Current Topics in Microbiology and Immunology

139

Carbohydrate-Protein Interaction

Edited by
A.E. Clarke and I.A. Wilson

With 35 Figures

Springer-Verlag
Berlin Heidelberg NewYork
London Paris Tokyo

ADRIENNE E. CLARKE, Ph.D.
Plant Cell Biology Research Centre
School of Botany
University of Melbourne
Parkville, Victoria 3052, Australia

IAN A. WILSON, D. Phil.
Department of Molecular Biology
Research Institute of Scripps Clinic
10666 North Torrey Pines Road
La Jolla, Ca 92037
USA

ISBN-13: 978-3-642-46643-4 e-ISBN-13: 978-3-642-46641-0
DOI: 10.1007/978-3-642-46641-0

Softcover reprint of the hardcover 1st edition 1988

Library of Congress Catalog Card Number 15-12910

2123/3130-543210 – Printed on acid-free paper

Preface

Carbohydrate as the primary product of photosynthesis has a vital role in the maintenance of life on this planet. Until relatively recently, interest in complex carbohydrates focussed on their structural role in the extracellular matrix/cell wall of animal, plant, and microbial cells and on their role as energy sources (e.g., starch and glycogen) and structural components (e.g., cellulose) in natural products. There was, however, indirect evidence that carbohydrates could play an informational role; this evidence was from the finding last century that plant lectins caused specific agglutination of certain animal cells and, more recently, that the agglutination was mediated by interactions between the plant lectin and cell surface carbohydrates.

It is now clear that endogenous carbohydrate binding proteins are important in cell-cell recognition phenomena in animal systems. Recently, impressive evidence has been presented that complex oligosaccharides, derived from cell walls, are also important in plant recognition events, for example in signalling the defence mechanisms of a plant to respond to attack by insects and microbial pathogens. Plant biologists have consequently become interested in the ways in which these oligosaccharides can interact with other molecules. Another aspect of biology which has generated interest in carbohydrate-protein interactions is the finding that cell surface saccharides are antigenic in animals and that in some diseases the cell surface antigens of the pathogen are the antigenic determinants. Thus, there is wide interest in complex carbohydrates and the way in which they interact with other cellular components.

The purpose of this book is to bring together information on the interaction of carbohydrates with proteins which will be of interest to all biologists, regardless of whether their experimental interests are in plant, microbial, or animal systems. We start with a review of the structure of carbohydrates found in plant and animal systems. This is followed by chapters on the carbohydrate binding sites of lectins, monoclonal antibodies, enzymes, and sugar

transport proteins. This collection of papers gives a comprehensive account of carbohydrate-protein interactions and will be a valuable resource for cell and molecular biologists and structural biochemists alike.

March, 1988

ADRIENNE E. CLARKE
IAN A. WILSON

Table of Contents

Indexed in Current Contents

List of Contributors

You will find the addresses at the beginning of the respective contribution

Complex Carbohydrates of Plants and Animals – A Comparison

Paul A. GLEESON

1 Introduction

Carbohydrate performs a diverse range of cellular functions, from those as structural components and storage macromolecules to playing a central role in cellular development and cell–cell recognition. This functional diversity is paralleled by the large variety of macromolecules that contain carbohydrate. These can be divided into four main groups: (a) polysaccharides, (b) proteoglycans, (c) glycoproteins, and (d) glycolipids. Although some "classical" polysaccharides are now known to contain small amounts of covalently bound protein, for example, glycogen (BUTLER et al. 1977; RODREGUEZ and WHELAN 1985), the common names are used in this review. The term proteoglycan refers to polysaccharide chains covalently attached to protein, and the term glycoprotein refers to monosaccharide residues or oligosaccharide chains covalently attached to protein (HUGHES 1983).

In this chapter an overview of the complex carbohydrates of plants and animals is presented as background for subsequent chapters. The emphasis is on glycoproteins of higher plants and animals; polysaccharides and proteogly-

Department of Pathology and Immunology, Monash Medical School Commercial Road, Prahan, Vic 3181, Australia

Current Topics in Microbiology and Immunology, Vol. 139

cans are also considered, but glycolipids are not discussed. Recent reviews of glycolipid structures are presented in HAKOMORI (1981), ELBEIN (1982), and WIEGANDT (1985). For the sake of brevity and clarity a number of generalizations are made, and some detail is omitted; for readers wishing to obtain more detailed information, appropriate reviews are cited.

One of the fundamental differences between plant and animal cells is the presence in plants of a rigid cell wall. The primary cell wall, which overlays the plasma membrane, consists of cellulose microfibrils embedded in a gel-like matrix of polysaccharides and proteins. The presence of this cell wall not only results in differences in architecture between the plant and the animal cell surfaces but also poses significant problems unique to the plant cell, e.g., secretion of extracellular macromolecules and reception of messages and signals. The cell wall of plants contains a wealth of extremely complex carbohydrates that are not encountered in animal cells. Other differences between the two cell types pertinent to their carbohydrate biochemistry include the intracellular plant storage granules and protein bodies.

2 Monosaccharides of Plant and Animal Glycans

A variety of monosaccharides are present in plant and animal glycans. Monosaccharides exist in a particular ring form, either a six-membered pyranose ring or the less common five-membered furanose ring (Fig. 1). The six-carbon sugars (hexoses) usually occur in the pyranose ring form whereas the five-carbon sugars (pentoses) are commonly in the furanose form. An exception is the pentose xylose, which occurs exclusively in the pyranose ring. Although each monosaccharide usually exists in one particular ring form, there are some exceptions, for example, the hexose galactose and the pentose arabinose occur in both ring forms in plant carbohydrates. In solution, the pyranose ring adopts the energetically favorable chair conformation ($^4C_1$), whereas the furanose ring

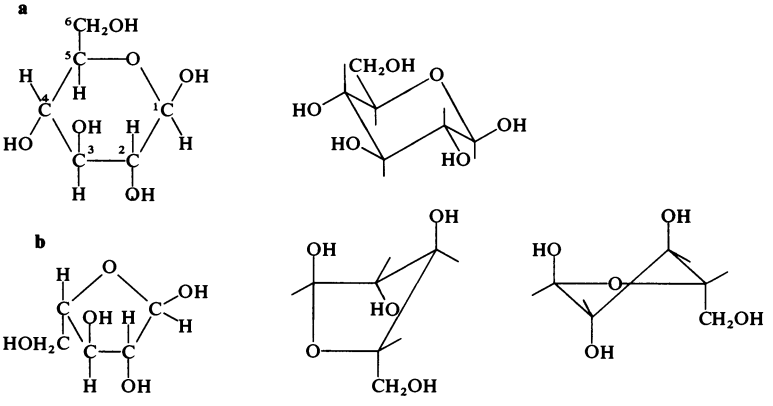


Fig. 1. a Representation of the hexose β-D-glucopyranose. *Left*, pyranose ring; *right*, chair conformation. **b** The pentose α-L-arabinofuranose. *Left*, furanose ring; *center*, envelope conformation, *right*, twist conformation

Table 1. Monosaccharide constituents of higher plant and animal complex carbohydrates

Monosaccharides found in both plant and animal carbohydrates	Monosaccharides found only in plants	Monosaccharides found only in animals
Glucose	Arabinose	Sialic acid[c]
Galactose	Apiose	Iduronic acid[d]
Mannose	Fructose	
Fucose	Galacturonic acid	
Xylose[a]	Rhamnose	
N--Acetylglucosamine	KDO (3-deoxymannooctulosonic acid)	
N-Acetylgalactosamine[b]		
Glucuronic acid	Aceric acid (3-C-carboxy-5-deoxyL-xylose)	

[a] Xylose is a common constituent of plant carbohydrates but in animals is restricted to the matrix proteoglycans

[b] Therer are only a few reports of *N*-acetylgalactosamine in plants (see text)

[c] There is, however, a single report of sialic acid in plants (see text)

[d] Although iduronic acid has not been found in higher plant carbohydrates it has been reported in lower plants (see text)

exists in a number of conformations known as the envelope and twist forms (Fig. 1) (REES 1977).

The monosaccharide units in glycan chains are joined by covalent bonds (glycosidic linkages) that connect the anomeric carbon atom (C1) through an oxygen atom to a carbon atom of the neighboring monosaccharide. This anomeric linkage can either be an α- or a β-glycosidic configuration. The most important monosaccharides and their structures are listed in Table 1 and Fig. 2. Some monosaccharides exist in both the α and β anomeric configurations (e.g., glucose, mannose, and galactose) whereas others occur in only one of the two possible configurations (e.g., α-L-fucose, α-sialic acid, and β-*N*-acetylglucosamine. Many monosaccharides are found in both animal and plant glycans (Table 1). However, there are differences in the relative importance and distribution of some of these shared monosaccharides. For example, although *N*-acetylgalactosamine is very common in animal glycoproteins, most plant glycoproteins analyzed to date are devoid of galactosamine (SELVENDRAN and O'NEILL 1982). *N*-Acetylgalactosamine has, however, been reported in a glycoprotein extract from *Cannabis sativa* leaves (HILLESTAD et al. 1977), and galactosamine was present in a polymer isolated from the culture medium of tobacco cells (HORI 1978). No information is available regarding the linkage of this amino sugar in these plant glycans. Another example is xylose, which is a common constituent of plant polysaccharides and glycoproteins, but is restricted to the carbohydrate–protein linkage of matrix proteoglycans in higher animals. Xylose has, however, recently been found in two glycoproteins from lower animals, the glycoprotein hemocyanin from *Helix pomatia* (VAN KUIK et al. 1985) and a cell surface glycoprotein of *Trypanosoma cruzi* (FERGUSON et al. 1983).

Sialic acids (a family of more than 20 sugars) occur throughout the higher and the lower animal kingdom (see reviews by WARREN 1963; SCHAUER 1982) and are an important source of charge in oligosaccharide chains. To date, there

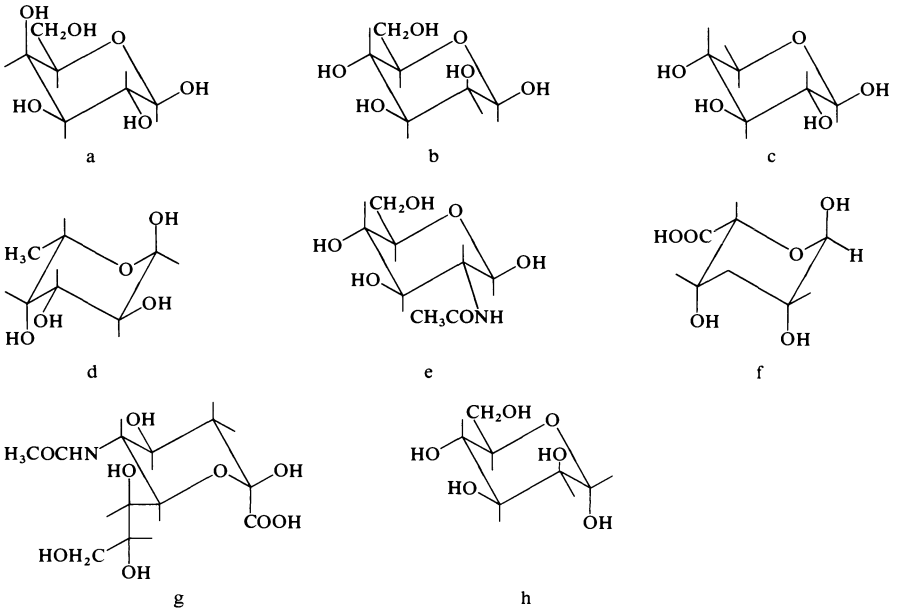


Fig. 2a–h. The ring structures of some of the monosaccharides present in plant and animal carbohydrate chains (cf. Fig. 1). **a** β-D-galactopyranose (galactose). **b** β-D-mannopyranose (mannose). **c** β-D-xylopyranose (xylose). **d** α-L-fucopyranose (fucose). **e** 2-acetamido-2-deoxy-β-D-glucopyranose (*N*-acetylglucosamine). **f** α-L-Idopyranosyluronic acid (iduronic acid). **g** *N*-acetylneuraminic acid (sialic acid). **h** α-D-mannopyranose (mannose)

has been only a single report of sialic acid in plants which has involved chemical characterization (BOURBOUZE et al. 1982). (A number of other reports of plant sialic acid are based on colorimetric assays; the color reaction may be caused by interferring substances; see SCHAUER 1982). L-Iduronic acid (5-epimer of D-glucuronic acid), a major component of animal matrix proteoglycans, has not been reported in higher plants, although it has recently been found in fungal cell walls (MIYAZAKI et al. 1984). Conversely, there are a number of monosaccharides that are restricted to the plant kingdom (Table 1). The most important of these is arabinose (Fig. 1), a common component of many plant polysaccharides, proteoglycans, and glycoproteins (Table 2). From Table 2 it is also clear that there are many differences at the compositional level between the three classes of plant and animal glycans. The most striking is the large number of monosaccharides found in plant polysaccharides compared to the number of those in animal polysaccharides. These differences will be discussed in structural terms in the following sections.

3 Modifications of Monosaccharides

Modifications of monosaccharides within the glycan chains pertinent to their physical and biological properties are listed in Table 3. These modifications

Table 2. Monosaccharide constituents of higher plant and animal polysaccharides, proteoglycans, and glycoproteins

	Higher Plants	Animals
Polysaccharides	Glucose Mannose Xylose Galactose Fucose Glucuronic acid Galacturonic acid Rhamnose Apiose Arabinose KDO Aceric acid	Glucose Galactose GlcNAc Glucuronic acid
Proteoglycans[a]	Arabinose Galactose	Xylose Galactose Glucuronic acid Iduronic acid GlcNAc GalNAc
Glycoproteins	Mannose GlcNAc Galactose Arabinose Xylose Fucose Glucose[b]	Mannose GlcNAc Galactose Fucose GalNAc Sialic acid Glucose

[a] Carbohydrate chains of plant proteoglycans can also contain additional minor constituents, e.g., rhamnose, fucose, uronic acids (see FINCHER et al. 1983). Carbohydrate chains of animal proteoglycans can be extensions from *N*-glycan core structures. These *N*-glycan monosaccharides have not been included

[b] Glucose has been reported only at the level of monosaccharide analysis

Table 3. Modifications of sugar residues within carbohydrate chains

Phosphorylation	Acetylation	Epimerization
Sulfation	Methylation	Anhydride formation

add a further dimension to the complexity of carbohydrate structures. Two modifications, namely the addition of phosphate and sulfate groups, are especially important since these constituents confer negative charge to the glycans. Furthermore, phosphorylated sugar residues of mammalian glycoproteins are involved in recognition interactions. For example, the specific transport

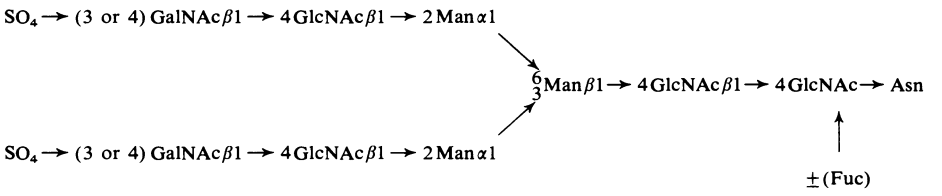


Fig. 3. Structure of disulfated *N*-linked oligosaccharide from bovine lutropin. (GREEN et al. 1985a)

of newly synthesized lysosomal hydrolases to the lysosomes in fibroblast cells involves a mannose 6-phosphate recognition marker on the high-mannose *N*-glycan chains of the lysosomal enzymes (GOLDBERG et al. 1984). This mannose 6-phosphate marker allows the binding of lysosomal enzymes to specific receptors in the Golgi complex and their subsequent transport to the lysosomes (CREEK and SLY 1984). Another recognition system involving mannose 6-phosphate residues has been described for glycoproteins that bind to the baseplate protein ligatin in the embryonic chick neural retina (MARCHASE et al. 1982). These glycoproteins contain mannose 6-phosphate residues in a diester linkage to C1 of terminal glucose residues (MARCHASE et al. 1982). Recently a novel type of phosphorylated sugar residue involving an *N*-acetylglucosamine-6-phosphate residue was described in a sialylated oligosaccharide from bovine colostrum (PARKKINEN and FINNE 1985). Phosphorylated mannose residues are also found in lower plants, for example, the mannoproteins of yeast (COHEN and BALLOU 1981); however, there have been no reports of phosphorylated sugar residues in the complex carbohydrates of higher plants.

Sulfation of the proteoglycans or glycosylaminoglycans of the extracellular matrix and connective tissues of animals has been known for some time and had even been considered an exclusive property of this group of glycans within the animal kingdom (RODEN and HOROWITZ 1978). The sulfate is found in either an ester sulfate linkage (O-sulfate) in a variety of positions on the monosaccharide ring or is linked to the amino groups of glucosamine (*N*-sulfate), as in heparin and heparan sulfate (RODEN and HOROWITZ 1978; HOOK et al. 1984). The degree of sulfation can be as high as three sulfate groups per disaccharide, as in some of the repeating units of heparin. The high sulfate content together with uronic acid residues results in a very high charge density in these polymers, and this contributes profoundly to their physical properties and to their potential to interact with other molecules (MUIR 1983; HOOK et al. 1984). Over the past few years the oligosaccharide chains of a number of mammalian glycoproteins have been found to bear sulfate constituents. Of particular interest are the human and bovine pituitary glycoprotein hormones that carry sulfated *N*-linked oligosaccharides (PARSONS and PIERCE 1980; GREEN et al. 1984; GREEN et al. 1985a). These *N*-linked oligosaccharides are also particularly unusual in their containing the amino sugar *N*-acetylgalactosamine. The structure of one sulfated oligosaccharide of bovine lutropin is shown in Fig. 3; the sulfated chains have *N*-acetyl-D-galactosamine (GalNAc) and sulfate at their nonreducing termini rather than the common sialic acid–galactose-*N*-acetylglucosamine sequences (see Sect. 6.3.1).

BAENZIGER and coworkers have recently examined the structural requirements for sulfation of these glycoprotein hormones and found the ability to be sulfated correlates with the presence of *N*-acetylgalactosamine (GREEN et al. 1985b). The addition of *N*-acetylgalactosamine, in turn, to the oligosaccharide chains appears to be a pituitary-specific and hormone-specific event (GREEN et al. 1985b). The carbohydrate chains of these hormones are important in the expression of their biological activity (SAIRAM and BHARGAVI 1985); exploring the role of the sulfate groups in hormone function will be of great interest. Other examples of sulfated *N*-linked animal oligosaccharides include sulfated mannose residues in the glycopeptide from ovalbumin (YAMASHITA et al. 1983), sulfated *N*-glucosamine residues of the *N*-glycans of calf thyroid plasma membrane glycoproteins (EDGE and SPIRO 1984), sulfated viral coat glycoproteins (PREHM et al. 1979), glycoproteins of the liver and lung of chick embryo (HEIFETZ et al. 1980), and glycoproteins of *Dictyostelium* (HOHMANN et al. 1985). There are also numerous examples of sulfated polysaccharides in lower plants, algae being a particularly rich source (PERCIVAL and MCDOWELL 1981; PAINTER 1983). The sulfated fucoidans of brown algae, for example, are a family of heteromolecules based on fucose, xylose, glucuronic acid, and half-ester sulfate (PERCIVAL and MCDOWELL 1981). Sulfate groups have not been reported in higher plant carbohydrates.

Other postpolymerization modifications include *O*-acetylation of sialic acids (SCHAUER 1982), *O*-acetylation and *O*-methylation of monosaccharide residues of cell wall polysaccharides of higher plants (MCNEIL et al. 1984), and the conversion of galactose 2,6 disulfate residue to 3,6 anhydrogalactose 2-sulfate in the polysaccharide chains of carrageenans (LAWSON and REES 1970). Finally, the monosaccharide L-iduronic acid is itself formed from D-glucuronic acid residues within glycan chains by C5 epimerization (MALMSTROM et al. 1975).

It is likely that the modifications described here are more widespread than is currently appreciated since they are often not given consideration in design of analytical procedures. One problem is that many analytical methods result in the removal of these additional constituents. It is important that the properties of the modified monosaccharides are appreciated when handling and analyzing carbohydrate samples.

4 Carbohydrate–Protein Linkages

The *N*-glycosidic linkage involving the C(1) of *N*-acetylglucosamine and the amide group of asparagine (GlcNAc–Asn) is the only type of *N*-glycosidic linkage described (Table 4). This *N*-linkage was first demonstrated in the glycoprotein ovalbumin by MARSHALL and NEUBERGER in 1964; it was, in fact, the first carbohydrate–protein linkage to be identified. We now realize that this linkage is widely distributed throughout eukaryotic organisms, and, furthermore, that the pathway for its biosynthesis appears to be conserved in all eukaryotes (see Sect. 6.3.1).

Table 4. Carbohydrate-protein linkages found in plants and animals

Linkage	Plants		Animals
	Lower	Higher	
N-Glycosidic			
β-D-GlcNAc-Asn	+[1]	+ [2]	+ [3]
O-Glycosidic			
α-D-GalNAc-Ser(Thr)	–	–	+ [4]
β-D-Xyl-Ser	–	–	+ [5]
α-D-Gal-Ser	–	+ [6, 7, 8]	+ [a] [9]
D-Man-Ser(Thr) [b]	+ [10]	–	–
D-Xyl-Thr [b]	+[11]	+ [12]	–
β-D-Ara(f)-Hyp		+ [13, 14]	–
Ara-Hyp [b]	+ [15]		–
β-D-Gal-Hyp		+ [16, 17]	–
Gal-Hyp [b]	+ [15]	+ [18]	–
β-D-Gal-Hyl	–	–	+ [19]
GlcNAc-Ser(Thr) [b]	–	–	? [20]
Glc-Tyr [b]	–	–	+ [21, 23]

[a] α-D-Gal residue attached to an alkali sensitive bond; the linkage has not been chemically characterized

[b] The anomeric linkage has not been established in these cases

References: Lee and Scocca (1972); [2] Sharon and Lis (1979); [3] Marshall and Neuberger (1970); [4] Sadler (1984); [5] Lindahl and Roden (1965); [6] Lamport et al. (1973); [7] O'Neill and Selvendran (1980); [8] Allen et al. (1978); [9] Muir and Lee (1969); [10] Sentandreu and Northcote (1968); [11] Heaney-Kieras et al. (1977); [12] Green and Northcote (1978); [13] Lamport (1967); [14] Akiyama and Kato (1976); [15] Miller et al. (1972); [16] McNamara and Stone (1981); [17] Strahm et al. (1981); [18] Pope (1977); [19] Butler and Cunningham (1966); [20] Torres and Hart (1984); [21] Rodriguez and Whelan (1985); [22] Aon and Curtino (1985)

There are a number of differences between the *O*-glycosidic linkages of plant and animal glycoproteins (Table 4). The α-D-*N*-acetylgalactosamine-serine (threonine) (GalNAc-Ser/Thr) linkage is the most important *O*-glycosidic linkage in animal glycoproteins; in view of the rarity of *N*-acetylgalactosamine in plant glycoproteins it is hardly surprising that this linkage has not been detected in plants. Other linkages found in animal glycoconjugates but thus far not in those of plants include β-D-xylosyl–serine in the matrix proteoglycans and β-D-galactosyl–hydroxylysine of collagen. In addition, there is indirect evidence for an *N*-acetylglucosamine-serine/threonine *O*-glycosidic linkage derived from glycoproteins on the surface of lymphocytes (Torres and Hart 1984). Hydroxyproline is found in a number of animal proteins (e.g., collagen and elastin), but in contrast to plants hydroxyproline is never glycosylated in animals. β-L-arabinosyl–hydroxyproline and β-D-galactosyl–hydroxyproline (Table 4) are important glycosidic linkages in glycoproteins and proteoglycans of both higher and lower plants. A characteristic feature of glycosidic linkages involving hyd-

roxyproline (and hydroxylysine) is their alkali stability; this distinguishes them from the alkali-labile *O*-glycosidic linkages involving serine and threonine. Plant *O*-glycosidic linkages involving these serine and threonine residues include α-D-galactosyl–serine found in the cell wall glycoprotein (extensin) and lectins from the Solanaceae, xylosyl–threonine in a root cap slime glycoprotein, and mannosyl–serine (threonine) in the yeast mannoproteins.

5 Complex Carbohydrates of Lower Plants and Animals

A detailed description of the carbohydrates of lower plants and animals is beyond the scope of this review. Sources of information are cell walls of lower plants, for example, algal cell walls (PERCIVAL and MCDOWELL 1981; PAINTER 1983) and fungal cell walls (WESSELS and SIETSMA 1981), reserve carbohydrates of lower plants (MANNERS and STURGEON 1982), and yeast mannoproteins (COHEN and BALLOU 1981).

There is less information available on the carbohydrates of lower animals. Glycoprotein structure and biosynthesis have been examined in mosquito cells (BUTTERS and HUGHES 1981; HSIEH and ROBBINS 1984) and the slime mould *Dictyostelium discoideum* (Ivatt et al. 1984). The protozoans are an important group of organisms since they are responsible for many parasitic infections. The cell surface glycoconjugates of a number of these organisms are known to play important roles in the biology of the parasite and its survival in the mammalian host, for example African trypanosomes. For a recent review on the glycosylation of the surface antigens of *Trypanosoma brucei* see HOLDER (1985).

6 Structures of Complex Carbohydrates of Higher Plants and Animals

6.1 Polysaccharides

The variety and complexity of plant polysaccharides are striking compared to those of animal polysaccharides. Plant polysaccharides can be classified into five structural classes: (a) linear homopolymers, (b) mixed-linked linear homopolymers, (c) linear heteropolymers, (d) branched homopolymers, and (e) branched heteropolymers. Selected examples of each class are listed in Table 5. The linear homopolymers are dominated by the glucans, for example, cellulose, amylose, and callose. The nature of the linkage between the glucosyl residues determines the overall conformation of the polymer, for example, cellulose with $\beta(1 \rightarrow 4)$glucosyl linkages adopts a flat extended-ribbon type conformation resulting in a tight packing of the individual polysaccharide chains, via hydrogen binding, to provide the high strength and fibrous characteristic of cellulose

Table 5. Classification of plant polysaccharides

Class	Polysaccharide	
Linear homopolymer	Cellulose	$\beta(1 \rightarrow 4)$glucan
	Amylose	$\alpha(1 \rightarrow 4)$glucan
	Callose	$\beta(1 \rightarrow 3)$glucan
	Homogalacturonans	$\alpha(1 \rightarrow 4)$galacturonyl residues
Mixed-linked linear homopolymer		$\beta(1 \rightarrow 3, 1 \rightarrow 4)$glucan
Linear heteropolymer	Glucomannans	Varying proportions of $\beta(1 \rightarrow 4)$linked manno-pyranosyl and $\beta(1 \rightarrow 4)$linked β-glucopyranosyl residues
Branched homopolymer	Amylopectin	Linear chains of $\alpha(1 \rightarrow 4)$linked glucopyranosyl residues containing $\alpha(1 \rightarrow 6)$linked glucopyrano-syl interchain branch linkages
	Arabinans	Backbone of $\alpha(1 \rightarrow 5)$linked L-arabino-furanosyl residues with single α-L-arabinofuranosyl resi-dues attached at O2 or O3 atoms
Branched heteropolymer	Arabinoxylan	$\beta(1 \rightarrow 4)$xylan backbone with side branches at-tached through O2 or O3 atoms of xylosyl resi-dues. The most common side branches are single α-L-arbinofuranosyl residues.
	Xyloglucans	$\beta(1 \rightarrow 4)$glucan backbone substituted with α-xy-losyl residues linked to the O6 atom of some of the glucosyl residues. The side branches can be further extended with galactose and fucose residues.
	Rhamnogalacturonin I	Backbone composed of alternating 2-linked L-rhamnosyl and 4-linked galacturonyl residues. Complex side branches are attached to the O4 atom of the L-rhamnosyl residues.
	Rhamnogalacturonan II	Very complex polysaccharide containing many unusual residues (see text)

With the exception of amylose and amylopectin all the polysaccharides listed are from cell walls of higher plants.

[a] Represents a portion of the callose polysaccharide

General references: Aspinall (1980); McNeil et al. (1984); Fincher and Stone (1986); Bacic et al. (1987)

(Rees 1977). On the other hand, $\alpha(1 \rightarrow 4)$glucan (amylose) and $\beta(1 \rightarrow 3)$glucan (callose) chains adopt a left-handed helical conformation, which results in a more flexible and open packing of the chains compared to cellulose (Rees 1977).

There is an increasing degree of structural complexity from the linear homopolymers to the branched heteropolymers (Table 5). Many of the polysaccharides listed in Table 5 are found in the primary and secondary cell walls of plants. Rhamnogalacturonan II, a pectic polysaccharide, is one of the most complex polysaccharides characterized to date; it has about 65 glycosyl residues and is composed of at least ten different sugars in some 20 different linkages (Darvill et al. 1985). Two heptasaccharides from rhamnogalacturonan II have

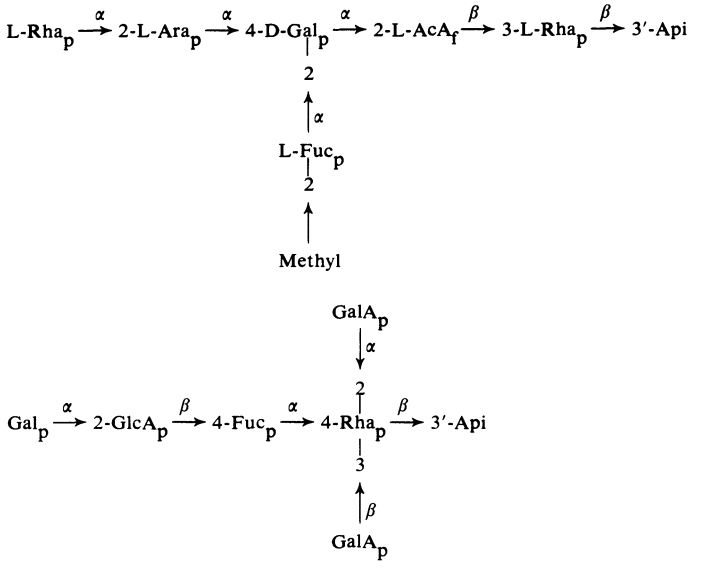


Fig. 4. The structures of two heptasaccharides isolated from rhamnogalacturonan II. (McNeil et al. 1984)

been characterized (McNeil et al. 1984; Darvill et al. 1985; York et al. 1985); their structures are shown in Fig. 4 to illustrate the amazing complexity of these polymers.

One of the most exciting discoveries in the field of plant polysaccharides is the potent biological activity of oligosaccharide fragments of cell wall polymers (for a reviews see McNeil et al. 1984; Ryan 1984; Albersheim and Darvill 1985). Fragments of the polysaccharide homogalacturonan (Table 5) induce plant tissues to synthesize antibiotics, called phytoalexins, which are thought to provide a general defense mechanism for plants (Darvill and Albersheim 1984). Oligosaccharides also induce proteinase inhibitors after tissue wounding. These inhibitors are thought to protect the plant against invading insects and microbes (Ryan 1978; Walker-Simmons et al. 1983). There is also evidence that oligosaccharides derived from plant cell walls play an important role in cell development and differentiation (Tran Thanh Van et al. 1985). These findings have begun to unravel the biological functions of these complex polysaccharides and have thereby transformed the concept of cell wall polysaccharides from one of purely structural entities to one of cellular components with a diverse range of biological activities.

In contrast to the numerous plant polysaccharides there are only a few different animal polysaccharides. This is largely due to the absence of a cell wall in animal cells. The major animal polysaccharide is the homopolymer glycogen, which has a similar structure to that of amylose (Table 5). This is the principal reserve polysaccharide of the animal kingdom. Hyaluronic acid, which is a linear repeating disaccharide of *N*-acetylglucosamine and glucuronic acid found in connective tissues and the extracellular matrix, can also be considered

Table 6. Dissacharide components of animal proteoglycans

Proteoglycan	Disaccharide
Chondroitin sulfate	D-glucuronyl $\beta(1 \rightarrow 3)$*N*-acetylgalactosamine 4 or 6 sulfate $\beta(1 \rightarrow 4)$
Dermatan sulfate	L-iduronyl $\beta(1 \rightarrow 3)$*N*-acetylgalactosamine 4 or 6 sulfate $\beta(1 \rightarrow 4)$
Heparan sulfate[a] Heparin[a]	L-iduronyl $\alpha(1 \rightarrow 4)$ *N*-acetylglucosamine / *N*-sulfoglucosamine $\}\alpha(1 \rightarrow 4)$
	D-glucuronyl $\beta(1 \rightarrow 4)$ *N*-acetylglucosamine / *N*-sulfoglucosamine $\}\alpha(1 \rightarrow 4)$
Keratan sulfate[a]	galactose $\beta(1 \rightarrow 4)$*N*-acetylglucosamine $\beta(1 \rightarrow 3)$

[a] See text for discussion of the position of *O*-sulfate in these proteoglycans

a polysaccharide since no covalent linkage to protein has been demonstrated. Other minor polysaccharides have also been described, for example, beef lung galactan, but these will not be considered here.

6.2 Proteoglycans

Proteoglycans are composed of a protein core with covalently attached polysaccharide chains. Both animal and plant cells secrete proteoglycans and, although chemically dissimilar, may have comparable biological functions.

