,** 1330 (1962).
24. Seegmiller, J. E., Rosenbloom, R. M., and Kelley, W., *Science* **155,** 1682 (1967).
25. Lyon, M. F., *Am. J. Hum. Genet.* **14,** 135 (1962).
26. Danes, B. S., and Bearn, A. G., *J. Exp. Med.* **126,** 509 (1967).
27. Rosenbloom, F. M., Kelley, W. N., Henderson, J. F., and Seegmiller, J. E., *Lancet* **2,** 305 (1967).
28. Riccuiti, F. C., Gelehrter, T. D., and Rosenberg, L. E., *Am. J. Hum. Genet.* **28,** 332 (1976).
29. Motulsky, A., *Nature (London)* **194,** 607 (1962).
30. Hers, H. G., and Van Hoof, F., eds. "Lysosomes and Storage Diseases," Academic Press, New York, 1973.
31. Shih, V. E., "Laboratory Techniques for the Detection of Hereditary Metabolic Disorders," CRC Press, Cleveland, 1975.
32. Anderson, N. G., and Anderson, N. L., *Anal. Biochem.* **85,** 331 (1978).

33. Anderson, N. L., and Anderson, N. G., *Anal. Biochem.* **85,** 341 (1978).
34. Purkiss, P., Hughes, R. C., and Watts, R. W. E., *Clin. Chem.* **24,** 669 (1978).
35. Chester, M. A., Lennartson, G., Lundblad, A., Lundsten, J., Norden, N. E., Sjöblad, S., Svensson, S., and Öckerman, P. A., *Monog. Hum.Genet.* **10,** 2, 1978.
36. Efron, M. L., *N. Engl. J. Med.* **272,** 1243 (1965).
37. Tanaka, K., Budd, M. A., Efron, M. L., and Isselbacher, K. J., *Proc. Natl. Acad. Sci.* **56,** 236 (1966).
38. Hommes, F. A., *in* "Models for the study of Inborn Errors of Metabolism" (F. A. Hommes, ed.) p. 1, Elsevier/Holland Biomedical Press, Amsterdam, 1979.
39. Danes, B. S., Scott, J. E., and Bearn, A. G., *Proc. Natl. Acad. Sci. U.S.A.* **67,** 357 (1970).
40. Pena, S. D. J., and Wrogemann, K., *Pediatr. Res.* **12,** 887 (1978).
41. Mattieu, J. M., Brady, R. O., and Quarles, R. H., *J. Neurochem.* **22,** 291 (1974).
42. Phillips, P. R., Jenkins, C. S. P., Luscher, E. F., and Larrieu, M. J., *Nature (London)* **257,** 599 (1975).
43. "Genetic Screening: Programs, Principles and Research." Washington, D.C., Nat. Acad. Sci., 1975.
44. Bickel, H., Gerrard, J., and Hickmans, E. M., *Lancet* **2,** 812 (1953).
45. Guthrie, R., and Susi, A., *Pediatrics* **32,** 338 (1963).
46. Guthrie, R., *Hosp. Pract.* **7,** 93 (1972).
47. Efron, M. L., Young, D., Moser, H. W., and MacCready, R. A., *N. Engl. J. Med.* **270,** 1378 (1964).
48. Ireland, J. T., and Read, R. A., *Ann. Clin. Biochem.* **9,** 129 (1972).
49. Levy, H. L., *Adv. Hum. Genet.* **4,** 1 (1973).
50. Holtzman, N. A., *Pediatr. Clin. North Am.* **25,** 411 (1978).
51. Watts, R. W., *Monogr. Hum. Genet.* **9,** 204 (1978).
52. Levy, H. L., Sepe, S. J., Shih, V. E., Vawter, G. F., and Klein, J. O., *N. Engl. J. Med.* **297,** 823 (1977).
53. Holtzman, N. A., *Pediatrics* **62,** 1056 (1978).
54. Paul, T. D., and Guthrie, R., *J. Pediatr.* **96,** 653 (1980).
55. Aebi, H., and Suter, H., *Adv. Hum. Genet.* **2,** 143 (1971).
56. Kalow, H., "Pharmacogenetics, Heredity and the Response to Drugs." Saunders, Philadelphia, Pennsylvania, 1962.
57. Rosenberg, L. E., Durant, J. L., and Elsas, L. J., *N. Engl. J. Med.* **278,** 1407 (1968).
58. Weinberg, A. N., *Metab., Clin. Exp.* **10,** 728 (1961).
59. Westwood, A., and Raine, D. N., *J. Med. Genet.* **12,** 327 (1975).
60. Lowden, J. A., Skomorowski, M. A., Henderson, F., and Kaback, M. M., *Clin. Chem. (Winston-Salem, N.C.)* **19,** 1345 (1973).
61. Galjaard, H., *Ann. Clin. Biochem.* **16,** 343 (1979).
62. Kaback, M. M., Greenwald, S., and Brossman, M. S. W., "Report of the California Tay-Sachs Disease Prevention Program, Division of Medical Genetics," Torrance, California, 1979.
63. Hobbins, J. C., and Mahoney, M. J., *N. Engl. J. Med.* **290,** 1065 (1974).
64. Kan, Y. W., Golbus, M. S., and Trecartin, R., *N. Engl. J. Med.* **294,** 1039 (1976).
65. Raine, D. N., "Medicosocial Management of Inherited Metabolic Disease." Med. Tech. Publ. Co., Ltd. Lancaster, England, 1977.
66. Miles, J. H., and Kaback, M. M., *Pediatr. Clin. North Am.* **25,** 593 (1978).
67. Sampson, J. L., *Clin. Obstet. Gynecol.* **2,** 259 (1979).
68. Mahoney, M. J., *Clin. Perinatol.* **2,** 255 (1979).
69. Nadler, H. L., *Pediatrics* **42,** 912 (1968).

70. Kaback, M. M., Leisti, J. T., and Levine, M. D., "Prenatal Genetic Diagnosis in Endocrine and Genetic Diseases of Childhood and Adolescence," p. 1241. Saunders, Philadelphia, 1975.
71. Galjaard, H., *Cytogenet. Cell Genet.* **16,** 453 (1976).
72. Epstein, C. J., and Golbus, M. S., *Am. Sci.* **65,** 703 (1977).
73. NICHD National Registry for Amniocentesis Study Group: Midtrimester Amniocentesis for Prenatal Diagnosis: Safety and Accuracy, *J. Am. Med. Assoc.* **236,** 1471 (1976).
74. Golbus, M. S., Loughman, W. D., Epstein, C. J., Halbasch, G., Stephens, J. D., and Hall, B. D., *N. Engl. J. Med.* **300,** 157 (1979).
75. Powledge, T. M., and Fletcher, J., *N. Engl. J. Med.* **300,** 168 (1979).
76. Kleijer, W. J., Niermeijer, M. F., and Galjaard, H., *Monogr. Hum. Genet.* **9,** 217 (1978).
77. Rodeck, C. H., Campbell, S., and Eze, L., *in* "Proceedings of the Third European Conference on Prenatal Diagnosis" (J. D. Murken, S. Stengel-Rutowski, and E. Schwinger, eds.), p. 229. Enke, Stuttgart, 1979.
78. Mahoney, M. J., Haseltine, F. P., Hobbins, J. C., Banker, B. O., Caskey, C. T., and Golbus, M. S., *N. Engl. J. Med.* **297,** 968 (1977).
79. Galjaard, H., "Genetic Metabolic Disease; Early Diagnosis and Prenatal Analysis." Elsevier/North Holland Biomedical Press, Amsterdam, 1980.
80. Kan, Y. W., and Dozy, A. M., *Lancet* **2,** 910 (1978).
81. Williamson, B., *Nature (London)* **276,** 114 (1978).
82. Marx, J. L., *Science* **202,** 1068 (1978).
83. Berg, P., Baltimore, D., Berger, H. W., Cohen, S. N., Davis, R. W., Hogness, D. S., Nathans, D., Roblin, R., Watson, J. D., Weissman, S., and Zinder, N. D., *Science* **185,** 303 (1974).
84. Recombinant DNA Research Guidelines, *Fed. Regist.*, July 7, 1976.
85. O'Brien, D., *Hum. Nutr.* **1,** 441 (1979).
86. O'Brien, D., *Am. J. Clin. Nutr.* **32,** 482 (1979).
87. Raine, D. N., "The Treatment of Inherited Metabolic Disease." Med. Tech. Publ. Co., Ltd. Lancaster, England, 1975.
88. Boyer, S. H., Siggers, D. C., and Krueger, L. J., *Lancet* **2,** 654 (1973).
89. Brady, R. O., Pentchev, P. G., Gal, A. E., Hibbert, S. R., and Dekaban, A. S., *N. Engl. J. Med.* **291,** 989 (1974).
90. Neerunjun, E. D., and Gregoriadis, G., *Biochem. Soc. Trans.* **4,** 133 (1976).
91. Gregoriadis, G., and Allison, A., *FEBS Lett.* **45,** 71 (1974).
92. Gregoriadis, G., Putman, D., Louis, L., and Neerunjun, E. D., *Biochem. J.* **140,** 323 (1974).
93. Gregoriadis, G., and Buckland, R. A., *Nature (London)* **244,** 170 (1973).
94. Colley, C. M., and Ryman, B. E., *Biochem. Soc. Trans.* **2,** 871 (1974).
95. Braidman, I., and Gregoriadis, G., *Biochem. Soc. Trans.* **4,** 259 (1976).
96. Gregoriadis, G., *N. Engl. J. Med.* **295,** 765 (1976).
97. Gregoriadis, G., and Dean, M. F., *Nature (London)* **278,** 603 (1979).
98. Lanse, S. B. Lee, J. C., Jacobs, E. A., and Brody, H., *Aviat. Space Environ. Med.* **49,** 890 (1978).
99. Arias, I. M., Gartner, L. M., and Cohen, M., *Trans. Assoc. Am. Physicians* **81,** 66 (1968).
100. Kreek, M. J., and Sleisenger, M. H., *Lancet* **1,** 73 (1968).
101. Moses, S. W. Levin, S., Chayoth, R., and Steinitz, K., *Pediatrics* **38,** 111 (1966).
102. Rosenberg, L. E., *Adv. Hum. Genet.* **6,** 1 (1976).

103. Harris, H., and Hopkins, D. A., *Ann. Hum. Genet.* **36,** 9 (1972).
104. Harris, H., *Johns Hopkins Med. J.* **138,** 245 (1976).
105. Yoshida, A., *J. Mol. Biol.* **52,** 483 (1970).
106. Tashian, R. E., Riggs, S. K., and Yu, Y. S. L., *Arch. Biochem. Biophys.* **117,** 320 (1966).
107. Frantantoni, J. C., Hall, C. W., and Neufeld, E. F., *Science* **162,** 570 (1968).
108. Galjaard, H., Hoogeveen, A., de Wit-Verbeek, H. A., Reuser, A. J. J., Keijzer, W., Westerveld, A., and Bootsma, D., *Exp. Cell Res.* **87,** 444 (1974).
109. Gravel, R. A., Mahoney, M. J., Ruddle, F. H., and Rosenberg, L. E., *Proc. Natl. Acad. Sci. U.S.A.* **72,** 3181 (1975).
110. Folling, A., *Hoppe-Seyler's Z. Physiol. Chem.* **277,** 169 (1934).
111. Penrose, L., and Quastel, J. H., *Biochem. J.* **31,** 266 (1937).
112. Tourian, A. Y., and Sidbury, J. B., *in* "Metabolic Basis of Inherited Disease" (J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson, eds.), 4th ed., p. 240. McGraw-Hill, New York, 1978.
113. Rosenberg, L. E., and Scriver, C. R., *in* "Metabolic Control and Disease" (P. K. Bondy and L. E. Rosenberg, eds.), p. 583. Saunders, Philadelphia, Pennsylvania, 1980.
114. Knox, W. E., *in* "The Metabolic Basis of Inherited Disease" (J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson, eds.), 3rd ed., p. 266. McGraw-Hill, New York, 1972.
115. Jagenburg, R., and Rodjer, S., *Clin. Chem. (Winston-Salem, N.C.)* **23,** 1661 (1977).
116. Perry, T. L., Teschler, B., Hansen, S., and MacDougall, L., *Clin. Chim. Acta* **15,** 47 (1967).
117. Rosenblatt, D., and Scriver, C. R., *Nature (London)* **218,** 677 (1968).
118. Mitoma, C., *Arch. Biochem. Biophys.* **60,** 476 (1956).
119. Kaufman, S., *J. Biol. Chem.* **226,** 511 (1957).
120. Kaufman, S., *J. Biol. Chem.* **230,** 931 (1958).
121. Searle, A. F., *Heredity* **6,** 395 (1952).
122. Zannoni, V. G., Weber, W. W., Van Valen, P., Rubin, A., Bernstein, R., and La Du, B. N., *Genetics* **54,** 1391 (1966).
123. Woolf, L. I., Jakubovic, A., Woolf, F., and Bory, P., *Biochem. J.* **119,** 895 (1970).
124. Prensky, A. L., Fishman, M. A., and Daftari, B., *Brain Res.* **73,** 51 (1974).
125. Lipton, M. A., Gordon, R., Guroff, G., and Udenfriend, S., *Science* **156,** 248 (1967).
126. Andersen, A. E., Rowe, V. D., and Guroff, G., *Proc. Natl. Acad. Sci. U.S.A.* **71,** 21 (1974).
127. Greengard, O., Yoss, M. S., and Delvalle, J. A., *Science* **192,** 1007 (1976).
128. Greengard, O., Dienel, G., and Valle, J. D., *in* "Models for the Study of Inborn Errors of Metabolism" (F. A. Hommes, ed.), p. 125. Elsevier/North-Holland Biomedical Press, Amsterdam, 1979.
129. Lane, J. D., Schone, B., Langenbeck, U., and Neuhoff, V., *in* "Models for the Study of Inborn Errors of Metabolism" (F. A. Hommes, ed.), p. 145. Elsevier/North-Holland Biomedical Press, Amsterdam, 1979.
130. Delvalle, J. A., Dienel, G., and Greengard, O., *Biochem. J.* **170,** 449 (1978).
131. Dhondt, J. L., Farriaux, J. P., and Dautrevaux, M., *in* "Models for the Study of Inborn Errors of Metabolism" (F. A. Hommes, ed.), p. 141. Elsevier/North-Holland Biomedical Press, Amsterdam, 1979.
132. Perry, T. L., Hansen, S., Tischler, B., Bunting, R., and Diamond, S., *N. Engl. J. Med.* **282,** 761 (1970).
133. Loo, Y. H., Scotto, J. S., and Wisniewski, H. M., *Adv. Exp. Med. Biol.* **100,** 453 (1978).

134. Auerbach, V. H., DiGeorge, A. M., and Carpenter, G. G., *in* "Amino Acid Metabolism and Genetic Variation" (W. C. Nyhan, ed.), p. 11. McGraw-Hill, New York, 1967.
135. Menkes, J. H., and Holtzman, N. A., *Neuropaediatrie* **1,** 434 (1970).
136. Smith, I., *Arch. Dis. Child.* **49,** 245 (1974).
137. Smith, I., Clayton, B. E., and Wolff, O. H., *Lancet* **1,** 328 (1975).
138. Kaufman, S., Holtzman, N. A., Milstein, S., Butler, I. J., and Krumholz, A., *N. Engl. J. Med.* **293,** 785 (1975).
139. Milstein, S., Kaufman, S., and Summer, G. K., *Pediatrics* **65,** 806 (1980).
140. Kaufman, S., Berlow, S., Sumner, G. K., Milstein, S., Schulman, J. D., Orloff, S., Spielberg, S., and Pueschel, S., *N. Engl. J. Med.* **299,** 673 (1978).
141. Curtius, H. C., Niederwieser, A., Viscontini, M., Otten, A., Schaub, J., Scheibenreiter, S., and Schmidt, H., *Clin. Chim. Acta* **93,** 251 (1979).
142. Bartholome, K., and Byrd, D. J., *Lancet* **2,** 1042 (1975).
143. Kaufman, S., *Adv. Enzymol.* **35,** 245 (1971).
144. Dobson, J., Koch, R., Williamson, M., Spector, K., Frankenburg, W., O'Flynn, M., Warner, R., and Hudson, F., *N. Engl. J. Med.* **278,** 1142 (1968).
145. Horner, F. A., Streamer, C. W., Alejandrino, L. L., Reed, L. H., and Ibbot, F., *N. Engl. J. Med.* **266,** 79 (1962).
146. Vanderman, P. R., *Am. J. Dis. Child.* **106,** 492 (1963).
147. Solomons, G., Keleske, L., and Opitz, E., *J. Pediatr.* **69,** 596 (1966).
148. Cabalska, B., Duczynska, A., Golabek, B., Koslacz-Folga, A., and Zorska, K., *Proc. Int. Assoc. Sci. Study Ment. Defic., 3rd.* p. 548 (1973).
149. O'Grady, D. J., Berry, H. K., and Sutherland, B. S., *Am. J. Dis. Child.* **121,** 20 (1971).
150. Smith, I., Lobascher, M. E., Stevenson, J. E., Wolff, O. H., Schmidt, H., Grubel-Kaiser, S., and Bickel, H., *Br. Med. J.* **2,** 723 (1978).
151. Thier, S. O., and Segal, S., *in* "Metabolic Basis of Inherited Disease" (J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson, eds.) p. 1578. McGraw-Hill, New York, 1978.
152. Yeh, H. L., Frankl, W., Dunn, M. S., Parker, P., Hughes, B., and György, P., *Am. J. Med. Sci.* **214,** 507 (1947).
153. Stein, W. H., *Proc. Soc. Exp. Biol. Med.* **78,** 705 (1951).
154. Milne, M. D., Asatoor, A. M., Edwards, K. D., and Loughridge, L. W., *Gut* **2,** 323 (1961).
155. Levy, H. L., *Prog. Hum. Genet.* **4,** 1 (1973).
156. Lewis, H. B., *Yale J. Biol. Med.* **4,** 437 (1932).
157. Kinoshita, K., Takemoto, M., Itatani, H., Yachiku, S., Kotake, T., and Sonoda, T., *Urol. Int.* **34,** 369 (1979).
158. Crawhall, J. C., Scowen, E. F., and Watts, R. W. E., *Br. Med. J.* **1,** 585 (1963).
159. Jaffe, I. A., Altman, K., and Merryman, P., *J. Clin. Invest.* **43,** 1969 (1964).
160. Harris, E. D., and Sjoerdsma A., *Lancet* **2,** 996 (1966).
161. Rosenberg, L. E., Downing, S., Durant, L. L., and Segal, S., *J. Clin. Invest.* **45,** 365 (1966).
162. Bovee, K. C., Thier, S. O., Rea, C., and Segal, S., *Metab., Clin. Exp.* **23,** 51 (1974).
163. Segal, S., and Bovee, L., *in* "Models for the Study of Inborn Errors of Metabolism" (F. A. Hommes, ed.), p. 19. Elsevier/North-Holland Biomedical Press, Amsterdam, 1979.
164. Brown, R. R., *Science* **157,** 432 (1967).
165. Mason, H. H., and Turner, M. E., *Am. J. Dis. Child.* **50,** 359 (1935).
166. Donnel, G. N., Koch, R., and Bergren, W. R., *in* "Galactosemia" (D. Y. Y. Hsia, ed.), p. 247. Thomas, Springfield, Illinois, 1969.

167. Segal, S., *in* "Metabolic Basis of Inherited Disease" (J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson, eds.), p. 160. McGraw-Hill, New York, 1978.
168. Felig, P., *in* "Metabolic Control and Disease" (P. K. Bondy and L. E. Rosenberg, eds.), p. 276. Saunders, Philadelphia, Pennsylvania, 1980.
169. Levy, H. L., and Hammersen, G., *J. Pediatr.* **92,** 871 (1978).
170. Paigen, K., and Pacholec, E., unpublished results.
171. Beutler, E., and Baluda, M. C., *J. Lab. Clin. Med.* **68,** 137 (1966).
172. Shih, V. E., Levy, H. L., Karol-Kewicz, V., Houghton, S., Efron, M. L., Isselbacher, K., Beutler, E., and MacCready, R. A., *N. Engl. J. Med.* **284,** 753 (1971).
173. Tedesco, T. A., and Mellman, W. J., *J. Clin. Invest.* **48,** 2390 (1969).
174. Anderson, E. P., Kalckar, H. M., and Isselbacher, K. H., *Science* **125,** 113 (1957).
175. Pesce, M. A., and Bodourian, S. H., *Ann. Clin. Lab. Sci.* **10,** 26 (1980).
176. Donnell, G. M., Bergren, W. R., Bretthauer, R. K., and Hansen, R. G., *Pediatrics* **25,** 572 (1960).
177. Fensom, A. H., Benson, P. F., and Blunts, S., *Br. Med. J.* **4,** 386 (1974).
178. Kohn, G., Cohen, M. M., Segal, M., Bach, G., and Beyth, Y., *Isr. Med. J.* **13,** 394 (1977).
179. Ng, W. G., Donnell, G. N., Bergren, W. R., Alfi, O., and Golbus, M. S., *Clin. Chim. Acta* **74,** 227 (1977).
180. Schapira, F., Gregori, C., Boué, J., Henrion, R., Vigneron, C., and Vidailhet, M., *Biomedicine* **29,** 136 (1978).
181. Kromrower, F. M., and Lee, D. H., *Arch. Dis. Child.* **45,** 367 (1970).
182. Fishler, K., Koch, R., Donnel, G. N., and Wenz, E., *Clin. Pediatr.* **19,** 38 (1980).
183. Hsia, D. Y. Y., and Walker, F. A., *J. Pediatr.* **59,** 872 (1961).
184. Schwarz, V., *Arch. Dis. Child.* **35,** 428 (1960).
185. Baker, L., Mellman, W. J., Tedesco, T. A., and Segal, S., *J. Pediatr.* **68,** 551 (1966).
186. Beutler, E., Baluda, M. C., Sturgeon, P., and Day, R. W., *Lancet* **1,** 353 (1965).
187. Mathai, C. K., and Beutler, E., *Science* **154,** 1179 (1966).
188. Shapira, F., and Kaplan, J. C., *Biochem. Biophys. Res. Commun.* **35,** 451 (1969).
189. Chacko, C. M., Christian, J. C., and Nadler, H. J., *J. Pediatr.* **78,** 454 (1971).
190. Ng, W. G., Bergren, W. R., and Donnell, G. N., *Ann. Hum. Genet.* **37,** 1 (1973).
191. Segal, S., and Bernstein, H., *J. Pediatr.* **62,** 363 (1963).
192. Spatz, M., and Segal, S., *J. Pediatr.* **67,** 438 (1965).
193. Dische, Z., Selmenis, G., and Youlous, J. *Am. J. Ophthalmol.* **44,** 332 (1967).
194. Kinoshita, J. H., Merola, L. O., and Hayman, S., *J. Biol. Chem.* **240,** 310 (1965).
195. Korc, I., *Arch. Biochem. Biophys.* **94,** 196 (1961).
196. Wells, W., Piltman, T., Wells, H., and Egan, T., *J. Biol. Chem.* **240,** 1002 (1965).
197. Woolley, D. W., "Biochemical Basis of Psychosis or the Serotonin Hypothesis About Mental Disease." Wiley, New York, 1962.
198. Woolley, D. W., and Gommi, B. W., *Proc. Natl. Acad. Sci. U.S.A.* **52,** 14 (1964).
199. Tedesco, T. A., and Mellman, W. J., *Science* **172,** 727 (1971).
200. Gitzelmann, R., *Pediatr. Res.* **1,** 14 (1967).
201. Segal, S., Rutman, J. Y., and Frimpter, G. W., *J. Pediatr.* **95,** 750 (1979).
202. Mayes, J. S., and Guthrie, R., *Biochem. Genet.* **2,** 219 (1968).
203. Beutler, A., Krill, A., Comings, D., and Trinidad, F., *J. Lab. Clin. Med.* **76,** 1006 (1970).
204. Gitzelmann, R., Steinmann, B., Mitchell, B., and Haigis, E., *Helv. Paediatr. Acta* **31,** 441 (1976).
205. Tay, W., *Trans. Ophthalmol. Soc. U. K.* **1,** 55 (1881).
206. Sachs, B., *J. Nerv. Ment. Dis.* **14,** 541 (1887).
207. Klenk, E., *Ber. Dtsch. Chem. Ges.* **75,** 1632 (1942).

208. Svennerholm, L., *Biochem. Biophys. Res. Commun.* **9,** 436 (1962).
209. Ledeen, R., and Salsman, K., *Biochemistry* **4,** 2225 (1965).
210. Robinson, D., and Stirling, J. L., *Biochem. J.* **107,** 321 (1968).
211. Sandhoff, K., Andreae, U., and Jatzkewitz, H., *Life Sci.* **7,** 283 (1968).
212. Okada, S., and O'Brien, J. S., *Science* **165,** 698 (1969).
213. O'Brien, J. S., Okada, S., Chen, A., and Fillerup, D. L., *N. Engl. J. Med.* **283,** 15 (1970).
214. O'Brien, J. S., *in* "Metabolic Basis of Inherited Disease" (J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson, eds.), p. 841. McGraw-Hill, New York, 1978.
215. Okada, S., Veat, M. L., Leroy, J., and O'Brien, J. S., *Am. J. Hum. Genet.* **23,** 55 (1971).
216. Schneck, L., Friedland, J., Valenti, C., Adachi, M., Amsterdam, D., and Volk, B. W., *Lancet* **1,** 582 (1970).
217. O'Brien, J. S., Okada, S., Fillerup, D. L., Veath, M. L., Adornato, B., Brenner, P. H., and Leroy, J. G., *Science* **172,** 61 (1971).
218. Navon, R., and Padek, B., *Br. Med. J.* **4,** 17 (1971).
219. Navon, R., and Padek, B., *J. Pediatr.* **80,** 1026 (1972).
220. Carmody, P. J., Rattazi, M. C., and Davidson, R. G., *N. Engl. J. Med.* **289,** 1072 (1973).
221. Saifer, A., and Perle, G., *Clin. Chem.* (*Winston-Salem, N.C.*) **20,** 538 (1974).
222. Friedland, J., Schneck, L., Saifer, A., Pourfar, M., and Volk, B. W., *Clinical. Chim. Acta* **43,** 417 (1973).
223. Nakagawa, S., Kumin, S., Chandra, P., and Nitowsky, H. M., *Clin. Chim. Acta* **88,** 249 (1978).
224. Kaback, M. M., Rimoin, D. L., and O'Brien, J. S., eds., "Tay Sachs Disease; Screening and Prevention," Prog. Clin. Biol. Res., Vol. 18. Alan R. Liss, Inc., New York, 1977.
225. Brett, E. M., Ellis, R. B., Haas, L., Ikonne, J. U., Lake, B. D., Patrick, A. D., and Stephens, R., *Arch. Dis. Child.* **48,** 775 (1973).
226. Sandhoff, K., Harzer, K., Wassle, W., and Jatzkewitz, H., *J. Neurochem.* **18,** 2469 (1971).
227. Rapin, I., Suzuki, K., Suzuki, K., and Valsamis, M. P., *Arch. Neurol.* (*Chicago*) **33,** 120 (1976).
228. Navon, R., Padeh, B., and Adam, A., *Am. J. Hum. Genet.* **25,** 287 (1973).
229. Vidgoff, J., Buist, N., and O'Brien, J. S., *Am. J. Hum. Genet.* **25,** 372 (1973).
230. Tallman, J. F., Brady, R. O., Navon, R., and Padeh, B., *Nature* (*London*) **252,** 254 (1974).
231. Dreyfus, J. C., Poenaru, L., and Svennerholm, L., *N. Engl. J. Med.* **292,** 61 (1975).
232. Srivastava, S. K., Yoshida, A., Awasthi, Y., and Beutler, E., *J. Biol. Chem.* **249,** 2049 (1974).
233. Bach, G., and Suzuki, K., *J. Biol. Chem.* **250,** 1328 (1975).
234. Li, Y. T., Mazzolta, M. Y., Wan, C. C., Orth, R., and Li, S. C., *J. Biol. Chem.* **248,** 7512 (1973).
235. Srivastava, S. K., and Beutler, E., *J. Biol. Chem.* **239,** 2054 (1974).
236. Desnick, R. J., and Goldberg, J. D., *Prog. Clin. Biol. Res.* **18,** 129 (1977).
237. Rattazi, M. C., Baker, H. A., Cork, L. C., Cox, N. R., Lanse, S. B., McCullough, R. A., and Munnell, J. F., *in* "Models for the Study of Inborn Errors of Metabolism" (F. A. Hommes, ed.), p. 57. Elsevier/North-Holland Biomedical Press, Amsterdam, 1979.
238. Brooks, S. E., Hoffman, L. M., Adachi, M., Amsterdam, D., and Schneck, L., *Acta Neuropathol.* **50,** 9 (1980).
239. Klein, R. Z., *Adv. Pediatr.* **26,** 417 (1979).