6.2.1 Animal Proteoglycans

Animal proteoglycans are present in high concentrations in connective tissues and basement membranes and are also found at the cell surface of many mammalian tissues and cultured cells (HOOK et al. 1984). In addition, there is evidence that the common cell surface proteoglycan of adherent mammalian cells, heparan sulfate, is anchored into the plasma membrane via a hydrophobic region of the core protein (HOOK et al. 1984; RAPRAEGER and BERNFIELD 1985). Cell surface proteoglycans interact with other components of the extracellular matrix, for example, fibronectin, laminin, hyaluronic acid, and collagen, and these are thought to be involved in cellular adhesion, differentiation, and cell growth (TOOLE 1981; APLIN and HUGHES 1982; HOOK et al. 1984). For example, a cell surface heparan sulfate proteoglycan on dorsal root ganglion neurons is required for Schwann cell proliferation (RATNER et al. 1985).

Animal proteoglycans have traditionally been classified in terms of the structure of repeating disaccharide units of their polysaccharide chains (Table 6) (RODEN and HOROWITZ 1978; HASCALL and HASCALL 1981; HOOK et al. 1984; FRANSSON 1985). The chondroitin sulfate and dermatan sulfate of connective tissues are closely related and are composed of either β-D-glucuronic acid or

α-L-iduronic acid and β-*N*-acetylgalactosamine 4 or 6 sulfate. The polysaccharide chains are attached to a unique tetrasaccharide unit [GlcA $\beta(1 \rightarrow 3)$ Gal $\beta(1 \rightarrow 3)$ Gal $\beta(1 \rightarrow 4)$ Xyl], which in turn is linked to serine residues in the polypeptide core (DORFMAN 1981). Heparan sulfate and heparin consist of a mixture of disaccharides (Table 6) with some of the glucosamine residues containing *N*-sulfate groups. In addition, *O*-sulfation occurs at carbon atom 2 of iduronic acid residues and carbon atom 6 of glucosamine residues and also at carbon atom 3 of some glucosamine residues. Heparan sulfate is widely distributed; it is found on many cell surfaces and, as mentioned, is possibly an integral membrane component. Heparin, on the other hand, appears to be restricted to the granules of mast cells. Heparin has a characteristically higher sulfate content than heparan sulfate and, in addition, has potent anticoagulant activity. The polysaccharides of both these proteoglycans are also attached to their protein core via the xylose–serine *O*-glycosidic linkage.

Keratan sulfate differs from the other proteoglycans in lacking uronic acid residues and in the nature of its linkage to protein. It is composed of the repeating disaccharide [Gal $\beta(1 \rightarrow 4)$ GlcNAc $\beta(1 \rightarrow 3)$], which is *O*-sulfated on the C6 position of *N*-acetylglucosamine and/or the C6 position of galactose residues (BHAVANANDAN and MEYER 1968). Two types of keratan sulfate proteoglycans have been found. In cartilage the keratan sulfate polysaccharide chains are covalently linked to the core protein of the chondroitin sulfate proteoglycan via a GalNAc-Ser(Thr) *O*-glycosidic linkage (LOHMANDER et al. 1980). The second type is found in the cornea, where the keratan sulfate chains are attached to a core protein through a GlcNAc-Asn *N*-glycosidic linkage. These keratan sulfate chains are, in fact, elongations of the outer branches of complex type *N*-glycans (KELLER et al. 1981; NILSSON et al. 1983; see Sect. 6.3.2.1). The repeating disaccharides of the keratan sulfates are structurally identical to the unsulfated polylactosamine chains present on the *N*- and *O*-glycans of a number of glycoproteins (FUKUDA 1985).

The proteoglycan core proteins are only now beginning to receive attention. There may be several different core proteins for each major class of proteoglycan (BUMOL and REISFELD 1982; BRENNAN et al. 1984; KANWAR et al. 1984; HASSELL et al. 1986). The cDNA for the rat yolk sac chondroitin sulfate protein has recently been cloned; the deduced amino acid sequence has a central 49 amino acid region composed of alternating serine and glycine residues (BOURDON et al. 1985). A serine-glycine dipeptide repeat sequence was also previously shown for the rat heparin proteoglycan (ROBINSON et al. 1978). This serine–glycine repeat undoubtedly functions as the acceptor site for the attachment of the polysaccharide chains; synthetic peptides containing serine and glycine act as acceptors for glycosylation by xylosyl transferase (COUDRON et al. 1980), which is the initiation step for polysaccharide synthesis. Partial cDNA clones have also been recently obtained for the rat cartilage (DOEGE et al. 1986) and the chicken cartilage (SAI et al. 1986) proteoglycans. The partial sequences of these cartilage core proteins show extensive homology thus far; however, they appear to be unrelated to the rat yolk sac tumor proteoglycan discussed above. Clearly, more detailed information on the core protein structures is required for a better understanding of the tissue specific expression of different proteoglycans.

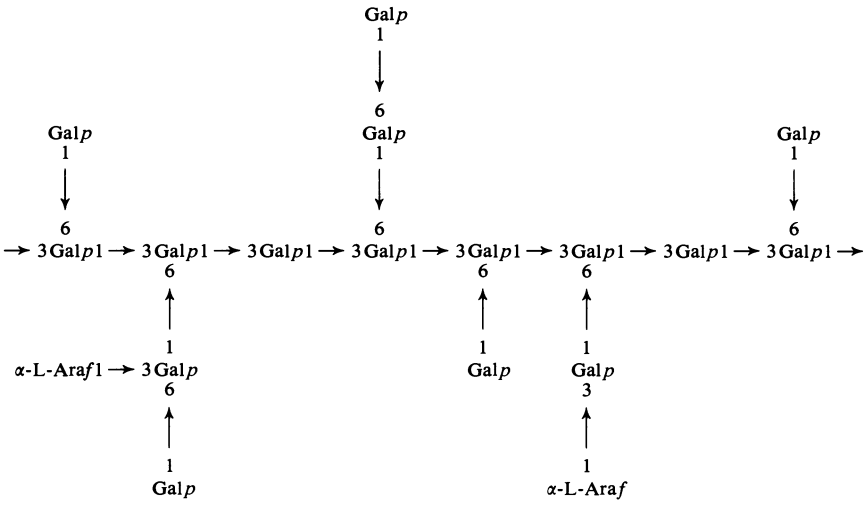


Fig. 5. A proposed structure for a portion of the *Gladiolus* style arabinogalactan protein. (GLEESON and CLARK 1979)

6.2.2 Plant Proteoglycans

Plant proteoglycans belong to a single group of macromolecules known as arabinogalactan proteins (AGPs). AGPs are widely distributed and are found in the majority of tissues and secretions of higher plants, as well as being synthesized and secreted by many cultured cells (CLARKE et al. 1979; FINCHER et al. 1983). AGPs are characterized by a high content of carbohydrate in which galactose and arabinose are the major monosaccharides; they are associated with a low content (usually 2%–10%) of protein, that is typically rich in hydroxyproline, alanine, and serine (FINCHER et al. 1983). The carbohydrate–protein linkage has been identified as a β-galactosyl–hydroxyproline linkage in wheat AGP (MCNAMARA and STONE 1981; STRAHM et al. 1981), and a similar linkage has been implicated for the AGP from cultured ryegrass endosperm cells (BACIC et al. 1987). The structures of the polysaccharide chains of a number of AGPs have been characterized (FINCHER et al. 1983); they are a structurally related group consisting of a linear $\beta(1 \rightarrow 3)$galactan backbone with short $\beta(1 \rightarrow 6)$ linked galactosyl oligosaccharide side branches that, in turn, are substituted by arabinose residues and other minor substituents which are often in terminal positions (FINCHER et al. 1983). The tentative structure of the AGP isolated from *Gladiolus* style mucilage is shown in Fig. 5. Evidence from a variety of studies suggests that the β-galactan backbone also contains periodate-sensitive residues interspersed at regular intervals along the backbone, resulting in discrete blocks of $\beta(1 \rightarrow 3)$ linked galactosyl residues (STEPHEN 1980, FINCHER et al. 1983).

As with animal proteoglycans, relatively little is known about the structure and organization of the protein core of AGPs. Ryegrass AGP has recently been deglycosylated to yield a single major protein component with an apparent

molecular weight of 35000 daltons (GLEESON et al. 1985a, b). Partial amino acid sequence analysis of deglycosylated ryegrass AGP (GLEESON et al. 1985a, b) as well as carrot AGP (JERMYN and GUTHRIE 1985) shows a characteristic repeating dipeptide of alanine-hydroxyproline. Therefore, both the animal and the plant proteoglycans have a repetitive dipeptide sequence that is glycosylated. AGPs from many tissues show heterogeneity based on size and charge (CLARKE et al. 1979; GLEESON and CLARKE 1980); furthermore AGPs from the sexual tissues of flowering plants show developmentally regulated changes (GELL et al. 1986). The nature of these differences has yet to be elucidated.

6.3 Glycoproteins

6.3.1 *N*-Glycans

Many animal and plant proteins are *N*-glycosylated; these include intracellular, membrane, and secreted proteins that are functionally diverse (Table 7). For example, the majority of animal plasma membrane proteins are *N*-glycosylated, as are many secreted glycoproteins, including immunoglobulins, α_1-acid glycoprotein and many hormones (Table 7). Plant glycoproteins known to carry *N*-glycans include enzymes, lectins, toxins, and intracellular storage proteins (Table 7). In contrast to the extensive studies of animal membrane glycoproteins, it is quite astonishing that plant biochemistry is unable to offer a single example of a characterized plant membrane glycoprotein (for review on this area see BOWLES 1982). Biosynthetic studies using pea cotyledon membranes (NAGAHASHI et al. 1980) and soybean cultured cells (HORI and ELBEIN 1983; HORI et al. 1985) have indeed indicated that plant membrane proteins are *N*-glycosylated, but there is as yet no direct structural characterization of any plant membrane glycoprotein.

The structure and biosynthesis of the *N*-glycans of animal glycoproteins have been extensively studied. New techniques for the structural analysis of oligosaccharides, in particular, ^{1}H-NMR (nuclear magnetic resonance) spectroscopy, have had a dramatic impact on the field of glycoprotein research and have resulted in the elucidation of the complete structures of a large number of *N*-glycans (for reviews see MONTREUIL 1982; VLIEGENTHART et al. 1983; KOBATA 1984; CARVER and BRISSON 1984). These structures can be classified into three main types: (a) high-mannose (b) hybrid, and (c) complex *N*-glycans (Fig. 6). High-mannose *N*-glycans usually have between five and nine mannose units, whereas hybrid *N*-glycans have features shared by both the high-mannose and the complex oligosaccharides. Complex *N*-glycans have a minimum of two outer branches with the terminal sequence of [SAα(2$\rightarrow$3 or 2$\rightarrow$6) Galβ(1$\rightarrow$4) GlcNAcβ(1$\rightarrow$4)]. Also an α(1$\rightarrow$6)linked fucose residue may be present on the Asn-linked GlcNAc of the chitobiose core. In addition, the hybrid and complex *N*-glycans may have a "bisecting" *N*-acetylglucosamine linked β(1$\rightarrow$4) to the β-linked mannose of the trimannosyl core. A large number of different complex *N*-glycan structures have been described; these are summarized in Fig. 7. There are variations in the number of outer branches (usually between two and four)

Table 7. Distribution of *N*-glycosylated glycoproteins in plants and animals

	Plant	Animal
Membrane glycoproteins	?	Many membrane proteins are *N*-glycosylated, for example, glycophorin [1] epidermal growth factor receptor [2], murine major histocompatibility antigens [3]
Enzymes	Bromelain (pineapple stem) [4] Laccase (secreted by sycamore cells) [5] Ficin (Fig latex) [6] Peroxidase (horseradish) [7]	α_1-antitrypsin [8] Lysosomal hydrolases [9] (intracellular)
Hormones		Lutropin [10] [11] Gonadotropin [12]
Toxins	Ricin Abrin [13, 14, 15]	
Lectins	Soybean agglutinin [16] *Vicia graminea* lector [17] Lima bean lectin [18] Phytohemagglutinin (*Phaselus vulgaris*) [19, 35] Tora bean lectin [20] *Clerodendron trichotomum* lectin [21]	Asialoglycoprotein receptor [22]
Storage glycoproteins	Physeolin (kidney bean) [23] 7S protein (soy bean) [24] Pea vicilin [25]	
Others	S-allele associated glycoprotein [26, 27]	Serum glycoproteins, e.g.: Transferrin [28, 29] α_1-Acid glycoprotein [30] Immunoglobulins G [31] M [32, 38] and A [34]

References: [1] Yoshima et al. 1980; [2] Cummings et al. 1985; [3] Swiedler et al. 1985; [4] Ishihara et al. 1979; [5] Takahashi et al. 1986; [6] Friedenson and Liener 1974; [7] Clarke and Shannon 1976; [8] Hodges et al. 1979; [9] Goldberg et al. 1984; [10] Parsons and Pierce 1980; [11] Green et al. 1985; [12] Kessler et al. 1979; [13] Olsnes et al. 1975; [14] Nanno et al. 1975; [15] Olsnes and Pihl 1982; [16] Dorland et al. 1981; [17] Prigent et al. 1984; [18] Misaki and Goldstein 1977; [19] Vitale et al. 1984; [20] Ohtani and Misaki 1980; [21] Kitagaki-Ogawa et al. 1986; [22] Kawasaki and Ashwell 1976; [23] Davies and Delmer 1979; [24] Yamauchi and Yamagishi 1979; [25] Badenach-Jones et al. 1981; [26] Anderson et al. 1986; [27] Takayama et al. 1987; [28] Spik et al. 1975; [29] Hatton et al. 1979; [30] Fournet et al. 1978; [31] Narasimhan et al. 1980; [32] Chapman and Kornfeld 1979; [33] Anderson and Grimes 1982; [34] Baenziger and Kornfeld 1974; [35] Sturm and Chrispeels 1986

and in the terminal sequences found on these branches. (For further details see VLIEGENTHART et al. 1983; KOBATA 1984; KORNFELD and KORNFELD 1985.) The type of structure found depends on the individual glycoprotein and the tissue of synthesis. In addition, the carbohydrate structures of some glycoproteins show changes during cell development and differentiation, for example, band 3 glycoprotein (FUKUDA 1985) and Thy 1 antigen (CARLSSON 1985; MORRISON et al. 1986). For a multiply glycosylated glycoprotein, such as murine histo-

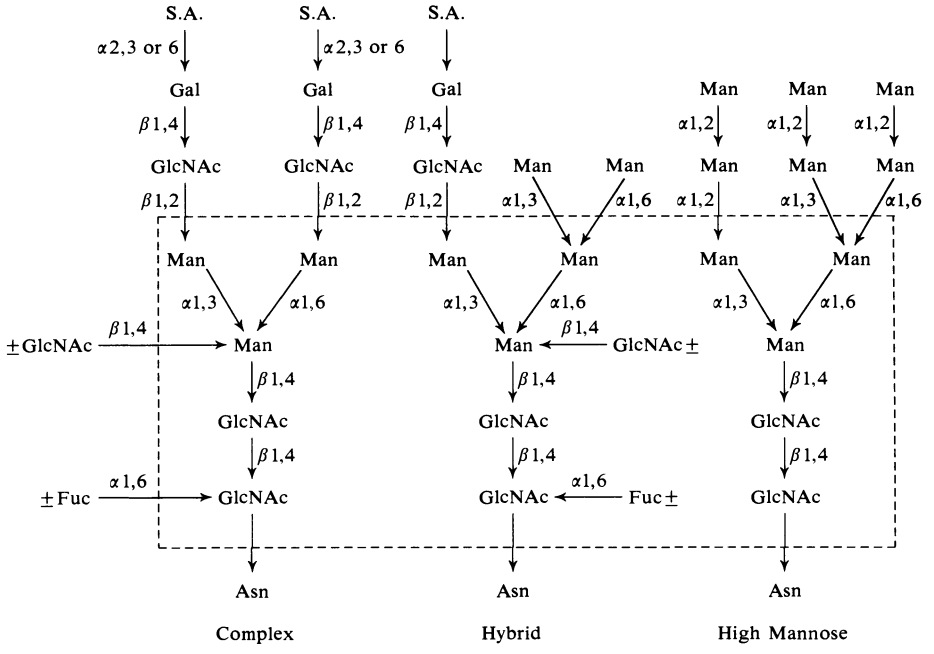


Fig. 6. Structures of the major types of *N*-linked oligosaccharides. *Boxed area* encloses the pentasaccharide core common to all *N*-glycans. (Adapted from KORNFELD and KORNFELD 1985)

compatibility antigens (SWIEDLER et al. 1985), each glycosylation site usually has a characteristic limited set of structures that may differ from the sets found at other glycosylation sites. Hence, glycosylation can be site-specific on individual glycoproteins.

The biosynthesis of animal *N*-glycans is now understood in some detail (for reviews see SCHACHTER et al. 1983; KORNFELD and KORNFELD 1985); this will be outlined briefly since it provides a framework for comparing the structures of the *N*-glycans of plants. The common pentasaccharide core seen in Fig. 6 arises because all *N*-glycans share a common biosynthetic pathway involving, initially, the assembly of a lipid-linked oligosaccharide precursor. The structure of this oligosaccharide precursor ($Glc_3Man_9GlcNAc_2$) is shown in Fig. 8. This oligosaccharide is then transferred from the lipid carrier (dolichol phosphate) to an asparagine residue of the tripeptide acceptor sequence -Asn-X-Thr(Ser)- (X can be any amino acid except proline) of nascent polypeptide chain. After transfer to protein the glucose residues and some of the mannose residues are removed by a number of processing enzymes in the rough and smooth endoplasmic reticulum. On arrival at the Golgi, further mannose residues are removed to yield a $Man_5GlcNAc_2$ structure (Fig. 9). This high-mannose oligosaccharide can then be acted on by a variety of glycosyl transferases and processing enzymes to convert the high-mannose *N*-glycan into the hybrid and complex type *N*-glycan structures outlined above (Fig. 9).

Plant *N*-glycans were largely unexplored until very recently. Biosynthetic studies by LEHLE (1981) and STANELONI et al. (1981) have indicated that plant

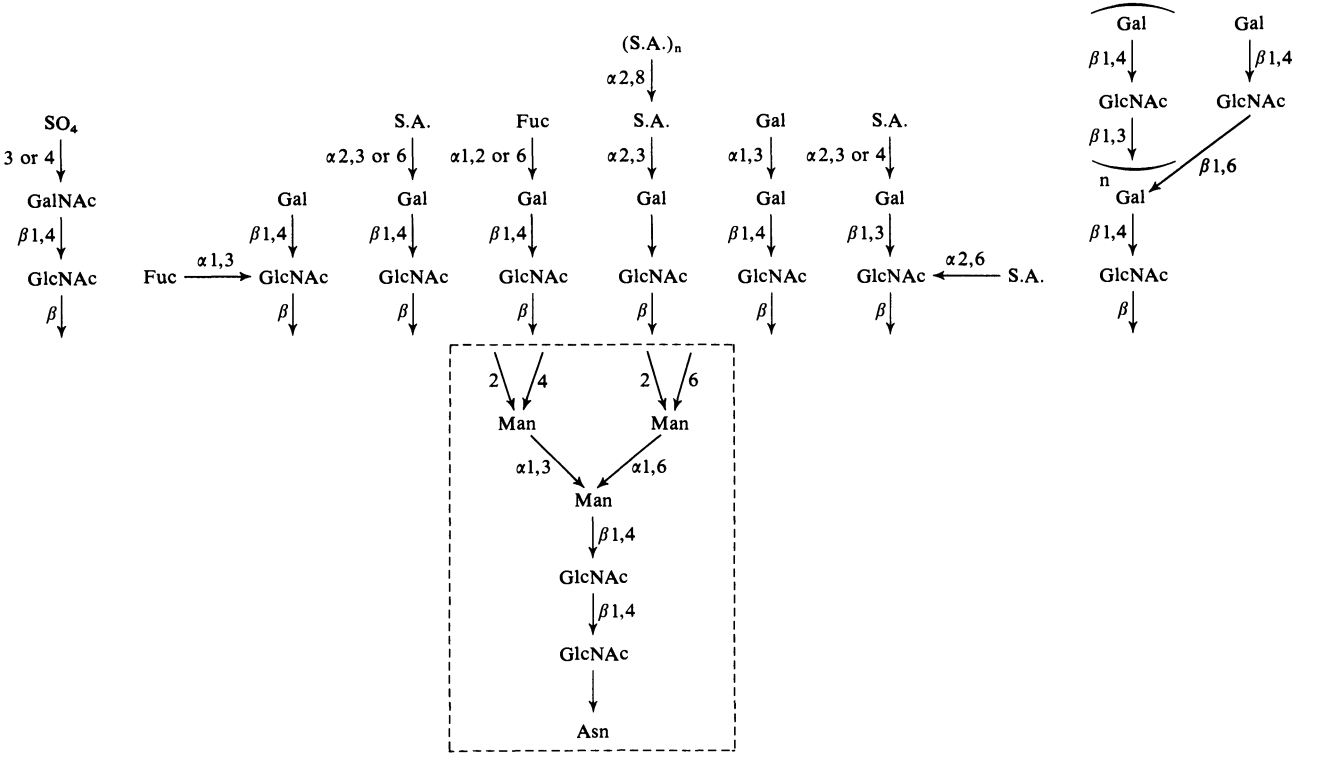


Fig. 7. Outer chain sequences found in complex *N*-glycans. *Boxed area* encloses the common pentasaccharide core to which the outer branches are attached. (Adapted from KORNFELD and KORNFELD 1985)

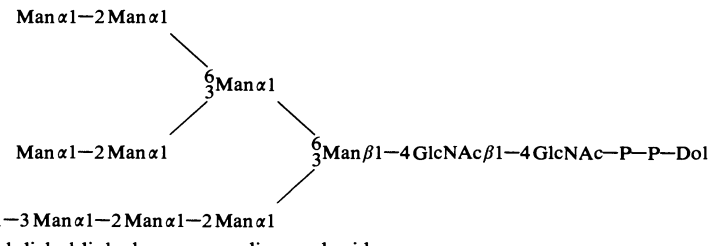


Fig. 8. Structure of the dolichol-linked precursor oligosaccharide

cells synthesize a similar, if not identical, oligosaccharide-lipid intermediate to that of animal cells. The first higher plant *N*-glycan to be characterized by means of ^{1}H-NMR was soybean agglutinin (DORLAND et al. 1981). This glycoprotein has a high-mannose oligosaccharide containing nine mannose residues whose structure is identical to that found in animal systems, again indicating a common biosynthetic pathway in plants and animals. A number of other plant glycoproteins also carry high-mannose *N*-glycans, for example, phytohemagglutinin (VITALE et al. 1984). Relatively little, however, is known about the subsequent biosynthetic processing events of plant high-mannose oligosaccharides. A number of plant *N*-glycans contain xylose, a monosaccharide absent from higher animal *N*-glycans. The structure of the xylose-containing *N*-glycan from pineapple stem bromelain (Fig. 10a) was first proposed by ISHIHARA et al. (1979) and recently supported by HOFFMANN et al. (1984). It differs from the animal *N*-glycans in the following respects: (a) the presence of a xylose residue linked $\beta(1 \rightarrow 2)$ to the β-mannosyl residue of the core, (b) the absence of an $\alpha(1 \rightarrow 3)$ linked mannose residue; and (c) the presence of an $\alpha(1 \rightarrow 3)$ linked fucose residue [rather than an $\alpha(1 \rightarrow 6)$ linkage] to the Asn-linked GlcNAc. This structure indicates that the processing of high-mannose oligosaccharides in plant cells can differ from that in animal cells. The presence of a $\beta(1 \rightarrow 2)$ linked xylose residue attached to the β-mannosyl unit may be a common feature of plant *N*-glycans since it has also been reported in the oligosaccharide structures from Tora bean lectin (OHTANI and MISAKI 1980), *Vicia graminea* lectin (PRIGENT et al. 1984), the lectins from *Clerodendron trichotomum* (KITAGAKI-OGAWA et al. 1986), and the glycoprotein laccase secreted by sycamore cells (TAKAHASHI et al. 1986). Xylose-containing *N*-glycans have also been reported from lower animals, the glycoprotein hemocyanins of the gastropods *Helix pomatia* (VAN KUIK et al. 1985) and *Lymnaea stagnalis* (VAN KUIK et al. 1986), and a cell surface glycoprotein from *Trypanosoma cruzi* (FERGUSON et al. 1983); the structure may not, therefore, be restricted to the plant kingdom.

An important recent finding is the presence of complex-type *N*-glycans in plants containing the *N*-acetyllactosamine terminal sequence similar to that of animal glycoproteins. The proposed structures of some of these complex-type oligosaccharides, which were isolated from the glycoprotein laccase (TAKAHASHI et al. 1986), are shown in Fig. 10b. These structures were determined by a variety of methods in combination with ^{1}H-NMR spectroscopy. They are biantennary complex *N*-glycans that differ from animal complex *N*-glycans in three respects: (a) all contain a $\beta(1 \rightarrow 2)$ xylose residue; (b) an $\alpha(1 \rightarrow 3)$ linked fucose residue

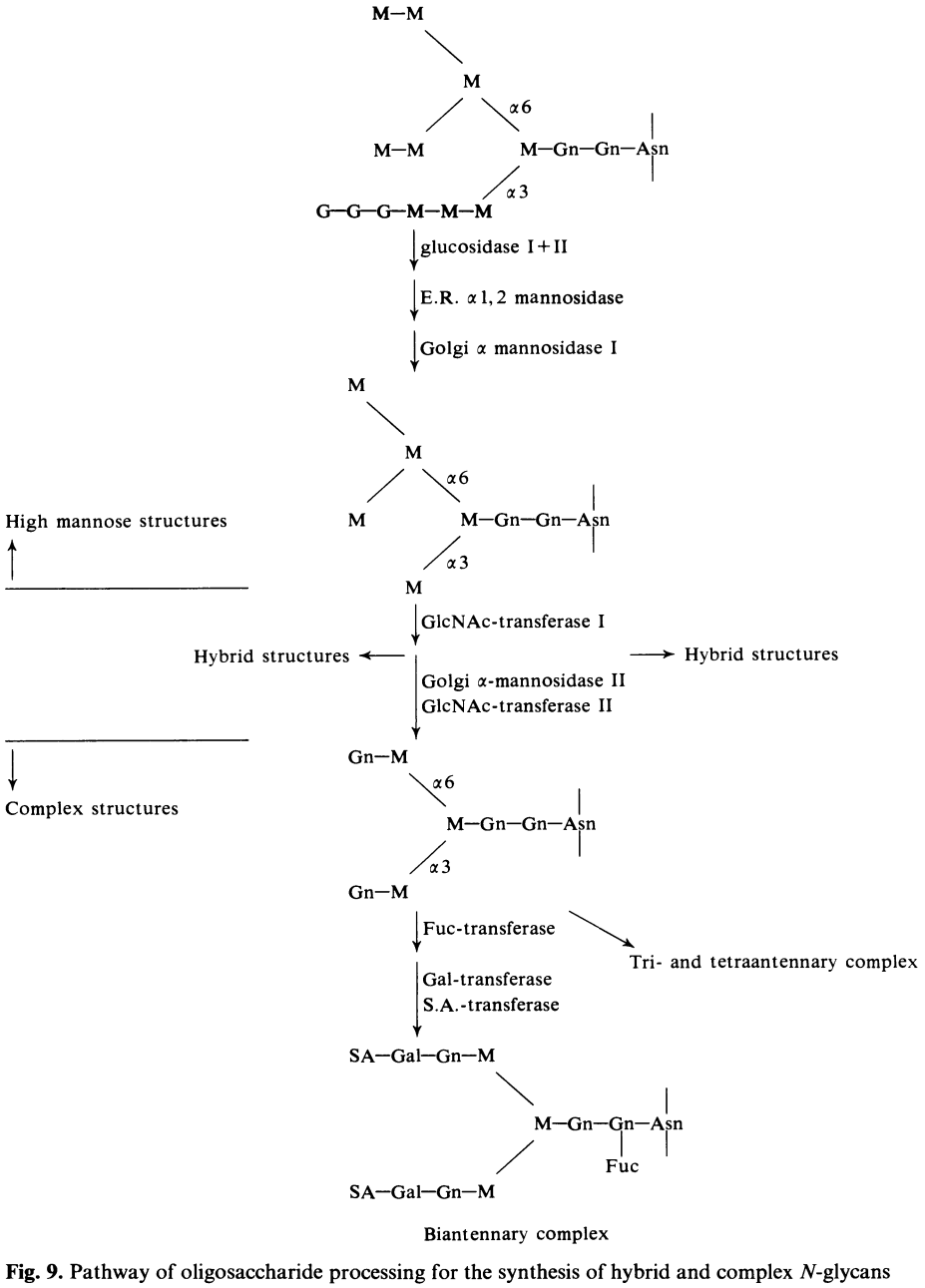


Fig. 9. Pathway of oligosaccharide processing for the synthesis of hybrid and complex *N*-glycans

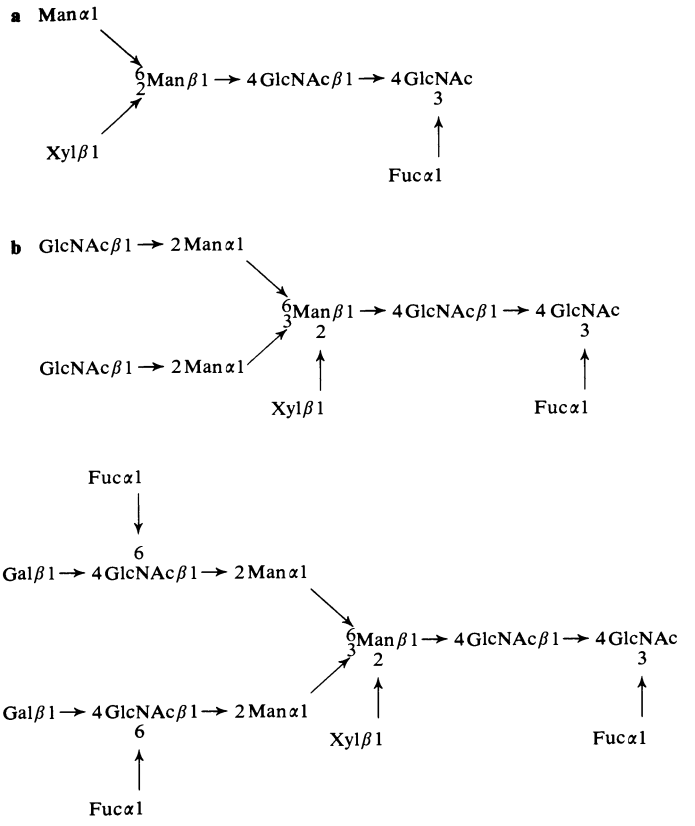


Fig. 10a, b. Proposed structures of some of the *N*-glycans from **a** stem bromelain (ISHIHARA et al. 1979), **b** laccase (TAKAHASHI et al. 1986)

is attached to the Asn-linked GlcNAc residue (as in the stem bromelain oligosaccharides); and (c) some oligosaccharides contain a second fucose residue attached to the outer "arm" GlcNAc residue in an $\alpha(1 \rightarrow 6)$ linkage. In animal *N*-glycans fucose is always attached to the arm GlcNAc residue in an $\alpha(1 \rightarrow 3)$ rather than an $\alpha(1 \rightarrow 6)$ linkage. The linkages of the two fucose residues in these plant *N*-glycans are therefore reversed compared to those in the animal structures. The identification of complex *N*-glycans in plants confirms earlier suggestions by POWELL and BREW (1974), who detected a galactosyl transferase activity in membrane preparations of onion stems, with characteristics similar to those of the $\beta(1 \rightarrow 4)$ galactosyl transferase in animal tissues involved in the synthesis of complex *N*-glycans.

Overall, the structures of the plant *N*-glycans elucidated to date suggest that the processing pathway of the *N*-linked plant oligosaccharides has features that are specific to plant tissues. With the technology currently available the stage is now set for a rapid expansion of knowledge concerning the structures and biosynthesis of plant *N*-glycans.

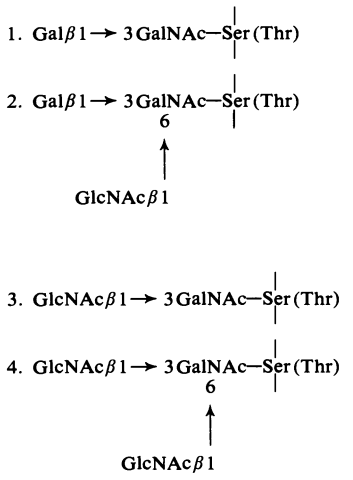


Fig. 11. The four basic core structures of animal *O*-glycans

6.3.2 *O*-Glycans

Since no structural similarity has so far been detected between plant and animal *O*-glycans these will be considered separately.