240. Fisher, D. A., Foley, T. P., LaFranchi, S., Mitchell, M. L., and Walfish, P. G., *J. Pediatr.* **94,** 700 (1979).
241. Klein, A., Meltzer, S., and Kenney, M. D., *J. Pediatr.* **81,** 912 (1972).
242. Raiti, S., and Newns, G. H., *Arch. Dis. Child.,* **46,** 692 (1971).
243. Dussault, J. H., Coulombe, P., Laberge, C., Letarte, J., Guyada, H., and Khoury, K., *J. Pediatr.* **86,** 670 (1975).
244. Committee of the American Thyroid Association, Recommendation for Screening Programs for Congenital Hypothyroidism, *J. Pediatr.* **89,** 692 (1976).
245. Robbins, J., and Gorden, P., *in* "Metabolic Control and Disease" (P. K. Bondy and L. E. Rosenberg, eds.), p. 1325. Saunders, Philadelphia, Pennsylvania, 1980.
246. Stanbury, J. B., *in* "Metabolic Basis of Inherited Disease" (J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson, eds.), 4th ed., p. 206. McGraw-Hill, New York, 1978.
247. Fisher, D. A., *Pediatr. Clin. North Am.* **25,** 423 (1978).
248. Fisher, D. A., *in* "Diabetes and Other Endocrine Disorders During Pregnancy and in the Newborn" (M. I. New and R. H. Fiser, Jr., eds.), p. 221. Alan R. Liss, Inc., New York, 1976.
249. Fisher, D. A., Dussault, J. H., and Sack, J., and Chopra, I. J., *Recent Prog. Horm. Res.* **33,** 59 (1977).
250. Stanbury, J. B., and Hedge, A. N., *J. Clin. Endocrinol. Metab.* **10,** 1471 (1950).
251. Salvatore, F., Stanbury, J. B., and Rall, J. E., *in* "Thyroid Gland" (M. Devisscher, ed.), p. 443. Raven, New York, 1980.
252. Committee on Drugs of the American Academy of Pediatrics, *Pediatrics* **62,** 413 (1978).

6

Biochemistry of Aging

JOHAN BJORKSTEN

I would not give a fig for
simplicity before complexity,
but I would give my life for
simplicity after complexity.

Justice Oliver Wendell Holmes

CLINICAL BIOCHEMISTRY
Contemporary Theories and Techniques, Vol. 2

ISBN 0-12-657102-3

I. THE EVOLUTION OF AGING

One of the most widely accepted ideas about aging is that the best way to assure oneself of a long life is to have long-lived parents. The heredity factor shows a strong correlation with actual longevity (*1*).

Which factors govern this?

In any organism such as man, in whom the causes of death are many, it is obvious that there is no single hereditary factor that determines the length of life (*201*). The efficacy of the immunological system is an important factor (*122*), the concentration and quality of repair enzymes another (*4*), and the uniformity and dimensioning of the arterial system a third. The endocrine and immune responses are hereditary and represented by a whole family of genes. If to this the familial factors are added that are nonhereditary, such as food and exercise habits (*1*), it becomes obvious that for man there cannot be any such thing as a directly pre-programmed death. Otherwise, the cause of death would be much less variable.

In the salmon, for example, in which death invariably follows the strenuous swim up a river to the spawning ground, there is one dominant cause of death: atrophy of the digestive tract. Unless evolutionary forces were to prevent it, the large fish, having completed a strenuous voyage, would be very hungry after spawning, and would be prone to eat their own eggs, which would eventually result in the disappearance of the species. Many other such examples exist (*259*).

No similar evolutionary need exists in humans who, like other mammals, have a strong parental instinct and take care of their offspring. Giant reptiles, so formidable in size, strength, and natural armament, may have vanished from the earth because they left their eggs unattended and their offspring unprotected as easy prey for small mammals.

A possible influence of parental attention on evolution is the endowment of a specific life span for each mammaliam species enabling the parents to rear and protect their offspring until the offspring are fully capable of self-protection. For mice this period is a few months; for man it can extend to 60 years.

Although there is no protection from lethal factors that extend beyond what is desirable from an evolutionary viewpoint, the freedom exists to improve, supplement, and reorganize a natural endowment, so as to maximize an individual's life.

II. PRIMARY AND SECONDARY CAUSES OF AGING

By systematically applying present knowledge, it should be possible to achieve an extension in life span of about 5–20 years. For example, the

rise in the female life expectancy in recent years might be related to the acceptance of the hormone supplementation in "the pill." The artificial supplementation in the contraceptive pill appears to have brought about reductions in several factors in the aging rate, which in their longevity consequences have outweighed the negative influence of an increased mortality due to coronary thrombosis.

This may indicate what a supplementation of a dwindling supply of one specific hormone can accomplish in relation to the aging progress.

III. THE CHAIN OF LIFE AND THE FIRST LAW OF LECOMPTE

An analogy that has proven useful in elucidating some aspects of life is that of a series of chains in which each link represents an enzyme. Life ends if any one of the links in the vital chain breaks.

This analogy makes it easy for us to accept LeCompte's first law, which states that "Aging is faster, the greater the number and the extent of deficiencies" (*181*). Considering the chain concept (*149,175,294*), it becomes plain enough that the more the links are poorly made or badly damaged, the greater is the chance that one of them is going to fail completely under even the slightest stress.

The chain analogy shows us also the futility of strengthening only one of the links in a chain, when several are critically weak. Yet this is being done as a rule rather than as an exception. Again and again, publicity is given to the importance of some single vitamin or alleged cure, so as to cause the public to buy it and feel secure, when strengthening over a much broader front was needed instead. A sailor worrying about an anchor chain would hardly consider spending all of his repair money on making a single link enormously strong while neglecting the others. Yet, when it comes to the vital chains of life, such is now common practice!

In view of the foregoing, we may conclude the following: It is inadvisable to depend too much on any one factor in deferring the symptoms of aging. Everything possible should be done to ensure at all times a constant, dependable, and entirely sufficient supply of all of the key factors necessary for life.

To do this successfully challenges the breadth of vision of the clinical gerontologist and the skills of the laboratory staff on which he depends.

IV. THE VULNERABILITY OF MOLECULES IN TRANSIT

Regardless of where a molecule comes from, it cannot perform its function until it has reached and been accepted at its target organ. More often

than not, the molecule spends more time in transit than in performing its function. During this transit time, it is exposed to unwanted random reactions such as the reaction with a receptor group other than the normal target, the capture by absorption to a surface (*141,142*), or being trapped in a network of cross-linked elongated molecules (*29,* see p. 49).

As an example, food is first chewed and mixed with saliva, in which it encounters a wide range of substances including amylolytic enzymes, still poorly defined proteins, and trace minerals. It then passes on to the stomach where it encounters hydrochloric acid, hydrolytic enzymes that are active under acid conditions, the sulfonic acid types of surface-active agents produced by the liver, and a host of additional enzymes, as well as proteins and other substances. These are also contained in the chyme leaving the stomach. In the small intestine, the chyme is processed with pancreatic enzymes and bile constituents, and is either absorbed or rejected. It may be partly decomposed to become absorbable, and then recombined. Bacteria and their metabolites are also present. Mistakes in the recombination are possible, and the ingested substances have still only moved a part of the way to their destination. After passing the intestinal wall, they enter the lymph, from which they finally reach the blood.

The ingested substances now have to face the scrutiny of the security forces of the immune system, particularly granulocytes and B and T lymphocytes, as well as a large variety of protective substances. More important, however, may be the "casual," randomly present chemicals, of which an almost infinite variety has been formed in metabolism. Some of these are now circulating in the blood enroute to being excreted, others are static, and still others are on their way to definite receptors. All of these are at least somewhat reactive. Any chance collision can give rise to unwanted compounds, or to rare events not intended by evolution, and for which no defense has been prepared (*20,25*). Such relatively rare events may play a decisive role in senescence (*27,29,36,273*).

The human organism would not have survived unless it had found means for removing or neutralizing those compounds that would prove fatal before progeny could be produced and reared. We have no natural defense, however, against those chance reactions that cause formation of insoluble, irremovable substances at a rate too slow to matter until after the age of 50–70.

Some types of compounds that are critical for aging are therefore substantially different for man and for short-lived mammals. For example, the rat needs no resistance against influences that would require 5 years to accumulate to the point critical for life, whereas for the human, anything that would automatically kill before 30 years of age has been ruled out already or neutralized in the course of human development. Some such

cause may explain why, for example, aspirin is a deadly poison to rats, whereas it is beneficial to humans.

But let us return to the immediate problem: that of transit. The reactive molecule that we are following now enters the blood stream in its quest for a receptor in another part of the body. If our molecule is water-soluble, it can dissolve directly in the blood, but if it is not, it must combine with a lipoprotein. All of these contain a phosphatide moiety, to which fat-soluble molecules, such as vitamins A, E, and D, many steroid hormones, and prostaglandins can attach themselves.

In recent years, much distinction has been drawn between the high-density and the low-density lipoproteins. The work of Blondin and Green (*46*) and of Van Gent (*284*) indicates strongly that the high-density lipoproteins owe their higher densities to a higher content of phospholipids, such as lecithin.

In clinical gerontology the phospholipid status deserves increased attention. It is true that the organism can synthetize its own lecithin if the components are there, but lecithin remains our best source of choline, commonly occurring in phospholipids.

A high phospholipid status shields many sensitive molecules from damage in transit. Already Mattill has shown that an unprotected oil, which on an accelerated test would become rancid in 1 hr, could be protected for 8 hr with α-tocopherol alone, for 50 hr if lecithin were also added, and for 200 hr if ascorbic acid was superimposed; this could be further extended if tannins were added to the above (*141*, p. 117).

A much better understanding of the role of phospholipids in membranes and in energy metabolism has been gained in relatively recent years (*46,125,126*). Lecithin, uniquely, has the capacity to transport both Mg^{2+} and inorganic phosphorus or Mg^{2+} and adenosine diphosphate (ADP). Further, it has the capacity to transport a pair of molecules in addition to the metal ion, Mg^{2+} + ATP + AMP, Mn^{2+} + citrate + nicotinamide adenine dinucleotide (NAD), Mg^{2+} + ATP + glucose, and Zn^{2+} + β-hydroxy-γ-butyrate + NAD. The rates of such a triple transport exceed by an order of magnitude the rates for metal ion with a single anionic species (*46*).

Thus, the lecithin system represents an altogether different type of ion carrier in which the metal ion is encapsulated with two molecules, with which it can form a coordination complex. The active ionophore is not limited to a single lecithin molecule, but can be a set (*126*).

Given the ideal combination of metal ion and an anionic molecule with which it can coordinate, lecithin can transport and protect in transit nucleotides, sugars, amino acids, citric cycle intermediates, and mono or divalent metal ions. Thus, it is uniquely adapted to transport and to protect in transit the gamut of reactive polar compounds that are known to move

within the body and to pass through biological membranes. It is likely to play an important role not only in transport, but also as the prosthetic ionophore in oxidative phosphorylation, in coupled ATP hydrolysis, and in ATP-driven systems. Each electron transfer complex, vital for the function of cell energetics, contains a set of ionophoroproteins imbedded in a matrix of phospholipid. This phosphatide, a lecithin in the broad sense, has the multiple functions of transport and protection and of serving as a vehicle compatible with the desired adjustments and controlled mobility required for optimizing cellular energetics.

In cases where no deficiency of vitamins appears to be present in the diet, but clinical symptoms still indicate that a lack of some essentials exists, it may be appropriate to give careful consideration to the transit aspects. Most active biomolecules spend a longer time in transit than in performing their function at the point of their destination. Lecithin accompanies cholesterol in most of the nutritional sources. A rigid low-cholesterol diet might well be reviewed to make certain that any phospholipid deficiency is compensated for.

V. APPROACHES TO MEASUREMENT OF PHYSIOLOGICAL AGE

A reliable method to judge the physiological age of a person would be highly desirable because (a) it would improve our ability to judge whether a proposed long-term treatment is likely to affect the overall longevity and (b) it would enable us to foretell more accurately whether or not the dosage of a given drug would be tolerated by an aged patient, under given conditions of diet, caloric intake, vitamins, trace metals, hormonal balance, heredity, clinical history, and accidental or environmental exposures.

Many methods have been proposed. A composite of a large number of parameters was developed by Comfort (*64*). This is a useful research tool but is rather complex for clinical use.

Assuming that cross-linkage is a principal common denominator, a useful practical index could be based on properties directly tied to cross-linkage: the loss of elasticity, the increase in viscosity of cytoplasm, and progressive change of light absorption by translucent tissue.

Some test of the speed of neurologic processes might prove very valuable, if the variability can be sufficiently controlled (*65,297*).

The following approaches do not seem sensitive enough:

1. Any test involving soluble substances, including blood and serum tests: The important parameters in blood are continually monitored by physiological sensors, and maintained within narrow margins by feedback

neuroendocrine mechanisms (*222*). The age-dependent changes related to insoluble aggregates are therefore slow to appear in the circulating body liquids until the regulating mechanisms begin to break down late in life (*123*).

2. Any test that is greatly influenced by short-term changes in diet and in fluctuating environmental changes.

3. Tests involving tissue or cell culture to determine capability of division may yield very interesting results, but appear too touchy for standard use in ordinary clinical practice.

The most promising approaches are those which reflect molecular changes toward agglomeration, preliminary to precipitations, shrinkage, or complete insolubilization. These include changes in the eye (*89*) and changes of elasticity.

1. Changes in the eye: These have the great advantages of permitting direct observation and measurement by standard, widely known, and developed methods. They include changes in accommodation, in the field of view, in the deformability of the lens, and in changes of color and opacity of any transparent substance.

The changes in vision in man are well known, and appear to be applicable for physiological age determinations in rats and probably in all other mammals (*89*).

Birds, however, do not show the same ocular changes in aging (*171,194*). From an evolutionary standpoint this is understandable. The ability of a fast-flying bird to brake almost instantly from a high speed to a dead stop in landing on a telephone wire, or in catching flying insects in fractions of a second, must indeed place an enormous premium on the undeviating reliability of ocular accommodation.

The further study of related biochemistry of the visual systems of birds might give new clues to possible defenses against aging.

2. Changes of elasticity: The resilience and extensibility of proteins, and of tissues generally, show early changes with much greater sensitivity than any of the precipitation or fluorescence tests. Furthermore, they measure directly a property of great importance to the organism. Loss of elasticity of the endothelium is an early prelude to scleroses, and may be the immediate cause of microfractures that could develop into arterial disease and perhaps also cancer.

Changes in elasticity can be determined by the skin-fold test, by the oscillographic measurement of the pulse wave, or by a setup that by contact or radiation quickly raises the surface temperature of any skin surface, and detects how fast the original temperature is restored with a high-speed recording, surface-temperature measuring optical device. This can be a function of the elasticity of the capillary bed.

3. Radioisotope assay: For research purposes, very rapid determinations can be made by the use of animals that have received huge doses of radioisotopes paranatally and have survived at least two-thirds of their normal life span, so that the retained radioactivity (a small fraction of the original dose) is mostly firmly bound. The sudden release of such retained activity can sharply and rapidly pinpoint treatment of a likely candidate, but it takes 2 years to prepare rats for this assay, and the cost and effort is quite out of line for any normal clinical use (*35,38,40*).

Age-Dependent Sensitivity to Drugs

A principal cause of sensitivity to drugs in aging is that the amount of functional cytoplasm in all cells is reduced by progressive cross-linkage or otherwise.

At birth, the average human organism has enough reserve space to tolerate the cumulative molecular damage predictable to 60 years of age. As this age of each individual is approached, the reserve margins are very rapidly reduced.

In the absence of detailed knowledge of the patient's weak points, it will be difficult even for the experienced clinician to be certain of anything. An increased supply of key vitamins can increase numerous safety margins—the literature reports many instances of increase in resistance to trauma by enhancement of niacin, ascorbic acid, pyridoxine, tocopherol, etc. as well as by hormone therapies.

VI. THE EXOGENOUS FACTORS

Returning now to the consideration of the chain-of-life concept, we will scrutinize its various "links." These can be divided in two groups: the exogenous factors, including vitamins and trace elements; and the endogenous factors, which depend on substances synthesized within the body, ranging from generally available substances such as hormones to essential structural elements such as collagen and elastin. The boundaries are somewhat overlapping, and some of the endogenous factors may depend on the availability of some exogenous ingredients, and vice versa.

On the whole, however, this subdivision will simplify description.

A. Vitamins

Vitamin is a generic term for those organic substances that are necessary in small amounts for life and cannot be synthesized by the organism.

Most of these are either parts or activators of enzymes that form links in one of the chains of life.

Taking LeCompte's Law as a guideline, which states that the rate of aging is roughly proportional to the number and severity of the deficiencies present, we have to be concerned with those deficiencies that are caused by a suboptimal vitamin supply. The determination and evaluation of such deficiencies, therefore, becomes crucial to the clinical aspects of needlessly rapid aging.

Such determinations are complicated by the fact that almost all of the vitamins are essential for more than one metabolic link. For example, nicotinic acid is a part of, or activates, at least 54 enzymes in known reactions. Most likely, it is also needed for at least an equal number of reactions that have not yet been identified. Pyridoxine has almost as many known functions, among which are many of the transaminations, essential in brain metabolism. Ascorbic acid (vitamin C) is essential for syntheses of collagen (*14*), dentine, adrenaline, corticosteroids, antioxidation in the water phase, generation of free hydroxyl radicals, blood coagulation, and the metabolism of several amino acids and the principal neurotransmitters (*47*).

The reason for such variety of functions is easy to grasp. For example, vitamin E may have been developed and adopted as an antioxidant by a unicellular organism such as *Scenedesmus obliquus* or *Ochromonas malhamensis* somewhere in the shallow warm seas where life may have originated (*30*). Oxygen was scarce at the time, as indicated by the presence of iron in the reduced bivalent form in all pre-Cambrian deposits.

The antioxidant enabled its possessor to survive in lower oxygen concentrations than its competitors. Furthermore, it could avoid forming deleterious intermidiates, by completing their oxidation. With these evolutionary advantages, the possessor of these antioxidant vitamins prevails.

Next time a new chemical need arose in the course of evolution, the organism would first try those chemicals it already had available. Once every couple of million years or so, one of these would fit another, quite unrelated need. Repeating this trial-and-error procedure for hundreds of millions of years, each vitamin gradually multiplied its functional range, so that the more than 50 different physiological functions of pyridoxine and nicotinic acid also appear reasonable and maybe even to be expected.

The products present in nature have had the benefit of first trial in every other application which the organism has had to solve in the countless millions of years of evolution. Each of these products can therefore be presumed to have many uses. Any proposed replacement or curtailment of them must be made with caution, lest some quite different link be weakened.

Therefore, it is unwise to base estimates of how much of a given compound the organism needs by relying mainly on one symptom, however spectacular. For example, the officially approved recommended dosages for normal adults for vitamin C are much too low and obsolete in the light of recent findings. These dosages are apparently based mainly on the dosages needed for preventing and/or curing scurvy. However, they do not take sufficiently into account the additional functions of ascorbic acid mentioned above. As Ritzel showed in 1961, in a careful, double-blind study (*286*), even as little as 1 g/day of ascorbic acid cuts in half the number of days lost to respiratory infections, and Wilson *et al.* have shown in some detail how the ascorbic acid is taken up by lymphocytes and used by them rapidly in case of infection (*189–192,298–301*).

The mechanism for this also seems clear. It has been shown in detail and by many researchers that ascorbic acid and five other less well-known representatives of the enediol chemical class, with one proximate carbonyl group (*83,84*), have the capacity to form free hydroxyl radicals in the presence of catalytic amounts of iron and peroxides, both normally present in the organism, and that the free hydroxyl radicals can break down almost any natural and several synthetic polymeric compounds into removable fragments (*33,83,84,260*). The known facts are consistent with the assumption that the ascorbic acid thus prepares the ammunition that the leucocytes have available and probably use to attack any adversary they can define, and break down its vital molecules to useless fragments. *In vitro* this weapon system has broken down such diverse large molecules as cellulose, starch (*229*), enzymes (*220*), polypeptides (*237*) (attacking any peptide linkage), proteins (*139,140,214,220,229,240*), and gerogenic insoluble aggregates (*33*).

In addition to this, Nobel laureate Albert Szent-Gyorgyi theorizes that ascorbic acid mobilizes proteins by attracting electrons from them, thereby increasing their mobility, and that this is a function vital to life processes (*271*).

On withdrawal of ascorbic acid from the diet, the immediate life-threatening symptoms of scurvy will not appear until a withdrawal has been made of most of the ascorbic acid reserved by the organisms for the other functions mentioned above. Many more examples could be given of this. Such withdrawals cirtically weaken the functions stated, as well as many others, long before the symptoms appear that are the basis for the "official recommended dosages." The neurotransmitters, including noradrenaline, adrenaline, and serotonin, require ascorbate for their synthesis and for protection from oxidation in transit and *in situ*. In the formation of collagen, ascorbate is essential for hydroxylation of proline and lysine (*14-47,188*). Space does not permit discussion here of other food essentials, which conform to the examples given.

B. Trace Elements

Most of the research with trace elements has been directed to those applications in which the elements are required, of which small additions will dramatically increase agricultural yields or prevent cattle disease.

In relation to aging, however, we are still more concerned with those instances in which small amounts of metals have a toxic effect and may be instrumental in accelerating aging.

In 1950, Albert Tyler was interested in factors affecting the life span of sperm cells, initially working with sea urchins (Holothuria). He tried various nutrient solutions added to the sea water into which the sea urchin sperm are normally released, and found indeed an appreciable life-extending effect. However, he was struck by the fact that this effect was most marked with those amino acids that were also chelating agents, with ethylenediamine tetraacetic acid and diethyldithiocarbamic acid by far the best (more than 50-fold). This led him to extend his work to other species including mammals. Chelation turned out to be the key in determining the functional life of sperm cells. The nutrient aspect of the additives was almost negligible in comparison. Life extension could also be gained by using a synthetic sea water, free from polyvalent trace metals. Tyler's work is thorough and convincing (*279,280*). It was unfortunately interrupted by his untimely death and has not yet been resumed.

The efficacy of chelating agents in prolonging the life span of short-lived organisms has been confirmed by others, particularly Lansing (*176–178*) and Sincock (*258*).

Like the sperm cells, all long-lived organisms accumulate, to the extent of many times their original content, certain metals in the course of a life span. Those metals that are polyvalent, and thus capable of reacting with any of the important biomolecules by cross-linking or chelating, can in very small amounts seriously interfere with longevity. In addition to the well-known poisons lead, mercury, and cadmium, this is also true of copper, a well-known oxidation catalyst, even though only small quantities of copper are normally needed. Copper is also an inhibitor of protein synthesis (*148*). There is a clear positive correlation between copper content in groundwater and death from circulatory disease.

However, the possibility of a much stronger relation between a polyvalent metal and age-dependent failure of the nervous system is indicated by the work of Crapper and co-workers, who demonstrated the possibility of inducing symptoms closely resembling those of Alzheimer–Pick disease in cats, by a single injection of a 1-mg quantity of aluminum chloride into the hippocampic ventricle of the brain (*70,75,85,86*). It would be futile to modify aluminum intake, since aluminum is 8.4% of the earth's crust, and man ingests an average of 30 mg aluminum daily. Data on the accumulation of the aluminum in man with age have been tabulated (*305*). This ref-

erence also lists chelating agents known to be effective on aluminum (*305*, p. 482). The freedom of licensed physicians to use chelation when proper in their judgment was importantly confirmed in the FDS versus Evers court decision of June, 1978 in Alabama (*97,214*).

The human body has been forced by evolution to develop an excellent screening system for aluminum: Almost all the ingested aluminum compounds pass unabsorbed through the digestive tract. However, in 60 years, even the best system permits a few milligrams of aluminum to escape and to reappear in potentially harmful places. Aluminum oxide is virtually insoluble, so that contact with metallic aluminum surfaces should normally be quite safe. However, this does not apply to ingestion of aluminum hydroxide, carbonate, or to other compounds that react with stomach acid to form soluble aluminum chloride, or which are otherwise soluble in water. The use of soluble aluminum compounds (alum) in purification of water supplies seems injudicious, not only from this standpoint, but also because the aluminum ion disastrously affects the flocculation of biocolloids, of which the human brain is a prime example (*238;* also see pp. 221–237).

C. Antioxidants

Oxidation is the mainstay of all higher life on earth. We depend on it for the most economical energy supply. An elaborate, highly developed sequence of reactions makes possible the gradual release of energy and the stepwise formation of adenosine triphosphate, the common currency of bioenergetics (*46,206*). This entails the metabolism organism of easily oxidized, intermediate products as well as nutrients. Among the more sensitive of these are the unsaturated fatty acids, the phospholipids containing these, the retinoids, and the isoprenoid compounds (*25,30*).

If no protective mechanisms existed, these labile compounds would sustain heavy losses by random, uncontrollable unscheduled oxidations. These reactions would not only damage much needed important molecules, but also waste oxygen, and further give rise to active cross-linkers (*25,26,41,141,197,198,232,268,273;* also see *129*, pp. 295–307).

Therefore, the antioxidants are vitally important. Hickman (*141*) stated, "As I view the longevity problem, it contains a vital question: How much can we have of the preservative factors without depressing active metabolism, making the organism as a whole lethargic?" McCay *et al.* (*195*) showed that addition of wheat germ oil containing tocopherol prolonged the life of female rats and the virility of male rats. H. Kaunitz is quoted as saying at the New York Academy of Sciences' Second International Meeting on Vitamin E (1949) that it does not require much intelligence to realize that antioxidants will have a favorable effect on longevity (*162*).

The effective antioxidant system is largely built on tocopherols, possibly because of the fine tuning made possible by a multiplicity of very similar structures available and their relative nontoxicity (*142*). Machlin *et al.* showed that at least for chicken encephalomalacia any nontoxic antioxidant could substitute for vitamin E (*196–198*).

However, the efficacy is enhanced by the synergists lecithin, ascorbic acid, tannin, and, perhaps most importantly, selenium. Selenium is more than a synergist. It is antioxidant in its own right, widely and successfully used in animal feeds, and the most economical antioxidant. However, it must be handled with good control of dosage, inasmuch as it, like vitamin A, is toxic in high dosages [*58,82,109,110,111* (164 references listed) *160,169* (34 references listed), *170,256,269,275,297*].

In addition to its antioxidant properties, selenium is also a detoxifier for cross-linking polyvalent metals such as cadmium, mercury, and lead, with which it forms very insoluble compounds. Space restrictions preclude detailed discussion here; the references given provide further assistance to those desiring more detailed information.

VII. THE ENDOGENOUS FACTORS

The hypothesis of a "death hormone," genetically controlled, which at a given point should cut the chains of life, is not founded on facts. If such a hormone existed, all humans would die of some definite, uniform cause, just as the salmons that are programmed to die after spawning. No similar reason for programmed death exists for humans. After human progeny are brought up and helped to a stage of independent existence, evolution no longer provides for replacement of organs or chemicals that become exhausted in the parents. Among the organs thus left to deteriorate are the glands which secrete hormones, as well as hormone-controlling feedback sensors (*88*) and other ancillary organs or chemicals. These are complex and sensitive, and are thus particularly vulnerable to damage by neglect (*87,98,103,127,222,227*). It then becomes a task for the clinician and his supporting laboratory staff to determine which hormones are becoming altered or deficient in a patient, and to furnish a supplement or replacement of these particularly vulnerable links, wherever possible.

A. Genetic

One of the most thoroughly established facts of aging is that long-lived persons mostly have long-lived parents (*1*). Even if the household habits and style of life are factors, the correlation between genetic inheritance and longevity stands out strongly. The reasons for this are clear: every

single trait, every innate physical and mental characteristic is genetically predisposed. Since every one of these singly has some influence on the course and duration of life, their total sum also must have such an influence. Could the natural predisposition to a longer life be cut short by a biochemical accident that affects the DNA itself in a critical cell? Yes, this could actually be happening. Life spans are several times shorter than they could be (*23,24,257*). It has been shown with illustrated examples (*29;* also see *24*, p. 45) exactly how a small bifunctional molecule can immobilize two macromolecules by cross-linkage, so as to jeopardize their function. The number of such small, bifunctional molecules is so large (*25*) that the probability of the cross-linking reaction overshadows the probability of any other reaction apt to cause the progressive immobilization and functional neutralization of large, vital molecules. In the course of 60 years, every reaction that is thermodynamically plausible will actually occur some time. Anyone who seeks for evidence of any such reactions will therefore actually find it, provided that his analytical technique is sufficiently sensitive. The *probability* of these various reactions will determine their relative significance. On this basis, cross-linkage emerges as the leading cause (*21,26,34,129,273*).

B. Somatic

There is another reason as well why the cross-linkage reaction appears significant (*165*). The age dependence of resistance to all trauma and the mortality curve follow a logarithmic probability curve (*39*). To be plausible as a leading cause of aging, any mechanism postulated should explain the self-accelerating progress of the aging process. Cross-linking meets this test. For example, an accidental cross-linkage can easily happen between a DNA molecule and an RNA molecule lined up in contact with it for a message. These two molecules thus become attached to each other, but at that point only one strand on the DNA is involved. The damage is still repairable, but some time will elapse before the message for help has been sent and received and before the repair enzymes are transported to the site. In the meantime, the normally short stay of the messenger RNA is being prolonged. The two molecules are in the closest proximity, and thus good targets for additional cross-linking molecules. During this time, the DNA molecule may function almost normally, since only one site is blocked. However, the probability for additional cross-linkage between the same molecules has been increased materially by their spatial fixation in proximity to each other.

If a second cross-link is formed, tying together the same two molecules in a second spot, this will fix a larger area of both next to each other, and

will thereby make the access more difficult for the relatively large and clumsy repair enzymes. With two cross-links established, the DNA and RNA molecules involved are now much closer together. A larger part of the DNA is covered by the attached RNA, so that a part of its sites are no longer readily accessible. Repair is thus impeded. This fixes the two principal molecules still closer together. The probability of further linkages increases logarithmically with each new linkage established. In this way, cross-linking generates continuously enlarging probabilities for more damage, leading to a logarithmic curve for mortality, partly resembling the Gompertz type, and more accurately described by Economos (*93*). Cross-linking could, therefore, be the molecular event constituting the "commitment" to senescence and death postulated by Kirkwood and Holliday (*165*).

Soon the structure can become so tight that excizing enzymes no longer will be able to operate on it. The damage will be permanent, then, so far as the resources of the organism are concerned. The possibility of releasing such structures by extraneous means will be discussed hereafter.

It is interesting to note that in plotting the progress of cross-linking *in vitro,* we can obtain a cross-linkage time curve, which can be made to resemble the human mortality curves (*23,39*). The effects of random, unwanted cross-linkages is by no means confined to the genetic system involving nucleic acids; any large molecule containing more than onr reactive electron donor or receiver can undergo cross-linking. This includes almost all large molecules present in a living cell: all proteins, to which every enzyme belongs; all carbohydrates; all polycarboxylic compounds; all di-, poly-, or halo- compounds; oxy- (including enol forms of ketones) compounds; amines; and aldehydes, to name a few examples.

DNA is the most important single molecule relative to aging, but for each DNA molecule there are thousands of large somatic molecules, which, taken together, may be equal and perhaps even more important in the overall aging syndrome. The work of Hart and Setlow proves that the speed of DNA repair in species of mammals roughly correlates with their maximum life span (*132,134*). However, it seems possible that the mechanisms which act rapidly in excise repairing DNA will also be fast in breaking up nongenetic cross-linked aggregates, until these cross-linkages have increased the density to the point that the repair—or hydrolytic—enzymes no longer can penetrate. This form of cross-linking will necessarily occur in all cells, because the cross-linking agents are too many and too diverse to be screened out, and the large molecules are essential to all living cells. However, the consequent changes can take many paths.

Where the primary involvement is with DNA, the effect of a blockage by any mechanism can be the production of errors, which may be immedi-

ately lethal or may be localized. Where the cross-linkage is more widespread, the viable space in the cell gradually becomes crowded. The so-called "lipofuscins" have received more attention than they perhaps deserve because of fluorescence, which attracts the observer. However, the really important property is insolubility. The quantities of insoluble material isolated bears little apparent correlation to fluorescence, but a strong correlation to age. It is not difficult to understand that even a loose network of cross-linked macromolecules will severely interfere with intracellular transport, thus causing increasing numbers of malfunctions in all cells in the course of 60 years plus. Most frequently evident are such malfunctions in those cells which have the greatest biologic leverage: cells of the immune system, cells producing hormones, and neurons. The clinician, being as yet unable to reverse the inexorably progressing increase of cross-linked or otherwise insoluble aggregates, may be able to give many patients quite a few additional years of relatively good health by supplementing or artifically replacing those hormones that threaten to decline below normal levels (*12,13,96,222,266,288*).

C. Hormones

As life developed, and several cells joined forces to form a slightly more complex organism, it became necessary for these cells to communicate with each other in order to achieve some degree of coordinated action. In the total absence of nervous connections, these primitive multicellular organisms depended on mechanisms inherited from monocellular organisms (*215*), which probably had, to some degree, an awareness of light, acidity, oxygen, and perhaps carbon dioxide, as well as the metabolic excretions of their neighbors. At a very early stage, a chemical language developed, with more specific signals for defined purposes. As the cells lacked any computer organ comparable even to the ganglia of ants, the response to the chemical signals took the form of automatically obeyed commands: Seek! Avoid! Receive! Refuse! This system of signals was too useful to abandon, even when our remote ancestors developed the senses that guide us now: the intelligence that shapes present choices, nerves, and synapses. The chemical language of the hormones carries communications and orders that regulate the chemistry of our bodies, our needs, and the husbanding of our resources, which guide everything in our lives that can in this way be automated.

At the age when evolution begins to select against an individual as no longer needed, that individual may be able to gain additional functional years if this automated system is monitored, and those chemical signals that for any reason begin to falter are supplemented. This applies to the

hormones, as well as to the receptors for the hormones that may also fail with age.

The word *hormone* is derived from the Greek word *horman,* meaning attack. This is now recognized as a misnomer, for hormone function is decidedly not aggressive. Hormones are messengers, though some of their messages may have the force of command.

In 1916, Schaffer suggested that the name accordingly be changed to "hermon," for Hermes, who was the messenger of the Greek gods. The name may even be more appropriate than Schaffer surmised, for the divine messenger (according to contemporary illustrations) had his wings on his feet, and was the only bewinged creature so endowed. This must have required extraordinary powers of balance, and indeed, several of the hormone systems have, in recent years, been found to have sophisticated balancing controls.

The only reason why Schaffer's logical suggestion is not adopted here and henceforth hormones not referred to as the "hermones," is that in the Finnish language "hermo" means nerve. The term hermone would thus be misleading for those more familiar with Finno–Ugric languages than with Greek mythology. The hormones most definitely did not develop from nerves; on the contrary, evolution used their precursors hundreds of millions of years before the first slender nerves originated.

The controls and balances of a hormone system have been most thoroughly explored for the pituitary adrenocortical system in which the hypothalamus and some undefined brain centers also enter the control and the pituitary–thyroid systems, which is tied via transcortin, a corticosteroid-binding globulin. The details of these relationships make interesting reading, which cannot be dwelled upon here. A thorough treatment is available in references *95, 96, 101, 222,* and *241.*

This system strikingly exemplifies a general rule in relation to the hormones: If any one link is eliminated, those links which depend on it entirely will atrophy, and those which depend on it partly will weaken.

Another system, which was known to mankind long before the scientific era, is that of sex hormones: The sex glands, as well as organs, atrophy following castration, and primitive cannibals believed that eating enemies' testicles improves sexual performance.

More recently, it has been established that when a neuron is destroyed, its target cells will atrophy, and that the reverse also is true. In general, any damage to any part of a hormone system will weaken its receptor organ in some proportion to the extent of the damage, and to the degree of the dependence on the damaged organ.

This is one reason why any serious decrease in hormone production of a patient should command immediate attention, lest it spread atrophy to

other systems to which it relates. The therapy might include administration of the hormone itself and/or its precursor, releasing agent, controller, or whichever link in the system may seem about to fail. In using such replacement therapy, care should be taken not to render it so strong or so constant that the feedback organs cease entirely to respond to the hormone levels, thus causing permanent atrophy of hormone-controlling and/or -producing cells. A corollary to this is that it is preferable to respond at the highest possible level of organization.

Because hormones are messengers, evolution requires that the individual hormone molecules not be allowed to remain in the circulation. For any kind of hormone signal, each command is carried by the same chemical structure. This must be cleansed from the blood before transmission of the next command is possible.

It would create confusion if a business message received were allowed to circulate again and again after it was acted upon. Channels must be kept clear for new messages. The necessity of a disposal process for every hormone is just as necessary to a living organism as filing or destroying old messages is for a well-organized office.

Herein lies a basic difference between vitamins and hormones. Vitamins may be stored, but hormones must be removed from the circulation when their message has been acted upon. This is mostly done by the liver. Hormones must be synthesized or supplied continuously, because they are necessarily destroyed by the organism itself. Failure or impairment of a hormone-producing gland will therefore quickly become apparent in the decline of a function or group of functions. Such failure can be induced by cross-linking within the cell, impeding intracellular transport.

The peptide-type brain hormones may be those most sensitive to impairment because the mechanism of producing them is generally the most complex. At least in the adrenocortical and thyroid system, it appears that age-dependent failures usually occur higher up in the hierarchy than the adrenal or thyroid glands, particularly in the control or release systems, but also in the sensors (*88*). For example, according to Starr (*265*), the thyroid gland does not show any demonstrable retrogression in thyroid secretory powers with advancing years, and the decline in output originates elsewhere.

Volumes have been written about each individual hormone, so it is impossible here to even try to cover this subject in any detail. As examples, I have chosen five situations, each of which is clinically important and generally illustrative.

1. The Thyroid

The thyroid hormones control the rate of metabolism in the body. As pacesetters, they influence the entire rate of events. When there is a general lack of energy, lethargy, or lack of enterprise, it may be that the thy-

roid function has declined. Correction of a thyroid deficiency can also bring enhancement of the activity of other glands. Too much thyroid hormone can be as bad as too little. Thyroid experimentation therefore requires the supervision of a physician. The onset of deficiency sumptoms is slow, and may escape attention.

There is a general agreement that thyroid correction often improves heart function and reduces fat deposits in the arteries. On the other hand, thyroid hormone can increase body temperature at rest, and this is a negative influence on longevity. Where the hormone is deficient, the net result of supplying thyroid can be striking (*12*). Until the thyroid hormones have been mapped more accurately than they have been, the safest course may be to supply thyroid in the form of dried whole gland rather than pure thyroxine and triiodothyronine.

The thyroid secretion is governed by a pituitary hormone, the thyroid stimulating hormone (TSH). This, in turn, is affected by hypothalamic thyroid releasing hormone (TRH), which, however, is not always effective, and is not now considered as good a choice for clinical use as is the TSH. Indications are that the thyroid gland is not particularly sensitive to changes in aging, and that the drop in thyroid activity, when this occurs, is more often due to a lag in the pituitary TSH system (*114,265,281*). However, here too, the interdependence of many systems and cybernetic feedbacks are in evidence (*121,153,241*), including also the adrenocortical system (*222*). Quite generally, many phenomena of aging are probably due to neuroendocrine and autonomic loci (*45,88,95,222*).

2. The Sex Hormones

It is almost a universal rule that the female of the species outlives the male. This has definitely been tied to one of the female sex hormones—estrone. If castrated animals are fed testosterone, they will show little change in life spans, while estrone clearly increases longevity, possibly by affecting genetic transcription (*131*) and/or blockages (*167*). In humans, estrone counteracts the deterioration of the arterial system. The sharp increase in apparent aging rate following the menopause is well-documented, and the increased longevity of females, as compared to males, in recent years (*281a*) coincides convincingly with the advent and widespread use of the contraceptive pill. This is true in spite of its negative effect on heart infarct rate, especially in vitamin C-deficient individuals.

The male hormone, testosterone, in spite of some early hopes, has no comparable life-extending property. One of the most challenging tasks of sterol chemistry would be to find a steroid that has the life-extending qualities of estrone, without its feminizing properties. So far, nothing of this nature has been generally accepted, though a number of anabolic steroids have been marketed.

Two mechanisms of action are common to these anabolic steroids and the other metabolic hormones. One is by way of the genetic equipment that regulates the synthesis of the specialized proteins to which these hormones are bound, and the other, a finer and more rapid adjustment, is by the generation of cyclic adenosine 3′,5′-monophosphate from adenosine triphosphate (ATP), in a reaction catalyzed by the enzyme adenylcyclase and inhibited by phosphodiesterase. The first mentioned enzyme is present in cell membranes, and can thus be turned on by chemicals that make contact with the exterior, without having to penetrate the membrane. The cyclic adenosine monophosphate thus generated would then, as a second messenger, effect the hormone action. High-molecular protein or protein-bound hormones can act without entering the cells by this triggering mechanism, which extends through the cell membrane. If a lack of hormones is the primary cause of failing short-range memory, ascribed to a deficiency in protein synthesis, administration of the appropriate steroid would help (*127*). This treatment may be effective for many years, but it must be supported by a functioning genetic mechanism and by a functioning energy-release chain. The latter, in particular, presupposes a steady, fully adequate simultaneous supply of all the vitamins that are a part of any enzymes of the respiratory chain, including those needed for production of high-energy phosphate. Failure of short-term memory is a warning sign that protein synthesis is weakening, and that the diagnostic identification of the weak link in a long chain should be attempted.

These hormonal effects and their dependence on reactions that in turn depend on vitamins lead to the study of differences in results with certain vitamins in males and females. For example, McCay *et al.* (*195*) showed in 1943 that the life span of female rats could be prolonged 21% by administration of vitamin E. This was a statistically significant result.

The males only showed a 4% life extension, which is not significant. However, the control males lost their sperm motility in 150 days, whereas those receiving vitamin E retained motility after 750 days. Tappel (*275*) investigated whether much higher dosages of vitamin E might cause substantial life extension in mice. His results indicate that this is not the case. The relation of vitamin E to life extension appears to follow the curve predicted by Hickman when he asked how much of the preservative factors we can have without depressing active metabolism, making the organism as a whole lethargic (*141*).

3. Insulin

The clinical use of insulin is multifaceted and too complex for entering into all of the ramifications. A central fact that explains many aspects of the metabolism of a diabetic is that the brain normally uses only glucose

for its energy source. It cannot use any other carbohydrate, nor any fat. As space is at a premium within the brain case, volume can not be spared for avoidable carbohydrate processing. Glucose became the logical choice in evolution. Viewed sterically, it is the most elementary carbohydrate structure with the maximal number of reactive groups readily accessible in the equatorial plane of the molecule.

The brain depends on glucose, and jealously guards its supply. If any shortage of glucose seems to threaten, the brain will issue hormonal commands to assert its top priority for glucose. These commands, given by the hormone glucagon, have the powers to inactivate insulin already in the circulation, and of reducing its formation. They thereby prevent the muscles from using glucose, leaving substantially all of this critical sugar to the brain. When this happens, it seems vital that glucose or its polymers, (e.g., boiled starches, dextrins) be continued in the diet whenever possible and that other means than glucose deprivation be used as the mainstay of the treatment (*6,55,56,209,231,233*).

Diabetes is a principal ailment of old age. At the Tokyo Congress of Gerontology (1978), it was stated that over 90% of the 60-year-olds in a large United States institution would be classed as diabetics if judged by the criteria as applied to the 20-year-old. In Japan the corresponding figure is only 45%. A reason for this difference could well be the preponderance in the Japanese diet of complex carbohydrates that on hydrolysis principally form glucose, thus satisfying the demand by the brain.

4. Adrenocortical System

The adrenocortical system has been more thoroughly explored than any other endocrine system, particularly with regard to the cybernetic details. Due to space limitations, only key literature, particularly the three-level model of communication and control hierarchy of this system will be referred to in this chapter (*222*, p. 160, Figs. 6–8; also *88,105,121,127*).

5. Placenta

The importance of the placenta as an endocrine organ may not have been fully appreciated. There are, however, strong indications pointing to such a function. In pregnant animals, the adrenal cortex continues to be stimulated even after removal of the pituitary, in a manner strongly indicative of the secretion by the placenta of a substance that is similar in function to ACTH (*166,248,261*). Canivenc (*59*) and Ray *et al.* (*235*) likewise indicate that the placenta has a related secretory function.

It seems plausible that the protein and the large molecular polypeptide hormones or growth factors from the brain, pituitary, and hypothalamus are unable to pass from the mother's circulation to the fetus. Inasmuch as

at least a couple of months may be necessary for the human embryo to develop the corresponding biochemistry, it seems highly probable that the the placenta is equipped to furnish the growth factors and hormones needed in the early stages of fetal development.

That this is so is strongly indicated by the work of Soviet scientists, particularly at the V. P. Filatov Institute of Ophthalmological Diseases and Tissue Therapy in Odessa (*128,212,262–264*). Working with corneal transplants from deceased persons, academician V. P. Filatov (in *262*) and co-workers observed that preexisting ocular infections of transplantees often cleared up radially from the transplant when treated with placental extracts, and that this therapeutic effect could be enhanced by preservation of the corneas at low temperatures several days prior to transplantation [at temperatures low enough to preclude bacterial growth, while still permitting some autolytic enzyme reactions to release bound bioactive substances (*263*)].

Having developed this procedure, they tested the effects of extracts from numerous other tissues of man, animals, and plants, but they found that the placenta is a most favorable source material (*128*). Substantial clinical material with humans has been published (*212,262–264*). Taken together, an impressive body of evidence indicates potential clinical utility of the placental preparations. It appears that the placenta also can supply the fetus with an ACTH-like substance and with a growth factor, in addition to other hormones.

In addition, the aromatization of steroid occurs in the microsomal fraction of human placenta (*248*). This is notable also because cortisone and hydrocortisone can extend the life span of human fibroblasts *in vitro* by about 20% (*199*).

Mayer and Canivenc (*59,205*) add observations which strongly imply that the placenta secretes hormones closely similar in function to some of the hormones of the pituitary gland. It also is a supplemental source of estrogens (particularly estriol) and progesterone.

Thus the placenta is a versatile gland, capable of serving as a producer, not merely a filter, to supply a wide range of those hormones and growth factors that the fetus is not yet ready to synthesize and that do not pass the mother–fetus blood barrier. It also supplies some estrogens in order to aid the maternal peak requirements.

6. *Nerve Growth Hormone*

Santiago Ramón y Cajál postulated the existence of neurotropic agents, directing nerve growth to receptive targets. The chain of events leading to the finding of sources rich in the postulated factor in mouse salivary gland sarcoma and in snake venom has been recently described by the pioneer in this field, Rita Levi-Montalcini, and P. Calissano (*187*).

A breakthrough in human nerve growth factor was made in 1978, when L. D. Goldstein, C. P. Reynolds, and J. R. Perez-Polo (*119*) isolated the corresponding nerve growth stimulant from human placenta as an endocrine hormone. The product purified by them by isoelectric focusing following chromatography is a basic protein of about 140,000 daltons. The amnion had the highest specific activity, but the placental cotyledon had the highest total activity and was therefore chosen as the starting material for purification. Isoelectric focusing yielded a basic protein in which activity seemed centered. Though not identical with the mouse sarcoma growth factor, this human placental hormone is chemically related to the mouse product; it partially cross-reacts with antimouse β-nerve growth factor IgG and has a similar isoelectric point and specific biological activity. The activity of the human placental hormone is very high—1 to 3 ng per biological unit (*16*, p. 177).

The work in this field is still in a state of rapid progress. If one may presume to attempt piecing together scattered information into a coherent picture, the following speculation might not prove too far afield: Neurons need nerve growth hormone to exist. They receive these hromones from their target organs or connections. This tropism is the force that guides the growing axon or dendrite to its target.

The circuitry of the nervous system in the fetus is established through some combination of genetic and extrinsic factors (*173*, p. 50), of which this tropism toward sources of nerve growth hormones is a principal part. The death of neurons that have not made the neural connections prevents overpopulation and leaves space for growth.

The fetus, however, needs a certain lead time for its nervous circuitry to become established. Therefore, nerve growth hormone must be supplied to the fetus from an external source in order to keep the competing neurons alive long enough for the essential connections to be made. The fetus cannot synthesize this hormone during its first months; evidently the placenta supplies this need by its endocrine function.

At the other end of human life, when hormone syntheses weaken, the most complex hormones may be among the first to decline. Whenever an organ is not used for some tiem, corresponding neurons may perish from lack of nerve growth hormone from that source. Then there is no regrowth, and a link in the chain of life is broken. Timely administration of the hormone might prove a potent means for avoiding premature debilities of any nerve-connected system, particularly those of a cybernetic character (*267*).

It is not yet known whether or not atrophied dendrites could be restored and synapses reinstated by timely administration of nerve growth hormone (purified or as an extract of assayed strength). Dosage is critical, because overdosage in experimental animals leads to dangerous in-

growths. We cannot discount the possibility that the active part of the nerve growth hormone is a much smaller molecule than now surmised, when freed from a carrier protein. The homeostatic properties of injectable placenta suspensions or extracts (*263,264*) might point in this direction.

VIII. DIRECT CAUSES OF PREMATURE DECLINE AND DEATH

The causes of death stated on a death certificate of older persons are seldom the principal causes of demise. With increased resistance to traumata in the young, these causes could have been overcome. A review of some of the most frequent series of events leading to decline and death will illustrate this.

A. Neurological Causes

The evident anatomic changes in the aging of the nervous system include a withering of the horizontal dendrite system, which brings the cortical neurons into extensive synaptic contact with each other. As a consequence of these changes, the neuronal volume becomes filled with inert, cross-linked aggregates, a distortion which causes distortion of the cells, displacement, and ultimately the atrophy of the nucleus (*48,49*).

However, there is evidence that this process can be countered if necessary to prevent a racially disadvantageous early death, even of the nondividing neurons. Von Buttlar-Brentano has shown that, in specified vital areas of the brain, there is virtually no cell death of the pyramidal neurons, which, short of division, accomplish changes to achieve functional survival. The size of some neurons is increased so that the cell surface area increases 8 to 10 times, the nucleoli undergo division (there may be up to six nucleoli in the nucleus of the cell) and the membranes fold to increase their surface area (*285*).

It thus appears possible for neurons to avoid cell death and remain functional by changes short of cell division. How this change is biochemically triggered is not known, and this may well prove to be a fruitful area for research.

There is an old Roman saying which states that "Whom the gods want to destroy, they first make very crazy." In modern form, the cybernetic theory of aging by Still (*267*) points to the central significance of the coordination of the numerous electric or other messages that cross the system, in an orderly, precise sequence, not to be disturbed without risk (*267*).

The understanding of these processes may be substantially advanced by recent attention to the onset of senile dementia, or Alzheimer's presenile disease. The latter is a loss of short-term memory, followed by motor disfunction, which progresses inexorably and terminates with death, usually in about 5–10 years. D. R. Crapper in Toronto has now succeeded in inducing the same sequence of symptoms in a compressed time sequence, by injecting a single dose of 100 nmol of aluminum chloride in the hippocampal ventricle of cat brain (*70–74,76,77*). Such a simulation of a disease often signals a breakthrough. A histochemical analysis following autopsy showed that aluminum is concentrated selectively (*85,86*) in the chromatin of the pyramidal neurons. There have been reports in the literature of similar aluminum deposits being found upon autopsy of persons who died from cerebral diseases (*92,179*). Furthermore, aluminum is known to accumulate with progressive age up to about age 55. The analysis after autopsy of brains of age 60 and older shows declining amounts of aluminum (*163*). Could this be, as Zinsser *et al.* (*305*) suggested in 1962, because those who accumulate aluminum at a more rapid rate tend not to survive their 60th year?

The mechanism of the damage seems clear. Aluminum is a powerful cross-linking agent for proteins and presumably also for nucleic acids, which contain several sites to which aluminum can bind. Indeed, aluminum is used industrially in tanning, to cross-link collagen (*17,104,221;* also *129,* p. 335). It could thus easily inactivate the sensitive end organs of the neurons, as well as cause blockages in the genetic molecules and disturb intracellular transports. Furthermore, aluminum is a powerful flocculant (*238*) and could by precipitation of colloids contribute to shrinkage of the brain, leading to the widening of fissures and ventricles, symptoms often encountered in advanced years. The observations of Crapper *et al.* thus add weight to the concept of exogenous causes of age-dependent mental symptoms.

That this actually occurs at least in some cases is indicated by Lapresle's autopsies (*179*) and the finding of aluminum in Parkinson's disease (*92*) and in Down's syndrome as well (*68*). How common is the aluminum induced syndrome in man? For all we know, as yet, it could be an important cause. Aluminum is 8.4% of the earth's curst, and one can hardly avoid ingestion of about 30 mg/day. It is commonly believed that aluminum is not absorbed from the intestines, but these analyses have not been done with nanogram accuracy. It appears that during a lifetime a few milligrams might indeed have passed through to heart and brain. Could other cross-linking agents cause the same syndrome? If aluminum were a major cause, would it be possible to remove it before damage becomes serious, either by one of the chelating agents, listed by Zinsser *et al.* (*305*)

as being effective on aluminum *in vitro,* or by electrolysis (*43*) or could careful use of nerve growth hormone help restore synapses? The coming years should answer these questions.

The monoamines play an important role in brain physiology. The monoamine oxidases tend to increase with age. This provides a rational for administering a monoamine oxidase inhibitor (*8,57*) efficient enough to restore the balance in monoamine metabolism but not powerful enough to block necessary reactions. Indications exist that procaine, in proper dosage and formulation, can meet both requirements (*9,57*). Several of the more powerful monoamine oxidase inhibitors are less safe to use, as they have caused death when given together with cold remedies, some tranquilizers, and tyramine-containing foods such as cheese (*60,67,202,277*). The phosphatides have a great influence on brain physiology. In particular, lecithin and cephalin are essential in carrying water-insoluble hormones, vitamins, and raw materials in the bloodstream and also are essential constituents of all the cell membranes (*46,141,142*). Of the vitamins, pyridoxine is an essential part of the transaminases, a most important group of enzymes in the brain, and nicotinic acid, which is a constituent or activator of at least 54 known enzymes and probably of as many more not yet discovered. The evidence indicates that the brain has other functions not generally considered until quite recently. These deal with electrical function and the generation of potentials that may be of importance in the healing of injuries, as well as with feedback mechanisms related to endocrine systems (*222,267*).

The oxygen consumption of the brain is about 25% of the total oxygen consumption of the body and there is a balance between the oxygen content entering the brain and that leaving it; the oxygen utilization declines with advancing age much more than can be accounted for by changes in blood circulation. More often than not, severe and fatal neurological disturbances occur even in spite of substantially undamaged circulation, as determined from autopsies (*2,3,15,120*). Evidently, the fault in these cases is not so much due to changes in the capillary bed (*206*), (Fig. 2) but rather to cellular changes, which is most likely due to the failure of one or several links in the electron transfer or oxidation chains (*37,230;* see *206*, Fig. 6). The recent substantial progress in our understanding of the details of these processes (*125,126*) foreshadows major advances in application.

Particularly important is the breakthrough made by F. F. Jöbsis, A. L. Sylvia, J. C. Lamanna, and their co-workers in the methodology of following redox reactions *in vivo* by infrared transillumination and dual beam spectrophotometry. The method has been used to monitor noninvasively from moment to moment cerebral and myocardial sufficiency and reduction states, even in human patients, as well as for detailed study of the cytochrome systems *in vivo* (*154–157,269,270*).

It has been pointed out that the amount of blood passing through the capillary bed of the brain is not much changed with increasing age because the increase in length of the capillaries is compensated for by an increase in diameter, so that the total passage may remain approximately constant (*206*, Fig. 2). However, the progressive cross-linkage that takes place over the years may result in a reduced permeability of the capillaries, and the greater diameter of these in the aged brain may be accompanied by increased wall thickness; both factors should tend to impair oxygen transfer.

Considering the enzyme, fructose-6-phosphate kinase activity decreases significantly with age, and Meier-Ruge and co-workers (*206*) present this as a probably important factor in aging. The next question is, if the formation of more fructose-6-phosphate kinase can be successfully supplied or induced, how great will be the improvement, how general, and what will be the next bottleneck in the oxidative pathway? This is now one of the most promising approaches in geriatric chemistry, for the cell energetics may well hold the key to early progress (*37,87,126,156,167,270*).

B. Circulatory Deficiencies

Foremost of the circulatory deficiencies are the sclerotic changes that lead to progressive occlusion and ischemia. A great deal of research has been devoted to this syndrome, which is the immediate cause of about 50% of all deaths in the United States. It has been conclusively proven that the first measurable step toward sclerotic development is leakage that has developed in the endothelium of an artery (*65,66,106,107*). This can occur in several ways. The most apparent mechanism is as follows: The large arteries consist of three layers—the intima, the media, and the adventitia. The intima is lined, on the side facing the circulating blood, by a thin dense film—the endothelium, which is impermeable to liquid flow. The two outer layers are porous, i.e., blood serum passes through them easily. Circulating in the blood are a large number of cross-linkage agents. As the years pass, the cross-linking agents will gradually harden the proteins in the endothelium, so that it no longer is elastic enough to follow the pulsations. Finally, the endothelium cracks, usually at a point of maximal hydrostatic pressure, such as the concave side of a bend, or at a point of branching of an artery. The blood leaks through the break and blood serum oozes through the porous outer layers, leaving behind a filter cake, consisting of everything insoluble and filterable in the blood. This includes blood corpuscles, fats (mostly in the form of chylomicrons, or fats carried by low-density lipoproteins, and sterols, including cholesterol. If this filter cake consists mainly of blood corpuscles and unsaturated fats, it

may be possible for the defense mechanisms to resolubilize it. The lesion may then regress. It will not heal if it consists largely of difficult-to-dissolve substances, such as fats with a melting point above body temperature, which includes most triglycerides or cholesterol. The lesion will then be invaded by cells such as fibroblasts and cells resembling phagocytes, and will thicken and calcify progressively. Plaques will become more developed, the process will repeat itself, and will finally lead to angina, thrombosis, or fatal occlusion.

As already mentioned, leakage of the endothelium has been fully established as a common step in the formation of intimal plaques, the earliest visible step leading to sclerosis. The initial break has been seen often (*124*), and the plaques can be provoked by damaging the endothelium artifically. Some modifications of this general process have been described and illustrated (*27*). The pulse wave at bends and branches may cause disruptions that kill small areas of endothelial cells, causing permeability, or the endothelial cells may part at their boundaries, forming microscopic cracks that are difficult to observe optically but that are readily measurable by isotope penetration (*66*).

The role of cholesterol in this syndrome is thus secondary to cross-linkages, the loss of elasticity the latter cause, and consequent permeability. Nonetheless, cholesterol is important, both as a rough indicator of the state of fat metabolism and intake, and also an inhibitor of healing of the initial lesions. Cholesterol is a precursor in the synthesis of steroid hormones, and is thus necessarily present in the body. It is regularly synthesized in the liver. An excellent discussion of relevant mechanisms in greater detail is given by Wolinsky (*303*). Considerable differentiation is now found between cholesterol bound to low-density lipoproteins (LDL) and cholesterol bound to high-density lipoproteins (HDL). The HDL owe their high density mainly to their higher phospholipid content. The phosphoric acid groups in the phospholipids contribute essentially to the higher density of the HDL (*239,284*). Cholesterol carried by HDL is considered to be exiting from the arteries, or not readily deposited, whereas the LDL transport cholesterol on the way to being deposited, and thus are detrimental. It appears that the HDL content is a favorable indication of reduced risk of circulatory disease, whereas LDL is unfavorable (*61*).