6.3.2.1 Animal O-*Glycans*

Many membrane and secreted glycoproteins of higher animals carry *O*-glycosidically linked oligosaccharides. The majority of these *O*-glycans are linked to the polypeptide through an α-D-*N*-acetylgalactosamine–serine (threonine) *O*-glycosidic linkage (collagen is an exception with a β-D-galactosyl–hydroxylysine *O*-glycosidic linkage; see Table 4). Monosaccharides commonly found in these GalNAc–Ser (Thr) *O*-glycans are GalNAc, Gal, GlcNAc, sialic acid, and fucose. Much of the structural information on the *O*-glycans comes from study of mucin glycoproteins, which are the major components of the viscous secretions lining the epithelium of the respiratory, gastrointestinal, and urogenital tracts. The simplest *O*-glycans carry the monosaccharide GalNAc and the disaccharide NeuAcα(2→6)GalNAc-Ser(Thr), as found in ovine submaxillary gland mucin (Hill et al. 1977). Larger *O*-glycan structures have been grouped into various classes according to their core structure (Schachter and Williams 1982; Hounsell et al. 1982). There are at least four different core classes (Fig. 11). The type 1 core, Galβ(1→3)GalNAc-Ser(Thr) is the most common, being found in a variety of mucins as well as in membrane and secretory glycoproteins (Sadler 1984) such as human bronchial mucin (Van Halbeek et al. 1982), porcine submaxillary mucin (Van Halbeek et al. 1980), rat small intestinal mucin (Carlsson et al. 1978), fetuin (Spiro and Bhoyroo 1974), glycophorin A from a variety of species (Thomas and Winzler 1969; Fukuda et al. 1981; Fukuda et al. 1982), human chorionic gonadotropin (Kessler et al. 1979), low

density lipoprotein receptor (CUMMINGS et al. 1983), immunoglobulin A (BAENZIGER and KORNFELD 1974; PIERCE-CRETEL et al. 1981), and the antifreeze glycoprotein from Antarctic fish (VANDENHEEDE et al. 1972). The type 2 trisaccharide core is an extension of the type 1 core (Fig. 11) by the addition of a $\beta(1 \rightarrow 6)$ linked GlcNAc residue to the GalNAc residue. This core type has been found in a number of mucins as well as in bovine glycoprotein (FUKUDA et al. 1982) human milk immunoglobulin A (PIERCE-CRETEL et al. 1981), and some tumor cells (NILSSON et al. 1982; FUNAKOSHI and YAMASHINA 1982). Types 3 and 4 cores are less commonly found; these do not have a galactose residue but rather a $\beta(1 \rightarrow 3)$ linked GlcNAc residue attached to GalNAc. The type 3 core has been found in the oligosaccharides of mucins such as rat colonic (SLOMIANY et al. 1980) and small-intestinal mucins (CARLSSON et al. 1978), human bronchial mucin (VAN HALBEEK et al. 1982), and bovine submaxillary gland mucin (GLEESON et al. 1984). Type 4 core structures have been found in sheep gastric mucin (HOUNSELL et al. 1981) and human bronchial mucin (VAN HALBEEK et al. 1982).

The four core structures can be elongated by the addition of further monosaccharides, which can result in large *O*-glycans containing more than 15 sugar residues. For example, the cores can be elongated with the repeating lactosamine disaccharide $[Gal\beta(1 \rightarrow 4)GlcNAc\beta(1 \rightarrow 3)]_n$ to generate linear and branched structures representing, respectively, the i and I antigenic determinants. These extensions can often be terminated by other antigenic determinants such as the human blood group ABH, Lewis determinants, and the stage-specific embryonic antigen (SSEA-I) (for further discussion see FEIZI 1985; FEIZI and CHILDS 1985).

The biosynthesis of Ser(Thr) linked oligosaccharides occurs by a direct transfer of sugar from the nucleotide intermediate to the polypeptide chain. The assembly of *O*-glycans does not involve preassembled, oligosaccharides or lipid intermediates (SCHACHTER and WILLIAMS 1982; SCHACHTER et al. 1985) and thus differs from the biosynthesis of the *N*-glycans. The synthesis of all *O*-glycans is initiated by incorporation of a GalNAc residue, from UDP-GalNAc, into the polypeptide chain by a UDP-GalNAc:polypeptide transferase. Unlike the synthesis of the GlcNAc-Asn linkage, a specific primary amino acid sequence for the attachment of GalNAc to serine or threonine cannot be identified. However, a high content of proline residues around the acceptor hydroxyamino acid residues appears to be important for effective glycosylation. (YOUNG et al. 1979). The subcellular site of *O*-glycosylation is probably initiated in the Golgi complex, in contrast to *N*-glycosylation, which is initiated in the rough endoplasmic reticulum. This conclusion is supported by a variety of studies, including those on the kinetics of *O*-glycosylation (HANOVER et al. 1982; JOHNSON and SPEAR 1983; CUMMINGS et al. 1983), the distribution of GalNAc transferase in subcellular membrane fractions (ELHAMMER and KORNFELD 1984), and a cytochemical investigation using colloidal gold conjugated to *Helix pomatia* lectin to localize GalNAc containing glycoproteins (ROTH 1984).

The enzymes involved in the synthesis of the four core structures have been detected and characterized (SCHACHTER and WILLIAMS 1982; SCHACHTER et al. 1985; BROCKHAUSEN et al. 1985) as well as many of the enzymes involved in the elongation steps (see SADLER 1984).

Animal *O*-glycans are known to play important structural roles. For example, the *O*-glycans of mucin are essential in determining their physical properties (SADLER 1984). It is of interest that a number of membrane and secreted glycoproteins contain sialylated type 1 *O*-glycans which are found clustered in serine- and threonine-rich domains of the protein. These small acidic oligosaccharides are likely to cause the protein to adopt an extended conformation. The clustered *O*-glycans of the low density lipoprotein receptor (CUMMINGS et al. 1983; GOLDSTEIN et al. 1985) and the interleukin-2 cell surface receptor (LEONARD et al. 1984; NIKAIDO et al. 1984) are found immediately external to the membrane-spanning region, and, furthermore, it has been suggested that these oligosaccharides may function as struts to keep the receptors extended from the membrane surface (GOLDSTEIN et al. 1985).

6.3.2.2 Plant O*-glycans*

There are two different *O*-glycosidic linkages described so far in higher plant glycoproteins: α-D-galactosyl-serine and β-L-arabinofuranosyl-hydroxyproline (Table 4). Both these linkages have been found in two distinct classes of glycoproteins: (a) cell wall glycoproteins (extensin) of dicotyledonous plants (LAMPORT 1977; LAMPORT and CATT 1981) and (b) glycoprotein lectins (chitobiose-specific) of the family Solanaceae, e.g., lectins from potato (ALLEN et al. 1978; MURRAY and NORTHCOTE 1978) and thorn apple (*Datura stramonium*) (HOREJSI and KOCOUREK 1978; DESAI et al. 1981). Both classes are hydroxyproline-rich macromolecules, containing about 50% carbohydrate with galactose and arabinose as the major monosaccharides (for review see SHOWALTER and VARNER 1987). Carbohydrate structures are very similar in the two classes of glycoproteins, but there are some major differences in the amino acid composition of cell wall glycoproteins and that of lectins (ALLEN et al. 1978; SMITH et al. 1984; CHEN and VARNER 1985b). In cell wall glycoproteins (LAMPORT and MILLER 1971) and potato lectin (ASHFORD et al. 1982) the carbohydrate linked through Ara–Hyp consists of mono-, tri- and tetra-arabinosides. The tri- and tetrasaccharides from the cell wall of *Nicotiana tabacum* (AKIYAMA et al. 1980) and potato lectin (ASHFORD et al. 1982) are identical; the structure of the tetrasaccharide is Ara f α(1 → 3)Ara f β(1 → 2)Ara f β(1 → 2)-Ara f β1 → Hyp while the trisaccharide lacks the terminal Ara f α(1 → 3) residue. The Gal–Ser linkage is also found in both classes of glycoproteins, present as single α-D-Gal residues in plant cell wall glycoproteins (LAMPORT et al. 1973; CHO and CHRISPEELS 1976; O'NEILL and SOLVENDRAN 1980) as well as potato and thorn apple lectins (ASHFORD et al. 1982). In addition, thorn apple lectin has the disaccharide Gal β(1 → 3)Gal α-Ser. The structure of these oligosaccharide chains sets these two classes of hydroxyproline-rich macromolecules distinctly apart from the AGP proteoglycans discussed previously, since the latter usually contain a higher carbohydrate content, present as polysaccharide chains attached to the protein through a Gal-Hyp linkage (Sect. 6.2).

Very little is known about the biosynthesis of these Ara–Hyp and Gal-Ser *O*-glycans. An arabinoside transferase activity reported in a membrane preparation from potato tissue culture cells transferred arabinose to deglycosylated

potato lectin (OWENS and NORTHCOTE 1981). Characterization of the products indicated the synthesis of hydroxyproline arabinosides (OWENS and NORTHCOTE 1981). The primary amino acid sequence requirements for arabinosylation of the hydroxyproline residues in these glycoproteins are not known. Of particular interest is the partial amino acid sequence of cell wall glycoprotein from tomato (LAMPORT 1977), with the unusual repeat pentapeptide Ser-Hyp-Hyp-Hyp-Hyp. cDNA and genomic clones have been isolated for carrot cell wall glycoprotein (CHEN and VARNER 1985a; 1985b), and the derived amino acid sequence contains 25 $Ser(Pro)_4$ pentapeptide repeats which are presumably post-translationally modified to the $Ser\text{-}(Hyp)_4$ sequence. The AGP proteoglycans, which contain the Gal-Hyp linkage rather than the Ara-Hyp linkage, have a repeat sequence of the dipeptide Ala-Hyp (Sect. 6.2). It is very likely that these characteristic sequences account for the differential glycosylation of the hydroxyproline residues in these classes of macromolecules.

Although no amino acid sequence information is available for the Solanaceae lectins, it has been reported that potato lectin adopts a polyproline type II helical conformation similar to that found in the cell wall glycoproteins (VAN HOLST et al. 1986). In addition, the potato lectin is known to contain two domains, only one of which is glycosylated. This glycosylated domain is extremely rich in hydroxyproline and serine residues, present at the ratio of 3 to 1 (ALLEN 1983), suggestive of a similar sequence to that of cell wall glycoproteins. The second, nonglycosylated, domain is rich in cystine and is involved in ligand binding (ALLEN 1983).

7 Concluding Remarks

Knowledge of the fine structure of complex carbohydrates is essential to study their structure-function relationship. Recent developments in ^{1}H-NMR spectroscopy and fast atom bombardment mass spectrometry (SWEELEY and NUNEZ 1985) have made a dramatic impact on the resolution and sensitivity of carbohydrate structural analysis. Attention is now likely to focus predominantly on *fine* structural aspects, especially those involving modifications of monosaccharide residues, because these have been shown to be of biological importance.

The interaction of proteins with carbohydrate depends on the complementary conformation of the protein and the interacting carbohydrate. An understanding of the specificity of protein–carbohydrate interactions therefore requires knowledge not only of the primary structure but also the three-dimensional structure of the carbohydrate chains. Considerable intererst is now centered on establishing the conformation of carbohydrates (CARVER and BRISSON 1984; CARVER et al. 1987), and this area of research will undoubtedly make significant contributions in the coming years to our appreciation of protein–carbohydrate interactions.

The techniques of recombinant DNA technology offer a new gateway for a more complete understanding of the biosynthesis of carbohydrates – especially the control, expression, and organization of the biosynthetic enzymes. In addi-

tion, it will ultimately provide a powerful direct approach to probe the biological function of the carbohydrate chains of glycoproteins.

Acknowledgements. I would like to thank Professor Bruce Stone, Dr. Tony Bacic and Dr. Geoff Fincher for thoughtful criticism of this manuscript, and Ms. Jenny Cooke for excellent typing.

References

Akiyama Y, Kato K (1976) Hydroxyproline arabinosides from suspension-cultured tobacco cell wall. Agric Biol Chem 40:2343–2348

Akiyama Y, Mori M, Kato K (1980) ^{13}C-NMR analysis of hydroxyproline arabinosides from *Nicotiana tabacum*. Agric Biol Chem 44:2487–2489

Albersheim P, Darrill AG (1985) Oligosaccharins. Sci Amer 253 (September):44–50

Allen AK (1983) Potato lectin – a glycoprotein with two domains. In: Goldstein IJ, Etzler ME (eds) Chemical taxonomy, molecular biology and function of plant lectins. Prog Clin Biol Res 138:71–85

Allen AK, Desai AN, Neuberger A, Creeth JM (1978) Properties of potato lectin and the nature of its glycoprotein linkages. Biochem J 171:665–674

Anderson DR, Grimes WJ (1982) Heterogeneity of asparagine-linked oligosaccharides of five glycosylation sites on immunoglobulin M heavy chain from mineral oil plasmacytoma 104E. J Biol Chem 257:14858–14864

Anderson MA, Cornish EC, Mau S-L, Williams EG, Hoggart R, Atkinson A, Bonig I, Grego B, Simpson R, Roche PJ, Haley JD, Penschow JD, Niall HD, Tregear GW, Coghlan JP, Crawford RJ, Clarke AE (1986) Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature 321:38–44

Aplin JD, Hughes RC (1982) Complex carbohydrates of the extracellular matrix structures, interactions and biological roles. Biochim Biophys Acta 694:315–418

Aon MA, Curtino JA (1985) Protein-bound glycogen is linked to tryosine residues. Biochem J 229:269–272

Ashford D, Desai NN, Allen AK, Neuberger A, O'Neill MA, Selvendran RR (1982) Structural studies of the carbohydrate moieties of the lectins from potato (*Solanum tuberosum*) tubers and thorn-apple (*Datura stramonium*) seeds. Biochem J 201:199–208

Aspinall GO (1980) Chemistry of cell wall polysaccharides. In: Priess J (ed) The biochemistry of plants. A comprehensive treatise, vol 3. Carbohydrates: structure and function. Academic, New York, pp 473–500

Bacic A, Churms SC, Stephen AM, Cohen PB, Fincher GB (1987) Fine structure of the arabinogalactan-protein from *Lolium multiflorum*. Carbohydr Res 162:85–93

Bacic A, Harris PJ, Stone BA (1988) Structure and function of plant cell walls. In: Preiss J (ed) The biochemistry of plants, vol 14, Academic, New York, pp 297–371

Badenoch-Jones J, Spencer D, Higgins TJV, Millerd A (1981) The role of glycosylation in storage-protein synthesis in developing pea seeds. Planta 153:201–209

Baenziger J, Kornfeld S (1974) Structure of the carbohydrate units of IgA immunoglobulin. J Biol Chem 249:7270–7281

Bhavanandan VP, Meyer K (1968) Studies on keratosulfates. Methylation, desulfation and acid hydrolysis studies on old human rib cartilage kerato sulfate. J Biol Chem 243:1052–1059

Bourbouze R, Akiki C, Chardon-Loriaux I, Percheron F (1982) Mise en evidence de derives de l'acide neuraminique dans des glycoproteines vegetales (1982) Carbohydr Res 106:21–30

Bourdon MA, Oldberg A, Pierschbacher M, Ruoslahti E (1985) Molecular cloning and sequence analysis of a chondroitin sulphate proteoglycan cDNA. Proc Natl Acad Sci USA, 82:1321–1325

Bowles DJ (1982) Membrane glycoproteins. In: Loewus FA, Tanner W (eds) Plant carbohydrates I. Springer, Berlin Heidelberg New York, pp 584–600 (Encyclopedia of plant physiology, new series, vol 13A)

Brennan MJ, Oldberg A, Pierschbacher MD, Ruoslahti E (1984) Chrondroitin/dermatan sulfate proteoglycan in human fetal membranes. J Biol Chem 259:13742–13750

Brockhausen I, Matta KL, Orr J, Schachter H (1985) Mucin synthesis. UDP-GlcNAc:GalNAc-R β3-*N*-acetyl-glucosaminyltransferase and UDP-GlcNAc:GlcNAc β1–3 GalNAc-R (GlcNAc to GalNAc) β6-*N*-acetyl-glucosaminyltransferase from pig and rat colon mucosa. Biochemistry 24:1866–1874

Bumol TF, Reisfeld RA (1982) Unique glycoprotein-proteoglycan complex defined by monoclonal antibody on human melanoma cells. Proc Natl Acad Sci USA, 79:1245–1249

Butler NA, Lee EYC, Whelan WJ (1977) A protein-bound glycogen component of rat liver. Carbohydr Res 55:73–82

Butler WT, Cunningham LW (1966) Evidence for the linkage of a disaccharide to hydroxylysine in tropocollagen. J Biol Chem 241:3882–3888

Butters TD, Hughes RC (1981) Isolation and characterization of mosquito cell membrane glycoproteins. Biochim Biophys Acta 640:655–671

Carlsson SR (1985) Changes in glycan branching and sialylation of the Thy-1 antigen during normal differentiation of mouse T-lymphocytes. Biochem J 226:519–525

Carlsson HE, Sundblad G, Hammarstrom A, Lonngren J (1978) Structure of some oligosaccharides derived from rat-intestinal glycoproteins. Carbohydr Res 64:181–188

Carver JP, Brisson J-R (1984) The three-dimensional structure of *N*-linked oligosaccharides. In: Ginsburg V, Robbins PW (eds) Biology of carbohydrates, vol 2. Wiley, New York, pp 289–331

Carver JR, Cumming DA, Grey AA (1988) Oligosaccharide three-dimensional structure determination by NMR. In: Ivatt R (ed) Glycoprotein Structure and Function, Plenum Press, New York (in press)

Chapman A, Kornfeld R (1979) Structure of the high mannose oligosaccharides of a human IgM myeloma protein. J Biol Chem 254:816–823

Chen J, Varner JE (1985a) Isolation and characterization of cDNA clones for carrot extension and a proline-rich 33-KDa protein. Proc Natl Acad Sci USA, 82:4399–4403

Chen J, Varner JE (1985b) An extracellular matrix protein in plants: characterization of a genomic clone for carrot extension. EMBO J 4:2145–2151

Cho YP, Chrispeels MJ (1976) Serine-*O*-galactosyl linkages in glycopeptides from carrot cell walls. Phytochemistry 15:165–169

Clarke J, Shannon LM (1976) The isolation and characterization of the glycopeptides from horseradish peroxidase isoenzyme C. Biochim Biophys Acta 427:428–442

Clarke AE, Anderson RL, Stone BA (1979) Form and function of arabinogalactans and arabinogalactan-proteins. Phytochemistry 18:521–540

Cohen RE, Ballou CE (1981) Mannoproteins: structure. In: Loewus FA, Tanner W (eds) Plant carbohydrates II. Springer, Berlin Heidelberg New York, pp 441–458 (Encyclopedia of plant physiology, new series, vol 138)

Coudron C, Koerner T, Jacobson I, Rodin L, Schwartz NB (1980) Peptides with alternating serine and glycine residues as substrates for xylosyltransferase (Abstract 346). Fed Proc Am Soc Exp Biol 39:1671

Creek KE, Sly WS (1984) The role of the phosphomannosyl receptor in the transport of acid hydrolases to lysosomes. In: Dingle JT, Dean RJ (eds) Lysosomes in biology and pathology, vol 7. Elsevier, Amsterdam, pp 63–82

Cummings RD, Kornfeld S, Schneider WJ, Hobgood KK, Tolleshaug H, Brown MS, Goldstein JL (1983) Biosynthesis of *N*- and *O*-linked oligosaccharides of the low density lipoprotein receptor. J Biol Chem 258:15261–15273

Cummings RD, Soderquist AM, Carpenter G (1985) The oligosaccharide moieties of the epidermal growth factor receptor in A-431 cells. J Biol Chem 260:11944–11952

Darvill AG, Albersheim P (1984) Phytoalexins and their elicitors-A defence against microbial infection in plants. Annu Rev Plant Physiol 35:243–275

Darvill AG, Albersheim P, Lau JM, McNeil M, Stevenson TT, Thomas J, York WG (1985) The structure of plant cell wall polysaccharides (Abstract OR-10-06). 1st international congress plant mol biol, Savannah

Davies HM, Delmer DP (1979) Seed reserve-protein glycosylation in an in vitro preparation from developing cotyledons of *Phaseolus vulgaris*. Plantamed 146:513–520

Desai NN, Allen AK, Neuberger A (1981) Some properties of the lectin from *Datura stramonium* (thorn-apple) and the nature of its glycoprotein linkages. Biochem J 197:345–353

Doege K, Fernandez P, Hassell JR, Sasaki M, Yamada Y (1986) Partial cDNA sequencing encoding a globular domain at the C terminus of the rat cartilage proteoglycan. J Biol Chem 261:8108–8111

Dorfman A (1981) Proteoglycan biosynthesis. In: Hay ED (ed) Cell biology of the extracellular matrix. Plenum, New York, pp 115–138

Dorland L, van Halbeek H, Vliegenthart JFG, Lis H, Sharon N (1981) Primary structure of the carbohydrate chain of soybean agglutinin. J Biol Chem 256:7708–7711

Edge ASB, Spiro RG (1984) Presence of sulfate in *N*-glycosidically linked carbohydrate units of calf thyroid plasma membrane glycoproteins. J Biol Chem 259:4710–4713

Elbein AD (1982) Glycolipids and other glycosides. In: Loewus FA, Tanner W (eds) Plant carbohydrates I. Springer, Berlin Heidelberg New York, pp 601–612 (Encyclopedia of plant physiology, new series, vol 13A)

Elhammer A, Kornfeld S (1984) Two enzymes involved in the synthesis of *O*-linked oligosaccharides are localized on membranes of different densities in mouse lymphoma BW 5147 cells. J Cell Biol 98:327–331

Feizi T (1985) Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. Nature 341:53–57

Feizi T, Childs RA (1985) Carbohydrate structures of glycoproteins and glycolipids as differentiation antigens, tumour-associated antigens and components of receptor systems. Trends Biochem Sci 10:24–29

Ferguson MAJ, Allen AK, Snary D (1983) Studies on the structure of a phosphoglycoprotein from the parasitic protozoan *Tryponosoma cruzi*. Biochem J 213:313–319

Fincher GB, Stone BA (1987) Cell walls and their components in cereal grain technology. In: Pomeranz Y (ed) Advances in Cereal Science and Technology. American Association of Cereal Chemists, St. Paul, Minnesota, USA, pp 207–295

Fincher GB, Stone BA, Clarke AE (1983) Arabinogalactan-proteins: structure, biosynthesis and function. Annu Rev Plant Physiol 34:47–70

Fournet B, Montreuil J, Strecker G, Dorland L, Haverkamp J, Vliegenthart JFG, Binette JP, Schmid K (1978) Determination of the primary structures of 16 asialo-carbohydrate units derived from human plasma α_1-acid glycoprotein by 360-MHz ^{1}H NMR spectroscopy and permethylation analysis. Biochemistry 17:5206–5214

Fransson L-A (1985) Mammalian glycosaminoglycans. In: Aspinall GO (ed) The polysaccharides, vol. 3, Academic, New York, London Toronto, pp 337–415

Friedenson B, Liener IE (1974) Evidence that ficin is a glycoprotein. Biochim Biophys Acta 342:209–212

Fukuda M (1985) Cell surface glycoconjugates as onco-differentiation markers in hematopoietic cells. Biochim Biophys Acta 780:119–150

Fukuda K, Tomita M, Hamada A (1981) Isolation and structural studies of the neutral oligosaccharide units from bovine glycophorin. Biochim Biophys Acta 677:462–470

Fukuda K, Kawashima I, Tomita M, Hamada A (1982) Structural studies of the acidic oligosaccharide units from bovine glycophorin. Biochim Biophys Acta 717:278–288

Funakoshi I, Yamashina I (1982) Structure of *O*-glycosidically linked sugar units from plasma membranes of an ascites hepatoma, AH 66. J Biol Chem 257:3782–3787

Gell A, Bacic A, Clarke AE (1986) Arabinogalactan-protein of the female sexual tissues of *Nicotiana alata*. I. Changes during flowering development and pollination. Plant Physiol 82:885–889

Gleeson PA, Clarke AE (1979) Structural studies on the major component of *Gladiolus* style mucilage, an arabinogalactan protein. Biochem J 181:607–621

Gleeson PA, Clarke AE (1980) Arabinogalactans of sexual and somatic tissues of *Gladiolus* and *Lilium*. Phytochemistry 19:1777–1782

Gleeson PA, Feeney J, Mills G, Hughes RC (1984) Galactosyl transferases of baby hamster kidney (BHK) cells: characterization of two oligosaccharide products synthesized using bovine asialo-submaxillary-gland mucin as acceptor. Eur J Biochem 144:143–150

Gleeson PA, Stone BA, Fincher GB (1985a) Cloning the cDNA for the ryegrass arabinogalactan-protein (Abstract OR-10-07). 1st international congress on plant mol biol, Savannah

Gleeson PA, Stone BA, Fincher GB (1985b) Cloning the cDNA for the arabinogalactan-protein from *Lolium multiflorum* (ryegrass). AGP News 5:30–36

Goldberg D, Gabel C, Kornfeld S (1984) Processing of lysosomal enzyme oligosaccharide units. In: Dingle JT, Dean RJ (eds) Lysosomes in biology and pathology, vol 7. Elsevier, Amsterdam, pp 45–62

Goldstein JL, Brown MS, Anderson RGW, Russell DW, Schneider WJ (1985) Receptor-mediated endocytosis: Concepts emerging from the LDL receptor system. Ann Rev Cell Biol 1:1–39

Green ED, Gruenebaum J, Bielinska M, Baenziger JU, Boime I (1984) Sulfation of lutropin oligosaccharides with a cell-free system. Proc Natl Acad Sci USA, 81:5320–5324

Green ED, Van Halbeek H, Boime I, Baenziger JU (1985a) Structural elucidation of the disulfated oligosaccharide from bovine lutropin. J Biol Chem 260:15623–15630

Green ED, Morishima C, Boime I, Baenziger JU (1985b) Structural requirements for sulfation of asparagine-linked oligosaccharides of lutropin. Proc Natl Acad Sci USA, 82:7850–7854

Green JR, Northcote DH (1978) The structure and function of glycoproteins synthesized during slime-polysaccharide production by membranes of the root-cap cells of maize (*Zea mays*). Biochem J 170:599–608

Hakomori S (1981) Glycosphingolipids in cellular interaction, differentiation and oncogenesis. Annu Rev Biochem 50:733–764

Hanover JA, Elting J, Mintz GR, Lennarz WJ (1982) Temporal aspects of the *N*- and *O*-glycosylation of human chorionic gonadotropin. J Biol Chem 257:10172–10177

Hascall VC, Hascall GK (1981) Proteoglycans. In: Hay ED (ed) Cell biology of the extracellular matrix. Plenum, New York, pp 39–63

Hassell JR, Kimura JH, Hascall VC (1986) Proteoglycan core protein families. Ann Rev Biochem 55:539–567

Hatton MWC, Marz L, Berry LR, Debanne MT, Regoeczi E (1979) Bi- and tri-antennary human transferrin glycopeptides and their affinities for the hepatic lectin specific for asialo-glycoproteins. Biochem J 181:633–638

Heaney-Kieras J, Roden L, Chapman DJ (1977) The covalent linkage of protein to carbohydrate in the extracellular protein-polysaccharide from the red alga *Porphyridium cruentum*. Biochem J 165:1–9

Heifetz A, Kinsey WH, Lennarz WJ (1980) Synthesis of a novel class of sulfated glycoproteins in embryonic liver and lung. J Biol Chem 255:4528–4534

Hill HD, Reynolds JA, Hill RL (1977) Purification, composition, molecular weight and subunit structure of ovine submaxillary mucin. J Biol Chem 252:3791–3798

Hillestad A, Wold JK, Paulsen BS (1977) Structural studies of water-soluble glycoproteins from *Cannabis sativa*. Carbohydr Res 57:135–144

Hodges LC, Laine R, Chan SK (1979) Structure of the oligosaccharide chains in human α_1-protease inhibitor. J Biol Chem 254:8208–8212

Hoffman RA, Mutsaers JHGM, van Kuik JA, van Halbeek H, Kamerling JP, Vliegenthart JFG (1984) A 500-MHZ ^{1}H-MNR study of the oligosaccharide chain from stem bromelain, having fucose in (1 → 3)-linked to asparagine-bound *N*-acetylglucosamine. XIIth international carbohydrates symposium, Utrecht, The Netherlands, p 431

Hohmann H-P, Gerish G, Lee RWH, Huttner WB (1985) Cell-free sulfation of the contact site A glycoprotein of *Dictyostelium discoideum* and of a partially glycosylated precursor. J Biol Chem 260:13869–13878

Holder AA (1985) Glycosylation of the variant surface antigens of *Trypanosoma brucei*. Curr Top Microbiol Immunol 117:57–74

Hook M, Kjellen L, Johansson S, Robinson J (1984) Cell-surface glycosaminoglycans. Annu Rev Biochem 53:847–869

Horejsi V, Kocourek J (1978) Studies on lectins XXXVII. Isolation and characterization of the lectin from jimson-weed seeds (*Datura stramonium L*). Biochim Biophys Acta 532:92–97

Hori H (1978) The demonstration of galactosamine in extracellular hydroxyproline-rich macromolecule purified from the culture media of tobacco cells. Plant Cell Physiol 19:501–505

Hori H, Elbein AD (1983) Processing of *N*-linked oligosaccharides in soybean cultured cells. Arch Biochem Biophys 220:415–425

Hori H, Kaushal GP, Elbein AD (1985) Fucosylation of membrane proteins in soybean cultured cells. Plant Physiol 77:687–694

Hounsell EF, Wood E, Feizi T, Fukuda M, Powell ME, Hakomori S (1981) Structural analysis of hexa- to octa-saccharide fractions isolated from sheep gastric glycoprotein having blood group I and i activities. Carbohydr Res 90:283–307

Hounsell EF, Lawson AM, Feizi T (1982) Structure and antigenic diversity in mucin carbohydrate chains. In: Chantler EN, Elder JB, Elstein M (eds) Mucin in health and disease, vol II. Plenum, New York, pp 39–41

Hsieh P, Robbins PW (1984) Regulation of asparagine-linked oligosaccharide processing. Oligosaccharide processing in *Aedes albopictus* mosquito cells. J Biol Chem 259:2375–2382

Hughes RC (1983) Glycoproteins. Chapman and Hill, London

Ishihara H, Takahashi N, Oguri S, Tejima S (1979) Complete structure of the carbohydrate moiety of stem bromelain. J Biol Chem 254:10715–10719

Ivatt RL, Dao OP, Henderson EJ, Robbins PW (1984) Glycoprotein biosynthesis in *Dictyostelium discoideum*. Developmental regulation of the protein-linked glycans. Cell 38:561–567

Jermyn MA, Guthrie R (1985) A final assult on the structure of carrot AGPs. AGP News 5:4–25

Johnson DC, Spear PG (1983) *O*-linked oligosaccharides are acquired by Herpes simplex virus glycoproteins in the Golgi apparatus. Cell 32:987–997

Kanwar YS, Veis A, Kimura JH, Jakubowski ML (1984) Characterization of heparan sulfate-proteoglycan of glomerular basement membranes. Proc Natl Acad Sci USA 81:762–766

Kawasaki T, Ashwell G (1976) Chemical and physical properties of an hepatic membrane protein specifically binds asialoglycoproteins. J Biol Chem 251:1296–1302

Keller R, Stein T, Stuhlsatz HW, Greiling H, Ohst E (1981) Studies on the characterization of the linkage-region between polysaccharide chain and core protein in bovine corneal proteokeratan sulfate Hoppe-Seyler's Z. Physiol Chem 362:327–336

Kessler MJ, Mise T, Ghai RD, Bahl OP (1979) Structure and location of the *O*-glycosidic carbohydrate units of human chorionic gonadotropin. J Biol Chem 254:7909–7914

Kitagaki-Ogawa H, Matsumoto I, Seno N, Takahashi N, Endo S, Arata Y (1986) Characterization of the carbohydrate moiety of *Clerodendron trichotomum* lectins. Eur J Biochem 161:779–785

Kobata A (1984) The carbohydrates of glycoproteins. In: Ginsburg V, Robbins PW (eds) Biology of carbohydrates, vol 2. Wiley, New York, pp 87–162

Kornfeld R, Kornfeld S (1985) Assembly of asparagine-linked oligosaccharides. Annu Rev Biochem 54:631–664

Lamport DTA (1967) Hydroxyproline-*O*-glycosidic linkage of the plant cell wall glycoprotein extensin. Nature 216:1322–1324

Lamport DTA (1977) Structure, biosynthesis and significance of cell wall glycoproteins. In: Loewus FA, Runeckles BC (eds) Recent advances in phytochemistry, vol 11. Plenum, New York, pp 79–115

Lamport DTA, Catt JW (1981) Glycoproteins and enzymes of the cell wall. In: Tanner W, Loewus FA (eds) Plant carbohydrates II. Springer, Berlin Heidelberg New York, pp 133–165 (Encyclopedia of plant physiology, new series, vol 138)

Lamport DTA, Miller DM (1971) Hydroxyproline arabinosides in the plant kingdom. Plant Physiol 48:454–456

Lamport DTA, Katona L, Roerig S (1973) Galactosylserine in extensin. Biochem J 133:125–131

Lawson CJ, Rees DA (1970) An enzyme for the metabolic control of polysaccharide conformation and function. Nature 227:193

Lee YC, Scocca JR (1972) A common structural unit in asparagine-oligosaccharides of several glycoproteins from different sources. J Biol Chem 247:5753–5758

Lehle L (1981) Plant cells synthesize glucose-containing lipid-linked oligosaccharides similar to those found in animals and yeast. FEBS Lett 123:63–66

Leonard WJ, Depper JM, Crabtree GR, Rudikoff S, Pumphrey J, Robb RJ, Kronke M, Svetlik PB, Peffer NJ, Waldmann TA, Greene WC (1984) Molecular cloning and expression of cDNAs for the human interleukin-2 receptor. Nature 311:626–631

Lindahl V, Roden L (1965) The role of galactose and xylose in the linkage of heparin to protein. J Biol Chem 240:2821–2826

Lohmander LS, DeLuca S, Nilsson B, Hascall VC, Caputo CB, Kimura JH, Heinegard D (1980) Oligosaccharides on proteoglycans from the swarm rat chondrosarcoma. J Biol Chem 255:6084–6091

Malmstrom A, Fransson LA, Hook LA, Lindahl V (1975) Biosynthesis of dermatan sulfate I. Formation of L-induronic acid residues. J Biol Chem 250:3419–3425

Manners DJ, Sturgeon RJ (1982) Reserve carbohydrates of algae, fungi and lichens. In: Loewus FA, Tanner W (eds) Plant carbohydrates II. Springer, Berlin Heidelberg New York, pp 472–514 (Encyclopedia of plant physiology, new Series, vol 13A)

Marchase RB, Koro LA, Kelly CM, McClay DR (1982) Retinal ligatin recognizes glycoproteins bearing oligosaccharides terminating in phosphodiester-linked glucose. Cell 28:813–820

Marshall RD, Neuberger A (1964) Carbohydrates in protein VIII. The isolation of 2-acetamido-1-(L-4′-aspartamido)-1,2-dideoxy-β-D-glucose from hen's egg albumin. Biochemistry 3:1596–1600

Marshall RD, Neuberger A (1970) Aspects of the structure and metabolism of glycoproteins. Adv Carbohydr Chem Biochem 25:407–478

McNamara MK, Stone BA (1981) Isolation, characterization and chemical synthesis of a galactosyl-hydroxyproline linkage compound from wheat endosperm arabinogalactan-peptide. Lebensm Wiss Technol 14:182–187

McNeil M, Darvill AG, Fry SC, Albersheim P (1984) Structure and function of the primary cell walls of plants. Annu Rev Biochem 53:625–663

Miller DH, Lamport DTA, Miller M (1972) Hydroxyproline hetero-oligosaccharides in Chamydomonas. Science 176:918–920

Misaki A, Goldstein IJ (1977) Glycosyl moiety of the lima bean lectin. J Biol Chem 253:6995–6999

Miyazaki T, Yadomae T, Tsuchihashi H (1984) Distribution and structures of fungal glycuronans. XIIth international carbohydrates symposium, Utrecht, The Netherlands, p 389

Montreuil J (1982) Glycoproteins. In: Florkin M, Stotz EH (eds) Comprehensive biochemistry, vol 19B, part II. Elsevier, Amsterdam, pp 1–188

Morrison MH, Lynch RA, Esselman WJ (1986) Poly-N-acetyllactosamine alterations of Thy-1 glycoprotein in lymphocyte differentiation. Molec Immunol 23:63–72

Muir H (1983) Proteoglycans as organizers of the intercellular matrix. Biochem Soc Trans 11:613–622

Muir L, Lee YC (1969) Structures of the D-galactose oligosaccharides from earthworm cuticle collagen. J Biol Chem 244:2343–2349

Murray RH, Northcote DH (1978) Oligoarabinosides of hydroxyproline isolated from potato lectin. Phytochemistry 17:623–629

Nagahashi J, Browder SK, Beevers L (1980) Glycosylation of pea cotyledon membranes. Plant Physiol 65:648–657

Nanno S, Ishiguro M, Funatsu G, Funatsu M (1975) The mode of binding of carbohydrate in ricin D. Agric Biol Chem 39:1615–1654

Narasimhan S, Harpaz N, Longmore G, Carver JP, Grey AA, Schachter H (1980) Control of glycoprotein synthesis. The purification by preparative high voltage paper electrophoresis in borate of glycopeptides containing high mannose and complex oligosaccharide chains linked to asparagine. J Biol Chem 255:4876–4884

Nikaido T, Shimizu A, Ishida N, Sabe H, Teshigawara K, Maeda M, Uchiyama T, Yudoi J, Honjo T (1984) Molecular cloning of cDNA encoding human interleukin-2 receptor. Nature 311:631–635

Nilsson B, De Luca S, Lohmander S, Hascall VC (1982) Structures of *N*-linked and *O*-linked oligosaccharides on proteoglycan monomer isolated from the swarm rat chondrosarcoma. J Biol Chem 257:10920–10927

Nilsson B, Nakazawa K, Hassell JR, Newsome DA, Hascall VC (1983) Structure of the oligosaccharide and the linkage region between keratan sulphate and the core protein on proteoglycans from monkey cornea. J Biol Chem 258:6065–6068

Ohtani K, Misaki A (1980) The structure of the glycan moiety of tora bean (*Phaseolus vulgaris*) lectin. Carbohydr Res 87:275–285

Olsnes S, Pihl A (1982) Toxic lectins and related proteins. In: Cohen P, van Heynmagen S (eds) Molecular action of toxins and viruses, Elsevier Biomedical, Amsterdam, pp 51–105

Olsnes S, Refsnes K, Christensen TB, Pihl A (1975) Studies on the structure and properties of the lectins from *Abrus precatorius and Ricinus communis*. Biochim Biophys Acta 405:1–10

O'Neill MA, Selvendran RR (1980) Glycoproteins from the cell wall of *Phaseolus coccineus*. Biochem J 187:53–63

Owens RJ, Northcote DH (1981) The location of arabinosyl:hydroxyproline transferase in the membrane system of potato tissue culture cells. Biochem J 195:661–667

Painter TJ (1983) Algal polysaccharides. In: Aspinall GO (ed) The Polysaccharides, vol. 2, Academic, London New York, pp 195–285

Parkkinen J, Finne J (1985) Occurrence of *N*-acetylglucosamine 6-phosphate in complex carbohydrates. J Biol Chem 260:10971–10975

Parsons TF, Pierce JG (1980) Oligosaccharide moieties of glycoprotein hormones. Bovine lutropin resists enzymatic deglycosylation because of terminal *O*-sulfated *N*-acetylhexosamines. Proc Natl Acad Sci USA 77:7089–7093

Percival E, McDowell RH (1981) Algal walls – composition and biosynthesis. In: Loewus FA, Tanner W (eds) Plant carbohydrates II. Springer, Berlin Heidelberg New York, pp 441–458 (Encyclopedia of plant physiology, new Series, vol 13B)

Pierce-Cretel A, Pamblanco M, Strecker G, Montreuil J, Spik G (1981) Heterogeneity of the glycans *O*-glycosidically linked to the hinge region of the secretory immunoglobulins from human milk. Eur J Biochem 114:169–178

Pope DG (1977) Relationship between hydroxyproline-containing proteins secreted into the cell wall and medium by suspension cultured *Acer pseudoplatanus* cells. Plant Physiol 59:894–900

Powell JT, Brew K (1974) Glycosyltransferases in the Golgi membranes of onion stem. Biochem J 142:203–209

Prehm P, Schied A, Choppin PW (1979) The carbohydrate structure of the glycoproteins of the paramyxovirus SV5 grown in bovine kidney cells. J Biol Chem 254:9669–9677

Prigent M-J, Montreuil J, Strecker G (1984) Preliminary investigation of the structure of the carbohydrate component of *Vicia graminea* lectin, a plant glycoprotein. Carbohydr Res 131:83–92

Rapraeger A, Bernfield M (1985) Cell surface proteoglycan of mammary epithelial cells. J Biol Chem 260:4103–4109

Ratner N, Bunge RP, Glaser L (1985) A neuronal cell surface heparan sulphate proteoglycan is required for dorsal root ganglion neuron stimulation of Schwann cell proliferation. J Cell Biol 101:744–754

Rees DA (1977) Polysaccharide shapes. Chapman and Hall, London

Robinson HC, Horner AA, Hook M, Ogren S, Lindahl U (1978) A proteoglycan form of heparin and its degradation to single-chain molecules. J Biol Chem 253:6687–6693

Roden L, Horowitz MI (1978) Structure and biosynthesis of connective tissue proteoglycans. In: Pigman W, Horowitz MI (eds) The glycoconjugates, vol 2. Academic, New York, pp 3–32

Rodreguez IR, Whelan WJ (1985) A novel glycosyl-amino acid linkage: Rabbit-muscle glycogen is covalently linked to a protein via tyrosine. Biochem Biophys Res Commun 132:829–836

Roth J (1984) Cytochemical localization of terminal *N*-acetyl-D-galactosamine residues in cellular compartments of intestinal goblet cells: implications for the topology of *O*-glycosylation. J Cell Biol 98:399–406

Ryan CA (1978) Proteinase inhibitors in plant leaves. A biochemical model for pest-induced natural plant protection. Trends Biochem Sci 3:148–150

Ryan CA (1984) In: Kosuge T, Nester EW (eds) Plant-microbe interactions: Molecular and genetic perspectives, vol. 1, Macmillan, New York, pp 307–320

Sadler JE (1984) Biosynthesis of glycoproteins: formation of *O*-linked oligosaccharides. In: Ginsburg V, Robbins PW (eds) Biology of carbohydrates, vol 2. Wiley, New York, pp 199–288

Sai S, Tanaka T, Kosher RA, Tanzer ML (1986) Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein. Proc Natl Acad Sci USA 83:5081–5085

Sairam MR, Bhargavi GN (1985) A role for glycosylation of the subunit in transduction of biological signal in glycoprotein hormones. Science 229:65–67

Schachter H, Williams D (1982) Biosynthesis of mucus glycoproteins. In: Chantler EN, Elder JB, Elstein M (eds) Mucus in health and disease, vol II. Plenum, New York, pp 3–28

Schachter H, Narasimhan S, Gleeson P, Vella G (1983) Control of branching during the biosynthesis of asparagine-linked oligosaccharides. Can J Biochem Cell Biol 61:1049–1066

Schachter H, Narasimhan S, Gleeson P, Vella G, Brockhausen I (1985) Glycosyltransferase involved in the biosynthesis of protein-bound oligosaccharides of the asparagine-*N*-acetyl-D-glucosamine and serine (threonine)-*N*-acetyl-D-galactosamine types. In: Martonosi AN (ed) The enzymes of biological membranes, 2nd edn, vol 2. Plenum, New York, pp 227–277

Schauer R (1982) Chemistry, metabolism and biological functions of sialic acids. Adv Carbohydr Chem Biochem 40:131–234

Selvendran RR, O'Neill MA (1982) Plant glycoproteins. In: Loewus FA, Tanner W (eds) Plant carbohydrates I. Springer, Berlin Heidelberg New York, pp 515–583 (Encyclopedia of plant physiology, vol 13A

Sentandreu R, Northcote DH (1968) The structure of a glycopeptide isolated from the yeast cell wall. Biochem J 109:419–432

Sharon N, Lis H (1979) Comparative biochemistry of plant glycoproteins. Biochem Soc Trans 7:783–799

Showalter AM, Varner JE (1988) Biology and molecular biology of plant hydroxyproline-rich glycoproteins. In: Marcus A (ed) The biochemistry of plants: A comprehensive treatise, vol. 11, Molecular biology, Academic, New York (in press)

Slomiany BL, Murty VLN, Slomiany A (1980) Isolation and characterization of oligosaccharides from rat colonic mucosa glycoprotein. J Biol Chem 255:9719–9723

Smith JJ, Muldoon EP, Lamport DTA (1984) Isolation of extensin precursors by direct elution of intact tomato cell suspension cultures. Phytochemistry 23:1233–1239

Spik G, Bayard B, Fourret B, Strecker G, Bouquelet S, Montreuil J (1975) Studies on glycoconjugates. LXIV complete structure of two carbohydrate units of human serotransferrin. FEBS Lett 50:296–299

Spiro RG, Bhoyroo VD (1974) Structure of the *O*-glycosidally linked carbohydrate units of fetuin. J Biol Chem 249:5704–5714

Steneloni RJ, Tolmasky ME, Petriella C, Leloir LF (1981) Transfer of oligosaccharide to protein from a lipid intermediate in plants. Plant Physiol 68:1175–1179

Stephen AM (1980) Plant carbohydrates. In: Bell EA, Charlwood BV (eds) Secondary plant products. Springer, Berlin Heidelberg New York, pp 555–584 (Encyclopedia of plant physiology, new series, vol 8)

Strahm A, Amado R, Neukom H (1981) Hydroxyproline-galactoside as a protein-polysaccharide linkage in a water soluble arabinogalactan-peptide from wheat endosperm. Phytochemistry 20:1061–1063

Sturm A, Chrispeels MJ (1986) The high mannose oligosaccharide of phytohaemagglutinin is attached to asparagine 12 and the modified oligosaccharide to asparagine 60. Plant Physiol 80:320–322

Sweeley CC, Nunez HA (1985) Structural analysis of glycoconjugates by mass spectrometry and nuclear magnetic resonance spectroscopy. Ann Rev Biochem 54:765–801

Swiedler SJ, Freed JH, Tarentino AL, Plummer TH, Hart GW (1985) Oligosaccharide microheterogeneity of the murine major histocompatibility antigens. J Biol Chem 260:4046–4054

Takahashi N, Hotta T, Ishihara H, Mori M, Tejima S, Bligny R, Akazawa T, Endo S, Arata Y (1986) Xylose-containing common structural unit in *N*-linked oligosaccharides of laccase from sycamore cells. Biochemistry 25:388–395

Takayama S, Isogai A, Tsukamoto C, Ueda Y, Hinata K, Okazaki K, Suzuki A (1987) Sequences of S-glycoproteins, products of the *Brassica campestris* self-incompatibility locus. Nature 326:102–105

Thomas DB, Winzler RJ (1969) Structural studies on human erythrocyte glycoproteins. Alkaline-labile oligosaccharides. J Biol Chem 244:5943–5946

Toole BP (1981) Glycosaminoglycans in morphogenesis. In: Hay ED (ed) Cell biology of the extracellular matrix. Plenum, New York, pp 259–294

Torres C-R, Hart GW (1984) Topography and polypeptide distribution of terminal *N*-acetylglucosamine residues on the surfaces of intact lymphocytes. J Biol Chem 259:3308–3317

Tran Thanh Van K, Toubart P, Cousson A, Darvill AG, Gollin DJ, Chelf P, Albersheim P (1985) Manipulation of the morphogenetic pathways of tobacco explants by oligosaccharides. Nature 314:615–617

Vandenheede JR, Ahmed AI, Feeney RE (1972) Structure and role of carbohydrate in freezing point-depressing glycoproteins from an antarctic fish. J Biol Chem 247:7885–7889

Van Halbeek H, Dorland L, Haverkamp J, Veldink GA, Vliegenthart JFG, Fournet B, Ricart G, Montreuil J, Gathmann WD, Aminoff D (1980) Structure determination of oligosaccharides isolated from A^+, H^+ and A-H-hog-submaxillary-gland mucin glycoproteins, by 360-MHz ^{1}H-NMR spectroscopy, permethylation analysis and mass spectrometry. Eur J Biochem 118:487–495

Van Halbeek H, Dorland L, Vliegenthart JFG, Hull WE, Lamblin G, Lhermitte M, Boersma A, Roussel P (1982) Primary-structure determination of fourteen neutral oligosaccharides derived

from bronchial mucus glycoproteins of patients suffering from cystic fibrosis, employing 500-MHz ^{1}H-NMR spectroscopy. Eur J Biochem 127:7–20

van Holst G-J, Martin SR, Allen AK, Ashford D, Desai NN, Neuberger A (1986) Protein conformation of potato (*Solanum tuberosum*) lectin determined by circular dichroism. Biochem J 233:731–736

van Kuik JA, van Halbeek H, Komerling JP, Vliegenthart JFG (1985) Primary structure of the low-molecular-weight carbohydrate chains of *Helix pomatia* α_1 hemocyanin. J Biol Chem 260:13984–15988

van Kuik JA, Sijbesma RP, Kamerling JP, Vleigenthart JFG, Wood EJ (1986) Primary structure of a low-molecular-mass *N*-linked oligosaccharide from hemocyanin of *Lymnaea stagnalis*. Eur J Biochem 160:621–625

Walker-Simmons M, Hadwiger L, Ryan CA (1983) Chitosans and pectic polysaccharides both induce the accumulation of the antifungal phytoalexin in pisatin in pea pods and antinutrient proteinase inhibitors in tomato leaves. Biochem Biophys Res Commun 110:194–199

Warren L (1963) The distribution of sialic acids in nature. Comp Biochem Physiol 10:153–171

Wessels JGH, Sietsma JH (1981) Fungal cell walls: a survey. In: Loewus FA, Tanner W (eds) Plant carbohydrates II. Springer, Berlin Heidelberg New York, pp 352–394 (Encyclopedia of plant physiology, new series, vol 13B)

Wiegandt H (ed) (1985) Glycolipids. New comprehensive biochemistry, vol 10. Elsevier, Amsterdam

Yamashita K, Veda I, Kobata A (1983) Sulfated asparagine-lined sugar chains of hen egg albumin. J Biol Chem 258:14144–14147

Yamauchi F, Yamagishi T (1979) Carbohydrate sequence of a soybean 7S protein. Agric Biol Chem 43:505–510

York WS, Darvill AG, McNeil M, Albersheim P (1985) 3-Deoxy-D-manno-2-octulosonic acid (KDO) is a component of rhamnogalacturonan II, a peptic polysaccharide in the primary cell walls of plants. Carbohydr Res 138:109–126

Yoshima H, Furthmayr H, Kobata A (1980) Structures of the asparagine-linked sugar chains of glycophorin A. J Biol Chem 255:9713–9718

Young JD, Tsuchiya D, Sandlin DE, Holroyde MJ (1979) Enzymic *O*-glycosylation of synthetic peptides from sequences in bacic myelin protein. Biochemistry 18:444–448

Carbohydrate-Binding Sites of Plant Lectins

G.N. REEKE Jr. and J.W. BECKER

1 Introduction

Lectins are a diverse group of proteins and glycoproteins that exhibit specific binding for certain carbohydrates. Proteins with this property have been described in a wide variety of taxa, ranging from bacteria (e.g. NETER 1956) to slime molds (BARONDES and HAYWOOD 1979) and lower vertebrates (SIMPSON et al. 1978), but the best characterized examples are from plants. These proteins were originally characterized as cell agglutinins, until the term lectin was introduced by BOYD (1954, 1963) to encompass a larger range of activities involving selectivity for specific saccharides. There is no universally accepted definition

Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

Current Topics in Microbiology and Immunology, Vol. 139

that specifies exactly the functional classes of proteins included in the lectins, but it is generally agreed that such well-known groups as antibodies, enzymes, and transport proteins are *not* included. These groups are covered separately in later chapters of this book.

The sources and properties of plant lectins have been extensively reviewed (MÄKELÄ 1957; SHARON and LIS 1972; LIS and SHARON 1973, 1977; BROWN and HUNT 1978; LUTSIK et al. 1981). The great functional diversity of these proteins suggests that they by no means constitute a monolithic class, and what is known of their structures tends to confirm this picture. The major exception at present is the lectins from the legumes, which include three of the four examples for which crystal structures are known. These proteins are very similar to each other in three-dimensional structure and also in amino acid sequence when an extraordinary circular permutation of their sequences is taken into account. They are undoubtedly all homologous, making a comparison of their structures particularly interesting. Such comparisons will be made following the discussions of the individual structures in this chapter. As additional structures become known for lectins from nonleguminous plants, the relationships among these lectins should be further clarified and new structural and functional classes revealed.

1.1 Activities of Lectins

The actual functions of lectins in the organisms that produce them are for the most part unknown. The functions are undoubtedly as diverse as the proteins themselves, having in common only their dependence on the carbohydrate-binding activity, which is in all cases specific and reversible but not catalytic (the bound molecules do not undergo chemical changes). Proposed functions for plant lectins include a storage or transport role for carbohydrates in seeds (ENSGRABER 1958; KAUSS and GLASER 1974), binding of nitrogen-fixing bacteria to root hairs (HAMBLIN and KENT 1973), and inhibition of fungal growth (ALBERSHEIM and ANDERSON 1971) or insect feeding (JANZEN et al. 1976).

Whatever the natural functions of lectins may be, their carbohydrate-binding properties imbue them with many other activities of a more or less adventitious nature. These activities have made them quite useful reagents in biochemistry and cell biology (see e.g., BITTIGER and SCHNEBLI 1976; LIS and SHARON 1986). We mention here only a sample of these applications.

1. Lectins have been used as affinity reagents to isolate saccharide-containing biological molecules (BITTIGER and SCHNEBLI 1976) and viruses (STEWART et al. 1973).
2. Lectins bind to and may be used as labels for a variety of subcellular structures including chromatin (RIZZO and BUSTIN 1977), mitochondria (GLEW et al. 1973), ribosomes (HOWARD and SCHNEBLI 1977), neurites (FELDMAN et al. 1982), and synaptic vesicles (MARCH and THORNTON 1983).
3. Lectins bind to cell surfaces and may be used to distinguish cells based on their blood-type antigens (BIRD 1959; BOYD 1963), developmental or transformed state, or numerous other variables (NICOLSON 1974).
4. Lectin binding induces changes in plasma membrane fluidity (BARNETT et al. 1974) and receptor mobility (SCHLESSINGER et al. 1976; SCHLESSINGER et al. 1977). These changes are associated with rearrangements in the distribution of lectin receptors (UNANUE et al. 1972; YAHARA and EDELMAN

1973; DEPETRIS et al. 1973), receptors for other cell-surface ligands (YAHARA and EDELMAN 1972), and submembranous cytoskeletal assemblies (EDELMAN et al. 1973; ALBERTINI and CLARK 1976).
5. Lectins compete with hormones such as epidermal growth factor (EGF) (BALLMER and BURGER 1980) and insulin (CUATRECASAS 1973a) for binding to their cell-surface receptors and as a result may mimic or inhibit the activities of the hormones (CUATRECASAS and TELL 1973) and cause changes in cyclic nucleotide levels (HADDEN et al. 1972).
6. Lectins induce maturational events in cells such as oocytes (KUBOTA and KANATANI 1975), macrophages (SMITH and GOLDMAN 1972; LEAK and SUN 1984), and platelets (SANTORO 1983).
7. Lectins induce blast transformation and mitosis in lymphocytes (BECKERT and SHARKEY 1970; ANDERSSON et al. 1972) and other cells.

Competing stimulatory and inhibitory effects with different mechanisms (WANG et al. 1975b; MCCLAIN and EDELMAN 1976) may be distinguished with chemically modified lectins of reduced valency (GUNTHER et al. 1973; WANG and EDELMAN 1978; TANAKA et al. 1981). In these processes cell-surface lectin receptors interact with cytoskeletal components and affect their motility and state of assembly (EDELMAN et al. 1973; AUBIN et al. 1975; ALBERTINI and CLARK 1976; ASH and SINGER 1976; MCCLAIN et al. 1977). A general theory of surface modulation has been proposed to unify and explain these phenomena (EDELMAN et al. 1973; EDELMAN 1976).

1.2 Classes of Plant Lectins

Lectins may be classified in various ways, perhaps most obviously according to the taxonomy of the plants from which they are derived. For the design of experimental applications, a classification by saccharide-binding specificity is useful, and, indeed, large compendia of such information are available (e.g., GOLDSTEIN and HAYES 1978). However, for our purposes it will be more instructive to compare the known three-dimensional structures from the point of view of their evolutionary relationships. The differences in binding specificity will then be seen as variations on a common structural theme.

The available data on the amino acid compositions and sequences of plant lectins have largely been derived from the lectins of the *Leguminosae*, so it is perhaps not surprising that the most interesting similarities have been found in that family. Despite differences in saccharide-binding specificity and other properties, these lectins are very similar in amino acid sequence and three-dimensional structure. They probably have a common biological function and evolutionary origin as well. We shall discuss these similarities as a prelude to a comparison of the known three-dimensional structures; a separate section is devoted to the less well known lectins from nonleguminous plants.

1.2.1 Amino Acid Sequence Homology and Subclasses in Leguminous Lectins

The first lectin to be characterized in detail (EDELMAN et al. 1972) was concanavalin A (Con A), the glucose- and mannose-binding lectin from the jack bean (*Canavalia ensiformis*). Like Con A, the other known lectins from legumes have protomers of approximately 230 amino acid residues and contain binding sites

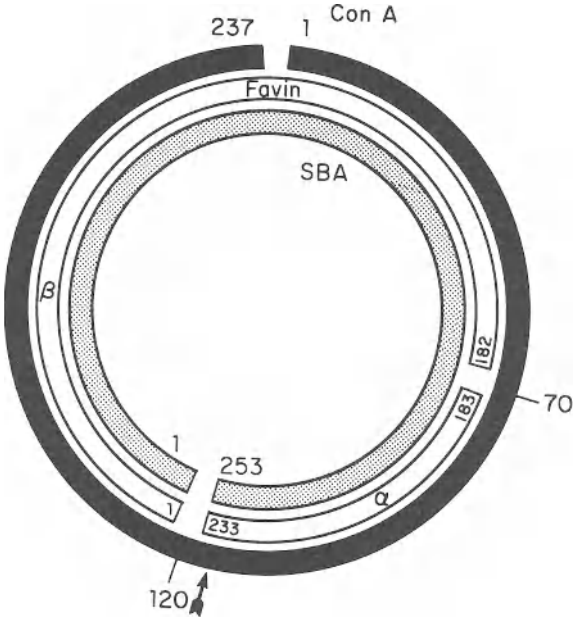


Fig. 1. Schematic drawing of favin α and β chains (□), soybean agglutinin (▨), and Con A (■) showing the circular permutation that relates Con A to the other two sequences. The site of the naturally occurring peptide cleavage (failure of ligation) in Con A is shown by the *arrow*. (Adapted from BECKER et al. 1983)

for Mn^{2+}, Ca^{2+}, and specific carbohydrate. The protomers can be divided into three structural subclasses (BECKER et al. 1983). The first includes only Con A itself (WANG et al. 1975a; CUNNINGHAM et al. 1975), which is a tetramer of four identical 237-residue chains, and the lectin from *Dioclea grandiflora* (RICHARDSON et al. 1984). The second includes the glucose- and mannose-binding lectins from fava bean (*Vicia faba*) (HOPP et al. 1982), lentil (*Lens culinaris*) (FORIERS et al. 1978; FORIERS et al. 1981), garden pea (*Pisum sativum*) (HIGGINS et al. 1983), sweet pea (*Lathyrus odoratus*) (SLETTEN et al. 1983), *Vicia sativa* (GEBAUER et al. 1981), and *Vicia cracca* (BAUMANN et al. 1982). These proteins are dimers, the subunits of which are composed of an α chain ($M_r = 5600$) and a β chain ($M_r = 20000$), which together are equivalent to a Con A monomer. The two polypeptide chains are homologous to the single Con A chain, but the amino acid sequence alignment exhibits a highly unusual circular permutation of chain segments, i.e. NH_2-β-α-COOH (Fig. 1). In the case of favin the α chains are homologous to residues 70–119 of Con A, whereas the β chain homology begins at residue 120 of Con A and continues to the carboxyl terminus of that molecule and then without interruption from the amino terminus to the point where the α chain begins. The proteins of the third subclass, which includes the lectins from soybeans, peanuts, and red kidney beans (FORIERS et al. 1977), sainfoin (*Onobrychis viciifolia* Scop.) (KOUCHALAKOS et al. 1984), and the GalNAc-binding lectins from *Vicia cracca* (BAUMANN et al. 1979) and *Dolichos biflorus* (ETZLER et al. 1977) are composed of single chains similar in size to the Con A monomer, but they too are circularly permuted, being equivalent to the two chains of the second subclass covalently linked in the order NH_2-β-α-COOH.

It has recently been shown that the three subclasses arise as a result of different posttranslational modifications of similar precursor molecules, all of which consist of a soybean-type sequence preceded by a signal sequence. In the processing of favin, the signal sequence is removed, and the remaining chain is cleaved to produce the mature α and β chains (HEMPERLY et al. 1982). The processing of Con A is quite unusual: the signal peptide is removed, and the chain is cleaved into smaller chains and is then reannealed to produce the circularly permuted sequence of the mature protein (CARRINGTON et al. 1985; BOWLES et al. 1986). Diffraction studies on Con A (BECKER et al. 1975; REEKE et al. 1975; HARDMAN et al. 1982) and favin (REEKE and BECKER 1986) as well as recent studies on pea lectin (EINSPAHR et al. 1986) indicate that the mature forms of these three lectins are extremely similar in three-dimensional structure despite their differences in subunit structure and posttranslational processing. The differences in saccharide-binding specificity between Con A and favin are consistent with the differences in three-dimensional structure that do exist between the two proteins (see below).

1.2.2 Amino Acid Sequences of Other Lectins

Amino acid sequence data on lectins from nonleguminous plants are much sparser than for leguminous plants, and as a result much less is known about the possible homology relationships of these proteins. Apparent internal gene duplications have been reported in wheat germ agglutinin (WGA) (WRIGHT et al. 1984, 1985) and ricin (VILLAFRANCA and ROBERTUS 1981). The folding pattern of WGA is similar to that of certain other cysteine-rich small proteins such as snake venom toxins (DRENTH et al. 1980), but the relationship may be one of supersecondary structure or convergent evolution rather than homology.

1.3 Crystallographic Studies of Plant Lectins

Con A was among the first proteins to be obtained in crystalline form (SUMNER 1919). It was also the first lectin whose amino acid sequence (EDELMAN et al. 1972) and three-dimensional structure (EDELMAN et al. 1972; HARDMAN and AINSWORTH 1972) were determined. The high-resolution structure of Con A was followed by that of WGA (WRIGHT 1977). Saccharide binding was difficult to study in both lectins because of disruption of crystal lattice interactions by bound saccharide. Until very recently the best results available were for WGA–saccharide complexes studied by difference Fourier methods at 2.8- to 3.0-Å resolution (WRIGHT 1980) and Con A-saccharide complexes studied in cross-linked crystals at 4.0 Å resolution (BECKER et al. 1976) and in a different crystal form at 6.0-Å resolution (HARDMAN and AINSWORTH 1976). Two new structures have now added significant information to what was known from these earlier studies. The structure of pea lectin has recently been determined at 2.8-Å resolution (EINSPAHR et al. 1986) and that of favin at the same resolu-

tion (REEKE and BECKER 1986). Both of these lectins are in the second subclass of leguminous lectins; the structures are very similar to that of Con A and show clearly the minimal nature of the changes required to accommodate the cyclic permutation in amino acid sequence between the two subclasses. In addition, favin contains bound glucose in the native crystals. Its structure is thus the first at high resolution of a leguminous lectin with bound saccharide. In the remainder of this chapter we shall present in some detail the structures of these four lectins, which are the best known, along with a summary of other lectins whose structures are in some stage of determination. In the course of these presentations we shall also compare the various structures and discuss the similarities and differences in relation to the different saccharide-binding specificities of the different molecules.

2 Concanavalin A

2.1 Saccharide-Binding Specificity

The stereochemical requirements for the interaction of saccharides with Con A in solution have been extensively studied. D-Mannose, D-glucose, and related compounds bind best (e.g., GOLDSTEIN et al. 1965b; SO and GOLDSTEIN 1967), whereas D-galactose and its derivatives are inactive as inhibitors of the biological activities of Con A (GOLDSTEIN et al. 1965b). The protein binds to glucosyl or mannosyl residues at the nonreducing termini of oligo- or polysaccharides (GOLDSTEIN et al. 1965a; SO and GOLDSTEIN 1968) and to nonterminal 2-*O*-substituted α-D-mannopyranosyl residues (HEHRE 1960; GOLDSTEIN et al. 1974). Di- and trisaccharides such as trehalose and isomaltose are excellent inhibitors (SO and GOLDSTEIN 1967), suggesting the presence of an extended binding site on the protein molecule. Certain aryl and alkyl β-glycosides of D-Glc and D-Man are more active than the unsubstituted sugars (IYER and GOLDSTEIN 1973; LOONTIENS et al. 1973), but in at least some cases these compounds bind to a hydrophobic binding site unrelated to the specific saccharide-binding site of the lectin (HARDMAN and AINSWORTH 1973; BECKER et al. 1976).

An analysis of the combined data (GOLDSTEIN et al. 1974) leads to the following conclusions. The hydroxyl at carbon 2 is not essential, but when present is more favorable in the axial (Man) than in the equatorial (Glc) position. Equatorial hydroxyls are required at carbons 3 and 4, and the hydroxyl at carbon 6 must be unsubstituted. These requirements are somewhat different than those for optimal binding to favin (ALLEN et al. 1976); as we shall see, the differences can be rationalized in terms of differences in the three-dimensional structures of the two proteins.