In nature, lecithin and/or choline are the principal nutrients for producing phosphatides. Lecithin usually occurs together with cholesterol, e.g., eggs contain on the average 0.32 g of cholesterol and 1.70 g of lecithin. The high lecithin content may be a principal reason why intake of cholesterol in eggs appears much less favorable than the same quantity of cholesterol in meat. If a patient is on a strictly cholesterol-free diet, it may be advisable to check his choline or lecithin intake, to assure their adequacy.

However, hyperlipemia (including cholesterolemia) is by no means the only cause of sclerosis. Its greatest importance occurs in the 40–70-year age group. Yet the study of Goldstein *et al.* (*118*) of lipid levels in 500 survivors of myocardial infarction indicates that most of the hyperlipemic men and 75% of the hyperlipemic women are dead before reaching age 70, and that hardly any hyperlipemics survive beyond the age of 80. Yet, heavy sclerosis still occurs at that age.

Cross-linkages other than those formed by autoxidation of fats appear to assume a higher degree of significance at ages above 80. These include, for example, quinones, polyvalent metals, many aldehydes, and many free radicals, which will render proteins inelastic and lipophilic, regardless of the presence or absence of oxidizing lipids. The processes involved are well known, because in industry planographic printing is based on converting hydrophilic proteins to lipophilic by cross-linkage (*21*, p. 346).

C. The Immune System

It is no longer doubtful that the immune system is a principal factor in the aging syndrome, as Walford postulated (*289,290;* also see *29*, p. 54). The close relationship of the immune system to the cross-linkage reactions has also been fully established. The immunological precipitants are bifunctional and, thus, inherently cross-linkers (*234*). The most concrete pathologic immunoprecipitates, the amyloids, have been shown to be protein aggregates cross-linked by the bifunctional precipitins. Since the easiest way nature can find to precipitate a large molecular substance is to cross-link it to itself or to other large molecules, it is natural that the immune system should have developed on the basis of cross-linkage. When cross-linkage occurs to such a degree that the aggregates become inaccessible to lytic enzymes, "amyloidoses" develop (*159*). Mutually cross-linked chains of β protein structure are often encountered (*292*). The hydrogen bonded areas are more easily analyzed than the complex of aminoglycan, collagen, reticulin, fibrinogen, and complement components and lipoproteins, which occur in amyloid and which appear to be covalently bonded (*292*).

The immunologic significance of the thymus and its history have been skillfully traced by Good (*122*, p. 1420). Thymus atrophy is another important cause of diminution of the immunologic defenses (*16,96,201,222*). The most likely extrinsic cause of this atrophy is a corresponding regulatory breakdown in the hypothalamic–pineal–pituitary–neuroendocrine axis in relation to the thymus. This, in turn, might well have its roots in interference with intracellular transports due to developing net structures of macromolecules, interlinked by progressive cross-linkage.

D. Other Degenerative Diseases

The foregoing instances highlight some clinically important cases. Rheumatism is often an immunologic disease. Osteoporosis may have its root in the embrittlement of the collagen matrix and in a loss of adhesion, both being consequences of cross-linkages involving the collagen fibers (*115,130,293*). Senile cataract has been discussed by Bellows and Bellows (*18*), who ascribe it to progressive cross-linkage, including cataracts caused by frequent exposure to intense light. Ogino *et al.* (*217,218*) have shown a connection of senile cataract with quinone compounds (*282*). Cancer is much too complex a subject to be discussed within the present framework; however, it may be appropriate to recall that every known radiomimetic chemical is either a cross-linking agent or has a metabolite which is (*5*). Cross-linkages between the two strands of DNA have been proven by isolation of certain dimers (*133*); the essential factor here is probably the bond itself, rather than the specific cross-linking structure.

IX. APPROACHES TO THE EXTENSION OF A HEALTHY LIFE

A. Less Than 30-Year Extension, Based on Present Knowledge and Resources

Nutrition and Exercise

Taking as a guideline LeCompte's Law (see Section III), we can even now proceed to correct deficiencies, so that all of the known metabolic links can meet peak requirements with a wide safety margin. Any one of our metabolic pathways runs a measurable risk of being blocked unexpectedly by some physical or biochemical accident. There should then be enough of a safety margin on each of the possible alternate pathways to handle a sudden call for excess capacity. In a sudden crisis, redundancy is the key to survival.

This means nutritionally maintaining a constant excess supply of those substances that the body might need in any emergency and that it cannot manufacture itself, such as vitamins and essential minerals. Physiologically, exercise should be engaged in regularly, and to such an extent that the body is maintained in a condition to meet greater demands than are likely to be placed on any part of it during normal functioning. The condition should be monitored at reasonable intervals, so that any deviation is observed and corrected before an emergency develops. As an example, two years ago I had an escape from a situation that could have had serious consequences unless detected. I used to spend a couple of months every

summer at the sea shore, where the diet included a good deal of fresh, salted fish. A spectrographic analysis of the ash of my hair suddenly revealed a thirty times higher sodium content than the previous steady figure at the low side of the normal range. Although I found it difficult to believe this value, I immediately underwent a physical examination. The enzyme profile and other biochemical indicators were normal, but my blood pressure had jumped from the customary 125/75 to 175/110! It was still correctable. No observable damage had been done, but had this remained undetected, it might well have proven disastrous.

Furthermore, it is possible to avoid many poisons and to remove some that are found to be accumulating. However, our knowledge is still incomplete. Protection against unknown deleterious factors is a diversity in nutrition and habit as well as avoidance of placing too much reliance on any given treatment or therapy.

B. Large Gains in Healthy Life—How Might a Breakthrough Be Achieved?

In a time span of 60 years, at body temperature, every chemical reaction that is theoretically possible will actually have taken place to some extent. If any of these reactions produces any amount of an insoluble, irremovable compound, that compound will accumulate in 60 years to an extent that must be reckoned with. Therefore, to prevent progressive deterioration beyond present age limits, we must be able to remove from all parts of the body, without exception, a very wide spectrum of insoluble, normally irremovable compounds. Most commonly these compounds will be those formed by cross-linkage, for this is the type of insolubilizing reaction in which the smallest input causes the largest change. In aging, it is irrelevant what the compounds are, as long as they accumulate and the body cannot excrete them or break them down to excretable fragments. Nor is any specific solution important, even if it could be applied to many of the presently most conspicuous compounds, because these, too, represent a wide diversity. We now need a breakthrough in finding a generic reaction or force applicable to the living organism and capable of breaking any carbon-to-carbon or perhaps any carbon-to-nitrogen bond, so as to reduce to excretable fragments any organic compound against which this force is directed. We further need a directive system such that the action of this treatment can be controlled, so that only a tolerable, *continually replaceable* amount of the normally present, essential molecules are removed, while at the same time the unwanted compounds, formed slowly by random accidents, will not be similarly replaced. We now have three possible methods for this, each of which has its advantages and limitations (*33*).

1. Methods Involving Metal-Based Bonds

Methods involving metal-based bonds are not general, but they apply only to the removal of molecules that contain metal atoms as the central, cross-linking components. However, this is an important group of age-dependent accretions, as already pointed out in connection with the neurological symptoms in the aged, the Alzheimer syndrome, and Dr. Crapper's cat experiments. They also represent some of the first experiments in this direction. In 1960, it was shown that, using chromate as a cross-linker, it is possible in a model system to produce a time–viscosity curve that very closely resembles in shape the mortality curve for humans (*39*). This, coupled with the observation of the increase in metals accumulating in the organism with age (*163,305*) provided good reasons for considering a means of removal of such metals.

Chelation provides such a means. The principal chelating agents are not absorbed well from the intestines and have to be administered by either injection or infusion. Infusion of derivatives of ethylenediaminetetraacetic acid (the disodium salt is now mostly preferred) is approved therapy for removal of lead in cases of lead poisoning, and also for removal of radioactive, bone-seeking elements such as plutonium, strontium, and yttrium (*174*). In the last 10 years, thousands of patients have been treated for scleroses by chelation, with an impressive number of successes. The standard method of treatment (*62,63*) is now infusion, over a period of 3.5–4 hr daily, of 500 ml of a solution of 3 g of disodium EDTA in Ringer's solution. Each treatment could theoretically solubilize about 800 mg of apatite or equivalent amounts of other calcium compounds. It is, however, not very specific, and will also "pull out" iron, aluminum, magnesium, some cobalt, and other polyvalent metals. These must be continuously monitored and replaced orally. Clinical attention must be given in particular to the possibility of depleting calcium dangerously, because the parathyroid might become deficient and cause cramps. To avoid this hazard, a good practice is to add a calcium gluconate solution to the infusion liquid about 10 min before termination of the treatment and to supply needed trace elements orally.

With proper care, the hazards can be avoided; however, the continual replenishment of the normally needed metals cuts down the theoretical efficiency of the treatment, so that 20 treatments (about 80 hr of infusion) are considered minimum, and a temporary cleansing of a serious sclerosis may require several times more. An alternative to the chelation by infusion is injection of a chelator, which is tolerated when so administered. Deferrioxamine has been so used and some favorable results reported (*208*). The principal danger of chelation is that rapid treatment may undercut lime deposits so as to loosen pieces of the plaques and to cause throm-

bosis. Some deaths have been ascribed to this. On the other hand, it is reported that the survival rate as a whole is better for those taking chelation treatment where this is indicated than for an otherwise comparable control group. The celebrated Evers case of June, 1978 was decided in favor of Dr. Evers on the basis of copious evidence of favorable results, justifying wider clinical use of chelation (*97*).

2. Lytic Enzymes of Low Molecular Weight

Soil bacteria probably contain enzymes capable of dissolving most of the cross-linked aggregates that defy enzymes normally present in organisms. Since the original components of the aggregates are metabolizable, it seems clear that the insolubility is due to the formation of cross-linked cage or net structures of such density that the active enzymes cannot position themselves at the active sites or may, at the most, be able to nibble at the edges of the aggregates at a rate slower than the growth of the aggregates by new large molecules becoming enmeshed.

The target of the search on this approach was therefore to be found among microorganisms present in the soil, producing lytic enzymes characterized by extremely low molecular weight.

Screening for organisms was done by the method introduced by Dubos (*91*). The candidate organisms were subcultured and the extracts passed through an ultrafilter, filtering out molecules larger than 10,000 MW. This greatly simplified purification, since it removed complicating macromolecules of higher molecular weight. A considerable number of enzyme-forming organisms were found in 4 years of screening. The most promising of these was a strain of *Bacillus cereus*. Both ease of culture and yield of enzyme were taken into cinsideration (*44*). The low-molecular-weight bacterial protease was isolated and is now commercially available as a laboratory chemical (*303a*). It is being evaluated, principally for the lysis of blood clots.

Any evaluation of effects on longevity of man may require several decades for completion. Chelation is a useful tool for attacking those aggregates in which metal atoms are the central binding agents. The low-molecular-weight enzymes of bacterial origin are relatively mild, and are more penetrating than any enzyme previously reported by virtue of their small molecular size. These compounds are active on aggregates that could not previously be attacked. For example, in tests on bird feathers, they digested 65% of the feathers, as compared to 15% with Pronase and 10% with chymotrypsin. The lowest-molecular-weight enzyme fractions occur in conjunction with enzymes of higher molecular weights, as shown in Schenk and Bjorksten (*253*) and Holmquist (*144*). The relative quantities and stabilities of these low-molecular-weight enzymes depend greatly on

the ions present. Removal of polyvalent metals with EDTA results in a rapid inactivation in solution at room temperature (*253*).

The high-molecular-weight *Bacillus cereus* proteases have been variously described (*99,144,152*). Only the low-molecular-weight active fractions (*253;* also *144*, pp. 4592, 4593) appear of possible interest in geriatric contexts.

3. Lysis by Free Hydroxyl Radicals

Another means of breaking up those very dense aggregates, which even a 1500-MW enzyme fragment could not penetrate, is lysis by free hydroxyl radicals. For this purpose we turned to the six enol–diene compounds with proximal carbonyls that Daubenmerkl singled out (*83*) as being able to depolymerize hyaluronic acid. Among these, ascorbic acid is the best known and so likely the safest to use in combination with an oxido-reductive system for producing free hydroxyl radicals *in vivo*. Oxidative breakdown of hyaluronic acid by peroxides and ferrous salts was previously known (*260*). Pigman *et al.* (*229*) showed that peroxides react to form free hydroxyl radicals; this reaction could be made continuous by supplying ascorbic acid with ferrous (*203*) or cuprous salts. In using ascorbic acid and a ferrous salt to depolymerize hyaluronic acid, Pigman *et al.* (*228*), based on Ingold (*150*), envisioned the reaction sequence as follows:

$$O_2 + Fe^2 \rightleftharpoons (Fe^3O_2^-)$$
$$(Fe^3O_2^-) + RH \longrightarrow Fe\ ^3O_2\ H + R \longrightarrow$$
$$Fe^2 + HO_2\cdot + R\cdot$$

$$R\cdot + HO_2\cdot \longrightarrow RO\cdot + Fe^3 + OH^-$$
$$R'\cdot H + RO\cdot \longrightarrow R'\cdot + ROH$$

where RH is ascorbic acid and R'H is hyaluronic acid.

Richheimer and Robinson (*237*) postulate that free hydroxyl radicals are the active agents.

Lagercrantz (*172*) showed that monohydroascorbic-acid free radical (MDA) is formed in autooxidizing solutions of ascorbate around the neutral pH range. With hydrogen peroxide it forms free hydroxyl radical and dehydroascorbic acid (DHA).

$$MDA\cdot + H_2O_2 \rightarrow HO\cdot\ H_2O + DHA$$

That the free hydroxyl radical is active in the depolymerization reactions is supported by numerous findings (*237*, p. 15).

The free hydroxyl radical is about as penetrating as any substance one could hope to generate in the organism. With a 17 MW, it is two orders of

magnitude smaller than our smallest possible enzyme fragment, but more importantly, it is electrically neutral, so that it is not absorbed as hyroxyl ions would be. It can move maong the molecules as freely as a neutron can move among protons and electrons. The free hydroxyl radical can cause cross-linking as well as fission. When the substrate molecules are close together, as in solids or high concentrations, cross-linkage will predominate, but in the biological dilutions here in question, the net effect is one of fission, solubilization, and depolymerization. Pigman *et al.* showed the use of this system in breaking down all carbohydrate polymers tried, including starches and celluloses (*139,203*). Herp *et al.* (*140*) extended this to the breakdown of synthetic polymers, Orr (*220*) to the enzyme catalase, Robinson *et al.* (*240*) to transferrin, and other proteins and polypeptides. Richheimer and Robinson proved that the free radicals can directly break peptide bonds, which means that they are able to destroy proteins, including enzymes, as well as nucleic acids and a host of other biochemically active substances (*237*).

Thus, ·OH should be able to break the most refractory organic aggregates formed in aging. That this is actually the case has been demonstrated *in vitro* with brain and liver insoluble fractions from old rats (*33,40*).

It seems plausible that an important biological function of vitamin C may be its utilization by leukocytes in destroying bacteria and other foreign invaders. The "bullet" may consist of free hydroxyl radicals, and the "ammunition" may be ascorbic acid in connection with peroxides and catalytic amounts of iron. Although this has not been proven, it is most consistent with the following known facts: Ascrobic acid, on injection, is stored primarily in the luekocytes; the consumption of ascorbic acid increases enormously in an organism fighting infections (*189,298*); and the system is capable in vitro of destroying even the most refractory organic substance it might encounter *in vivo*.

The main problem in using this "ORD" reaction [Pigman *et al.* (*229*) named it "oxido-reductive depolymerization"] is that the free hydroxyl radicals will destroy normal tissues as well. However, this can be countered by carefully modulating the dosage. The gerogenic aggregates have been formed slowly by random reactions. The normal tissues and body chemicals, when consumed, are regenerated at a much faster rate. Even the DNA in a nondividing neuron can be gradually renewed, atom for atom, and the speed of this renewal, as illustrated by the rate of repair, is proportional to the specific life spans of the various species (*132,134*).

Together the three pathways here outlined represent an "arsenal" that should be effectively applicable to the problem of gerontology: a breakthrough in extension of specific lifetime, consistent with health (*33*).

X. THE CENTRAL FACTS

The following facts are central: If a nerve is destroyed, its target organ will atrophy. If an organ is destroyed, the neurons wholly dependent on it will atrophy, and those partially dependent on it will suffer damage, proportional to the degree of dependence. If an endocrine gland ceases to function, its target organs will atrophy. If a target organ is destroyed, the endocrine gland wholly serving it will cease to function, and other organs interconnected with it will suffer damage, dependent on the degree of interconnection.

The organism can be viewed as a web of interconnected metabolic chains, each consisting of numerous links. The failure of *one single link* causes the chain of which it is a part to fail totally. If the chain is a vital one, this may mean death. In many, if not most, cases, compensating changes will take place, and these may or may not be sufficient.

If a link in one chain is identified and replaced or supplemented by medication or nutrition before it has wholly failed, then no damage is done and one danger has been averted. Enough is now known in order to identify many, though not all, of the links. By making better use of knowledge than is now generally done, it should be possible to add at least 5 years, and possibly up to 20 years, to the present healthy part of life. By adding to the knowledge of the prevention of aging, we might possibly succeed in doubling this.

Evolution has provided us with a genetically programmed, semiautomatic maintenance of the chains, designed to last about 60 years. After 60 years, this repair system progressively wears down. One of the links will fail first, and regardless of whichever it is, many others will weaken. For each weakening, death comes closer.

A greatly increased specific life span would require an automatic means for completely renewing all of the links in all of the chains periodically. This would require a breakthrough in prevention. Such a breakthrough could multiply our present life span by a factor that is impossible to predict at this time (*33*, p. 398).

REFERENCES

1. Abbott, M., Murphy, E., Bolling, D., and Abbey, H., *Johns Hopkins Med. J.* **135**, 1–16 (1974).
2. Adams, G. F., *in* "Geriatric Medicine" (W. F. Anderson and T. G. Judge, eds.), pp. 135–145. Academic Press, New York, 1974.
3. Adams, R. D., Fischer, C. M., Haskim, S., Ojemann, R. G., and Sweet, W. H., *N. Engl. J. Med.* **273**, 117–126 (1965).

4. Adelman, R., *Exp. Gerontol.* **6,** 75–87 (1971).
5. Alexander, P., *in* "The Initial Effects of Ionizing Radiation on Cells" (R. J. C. Harris, ed.), pp. 3–19. Academic Press, New York, 1963.
6. Anderson, J. W., Herman, R. H., and Zakim, D., *Am. J. Clin. Nutr.* **26,** 600–607 (1973).
7. Andrew, W., "The Anatomy of Aging in Man and Animals," pp. 218–237. Grune and Stratton, New York, 1971.
8. Aslan, A., *in* "Theoretical Aspects of Aging" (M. Rockstein, M. L. Sussman, and J. Chesky, eds.), pp. 145–156. Academic Press, New York (1974).
9. Aslan, A., *Conf. Aging, 11th* (1978).
10. Austin, J. H., Rinehart, R., Williamson, T., Burcar, P., Russ, K., Nikaĭdo, T., and LaFrance, M. *Prog. Brain Res.* **40,** 485–495 (1973).
11. Bakay, L., *AMA Arch. Neurol. Psychiatry* **73,** 2–12 (1955).
12. Barnes, B. O., and Barnes, C. W., *in* "Heart Attack Rareness in Thyroid Treated Patients." Thomas, Springfield, Illinois, 1972.
13. Barnes, B. O., and Barnes, C. W., *in* "Solved: The Riddle of Heart Attacks," p. 48. Robinson Press, Inc., Fort Collins, Colorado, 1976.
14. Barnes, M. J., and Kodicek, E., *Vitam. Horm. (N.Y.)* **30,** 439–442 (1972).
15. Barr, *in* "Cerebral Vascular Diseases" (I. S. Wright and E. H. Lucky, eds.), p. 71. Grune and Stratton, New York, 1955.
16. Bearn, J. G., *Br. J. Exp. Pathol.* **49,** 136–144 (1968).
17. Beebe, C. W., Rogers, J. S., and Happich, W. F., *J. Am. Leather Chem. Assoc.* **44,** 204 (1949); **46,** 659 (1951).
18. Bellows, J. G., and Bellows, R. T., *Ann. Ophthalmol.* **8,** 129–134 (1976).
19. Berlin, M., and Wallace, R. B., *Exp. Aging Res.* **2,** 125–164 (1976).
20. Bjorksten, J., *Cm. Ind.* **48,** 746–751 (1941).
21. Bjorksten, J., *Adv. Protein Chem.* **6,** 357 (1951).
22. Bjorksten, J., *Proc. Soc. Exp. Biol. Med.* **81,** 350–353 (1952).
23. Bjorksten, J., *Tex. Rep. Bio. Med.* **18,** 347–357 (1960).
24. Bjorksten, J., *Chemist (Washington, D.C.)* **38,** 11–13 (1961).
25. Bjorksten, J., *Gerontologia* **8,** 179–192 (1963).
26. Bjorksten, J., *J. Am. Geriatr. Soc.* **16,** 411–413 (1968).
27. Bjorksten, J. *Suom. Kemistsiseuran Tied.* **80,** 28–32 (1971).
28. Bjorksten, J., *Relevant Sci.* **1,** 29–32 (1971).
29. Bjorksten, J., *in* "Theoretical Aspects of Aging" (M. Rockstein, M. L. Sussman, and J. Chesky, eds.), pp. 43–59. Academic Press, New York, 1974.
30. Bjorksten, J., *Rejuvenation* **3,** 37–52 (1975).
31. Bjorksten, J., *Comp. Ther.* **2,** 68 (1976).
32. Bjorksten, J., *Rejuvenation* **4,** 4 (1976).
33. Bjorksten, J., *J. Am. Geriatr. Soc.* **25,** 396–399 (1977).
34. Bjorksten, J., *in* "Protein Crosslinking - B. Nutritional and Medical Consequences" (M. Friedman, ed.), pp. 579–602. Plenum, New York, 1977.
35. Bjorksten, J., *Rejuvenation* **5,** 2 (1977).
36. Bjorksten, J., *Comp. Ther.* **4,** 44–52 (1978).
37. Bjorksten, J., *in* "Longevity: A Quest," pp. 242–244. Bjorksten Research Foundation, Madison, Wisconsin, 1981.
38. Bjorksten, J., Acharya, P. V. N., Ashman, S., and Wetlaufer, D. B., *J. Am. Geriatr. Soc.* **19,** 561–574 (1971).
39. Bjorksten, J., and Andrews, F., *J. Am. Geriatr. Soc.* **8,** 632–637 (1960).
40. Bjorksten, J., and Ashman, S., *J. Am. Geriatr. Soc.* **18,** 115–123 (1970).

41. Bjorksten, J., and Collbring, T., *Proc. Sci. Sect. Toilet Goods Assoc.* **42,** 1–5 (1964).
42. Bjorksten, J., and Gottlieb, H., *Suom. Kemistsiseuran Tied.* **63,** 74–77 (1954).
43. Bjorksten, J., Schenk, R. U., and Yaeger, L. L., *Rejuvenation* **6,** 73–75 (1978).
44. Bjorksten, J., Weyer, E. R., and Ashman, S. M., *Suom. Kemistsiseuran Tied.* **80,** 70–87 (1971).
45. Blichert-Toft, M., *Acta Endocrinol. (Copenhagen)* Suppl., **78,** 195 (1975).
46. Blondon, G. A., and Green, D. E., *Chem. Eng. News* **53,** 26–42 (1975).
47. Blumberg, W. E., Goldstein, M., Lauber, E., and Peisach, J., *Biochim. Biophys. Acta* **99,** 187–190 (1965).
48. Bondareff, W., *in* "Development and Aging in the Nervous System" (M. Rockstein and M. L. Sussman, eds.), pp. 1–9. Academic Press, New York, 1973.
49. Bondareff, W., and Narotzky, R., *Science* **176,** 1135 (1972).
50. Bonner, J., Shih, T. Y., Bishop, D. W., and Van Buygen, J. T., *in* "Metabolism" (P. L. Altman and D. S. Dittmer, eds.), pp. 425–426. Fed. Am. Soc. Exp. Biol., Bethesda, Maryland, 1968.
51. Bornstein, M. B., *U.S. Public Health Serv. Publ.* **1378,** 175–185 (1965).
52. Bornstein, M. B., and Raine, C. S., *Lab. Invest.* **23,** 536 (1970).
53. Breen, M., Weinstein, H. G., Johnson, R. L., Veis, A., and Marshall, R. T., *Biochim. Biophys. Acat* **201,** 54 (1970).
54. Brody, H., *in* "Development and Aging in the Nervous System" (M. Rockstein and M. L. Sussman, eds.), pp. 121–134. Acedemic Press, New York, 1973.
55. Brunzell, J. D., Lerner, R. L., Hazzard, W. R., Porte, D., Jr., and Bierman, E. I., *N. Engl. J. Med.* **284,** 521–524 (1971).
56. Buber, V., *Schweiz. Med. Wochenschr.* **98,** 711–712 (1968).
57. Bucci, L., and Saunders, J. C., *N. Neuropsychiatry* **5,** 276 (1960).
58. Burns-Biotechnical Laboratory, Division of Chromalloy Pharmaceutical, Inc., "Selenium-Tocopherol, A Biomedical Summary for Veterinarians." Oakland, California, 1977.
59. Canivenc, R., *Arch. Anat., Strasbourg* **34,** 105 (1951).
60. Cannon, P. J., Ames, R. R., and Laragh, J. H., *JAMA, J. Am. Med. Assoc.* **185,** 854–863 (1963).
61. Castelli, W. P. *et al., Coron. Heart Dis.* **55,** 767 (1977).
62. Clarke, N. E., Sr., Clarke, N. E., Jr., and Mosher, R. E., *Am. J. Med. Sci.* **232,** 654–666 (1956).
63. Clarke, N. E., Sr., Clarke, N. E., Jr., and Mosher, R. E., *Am. J. Med. Sci.* **239,** 732–744 (1960).
64. Comfort, A., *Lancet,* 1411–1415 (1969).
65. Constantinides, P., *in* "Atherogenesis" (T. Shimamoto and F. Numano, eds.), pp. 99–102. Excerpta Medi. Found., Amsterdam, 1969.
66. Constantinides, P., and Wiggers, K. D., *Am. J. Pathol.* **70,** 81a (1973).
67. Cooper, A. J., Magnus, R. V., and Ross, M. J., *Lancet* **1,** 527–529 (1964).
68. Cragg, *J. Anat.* **101,** 639–654 (1967).
69. Crain, S. M., *Int. Rev. Neurobiol.* **9,** 1 (1966).
70. Crapper, D. R., *in* "Frontiers in Neurology and Neuroscience Research 1974" (P. Seeman and G. M. Brown, eds.), pp. 97–111. Univ. of Toronto Press, Toronto, 1974.
71. Crapper, D. R., *in* "Neurobiology of Aging" (R. D. Terry and S. Gershon, eds.), pp. 405–432. Raven, New York, 1976.
72. Crapper, D. R., *Int. Conf. Gerontol., 1978* (1978).
73. Crapper, D. R., and DeBoni, U., *Symp. Aging Brain Senile Dementia, 1976* (1976).

74. Crapper, D. R., and DeBoni, U., *in* "Symposium on the Aging Brain and Senile Dementia" (K. Nandy and I. Sherwin, eds.). Plenum. New York, 1977.
75. Crapper, D. R., Krishnan, S. S., and Dalton, A. J., *Science* **180,** 511–513 (1973).
76. Crapper, D. R., Krishnan, S. S., DeBoni, U., and Tomko, G. J., *Arch. Neurol. (Chicago)* **32,** 356 (abstr.).
77. Crapper, D. R., Quittkat, S., and Krishnan, S. S., Brain **99,** 67–79 (1976).
78. Cristofalo, V. J., *in* "Aging in Cell and Tissue Culture" (E. Holeckŏvá and V. J. Cristofalo, eds.), pp. 83–119. Plenum, New York, 1976.
79. Curran, G. L., *J. Biol. Chem.* **210,** 765 (1954).
80. Curran, G. L., and Azarnoff, D. L., *Arch. Intern. Med.* **101,** 685 (1958).
81. Curran, G. L., Azarnoff, D. L., and Bolinger, R. E., *J. Clin. Invest.* **38,** 1251 (1959).
82. Data Sheet for Physicians, "Selenium-Tocopherol in the Human." Burns-Biotech. Lab., Div. of Chromalloy America Corp., St. Louis, Missouri and Oakland, California.
83. Daubenmerkl, W. A., *Acta Pharmacol. Toxicol.* **7,** 153 (1951).
84. Daubenmerkl, W. A., *Acta Pharmacol. Toxicol.* **9,** 1 (1953).
85. DeBoni, U., Otvos, A., Scott, J. W., and Crapper, D. R., *Acta Neuropathol.* (1976).
86. DeBoni, U., Scott, J. W., and Crapper, D. R., *Histochemistry* **40,** 31–37 (1974).
87. Denckla, W. C., *J. Clin. Invest.* **53,** 572–581 (1974).
88. Dilman, V. M., *Lancet* **1,** 1211–1219 (1971).
89. Dische, A., Borenfreund, E., and Zelmenis, G., *AMA Arch. Ophthalmol.* [N.S.] **55,** 471–483 (1956).
90. Dubina, T. L., and Razumovich, A. N., "Introduction to Experimental Gerontology," Chapter 5. Izd. Nauka i Tehknika, Minsk, 1975.
91. Dubos, R. J., *in* "The Bacterial Cell." Harvard Univ. Press, Cambridge, Massachusetts, 1945.
92. Duckett, S. and Galle, P., *Electron Microsc., Proc. Eur. Reg. Conf., 6th, 1976,* pp. 210–211 (1976).
93. Economos, A., "Refutation of Gompertz's Law: A Paradigm for Mortality Kinetics Pertaining to Metazoan Animals and Manufactured Products." Technology, Inc., Life Sci. Div., Mountain View, California, 1977.
94. Everitt, A. V., *Gerontologia* **2,** 21 (1958).
95. Everitt, A. V., *Exp. Gerontol.* **8,** 265–277 (1973).
96. Everitt, A. V., and Burgess, J. A., "Hypothalamus, Pituitary and Aging" (A. Everitt, ed.), pp. 35–42. Thomas, Springfield, Illinois, 1976.
97. Evers, H. R., Judgment, Civil Action No. 89-93-N, U.S. District Court for the Middle District of Alabama, Northern Division; United States of America, Plaintiff, vs. H. Ray Evers, M. D., Defendant, June 27, 1978.
98. Fabris, N., Pierpaoli, W., and Sorkin, E., *Nature (London)* **240,** 557–559 (1972).
99. Feder, J., Keay, L., Garrett, L. R., Cirulis, N., Moseley, M. H., and Wildi, B. S., *Biochim. Biophys. Acta* **251,** 74–79 (1971).
100. Feldman, M. L., *in* "The Aging Brain and Senile Dementia" (K. Nandy and I. Sherwin, eds.), pp. 23–38. Plenum, New York, 1976.
101. Fernández-Crus, A., *in* "Basic Applications and Clinical Uses of Hypothalamic Hormones" (A. L. Charro Salgado. R. Fernández-Durango, and J. G. López del Campo, eds.). Excerpta Medica, Amsterdam, 1976.
102. Fernandez-Moran, H., *in* "Metabolism of the Nervous System" (E. Richter, ed.), pp. 1–34. Pergamon, Oxford, 1956.
103. Finch, C. E., *in* "Proceedings of the First European Conference on Cell Differentiation" (R. Harris and D. Viza, eds.), pp. 123–126. Munksgaard, Copenhagen, 1971.
104. Finch, C. E., *Brain Res.* **52,** 261–276 (1973).

105. Finch, C. E., *Q. Rev. Biol.* **51,** 49–83 (1976).
106. Fishman, J. A., Graeme, B. R., and Karnovsky, M. J., *Lab. Invest.* **32,** 339 (1975).
107. Friedman, F. J., Moore, S., and Singal, D. P., *Lab. Invest.* **30,** 404 (1975).
108. Friedman, S. M., and Friedman, C. L., *Gerontologia* **1,** 107 (1957).
109. Frost, D. V., and Ingvoldstad, D., *Chem. Scr.* **8A,** 96–107 (1975).
110. Frost, D. V., and Lish, P. M., *Annu. Rev. Pharmacol.* **15,** 259–284 (1975).
111. Frost, D. V., and Van Poucke, R., *Trace Subst. Environ. Health* **6,** 259–266 (1973).
112. Fry, J. M., Lehrer, G. M., and Bornstein, M. B., *Science* **175,** 192 (1972).
113. Fry, J. M., Lehrer, G. M., and Bornstein, M. B., *Trans. Am. Soc. Neurochem.* **3,** 79 (1972).
114. Gallagher, T. F., *Annu. Meet. Am. Geriatr. Soc., 1954,* 1954.
115. Garn, S. M., *in* "The Physiology and Pathology of Human Aging" (R. Goldman and M. Rockstein, eds.). Academic Press, New York, 1975.
116. Gerlach, J. L., and McKewen, B. S., *Science* **175,** 1133–1136 (1972).
117. Glomset, J. A., *Biochim. Biophys. Acta* **65,** 128 (1962).
118. Goldstein, J., Hazzard, W. R. *et al., J. Clin. Invest.* **52,** 1533–1543 (1973).
119. Goldstein, L. S., Reynolds, C. P., and Perez-Polo, J. R., *Neurochem. Res.* **3,** 175–183 (1978).
120. Goldstein, S., *N. Engl. J. Med.* **285,** 1120–1129 (1971).
121. Gomez-Pan, A., Hall, B., Besser, G. M., Mortimer, C. H., Schally, A. V., Kasten, A., and Coy, D. H., *Proc. Int. Symp. Basic Appl. Clin. Uses Hypothalamic Horm., 1st, 1975,* pp. 313–318, 1975.
122. Good, R. A., Presidential address to the American Association of Immunologists, delivered in Anaheim, California, April 13, 1976. Runestones in immunology: Inscriptions to journeys of discovery and analysis. *J. Immunol.* **117,** 5, Part 1 (1976).
123. Gortner, W. A., *J. Gerontol.* **9,** 251–261 (1954).
124. Gottlieb, H. L., Bjorksten Research Foundation, Madison, Wisconsin (personal communication, 1954).
125. Green, D. E., *Proc. Natl. Acad. Sci. U.S.A.* (1978).
126. Green, D. E., and Blondin, G. A., *BioScience* **28,** 18–24 (1978).
127. Gregerman, R. I., and Bierman, E. L., *in* "Textbook of Endocrinology" (R. H. Williams, ed.), 5th ed., pp. 1059–1070. Saunders, Philadelphia, Pennsylvania, 1974.
128. Grimmer, G., and Makaroo, M. A., Quoted from "Tissue Therapy and Tissue Preparations," according to V. P. Filatov, c/o Medexport, Moscow (undated, p. 14).
129. Gustavson, K. H., *in* "The Chemistry of Tanning Processes" (K. H. Gustanson, ed.), pp. 295–307, 360. Academic Press, New York, 1956.
130. Gustavson, K. H., "The Chemistry and Reactivity of Collagen." Academic Press, New York, 1956.
131. Hamilton, T. H., *Science* **161,** 649–661 (1968).
132. Hart, R. W., and Setlow, R. B., *Proc. Natl. Acad. Sci. U.S.A.* **71,** 2169–2173 (1974).
133. Hart, R. W., and Setlow, R. B., *in* "Molecular Mechanisms for Repair of DNA" (P. C. Hanawalt and R. B. Setlow, eds.). Plenum, New York, 1975.
134. Hart, R. W., and Setlow, R. B., *Mech. Ageing Dev.* **5,** 67–77 (1976).
135. Hawker, R. W., "Endocrinology: With Aspects of Maternal, Fetal and Neonatal Physiology," pp. 140–144. Churchill-Livingstone, Edinburgh and London, 1978.
136. Hayflick, L., *Exp. Cell Res.* **37,** 614–636 (1965).
137. Hayflick, L., and Moorhead, P. S., *Exp. Cell Res.* **25,** 585–621 (1961).
138. Helmreich, E., and Cori, F. C., *Adv. Enzyme Regul.* **3,** 91 (1965).
139. Herp, A., and Rickards, T., *in* "The Carbohydrates" (W. Pigman and D. Horton, eds.), Vol. 1B. Academic Press, New York, 1977.

140. Herp, A., Rickards, T., Matsumura, G., Jakosalem, L. B., and Pigman, W., *Carbohydr. Res.* **4,** 63–71 (1967).
141. Hickman, K. C. D., *in* "Josiah Macy Conference on Physiologic Antioxidants," p. 122. Josiah Macy, Jr. Found., New York, 1949.
142. Hickman, K. C. D., and Harris, P. L., *Adv. Enzymol.* **6,** 469–524 (1946).
143. Hoffman, J. L., *Nature (London)* **251,** 260 (1974).
144. Holmquist, B., *Biochemistry* **16,** 4591–4594 (1977).
145. Housley, T., and Tanzer, M. L., *Biochem. Biophys. Res. Commun.* **67,** 2 (1975).
146. Housley, T., Tanzer, M. L., Henson, E., and Gallop, P. M., *Biochem. Biophys. Res. Commun.* **67,** 824–829 (1975).
147. Hove, E. L., Hickman, K., and Harris, P. L., *Arch. Biochem.* **8,** 395–404 (1945).
148. Hovytzjez, M. N., Korolev, P. N., and Bulzakova, V. G., *Microbiology* **47,** 442–443 (*Engl. Transl.*) (1978).
149. Hurlbert, R. B., *in* "Encyclopedia of Biochemistry" (R. J. Williams and E. M. Lansford, Jr., eds.), p. 713. Van Nostrand-Reinhold, Princeton, New Jersey, 1967.
150. Ingold, K. U., *Chem. Rev.* **61,** 563 (1961).
151. Isaacs, A., U.S. Patent 151,951 (1915).
152. Islam, M. A., and Blanshard, J. M. V., *J. Dairy Res.* **40,** 427–440 (1973).
153. Jackson, I., and Reichlin, S., *Endocrinology* **95,** 854–862 (1974).
154. Jöbsis, F. F., *Science* **198,** 1264–1267 (1977).
155. Jöbsis, F. F., Keizer, J. H., LaManna, J. C., and Rosenthal, M., *J. Appl. Physiol.: Respir., Environ. Exercise Physiol.* **43,** 858–872 (1977).
156. Jöbsis, F. F., and Rosenthal, M., *Ciba Found. Symp.* [N.S.] **56,** 129–148 (1978).
157. Jöbsis, F. F., and Rosenthal, M., *Ciba Found. Symp.* [N.S.] **56,** 149–169 (1978).
158. Johnson, P. E., *in* "Biological Data Book III" (P. L. Altman and D. S. Dittmer, eds.), pp. 1446–1450. Fed. Am. Soc. Exp. Biol., Bethesda, Maryland, 1974.
159. Jones, N. F., and Tigh, J. R., *in* "Amyloidosis 1974" (O. Wegelius and A. Pasternack, eds.), p. 565. Academic Press, New York, 1976.
160. Kääntee, E., Kurkela, P., and Korhonen, I., *Suom. Elainlaakaril.* **84,** 393–397 (1978).
161. Kamimura, M., Takahashi, S., and Henmi, I., *Sapporo Med. J.* **21,** 71–77 (1962).
162. Kaunitz, H., *Ann. N.Y. Acad. Sci.* **52,** Artic. 3, 200–201 (1949).
163. Kehoe, R. A., Cholak, J., and Story, R. V., *J. Nutr.* **19,** 579–592 (1940).
164. King, E. J., *Biochem. J.* **33,** 944–954 (1939).
165. Kirkwood, T. B. L., and Holliday, R., *Radiat. Biol. Cent. Int. Symp., 2nd, 1978,* 1978.
166. Knobil, E., and Briggs, N. F., *Endocrinology* **57,** 147 (1955).
167. Komarov, L. V., *Rejuvenation* **6,** 72 (1978).
168. Konu, V., Thesis. *Acta Med. Scand., Suppl.* **604** (1977).
169. Kurkela, P., *IFAP 22nd Gen. Conf., 1977, Eripainos, Ympäristö ja Terveys-lehti* **6,** (1977).
170. Kurkela, P., and Kääntee, E., *Suom. Eläinlääkaril.* (1977).
170a. Kurkela, P., and Kääntee, E., *Zentralbl. Veterinaer Med. Reihe B* **26,** 168–173 (1979).
171. Labisky, R. F., Mann, S. H., and Lord, R. D., Jr., *J. Wildl. Manage.* **33,** 2 (1969).
172. Lagercrantz, C., *Acta Chem. Scand.* **18,** 562 (1964).
173. Lagercrantz, C., and Yhlard, M., *Acta Chem. Scand.* **16,** 1807–1809 (1962).
174. Langham, W. H., *Health Phys.* **2,** 172–185 (1959).
175. Lansford, E. M., Jr., *in* "Biology Data Book III" (P. L. Altman and D. S. Dittmer, eds.), p. 1568. Fed. Am. Soc. Exp. Biol., Bethesda, Maryland, 1974.
176. Lansing, A. I. A., *J. Appl. Zool.* **91,** 195–211 (1942).

177. Lansing, A. I. A., *Ann. N.Y. Acad. Sci.* **57,** 455–464 (1947).
178. Lansing, A. I. A., *J. Gerontol.* **2,** 228–239 (1954).
179. Lapresle, J., Duckett, S., Galle, P., and Cartier, L., *Biol.* **169,** 282 (1975).
180. Larner, J., *Trans. N.Y. Acad. Sci.* **11,** 192 (1966).
181. LeCompte, H., *Z. Alternsforsch.* **18,** 2 (1965).
182. Lehrer, G. M., and Bornstein, M. B., *Trans. Am. Neurol. Assoc.* **93,** 174 (1968).
183. Lehrer, G. M., Bornstein, M. B., Weiss, C., Furman, M., and Lichtman, C., *Exp. Neurol.* **27,** 410–425 (1970).
184. Lehrer, G. M., Bornstein, M. B., Weiss, C., and Silides, D. J., *Exp. Neurol.* **26,** 595–606 (1970).
185. Lehrer, G. M., Fry, J. M., and Bornstein, M. B., *2nd Annu. Meet. Soc. Neurosci.* p. 181, 1972.
186. Leloir, L. F., Cardini, C. E., and Cabib, E., *in* "Comparative Biochemistry" (M. Florken and H. S. Mason, eds.), Vol. 2, p. 97. Academic Press, New York, 1960.
187. Leve-Montcalcini, R., and Calissano, P., *Sci. Am.* **240,** 44–53, 150 (1979).
188. Lewin, S. "Vitamin C: Its Molecular Biology and Medical Potential." Academic Press, New York, 1976.
189. Loh, H. W., Odomosu, A., and Wilson, C. W. M., *Clin. Pharmacol. Ther.* **16,** 390–408 (1974).
190. Loh, H. W., and Wilson, C. W. M., *Br. Med. J.* **3,** 733–735 (1971).
191. Loh, H. W., and Wilson, C. W. M., *Int. J. Vitam. Nutr. Res.* **41,** 259–267 (1971).
192. Loh, H. W., and Wilson, C. W. M., *Lancet,* 317 (1973).
193. Lowry, O. H., and Passonneau, J. V., "A Flexible System of Enzymatic Analysis." Academic Press, New York, 1972.
194. Ludke, J. L., and Henny, C. J., *J. Wildl. Manage.* **38,** 138 (1974).
195. McCay, C. M., Sperling, G., and Barnes, K. L., *Arch. Biochem.* **2,** 474 (1943).
196. Machlin, L. J., *J. Am. Oil Chem. Soc.* **40,** 386 (1963).
197. Machlin, L. J., Gordon, R. S., Marr, J. E., and Pope, C. W., *J. Nutr.* **76,** 284 (1962).
198. Machlin, L. J., Gordon, R. S., and Miesky, K. H., *J. Nutr.* **67,** 333 (1959).
199. Macierira-Coelho, A., *Experientia* **22,** 390–391 (1966).
200. Maker, H. S., Lehrer, G. M., Weissbarth, S., and Bornstein, M. B., *Prog. Brain Res.* **44,** 189 (1972).
201. Makinodan, T., *in* "Biology of Aging" (C. E. Finch and L. Hayflick, eds.), p. 399. Van Nostrand-Reinhold, Princeton, New Jersey, 1977.
202. Mann, A. M. and Laing, W. A. R., *Can. Med. Assoc. J.* **89,** 1115–1118 (1963).
203. Matsumura, G., and Pigman, W., *Arch. Biochem. Biophys.* **110,** 526 (1965).
204. Mattock, G. L., *J. Chem. Soc.* pp. 4728–4735 (1965).
205. Mayer, G., and Canivenc, R., *C.R. Seances Soc. Biol. Ses Fil.* **144,** 410 (1950).
206. Meier-Ruge, W., Hunziker, O., Iwangoff, P., Reichlmeier, K., and Sandoz, P., *in* "Senile Dementia: A Biomedical Approach" (K. Nandy, ed.), Dev. Neurosci., Vol. 3. Elsevier/North-Holland Bimoedical Press, Amsterdam and New York, 1978.
207. Metropolitan Insurance Company, Statistics, 1951.
208. Meyer-Brunot, H. G., and Keberle, H., *Biochem. Pharmacol.* **16,** 527–535 (1967).
209. Miettinen, T., *in* "Medica Company Diabetes Symposium," pp. 37–41. Weilin and Göös, Tapiola, Finland, 1969.
210. Mills, C. A., *Blood* **4,** 153–159 (1949).
211. Moschocowitz, E., *JAMA, J. Am. Med. Assoc.* **143,** 861 (1950).
212. Muchnik, S. R., *Probl. Oncol.* (translation from Russian) **12,** 98–99 (1966).
213. Natelson, S. J., Pincus, J., and Lugovoy, J. K., *J. Clin. Invest.* **27,** 446–449 (1948).
214. National Health Federation Bulletin, pp. 27–28, October, 1978.

215. Newell, P. C., *Endeavour* [N. S.] **1,** 63–68 (1978).
216. Norwood, W. D., Fuqua, P. A., and Scudder, B. C., *Ind. Med.* **25,** 135–139 (1956).
217. Ogino, S., *Acta Soc. Ophthalmol. Jpn.* **56,** 1342 (1952).
218. Ogino, S., and Ichihara, T., *Am. J. Ophthalmol.* **43,** 754 (1957).
219. Orgel, L. E., *Proc. Natl. Acad. Sci. U.S.A.* **49,** 517–521 (1963).
220. Orr, C. W. M., *Biochemistry 6,* 3000 (1967).
221. Otin, C., and Alexa, G., *J. Soc. Leather Trades' Chem.* **22,** 339 (1938).
222. Papaikonomou, E., "Biocybernetics. Biosystems Analysis, and the Pituitary Adrenal System." Nooy's Drukkerij-Pormerend, Holland, 1974.
223. Parmeggiani, A., and Bowman, R. H., *Biochem. Biophys. Res. Commun.* **12,** 268 (1963).
224. Passonneau, J., and Lowry, O. H., *Biochem. Biophys. Res. Commun.* **7-8,** 10 (1962).
225. Passonneau, J., and Lowry, O. H., *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **21,** 87 (1962).
226. Passonneau, J., and Lowry, O. H., Phosphofructokinase and the control of the citric acid cycle. *Biochem. Biophys. Res. Commun.* **13,** 372–379 (1963).
227. Pierpaoli, W., and Sorkin, E., *Nature (London)* **215,** 834–837 (1967).
228. Pigman, W., Matsumura, G., and Herp, A., *Proc. Int. Congr. Rheol., 4th, 1963,* pp. 505–519 (1965).
229. Pigman, W., Rizvi, S., and Holley, H. L., *Arthritis Rheum.* **4,** 240–252 (1961).
230. Pontén, J., *in* "The Biology of Aging" (C. E. Finch and L. Hayflick, eds.), pp. 536–546 Van Nostrand-Reinhold, Princeton, New Jersey, 1977.
231. Pritikin, N., *Pap., 37th Annu. Asembly Am. Acad. Phys. Med. Rehab., 1975* pp. 1–12.
232. Procter, H. R., "Principles of Leather Manufacture," Spon, London, 1904.
233. Rabinowich, I. M., *Can. Med. Assoc. J.* **33,** 136–144 (1935).
234. Raff, J., *Sci. Am.* **234,** 30–39 (1976).
235. Ray, E. W., Averill, S. C., Lyons, W. R., and Johnson, R. E., *Endocrinology* **56,** 359–373 (1955).
236. Ri, K., *J. Biochem. (Tokyo)* **31,** 215–241 (1940).
237. Richheimer, S. L., and Robinson, A. B., "Degradation of Transferrin in the Presence of Ascorbic Acid and Oxygen," Contrib. No. 65. Linus Pauling Institute of Science and Medicine, Menlo Park, California, 1977.
238. Riddick, T. M., *in* "Control of Colloid Stability Through Zeta Potential," pp. 266–275. Livingston Publ. Co., Wynnewood, Pennsylvania, 1968.
239. Rinse, J. *Am. Lab. (Fairfield, Conn.)* **10,** 68–71 (1978).
240. Robinson, A. B., Irving, K., and McCrea, M., *Proc. Natl. Acad. Sci. U.S.A.* **70,** 2122 (1973).
241. Robyn, C., *Proc. Int. Symp. Basic Appl. Clin. Uses Hypothalamic Horm., 1st, 1975* pp. 319–342 (1975).
242. Rockstein, M., and Baker, G., *Mech. Ageing Dev.* **3,** 271–278 (1974).
243. Rockstein, M., and Miquel, J., *in* "The Physiology of Insecta" (M. Rockstein, ed.), Vol. I, pp. 371–471. Academic Press, New York, 1973.
244. Ross, L. L., Bornstein, M. B., and Lehrer, G. M., *J. Cell Biol.* **14,** 19 (1962).
245. Roth, G. S., *Endocrinology* **94,** 82–90 (1974).
246. Roth, G. S., *Brain Res.* **107,** 345–354 (1976).
247. Russell, W. O., *Am. J. Clin. Pathol.* **56,** 259 (1971).
248. Ryan, K. J., *J. Biol. Chem.* **234,** 268 (1959).
249. Salway, A. H. *J. Chem. Soc.* **109,** 138 (1916).
250. Samis, H. V., Baird, M. B., and Massie, H. R., *in* "Molecular Genetic Mechanics in Development and Aging" (M. Rockstein and G. T. Baker, III, eds.), pp, 113–143. Academic Press, New York, 1972.

251. Sansum, W. D., Blatherwick, N. R., and Bowden, R., *JAMA, J. Am. Med. Assoc.* **86,** 178 (1926).
252. Scheibel, M. E., and Scheibel, A. B., *Aging* **1,** 11–37 (1978).
253. Schenk, R. U., and Bjorksten, J., *Suom. Kemistiseuran Tied.* **82,** 26–46 (1973).
254. Schiefelbein, S., *Saturday Rev.* Aug. 7, 1978.
255. Seelert, K., and Schenk, G., *Arch. Pharm. (Weinheim, Ger.)* **299,** 757–762 (1966).
256. Shamberger, R. J., Rukovena, E., Longfield, A. K., Tytko, S. A., Deodhar, S., and Willis, C. W., *JNCI, J. Natl. Cancer Inst.* **50,** 863–870 (1973).
257. Simms, H., *J. Gerontol.* **1,** 13–26 (1946).
258. Sincock, A. M., *J. Gerontol.* **29,** 514–517 (1974).
259. Sinex, M. F., *in* "Handbook of The Biology of Aging" (C. E. Finch and L. Hayflick, eds.), pp. 37–62. Van Nostrand-Reinhold, Princeton, New Jersey, 1977.
260. Skanse, B., and Sundblad, L., *Acta Physiol. Scand.* **6,** 37–51 (1943).
261. Smith, P. E., *Endocrinology* **56,** 271–284 (1955).
262. Solovieva, V. P., "Comparative Assessment of the Efficiency of Tissue Preparations in Oxygen Deficiency," p. 20. Theses of the Scientific Industrial Conference on the use of tissue preparation, after V. P. Filatov, in cattle breeding and veterinary practice, Kiev, 1964.
263. Solovieva, V. P., and Lipovetskaya, E. M., *in* "Tissue Therapy and Tissue Preparations According to V. P. Filatov" (S. R. Muchnik and V. P. Solovieva, ed's.), pp. 30–31, 1967.
264. Solovieva, V. P., and Stepanok, E. G., "The Effect of Placental Suspension on the Course of Experimental Myocarditis," p. 87. Materials of a Scientific Conference devoted to V. P. Filatov, 90th anniversary, Kiev, 1965.
265. Starr, P., *Annu. Meet. Am. Geriatr. Soc.*, 1954.
266. Steinach, E., "Verjüngung durch experimentelle Neibelebund der alternden Pubertätsdrüse." Springer-Verlag, Berlin and New York, 1920.
267. Still, Joseph W., *Med. Ann. D. C.* **25,** 199 (1956).
268. Sundholm, F., Visapää, A., and Bjorksten, J., *Lipids* **13,** 755–757 (1978).
269. Sylvia, A. L., LaManna, J. C., Rosenthal, M., and Jöbsis, F. F., *J. Phramacol. Exp. Ther.* **201,** 117–125 (1977).
270. Sylvia, A. L., and Rosenthal, M., *Brain Res.* **146,** 109–122 (1977).
271. Szent-Györgyi, A., *Exec. Health* **14,** 8 (1978).
272. Tablert, G., *Am. J. Obstet. Gynecol.* **102,** 451–457 (1968).
273. Tanzer, M. L., Cross-linking of collagen. *Science* **180,** 561–566 (1973).
274. Tappel, A. L., *Arch. Biochem. Biophys.* **54,** 266 (1955).
275. Tappel, A. L., Fletcher, A., and Deamer, D., *J. Gerontol.* **28,** 419 (Fig. 4) (1973).
276. Timiras, P. S., "Developmental Physiology and Aging," pp. 79–80. Macmillan, New York, 1972.
277. Tonks, C. M., and Livingston, D., *Lancet* **1,** 1323–1324 (1963).
278. Tsai, S. P., and Lee, E. S., *Tex. Rep. Biol. Med.* **36** (1978).
279. Tyler, A., *Biol. Bull. (Woods Hoke, Mass.)* **104,** 224–239 (1953).
280. Tyler, A., *in* "Aging and Levels of Biochemical Organization" (A. M. Bruder and G. A. Sacher, eds.), Sect. II, Part II, pp. 50–86. Univ. of Chicago Press, Chicago, Illinois, 1965.
281. Tyler, F. H. *Annu. Meet. Am. Geriatr. Soc., San Francisco,* 1954.
281a. U.S. Bureau of Vital Statistics, Washington, D.C., 1976.
282. Uyama, Y., and Ogino, S., *Med. J. Osaka Univ.* **6,** 813 (1956).
283. Uyama, Y., Ogino, S., and Ichihara, T. (1954). *Acta Soc. Ophthalmol. Jpn.* **58,** 1100 (1954); *Am. J. Ophthalmol.* **39,** 125 (1953).

284. Van Gent, C. M., *Chem. Weekbl.*, 614 (1977).
285. von Buttlar-Brentano, *J. Hirnforsch.* **1,** 337 (1954).
286. von Ritzel, *Helv. Med. Acta* **28,** 63 (1961).
287. Voronoff, S. C., "Etude des moyens de relever l'energie vitale et der prolonger la vie." Grasset, Paris, 1920.
288. Voronoff, S. C., "Testicular Grafting from Ape to Man; Operative Technique, Physiological Manifestations, Histological Evolution, Statistics." Brentano, Ltd., London, 1929.
289. Walford, R. L., *J. Gerontol.* **17,** 281 (1962).
290. Walford, R. L., "The Immunologic Theory of Aging." Williams and Wilkins, Baltimore, Maryland, 1969.
291. Watson, J. D., "The Molecular Biology of the Gene." Benjamin, New York, 1965.
292. Wegelius, O., *in* "Amyloidosis 1974" (O. Wegelius and A. Pasternack, eds.), pp. 1–16. Academic Press, New York, 1976.
293. Whedon, G. D., *in* "Hormones and the Aging Process" (E. T. Engle and G. Pincus, eds.), pp. 221–239. Academic Press, New York, 1956.
294. White, A., Handler, P., and Smith, E. L. *in* "Principles of Biochemistry," pp. 719–727. McGraw-Hill, New York, 1973.
295. Williams, G. C., *Evolution* **11,** 398–411 (1957).
296. Williams, G. C., *in* "The Biology of Aging" (B. L. Strehler, ed.), pp. 332–337. Washington, D.C., 1960.
297. Wilson, C. W., and Petrakis, N. L., *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **35,** 578 (abstr.) (1976).
298. Wilson, C. W. M., *in* "The Practitioner," p. 487. London, 1974.
299. Wilson, C. W. M., and Loh, H. S., *Acta Allergol.* **24,** 367–380 (1969).
300. Wilson, C. W. M., and Loh, H. S., *Lancet*, 638–641 (1973).
301. Wilson, C. W. M., Loh, H. S., and Odomosu, A., *Clin. Pharmacol. Ther.* **16,** 390–408 (1974).
302. Wiśniewski, H. M., and Terry, R. D., *Prog. Brain Res.* **40,** 167–186 (1973).
303. Wolinsky, H., *Cardiovasc. Med.* **1,** 41–54 (1976).
303a. Worthington Biochemical Catalog. *Bacillus cereus* Low Molecular Protease "Microprotease," p. 26. Worthington Biochemical Corp., Freehold, New Jersey, 1975.
304. Yudkin, J., *Lancet* **2,** 4 (1964).
305. Zinsser, H., Bjorksten, J., Bruck, E. M., Baker, R. F., Kaeburn, L., Kinnear, J., Cohen, A., Andrews, F., Sarfati, I., and Light, I., *in* "Medical and Clinical Aspects of Aging" (H. T. Blumenthal, ed.), pp. 460–483. Columbia Univ. Press, New York, 1962.

7

Quantitation and Interpretation of Serum Drug Concentrations

C. E. PIPPENGER

CLINICAL BIOCHEMISTRY
Contemporary Theories and Techniques, Vol. 2

ISBN 0-12-657102-3

I. INTRODUCTION

Both physicians and pharmacologists have a continuing interest in ascertaining why a fixed drug dosage is therapeutically effective in some individuals but not in others. For centuries appropriate drug dosage regimens were established clinically by trial and error. Modern analytical technology has provided new insights and approaches to patient therapy. We now possess a better understanding of the relationships between a given drug concentration and its pharmacological effect. History will confirm that one of the greatest advances in pharmacology was the development of a simple, rapid analytical technique for the quantitation of drug concentratinions in biological fluids. The ability to correlate drug concentrations in serum, and, by inference, tissue concentrations with the clinical effect observed following drug administration provides new approaches to all aspects of therapeutics. Investigators have established that the desired effect occurs only above a specific plasma concentration, and that there is an optimal plasma concentration range over which drug therapy is most often successful. Above this optimal range, patients can be expected to experience undesirable drug side effects.