2.2 General Description of the Structure

The tetrameric Con A molecule crystallizes in the orthorhombic space group I222 with one protomer per asymmetric unit. The structure was solved independ-

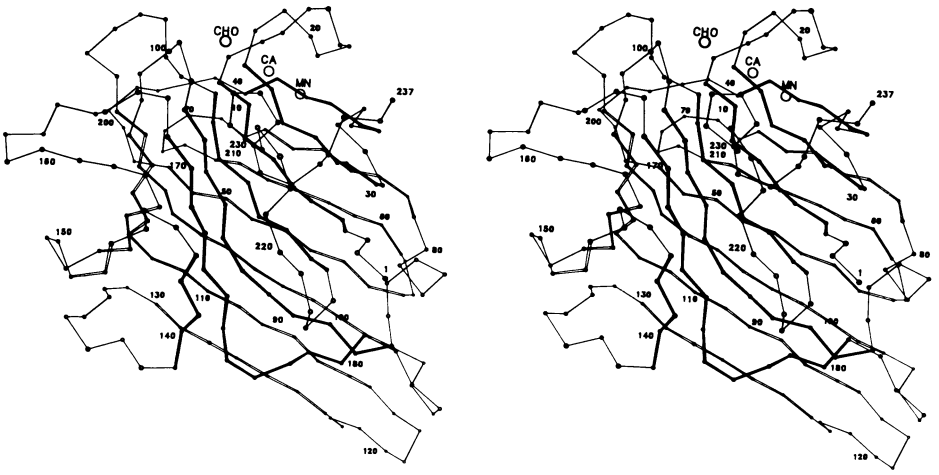


Fig. 2. Stereo α-carbon drawing of a Con A monomer with the two β structures highlighted (front β structure, *heavy black bonds*; back β structure, *open bonds*). The Mn^{2+}, Ca^{2+}, and saccharide binding positions are indicated by *MN*, *CA*, and *CHO*, respectively. The loop between the two lowest chains at the *bottom right* contains the site of the natural cleavage (failure of ligation) between residues 118 and 119

ently in two laboratories at a nominal resolution of 2.0 Å by multiple isomorphous replacement (EDELMAN et al. 1972; HARDMAN and AINSWORTH 1972; HARDMAN 1973) and was subsequently described in fuller detail (BECKER et al. 1975; REEKE et al. 1975; HARDMAN et al. 1982). Atomic parameters have been refined by a constrained difference Fourier procedure (BECKER et al. 1976) and by a least-squares method (HARDMAN et al. 1982). The protomer, containing a single polypeptide chain of 237 amino acid residues, is a compact dome-shaped molecule approximately 42 × 40 × 39 Å in size (Fig. 2). The chain folding is dominated by two extensive antiparallel β structures: a six-stranded sheet that makes up the entire back of the molecule and a seven-stranded sheet that passes through its center. Two of the six-stranded sheets interact end-to-end via main-chain hydrogen bonds to form an elongated dimer with a single 12-stranded sheet; two of these dimers in turn interact face-to-face, mainly through their side chains, to form the 222 tetramer (REEKE et al. 1975).

The native protomer contains Mn^{2+} and Ca^{2+} ions that are essential for saccharide-binding activity (SUMNER and HOWELL 1936; KALB and LEVITZKI 1968). These ions are bound 5 Å from each other in a complex site located between the central β structure and a loop of peptide that connects two strands of this structure. Each ion has four protein groups and two water molecules as ligands. One of the Ca^{2+} water ligands is hydrogen bonded to the side chain of an aspartic acid residue (Asp 208) that is joined to the preceding residue (Ala 207) by an unusual *cis* linkage. Asp 208 also makes up part of the specific carbohydrate-binding site, which lies in a depression in the molecular surface 10–15 Å from the metal ions and between the peptide loop over the metal ions and a second loop that is also attached to the central β structure.

2.3 Description of the Carbohydrate-Binding Site

Evidence on the location of the carbohydrate-binding site of Con A has been obtained independently by nuclear magnetic resonance (BREWER et al. 1973; VILLAFRANCA and VIOLA 1974; ALTER and MAGNUSON 1974) and crystallographic (BECKER et al. 1976; HARDMAN and AINSWORTH 1976) techniques. The binding of saccharides to Con A is accompanied by conformational changes in the protein (PFLUMM et al. 1971) that disrupt native crystals and preclude routine crystallographic difference Fourier analysis. This difficulty was overcome by binding the heavy-atom labeled inhibitory sugar 2-deoxy-2-iodo-methyl-α-D-mannopyranoside to Con A in crystals that had been stabilized by treatment with glutaraldehyde (BECKER et al. 1976). In the presence of this compound Con A crystals undergo a reversible solid-state structural change to give a crystalline complex whose structure was determined at 3.5-Å resolution by multiple isomorphous replacement (m.i.r.) and electron density modification methods.

The saccharide moiety in these crystals is found approximately 13 Å from the Mn^{2+} ion, in a shallow pocket near the top of the molecule (Fig. 2), where it could potentially interact with atoms from residues 14–16, 97–99, 168, 169, 207, 208, 224–228, and 235–237 of the Con A polypeptide chain. This position was confirmed by calculation of a difference electron density map between the iodinated sugar and the analogous chloro- derivative. Significant local changes in the protein conformation were observed in these complexes. The positions of chain segments 95–105 and 195–207 relative to the rest of the molecule were altered, along with the entire metal-binding region, residues 33–45, and the top two chains of the back β structure. However, the available resolution was insufficient to describe these changes with certainty, and it is likely that some of them are associated with the disruption of the crystal lattice and not with saccharide binding *per se*. Further details will have to await the analysis of another crystalline Con A–saccharide complex at higher resolution.

Crystals suitable for such an analysis have been obtained by BECKER et al. (1976) and by HARDMAN and AINSWORTH (1976). In both cases the crystals grow in space group $C222_1$ and have an asymmetric unit volume approximately six times that of native crystals. The latter authors studied a methyl-α-D-mannopyranoside complex at 6-Å resolution by m.i.r. methods. They found eight monomers per asymmetric unit in the electron density map, but only sufficient mass (by crystal density measurements) to account for four monomers, suggesting the existence of crystallographic disorder in the distribution of molecules among the three different molecular positions (two of them special positions of one-half the multiplicity of a general position) that were found. Nonetheless, the molecular images were found to be clear, and the protein was in a conformation very similar to that seen in native crystals. The bound saccharide was located by difference Fourier methods based on replacement of the methyl-α-D-mannopyranoside with two iodine-labeled analogs; it was found near the location described above (BECKER et al. 1976). Protein side chains positioned to interact with the saccharide are Tyr 12, Tyr 100, Asp 16, and Asp 208. Although the orientation of the pyranoside ring could not be unambiguously determined, it was suggested that carbon 1 is situated away from the center of the protein

monomer, based on the position of the iodine substituent in the difference maps.

More recently the interaction between Con A and a complex biantennary class glycopeptide has been modeled by computer graphics methods, based on the known structure of the native protein and on a structure for the glycopeptide derived from nuclear Overhauser effect measurements (CARVER et al. 1985). The regions of the protein most likely to be in contact with the ligand were identified as amino acid residues 12–18, 98–102, 205–208, and 226–229. Among these, several side chains appeared to be most responsible for the particular specificity of this lectin: Leu 99, which is in a position to block binding of sugars with a β substituent at oxygen 6; and Thr 15 and Tyr 12, which would block binding of sugars substituted at oxygen 4.

Confirmation of these assignments will require further crystallographic studies. A high-resolution analysis of Con A complexes with specifically bound methyl-α-D-glucopyranoside and methyl-α-D-mannopyranoside in a $C222_1$ crystal form (BECKER et al. 1976) is in progress. Data are now being collected to the full 2.8-Å resolution afforded by the diffraction patterns in order better to characterize the saccharide-binding interactions and saccharide-induced conformational changes of Con A and to provide a basis for comparison with the recently determined structure of a carbohydrate complex of favin.

2.4 Activation of Saccharide Binding by Metals

It has long been known that certain divalent metal ions are necessary for the saccharide-binding activity of Con A (SUMNER and HOWELL 1936; KALB and LEVITZKI 1968). Treatment of the native protein with acid reversibly removes the metal ions and destroys the ability to bind carbohydrates (AGRAWAL and GOLDSTEIN 1967; YARIV et al. 1968). Extensive studies in solution, particularly by nuclear magnetic resonance techniques, have revealed a complicated set of interlocking equilibria involving the apoprotein and various complexes with one or both metals and the saccharide (BREWER et al. 1983). One of the main conclusions of these studies is that the protein exists in at least two conformational forms, an "unlocked" form found mainly in the apoprotein that binds metals and saccharides only weakly and a "locked" form possessing full binding activities. Magnetic resonance data (BROWN et al. 1977) further suggest that readdition of metal ions to the apoprotein causes significant conformational changes, including a process with a rate constant appropriate for a *trans* to *cis* isomerization of a peptide bond. Comparison of the three-dimensional structures of the two forms (REEKE et al. 1978) shows the nature of these changes in some detail.

Crystals of demetallized Con A belong to the orthorhombic space group $P22_12_1$ and contain a dimer in the asymmetric unit (JACK et al. 1971). They diffract well to a resolution of 2.8 Å at room temperature. The structure was solved by molecular replacement using native Con A as the model (REEKE et al. 1978) and was refined by constrained difference Fourier methods to a crystallographic R factor of 0.33. The structures of the native and demetallized proteins are extremely similar with the exceptions of the loop of peptide that encloses

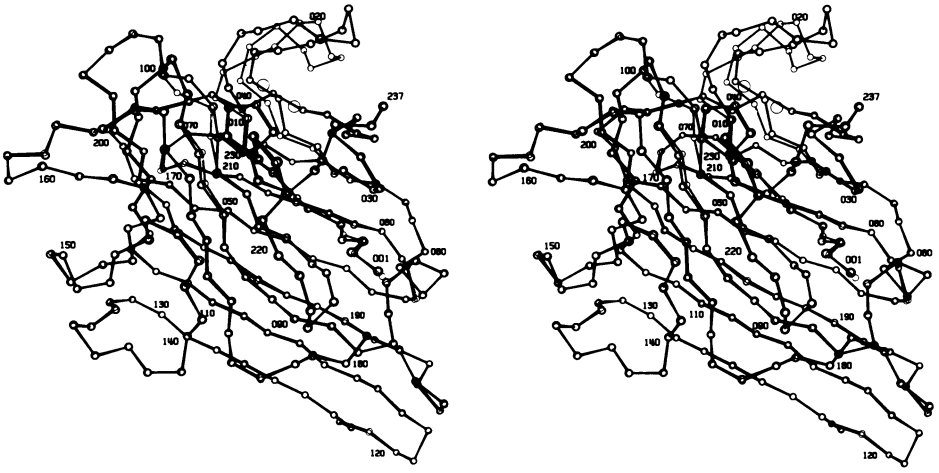


Fig. 3. Superposition of the entire α-carbon backbone of native Con A (*open bonds*) and one of the two crystallographically independent subunits of demetallized Con A (*heavy black bonds*). Orientation as in Fig. 2. The largest differences between the two structures are in the region of metal ion binding at the top of the figure. (Reproduced from REEKE et al. 1978, with permission)

the metal ions, which is generally less ordered and more accessible to solvent in demetallized Con A, and the *cis* aspartic acid, Asp 208, which is in the normal *trans* configuration in the apoprotein (Fig. 3). Because many of the residues whose conformational changes on demetallization also form part of the carbohydrate binding site, this model of demetallized Con A provides a straightforward explanation for the inability of the apoprotein to bind carbohydrate. The changes in the metal-binding loop as well as the *trans* to *cis* peptide isomerization are in full agreement with the magnetic resonance data discussed above, which have also suggested such rearrangements.

A subsequent independent crystallographic analysis of demetallized Con A (SHOHAM et al. 1979) has produced somewhat different results. That study, which was also based on molecular replacement, found less extensive conformational changes, mostly involving the side chains of Asp 19 and His 24, and found no evidence for a *cis/trans* isomerization at Asp 208. The differences in interpretation between the two studies significantly affect our overall understanding of the role of the metals in the activity of the native protein, but they cannot be definitively resolved with the present data. It is particularly important that the question of the *cis/trans* isomerization be clarified by further refinements of both structures of the demetallized protein.

3 Favin

3.1 Saccharide-Binding Specificity

As with Con A, agglutination of erythrocytes by fava bean extracts (LEE et al. 1974) is inhibited by glucose- and mannose-containing sugars, but not by galac-

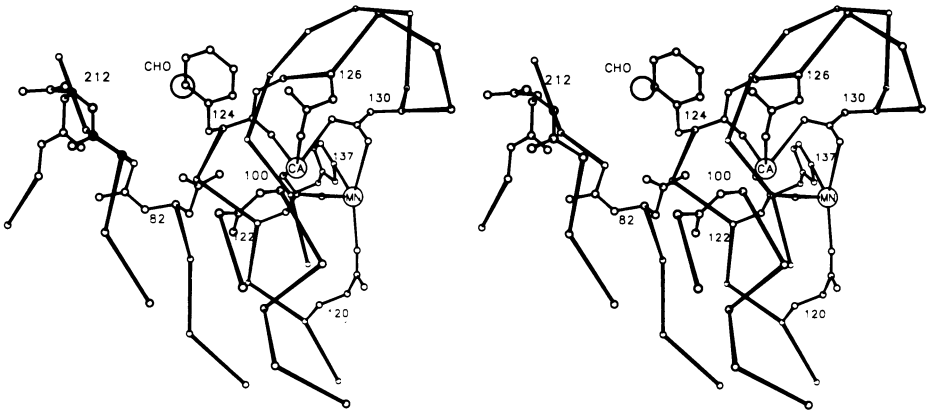


Fig. 4. Close-up stereo view of the region of metal ion and saccharide binding in favin. The Mn^{2+} and Ca^{2+} are at the *center right*, the *cis*-linked Ala 81–Asp 82 peptide bond is *just left of center*, and the bound glucose is at the *top*. Residues implicated in specific interactions with the bound saccharide are listed in the text

tose or arabinose and their derivatives. More detailed studies (ALLEN et al. 1976) indicate that the ability to tolerate glycosidic substitution, particularly in the α anomer, as well as the requirement for unsubstituted hydroxyls at carbons 4 and 6, is identical in Con A and favin, as well as in pea and lentil lectins. However, the latter three lectins differ from Con A in being more strongly inhibited by saccharides with small substituents (e.g., methyl) on oxygen 3 than by the unsubstituted sugars. The recently determined three-dimensional structure of favin (Fig. 4) suggests a plausible explanation for this behavior, as will be shown in Sect. 3.3.

3.2 General Description and Comparison with Concanavalin A

The complex of favin with glucose crystallizes in the orthorhombic space group $P2_12_12_1$ with a complete dimeric molecule in the asymmetric unit (WANG et al. 1974). The structure (REEKE and BECKER 1986) was determined at 2.8-Å resolution using molecular replacement based on pea lectin (EINSPAHR et al. 1986) as a model. After a preliminary restrained refinement the crystallographic R factor is 0.38. In places where the favin and pea lectin sequences differ the electron density map invariably agrees with the favin sequence, indicating a lack of bias from the starting model in the phasing.

The structure of favin is strikingly similar to that of Con A (Sect. 2.2) and that of pea lectin (Sect. 4.2). The promoter has the same domelike shape and the folding of the chains, including the topological arrangement of the strands in the two antiparallel β structures, is identical in the three molecules. In Con A the dimer is formed by the interaction of two six-stranded sheets across a perfect (crystallographic) twofold axis; in favin the same interaction occurs across a local (noncrystallographic) twofold axis, but the two protomers are nonetheless

very similar, the rms distance between corresponding α carbons being only 0.48 Å when the two protomers are aligned by rotation about the local twofold axis. The separate α and β chains of favin are interwoven through the two pleated sheets in such a way that the folding pattern of neither chain could apparently be stable in the absence of the other. The two chains together, like the single chain of Con A, thus form a single folding domain.

Favin and Con A also have very similar metal ion-binding sites (see Fig. 4). In both proteins the Mn^{2+} and Ca^{2+} are bound in a double site with two aspartate ligands shared by both metals. The protein ligands in both cases are a glutamate, a histidine, and two aspartates to the Mn^{2+}, and an asparagine, two aspartates, and a backbone carbonyl to the Ca^{2+}. In both cases the metal coordination is completed by two water molecules that are hydrogen bonded to the protein. One of these waters is hydrogen bonded in both lectins to the side chain of the *cis*-linked aspartate (Asp 82 in favin, Asp 208 in Con A) that links the metals to the saccharide-binding site.

3.3 Description of the Saccharide-Binding Site

The noncovalently bound glucose is seen as a peak of electron density in both of the independent subunits of favin. The glucose is bound in a cleft in the protein surface formed by the loop of polypeptide chain that encloses the metals, another loop that includes residues 99 and 100, and side chains extending from the front β structure. This binding site is at an analogous location to the one in Con A. Polar groups that are in position to interact with the carbohydrate include the main chain nitrogens of Gly 100, Ala 212, and Thr 213, the side chains of Asn 40 and Asn 126, and the side chain of the *cis*-linked Asp 82. As with Con A, it is thought that stabilization of this *cis*-peptide linkage by Ca^{2+} is a major element in the induction of saccharide-binding activity by the metal ions.

The major structural difference between Con A and favin in this region is the replacement of two large side chains, Leu 99 and Arg 228, in Con A with smaller groups, Ala 212 and Gly 100, in favin. These replacements render the favin site considerably more open than that of Con A and probably account for the fact that substitution of methyl or phenyl groups at oxygen 3 enhances carbohydrate binding in favin but decreases it in Con A. Near to this location are two hydrophobic side chains, Tyr 101 and Trp 129, and nonpolar interactions between these groups and hydrophobic aglycones at oxygen 3 would be consistent with the ability of such groups to enhance binding. Similarly, hydrophobic substituents at oxygen 1 may interact with Phe 126 (homologous to Tyr 12 in Con A), also enhancing the lectin–carbohydrate interaction.

3.4 Structural Consequences of Posttranslational Processing

The differences between Con A and favin resulting from their different pathways of posttranslational modification are accommodated in the three-dimensional

structures with minimal disruptions. The locations of these differences and of the amino acid sequence differences are nicely consistent with the similar binding activities but different quaternary structures of the two proteins. The largest difference in chain length is a loop of seven residues near position 160 in Con A that is not present in the corresponding region of favin, near residue 36. This loop is well away from any sites of metal or saccharide binding or quaternary structure interactions in the two proteins. The remaining differences in folding are accounted for by the different chain structures of the two proteins: a loop of polypeptide chain containing favin residues 108–115 that corresponds to the gap between the COOH and NH_2 termini of Con A; a reverse turn near residue 120 of Con A that is aligned with the NH_2 terminus of the favin β chain and the COOH terminus of the α chain; and a cleaved bond between favin residues 182 and 183 that replaces a β bend at residues 67–70 of Con A. A further significant difference is that most of the amino acid side chains (35 of 44 residues) projecting from the rear of the back β structure are different in the two lectins. It is through these groups that pairs of Con A dimers interact to form tetramers, an interaction that does not occur in favin. None of the residues involved in electrostatic bonds between the two Con A dimers is charged in favin, leaving favin without a possible major source of stabilization for tetramer formation. In addition, the covalently attached saccharide of favin is contained in this region, possibly hindering tetramer formation by its bulk.

The overall similarity in the three-dimensional folding of favin and Con A, together with the fact that all the sites of posttranslational modification in both proteins are in loops on the molecular surface, has led to the proposal that the favin and Con A precursors are synthesized with three-dimensional structures similar to each other and to the mature proteins, and that posttranslational modification takes place without gross modification of the structures (Bowles et al. 1986; Reeke and Becker 1986). The availability of all the sites of posttranslational modification to enzymatic attack at the molecular surface is consistent with this view. Both molecules are apparently synthesized as precursors similar in three-dimensional structure to favin, but with a signal sequence at the NH_2 terminus and with residues 182 and 183 joined by a peptide bond in a conformation similar to that existing at residues 69 and 70 of Con A.

4 Pea Lectin

The glucose- and mannose-binding lectin from pea is the only other legume lectin for which a high-resolution structure is now available (Einspahr et al. 1986), and this protein promises to become one of the most accurately known lectin structures, since native diffraction data extend to a resolution limit of 1.2 Å. Since the crystals studied do not contain specifically bound saccharide, we will describe the native structure and compare it with Con A and favin in order to see what deductions can be made about its saccharide binding. However, preliminary crystallographic data on a saccharide complex of pea lectin have recently been reported (Rini et al. 1986), and it can be anticipated that direct observation of pea lectin–carbohydrate binding will soon be possible.

4.1 General Description and Comparison with Con A and Favin

Pea lectin is similar to both Con A and favin but is more closely related to favin (BECKER et al. 1983). The molecule exists in solution and in the crystals as dimers of structure $(\alpha\beta)_2$, with a total of 226 residues in the two chains of each protomer. At 3.0-Å resolution the overall shape of the molecule, the secondary and tertiary structures, and the topology of the β structure strands are all the same as in Con A and favin. Pea lectin also has binding sites for Mn^{2+} and Ca^{2+} ions with protein ligands for both metals that are similar to those of the other two lectins. The peptide bond preceding Asp 81 is in the *cis* configuration, corresponding to Asp 82 in favin and Asp 208 in Con A. The carbonyl and carboxylate oxygens of this residue are at the appropriate distances from the Ca^{2+} for hydrogen bonding to a bridging water molecule, which, however, cannot be seen in the current electron density map.

As in Con A and favin, the sites of posttranslational modification are all at the surface of the molecule, where the necessary enzymatic reactions can proceed with no significant disruption of the structure of the presumed precursor molecule.

As with favin, the quaternary structure of pea lectin is dimeric rather than tetrameric. The twist of the back pleated-sheet structure is different in pea lectin than in Con A, which might be responsible for the lack of tetramer formation in pea lectin. In addition, the potential contact residues between the sheets are different, only 12 of the 54 residues of Con A that are involved in these interactions being conserved in pea lectin, and none of the interacting pairs.

Assuming that the saccharide-binding site of pea lectin will be found at a position corresponding to those in Con A and favin, only three of the residues postulated (CARVER et al. 1985) to be in direct contact with the bound saccharide in Con A are the same in pea lectin. Interpretation of this observation will require further crystallographic data on the binding of saccharides to pea lectin (RINI et al. 1986), as well as a more detailed characterization of the stereochemical requirements for saccharide binding than is now available.

5 Wheat Germ Agglutinin

WGA is the first lectin from the *Graminae* family to be characterized in molecular detail. At present it also remains the only nonlegume and only nonglucose- or nonmannose-binding lectin whose three-dimensional structure is known at high resolution. Like the other *Graminae* lectins (GOLDSTEIN and HAYES 1978), it has an unusual and highly stable structure distinguished by high cystine and glycine content and extensive internal sequence homology.

5.1 Saccharide-Binding Specificity

Specificity of WGA for membrane receptors containing the two acetylated saccharides *N*-acetyl-D-glucosamine (GlcNAc) and *N*-acetyl-D-neuraminic acid

(NeuNAc) has been inferred from studies of hapten inhibition of agglutination (LIS and SHARON 1977; GOLDSTEIN and HAYES 1978) and from enzymatic removal of terminal saccharides in intact cell systems (e.g., CUATRECASAS 1973b; GANGULY and FOSSETT 1979; BHAVANANDAN and KATLIC 1979). Terminal nonreducing NeuNAc has been implicated in these studies as the principal target for WGA binding to various types of cells. Well characterized receptor molecules include chicken egg-white ovomucoid (MONTREUIL 1975), chitin (MIRELMAN et al. 1975), and glycophorin A (ANDERSON and LOVRIEN 1981). Studies with simple oligosaccharides suggest that each WGA subunit possesses two independent and noncooperative binding sites for GlcNAc oligomers; the crystallographic results show that these sites are in fact located at the interfaces between domains of two monomers in the dimeric WGA molecule.

5.2 General Description of the Structure

WGA occurs as several genetically distinct isolectins (ALLEN et al. 1973). The two major species, WGA1 and WGA2, cocrystallize, and the structure of a space group C2 modification was determined at 2.2-Å resolution (WRIGHT 1977). Subsequently, separate data have been obtained for the two isolectins, and the structure of WGA2 has been refined at 1.8-Å resolution (WRIGHT 1987; WRIGHT and OLAFSDOTTIR 1986). The molecules are dimers approximately $40 \times 40 \times 70$ Å in size. The protomers have 171 amino acid residues, folded into four domains (A, B, C, and D) of 43, 43, 43, and 42 residues, respectively. The domains have extensive internal homologies, concentrated in five core clusters, which in all include 22 residues in each domain. These homologies are highly suggestive of gene duplication events leading to the present four-domain structure (WRIGHT et al. 1985; WRIGHT 1987).

The domain fold is highly ordered, with four interlocking disulfide bonds and numerous hydrogen bonds stabilizing each domain (Fig. 5). However, there is virtually no regular secondary structure of an identified type. Successive domains along the monomer polypeptide chain are related by approximate non-

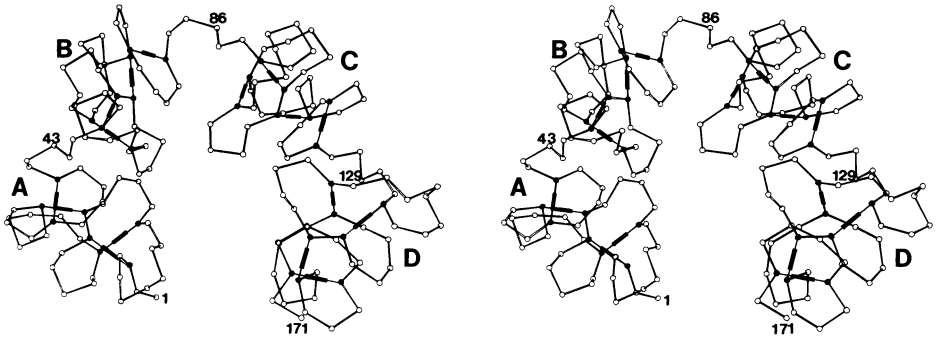


Fig. 5. Folding of the polypeptide chain in WGA. Structurally similar domains are labeled *A, B, C,* and *D*; disulfide bonds are indicated by *heavy bonds* and *darkened A-carbon positions.* (C.S. Wright, private communication, reproduced with permission)

integral screw axes, leading to an unusually elongated structure with very few interactions between domains. Neighboring domains in adjacent chains are more closely related by quasi-twofold axes to form the interlocked dimers. The saccharide-binding sites are located at these domain interfaces, to form in all four principal and four secondary binding sites per dimer.

The two isolectins studied differ at only four sites, three closely spaced in the B domain and the fourth in the C domain. All of these are accommodated with little or no disturbance in the monomer and dimer structures. One of them (Tyr 66 in WGA1 → His 66 in WGA2) is at a position that is functional in saccharide binding, and may be related to the tighter complex formation found in WGA1 for neuraminyl-lactosamine (KRONIS and CARVER 1982).

5.3 Description of the Saccharide-Binding Site

The nature of the saccharide-binding sites in WGA has been elucidated by difference Fourier analysis of complexes of the lectin with di-*N*-acetylglucosamine (2.8-Å data), 6-iodo-1,4-dimethyl-*N*-acetylglucosamine (3-Å data) and *N*-acetylneuraminic acid lactose (4-Å data) (WRIGHT 1980). There are four sites per dimer. One site lies in each area of contact between domains from opposite monomers, and side chains from both monomers contribute to each site. Thus, an isolated protomer would be unable to bind sugars. The sites have been classified as "primary" (at the B–C interfaces) and "secondary" (at the A–D interfaces) according to their accessibility in the crystal. Binding at the secondary sites appears to interfere with crystal lattice interactions, and sugars that bind there, such as *N*-acetylglucosamine, tend to disintegrate native crystals.

Figure 6 illustrates the composition and relative orientations of the two types of binding sites (P, primary site; S, secondary site). This figure shows all the amino acids in the vicinity of each binding site, the approximate positions of two bound $(GlcNAc)_2$ disaccharides (dashed ovals) and the pseudo-twofold axes relating domains of opposite protomers. The three amino acid residues most crucial to sugar binding are homologous in the two binding sites: Ser 62, Tyr 73, and Glu 115 in the P site to Ser 148, Tyr 159, and Asp 29 in the S site. In addition, 11 of the 14 aromatic amino acids lie within 15 Å of these binding sites, and the surface spanning the distance between them is lined with hydrophilic side chains.

The existence of extended binding sites with subsites for up to three saccharide monomers has been inferred from inhibition data (ALLEN et al. 1973). There is some crystallographic corroboration for the second subsites (Fig. 6) (WRIGHT 1984), but the strongest interactions are clearly with the main subsite at which monosaccharides bind in both the P and S sites. The characteristic interactions seen in both sites (P site shown in Fig. 7) are: (a) hydrogen-bonded contacts between the *N*-acetyl carbonyl group with the sidechain hydroxyl of Ser 62 in the P site (Ser 148 in the S site), the *N*-acetyl amino group and $O^{\varepsilon 2}$ of Glu 115 (Asp 29), and the ring hydroxyl oxygen of C3 with the hydroxyl of Tyr 73 (Tyr 159); (b) van der Waals contacts involving Tyr 73 (Tyr 159) with the *N*-acetyl methyl group, His 66 (Ser 152) with several atoms in the pyranoside ring,

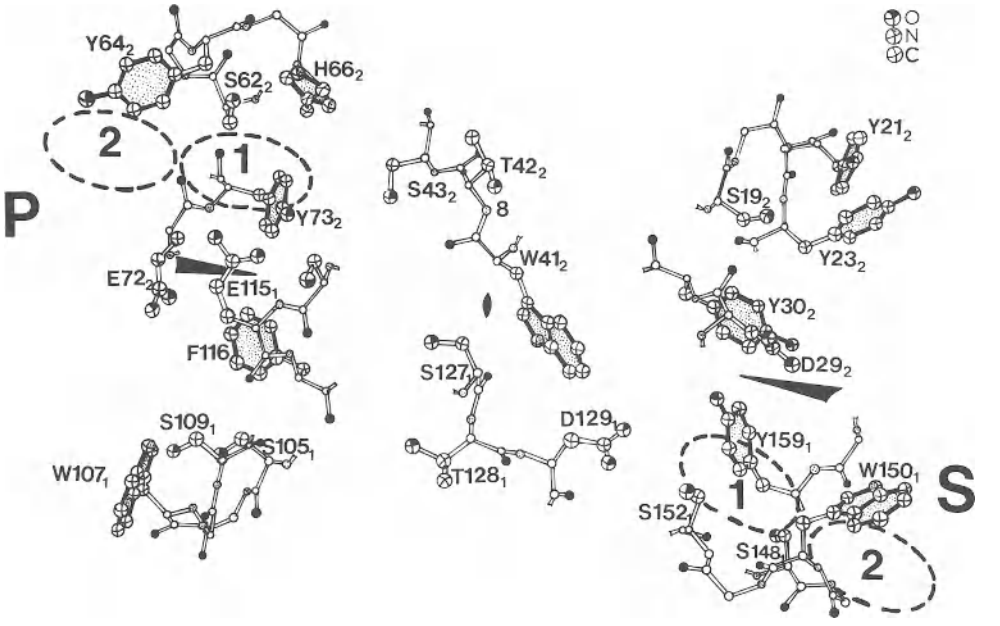


Fig. 6. Schematic view showing the relative positions and orientations of the two distinct saccharide-binding sites in WGA. The three pseudo-twofold axes relating domains of the two opposite protomers are indicated by the standard crystallographic symbols. *P*, primary site; *S*, secondary site; subscripts *1* and *2* refer to protomers 1 and 2; *oval areas* marked *1* and *2* represent approximate subsite positions. (C.S. Wright, private communication, reproduced with permission)

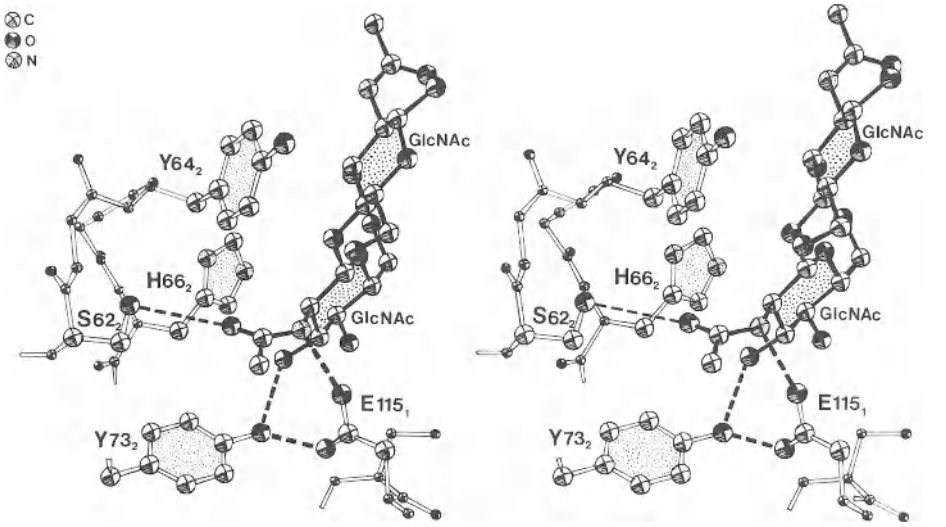


Fig. 7. Primary saccharide-binding site of WGA. $(GlcNAc)_2$, *filled bonds*; WGA, *open bonds*. *Broken lines* represent hydrogen bonds and *subscripts* denote protomers 1 and 2. (C.S. Wright, private communication, reproduced with permission)

and Tyr 64 (Trp 150, earlier thought to be a Gly, as its side chain is completely disordered in the crystal structure) with the second pyranoside ring; and (c) displacement of three ordered water molecules from the binding site by the bound saccharide. There are no significant differences in these specifics between the two isolectins WGA1 and WGA2 (WRIGHT 1984).