The value of therapeutic drug monitoring (TDM) as an aid to rational drug therapy in patients with various diseases is firmly established. Serum drug concentrations of numerous compounds, including antiepileptic, antibiotic, antiarrhythmic, antiasthmatic, antineoplastic, and antidepressant drugs are now measured routinely in order to establish an optimal therapeutic regimen for an individual patient at a given point in time. For the first time, the physician who monitors a patient's serum drug concentration is in a position to know why a patient exhibits an inappropriate response to a particular drug dosage or experiences undesirable side effects to a standard therapeutic drug dose (*1,2*).

Without question, TDM has significantly improved patient care. For example, a recent study by Reynolds *et al.* (*3*) has demonstrated that single drug therapy with phenytoin, in conjunction with close monitoring of serum phenytoin concentrations, is effective in many epileptic patients without significant side effects. This contrasts sharply with previously prescribed drug regimens for the management of the epilepses, which almost invariably involve the administration of at least two or three antiepi-

leptic drugs to produce the desired anticonvulsant effect. Sherwin and his colleagues have (*4*) clearly demonstrated that it is possible to significantly improve seizure control in epileptic children through routine monitoring. The monitoring of ethosuximide concentrations and appropriate dosage adjustments resulted in an increased percentage of patients with complete seizure control as compared to 47% seizure control prior to the availability of drug monitoring services. During the past decade, the rapid development of improved analytical techniques, particularly with respect to sensitivity (the capability to determine drug concentrations in small specimen volumes), has provided the tools which were previously unavailable for investigating drug disposition in the pediatric population. Consequently, in the past 6 years, clinical pharmacologists and pediatricians have become increasingly aware of the importance of therapeutic drug monitoring in children (*6,7*). Unfortunately, detailed information on age-related differences in drug utilization is still unavailable for many drugs commonly prescribed for children. This is particularly true with respect to information relating to drug disposition in children between the ages of 1 and 15. There are extensive studies available correlating age-related drug disposition patterns in neonates and adults.

Two recent monographs related specifically to pediatric pharmacology by Morselli (*7*) and Mirkin (*8*) are currently available. Extensive reviews on adult pharmacology are available also. Particularly valuable to both the clinician and the clinical chemist are the works by Avery (*9*), Melmon and Morelli (*10*), and Gilman *et. al.* (*11*). There are no monographs devoted to geriatric pharmacology currently on the market.

II. CLINICAL PHARMACOLOGY—AN OVERVIEW

The purpose of this brief overview of applied clinical pharmacology is to examine factors which influence drug disposition in newborns, children, adults, and geriatric patients. Those utilizing therapeutic drug monitoring in their practice should always be alert to the special therapeutic problems that arise as a consequence of age-related differences in drug disposition. Because of constant and sometimes surprisingly rapid physiological changes that occur at various intervals during growth and development, it is imperative that both the clinician and laboratory engaged in routine monitoring be alert for changes in drug disposition patterns in children of various ages. If altered rates of drug disposition are rapidly identified, appropriate adjustment of the child's maintenance dosage by the pediatrician can prevent the undesirable physiological and sociological sequelae that often are associated with inappropriate drug therapy in chil-

dren. Of equal importance is early recognition of decreasing drug disposition patterns in the elderly (*12*). Prompt recognition and appropriate adjustment of the medication regimen can prevent many untoward drug reactions which ultimately decrease the quality of life in the elderly.

Drug therapy is usually aimed at abolishing an acute or chronic pathologic state. Much more difficult to evaluate and treat, however, are diseases with only occasional clinical manifestations, e.g., hypertension, certain cardiac arrhythmias, asthma, and the epilepses. For centuries, optimal drug dosages were established by trial and error. A patient was given a fixed quantity of drug and, if the desired effect did not result, the dosage was increased until signs of toxicity appeared at which point the dosage was reduced. If the expected response still did not occur, a second drug was prescribed. If a response to the second drug was not forthcoming, the process was repeated until the desired effect was achieved or until all possible drugs and drug combinations were explored and exhausted. Trial and error therapy places both the patient and clinician at the mercy of an unknown factor—the kinetics of the prescribed drug in that particular patient.

Today, TDM allows more accurate titration of dosage to assure optimal serum concentrations, and thus greater individualization of drug therapy. Only within the last three decades has a clearer understanding of the relationship between a drug's concentration within a biological system and its therapeutic effectiveness become possible. Our current ability to monitor a wide variety of drug doses and to correlate their plasma concentrations (and by inference their tissue concentrations) with their therapeutic effects had to await the development of highly specific, reproducible analytical technologies. Utilizing these analytical techniques, investigators established that a minimum effective concentration (MEC) of a given drug in the plasma was necessary to elicit that particular drug's therapeutic ef-

TABLE I

Percentage of Steady-State Plasma Levels Achieved at Each Half-Life Interval

Number of half-lives	Percentage of steady-state levels
1	50
2	75
3	88
4	94
5	97
6	98
7	99

fect. Furthermore, it is now accepted that there are optimum plasma concentration ranges (see Table I) within which therapeutic effects can be expected to occur in most patients receiving a particular drug (*1,2*). Should plasma concentrations exceed the optimum plasma concentration range, undesirable side effects or toxicity, which may or may not be clinically demonstrable, can be expected in most patients. It is essential that the physician utilizing drug concentrations reported by the analytical laboratory as well as the clinical pathologist, clinical pharmacist, or clinical chemist providing routine monitoring services understand the fundamental principles and techniques of clinical pharmacology as they are applied to patient care. Such an understanding will enable more effective interpretation of TDM data, which is encountered in various clinical situations.

A. Site and Mechanism of Drug Action

The biological effect achieved following a given drug dose is a direct consequence of the formation of reversible bonds between the drug and tissue receptors controlling a particular response (*13*). For most drugs, the intensity and duration of a given pharmacological effect is proportional to the drug concentration at the receptor site (*13,14*). The exact mechanisms of drug receptor interactions, however, remain unclear. In order for a drug to exert the desired biological effect, it must reach and interact with the receptors regulating a specific response. In addition, disease, age, sex, compliance, drug interactions, and individual differences in drug metabolism and excretion contribute to inter-patient response differences (*7–11*). Figure 1 schematically depicts the factors which can alter the concentration of drugs, ultimately achieved and maintained at a given receptor site.

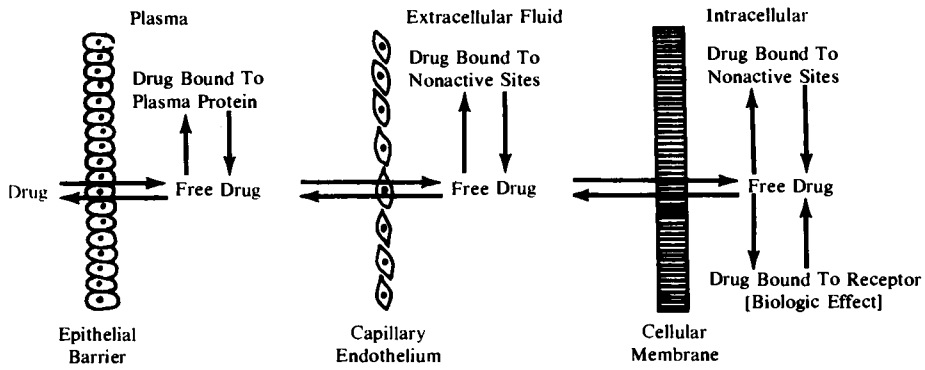


Fig. 1 Drug distribution scheme. Adapted from Pippenger (*39*).

The titration of drug dosage using TDM is the most precise method for achieving therapeutic plasma concentrations and compensating for these inter-individual variations in response (*1,2*). Every drug acts to produce a change in some known physiologic function or process. Any drug may increase, decrease, or return to normal the physiologic function of tissues, organs, or physiological systems. The biological effect observed following administration of a drug is the sum of the processes by which that drug creates changes in some physiologic or biochemical process. Such effects can only be measured and expressed in terms of alteration of a specific function or process. A change in function due to a drug's pharmacodynamic activity may return the function or physiologic process from an abnormal to a normal level of activity. Or, it may prevent deviation from the normal physiologic state of a given system (*1,2*, and *13*).

For most drugs, the intensity of a pharmacologic effect tends to be proportional to the drug concentration in extracellular fluid which can enter tissues and interact with specific receptors to elicit a biological effect (*1,2*, and *13*). For example, antiepileptic drugs are believed to prevent seizures by binding to neural membranes or altering the neurotrasmitter release. Alteration of these functions is thought to stabilize neuronal membranes against the excessive electrical activity that is responsible for precipitating a clinical seizure. Figure 1 is a schematic representation of drug distribution following absorption between the plasma and the various tissue compartments. Drug concentration in extracellular water is in equilibrium with the drug concentration in plasma water. The latter, representing the free drug concentration, is an indirect measure of drug concentration at the site of action.

Since many drugs are partially bound to plasma proteins, an equilibrium exists between the concentration of protein-bound drug and the free drug concentration in plasma water (*16a,17*). Only free drug is capable of crossing the various lipoprotein membranes that surround the receptor sites. It is impossible to directly monitor receptor site drug concentrations *in vivo;* therefore, monitoring the total plasma drug concentrations is a reflection of the equilibrium which exists between tissue, extracellular fluid, and plasma water drug concentrations. The site at which a given drug acts to initiate the events which lead to a specific biological effect is arbitrarily defined as that particular drug's *site of action*. A drug's biological effect may be elicitated by direct interaction with a receptor that controls a specific function, or by alteration of the physiologic process which regulates that specific function (*9–15*).

The *mechanism of action* of a drug refers to the actual biochemical or physical process which initiates a biological response at a specific site. The mechanism of action of most drugs depends upon their chemical in-

teraction with a functionally viable component of some physiologic system. However, since the exact molecular mechanism of action for most drugs remains obscure, theoretical models have been developed to explain this mechanism of action. The fundamental concept upon which these models are based is that intracellular macromolecular receptors exist which, when stimulated, elicit a specific biological response. More specifically, drugs are believed to combine reversibly with receptors by means of ionic bonds, hydrogen bonds, and van der Waals forces. Such a reversible combination is thought to form a drug–receptor complex of sufficient stability to alter the physiologic response of the target system, consequently producing the observed pharmacological effect. Both clinical and molecular studies of the pharmacological profiles of a wide variety of drugs have demonstrated that a much better correlation exists between the observed clinical effects of a drug and its plasma concentration than that observed between the clinical effect and total daily drug dosage.

Numerous factors, including individual differences in drug metabolism and excretion, age, sex, patient compliance, disease, and drug interactions (particularly during multiple drug therapy) regulate the disposition pattern of a drug within an individual patient(*1,2*). The rate of drug disposition regulates the amount of drug available to interact with a receptor; thus the therapeutic response observed in a given patient is dependent on the sum of all these processes. The observed clinical effect of a given drug in a specific patient is directly related to the drug concentration in that particular patient. Interactions between all the potential factors influencing drug disposition accounts for the broad inter-patient variability in plasma concentrations following either single or multiple drug doses. Individual patient response to a given drug dose, however, remains constant because the factors which can alter drug utilization within the individual are relatively fixed (*1,2*).

Generally, inter-individual variations of response, as demonstrated by the clinical response of a large population to a fixed drug dose, are more of a reflection on the relationship between total daily dose and plasma concentration than they are of the relationship between plasma concentration and the intensity of response. In other words, the probability of achieving a given plasma concentration from a given drug dose is much less than the probability of obtaining a specific biological effect from a given plasma concentration. This is why drugs administered at fixed doses produce marked variations within a population in the observed therapeutic response: When an average or standard drug dosage is administered to a large patient population, the desired therapeutic effect will be achieved in some patients; no therapeutic effect will occur in others, and toxicity, usually associated with drug overdosage, will be evident in still others (*1* –

14). The titration of drug dosage to obtain an optimal therapeutic plasma concentration in the individual patient can successfully eliminate undesirable therapeutic responses which are a direct consequence of inter-individual variations in drug disposition. TDM can aid clinicians in identifying the appropriate therapeutic regimen necessary to achieve optimal plasma concentrations of a drug and thus assure the prompt establishment and maintenance of the desired therapeutic response.

For most drugs, there is a direct linear relationship between the administered drug dosage and the plasma concentration achieved at steady state (*9–11*). However, one of the greatest misconceptions in this area is that such a linear relationship exists between the plasma concentration and dose for all drugs, i.e., that as the total drug dose is increased, there will be a concommitant, directly proportional linear increase in that specific drug plasma concentration. Unfortunately, this is not the case. In fact, certain drugs such as phenytoin and amitriptyline demonstrate an apparent linear dose–concentration relationship only over a given range (*1, 2*). Beyond this point, a marked elevation of plasma concentration, completely disproportionate to the dose administered, can follow what would be clinically considered a negligible dosage increase. The phenomenon responsible for nonlinear increases in drug concentration is termed *saturation* or *zero-order kinetics*. Saturation kinetics is a direct reflection of the limited capacity of some drug-metabolizing enzyme mechanisms. Clinically, saturation kinetics should be suspected when a patient rapidly and unexpectedly develops an adverse drug reaction following a small dosage increment (*1,2*). One of the major advantages of TDM is that it can "predict" most pharmacologic responses. It does so by either assuring the clinician that the plasma concentration is within an optimal range in a given patient, or signalling the presence of subtherapeutic or toxic concentrations in a specific patient. TDM is not a panacea, and drug concentration data must always be interpreted in conjunction with the patient's clinical status (*1,2,* and *10*). TDM as part of routine patient care provides physicians with a valuable, more precise tool for assessing the pharmacologic status of individual patients and helps them establish a regimen to achieve the desired therapeutic response.

B. Factors Which Alter Individual Drug Disposition Patterns

1. Patient Noncompliance

It has been suggested that over 60% of all patients do not take their medications in the manner prescribed by their physicians. The most common cause of suboptimal drug concentrations, and consequent failure to

achieve the desired therapeutic response, is patient noncompliance. Whenever a patient presents with consistently low plasma drug concentrations, noncompliance should be considered the probable cause. Noncompliance can usually be demonstrated by careful supervision of the patient's daily drug intake over a specified time interval (usually 5 half-lives of the drug) with routine monitoring of serum drug concentrations at appropriate intervals. If there is a progressive increase in serum drug concentration over the time interval selected, the patient has been noncompliant (*1,2*).

The administration of the recommended or average total daily dose of a given drug without taking into account the numerous factors which alter drug disposition in each patient can also lead to consistently low serum drug concentrations. Failure to individualize drug therapy (physician noncompliance) is often responsible for suboptimal drug concentrations (*1,2*). If the serum concentrations remain low under supervised intake, other causes such as drug malabsorption or rapid drug metabolism should be suspected.

2. Drug Absorption

The entrance of drugs into the general circulation following either i.v. or i.m. administration is generally rapid and circumvents the problems associated with drug absorption following oral administration (*9–14*). Most drugs are administered orally. Following oral drug administration, a number of factors can alter the amount of drug absorbed from the gastrointestinal tract into the circulatory system. The type of drug preparation, drug solubility, concomitant administration of other drugs, whether or not the drug is taken with meals, and the presence of diarrhea or constipation can all alter the amount of drug which will be absorbed following a single oral dose (*1,9–14*). Some patients receiving appropriate drug doses will have consistently low plasma drug concentrations. Generally, these patients are classified as either noncompliant or as fast drug metabolizers.

However, before classifying someone as a fast metabolizer, the patient's ability to absorb the administered drug should be evaluated. Malabsorption of an orally administered drug can often be confirmed by measuring serial plasma drug concentrations at given time intervals after parenteral administration of the prescribed dose. If altered absorption is present, the maximum plasma concentrations and observed drug half-life following the i.v. dose will be significantly higher than those achieved following the same dose administered orally. Conversely, if the patient problem is fast drug metabolism, there will be no significant differences in the plasma concentrations achieved or the observed half-life regardless of the route of administration (*9–14*).

3. Drug–Plasma Protein Binding

Upon entering the systemic circulation, any protein-bound drug will bind to plasma proteins and an equilibrium between free and bound drug will be established. By definition, the *bound drug* is that portion of a drug bound to plasma proteins. Bound drug is unable to cross cell membranes and consequently exerts no biological effect. The unbound or *free* drug is dissolved in the plasma water and can be transported across cell membranes. Only the free drug can cross biological membranes and interact with specific receptors to elicit a biological response. Each drug has its own characteristic protein-binding pattern which is dependent on its physical and chemical properties (*16–19*). As a general rule, acidic drugs are bound primarily to albumin, and basic drugs to globulins, particularly α-1-acid glycoprotein.

A drug may be either tightly or loosely bound, as determined by its affinity for plasma proteins. A weakly bound drug can be displaced from its protein sites by binding one with greater affinity for the plasma protein-binding site. Protein binding of a drug is also dependent on the physical characteristics of the plasma proteins, and on the presence or absence of fatty acids or other drugs in the blood. Fatty acids can interfere with protein binding by displacing a drug from its protein-binding sites. Tightly bound drugs will not be displaced, but a weakly bound drug can be displaced quite rapidly by elevated free fatty acids or another drug (*17*). It is important to recognize that even though the total serum drug concentration may remain unchanged, displacement of a drug from its plasma protein-binding site can elevate free-drug concentrations and result in clinical toxicity.

Certain disease states can significantly alter drug protein binding. For example, uremic patients lack the ability to completely bind drugs to the plasma proteins. In the case of phenytoin, this lack of binding capacity ranges from uremic patients who can bind no phenytoin to those who can bind only 60–70% of the phenytoin present in plasma. Clinically, this means that in a patient who lacks the capacity to bind phenytoin, concentrations of 1–2 μg/ml would be therapeutic and plasma concentrations above 2.5 μg/ml would result in phenytoin toxicity. Altered drug binding requires careful monitoring of all drugs administered in patients with abnormal renal function (*20*).

In patients who present with either clinical toxicity or a nontherapeutic response, when total plasma concentrations are known to be optimal, altered protein binding should be considered. Until recently, determination of protein binding was a time-consuming and tedious procedure. Since only the free drug crosses into the saliva, the protein-binding status of a patient can be assessed indirectly by measuring salivary drug concentra-

tions (*21,22*). Nevertheless, caution is indicated. Salivary levels are a good indicator of free drug levels for any drug that has an ionization constant (pK_a) significantly different from the pH of plasma, e.g., phenytoin. However, with phenobarbital, salivary concentrations will not reflect the true free-drug concentrations (*21*). Salivary drug levels will not reflect the actual free concentration of drugs which are actively transported into the saliva. The recent development of rapid ultrafiltration systems for determining free plasma drug concentrations should enhance our ability to identify patients who lack the ability to bind proteins (*23,24*).

4. Drug Metabolism.

Any foreign compound that enters the body must be eliminated. Phylogenetically, from fish to man drug elimination mechanisms become more complex as one proceeds up the phylogentic scale (*25,26*). There is a progressive increase in the ability of the body to alter foreign compounds into compounds which are more water-soluble and thus readily excreted. It is generally believed that the ability of the liver to metabolize drugs evolved as a mechanism for detoxifying poisionous substances ingested with food.

The drug-metabolizing enzymes of the liver are nonspecific and interact with a wide variety of chemical structures. The entire purpose of the hepatic drug-metabolizing systems is to make compounds more water-soluble; therefore, the degradation of orgnaic compounds leads to compounds that are less fat-soluble and more water-soluble (*27*). Metabolites of many drugs are conjugated within the liver to either glucuronic acid, amino acids, or sulfates, thus increasing water solubility even more, and consequently, the rate of renal excretion. For example, *p*-hydroxyphenytoin, the major metabolite of phenytoin, is conjugated with glucuronic acid. This conjugation increases its water solubility almost 100 times. Most drug metabolism takes place within the microsomal fraction of the hepatocyte.

The microsomal enzyme systems are also responsible for the metabolism of endogenous steriods (*9–14*). The enzymatic systems responsible for drug metabolism are not designed to recognize specific drugs; rather they act upon classes of compounds with similar structures. The same enzyme that is responsible for the hydroxylation of phenytoin is also responsible for the hydroxylation of many other drugs containing an appropriate phenyl ring. Therefore, when phenytoin is administered simultaneously with one of these drugs, there may be some clinically significant alterations of drug concentrations that are a direct consequence as competition for metabolic sites on the enzyme increases. Clinically, one would expect to see higher serum concentrations of the drug with the least affinity for

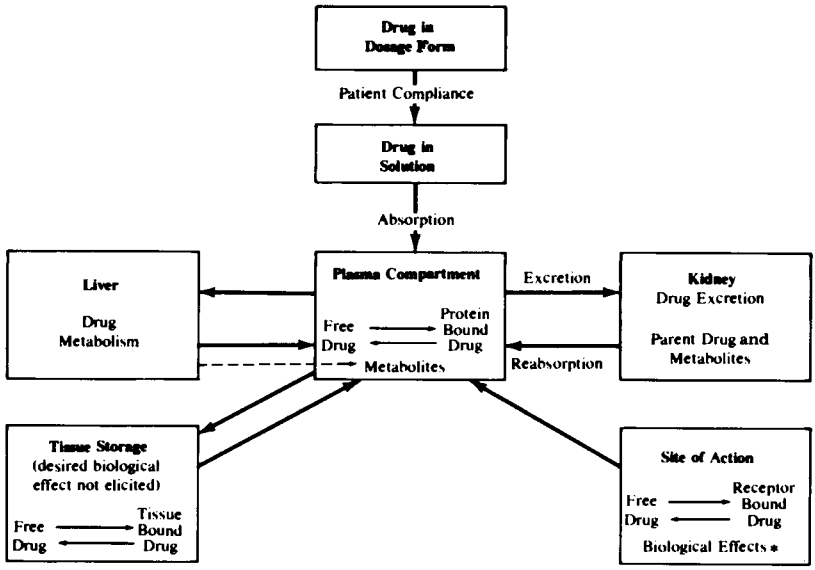
the enzyme. Phenytoin has a very low affinity for microsomal enzymes. Thus, administration of a drug with a greater affinity for the enzyme than phenytoin will decrease phenytoin's rate of metabolism, and plasma phenytoin concentrations will become elevated (*23*).

One characteristic of the hepatic microsomal system is that it can be induced to metabolize drugs at a faster rate. As increasing doses of drug are administered, the body, in its attempt to eliminate the drug, synthesizes new proteins, in the form of enzymes capable of metabolizing that agent (*9–14*). Increased activity of drug-metabolizing enzymatic systems is not necessarily induced with every dosage increment or with the addition of another drug to the patient's regimen. There is a maximum rate at which protein synthesis can occur (*2*; see also *28*).

Thus, if a patient has been regularly receiving a drug with known enzyme-induction properties, it does not follow that a second drug of similar structure added to the patient's therapeutic regimen will cause a marked increase in the rate of metabolism of both the first and second drug. Genetic factors play a major role in determining the ability of a patient to metabolize drugs (*29*). Individuals of different ethnic origins as well as individuals in certain families metabolize drugs, e.g., phenytoin or isoniazid, at a faster or slower rate than the general population (*30*). The importance of identifying fast or slow drug metabolizers cannot be overemphasized. A fast drug metabolizer will require a greater daily dose (mg/kg) than will a normal individual to achieve the same serum concentration necessary for eliciting the desired therapeutic response. A slow drug metabolizer given standard drug dosages will invariably exhibit drug toxicity.

Absolute identification of fast and slow drug metabolizers depends upon the quantitative identification of urinary drug-metabolite excretion profiles as well as on the serial determination of plasma drug concentrations. Use of plasma drug concentrations alone to identify fast and slow metabolizers can be misleading. Generally, plasma drug concentrations of slow metabolizers will be significantly higher than would be observed in the general population receiving the same mg/kg/day dosage. Consistently high plasma concentrations on normal or low drug doses is suggestive of slow drug metabolism. However, a drug interaction or disease process that blocks drug metabolism will also result in elevated plasma concentrations. On the other hand, fast drug metabolizers usually exhibit consistently low plasma concentrations on standard dosage regimens. Since plasma drug levels in noncompliant patients mimic those observed in fast metabolizers, there is a tendency to identify noncompliant patients as fast metabolizers.

Generally, drugs are metabolized from a pharmacologically active



Absorption

Drug must be formulated in a manner which assures bioavailability for absorption.

Metabolism

Drug converted to a more soluble compound which may be biologically active or inactive. Metabolism can also occur in other tissues.

Excretion

Usually more water-soluble drug metabolites are excreted in urine. Also drug excretion can occur via bile, feces, saliva, and expired air.

Tissue Storage

Distribution of drug to sites where the desired biological effect is not elicited. Undesirable effects may be elicited by drug interaction with a specific physiological system.

Site of Action

Free drug binds to receptor to elicit a biological effect (response). Number and type of receptors to which drug is bound determines the intensity and duration of the desired and undesired effects.

Fig. 2 Scheme depicting the total physiological activity of a sample drug.

agent to an inactive product, incapable of eliciting a given therapeutic response. There are exceptions to this rule: Some organic compounds (*9–14*), when metabolized, have a greater biological activity than the parent compound. For example, diazepam is rapidly metabolized to desmethyl diazepam, which is the most active antianxiety agent of all the diazepam metabolites. As a general rule, when a compound has a less polar active metabolite, the half-life of the active metabolite is significantly longer than that of the parent compound. Such is the case with procaina-

mide and *N*-acetylprocainamide (NAPA). The half-life of procainamide is 3–4 hr, whereas NAPA has a half-life of 6–9 hr in patients with normal creatinine clearance. This means that there will be an accumulation of NAPA, the active metabolite, within the system and at its site of action.

5. Renal Excretion

Urinary excretion is the major pathway for the elimination of drugs and their metabolites. For any drug which is not extensively metabolized, changes in renal function will alter that drug's plasma concentrations. If renal function is impaired, drug plasma concentrations can become elevated (*9–14,20*).

Uremic patients and those with congestive heart failure have decreased renal drug clearance. Interestingly, drug metabolites are so water-soluble that a significant decrease in urinary output will not result in increased plasma concentrations of most conjugated drug metabolites (*20*).

The clinical status of a patient can also dramatically alter drug utilization patterns. Hepatitis can impair the metabolism of drugs. If the liver has lost its reserve capacity, patients with hepatitis can become severely intoxicated when given drugs dependent upon hepatic degradation (*9–14*). Congestive heart failure can significantly alter the distribution of drugs to tissues, thus precipating altered drug utilization and response patterns.

III. PHARMACOKINETICS

Anyone utilizing routine therapeutic drug monitoring must constantly keep in mind that the plasma concentration achieved and maintained following the administration of a fixed drug dosage is a direct consequence of the interactions of a wide variety of processes. These include drug absorption, distribution, metabolism, and excretion, in addition to the physiologic status of the patient. All these factors are interrelated, each playing a role in determining the steady-state drug concentration which will be achieved on a fixed dosage regimen. The study of these interrelationships forms the basis of pharmacokinetics.

Pharmacokinetics is the study of the time-course of drug and metabolite levels in different fluids, tissues, and excreta of the body, and of the mathematical relationships which can be utilized to develop models for interpretation of the blood concentration patterns observed in a given patient. In the practical sense, pharmacokinetics as a discipline represents an attempt to utilize mathematical models to predict the distribution and excretion patterns of drugs, usually at steady-state concentrations, following a

given dosage regimen. Applied clinical pharmacokinetics has been developed to the point where it is applicable to the study of patients receiving a given drug, provided the theoretical limitations of the model are recognized.

One of the unfortunate limitations of pharmacokinetic models is that many models do not take into account multiple-drug therapy or the clinical status of the patient. Interactions between drugs can alter the kinetics of each and affect plasma drug concentrations as well. Therefore, unless specific clinical data from a given patient are available, these models should serve only as a general guideline. Availability of drug monitoring techniques in biological fluids resulted in attempts to correlate a given mg/kg dosage of drug with the observed plasma concentration and clinical response in large patient populations. The fundamental assumption of these studies was that the patient was at steady state, i.e., the intake of a drug was constant over a period of time, and drug elimination, as reflected in the rates of drug metabolism and excretion, was constant.

Based upon data derived from these studies, a number of computer programs have been developed that, given plasma concentration data with respect to time, will calculate the drug dosage necessary to achieve a given plasma drug concentration in a specific patient. Unfortunately, these programs, and the information derived from them, are not yet widely available to clinical chemistry laboratories or practicing clinicians.

The clinical establishment of a national therapeutic drug regimen does not require a detailed knowledge of pharmacokinetics. However, an awareness of the terminology and fundamental principles is essential. The following descriptions and discussion of some of the relevant terms are designed to alert those engaged in TDM to the terminology of pharmacokinetics.

A. First-Order Kinetics

A process associated with drug utilization (clearance) exhibits first-order kinetics. Figure 3A shows the linear relationship between plasma drug concentration and total daily dose (mg/kg). This graphically depicts that an increase in drug dose would be expected to result in a proportionate increase in plasma drug concentration.