The data are consistent with the following set of requirements for specific saccharide binding to WGA: (a) an unsubstituted acetamido group is required at carbon 2: (b) equatorial hydroxyls are required at carbons 3 and 4; and (c) substitution of the hydroxyl groups at carbons 1 (α or β), 4, and 6 is allowed. The oligosaccharide-binding mode, unlike that in lysozyme, is stabilized mainly through one subsite, with minor contributions coming from an adjacent subsite and no interactions from a third saccharide ring.

6 Other Studies of Lectin Structures in Progress

The lectins already discussed are the only ones for which high-resolution crystal structures are now available. In addition to these, crystallization or preliminary diffraction studies have been reported for lentil (LEBRUN et al. 1983), peanut (SALUNKE et al. 1985), and soybean (SHAANAN et al. 1984) lectins, and for ricin (VILLAFRANCA and ROBERTUS 1981) and abrin (WEI and EINSTEIN 1974). These and other lectin structures should provide additional insight into the relationships among lectins of different classes and the structural bases for the different saccharide-binding specificities of lectins.

7 Summary and Conclusions

The structural studies summarized in this chapter provide the beginnings of a general understanding of protein–saccharide interactions. At this stage there are probably too many differences and as yet too few examples to draw any general conclusions. Nature seems able to construct saccharide-binding sites on very different frameworks, as exemplified by the legume lectins with their large β structures and low cysteine content, on the one hand, and wheat germ agglutinin with its high cysteine content and virtual absence of regular secondary structures on the other. Chemical groups involved directly in binding are equally diverse, ranging from backbone amino groups to side-chain carboxylates. One provocative finding arising from comparison of the saccharide-binding sites of the legume lectins is that, because main-chain atoms form many of the specific polar interactions between the lectin and its carbohydrate ligands, general binding specificity can be conserved despite considerable variation in several of the amino acids that make up the binding site.

There are many outstanding questions waiting to be answered. We do not yet know how many classes of lectins are likely to exist, or to what extent they follow the taxonomic classes of the plants from which they are derived.

There are many suggestions that lectins interact with multiple sugar moieties in oligosaccharide substrates, but very little is known of the specific mechanisms of such binding, in particular the influence of neighboring subsites on the specificity of the primary binding sites. The role of hydrophobic binding sites in lectin function is another largely unexplored area. In the case of the lectins from legumes there are some more specific questions: do all share the same *trans* to *cis* isomerization upon the activation of saccharide-binding activity by the native metal ions? What is the orientation of the saccharide in these lectins, and is it the same for all of them? Does the circular permutation of amino acid sequence have any functional significance? Knowledge of the normal function of lectins in the species from which they are derived would undoubtedly be very helpful in putting together the clues that are at hand and coming to answer these questions, besides solving a longstanding puzzle that has intrigued all those who have joined in the study of these most fascinating molecules.

Acknowledgements. We would like to thank Dr. F.L. Suddath for making available to us the coordinates of pea lectin and a description of the structure prior to its publication. We would also like to thank Dr. C.S. Wright for providing Figs. 5, 6, and 7 and other information on wheat germ agglutinin prior to publication.

Note added in proof: Crystallographic studies on plant lectins which appeared too late for discussion in this chapter include the structure of the pea lectin – trimannoside complex (J.M. RINI, Ph.D. Dissertation, University of Toronto, 1987) and that of the ricin B subunit (E. RUTENBAR, M. READY, and J.D. ROBERTUS, Nature *326*:624–626, 1987). Several other lectins have been crystallized, and three-dimensional structure determinations are in progress.

References

Agrawal BBL, Goldstein IJ (1968) Protein–carbohydrate interaction. XV. The role of bivalent metal ions in concanavalin A–polysaccharide interaction. Can J Biochem 46:1147–1150

Albersheim P, Anderson AJ (1971) Proteins from plant cell walls inhibit polygalacturonases secreted by plant pathogens. Proc Natl Acad Sci USA 68:1815–1819

Albertini DF, Clark JI (1976) Membrane–microtubule interactions: concanavalin A capping induced redistribution of cytoplasmic microtubules and colchicine binding proteins. Proc Natl Acad Sci USA 72:4976–4980

Allen AK, Neuberger A, Sharon N (1973) The purification, composition, and specificity of wheat-germ agglutinin. Biochem J 131:155–162

Allen AK, Desai NN, Neuberger A (1976) The purification of the glycoprotein lectin from the broad bean (*Vicia faba*) and a comparison of its properties with lectins of similar specificity. Biochem J 155:127–135

Alter GM, Magnuson JA (1974) Characterization of concanavalin A sugar binding site by ^{19}F nuclear magnetic resonance. Biochemistry 13:4038–4045

Anderson RA, Lovrien RE (1981) Erythrocyte membrane sidedness in lectin control of the Ca^{2+}-A23187-mediated diskocyte⇌echinocyte conversion. Nature 292:158–161

Andersson J, Edelman GM, Möller G, Sjöberg O (1972) Activation of B lymphocytes by locally concentrated concanavalin A. Eur J Immunol 2:233–235

Ash JF, Singer SJ (1976) Concanavalin A-induced transmembrane linkage of concanavalin A surface receptors to intracellular myosin-containing filaments. Proc Natl Acad Sci USA 73:4575–4579

Aubin JE, Carlsen SA, Ling V (1975) Colchicine permeation is required for inhibition of concanavalin A capping in Chinese hamster ovary cells. Proc Natl Acad Sci USA 72:4516–4520

Ballmer K, Burger MM (1980) Modulation of EGF binding and action by succinylated concanavalin A in fibroblast cell cultures. J Supramol Struct 14:209–214

Barnett RE, Scott RE, Furcht LT, Kersey JH (1974) Evidence that mitogenic lectins induce changes in lymphocyte membrane fluidity. Nature 249:465–466

Barondes S, Haywood P (1979) Comparison of developmentally regulated lectins from three species of cellular slime molds. Biochim Biophys Acta 550:297–308

Baumann C, Rüdiger H, Strosberg AD (1979) A comparison of the two lectins from *Vicia cracca*. FEBS Lett 102:216–218

Baumann CM, Strosberg AD, Rüdiger H (1982) Purification and characterization of a mannose/glucose-specific lectin from *Vicia cracca*. Eur J Biochem 122:105–110

Becker JW, Reeke GN, Wang JL, Cunningham BA, Edelman GM (1975) The covalent and three-dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides. J Biol Chem 250:1513–1524

Becker JW, Reeke GN, Cunningham BA, Edelman GM (1976) New evidence on the location of the saccharide binding site of concanavalin A. Nature 259:406–409

Becker JW, Cunningham BA, Hemperly JJ (1983) Structural subclasses of lectins from leguminous plants. In: Etzler ME, Goldstein IJ (eds) Chemical taxonomy, molecular biology and function of plant lectins. Liss, New York, pp 31–45

Beckert WH, Sharkey MM (1970) Mitogenic activity of the jack bean (*Canavalia ensiformis*) with rabbit peripheral blood lymphocytes. Int Arch Allergy Appl Immunol 39:337–341

Bhavanandan VP, Katlic AW (1979) The interaction of wheat germ agglutinin with sialoglycoproteins. The role of sialic acid. J Biol Chem 254:4000–4008

Bird G (1959) Hemagglutinins in seeds. Br Med Bull 15:165–168

Bittiger H, Schnebli HP (eds) (1976) Concanavalin A as a tool. Wiley, London

Bowles DJ, Marcus SE, Pappin DJC, Findlay JBC, Eliopoulos E, Maycox PR, Burgess J (1986) Posttranslational processing of concanavalin A precursors in jackbean cotyledons. J Cell Biol 102:1284–1297

Boyd WC (1954) The proteins of the immune system. In: Neurath H, Bailey K (eds) The proteins, vol 2, part B. Academic, New York

Boyd WC (1963) The lectins: their present status. Vox Sang 8:1–32

Brewer CF, Sternlicht H, Marcus DM, Grollman AP (1973) Binding of ^{13}C-enriched α-methyl-D-glucopyranoside to concanavalin A as studied by carbon magnetic resonance. Proc Natl Acad Sci USA 70:1007–1011

Brewer CF, Brown RD, Koenig SH (1983) Metal ion binding and conformational transitions in concanavalin A: a structure-function study. J Biomol Struct Dynam 1:961–997

Brown J, Hunt R (1978) Lectins. Int Rev Cytol 52:277–349

Brown RD, Brewer CF, Koenig SH (1977) Conformation states of concanavalin A: kinetics of transitions induced by interaction with Mn^{2+} and Ca^{2+} ions. Biochemistry 16:3883–3896

Carrington DM, Auffret A, Hanke DE (1985) Polypeptide ligation occurs during post-translational modification of concanavalin A. Nature 313:64–67

Carver JP, Mackenzie AE, Hardman KD (1985) Molecular model for the complex between concanavalin A and a biantennary-complex class glycopeptide. Biopolymers 24:49–63

Cuatrecasas P (1973a) Interaction of concanavalin A and wheat germ agglutinin with the insulin receptor of fat cells and liver. J Biol Chem 248:3528–3534

Cuatrecasas P (1973b) Interaction of wheat germ agglutinin and concanavalin A with isolated fat cells. Biochemistry 12:1312–1323

Cuatrecasas P, Tell GPE (1973) Insulin-like activity of concanavalin A and wheat germ agglutinin – direct interactions with insulin receptors. Proc Natl Acad Sci USA 70:485–489

Cunningham BA, Wang JL, Waxdal MJ, Edelman GM (1975) The covalent and three-dimensional structure of concanavalin A. II. Amino acid sequence of cyanogen bromide fragment F_3. J Biol Chem 250:1503–1512

dePetris S, Raff MC, Mallucci L (1973) Ligand-induced redistribution of concanavalin A receptors on normal, trypsinized, and transformed fibroblasts. Nature [New Biol] 244:276–278

Drenth J, Low BW, Richardson JS, Wright CS (1980) The toxin-agglutinin fold. A new group of small protein structures organized around a four-disulfide core. J Biol Chem 255:2652–2655

Edelman GM (1976) Surface modulation in cell recognition and cell growth. Science 192:218–226

Edelman GM, Cunningham BA, Reeke GN, Becker JW, Waxdal MJ, Wang JL (1972) The covalent and three-dimensional structure of concanavalin A. Proc Natl Acad Sci USA 69:2580–2584

Edelman GM, Yahara I, Wang JL (1973) Receptor mobility and receptor-cytoplasmic interactions in lymphocytes. Proc Natl Acad Sci USA 70:1442–1446

Einspahr H, Parks EH, Suguna K, Subramanian E, Suddath FL (1986) The crystal structure of pea lectin at 3.0 Å resolution. J Biol Chem 261:16518–16527

Ensgraber A (1958) Die Phytohämagglutinine und ihre Funktion in der Pflanze als Kohlenhydrat-Transportsubstanzen. Ber Dtsch Bot Ges 71:349–361

Etzler ME, Talbot CF, Ziaya PR (1977) NH_2-terminal sequences of the subunits of *Dolichos biflorus* lectin. FEBS Lett 82:39–41

Feldman EL, Heacock AM, Agranoff BW (1982) Lectin binding to neurites of goldfish retinal explants. Brain Res 248:347–354

Foriers A, Wuilmart C, Sharon N, Strosberg AD (1977) Extensive sequence homologies among lectins from leguminous plants. Biochem Biophys Res Commun 75:980–986

Foriers A, de Neve R, Kanarek L, Strosberg AD (1978) Common ancestor for concanavalin A and lentil lectin? Proc Natl Acad Sci USA 75:1136–1139

Foriers A, Lebrun E, Van Rapenbusch R, de Neve R, Strosberg AD (1981) The structure of the lentil (*Lens culinaris*) lectin. Amino acid sequence determination and prediction of the secondary structure. J Biol Chem 256:5550–5560

Ganguly P, Fossett NG (1979) Role of surface sialic acid in the interaction of wheat germ agglutinin with human platelets. Biochem Biophys Res Commun 89:1154–1160

Gebauer G, Schiltz E, Rüdiger H (1981) The amino-acid sequence of the α subunit of the mitogenic lectin from *Vicia sativa*. Eur J Biochem 113:319–325

Glew RH, Kayman SC, Kuhlenschmidt MS (1973) Studies on the binding of concanavalin A to rat liver mitochondria. J Biol Chem 248:3137–3145

Goldstein I, Hayes J (1978) The lectins: carbohydrate-binding proteins of plants and animals. Adv Carbohydr Chem Biochem 35:127–340

Goldstein IJ, Hollerman CE, Merrick JM (1965a) Protein–carbohydrate interaction. I. The interaction of polysaccharides with concanavalin A. Biochim Biophys Acta 97:68–76

Goldstein IJ, Hollerman CE, Smith EE (1965b) Protein–carbohydrate interaction. II. Inhibition studies on the interaction of concanavalin A with polysaccharides. Biochemistry 4:876–883

Goldstein IJ, Reichert CM, Misaki A (1974) Interaction of concanavalin A with model substrates. Ann NY Acad Sci 234:283–295

Gunther GR, Wang JL, Yahara I, Cunningham BA, Edelman GM (1973) Concanavalin A derivatives with altered biological activities. Proc Natl Acad Sci USA 70:1012–1016

Hadden JW, Hadden EM, Haddox MK, Goldberg ND (1972) Guanosine 3′:5′-cyclic monophosphate: a possible intracellular mediator of mitogenic influences in lymphocytes. Proc Natl Acad Sci USA 69:3024–3027

Hamblin J, Kent SP (1973) Possible role of phytohemagglutinin in *Phaseolus vulgaris* L. Nature [New Biol] 245:28–30

Hardman KD (1973) Crystallography of a metal-containing protein, concanavalin A. Adv Exp Med Biol 40:103–123

Hardman KD, Ainsworth CF (1972) Structure of concanavalin A at 2.4 Å resolution. Biochemistry 11:4910–4919

Hardman KD, Ainsworth CF (1973) Binding of nonpolar molecules by crystalline concanavalin A. Biochemistry 12:4442–4448

Hardman KD, Ainsworth CF (1976) Structure of the concanavalin A-methyl-α-D-mannopyranoside complex at 6-Å resolution. Biochemistry 15:1120–1128

Hardman KD, Agarwal RC, Freiser MJ (1982) Manganese and calcium binding sites of concanavalin A. J Mol Biol 157:69–86

Hehre EJ (1960) Contribution of classical immunology to the development of knowledge of dextran structures. Bull Soc Chim Biol 42:1581–1590

Hemperly JJ, Mostov KE, Cunningham BA (1982) *In vitro* translation and processing of a precursor form of favin, a lectin from *Vicia faba*. J Biol Chem 257:7903–7909

Higgins TJV, Chandler PM, Zurawski G, Button SC, Spencer D (1983) The biosynthesis and primary structure of pea seed lectin. J Biol Chem 258:9544–9549

Hopp TP, Hemperly JJ, Cunningham BA (1982) Amino acid sequence and variant forms of favin, a lectin from *Vicia faba*. J Biol Chem 257:4473–4483

Howard GA, Schnebli HP (1977) Eukaryotic ribosomes possess a binding site for concanavalin A. Proc Natl Acad Sci USA 74:818–821

Iyer RN, Goldstein IJ (1973) Quantitative studies on the interaction of concanavalin A, the carbohydrate binding protein of the jack bean, with model carbohydrate-protein conjugates. Immunochemistry 10:313–322

Jack A, Weinzierl J, Kalb AJ (1971) An x-ray crystallographic study of demetallized concanavalin A. J Mol Biol 58:389–395

Janzen DH, Juster HB, Liener IE (1976) Insecticidal action of the phytohemagglutinin in black beans on a brucid beetle. Science 192:795–796

Kalb AJ, Levitzki A (1968) Metal-binding sites of concanavalin A and their role in the binding of α-methyl-D-glucopyranoside. Biochem J 109:669–672

Kauss H, Glaser C (1974) Carbohydrate-binding proteins from plant cell walls and their possible involvement in extension growth. FEBS Lett 45:304–307

Kouchalakos RN, Bates OJ, Bradshaw RA, Hapner KD (1984) Lectin from sainfoin (*Onobrychis viciifolia* Scop.). Complete amino acid sequence. Biochemistry 23:1824–1830

Kronis KA, Carver JP (1982) Specificity of isolectins of wheat-germ agglutinin for sialyloligosaccharides: a 360-MHz proton nuclear resonance binding study. Biochemistry 21:3050–3057

Kubota J, Kanatani H (1975) Concanavalin A: its action in inducing oocyte maturation-inducing substance in starfish follicle cells. Science 187:654–655

Leak LV, Sun DC (1984) Effect of concanavalin A on lymph node macrophages: stimulation of endocytic cysternae. J Ultrastruct Res 86:1–17

Lebrun E, Rapenbusch RV, Foriers A, Hoebeke J (1983) Crystallization and preliminary x-ray diffraction studies of the mitogenic lentil (*Lens culinaris*) lectin. J Mol Biol 166:99–199

Lee JKN, Pachtman EA, Frumin AM (1974) Structural configuration of sugars and their ability to inhibit fava bean hemagglutinin. Ann NY Acad Sci 234:161–168

Lis H, Sharon N (1973) The biochemistry of plant lectins (phytohemagglutinins). Annu Rev Biochem 42:541–574

Lis H, Sharon N (1977) Lectins: their chemistry and application to immunology. In: Sela M (ed) The antigens, vol 4. Academic, New York, pp 429–529

Lis H, Sharon N (1986) Lectins as molecules and as tools. Annu Rev Biochem 55:35–67

Loontiens FG, Van Wauwe JP, DeGussem R, DeBruyne CK (1973) Binding of *para*-substituted phenyl glycosides to concanavalin A. Carbohydr Res 30:51–62

Lutsik MD, Panasyuk EN, Lutsik AD (1981) Lektiny (in Russian). Lvov State University, Lvov

Mäkelä O (1957) Studies in hemagglutinins of *Leguminosae* seeds. Dissertation, University of Helsinki

March PE, Thornton ER (1983) Binding of concanavalin A to calf brain synaptic vesicles. Biochem Biophys Res Commun 110:804–810

McClain D, Edelman GM (1976) Analysis of the stimulation–inhibition paradox exhibited by lymphocytes exposed to concanavalin A. J Exp Med 144:1494–1507

McClain D, D'Eustachio P, Edelman GM (1977) The role of surface-modulating assemblies in growth control of normal and transformed fibroblasts. Proc Natl Acad Sci USA 74:666–670

Mirelman D, Galun E, Sharon N, Lotan R (1975) Inhibition of fungal growth by wheat-germ agglutinin. Nature 256:414–416

Montreuil J (1975) Recent data on structure of carbohydrate moiety of glycoproteins–metabolic and biological implications. Pure Appl Chem 42:431–477

Neter E (1956) Bacterial hemagglutination and hemolysis. Bacteriol Rev 20:166–188

Nicolson GL (1974) The interactions of lectins with animal cell surfaces. Int Rev Cytol 39:89–190

Pflumm MN, Wang JL, Edelman GM (1971) Conformational changes in concanavalin A. J Biol Chem 246:4369–4375

Reeke GN, Becker JW (1986) Three-dimensional structure of favin: saccharide binding and cyclic permutation in leguminous lectins. Science 234:1108–1111

Reeke GN, Becker JW, Edelman GM (1975) The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. J Biol Chem 250:1525–1547

Reeke GN, Becker JW, Edelman GM (1978) Changes in the three-dimensional structure of concanavalin A upon demetallization. Proc Natl Acad Sci USA 75:2286–2290

Richardson M, Campos FDAP, Moreira RA, Ainouz IL, Begbie R, Watt WB, Pusztai A (1984) The complete amino acid sequence of the major α subunit of the lectin from the seeds of *Dioclea grandiflora* (Mart). Eur J Biochem 144:101–111

Rini JM, Carver JP, Hardman KD (1986) Crystallization and preliminary x-ray diffraction studies of a pea lectin-methyl-3,6-di-*O*-(α-D-mannopyranosyl)-α-D-mannopyranoside complex. J Mol Biol 189:259–260

Rizzo WB, Bustin M (1977) Lectins as probes of chromatin structure. J Biol Chem 252:7062–7067

Salunke DM, Swamy MJ, Khan MI, Mande SC, Surolia A, Vijayan M (1985) Arrangement of subunits in peanut lectin. Rotation function and chemical cross-linking studies. J Biol Chem 260:13576–13579

Santoro SA (1983) Differential effects of concanavalin A and succinyl concanavalin A on the macromolecular events of platelet activation. Biochim Biophys Acta 757:101–110

Schlessinger J, Koppel DE, Axelrod D, Jacobson K, Webb WW, Elson EL (1976) Lateral transport on cell membranes: mobility of concanavalin A receptors on myoblasts. Proc Natl Acad Sci USA 73:2409–2413

Schlessinger J, Elson EL, Webb WW, Yahara I, Rutishauser U, Edelman GM (1977) Receptor diffusion on cell surfaces modulated by locally bound concanavalin A. Proc Natl Acad Sci USA 74:1110–1114

Shaanan B, Shoham M, Yonath A, Lis H, Sharon N (1984) Crystallization and preliminary x-ray diffraction studies of soybean agglutinin. J Mol Biol 174:723–725

Sharon N, Lis H (1972) Lectins: cell-agglutinating and sugar-specific proteins. Science 177:949–959

Shoham M, Yonath A, Sussman JL, Moult J, Traub W, Kalb (Gilboa) AJ (1979) Crystal structure of demetallized concanavalin A: the metal-binding region. J Mol Biol 131:137–155

Simpson DL, Thorne DR, Loh HH (1978) Lectins: endogenous carbohydrate-binding proteins from vertebrate tissues. Functional role in recognition processes? Life Sci 22:727–748

Sletten K, Kolberg J, Michaelson TE (1983) The amino acid sequence of the α-subunit of a mitogenic lectin from seeds of *Lathyrus odoratus*. FEBS Lett 156:253–256

Smith CW, Goldman AS (1972) Effects of concanavalin A and pokeweed mitogen *in vivo* on mouse peritoneal macrophages. Exp Cell Res 73:394–398

So LL, Goldstein IJ (1967) Protein–carbohydrate interaction. IX. Application of the quantitative hapten inhibition technique to polysaccharide–concanavalin A interaction. Some comments on the forces involved in concanavalin A–polysaccharide interaction. J Immunol 99:158–163

So LL, Goldstein IJ (1968) Protein-carbohydrate interaction. XIII. The interaction of concanavalin A with α-mannans from a variety of microorganisms. J Biol Chem 243:2003–2007

Stewart ML, Summers DF, Soeira R, Fields BN, Maizel JV (1973) Purification of oncornaviruses by agglutination with concanavalin A. Proc Natl Acad Sci USA 70:1308–1312

Sumner JB (1919) The globulins of the jack bean, *Canavalia ensiformis*. J Biol Chem 37:137–142

Sumner JB, Howell SF (1936) The role of divalent metals in the reversible inactivation of jack bean hemagglutinin. J Biol Chem 115:583–588

Tanaka I, Abe Y, Hamada T, Yonemitsu O, Ishii S (1981) Monovalent monomer derivative of concanavalin A produced by photochemically-induced alkylation. J Biochem (Tokyo) 89:1643–1646

Unanue ER, Perkins WD, Karnovsky MJ (1972) Ligand-induced movement of lymphocyte membrane macromolecules. I. Analysis by immunofluorescence and ultrastructural radiography. J Exp Med 136:885–906

Villafranca JE, Robertus JD (1981) Ricin B chain is a product of gene duplication. J Biol Chem 256:554–556

Villafranca JJ, Viola RE (1974) The use of ^{13}C spin lattice relaxation times to study the interaction of α-methyl-D-glucopyranoside with concanavalin A. Arch Biochem Biophys 160:465–468

Wang JL, Edelman GM (1978) Binding and functional properties of concanavalin A and its derivatives. I. Monovalent, divalent, and tetravalent derivatives stable at physiological pH. J. Biol Chem 253:3000–3007

Wang JL, Becker JW, Reeke GN, Edelman GM (1974) Favin, a crystalline lectin from *Vicia faba*. J Mol Biol 88:259–262

Wang JL, Cunningham BA, Waxdal MJ, Edelman GM (1975a) The covalent and three-dimensional structure of concanavalin A. I. Amino acid sequence of cyanogen bromide fragments F_1 and F_2. J Biol Chem 250:1490–1502

Wang JL, McClain DA, Edelman GM (1975b) Modulation of lymphocyte mitogenesis. Proc Natl Acad Sci USA 72:1917–1921

Wei CH, Einstein JR (1974) Preliminary crystallographic data for a new crystalline form of abrin. J Biol Chem 249:2985–2986

Wright CS (1977) The crystal structure of wheat germ agglutinin at 2.2 Å resolution. J Mol Biol 111:439–457

Wright CS (1980) Crystallographic elucidation of the saccharide-binding mode in wheat germ agglutinin and its biological significance. J Mol Biol 141:267–291

Wright CS (1984) Structural comparison of the two distinct sugar binding sites in wheat germ agglutinin isolectin II. J Mol Biol 178:91–104

Wright CS (1987) Refinement of the crystal structure of wheat germ agglutinin isolectin 2 at 1.8 Å resolution. J Mol Biol 195:501–529

Wright CS, Olafsdottir S (1986) Structural differences in the two major wheat germ agglutinin isolectins. J Biol Chem 261:7191–7195

Wright CS, Gavilanes F, Peterson DL (1984) Primary structure of wheat germ agglutinin isolectin 2. Peptide order deduced from x-ray structure. Biochemistry 23:280–287

Wright HT, Brooks DM, Wright CS (1985) Evolution of the multidomain protein wheat germ agglutinin. J Mol Evol 21:133–138

Yahara I, Edelman GM (1972) Restriction of the mobility of lymphocyte immunoglobulin receptors by concanavalin A. Proc Natl Acad Sci USA 69:608–612

Yahara I, Edelman GM (1973) The effects of concanavalin A on the mobility of lymphocyte surface receptors. Exp Cell Res 81:143–155

Yariv J, Kalb AJ, Levitzki A (1968) The interaction of Con A with methyl-α-D-glucopyranoside. Biochem Biophys Acta 165:303–305

Binding Sites of Monoclonal Anti-Carbohydrate Antibodies

J. THURIN

1 Introduction

This chapter presents an overview of monoclonal antibodies (mAbs) directed to carbohydrates in mammalian systems with well-characterized specificities, together with some of the information obtained from their use. Because of the rapidly increasing number of this kind of reagent not all available mAbs are included. Naturally occurring mAbs and embryological markers and their suspected immunogenicity have been extensively reviewed by HAKOMORI (1984) and FEIZI (1985).

With few expections, the binding epitopes of most known mAbs against carbohydrate antigens include the terminal saccharide chains of glycolipids and glycoproteins. The sugars most commonly involved are galactose/galactosamine, usually coterminated with fucose and/or neuraminic acid. For the mAbs that recognize human gastrointestinal tumor-associated antigens such binding epitopes imply aberrant blood group determinants (HAKOMORI and JEANLOZ 1964; MAGNANI et al. 1981; and HAKOMORI 1985) as well as normal blood group antigens (BROCKHAUS et al. 1981; HANSSON et al. 1983; BLASZCZYK et al. 1984a). For antibodies against human anti-melanoma antigens, these epitopes are on certain ganglioside antigens normally more abundant in embryonic tissues (YEH et al. 1982; PUKEL et al. 1982; NUDELMAN et al. 1982; WATANABE et al. 1982; CAHAN et al. 1982; TAI et al. 1983; CHERESH et al. 1984a; THURIN et al. 1985; DUBOIS et al. 1986).

In this review the apparent emphasis on carbohydrate determinants elucidated by studying glycolipids rather than glycoproteins, mucins, or proteoglycans reflects in part methodological considerations, because the separation technolo-

Wistar Institute of Anatomy and Biology, 3601 Spruce Street, Philadelphia, PA 19104, USA

Current Topics in Microbiology and Immunology, Vol. 139

gy for glycolipids has until recently been superior to that for released oligosaccharides. However, it might also be that some of the carbohydrate chains on glycolipids are in fact more immunogenic than those present on glycoproteins.

The focus of the discussion is on carbohydrate structures present on tumor cells as antigens, since most of the available antibodies have been derived by immunizing mice with live human tumor cells and/or material purified from them (KOPROWSKI et al. 1978, 1979). Screening procedures are usually directed to the selection of antibodies that recognize tumor-associated antigens expressed in large amounts and to the further selection for intrinsic properties of the immunoglobulin itself in order to obtain the most suitable mAbs for immunotherapy, imaging, serology, immunohistopathology, etc. (KOPROWSKI 1984; HERLYN and KOPROWSKI 1985; HERLYN, to be published). A second generation of antibodies now emerging involves selection with known tumor-associated antigens to optimize the antigen–antibody interaction. Most of the data enabling the detailed interpretation of protein–carbohydrate interactions have been derived by studying antibody–carbohydrate interactions using synthetic carbohydrate determinants (LEMIEUX 1982, 1985; HINDSGAUL et al. 1982; BAKER et al. 1983; LEMIEUX et al. 1984; SPOHR et al. 1985a, b; HINDSGAUL et al. 1985; LEMIEUX et al. 1985; CHEN and KABAT 1985).

2 Antigen Specificity

Tables 1 and 2 present the glycolipid structures of most carbohydrate antigens for which mAbs are available and information about the binding specificities of these mAbs. The antigens are listed according to their respective core structures and thus belong to either the globo-, lacto-, or ganglioseries (IUPAC-IUB 1978). The few sulfate-containing lipids for which antibody reactivity is known are listed separately. Structures that show reactivity with at least one mAbs are marked with an asterisk. The extent of this reactivity is indicated in Table 2, where the major antigens detected by the cited mAbs are shown. References are to antibody characterizations. Reactivity with carbohydrate structures, usually tested on glycolipids, is graded from 1 to 5 (with 5 the highest). This scale represents relative reactivity obtained in the particular test systems used in the respective reports, and therefore the strengths of different antibody-binding interactions cannot be compared with another. These data were included since a cross-reactivity to a certain epitope can be useful in a system where other epitopes to which the mAb binds are absent. They also indicate to what extent the mAbs have been characterized.

From Table 1 it is clear that most of the anti-carbohydrate mAbs available detect antigens on glycolipids of the lactoseries, because this category contains most of the reactive species marked with an asterisk. However, the relevance of this observation is unclear, since the tissues most often used in glycolipid studies are erythrocytes and normal and malignant gastrointestinal tract speciments, which are rich in this type of structure (HAKOMORI 1981; BREIMER et al. 1982; HAKOMORI and KANNAGI 1983; BREIMER 1984; HANSSON et al. 1984).

Table 1. Structures of carbohydrate antigens used to test specifities of mAbs

Structure	Assigned number	Abbreviation
Globoseries		
Galβ1 → 1Cer	1	GalCer
Glcβ1 → 1Cer	2	GlcCer
Galα1 → 4Galβ1 → 4Glcβ1 → 1Cer[a]	3	Gb_3Cer
Galα1 → 3Galβ1 → 4Glcβ1 → 1Cer	4	iGb_3Cer
GalNAcβ1 → 3Galα1 → 4Galβ1 → 4Galβ1 → 1Cer[a]	5	Gb_4Cer
GalNAcβ1 → 3Galα1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	6	iGb_4Cer
GalNAcβ1 → 3Galα1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	7	$iGb\text{-}nLc_6Cer$
GalNAcα1 → 3GalNAcβ1 → 3Galα1 → 4Galβ1 → 4Glcβ1 → 1Cer[a]	8	Forssman
Galβ1 → 3GalNAcβ1 → 3Galα1 → 4Galβ1 → 4Glcβ1 → 1Cer[a]	9	Gb_5Cer
Fucα1 → 2Galβ1 → 3GalNAcβ1 → 3Galα1 → 4Galβ1 → 4Glcβ1 → 1Cer[a]	10	$H\text{-}Gb_4Cer$
GalNAcα1 → 3Gal(2 ← 1αFuc)β1 → 3GalNAcβ1 → 3Galα1 → 4Galβ1 → 4Glcβ1 → 1Cer[a]	11	$A\text{-}Gb_4Cer$
GalNAcβ1 → 3GalNAcβ1 → 3Galα1 → 4Galβ1 → 4Glcβ1 → 1Cer	12	GalNAcβ1 → $3Gb_4Cer$
NeuNAcα2 → 3Galβ1 → 3GalNAcβ1 → 3Galα1 → 4Galβ1 → 4Glcβ1 → 1Cer[a]	13	$NeuNAc\text{-}Gb_5Cer$
Lactoseries		
Galβ1 → 4Glcβ1 → 1Cer[a]	14	LacCer
NeuNAcα2 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	15	GM3
NeuNGcα2 → 3Galβ1 → 4Glcβ1 → 1Cer	16	GM3(NeuNGc)
NeuNAcα2 → 8NeuNAcα2 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	17	GD3
9-O-acetyl-NeuNAcα2 → 8NeuNAcα2 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	18	9-O-Ac-GD3
NeuNAcα2 → 8NeuNAcα2 → 8NeuNAcα2 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	19	GT3
GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	20	Lc_3Cer
Galβ1 → 3GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	21	Lc_4Cer
Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	22	nLc_4Cer
GalNAcβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	23	GalNAcβ1 → $4Lc_3Cer$
Galα1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	24	Galα1 → $3nLc_4Cer$
Galβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer	25	Galβ1 → $3nLc_4Cer$
GlcNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	26	nLc_5Cer
Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	27	nLc_6Cer

Table 1 (continued)

Structure	Assigned number	Abbreviation
GalNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer	28	GalNAcβ1 → 3nLc$_4$Cer
NeuNAcα2 → 3Galβ1 → 3GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	29	NeuNAc-Lc$_4$Cer
NeuNAcα2 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	30	NeuNAc-3-nLc$_4$Cer
NeuNAcα2 → 6Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	31	NeuNAc-6-nLc$_4$Cer
NeuNAcα2 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer	32	NeuNAc-3-nLc$_6$Cer
NeuNAcα2 → 6Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	33	NeuNAc-6-nLc$_6$Cer
NeuNAcα2 → 8NeuNAcα2 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	34	DPG
Fucα1 → 2Galβ1 → 3GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	35	H-5-1
Fucα1 → 2Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	36	H-5-2
Fucα1 → 2Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	37	H-7-2
Fucα1 → 2Galβ1 → 4GlcNAcβ1 → 3Gal(6 ← 1βGlcNAc4 ← 1βGal2 ← 1Fuc)β1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	38	H-10-2
Galβ1 → 3GlcNAc(4 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	39	Le^a-5
Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	40	X-5
Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	41	X-7-term
Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	42	X-7-intern
Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	43	X-8
Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	44	X-9-term
Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer	45	X-10
Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	46	X-11
NeuNAcα2 → 3Galβ1 → 3GlcNAc(4 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	47	NeuNAc-Le^a-5
NeuNAcα2 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	48	NeuNAc-X-5
NeuNAcα2 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	49	NeuNAc-X-7-intern
NeuNAcα2 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	50	NeuNAc-7-term
NeuNAcα2 → 3Galβ1 → 3GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	51	NeuNAc-X-8
Fucα1 → 2Galβ1 → 3GlcNAc(4 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	52	Le^b-6
Fucα1 → 2Galβ1 → 4GlcNAc(3 ← 1αFuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer	53	Y-6
GalNAcα1 → 3Gal(2 ← 1αFuc)β1 → 3GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	54	A-6-1

GalNAcα1 → 3Gal(2 ← 1 α Fuc)β1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	55	A-6-2
GalNAcα1 → 3Gal(2 ← 1 α Fuc)β1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	56	A-8-2
Fucα1 → 2Galβ1 → 3GalNAcα1 → 3Gal(2 ← 1 α Fuc)β1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer	57	H-8-3
GalNAcα1 → 3Gal(2 ← 1 α Fuc)β1 → 3GlcNAc(4 ← 1 α Fuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	58	A-7-1
GalNAcα1 → 3Gal(2 ← 1 α Fuc)β1 → 4GlcNAc(3 ← 1 α Fuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	59	A-7-2
GalNAcα1 → 3Gal(2 ← 1 α Fuc)β1 → 3GalNAcα1 → 3Gal(2 ← 1 α Fuc)β1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	60	A-9-3
Galα1 → 3Gal(2 ← 1 α Fuc)β1 → 3GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	61	B-6-1
Galα1 → 3Gal(2 ← 1 α Fuc)β1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	62	B-6-2
Galα1 → 3Gal(2 ← 1 α Fuc)β1 → 3GlcNAc(4 ← 1 α Fuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	63	B-7-1
Galα1 → 3Gal(2 ← 1 α Fuc)β1 → 4GlcNAc(3 ← 1 α Fuc)β1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	64	B-7-2
Galβ1 → 4GlcNAcβ1 → 3Gal(6 ← 1βGlcNAc4 ← 1βGal)β1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	65	I-8
Ganglioseries		
NeuNAcα2 → 3Galβ1 → 1Cer	66	GM4
GalNAcβ1 → 4Gal(3 ← 2NeuNAc8 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer[a]	67	GD2
GalNAcβ1 → 4Gal(3 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer[a]	68	GM2
GalNAcβ1 → 4Galβ1 → 4Glcβ1 → 1Cer	69	Gg_3Cer
Galβ1 → 3GalNAcβ1 → 4Galβ1 → 4Glcβ1 → 1Cer	70	Gg_4Cer
Fucα1 → 2Galβ1 → 3Galβ1 → 3GalNAcβ1 → 4Galβ1 → 4Glcβ1 → 1Cer	71	B-Gg_4Cer
Fucα1 → 2Galβ1 → 3GalNAcβ1 → 4Galβ1 → 4Glcβ1 → 1Cer	72	Fuc-Gg_4Cer
Galβ1 → 3GalNAcβ1 → 4Gal(3 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer	73	GM1a
NeuNAcα2 → 3Galβ1 → 3GalNAcβ1 → 4Galβ1 → 4Glcβ1 → 1Cer[a]	74	GM1b
Fucα1 → 2Galβ1 → 3GalNAcβ1 → 4Gal(3 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer[a]	75	Fuc-GM1
NeuNAcα2 → 3Galβ1 → 3GalNAcβ1 → 4Gal(3 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer	76	GD1a
Galβ1 → 3GalNAcβ1 → 4Gal(3 ← 2NeuNAc8 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer[a]	77	GD1b
Fucα1 → 2Galβ1 → 3GalNAcβ1 → 4Gal(3 ← 2NeuNAc8 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer	78	Fuc-GD1b
NeuNAcα2 → 8NeuNAcα2 → 3Galβ1 → 3GalNAcβ1 → 4Gal(3 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer	79	GT1a
NeuNAcα2 → 3Galβ1 → 3GalNAcβ1 → 4Gal(3 ← 2NeuNAc8 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer	80	GT1b
NeuNAcα2 → 8NeuNAcα2 → 3Galβ1 → 3GalNAcβ1 → 4Gal(3 ← 2NeuNAc8 ← 2NeuNAc)β1 → 4Glcβ1 → 1Cer	81	GQ1b
Sulfate lipids		
SO_4-3Galβ1 → 1Cer[a]	82	Sulfatide
SO_4-3GlcUAβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	83	SO_4-GlcUAnLc_4Cer
SO_4-3GlcUAβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4GlcNAcβ1 → 3Galβ1 → 4Glcβ1 → 1Cer[a]	84	SO_4-GlcUAnLc_6Cer

[a] Indicates structure for which reactivity with antibody has been described
For nomenclature see IUPAC-IUB in references

Table 2. Anti-carbohydrate specificity of mAbs

Major antigen	Antibody code	Specificity[b]	Reference
Globoseries[a]			
3. Gb_3Cer	38.12	3(5), 4, 5, 14	Nudelman et al. (1983)
	KH-1	3(5), 28(1), 1, 2, 5, 14, 22	Ito et al. (1984)
5. Gb_4Cer	Several	5(5), 6(5), 3, 4, 7, 8	Brodin et al. (1986)
6. iGb_4Cer	4, 17, 44, 85	6(5), 7(2), 3, 4, 5, 8	Brodin et al. (1986)
	11 and 49	6(5), 7(5), 3, 4, 5, 8	Brodin et al. (1986)
7. iGb-nLc_6Cer	46 and 86	7(5), 6(2), 3, 4, 5, 8	Brodin et al. (1986)
	11 and 49	6(5), 7(5), 3, 4, 5, 8	Brodin et al. (1986)
8. Forssman	H1-C4	8(5), 3, 5, 14	Nowinski et al. (1980) Willison et al. (1982)
9. Gb_5Cer	SSEA-3/MC631	9(5), 5(2), 10(2), 13(2), 6(1), 8(1), 3, 12, 25, 28, 30	Kannagi et al. (1983b)
10. H-Gb_4Cer	MBr 1	10(5), 75(3), 9(2), 21, 22, 36, 71	Breimer et al. (1984)
13. NeuNAc-Gb_5Cer	SSEA-4/ MC813-70	13(5), 74(2), 76(1), 3, 5, 8, 9, 10, 30, 69, 70	Kannagi et al. (1983a)
Lactoseries			
14. LacCer	T5A7	14(5), 1, 2, 3, 5, 9, 15, 17, 21, 32, 68, 69, 70, 73, 75, 76	Symington et al. (1984a)
	KH-2	14(5), 22(5), 1, 2, 3, 5, 24	Ito et al. (1984)
15. GM3	M2590	15(5), 30(5), 14, 16, 17, 22, 66–68, 73, 74	Hirabayashi et al. (1985)
17. GD3	4.2	17(5), 34(5), 15, 18, 73, 76, 77, 79–81	Nudelman et al. (1982)
	R24	17(5), 34(5–3), 15, 18, 66–68, 73, 76, 77, 80, 81	Pukel et al. (1982) Siddiqui et al. (1984)
	MB3.6	17(5), 15, 67, 68, 73, 76	Cheresh et al. (1984b)
	2B2, 1F4, 109, (4.2)	17(5), 34(3), 15, 18, 30, 31, 67, 68, 80, 81	Brodin et al. (1985)
	ME24	17(5), 67(1), 15, 18, 19, 66, 68, 73, 76, 77, 79–81	Thurin et al. (to be published)
	ME356	17(5), 15, 18, 19, 66–68, 73, 76, 77, 79–81	Thurin et al. (to be published)
18. 9-O-Ac-GD3	D1.1	18(5), 15, 17, 67, 68, 73, 76	Cheresh et al. (1984a)
	ME311	18(5), 15, 17, 19, 30, 34, 66–68, 73, 76, 77, 79–81	Thurin et al. (1985)
19. GT3	18B8	19(5), 15, 17, 68, 76, 77, 80, 81	Dubois et al. (1986)
20. Lc_3Cer	J1	20(5), 23(2), 27(2), 5, 22, 40, 69	Symington et al. (1984b)
21. Lc_4Cer	FC10.2	21(5), 22 see text	Gooi et al. (1983)

Table 2 (continued)

Major antigen	Antibody code	Specificity[b]	Reference
22. nLc_4Cer	IB2	22(5), 27(3), 37(3), 20, 21, 30, 70	Young et al. (1981b)
	KH-2	22(5), 14(5), 1, 2, 3, 5, 24	Ito et al. (1984)
31. NeuNAc-6-nLc_4Cer	IB9	31(5), 33(5), 49(5), 15, 22, 30, 73	Hakomori et al. (1983b)
33. NeuNAc-6-nLc_6Cer	IB9	33(5), 31(5), 49(5), 15, 22, 30, 73	Hakomori et al. (1983b)
34. DPG	4.2, R24	34(5), 17(5), 15, 18, 73, 76, 77, 79–81	Brodin et al. (1985)
35. H-terminal	BE2	36(5), 37(5), 38(5), 35(1)	Young et al. (1981b)
36. H-5-2	102	36[c](5) see text	Fredman et al. (1983)
	G10	36[d](5), 53[d](2), 35[d], 39[d], 40[d], 52[d]	Kimmel et al. (1986)
39. Le^a-5	CF4-C4, CF4-F4, BC9-E5, DG4-1	39(5), 21, 22, 35, 36, 40	Young et al. (1983)
	CO514	39(5), 21, 22, 35, 36, 40	Blaszczyk et al. (1984)
40. X-5	Anti-SSEA-1	40(5), 41(5), 3, 5, 8, 14, 15, 20, 22, 30, 35–39 52, 53, 55, 62–64	Nudelman et al. (1980) Gooi et al. (1981) Hounsell et al. (1981) Brockhaus et al. (1982)
	D1-56-22	40(5), 41(5), 22, 35, 36, 39, 52, 53, 55, 62–64	Hansson et al. (1983)
	1G10	40(5), 41(5), 43(5), 44(5), 42	Urdal et al. (1983)
	My-1	see text	Huang et al. (1983)
	PMN-6, -29, -81	see text	Magnani et al. (1984)
	FH-2 and FH-3	40(5), 41(5), 43(5), 46(5), 42	Fukushi et al. (1984a)
	Anti-7A	see text	Yamamoto et al. (1985)
	Several antibodies	see text	Uemura et al. (1985)
41. X-7-term	FH-2 and FH-3	41(5), 40(5), 43(5), 46(5), 42	Fukushi et al. (1984a)
43. X-8	FH-4	43(5), 46(5), 37, 40–42	Fukushi et al. (1984a)
46. X-11	FH-1	46(5), 42(2), 43(2), 37, 40, 41	Fukushi et al. (1984a)
	FH-4	46(5), 43(5), 37, 40–42	Fukushi et al. (1984a)
	FH-5 and ACF-H18	46(5), 42(2), 43(2), 37, 40, 41	Fukushi et al. (1984a)
47. NeuNAc-Le^a-5	CO19-9	47(5), 5, 29, 39, 40, 48, 52–59, 61–64, 66–68, 73, 76, 77, 79–81	Magnani et al. (1982) Månsson et al. (1985)
	CA-50	47(5), 29(3), 48, 73, 76, 77, 79–81	Nilsson et al. (1985)
	CO29-11	47(5), 39(1), 21, 35, 40, 52, 53, 73, 76, 77, 79–81	Herlyn et al. (1985)
48. NeuNAc-X-5	CSLEX1	48(5), 50(5), 51(5), 30, 43, 47, 48	Fukushima et al. (1984)

Table 2 (continued)

Major antigen	Antibody code	Specificity[b]	Reference
50. NeuNAc-X-7-term	CSLEX1	50(5), 48(5), 51(5), 30, 43, 47, 48	Fukushima et al. (1984)
51. NeuNAc-X-8	FH-6	51(5), 43, 47–49	Fukushi et al. (1984b)
	CSLEX1	51(5), 48(5), 50(5), 30, 43, 47, 48	Fukushima et al. (1984)
52. Le^b-6	NS-10-17	52(5), 35(1), 21, 22, 36, 39, 40, 53–55, 61, 62	Brockhaus et al. (1981)
	CO-431	52(5), 21, 22, 35, 36, 39, 40, 53, 58, 63	Blaszczyk et al. (1984)
	CO-432	52(5), 35(2), 21, 22, 35, 36, 39, 40, 52, 58, 63	Blaszczyk et al. (1984)
	CO-512	52(5), 39(3), 21, 22, 35, 36, 39, 40, 53, 58, 63	Blaszczyk et al. (1984)
53. Y-6	AH-6	53(5), 22, 36, 40, 52	Abe et al. (1983)
	75.12	53[c](5), 8[c], 21[c], 22[c], 35[c], 36[c], 39[c], 40[c], 52[c], 55[c], 59[c], 61[c], 65	Blaineau et al. (1983)
	C14/1/46/10	53[d](5), 35[d], 52[d]	Brown et al. (1983)
	F3	53(5), 21, 22, 35, 36, 39, 40, 52	Lloyd et al. (1983)
	101	see text	LePendu et al. (1985)
	Several antibodies	see text	Uemura et al. (1985)
	BR55.2	53(5), 64(1), 5, 6, 8, 21, 22, 35, 36, 39, 40, 47, 52, 62, 63, 64	Thurin et al. (to be published)
	BR15-6a	53(5), 5, 6, 8, 21, 22, 35, 36, 39, 40, 47, 52, 62, 63, 64	Rodeck et al. (to be published)
54. A-6-1	AH-16	54(5), 55(2), 56(2), 8, 22, 36	Abe et al. (1984)
	AH-21	54(5), 8, 22, 36, 55, 56	Abe et al. (1984)
55. A-6-2	AbS12	55(5), 56(5), 11, 60	Furukawa et al. (1985)
56. A-8-2	AbS12	56(5), 55(5), 11, 60	Furukawa et al. (1985)
58. A-7-1	HH-3	58(5), 59(5), 54–56, 60	Clausen et al. (1985a)
59. A-7-2	HH-3	59(5), 58(5), 54–56, 60	Clausen et al. (1985a)
60. A-9-3	TH-1	60(5), 11(1), 37, 52, 53, 55–59	Clausen et al. (1985b)
62. B-6-2	E1-15-2 and E1-66-1	62(5), 64(5), 52(2), 63(2), 35, 36, 39, 40, 53, 61	Hansson et al. (1983)
	E2-83-52	62(5), 64(5), 61(3), 63(3), 52(2), 35, 36, 39, 40, 53	Hansson et al. (1983)
64. B-7-2	E1-15-2 and E1-66-1	64(5), 62(5), 63(3), 52(2), 35, 36, 39, 40, 53	Hansson et al. (1983)
	E2-83-52	64(5), 62(5), 61(3), 63(3), 52(2), 63(2), 35, 36, 39, 40, 53	Hansson et al. (1983)
65. I-8			Feizi (1985)[e]

Table 2 (continued)

Major antigen	Antibody code	Specificity[b]	Reference
Ganglioseries			
67. GD2	Anti-OFA-I-2	67(5), 15, 17, 68, 73, 76, 77, 80	Cahan et al. (1982)
	126	67(5), 17(2), 81(2), 15(1), 77(1), 80(1), 68, 73, 76	Cheresh et al. (1984b)
	3F8 (3A7, 3G6, 2F7)	67(5), 15, 17, 66, 68, 73, 76, 77, 80, 81	Saito et al. (1985)
	ME361	67(5), 17(3), 15, 18, 19, 66–68, 73, 76, 77, 79–81	Thurin et al. (to be published)
68. Gg_3Cer	Anti-Gg3Cer	see text	Young et al. (1979)
70. Gg_4Cer	NCC-LU-35	see text	Hirohashi et al. (1985)
75. Fuc-GM1	Anti-Fuc-GM1	see text	Fredman et al. (1986)
77. GD1b	Anti-GD1b	77(5), 15, 17, 67, 68, 73, 75, 76, 78–81	Fredman et al. (1985)
Sulfate lipids			
82. Sulfatide		82(5)	Goujet-Zalc (1986)
83. SO_4-GlcUA-nLc_4Cer	HNK-1 or Leu-7	83(5)	Chou et al. (1985a)
84. SO_4-GlcUA-nLc_6Cer	HNK-1 or Leu-7	84(5)	Chou et al. (1985a, b)

[a] Same numbers as in Table 1

[b] First number indicates structure tested from Table 1 and number in parentheses indicates relative reactivity on a scale of 1 to 5, with 5 the highest. Numbers with no parentheses indicate tested but nonreactive

[c] Specificity tested with oligosaccharides

[d] Specificity tested with synthetic saccharide identical to respective glycolipid terminal but with $-(CH_2)_8-CO_2Me$ substituted for an internal LacCer

[e] See reference for review of these antibodies and specificities

Table 1 also indicates that the antibody-binding epitopes are usually found on terminal portions of the saccharide chains and only infrequently on the internal or core structures, e.g., mAbs binding #4, 5, 10, 11, 27 in Table 2. Most mAbs are derived from the mouse, which in general is reactive to the same antigens as man, although there are exceptions. The carbohydrate terminals that typically give rise to mAbs are those of the human blood group systems ABO, Lewis, and Ii. This makes the grouping according to core structures in the two tables somewhat arbitrary, since blood group active terminals have been found on all three types of core structures, e.g., in structures H-Gb_4Cer, A-Gb_4Cer, and B-Gg_4Cer (#10, 11, 71 in Table 1), as well as in antigens based on the usual lactoseries core. These antigens are expressed on cells lining body cavities and are normally present in tissues that show a high degree of regeneration, i.e., gastrointestinal epithelium, vascular endothelium, and urinary tract epithelium

(SZULMAN 1960, 1964; ERNST et al. 1984a, b). The positional isomers of the Le^a and Le^b antigens (Table 1, #39, 52), the X and Y antigens (#40, 53) and especially the X antigen (GINSBURG et al. 1984, 1985) also frequently give rise to antibody-producing hybridomas. However, the fact that most carbohydrate-directed mAbs bind to the X antigen and not to other blood group type antigens might reflect a bias created by the secondary screening systems used in antibody production. These systems have been applied by most laboratories, since it was found that human blood group antigens give rise to tumor-associated antigens (BROCKHAUS et al. 1981; HANSSON et al. 1983). Negative screening for blood group activity using erythrocytes of different blood groups does not exclude the X antigen, which is genetically unrelated to the ABO and Lewis systems (WATKINS 1980). On the other hand, the X antigen in the mouse is an embryonic antigen referred to as SSEA-1 (stage-specific embryonic antigen; SOLTER and KNOWLES 1978; NUDELMAN et al. 1980; GOOI et al. 1981), which could explain the high yield of hybridomas derived, provided that embryonic antigens are particularly immunogenic.

The data in Tables 1 and 2 reveal that blood groups A and B on both type 1 (–Galβ1 → 3GlcNac–) and type 2 chains (–Galβ1 → 4GlcNAc–) (#54, 55, 58, 59, 61–64) as well as on blood group H type 2 chains (#36) frequently have mAb-directed epitopes. There is, however, not a single report of an anti-blood group H type 1 specific mAb, although some anti-Le^b mAbs do cross-react with blood group H type 1 (Table 2, #52, NS-10-17 and CO432), and an antibody directed against the type 1 core structure Lc_4Cer (Table 2, #21) has been reported. This might be owing to a restriction in the mouse specificity repertoire, indicating that a similar structure might exist in the mouse and make it tolerant to this particular antigen. This observation, among others, leads to an important point regarding interpretation of mAb specificity data. The glycolipids are selected for testing on the basis of carbohydrate sequence similarity. This does not rule out a priori binding to structures with different sequences. For instance, it was recently shown that the terminal saccharides of Le^b and Y antigen, for which the conformations are known (SPOHR et al. 1985a) (Table 1, #52, 53), share a common epitope, i.e., have a similar surface that interacts specifically with a protein, in this case with lectin IV from *Griffonia simplicifolia* II (LEMIEUX 1985a). In this case, two antigens with different carbohydrate sequences share a similar epitope, and it remains possible that other unknown and sequence-unrelated structures are also reactive. Before the actual structure responsible for the reaction has been characterized, no conclusions based on the sequence alone should be reached.

When comparing antibodies directed to similar or even identical epitopes, another difficulty becomes evident. Differences in isotypes that lead to different reactivities with the second antibody in the test system, and thus an error factor in the analysis of primary antibody binding, point to the advantage of using purified directly labeled antibodies. Unfortunately, such an approach is impractical in most cases, but even if it were practical, mAbs of different isotypes might react differently in, for example, solid-phase RIA than in immunochemistry (C. ERNST, personal communication; STEPLEWSKI et al. 1985); such differences could rest in the different degrees of steric hindrance involved in the

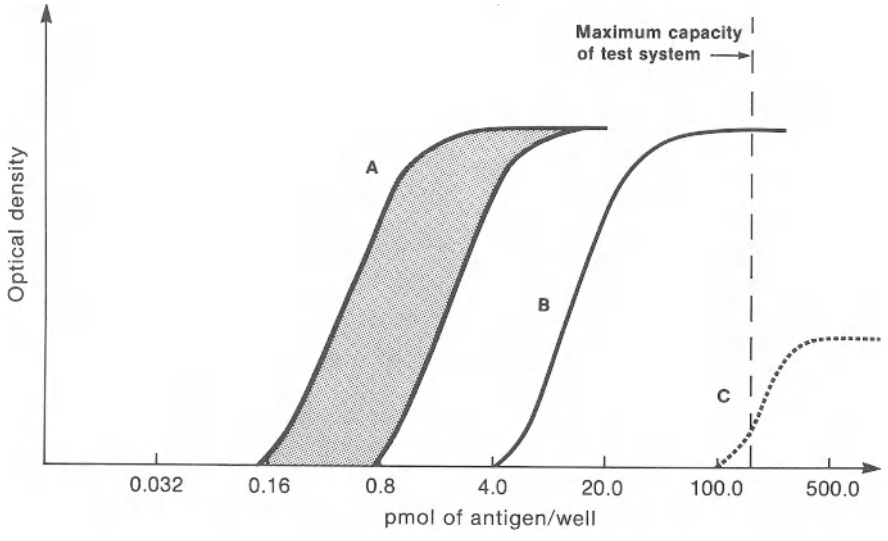


Fig. 1. Schematic presentation of typical results in ELISA (or RIA) performed on serially diluted glycolipid antigens coated on microtiter plates. The *y* axis displays amount of antibody binding, and the *x* axis the logarithmic concentration of antigen

reactivities of antibodies with different isotypes to a tissue section compared to the less complicated solid-phase system.

Figure 1 gives typical results of the most frequently used assay for antibody-binding specificity on glycolipids. The antibody described binds to three different hypothetical glycolipids, and the antibody dilution used was predetermined to give maximal binding to the "major" antigen (arbitrarily defined as grade 5-reactive, Table 2). The antibody is then tested against serially diluted antigens and transferred to 96-well microtiter plates. Reactivity of the typical mAb, added to the wells and detected with a labeled second antibody and in this case predominantly antigen A, becomes detectable at 0.1–1 pmol/well. Antigen B would be a "cross-reactive" species (grades 1–4) and antigen C, a nonreactive species (grade 0). If *affinity* is defined as the strength with which two different mAbs bind to the same epitope with different van der Waal interactions or degree of fit between antibody and antigen, and if *specificity* refers to binding to different epitopes but not necessarily involving differences in van der Waal interactions, then a mAb with lower affinity but with the same specificity as that in Fig. 1 would be characterized as binding only to glycolipid A. All curves would be shifted to the right to the extent of the affinity decrease, which in this case must be sufficiently large to move curve B out of the range of the assay. The same effect would be observed if the secondary test system (the labeled anti-mouse reagent) was correspondingly less efficient. For mAbs with higher affinity, a cross-reactivity might be as strong as the normal reactivity for major antigens; for example, the reactivity of anti-Le^b antibody CO512 (#52) with Le^a (#39) is about 1/1000 that with Le^b, but it still shows sufficient reactivity to be useful in an assay from which both reactivities are wanted.

These definitions, while arbitrary, might be helpful in explaining why antibodies directed to the same carbohydrate antigens (major antigens) frequently show great discrepancies in binding in the same assay system, e.g., immunohistochemistry. Several examples of this and the reverse phenomenon, i.e., the higher affinity situation, have been observed (BLASZCZYK 1984a; HERLYN et al. 1985; unpublished observations).

3 Production of Anti-Carbohydrate Antibodies

The majority of mAbs directed to glycolipids have been obtained by immunizing mice with live human tumor cells. To obtain anti-blood group antigen mAbs, human erythrocytes have been used. The immunogenic dominance of certain antigens, however, has made it difficult to obtain antibodies to other minor and/or less immunogenic antigens on a given cell. Several strategies have been proposed to solve this problem (HERLYN et al. 1985). One involves arranging the primary screening so that the immunodominant antigens are discarded, thus revealing less frequent hybridomas. Another strategy involves immunization with purified antigen. Synthetic antigens (BUNDLE et al. 1982), natural antigens together with different adjuvants, e.g., glycolipids coated on *Salmonella minnesota* polysaccharides (GALANOS et al. 1971; YOUNG et al. 1979) and incorporated into lipid A containing liposomes for more efficient antigen expression (BRODIN et al. 1986) have all been tried. Thus far, no one approach is clearly superior to obtain a high yield of hybridomas for a given antigen. One major obstacle seems to be the repertoire of antigen specificities of the mouse. The use of a different species or approaches to circumvent the tolerance of mice for antigens that might be exposed in the same species could resolve this problem.

4 mAbs Directed to Carbohydrate Epitopes in Tumor Immunology

4.1 The Gastrointestinal System

Results of analysis with CO 19-9 (formerly 1116NS19-9) illustrates the usefulness of a highly specific anti-carbohydrate mAb. This antibody was obtained by immunization of mice with the human colorectal carcinoma cell line SW 1116, and its major antigen was described as sialylated Le^a (Table 1, #51) by MAGNANI et al. (1982) and FALK et al. (1983). With the aid of the antibody, immunoperoxidase data revealed that the antigen was also present in normal tissues at localizations coinciding with mucin-covered surfaces, e.g., gall bladder, pancreas (ATKINSON 1982), saliva (BROCKHAUS et al. 1985), and seminal fluid (UHLENBRUCK et al. 1984). The epitope-bearing antigen was originally identified as a glycolipid but was also detected on mucins (MAGNANI et al. 1983) and on glycoproteins (HANISCH et al. 1984). The earlier relationship of this antigen

to human blood group Le^a has been seriously challenged by the discovery that the biosynthetic pathway for the production of the sialylated Le^a epitope is not by addition of sialic acid to the Le^a antigen (Table 1, #39) but, instead, by addition of Fucα1 $\rightarrow$ 4 to the NeuNAc-Lc_4Cer structure (#33), thus forming the CO19-9 defined epitope (#47) (HANSSON and ZOPF 1985). The general nature of this biosynthetic pathway is suggested by the finding that the NeuNAc-X structure (Table 1, #48), originally described by BLASZCZYK et al. (1984b) and FUKUSHIMA et al. (1984), is biosynthesized in the same manner (HOLMES et al. 1986). In this context, HANSSON and ZOPF (1985) suggested that an antibody directed toward the precursor NeuNAc-Lc_4Cer (#33) might be more tumor-specific because the presence of this antigen would be independent of the blood group status of the patient. Antibody CA-50 (Table 2, #47) actually cross-reacts with this epitope (#33), perhaps explaining the somewhat broader tumor specificity of this antibody (LINDHOLM et al. 1983) compared to that of CO19-9 in immunoperoxidase assay.

It is important to consider the normal distribution of the sialylated Le^a antigen together with the antigenic form in which the carbohydrate epitope is localized. For example, when the antigen is used as a serum marker for tumor patients, in whom it exists as a mucin (MAGNANI et al. 1983), false-positive cases arise if the mucin detected in the serum actually reflects certain benign diseases, e.g., those causing stasis of gallbladder drainage. The ganglioside form of the antigen might be more tumor-specific, because it has so far not been clearly demonstrated in normal tissues. Thus, at least in theory, the ganglioside antigen could be useful in colorectal carcinoma screening by analyzing, for example, feces, where its detection might signify an early stage of tumor progression when the lesion is very small, and theoretically even before the antigen can be detected in the circulation (H. VERILL, personal communication). The ganglioside form of the antigen has been detected in precancerous tissue, i.e., in ulcerative colitis (OLDING et al. 1985), indicating the value in probing selectively for one specific antigen form.

If the antigen is to be targeted both in imaging of tumor patients and in immunotherapy, the mucin form of the antigen covering the tumor cell surface might prevent efficient binding of the antibody to the actual tumor cell membrane. Thus, the ganglioside antigen might be superior for this purpose, being in closer proximity to the cell surface, although unfortunately both antigen forms are often present in the gastrointestinal tract. In melanoma, however, only the glycolipid form has been detected (see below), making it ideally suited for use in imaging and immunotherapy, as noted early by YOUNG and HAKOMORI (1981a).

Analysis of expression of the CO19-9 defined epitope in human fetal tissues (RAUX et al. 1983; OLDING et al. 1984) has revealed the presence of the antigen in fetal small intestine but not in fetal colon. Other regional differences of blood group antigen expression have been found in normal human colon (ERNST et al. 1984a), where the expression of Lewis blood group antigens Le^a and Le^b (#39 and #52, respectively) was studied. BLASZCZYK et al. (1985) showed that the extracted Le^a antigen was distributed along the full length of the large intestine in patients with this blood group, but that the Le^b antigen was

expressed only in the upper two-thirds of the normal colon in Le^b individuals. However, carcinomas of the lower third of the colon expressed the Le^b antigen in Le^b-positive individuals. Thus the enzyme responsible for the biosynthesis of Le^b from Le^a, the secretor enzyme (WATKINS 1980), could have been induced during tumor progression. Alternatively, a tumor-specific glycosyltransferase is induced. The possible relationship of this tumor progression event to the secretor gene is intriguing. KUMAZAKI and YOSHIDA (1984) suggested that the gene product for the secretor gene encoded enzyme is a glycosyltransferase with characteristics different from the enzyme that catalyzes addition of Fuc$\alpha 1 \rightarrow 2$ to secreted blood group ABH substances. This novel enzyme might be induced during colorectal adenocarcinoma progression and thus act on unsecreted blood group antigens such as glycolipids. The findings of SAKAMOTO et al. (1986) and ABE et al. (1986) support this possibility.

4.2 The Melanoma System

It is not surprising that the first mAbs applied to the study human malignant melanoma were produced against the GD3 ganglioside (Table 2, #17), since it is the major melanoma ganglioside (PORTOUCALIAN et al. 1979). Because of the abundance of gangliosides in melanoma cells and the high degree of tumor specificity provided by these antigens, much effort has been directed toward the production of additional anti-ganglioside reagents for use in diagnosis, imaging, and therapy. Promising results in the immunotherapy of human malignant melanoma have recently been obtained by HOUGHTON et al. (1985), who report a beneficial effect in three of 12 patients injected intravenously with purified mouse IgG3 anti-GD3 ganglioside antibody R24 (Table 2, #17).

The first suggestions that both GM2 and GD2 gangliosides were immunogenic in humans came from studies noting that transformation of human lymphocytes with Epstein-Barr virus produced cells that secreted antibodies against the GM2 and GD2 gangliosides (TAI et al. 1982; CAHAN et al. 1982) and from an observation by WATANABE et al. (1982) that sera of melanoma patients reacted with the GD2 ganglioside. Evidence supporting this notion came from studies in which an antibody response to these gangliosides, particularly against GM2, was detected in melanoma patients after immunization with ganglioside extracts from melanoma cells suggested that these gangliosides were immunogenic in man (TAI et al. 1985). Interestingly, only the GD2 ganglioside seems to be immunogenic in mice, since no mouse mAb currently available reacts with GM2, despite the presence of this ganglioside in many melanoma cell lines used for immunization.