B. Zero-Order Kinetics

When the rate of a process is independent of concentration it is said to follow zero-order kinetics. Zero-order kinetics become clearly apparent when a point is reached at which enzyme or transport mechanisms be-

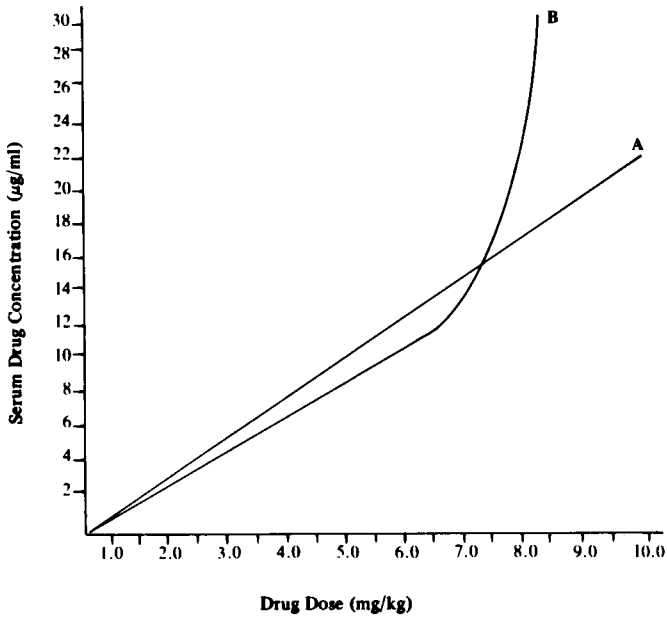


Fig. 3 Relationship between serum drug concentration and total daily drug dose. (A) Dose-response curve for a drug observing first-order kinetics (linear). (B) Dose-response curve for drug observing zero-order kinetics (nonlinear or saturation). Reproduced from Pippenger (*39*).

come saturated. While plotting of drug plasma concentration versus total daily dose (mg/kg) initially yields an apparently straight line indicative of first-order kinetics, a sharp upward curve is seen as the saturation point is reached (see Fig. 3B). The changes in the rates of drug clearance which occur beyond the saturation point, as represented by the disproportionate increase in plasma drug concentration following a given dosage increment, is the hallmark of zero-order kinetics (*31–33*).

Fortunately, in clinical practice, only a few drugs exhibit zero-order kinetics. For most drugs, the plasma concentrations achieved at therapeutic dosages are low relative to the concentration necessary to saturate the particular system involved (*31,32*). Therefore, first-order kinetics are observed throughout the therapeutic range. There are notable exceptions to this rule, however, since both phenytoin and aspirin exhibit saturation kinetics near the upper limits of the therapeutic range. For any drug that exhibits zero-order kinetics, a very small dosage increment may result in a clinically significant elevation of plasma concentrations. It is to be noted that even though the initial dose-response curve may appear linear in drugs with zero-order kinetics, the drug clearance is altered throughout

the dosage range and at all plasma concentrations and does not parallel the kinetics observed in a first-order relationship (*31–33*).

C. Drug Half-Life

Drug half-life is also referred to as the elimination half-time $t_{1/2}$. It is the time required for elimination of half the concentration of a drug present in the system, provided no additional drug is administered following a given point in time. For example, if the concentration of phenytoin ($t_{1/2}$ = 24 hr) were 20 μg/ml, the time required to clear the drug to a concentration of 10 μg/ml would be 24 hr, provided no additional doses of the drug had been given. It must be emphasized that drug half-life is, in reality, a reflection of the individual rates of the several different processes which regulate drug clearance. The rates of drug metabolism and excretion are the primary determinants of the drug half-life in any given patient (*31–33*).

D. Fate of A Single Drug Dose

Following the administration of a single drug dose, a peak plasma concentration is reached when the absorption phase is almost complete (see Fig. 4). The plasma concentration then begins to decline, even as the drug continues to be absorbed. The rate of this decline in plasma concentration is dependent upon the rates of absorption, metabolism, and excretion of

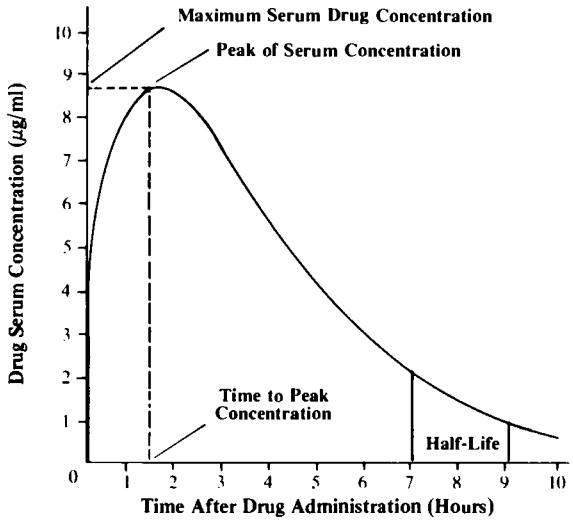


Fig. 4 Dose-response curve after oral administration of a single dose of a hypothetical drug. Reproduced from Pippenger (*39*).

the drug. Once the absorption phase is complete, the rate of decline in plasma concentration is a reflection of the clearance (elimination) rate which is the sum of the rates of excretion and metabolism of the drug. Following completion of the absorption phase, the half-life can be determined by measuring the decline in plasma concentration over fixed time intervals (*31–33*).

E. Steady States

When long-term oral therapy is initiated, the drug will continue to accumulate within the body unitl such time as the rate of clearance, which comprises all tissue distribution, metabolic, and renal processes involved in drug disposition, equals the rate of administration (see Fig. 5). When the equilibrium between drug clearance and intake is achieved the system is said to be at a *steady state*, i.e., the amount of drug ingested over a 24-hr period is equal to the amount of drug eliminated in the same 24-hr period. Over a period of time, body and plasma drug concentrations will increase exponentially until they reach a steady state or plateau. The time required to reach a steady state following institution of drug therapy is presented in Table I. It requires 7 half-lives of drug administration before a true steady state concentration is achieved and stabilized. Steady state processes are, however, 97% complete within 5 half-lives (*31–33*). As a practical rule, 5–6 times the half-life of any drug is the time required to achieve a steady

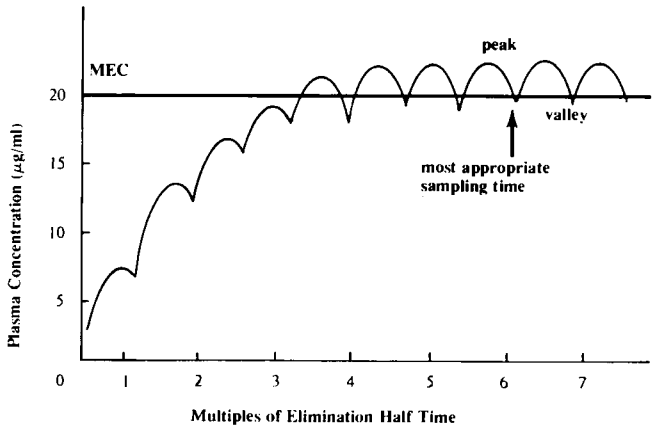


Fig. 5 Dose-response curve following repeated oral drug administration. MEC, minimum effective concentration. Reproduced from Pippenger (*39*).

state. For example, phenytoin, which has a half-life of 24 hr, requires a period of 24 × 5.5 or 132 hr (5–6 days) to achieve a steady state. In contrast, drugs such as primidone or valproate, which have half-lives of 6–8 hr, require only 33–44 hr to reach a steady state (*34*).

It is important to note that the same principles which govern the gradual accumulation of a drug to a steady state also apply when drug therapy is discontinued. For instance, if plasma phenytoin is at a steady state concentration, and drug administration is stopped, there will be a period of 5–6 days, or 5–6 half-lives, before the drug is completely eliminated from the body. This is why drugs with prolonged half-lives can still be detected in plasma for 3–4 weeks after administration of the last dose. For example, phenobarbital, with a half-life of 4 days, requires 22 days for complete elimination. Bromides, with a half-life of 12 days, require 66 days for complete elimination (or to reach steady state).

Plasma concentrations for drugs with first-order kinetics during the steady state are linearly related to dose. For any change of dosage, after a steady state has been achieved, the principles regulating the time required to achieve a new steady state stage still apply. To illustrate this important principle, if the maintenance dose of a drug were doubled, the new steady-state concentration would not double until the completion of 5–6 half-lives. If a plasma drug concentration is determined before achievement of a steady state, for example after 2 half-lives, it will not reflect the true steady-state concentration of the drug.

It is emphasized that the full therapeutic effect from a given dose is not achieved until steady-state concentrations are reached. Therefore, before a given dosage regimen is considered as a failure, the clinician should be sure that steady-state concentrations have been achieved (*1* and *31,32*).

IV. INDIVIDUALIZATION OF DOSAGE

The drug concentration during a steady state is directly proportional to the drug half-life (clearance). That is, if two patients differ in their respective rates of drug clearance, i.e., if one were a fast and the other a normal metabolizer, the daily dose necessary to maintain the same steady state concentration in the two patients will differ by the same factor that the rates of drug clearance differ. This is why individualization for drug therapy is absolutely essential to assure appropriate steady-state concentration. It is to be emphasized that every individual has a different drug clearance pattern (half-life) on a fixed dosage regimen which is dependent on his overall physiological status at that time. In other words, the clinical status of a patient is an important determinant of drug clearance.

Determining Dosage Intervals

A steady-state drug concentration is usually maintained in an individual patient by various combinations of total drug dosage and dosage intervals. Generally, in order to maintain a smooth, constant steady-state drug concentration, without excessive fluctuations, dosage intervals should be half of each particular drug's half-life. To illustrate this in practical terms, phenytoin would be given every 12 hr since its half-life is 24 hr, whereas primidone would be given every 3 hr since its half-life is 6 hr (*1* and *31*, *32*).

Short dosage intervals (3–4 hr) are often impractical in out-patients, but drugs with short half-lives should be administered at least once each half-life. The principal idea is to maintain the valley (or trough) drug concentration within the therapeutic range, i.e., above the minimum effective concentration (MEC), without the peak concentration reaching toxic levels. As long as the dosage schedule and the interval between doses are selected so that there are no significant fluctuations between peak and valley concentrations of drug during the dosage interval, an appropriate steady-state concentration will be maintained. If the dosage interval is too long relative to the half-life of the drug, the plasma concentration just prior to the next dose may be insufficient to provide the desired therapeutic effect. To apply this concept to epilepsy, for example, a patient whose phenytoin level falls from 14 μg/ml to 9 μg/ml may have a seizure at the lower level but not at the higher. For any drug, it is impossible to maintain a plasma concentration within the optimal therapeutic range at all times by adhering to appropriate dosage schedules.

If a patient's drug absorption is very rapid, or if dosage intervals are excessively short, he may experience periods of drug intoxication which may present clinically as the side effects normally associated with that drug. These symptoms usually appear transiently at fixed intervals following drug administration throughout the day. The toxicity is attributed to a peak plasma concentration above the optimal therapeutic range shortly after drug administration. Such side effects can often be eliminated by extending the dosage interval to assure that peak concentrations do not exceed the upper limits of the therapeutic range.

V. BASIC GUIDELINES FOR ROUTINE THERAPEUTIC DRUG MONITORING

The ultimate responsibility of the laboratory engaged in routine TDM is to assure that all information relevant to the patient's pharmacological profile be available for utilization by the clinician to individualize a given patient's therapeutic regimen.

Clinical pharmacokinetics is a valuable tool for understanding and interpreting the response of an individual patient to a given drug regimen. A number of texts containing detailed mathematical derivations of the fundamental principles of clinical pharmacokinetics are available (*31–33*). Computer programs which apply these principles to dosage calculations for individual patients have been developed (*31,32,35*). These programs can be utilized for any given drug to calculate the expected plasma concentration which will be achieved over a fixed time interval following a given dose, provided the patient's plasma concentration at a given time interval after the last drug dose is available. Unfortunately, these programs and the information derived from them are not widely available to the clinical chemistry laboratory or to the practicing clinician. A series of simple guidelines exist which will generate approximately the same information as the computer programs without the necessity of complex mathematical formulas or computer programs (*1,6,10*).

A. Types of Patient Information Needed for Therapeutic Drug Monitoring

Wide individual variability exists in patient utilization of drugs as a direct consequence of genetic factors, multiple drug therapy, age, and weight. It is extremely difficult and dangerous to generalize about the relationship between plasma concentrations and drug dose. Rather, it is necessary to apply the fundamental principles of clinical pharmacology to achieve the desired pharmacological effect without introducing unwanted drug side effects in each individual patient.

The importance of individualized drug therapy is particularly significant when administering drugs to children who utilize drugs at a faster rate than do adults. Conversely, geriatric patients generally utilize drugs at a slower rate than the general adult population and are more susceptible to the development of adverse side effects (*6,7,9–12*).

In order to derive as much information as possible about the pharmacological status of the patient, each laboratory engaged in therapeutic drug monitoring should have the following information available at the time any drug is monitored:

1. Patient's age. It is clearly established that there are marked age-dependent differences in drug utilization, in particular are the transition ages between neonate and infant, child and adolescent, and adult and geriatric.
2. Patient's weight. The weight of the patient is essential for mathematical calculations of the relationships between drug dose, plasma concentration, and drug clearance.
3. All drugs which the patient is receiving. Knowledge of the drugs which the patient is receiving, in addition to the agent being monitored, is

essential for identification of potential drug interactions that have been reported to alter plasma concentrations as well as for the identification of compounds which may interfere in a given analytical technique.

4. Total daily dosage of drugs. A knowledge of the total daily dosage for each drug administered is necessary to mathematically determine the patient's total daily drug dose in mg/kg. Without this information it is impossible to correlate the patient's actual plasma concentration with his expected drug concentrations. Knowledge of the mg/kg dose allows the prediction of the patient's expected plasma drug concentration by calculation with the concentration dose ratio (CDR). Predicted drug concentrations can then be correlated with the observed (measured) drug concentration to provide an indication of the patient's pharmacological status.

5. Critical time-intervals. The time at which the last dose of drug was administered and the time at which the blood specimen was drawn are essential bits of data. Without this information it is difficult to assess whether the actual plasma concentration represents a peak or trough level. A knowledge of the actual sampling time and dosage interval is extremely important for accurate interpretation of plasma concentrations of drugs with short half-lives, such as theophylline and lidocaine.

6. Clinical status of the patient. It is well established that acute or chronic disease can dramatically alter drug utilization patterns. Awareness of the patient's current clinical status is particularly important for regulation of drug therapy in patients with hepatitis or renal failure. Without the knowledge of the clinical status of the patient, it is impossible for those interpreting drug concentrations to distinguish an altered drug utilization pattern which is associated with a given disease state from other factors (noncompliance, drug interactions, etc.) which can be present with a similar pattern.

If information with respect to all of the components described above is available, measurement of drug concentrations can provide a great deal of insight into the pharmacological status of the patient. Clinical application of the information derived from TDM will allow for the individualization of drug therapy.

Calculation of Concentration Dose Ratio

For all drugs which exhibit first-order kinetics, there is a direct relationship between the total daily drug dose and the plasma concentration achieved at steady state (*30,31*). The contribution of each mg/kg of drug administered to the final steady-state plasma concentration in a given patient can be determined by calculation of the concentration dose ratio (CDR) for the administered drug as follows:

$$\text{CDR} = \frac{\text{Plasma concentration at steady state } (\mu\text{g/ml})}{\text{Total daily drug dose (mg/kg)}}$$

Therefore, multiplication of the patient's total daily drug dose (mg/kg) by the CDR will yield the anticipated steady-state plasma drug concentration as follows:

$$\text{CDR} \times \frac{\text{Total daily dose}}{(\text{mg/kg})} = \frac{\text{Steady-state drug concentrations}}{(\mu\text{g/ml})}$$

The CDR for many drugs has been established, based upon data obtained from routine monitoring of drug concentrations in large patient populations. Therefore, if one knows the CDR, the plasma concentration which would be expected in the average patient of a general population can be calculated on the basis of the patient's total daily drug intake (mg/kg). Utilizing the CDR, it is possible to calculate a patient's expected plasma concentration at steady state.

To predicate a blood level for phenobarbital, the following calculation is used:

$$\text{CDR} \times \text{dose} = \text{plasma concentration}$$

For example, each mg/kg of phenobarbital administered contributes 10 μg/ml to the steady-state plasma level in adults.

Thus, in a patient who receives a total daily phenobarbital dose of 180 mg and weighs 60 kg, the total daily dose = 180 mg/kg = 3 mg/kg/day. If the patient is 26 years old, the CDR is 10(6), and the expected phenobarbital level would be 30 μg/ml.

$$[(10\ \mu\text{g/ml})/(\text{mg/kg})] \times 3\ \text{mg/kg} = 30\ \mu\text{g/ml}$$

Thus, a phenobarbital concentration of 30 μg/ml at steady would be expected. If a patient, on the other hand, weighing 80 kg were receiving the same 180-mg phenobarbital dose, one would expect to achieve, on the basis of his total daily dose of 2.1 mg/kg/day, a plasma concentration of 21 μg/ml.

B. Monitoring Steady-State Drug Concentrations

When long-term oral therapy is initiated, the drug will continue to accumulate within the body until such time as the rate of drug clearance (elimination) is in equilibrium with the total daily drug intake. Drug clearance encompasses all absorption, distribution, and metabolic and renal processes involved in drug disposition. Over a period of time, body and plasma drug concentrations will increase exponentially until they reach a steady state or plateau (see Fig. 5).

We again emphasize that the time required to reach a stabilized steady state is 7 half-lives following institution of drug therapy; steady-state processes are, however, 97% complete within 5 half-lives. If the prescribed

drug dosage is changed after a steady state has been achieved, the principles regulating the time required to achieve a new steady state plateau still apply (5× the drug half-life). For instance, if the maintenance dose of a drug were doubled, the new steady-state drug concentration would not be doubled until the completion of 5–6 half-lives. Only at steady state is the full therapeutic effect of that dose achieved.

If a plasma drug concentration is determined before achievement of a steady state, for example, after only 2 half-lives, it will not reflect the true steady-state concentration of the drug. Measurement of plasma concentrations before a steady state is achieved does not yield as much clinically useful information with respect to the patient as do levels measured at steady state.

It is possible, if one knows the time of initiation of drug therapy, to extrapolate predicated steady-state plasma concentrations by correcting for the number of half-lives expired before sampling. This technique, however, provides only a rough estimate of the expected steady-state concentrations.

C. Consideration of Half-Life Variables

Drug half-life, by definition, is the time required for elimination of half the plasma concentration of drug present at an initial starting time. It must be remembered that half-life is in reality the elimination half-time of a drug and reflects the various factors which determine the clearance of a drug.

The half-life is one of the most important pieces of information for anyone attempting to interpret blood levels. Multiplication of the half-life × 5 yields the appropriate time required to achieve steady-state drug concentrations or the time required to eliminate the drug from the body following discontinuance of drug therapy.

There are many factors that can alter the half-life of a drug. Since half-life is dependent upon the rates of drug metabolism and excretion, it is the sum total of these two processes which define a drug's clearance (elimination) rate. Clearance rate does not necessarily refer to the actual elimination of the drug from the body, since drug metabolism can convert drugs from pharmacologically active to pharmacologically inactive compounds. Thus the pharmacological activity of the drug may be eliminated even though the drug's metabolite is still present in the body. It is to be emphasized that most analytical techniques for routine therapeutic drug monitoring today measure the parent compound and do not measure drug metabolites. Therefore, a sudden change in steady-state drug concentrations usually serves as an indicator of altered drug disposition of the pharmacologically active compound.

For example, the addition of sodium valproate to the regimen of a patient who is receiving phenytoin may result in a marked decrease in total phenytoin concentrations. This decrease is a direct consequence of the displacement of phenytoin from its plasma protein-binding sites. The displaced phenytoin is rapidly converted to its inactive metabolite: *p*-hydroxyphenytoin. The observed fall in total phenytoin levels is indicative of an altered rate of phenytoin disposition.

D. Effect of Multiple Drug Therapy on Drug Disposition

Most clinically significant drug interactions are readily identifiable in the presence of elevated plasma concentrations of a given drug. Generally it will be observed that the interfering drug has a metabolic pathway which is similar to that of the drug being monitored. In addition, multiple drug therapy can also alter absorption and protein binding renal clearance of a given agent. Change in any of these factors can result in altered steady-state concentrations.

Any factor which alters drug half-life (clearance) will alter the drug's steady-state concentration. During multiple drug therapy, two drugs may compete for the same metabolic site. This competition will decrease the rate of metabolism of the drug which is excluded from the site and prolong its half-life. Since the half-life is prolonged (clearance is decreased), a new, higher steady-state drug concentration will be achieved and maintained, as long as the multiple drug therapy is continued. For this reason it is necessary to be able to identify all of the drugs which a patient is receiving and to indentify those agents which are potentially capable of altering a given drug's half-life (*1,9–12,31,32*).

E. Disease States and Alteration in Drug Disposition

Drug clearance is dramatically altered during renal and liver disease because the elimination rates of the drugs are changed. Consequently, new steady-state levels will be achieved which may differ significantly from those observed in healthy individuals. One must always consider the clinical status of the patient when interpreting plasma drug concentrations.

F. Alterations of Drug Disposition in Healthy Individuals Due to Pharmacogentics

Drug clearance is significantly regulated by genetic factors. In a large population of patients one could predict that if the entire population were given the same mg/kg dosage of a drug, there would be marked differences in the ability of individuals within the population to utilize the drug.

These genetic differences will be reflected in a marked variability of the steady-state plasma concentrations observed in this population. For example, in a population of patients receiving phenytoin at a standard therapeutic dose of 5 mg/kg/day, one would theoretically expect all patients to have a therapeutic drug level of 15 μg/ml (*29,31,32*).

In reality, plasma concentrations will range from 0 μg/ml, which suggests drug malabsorption, patient noncompliance, or fast drug metabolism, to levels of 40–50 μg/ml, which may indicate patients who exhibit hepatic or renal disease, drug interactions, or who are genetically slow drug metabolizers.

The importance of genetic regulation of drug metabolism as a determinent of each individual patient's drug utilization pattern cannot be neglected. As an example, consider the incidence of fast and slow metabolism in patients receiving isoniazid, a drug commonly used in the treatment of tuberculosis. Approximately 40% of all Caucasians are rapid acetylators of isoniazid. In contrast, over 90% of Japanese and Eskimos are rapid acetylators. This genetic variability requires individualization of therapeutic regimens to assure the maintenance of optimal isoniazid concentrations in the different populations and individuals within the population (*1,9–12,31,32*).

VI. OPTIMAL THERAPY REQUIREMENT OF A MINIMUM EFFECTIVE CONCENTRATION

Following the administration of any drug, there is always a peak level which represents the point of maximum absorption, and a trough (valley) level, which represents the lowest point achieved following a given dose of drug.

The trough level occurs after the absorption process is complete. It is the lowest point achieved as a consequence of the process of drug elimination (drug metabolism and excretion) which occurs during each dosage interval.

The object of all drug therapy is to assure that a minimum effective concentration (MEC) of drug is present throughout a dosing interval since the desired therapeutic effect will not be achieved at concentrations below the MEC. If plasma concentrations fall below the MEC during any given dosing interval, an exacerbation of the patient's clinical status can be expected. Therefore, the blood specimens for routine therapeutic monitoring should be drawn immediately prior to the next drug dose. The actual plasma concentration observed at that time should be within the optimal therapeutic range. It is imperative to recognize that the MEC of a given

drug necessary to achieve the desired therapeutic effect can vary from patient to patient. The severity of the disease process plays a major role in determining the MEC for many drugs. Therefore, it is possible to achieve a therapeutic response at suboptimal concentrations in some patients, whereas others will require toxic concentration to achieve control. Plasma concentrations must always be interpreted in conjuction with the clinical status of the patient and the desired therapeutic end point (*1*,*2*,*7*,*9*, and *10*).

VII. PEDIATRIC CLINICAL PHARMACOLOGY

The many factors involved in determing the steady-state plasma concentrations of a drug after a fixed dosage regimen have been reviewed above. Detailed information related to this subject is available in any medical pharmacology text (*9–12*). Ultimately, the pharmacological (or biological) effect observed clinically after a given drug dosage is determined by the concentration of drug at the receptor site. This concentration can be changed by altering the rate of drug absorption, the degree of protein binding, the rate of drug metabolism, or the rate of renal excretion of the drug. Such changes can occur as a consequence of normal physiological development, a change in physiological systems as a consequence of disease, or alteration by administration of a pharmacologically active agent.

It is to be noted that essentially from birth until death the efficiency of the physiological systems constantly decreases in any given individual.

TABLE II

Factors Affecting Drug Absorption in Children

Physiological variables	Newborns	Infants	Children
Gastric emptying time	Increased (6–8 hr)	Increased (6–8 months' decrease)	Decreased
Intestinal motility	Decreased, irregular	Increased	Increased
Absorbing gastrointestinal surface	Reduced	Normal	Normal
Biliary function	Not fully developed	Developed	Developed
Muscular blood flow	Decreased	Increased	Increased
Microbial flora	Colonization phase	Adult pattern	Adult pattern
Skin permeability	Increased	Increased	Normal

TABLE III

Factors Affecting Drug Protein Binding in Children

Physiological variables	Newborns	Infants	Children
Total plasma protein	Reduced	Reduced	Normal
Plasma albumin	Reduced	Normal	Normal
Fetal albumin	Present	Absent	Absent
Plasma albumin	Present	Absent	Absent
Plasma globulins	Reduced	Reduced	Normal
Serum bilirubin	Elevated	Normal	Normal
Free fatty acids	Elevated	Normal	Normal
Blood pH	Reduced	Normal	Normal

This continual decrease occurs rapidly during some stages of growth and development, and slowly during others. Age-related differences in drug disposition are a reflection of changes in normal physiological function. These age-related changes in drug disposition are presented schematically in Fig. 1, in which the shadow represents the rate of drug disposition with respect to body size during various stages of development. It is to be emphasized that although the rate of drug disposition in children is increased, the optimal drug concentration of most therapeutic agents necessary to produce the desired therapeutic response is similar to that observed in adults. Therefore, because of the faster drug clearance in children, it is necessary to prescribe larger drug doses to the pediatric population in order to achieve and maintain optimal drug concentrations (*6,7, 10*). Tables II–V list the physiological variables which affect drug disposition in children.

TABLE IV

Factors Affecting Metabolic Degradation of Drugs in Children

Physiological variables	Newborns	Infants	Children
$\frac{\text{Liver weight}}{\text{Body weight}}$ ratio	Increased	Increased	Increased
Protein Y	Absent	Present	Present
Hepatic microsomal activities	Decreased (?)	Increased (?)	Increased (?)
Hepatic blood flow	Reduced Increased	Increased	Increased
Blood esterase activities	Reduced	Normal	Normal
Synthetic reactions	Reduced	Normal	Normal
β-Glucuronidase	Increased	(?)	Normal

TABLE V

Factors Affecting Renal Excretion of Drugs in Children

Physiological variables	Newborns	Infants	Children
Renal blood flow	Reduced or increased	Increased	Normal
Urinary circadian rhythm	Absent	Present (?)	Present
Glomerular filtration	Reduced	Reduced or normal	Normal
Tubular secretion	Reduced	Normal	Normal
Tubular reabsorption	Reduced or normal	Increased or reduced Normal	Normal
Urinary pH	Acid	Fluctuates	Fluctuates

A. Drug Disposition in the Neonate

During pregnancy, a limited capacity for drug metabolism develops in the fetal liver. However, the elimination of drugs from the fetus depends primarily on the mother's capacity to metabolize and excrete foreign compounds. Trans-placental equilibrium of drugs ingested by the mother depends on the physical properties of the drug, but in general it is rapid. However, one should remember that although the total concentration of a drug in plasma may be similar in fetal and maternal plasma, the concentration of the free drug may differ significantly, because fetal albumin does not have the same binding affinity for drugs as does the albumin in the plasma of adults. Generally, free drug concentrations are higher in neonates (*7,8*).

From the moment of birth the neonate must depend on its own mechanisms for drug disposition. During the first postnatal week, the rates of drug metabolism and of drug elimination are both very slow. The slow rate of drug disposition in the neonate is a direct consequence of hepatic and renal-functional immaturity, and it is reflected in a prolonged biological half-life of the drug. It is to be noted that the rate of development of drug disposition patterns in premature infants is significantly decreased from that observed in full-term infants.

Throughout the second to fourth weeks of postnatal life, an increasing rate of drug disposition develops as these systems mature, and this is reflected by a marked decrease of drug half-lives.

B. Drug Disposition in the Infant

As the physiological maturation process continues through the early stages of infancy, drug disposition rates reach the highest levels they will achieve throughout the entire life span. Unfortunately, there are few studies of the changes in the pharmacokinetic patterns associated with infancy. However, those studies that are available indicate increased rates of drug disposition. After 8–12 weeks of age, the changes in drug kinetics as compared to the changes observed in neonates are more gradual (*7,8*).

C. Drug Disposition in Prepubescent Children

It is an often-stated axiom that children utilize drugs twice as fast as adults. This is generally true in children between the ages of 6 and 10. Therefore, as a rule, many clinicians prescribe approximately twice the adult dosage (per unit of body weight) of a given drug to pediatric patients in order to achieve the same optimal plasma concentrations. Again, detailed studies of the pharmacokinetic patterns observed in this age group are sparse. In general, throughout this age range, there is a decrease in the rate of drug disposition as compared to these rates observed in infants (*6–8*).

D. Drug Disposition in Pubescence

Rapid physiological changes are associated with the onset of puberty. These changes are highly sex- and age-dependent. The individual variability in the onset of puberty is well-known. Therefore, it is difficult to predict exactly when the drug-disposition pattern will change in a pubescent child unless the patient is closely followed by routine therapeutic drug monitoring.

A major question related to pediatric drug utilization remains unresolved: When does a child's drug disposition pattern change from those of a child to those of an adult? Evidence based on therapeutic drug monitoring is beginning to emerge that strongly supports the concept that drug utilization in children changes rapidly during a few months. These changes are directly associated with the initial onset of puberty. Further study of this particular question is urgently needed. However, it can be stated with some degree of confidence that as a child is entering the earliest stages of puberty, the changes in drug metabolism appear to be directly correlated to changes in endocrine activity (*6*).

Obviously, the time at which this conversion will occur varies dramatically with respect to sex and age. The changes in drug-utilization patterns

are observed earlier in girls than in boys. However, the time required for the drug utilization patterns to change significantly depends on the individual characteristics of each child. Therefore, there is no hard and fast rule which allows one to predict the exact time at which this change will occur. In general, the conversion in drug-disposition patterns to patterns typical of adults is most likely to occur at the initial onset of the physiological changes associated with the onset of puberty.

This conversion to adult drug-metabolizing patterns is probably directly attributable to the increasing concentrations of sex hormones. The onset of puberty is characterized by increased rates of synthesis and maintenance of higher steady-state concentrations of sex hormones. A competition between sex steriods and drugs for the same metabolic sites in the hepatic microsomal enzymes may be the explanation for the decreased drug clearance rates associated with pubsecent changes. Because drugs are not a natural substrate of the microsomal enzyme system, one would expect that rates of drug metabolism would be decreased as the demand for metabolism of reproductive steriods (the natural substrates which possess a greater affinity for this system) increases. Indirect clinical evidence supports this hypothesis, but studies to elucidate the mechanisms of this phenomenon have not yet been reported (*6,7*).

As a consequence of the fact that the exact time interval during which utilization in pubescent children undergoes conversion to adult drug-disposition patterns cannot be clearly identified, special attention is drawn to this group of children. Any laboratory engaged in routine therapeutic drug monitoring or any clinician treating children entering puberty should be cognizant of the fact that during this stage of development fairly rapid fluctuations in drug concentrations in the plasma can be expected without any significant change in either physical factors such as age and weight or the prescribed drug dosage (mg/kg body weight).

Because of the changes in drug-disposition patterns observed in children of this age group, both the clinician and the laboratory engaged in routine monitoring should be alert for values that would not be expected on the basis of the prescribed drug dosage. To assure that no child unnecessarily develops a drug intoxication as a consequence of changing metabolic patterns, we recommend, as the child approaches puberty, routine monitoring of drug concentrations in any child on chronic drug therapy at least once every four months until drug disposition patterns have stabilized. Careful therapeutic drug monitoring will alert the clinician to change, and dosages can be adjusted to compensate for the decreased rate of drug disposition.

An example of altered drug disposition in pubescent children is encoun-

tered in the management of the epilepsies. With the onset of puberty, increased lethargy and dullness may be observed clinically. We suggest that many of the clinical problems associated with seizure control in pubescent children may be directly related to altered drug utilization patterns. The importance of this problem is demonstrated by the following hypothetical example based on our clinical experience (*6*).

Consider a female who, at the age of 9 is placed on the usual pediatric dosage of phenobarbital (6 mg/kg) and whose serum phenobarbital concentration is 30 μg/ml: CDR = [(5 μg/ml)/(mg/kg)] × 6 mg/kg/day = 30 μg/ml. Drug concentrations are carefully monitored, seizures are well controlled, and she is seen by her pediatrician once a year. Between her yearly visits, the parents and teachers become aware of a gradually increased lethargy in the child. At her next yearly examination, the serum phenobarbital concentrations are found to have increased to 60 μg/ml, reflecting the characteristic adult values observed with a phenobarbital dosage of 6 mg/kg: CDR = [(10 μg/ml)/(mg/kg)] × 6 mg/kg/day = 60 μg/ml. This increased phenobarbital concentration is attributable directly to a decreased drug clearance associated with the onset of puberty.

Chronic phenobarbital toxicity is subtle, often occuring without dramatic clinical changes, and may go undetected by the casual observer. As a consequence of a child's potential ability to learn (cognitive function) being diminished by increasing lethargy, antiepileptic drug concentrations should be monitored frequently (at least every 3 months), especially approaching early pubescence, regardless of whether or not the patient is seen by a physician. If serum drug concentrations increase, it is recommended that the dosage be appropriately regulated without delay. If such a procedure is followed, unnecessary phenobarbital toxicity can be prevented and the subtle side effects of drug intoxication can be completely avoided.

E. Drug Disposition in Adolescence

By the time the physical secondary sex characteristics have begun to appear in a child, and throughout the remaining adolescent period, observed adolescent drug-disposition patterns are essentially identical to those of an adult. Therefore, medication regimens should generally be based upon adult criteria. It is to be emphasized that the marked individual variability in the onset of puberty and thus alteration of drug disposition patterns exists. Therefore, careful monitoring of drug concentrations in early adolescence can be particularly valuable in regulating therapy (*6,77*).

VII. GUIDE TO THERAPEUTIC DRUG MONITORING

There are a number of advantages to TDM which provide the clinician with clinically useful information. Plasma drug concentrations in conjuction with a thorough assessment of the patient's clinical status and the therapeutic goals to be achieved provide a means of successfully and rapidly individualizing a patient's therapeutic regimen to assure optimal benefits with minimal risk.

1. Noncompliance can be identified. Many patients, in particular those who have a chronic disease requiring therapy over a prolonged period of time, tend not to take their medications as prescribed. Moreover, patients with a chronic disease that does not necessarily cause pain or other unusual discomfort (for example, the epilepsies, asthma, or hypertension), may easily neglect to take their medicine. The end result of such noncompliance is an exacerbation of the existing disorder some time in the future. Studies have clearly demonstrated that noncompliance is a major factor in treatment failures (*9–11*).

2. Individual variations in drug utilization patterns can be dealt with appropriately. In any population of individuals, a drug dosage based solely on body weight results in a fixed steady-state serum concentration. If the plasma concentrations following a specific dosage are analyzed in a large patient population, however, the distribution of drug levels will be Gaussian (a bell-shaped curve). The vast majority of patients will show levels within the range expected for a given mg/kg dosage. Patients who are genetically either "fast" or "slow" drug metabolizers will have levels at extreme ends of the curve. The fast drug metabolizers require significantly higher doses to achieve the same plasma concentrations and consequently the desired therapeutic effect. Patients who are slow drug metabolizers become intoxicated and experience side effects from standard therapeutic doses of the drugs; therefore, optimal drug levels can be maintained in these patients with dosages well below the standard regimen.

Therapeutic drug monitoring allows identification of individuals who are fast or slow metabolizers and ensures that their medication regimens can be appropriately adjusted to fit their own metabolic patterns. Without TDM, a prolonged period (sometimes lasting months) of trial-and-error therapy is required to achieve the appropriate dosage regimen, thus unnecessarily subjecting the patient to a time interval when the disease process is uncontrolled (*1* and *28,29*).

A common factor characteristic of a few drugs, including phenytoin and aspirin, associated with abnormal drug elevations, is the phenomenon of saturation (zero-order) kinetics. In TDM this is often reflected by a rapid

rise in serum drug concentrations following a small increment in drug dosage. Thus, a small dosage increment will cause a large increase in serum concentration. The therapeutic index, that is, the margin of safety between therapeutic and toxic drug concentrations, is extremely small for many drugs. If the enzymes responsible for the metabolism of these drugs become saturated, drug intoxication can develop quickly.

3. Altered drug utilization as a consequence of disease can be readily identified. Patients on long-term drug therapy may become acutely ill and require the administration of additional therapeutic agents. Drug interactions may then cause these patients to respond in an unexpected manner to a fixed dosage of some adjunctive therapy. Acute or chronic uremia can dramatically decrease the elimination of a drug that is primarily dependent on urinary excretion, and renal failure can alter the protein-binding characteristics of many drugs to albumin. In both situations the ratio of free drug to total drug may increase to the point where free drug concentrations are high enough to produce a clinically evident toxic drug response, although the total serum drug concentrations are well within optimal therapeutic range. Hepatic disease can extensively alter a given therapeutic response by impairing a patient's ability to metabolize drugs. Most drugs depend on liver detoxification for conversion to water-soluble products, which are easily eliminated from the body. Thus, a precipitous rise in parent drug concentrations can occur as the unmetabolized drug, which normally would have been eliminated from the system, accumulates (*9–11,27*).

Therapeutic drug monitoring provides a means of accurately calculating and correcting dosage regimens to coincide with the disease status of the patient. For example, if the patient is in chronic hepatic failure, it may be desirable to administer a drug that would normally be administered daily (such as phenytoin), only once every 4 or 5 days, thereby maintaining therapeutic concentrations without producing toxicity.

4. An altered physiologic state can be compensated for. Normal alterations in physiologic state also change drug utilization patterns. Three areas in which TDM is crucial to successful dosage regimen adjustments should be emphasized. Recent studies have shown that decreased drug absorption during pregnancy is associated with a dramatic fall in serum phenytoin concentration and exacerbation of seizures in epileptic gravidas. The use of TDM from the onset of pregnancy, with appropriate dosage regulation to maintain therapeutic drug concentrations, significantly decreases the number of seizures that occur, thus decreasing the potential harm to the fetus.

Most importantly, the normal process of maturation involves a large number of physiologic changes that can dramatically alter drug utiliza-

tion. Children utilize drugs at a faster rate than adults, and therefore, as a rule, require almost twice as much drug on a body weight basis as does an adult to achieve the same therapeutic drug concentration. As a child enters puberty, his or her drug utilization patterns rapidly change to those of adulthood, to the extent that by early pubescence the conversion to adult patterns is complete. These changes usually occur between the ages of 10 and 13, appearing earlier in girls than in boys. It is imperative that TDM be carried out carefully for any drug administered chronically to early pubescent and pubescent children. Failure to adjust the child's therapeutic regimen to compensate for the associated physiologic changes may result in exposure to unnecessary and prolonged drug toxicity, with its attendant sequelae (*6,7*).

As the maturation process continues, however, the efficiency of normal physiologic functions decreases, as does the ability to bind drugs to plasma protein. Geriatric patients often exhibit reduced rates of drug elimination, thereby requiring reduced drug dosages. It is possible for geriatric patients to have total drug plasma concentrations within the optimal therapeutic range, but elevated free drug concentrations that can produce adverse side effects. The clinical signs of drug intoxication in the elderly often present clinically as lethargy and confusion, and TDM provides a means of distinguishing drug-induced confusion from organic deterioration (*12*).

Interpretation of Plasma Concentration

Anyone involved in the utilization of information derived from therapeutic drug monitoring must always bear in mind that the interpretation of plasma drug concentrations must *always* be carried out in conjunction with an assessment of the clinical status of the patient! Therapeutic ranges

TABLE VIA

General Factors Influencing Interpretation of Assay Data or Therapeutic Drug Monitoring

Patient compliance, including dosage error and wrong medication
Absorption via route of administration
Drug distribution
Biotransformation
Excretion
Genetic variability
Pathophysiologic factors
Drug interactions
Drug tolerance
Inappropriate drug effects

TABLE VIB
Information Needed for Interpretation of Drug Levels

Patient age, weight, and sex
List of all of the drugs which the patient is receiving
Total daily dose of all drugs
Dosage regimen and dosage form of each drug
Time the last dose of drug, the level of which is being requested, was administered
Time the specimen was drawn
Clinical status of the patient

should more correctly be described as optimal concentrations. The optimal concentration (therapeutic range) of a drug is defined as that concentration of drug present in plasma or some other biological fluid or tissue which provides the desired therapeutic response in most patients. It is to be emphasized that the severity of the disease process determines the amount of drug necessary to achieve a given therapeutic effect. Thus, it is quite possible that a given patient may achieve the desired therapeutic effect at a plasma concentration well below the optimal range. Conversely, some patients will not achieve the desired therapeutic effect even when plasma concentrations are elevated into the toxic range. If the desired therapeutic effect is achieved at suboptimal plasma concentrations, every attempt should be made to avoid the prescription of additional drugs simply to increase the plasma concentration into what is commonly referred to as the therapeutic range. Obviously, the interpretation of plasma drug concentration must take into account the various factors which can alter the steady-state plasma concentration achieved on a given dosage regimen. These factors are summarized in Tables VIA and B.

TABLE VII

Indications for Monitoring Plasma Drug Levels

Plasma drug levels should be monitored for the following reasons:
When a drug has a narrow, well-defined therapeutic range
When noncompliance is suspected
When the desired therapeutic effect is not achieved or when symptoms to toxicity are observed
When there are large inter-individual variations in drug utilization or metabolism
When drug utilization is altered as a consequence of secondary disease or physiological state
When drug interactions are suspected
When there is a need for medico–legal verification of treatment

TABLE VIII

Time to Draw TDM Specimens

Selection of the time a specimen is drawn in relation to drug administration should be based on the pharmacokinetic properties of the drug and dosage form.
Patient should be at or near steady state when sample is drawn. This is achieved when the drug has been administered at a constant rate for 3 to 5 half-lives.
After dosage adjustment, time should be allowed for equilibrium to be re-established with the new dosage regimen before another specimen is drawn.
Specimens drawn immediately before administration of the next oral dose provide trough serum levels for drugs administered on a chronic basis; the trough level should ideally be above the minimum effective serum level.
Specimens for peak levels are generally drawn 15–30 min after i.v. administration, 1–2 hr after i.m. administration, and 1–5 hr after oral administration (depends on rate of drug distribution).
When the specimen is going to be drawn during an infusion, the sample should be taken from the opposite limb.

The question of when to monitor a patient's plasma concentrations often arises. As a general rule, monitoring should be undertaken whenever there is a marked change in the clinical status of the patient. Indications for monitoring are summarized in Table VII.

Another question which often arises is the time of drawing the specimen (see Table VIII). As a rule, specimens for therapeutic drug monitoring should be drawn at a trough value, since this concentration represents the lowest which will be observed in a dosing interval. Measurement of peak concentrations following oral administration is difficult because of the marked individual variability in drug absorption patterns. Peak levels are indicated in certain situations following i.v. drug administration. Peak concentrations have been reported to be of value in the monitoring of antibiotics, theophylline, and certain antiarrhythmic drugs.

The clinician should be aware of the pharmacological properties of those drugs which are theapeutically monitored. Table IX presents the clinically significant properties of these agents. Information contained within the table is based on information derived from large patient populations in the case of adults and often of very small populations of children. Although population studies provide useful general guidelines, successful TDM is based on individualization of drug therapy based upon assessment of the plasma concentration, the clinical status of the patient, and therapeutic goals. The clinician's ability to weigh the significance of each of these factors in a given patient and then establishing an appropiate therapeutic regimen assures optimal therapeutics (*1,2*).

TABLE IX

Pharmacological Parameters for Commonly Prescribed Drugs

Drug	Recommended dose (mg/kg/day)		Optimal range (μg/ml)		Toxic level (μg/ml)		Percentage dose absorbed	Time to peak plasma (hours)	Percentage protein bound	Volume of distribution (l/kg)	Half-life (hours)		Time to steady-state	
	Adults	Children	Adults	Children	Adults	Children					Adults	Children	Adults	Children
Antiarrhythmics														
Digixon	3–5	10–15	.08–2.0 (ng/ml)	0.8–2.0 (ng/ml)	>2.4 (ng/ml)	>2.4 (ng/ml)	60–75	1.5–5.0	20–40	5.0–10.0	36–51	11–50	7–11 days	2–10 days
Disopyramide	8.6	n/a[a]	2/5	n/a	>7	n/a	80	0.5–3.0	10–80	0.8	5–6	n/a	25–30 hr	n/a
Lidocaine	1.5	n/a	1.5–5	n/a	7	n/a	—	15–30 min	50–70	1.7	1–2	n/a	5–10 hr	n/a
Procainamide	32	n/a	4–10	n/a	>16	n/a	70–95	1–2	15	1.7–2.2	2.2–4.0	n/a	11–20 hr	n/a
								1–2	90–96	2–4.6	2–6	n/a	10–30 hr	n/a
Quinidine	10–20	n/a	2–5	n/a	>10	n/a	98	1.5–2	80–90	0.5–2.6	4–7	n/a	20–35 hr	n/a
Antibiotics														
Amikicin	10–15	10–15	10–25	10–25	>35 (peak) >5 (trough)	>35 (peak) >5 (trough)	—	Varies with dosage and clinical status	0–11	0.28	2–3	n/a	10–15 hr	n/a
Chloramphenicol	50–100	25 (<1 month) 50 (>1 month)	10–20	10–20	>25	>25	75–90	2	60–80	0.57	1.5–5	n/a	7.5–25 hr	n/a
Gentamycin	3–5	6–7.5	5–10	5–10	>12 (peak) >2 (trough)	>12 (peak) >2 (trough)	—	Varies with dosage and clinical status	0–10	0.25	2–3	2–3	10–15 hr	10–15 hr
Tobramycin	3–5	3–5	5–10	5–10	>12 (peak) >2 (trough)	>12 (peak) >2 (trough)	—	Varies with dosage and clinical status	0–10	0.22	2–3	n/a	10–15 hr	n/a

Antidepressants														
Amitriptyline	0.7	n/a	120–250 (ng/ml)	n/a	>500 (ng/ml)	n/a	56–70	2–8	82–96	n/a	17–40	n/a	4–8 days	n/a
Desmethylipramine	1–2	n/a	150–250 (ng/ml)	n/a	>500 (ng/ml)	n/a	90	2–8	73–92	28–60	12–54	n/a	2.5–11 days	n/a
Imipramine	0.7–1.4	n/a	150–250 (ng/ml)	n/a	>500 (ng/ml)	n/a	29–77	1–2	80–95	10–20	9–24	n/a	2–5 days	n/a
Lithium	10–20	n/a	0.8–1.4 (mEq/ liter)	n/a	2.0 (mEq/1)	n/a	97	1–3	0	0.4–1.4	8–35	n/a	2–7 days	n/a
Nortriptyline	0.7–1.4	n/a	50–150 (ng/ml)	n/a	>500 (ng/ml)	n/a	46–79	4–8	93–95	20–40	18–93	n/a	4–19 days	n/a
Antiepileptics														
Carbamazepine	10–20	10–20	8–12	8–12	>15	>15	70–80	6–18	65–85	0.8–1.4	10–30	8–19	2–6 days	2–4 days
Ethosuximide	20–30	30–60	40–100	40–100	>150	>150	100	1–2	0	0.7–0.9	40–60	30–50	8–12 days	6–10 days
Phenobarbital	2–4	4–8	15–40	15–40	>50	>50	80–100	6–18	45–50	0.7	50–120	40–70	11–25 days	8–15 days
Phenytoin	5–6	5–10	10–20	10–20	>20	>20	90	4–8	87–93	0.5–0.8	18–30	12–22	4–6 days	2–5 days
Primidone	10–20	15–30	5–12	5–12	>15	>15	80–90	2–4	0–10	0.6–1.0	3.3–2.15	4–6	16–60 hr	20–30 hr
Valproic acid	30–60	30–60	50–100	50–100	>200	>200	85–100	0.5–1.5	90–95	0.15–0.40	8–15	6–15	40–75 hr	30–75 hr
Miscellaneous														
Methotrexate	depends on route of administration		depends on therapeutic regimen				varies with dosage	1–2	50–70	0.75	varies	varies	varies	varies
Theophylline	13–18	16–24	10–20	10–20	>20	>20	95–100	2–3	55–65	0.3–0.7	3–8	1–8	15–40 hr	5–40 hr

VIII. AN OVERVIEW OF THE ROLE OF THE CLINICAL LABORATORY IN THERAPEUTIC DRUG MONITORING

Rapid advances in clinical pharmacology over the past decade are directly attributable to TDM; the availability and clinical utility of TDM in turn is directly related to the rapid advancement in clinical pharmacology associated with the quantitation of drug compounds. Once the basic methodologies for the analysis of drugs in plasma became available, u.v. spectrophotometric studies correlating drug concentration with therapeutic effects were conducted in the late 1950s and early 1960s. These procedures required large sample volumes, extraction techniques were time-consuming and complex, and the assays were subject to many interferences; therefore drug assays were usually conducted in basic research laboratories rather than in routine clinical chemistry laboratories.

Not until the late 1960s did TDM become widespread. Gas–liquid chromatography (GLC) represented a major breakthrough because it provided a method of separating classes of drugs, as well as individual drugs within a class, rapidly and quantitatively at the same time. Gas–liquid chromatographic techniques were further refined and improved so that by the early 1970s GLC analysis of various therapeutically monitored agents was performed routinely in many clinical chemistry laboratories. One of the major disadvantages of GLC had been the complexity of the instrumentation, which necessitated a highly trained and skilled analyst. More recent advances in the development of detectors, particularly the nitrogen–phosphorus detector, have increased the sensitivity of the instruments to such an extent that microsampling by GLC is now possible on a routine basis. Nitrogen detection serves as a successful means of monitoring nanogram quantities of drugs and has been applied to TDM of antiepileptic, antiarrhythmic, and antidepressant drugs (*37*).

The development of radioimmunoassay techniques permitted quantitation of drug concentrations in microvolumes of serum. Unfortunately, however, the complexity of the technique as well as the lack of radioimmunoassays for a wide variety of drugs prevent its widespread application to routine monitoring (*38*).

Making TDM available to all laboratories and physicians required a simple technology that could be performed by a technician without special training or instrumentation. This was achieved with the development of the homogeneous enzyme immunoassay system (EMIT) which is capable of performing five drug assays on a 50-μl serum specimen. Once the initial daily calibration is complete, each drug assay can be performed in 2 min. The major advantages of the system are its microcapability, accuracy, and the rapidity and ease of operation of the assays. The disadvan-

tage of the system is that it is limited to those drugs for which antibodies are available (*1*).

A large number of drugs exist for which antibodies are not available, but which must be therapeutically monitored. The most promising and practical method of monitoring these agents is by high-pressure liquid chromatography (HPLC). Within the last 5 years the development of HPLC has provided laboratories with a system having the same advantages as the homogeneous enzyme immunoassay system; it is capable of processing microsamples (100 μl), is rapid and specific, and the instrumentation is relatively simple to operate. In addition, HPLC can be adapted to simultaneously quantitate a large variety of drugs as well as their active metabolites. High-pressure liquid chromatography permits simultaneous drug analysis and is a valuable tool for establishing correlations between drug and drug metabolite concentrations in biological fluids. Obviously, over the next few years, HPLC will become of even more use in routine TDM (*1*).

ACKNOWLEDGMENTS

The material presented in this chapter represents a compilation of reviews by the author published over the past 3 years. Extensive portions of these reviews have been reprinted. The author gratefully acknowledges the cooperation of the American Association for Clinical Chemistry, Raven Press, and the Syva Company for allowing the reproduction of materials contained in their publications.

REFERENCES

1. Pippenger, C. E., *Ther. Drug Monit.* **1**, 3–9 (1979).
2. Koch-Weser, J., *Ther. Drug Monit.* **3**, 3–16 (1981).
3. Baer, D. M., and Dito, W. R. (eds.) "Interpretations in Therapeutic Drug Monitoring," American Society of Clinical Pathologists, Chicago, 1981.
4. Sherwin, A. L., "Clinical Pharmacology of Ethosuximide in Antiepileptic Drugs: Quantative Analysis and Interpretation" (C. E. Pippenger, J. K. Penny, and H. Kutt, eds.) pp. 283–295, Raven, New York, 1978.
5. Morselli, P. L. (ed.) "Drug Disposition During Development," Spectrum Publ., New York, 1977.
6. Pippenger, C. E., "Pediatric Clinical Pharmacology of Antiepileptic Drugs: A Special Consideration in Antiepileptic Drugs, Quantitative Analysis and Interpretation, (C. E. Pippenger, J. K. Penny, and H. Kutt, eds.) pp. 315–319, Raven, New York, 1978.
7. Morselli, P. L., "Antiepileptic Drugs in Drug Disposition During Development" (P. L. Morselli, ed.) pp. 311–360. Spectrum Publ., New York, 1977.
8. Mirkin, B. L. (ed.), "Perinatal Pharmacology and Therapeutics," Academic, New York, 1976.

9. Avery, G. S. (ed.) "Drug Treatment: Principles and Practice of Clinical Pharmacology and Therapeutics," 2nd ed., ADIS, Sydney, 1980.
10. Melmon, K. L., and Morelli, H. I. (eds.), "Clinical Pharmacology: Basic Principles in Therapeutics" 2nd ed., Macmillan, New York, 1978.
11. Gilman, A. G., Goodman, L. S., and Gilman, A., "The Pharmacological Basis of Therapeutics," 6th ed., Macmillan, New York, 1980.
12. O'Malley, K., Judge, T. G., and Crooks, J., "Geriatric Clinical Pharmacology and Therapeutics in Drug Treatment," G. S. Avery (ed.), 2nd ed., ADIS, Sydney, 1980.
13. Goldstein, A., Aronow, L., and Kalman, S. M., "Principles of Drug Action: The Basis of Pharmacology," 2nd ed., Wiley, New York, 1974.
14. Mayer, S. E., Melmon, K. L., and Gilman, A. G., "Introduction, The Dynamics of Drug Absorption, Distribution, and Elimination in the Pharmacological Basis of Therapeutics," A. G. Gilman, L. S. Goodman, and A. Gilman (eds.), pp. 1–27. Macmillan, New York, 1980.
15. Glaser, G. H., Penny, J. K., and Woodbury, D. M. (eds.), "Antiepileptic Drugs: Mechanisms of Action," Raven, New York, 1980.
16a. Koch-Weser, J., and Sellers, E. M., *N. Engl. J. Med.* **294,** 311–316 (1976). *N. Engl. J. Med.* **294,** 526–531 (1976).
16b. Koch-Weser, J., *N. Engl. J. Med.* **291,** 234 (1974).
17. Levy, R. H., *Ther. Drug Monit.* **2,** 199–201 (1980).
18. Sjoqvest, F., Borga, O., and Orme, M. L. E., "Fundamentals of Clinical Pharmacology in Drug Treatment," (G. S. Avery, ed.), 2nd ed., pp. 1–61. ADIS, Sydney, 1980.
19. Anton, A. H., and Solomon, H. M. (eds.) *Ann. N.Y. Acad. Sci.* **226,** 1–362 (1973).
20. Benet, L. Z. (ed.), "The Effect of Disease States on Drug Pharmacokinetics," Am. Pharm. Assoc., Washington, D.C., 1976.
21. McAuliffe, J. J., Sherwin, A. L., Leppick, I. E., Fayle, S. A., and Pippenger, C. E., *Neurology*, **27,** 409–413 (1977).
22. Danhof, M., and Breimer, D. D., *Clin. Pharmacokinet.* **3,** 39–57 1978.
23. Pippenger, C. E., *Pediatr. Clin. North Am.*, **27,** 891–925 1980.
24. Garlock, C. M., Pippenger, C. E., Desaulniers, C. W., and Sternberg, S., *Clin. Chem.* **25,** 1117 1979.
25. Lu, A. Y. H., *Fed. Proc.* **35,** 2460–2463 1976.
26. Conney, A. H., *N. Engl. J. Med.* **280,** 653–660 1969.
27. Conney, A. H., *Pharmacol. Rev.* **19,** 317–366 1967.
28. Kutt, H., "Evaluation of Unusual Antiepileptic Drug Concentrations in Antiepileptic Drugs: Quantitative Analysis and Interpretation," (C. E. Pippenger, J. K. Penny, and H. Kutt, eds.), pp. 307–314. Raven, New York, 1978.
29. Kromann, N., Christiansen, J., Flachs, H., Dom, M., and Hvidberg, E. F., *Ther. Drug Monit.* **3,** 239–246 1981.
30. Winter, M. E., "Basic Clinical Pharmacokinetics," Applied Therapeutics, San Francisco, 1980.
31. Rowland, M., and Tozer, T. N., "Clinical Pharmacokinetics: Concepts and Applications," Lea and Febiger, Philadelphia, 1980.
32. Wagner, J. G., "Fundamentals of Clinical Pharmacokinetics," Drug Intelligence Publications, Hamilton, Illinois, 1975.
33. Pippenger, C. E., Penny, J. K., and Kutt, H. (eds.), "Antiepileptic Drugs: Quantitative Analysis and Interpretation," Raven, New York, 1978.
34. Painter, M. J., Pippenger, C. E., and Pitlick, W., *J. Pediatr.*, **92,** 315–319 1978.
35. Van den Herivel, W. J., *Adv. Chromatogr.*, **13,** 265–303 1975.
36. Cook, C. E., "Radioimmunoassay in Antiepileptic Drugs: Quantitative Analysis and In-

terpretation," (C. E. Pippenger, J. K. Penny, and H. Kutt, eds.), pp. 163–173. Raven, New York, 1978.
37. Maggio, E. T. (ed.) "Enzyme Immunoassay," CRC Press, Boca Raton, Florida, 1980.
38. Hawk, G. L. (ed.), "Biological/Biomedical Applications of Liquid Chromatography," Marcel Dekker, New York, 1979.
39. Pippenger, C. E., "Principles of Therapeutic Drug Monitoring," *Syva Monitor,* Syva Company, Palo Alto, California, 1981.

Index

B

C

G

I

Q

R

S

U

V