A novel human ganglioside, the 9-O-acetyl GD3, was described by CHERESH et al. (1984a) and THURIN et al. (1985) with the aid of mAbs. This ganglioside was more tumor-specific than the GD3 ganglioside, the latter previously suggested as a prime candidate for immunotherapy (HOUGHTON et al. 1985). The usefulness of this novel antigen in melanoma immunotherapy is presently awaiting clinical trials.

In our analysis of anti-melanoma ganglioside antibodies and their reactivity on cell extracts derived from tissues representing human melanoma progression, we detected the onset of ganglioside GD2 biosynthesis between the early primary and advanced primary melanoma stages (THURIN et al., to be published). The activity of the glycosyltransferase responsible for GD2 biosynthesis from its precursor GD3 was also found to correlate with induction of GD2 synthesis. Thus, the formation of GD2 correlated with the increased growth rate and acquisition of metastatic competence, i.e., the key changes observed between these two stages in the progression of human malignant melanoma.

5 Conclusion

The detailed structural information obtained from carbohydrate antigens has given the field of tumor immunology new tools in the form of highly specific and well-characterized mAbs and synthetic carbohydrate tumor antigens for use in test systems and future cancer therapy.

Acknowledgements. The valuable discussions with Drs. Gunnar C. Hansson, Dorothee Herlyn, Meenhard Herlyn, Ole Hindsgaul, Karl-Anders Karlsson, Alonzo Ross, Zenon Steplewski and Magdalena Thurin are greatly appreciated as is the editorial assistance of Ms. Marina Hoffmann.

References

Abe K, McKibbin JM, Hakomori S-i (1983) The monoclonal antibody directed to diffucosylated type 2 chain. Fucα1 $\rightarrow$ 2Galβ1 $\rightarrow$ 4GlcNAc(3 $\leftarrow$ 1αFuc)–; Y determinant. J Biol Chem 258:11793–11797

Abe K, Levery SB, Hakomori S-i (1984) The antibody specific for type 1 chain blood group A determinant. J Immunol 132:1951–1954

Abe K, Hakomori S-i, Oshiba S (1986) Differential expression of difucosyl type 2 chain (LeY) defined by monoclonal antibody AH6 in different locations of colonic epithelia, various histological types of colonic polyps, and adenocarcinomas. Cancer Res 46:2639–2644

Atkinson B, Ernst C, Herlyn M, Steplewski Z, Sears H, Koprowski H (1982) Gastrointestinal cancer-associated antigen in immunoperoxidase assay. Cancer Res 42:4820–4822

Baker DA, Sugii S, Kabat KA, Ratcliffe RM, Hermentin P, Lemieux RU (1983) Immunochemical studies of the combining sites of Forssman hapten specific hemagglutinins from *Dolichos biflorus* and *Helix pomatia*. Biochemistry 22:2741–2750

Blaineau C, LePendu J, Arnaud D, Connan F, Avner P (1983) The glycosidic antigen recognized by a novel monoclonal antibody, 75.12, is developmentally regulated on mouse embryonal carcinoma cells. EMBO J 2:2217–2222

Blaszczyk M, Hansson OC, Karlsson KA, Larson G, Strömberg N, Thurin J, Herlyn M, Steplewski Z, Koprowski H (1984a) Lewis blood group antigens defined by monoclonal anti-colon carcinoma antibodies. Arch Biochem Biophys 233:161–168

Blaszczyk M, Ross AH, Ernst CS, Marchisio M, Atkinson B, Pak KY, Steplewski Z, Koprowski H (1984b) A fetal glycolipid expressed on adenocarcinomas of the colon. Int J Cancer 33:313–318

Blaszczyk M, Pak KY, Herlyn M, Sears HF, Steplewski Z (1985) Characterization of Lewis antigens in normal colon and gastrointestinal adenocarcinomas. Proc Natl Acad Sci USA 82:3552–3556

Breimer ME (1984) Tissue specificity of glycosphingolipids as expressed in pancreas and small intestine of blood group A and B human individuals. Arch Biochem Biophys 228:71–85

Breimer ME, Hansson GC, Karlsson K-A, Leffler H (1982) Glycosphingolipids of rat tissues. J Biol Chem 257:557–568

Bremer EG, Levery SB, Sonnino S, Ghidoni R, Canevari S, Kannagi R, Hakomori S-i (1984) Characterization of a glycosphingolipid antigen defined by the monoclonal antibody MBr 1 expressed in normal and neoplastic epithelial cells of human mammary gland. J Biol Chem 259:14773–14777

Brockhaus M, Magnani JM, Blaszczyk M, Steplewski Z, Koprowski H, Karlsson KA, Larson G, Ginsburg V (1981) Monoclonal antibodies directed against the human Le^b blood group antigen. J Biol Chem 256:13223–13225

Brockhaus M, Magnani JM, Herlyn M, Blaszczyk M, Steplewski Z, Koprowski H, Ginsburg Y (1982) Monoclonal antibodies directed against the sugar sequence of lacto-*N*-fucopentaose III are obtained from mice immunized with human tumors. Arch Biochem Biophys 217:647–651

Brockhaus M, Wysocka M, Magnani JL, Steplewski Z, Koprowski H, Ginsburg V (1985) Normal salivary mucin contains the gastrointestinal cancer-associated antigen detected by monoclonal antibody 19-9 in the serum mucin of patients.Vox Sang 48:34–38

Brodin T, Hellström I, Hellström K-E, Karlsson K-A, Sjögren H-O, Strömberg N, Thurin J (1985) Mouse monoclonal antibodies with specificity for melanoma-associated disialyllactosylceramide (GD3) also react with the structural analogue disialylparagloboside. Biochim Biophys Acta 837:349–353

Brodin T, Thurin J, Strömberg N, Karlsson K-A, Sjögren H-O (1986) Production of oligosaccharide-binding monoclonal antibodies of diverse specificities by immunization with purified tumor-associated glycolipids inserted into liposomes with lipid A. Eur J Immunol (in press)

Brown A, Feizi T, Gooi HC, Embleton MJ, Picard JK, Baldwin RW (1983) A monoclonal antibody against human colonic adenoma recognizes difucosylated type-2-blood-group chains. Biosci Rep 3:163–170

Bundle DR, Gidney MAJ, Kassam N, Rahman AFR (1982) Hybridomas specific for carbohydrates; synthetic human blood group antigens for the production, selection, and characterization of monoclonal typing reagents. J Immunol 129:678–682

Cahan LD, Irie RF, Singh R, Cassidenti A, Paulson JC (1982) Identification of a human neuroectodermal tumor antigen (OFA-1-2) as a ganglioside. Proc Natl Acad Sci USA 79:7629–7633

Chen H-T, Kabat EA (1985) Immunological studies of blood groups. The combining site specificities of mouse monoclonal hybridoma anti-A and anti-B. J Biol Chem 260:13208–13217

Cheresh DA, Varki AP, Varki NM, Stallcup WB, Levine J, Reisfeld RA (1984a) A monoclonal antibody recognizes an *O*-acetylated sialic acid in a human melanoma associated ganglioside. J Biol Chem 259:7453–7459

Cheresh DA, Harper JR, Schulz G, Reisfeld RA (1984b) Localization of the gangliosides GD2 and GD3 in adhesion plaques and on the surface of human melanoma cells. Proc Natl Acad Sci USA 81:5767–5771

Chou DKH, Ilyas AA, Evans CE, Quarles RH, Jungalwala FB (1985a) Structure of a glycolipid reacting with monoclonal IgM in neuropathy and with HNK-1. Biochem Biophys Res Commun 128:383–388

Chou DKH, Ilyas AA, Evans CE, Quarles RH, Jungalwala FB (1985b) Sulfated glucuronyl neolactoseries of glycoilipids in peripheral nerves. In: Proceedings of the VIIIth international symposium of glycoconjugates (additional Abstract). Preager, New York, p 49

Clausen H, Levery SB, McKibbin JM, Hakomori S-i (1985a) Blood group A determinants with mono-and di-fucosyl type I chain in human erythrocyte membranes. Biochemistry 24:3578–3586

Clausen H, Levery S, Nudelman E, Tsuchia S, Hakomori S-i (1985b) Repetitive A epitope (type 3 chain A) defined by blood group A1-specific monoclonal antibody TH-1: chemical basis of qualitative A1 and A2 distinction. Proc Natl Acad Sci USA 82:1199–1203

Dubois C, Magnani JL, Grunwald GB, Spitalnik SL, Trisler GD, Nirenberg M, Ginsburg Y (1986) Monoclonal antibody 18B8, which detects synapse-associated antigens, binds to ganglioside GT3 ($II^3(NeuNAc)_3LacCer$). J Biol Chem 261:3826–3830

Ernst C, Atkinson B, Wysocka M, Blaszczyk M, Herlyn M, Sears H, Steplewski Z, Koprowski H (1984a) Monoclonal antibody localization of Lewis antigens in fixed tissue. Lab Invest 50:394–400

Ernst C, Thurin J, Atkinson B, Wurzel H, Herlyn M, Strömberg N, Civin C, Koprowski H (1984b) Monoclonal antibody localization of A and B isoantigens in normal and malignant fixed human tissues. Am J Pathol 117:451–461

Falk K-E, Karlsson K-A, Larson G, Thurin J, Blaszczyk M, Steplewski Z, Koprowski H (1983) Mass spectrometry of a human glycolipid antigen being defined by mouse monoclonal antibody NS-19-9. Biochem Biophys Res Commun 110:383–391

Feizi T (1985) Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. Nature 314:53–57

Fredman P, Richert ND, Magnani JL, Willingham MC, Pastan I, Ginsburg V (1983) A monoclonal antibody that precipitates the glycoprotein receptor for epidermal growth factor is directed against the human blood group H type 1 antigen. J Biol Chem 258:11206–11210

Fredman P, Jeansson S, Lycke E, Svennerholm L (1985) A monoclonal antibody reacting specificially with ganglioside GD1b in human brain. FEBS Lett 189:23–26

Fredman P, Brezicka T, Holmgren J, Lindholm L, Nilsson O, Svennerholm L (1986) Binding specificity of monoclonal antibodies to ganglioside Fuc-GMI. Biochim Biophys Acta 875:316–323

Fukushi Y, Hakomori S-i, Nudelman E, Cochran N (1984a) Novel fucolipids accumulating in human adenocarcinoma. II. Selective isolation of hybridoma antibodies that differentially recognize mono-, di-, and tri-fucosylated type 2 chain. J Biol Chem 259:4681–4685

Fukushi Y, Nudelman E, Levery SB, Rauvala H, Hakomori S-i (1984b) Novel fucolipids accumulation in human cancer. III. A Hybridoma antibody (FH6) defining a human cancer-associated difucoganglioside ($VI^3NeuNAcY^3III^3Fuc_2nLc_6$). J Biol Chem 259:10511–10517

Fukushima K, Hirota M, Terasaki PI, Wakisaka A, Togashi H, Chia D, Suyama N, Fukushi Y, Nudelman E, Hakomori S-i (1984) Characterization of sialylated $Lewis^x$ as a new tumor associated antigen. Cancer Res 44:5279–5285

Furukawa K, Mattes MJ, Lloyd KO (1985) A1 and A2 erythrocytes can be distinguished by reagents that do not detect structural differences between the two cell types. J Immunol 135:4090–4094

Galanos C, Luderitz O, Westphal O (1971) Preparation and properties of antisera against the lipid-A component of bacterial lipopolysaccharides. Eur J Biochem 24:116–122

Ginsburg V, Fredman P, Magnani JL (1984) Cancer-associated carbohydrate antigens detected by monoclonal antibodies. Contrib Oncol 19:44–50

Ginsburg V, Magnani JM, Spitalnik SL, Spitalnik PF, Roberts DO, Dubois C (1985) Carbohydrate antigens detected by monoclonal antibodies. In: Proceedings of the VIII international symposium on glycocojugates. Preager, New York, pp 550–551

Gooi HC, Feizi T, Kapadia A, Knowles BB, Solter D, Evans HJ (1981) Stage specific embryonic antigen involves $\alpha 1 \rightarrow 3$ fucosylated type 2 blood group chains. Nature 292:156–158

Gooi HC, Williams LK, Uemura K, Hounsell EF, Mcllinney RAJ, Feizi T (1983) A marker of human endoderm defined by a monoclonal antibody involves type 1 blood groups chains. Mol Immunol 30:607–613

Goujet-Zalc C, Guerci A, Dubois G, Zalc B (1986) Schwann cell marker defined by a monoclonal antibody (224–58) with species cross-reactivity II. Molecular characterization of the epitope. J Neurochem 46:435–439

Hakomori S-i (1981) Blood group ABH and Ii antigens of human erythrocytes: chemistry, polymorphism and their developmental change. Semin Hematol 18:39–62

Hakomori S-i (1984) Monoclonal antibodies directed to cell-surface carbohydrates. In: Monoclonal antibodies and functional cell lines. Plenum, New York, pp 67–100

Hakomori S-i (1985) Aberrant glycosylation in cancer cell membranes as focused on glycolipids: overview and perspectives. Cancer Res 45:2405–2414

Hakomori S-i, Jeanloz RW (1964) Isolation of a glycolipid containing fucose, glucose, and glucosamine from human cancerous tissue. J Biol Chem 239:3606–3607

Hakomori S-i, Kanagi R (1983) Glycosphingolipids as tumor-associated and differentiation markers. J Natl Cancer Inst 71:231–251

Hakomori S-i, Patterson CM, Nudelman E, Seikguchi K (1983b) A monoclonal antibody directed to *N*-acetylneuraminosyl $\alpha 2 \rightarrow 6$ galactosyl residue in gangliosides and glycoproteins. J Biol Chem 258:11819–11822

Hanisch F-G, Uhlenbruck G, Dienst C (1984) Structure of tumor-associated carbohydrate antigen CA 19-9 on human seminal-plasma glycoproteins from healthy donors. Eur J Biochem 144:467–474

Hansson GC, Zopf D (1985) Biosynthesis of the cancer-associated sialyl-Le^a antigen. J Biol Chem 260:9388–9392

Hansson GC, Karlsson K-A, Larson G, McKibbin JM, Blaszczyk M, Herlyn M, Steplewski Z,

Koprowski H (1983) Mouse monoclonal antibodies against human cancer cell lines with specificities for blood group and related antigens. J Biol Chem 258:4091–4097

Hansson GC, Karlsson K-A, Thurin J (1984) Glycosphingolipid patterns of the epithelial and non-epithelial compartments of rat large intestine. Biochim Biophys Acta 792:281–292

Herlyn M, Koprowski H (1985) New tumor markers in human gastrointestinal cancer. In: Clinical laboratory molecular analyses. Grune and Stratton, New York

Herlyn M, Blaszczyk M, Benicelli J, Sears HF, Ernst C, Ross AH, Koprowski H (1985) Selection of monoclonal antibodies detecting serodiagnostic human tumor markers. J Immunol Methods 80:107–116

Herlyn M, Ross AH, Balaban G, Herlyn D, Thurin J, Koprowski H (to be published) Cell biology of melanoma and non-malignant lesions. In: Pigment cell: pathobiology of melanoma. Karger, Basel

Hindsgaul O, Norberg T, LePendu J, Lemieux RU (1982) Synthesis of type 2 human blood group antigenic determinants. The H, X, and Y haptens and variations of the H type 2 determinants as probes for the combining site of the lectin I of *Ulex europaeus*. Carbohydr Res 109:109–142

Hindsgaul O, Khare DP, Bach M, Lemieux RU (1985) Molecular recognition. III. The binding of the H-type 2 human blood group determinant by the lectin I of *Ulex europeus* 63:2653–2658

Hirabayashi Y, Hamaoka A, Matsumoto M, Matsubara T, Tagawa M, Wakabayashi S, Taniguchi M (1985) Syngeneic monoclonal antibody against melanoma antigen with interspecies cross-reactivity recognizes GM3, a prominent ganglioside of B 16 melanoma. J Biol Chem 260:13328–13333

Hirohashi S, Clause H, Yamada T, Shimosato Y, Hakomori S-i (1985) Blood group A cross-reacting epitope defined by monoclonal antibodies NCC-LU-35 and -81 expressed in cancer of blood group 0 or B individuals: its identification as Tn antigen. Proc Natl Acad Sci USA 82:7039–7043

Holmes EH, Ostrander GK, Hakomori S-i (1986) Biosynthesis of the sialyl-Lex determinant carried by type 2 chain glycosphingolipids ($IV^3NeuNAcIII^3FucnLc_4$, $VI^3NeuNAcV^3FucnLc_6$, and $VI^3NeuNAcIII^3V^3Fuc_2nLc_6$) in human lung carcinoma PC9 cells. J Biol Chem 261:3737–3743

Houghton AN, Mintzer D, Cordon-Cardo C, Welt S, Fliegel B, Vadhan S, Carswell E, Melamed MR, Oettgen HF, Old LJ (1985) Mouse monoclonal IgG3 antibody detecting GD3 ganglioside: a phase I trial in patients with malignant melanoma. Proc Natl Acad Sci USA 82:1242–1246

Hounsell EF, Gooi HC, Feizi T (1981) The monoclonal antibody anti-SSEA-1 discriminates between fucosylated type 1 and type 2 blood group chains. FEBS Lett 131:279–282

Huang LC, Civin CK, Magnani JL, Shaper JH, Ginsburg V (1983) My-1, the human myeloid specific antigen detected by mouse monoclonal antibodies, is a sugar sequence found in lacto-*N*-fucopentaose III. Blood 61:1020–1023

IUPAC-IUB Commission on Biochemical Nomenclature (1977) The nomenclature of lipids. Biochem J 171:21–35

Ito M, Suzuki E, Naiki M, Sendo F, Arai S (1984) Carbohydrates as antigenic determinants of tumor-associated antigens recognized by monoclonal anti-tumor antibodies produced in a syngeneic system. Int J Cancer 34:689–697

Kannagi R, Cochran N, Ishigami F, Hakomori S-i, Andrews P, Knowles BB, Solter D (1983a) Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells. EMBO J 2:2355–2361

Kanagi R, Levery SB, Ishigami F, Hakomori S-i, Shevinsky LH, Knowles BB, Solter D (1983b) New globoseries glycosphingolipids in human teratocarcinoma reactive with the monoclonal antibody directed to a developmentally regulated antigen, stage-specific embryonic antigen 3. J Biol Chem 258:8934–8942

Kimmel KA, Carey TE, Judd WJ, McClatchkey KD (1986) Monoclonal antibody (G10) to a common antigen of human squamous cell carcinoma: binding of the antibody to the H type 2 blood group determinant 76:9–19

Koprowski H (1984) The Gordon Wilson lecture: monoclonal antibodies in human cancer. Trans Am Clin Clinat Assoc 96:81–97

Koprowski H, Steplewski Z, Herlyn D, Herlyn M (1978) Study of antibodies against human melanoma produced by somatic cell hybrids. Proc Natl Acad Sci USA 75:3405–3409

Koprowski H, Steplewski Z, Mitchell K, Herlyn M, Herlyn D, Fuhrer P (1979) Colorectal carcinoma antigens detected by hybridoma antibodies. Somatic Cell Genet 5:957–972

Kumazaki T, Yoshida A (1984) Biochemical evidence that secretor gene, Se, is a structural gene encoding a specific fucosyltransferase Proc Natl Acad Sci USA 81:4193–4197

Lemieux RU (1982) The binding of carbohydrate structures with antibodies and lectins. In: IUPAC, frontiers in chemistry. Pergamon, Oxford, pp 3–24

Lemieux RU, Wong TC, Liao J, Kabat EA (1984) The combining site of anti-I Ma (group 1). Mol Immunol 21:751–759

Lemieux RU (1985) How oligosaccharides are recognized for specific noncovalent binding by antibodies and lectins. In: Proceedings of the VIIIth international symposium on glycoconjugates. Praeger, New York, p 177

Lemieux RU, Venot AP, Spohr U, Bird P, Mandal G, Morishima N, Hindsgaul O, Bundle DR (1985) Molecular recognition. V. The binding of the B human blood group determinant by hybridoma monoclonal antibodies 63:2664–2668

LePendu J, Fredman P, Richter N, Magnani JL, Willingham MC, Pastan I, Oriol R, Ginsburg V (1985) Monoclonal antibody 101 that precipitates the glycoprotein receptor for epidermal growth factor is directed against the Y-antigen, not the H type 1 antigen. Carbohydr Res 141:347–349

Lindholm L, Holmgren J, Svennerholm L, Fredman P, Nilsson O, Persson B, Myrvold H, Lagergård T (1983) Monoclonal antibodies against gastrointestinal tumor-associated antigens isolated as monosialogangliosides. Int Arch Allergy Appl Immunol 71:178–181

Lloyd KO, Larson G, Strömberg N, Thurin J, Karlsson K-A (1983) Mouse monoclonal antibody F-3 recognizes the difucosyl type-2 blood group structure. Immunogenetics 17:537–541

Magnani JL, Brockhaus M, Smith DF, Ginsburg V, Blaszczyk M, Mitchell KF, Steplewski Z, Koprowski H (1981) A monosialo-ganglioside is a monoclonal antibody-defined antigen in colon carcinoma. Science 212:55–56

Magnani JL, Nilsson B, Brockhaus M, Zopf D, Steplewski Z, Koprowski H, Ginsburg V (1982) A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-*N*-fucopentaose II. J Biol Chem 257:14365–14369

Magnani JL, Steplewski Z, Koprowski H, Ginsburg V (1983) Identification of the gastrointestinal cancer-associated antigen defined by monoclonal antibody 19-9 in the sera of patients as a mucin. Cancer Res 43:5489–5492

Magnani JL, Ball ED, Fanger MW, Hakomori S-i, Ginsburg V (1984) Monoclonal antibodies PMN 6, PMN 29, and PM-81 bind differently to glycolipids containing a sugar sequence occurring in lacto-*N*-fucopentaose III. Arch Biochem Biophys 233:501–506

Månsson J-E, Fredman P, Nilsson O, Lindholm L, Holmgren J, Svennerholm L (1985) Chemical structure of carcinoma ganglioside antigens defined by monoclonal antibody CA-50 and some allied gangliosides from human pancreatic adenocarcinoma. Biochim Biophys Acta 834:110–117

Nilsson O, Månsson J-E, Lindholm L, Holmgren J, Svennerholm L (1985) Sialosyllactotetraosylceramide, a novel ganglioside antigen detected in human carcinomas by a monoclonal antibody. FEBS Lett 182:398–402

Nowinski R, Berglund C, Lane J, Lostrom M, Bernstein I, Young WW, Hakomori S-i, Hill L, Cooney M (1980) Human monoclonal antibody against Forssman antigen. Science 210:537–539

Nudelman E, Hakomori S-i, Knowles BB, Solter D, Nowinski RC, Tam MR, Young WW (1980) Monoclonal antibody directed to the stage-specific embryonic antigen (SSEA-1) reacts with a branched glycosphingolipid similar in structure to Ii antigen. Biochem Biophys Res Commun 97:443–451

Nudelman E, Hakomori S-i, Kannagi R, Levery S, Yeh M-Y, Hellström K-E, Hellström I (1982) Characterization of a human melanoma-associated antigen defined by a monoclonal antibody, 4.2. J Biol Chem 257:12752–12756

Nudelman E, Kannagi R, Hakomori S-i, Parsons M, Lipinski M, Wiels J, Fellous M, Tursz T (1983) A glycolipid antigen associated with Burkitt lymphoma defined by a monoclonal antibody. Science 220:509–511

Olding L, Thurin J, Svalander C, Koprowski H (1984) Expression of gastrointestinal carcinoma-associated antigen (GICA) detected in human fetal tissues by monoclonal antibody NS-19-9. Int J Cancer 34:187–192

Olding L, Åhren C, Thurin J, Karlsson K-A, Svalander C, Koprowski H (1985) Gastrointestinal carcinoma-associated antigen detected by a monoclonal antibody in dysplasia and adenocacinoma associated with chronic ulcerative colites. Int J Cancer 36:131–136

Portoucalian J, Zwingelstein G, Doré J-F (1979) Lipid composition of human malignant melanoma tumors at various levels of malignant growth. Eur J Biochem 94:19–23

Pukel CS, Lloyd KO, Travassos LR, Dippold WG, Oettgen HF, Old LJ (1982) GD3, a prominent ganglioside of human melanoma. J Exp Med 155:1133–1147

Raux H, Labbe F, Fondeneche M-C, Koprowski H (1983) A study of gastrointestinal cancer-associated antigen (GICA) in human fetal organs. Int J Cancer 34:315–319

Rodeck U, et al. (to be published)

Saito M, Yu RK, Cheung N-KV (1985) Ganglioside GD2 specificity of monoclonal antibodies to human neuroblastoma cell. Biochem Biophys Res Commun 127:1–7

Sakamoto J, Furukawa K, Cordon-Cardo C, Yin BWT, Rettig WJ, Oettgen HF, Old LJ, Lloyd KO (1986) Expression of Lewis a, Lewis b, X, and Y blood group antigens in human colonic tumors and normal tissue and in human tumor derived cell lines. Cancer Res 46:1553–1561

Siddiqui B, Buehler J, DeGregorio MW, Macher BA (1984) Differential expression of ganglioside GD3 by human leukocytes and leukemia cells. Cancer Res 44:5262–5265

Solter D, Knowles BB (1978) Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). Proc Natl Acad Sci USA 75:5565–5569

Spohr U, Hindsgaul O, Lemieux RU (1985a) Molecular recognition. II. The binding of the Lewis b and Y human blood group determinants by the lectin iV of *Griffonia simplicifolia*. Can J Chem 63:2644–2652

Spohr U, Morishima N, Hindsgaul O, Lemieux RU (1985b) Molecular recognition. IV. The binding of the Lewis b human blood group determinant by a hybridoma monoclonal antibody. Can J Chem 63:2659–2663

Steplewski Z, Spira G, Blaszczyk M, Lubeck MD, Radbruch A, Illges H, Herlyn D, Rajewski K, Scharff M (1985) Isolation and characterization of anti-monosialoganglioside monoclonal antibody 19-9 class-switch variants. Proc Natl Acad Sci USA 82:8653–8657

Symington FW, Bernstein ID, Hakomori S-i (1984a) Monoclonal antibody specific for lactosylceramide. J Biol Chem 259:6008–6012

Symington FW, Fenderson BA, Hakomori S-i (1984b) Fine specificity of a monoclonal antitesticular cell antibody for glycolipids with terminal *N*-acetyl-D-glucosamine structure. Mol Immunol 21:877–882

Szulman AE (1960) The histological distribution of blood group substances A and B in man. J Exp Med 111:785–799

Szulman AE (1964) The histological distribution of blood group substances in man as disclosed by immunofluorescence. J Exp Med 119:503–516

Tai T, Paulson JC, Cahan LD, Irie RF (1983) Ganglioside GM2 as a human tumor antigen (OFA-I-1). Proc Natl Acad Sci USA 80:5392–5396

Tai T, Cahan LD, Tsuchida T, Saxton RE, Irie R, Morton DL (1985) Immunogenicity of melanoma-associated gangliosides in cancer patients. Int J Cancer 35:607–612

Thurin J, Herlyn M, Hindsgaul O, Strömberg N, Karlsson K-A, Elder D, Steplewski Z, Koprowski H (1985) Proton NMR and fast-atom bombardment mass spectrometry analysis of the melanoma-associated ganglioside 9-*O*-acetyl-GD3. J Biol Chem 260:14556–14563

Thurin J, Thurin M, Herlyn M, Elder DE, Clark WH, Steplewski Z, Koprowski H (to be published) GD2 ganglioside biosynthesis is a distinct biochemical event in human melanoma tumor progression

Thurin M, Thurin J, Hindsgaul O, Karlsson K-A, Steplewski Z, Koprowski H (to be published) Y and blood group B type 2 glycolipid antigens accumulate in a human gastric carcinoma cell line as detected by monoclonal antibody: Isolation and characterization by mass spectrometry and NMR spectroscopy

Thurin J, Thurin M, Kimoto Y, Lubeck M, Herlyn M, Elder DE, Smereszynska M, Karlsson K-A, Clark WH, Steplewski Z, Koprowski H (to be published) Antimelanoma antibody, directed against both antigens GD2 and GD3, indicates the significance of multiple specificity towards tumor-associated antigens in anti-tumor activity

Uemura K-i, Macher BA, DeGregorio M, Scudder P, Beuhler J, Knapp W, Feizi T (1985) Glycosphingolipid carriers of carbohydrate antigens of human myeloid cells recognized by monoclonal antibodies. Biochim Biophys Acta 846:26–36

Uhlenbruck G, van Meensel-Maene U, Hanish F-G, Dienst C (1984) Unexpected occurrence of

the CA 19-9 tumor marker in normal human seminal plasma. Hoppe-Seylers Z Physiol Chem 365:613–617

Urdal DL, Brentnall TA, Bernstein ID, Hakomori S-i (1983) A granulocyte reactive monoclonal antibody, IG10, identifies the Galβ1 → 4GlcNAc(3 ← 1αFuc)-(X determinant) expressed in HL-60 cells on both glycolipid and glycoprotein molecules. Blood 62:1022–1026

Watanabe T, Pukel CS, Takeyama H, Lloyd KO, Shiku H, Li LTC, Travassos LR, Oettgen HF, Old LJ (1982) Human melanoma antigen AH is an autoantigenic ganglioside related to GD2. J Exp Med 156:1884–1889

Watkins WM (1980) Biochemistry and genetics of the AB0, Lewis and P blood group systems. Adv Hum Genet 10:1–136

Willison KR, Karol RA, Suzuki A, Kundu SK, Marcus DM (1982) Neutral glycolipid antigens as developmental markers of mouse teratocarcinoma and early embryos: and immunologic and chemical analysis. J Immunol 129:603–609

Yamamoto M, Boyer AM, Schwarting GA (1985) Fucose-containing glycolipids are stage- and region-specific antigens in developing embryonic brain of rodents. Proc Natl Acad Sci USA 82:3045–3049

Yeh M-Y, Hellström I, Abe K, Hakomori S-i, Hellström K-E (1982) A cell-surface antigen which is present in the ganglioside fraction and shared by human melanomas. Int J Cancer 29:269–275

Young WW, Hakomori S-i (1981a) Therapy of mouse lymphoma with monoclonal antibodies to glycolipid: selection of low antigenic variants in vivo. Science 211:487–489

Young WW, Johnson HS, Tamura Y, Karlsson K-A, Larson G, Paker JMR, Khare DP, Spohr U, Baker DA, Hindsgaul O, Lemieux RU (1983) Characterization of monoclonal antibodies specific for the Lewis-a human blood group determinant. J Biol Chem 258:4890–4894

Young WW, MacDonald EMS, Nowinski RC, Hakomori S-i (1979) Production of monoclonal antibodies specific for two distinct steric portions of the glycolipids ganglio-*N*-tertaosylceramide (asialo-GM2). J Exp Med 150:1008–1019

Young WW, Portoukalian J, Hakomori S-i (1981b) Two monoclonal anticarbohydrate antibodies directed to glycosphingolipids with a lacto-*N*-glycosyl type II chain. J Biol Chem 256:10967–10972

Protein–Oligosaccharide Interactions: Lysozyme, Phosphorylase, Amylases

L.N. JOHNSON, J. CHEETHAM, P.J. MCLAUGHLIN, K.R. ACHARYA, D. BARFORD, and D.C. PHILLIPS

1 Introduction

Polysaccharides have many diverse biological roles. They comprise the insoluble structural and supportive elements of bacterial and plant cell walls. They serve

Laboratory of Molecular Biophysics, Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK

Current Topics in Microbiology and Immunology, Vol. 139

as storage macromolecules (e.g., glycogen and starch) to provide fuel for cells, and they form important components in the heteromacromolecules, proteoglycans and glycoproteins, which are involved in the structural elements of connective tissue, in the lubrication of skeletal joints, in cell–cell adhesion and in cell–surface recognition. The enzymes that catalyze the biosynthesis and degradation of these polysaccharides are highly specific, resulting, in the case of biosynthesis, in polymers of defined constitution and sequence and, in the case of degradation in the cleavage of specific linkages. In this article we describe the X-ray structural results for three of these enzymes: lysozyme, glycogen phosphorylase, and amylase. Each catalyzes a specific step in the degradation of different polysaccharides. (To date there are no structural data on any of the biosynthetic enzymes.) Lysozyme catalyzes the hydrolysis of a β-(1–4) linkage of bacterial cell wall polysaccharides; glycogen phosphorylase catalyzes the reversible phosphorylation of α-(1–4) linkages of glycogen and amylase catalyzes the hydrolysis of α-(1–4) linkages in starch. We use these results in an attempt to draw together some general principles that appear to be important in generating a protein–oligosaccharide recognition site and compare these features with protein monosaccharide sites.

Much has been written on protein–protein interactions and protein–nucleic acid interactions. The stereochemical features of protein–oligosaccharide interactions have received less attention. Are these sites associated with particular elements of secondary structure? Are they buried or exposed, and how do they complement both the polar and nonpolar components of the sugars? How is specificity achieved? Can we predict such sites? The studies with lysozyme and phosphorylase provide the answers to some of these questions, especially with regard to molecular recognition. These conclusions are summarized at the end of the article after a detailed discussion of the individual enzyme systems.

Oligosaccharides contain a mixture of polar and nonpolar groups of atoms. This is illustrated in Fig. 1 for α-(1–4) and β-(1–4) linked glucosyl polymers. For both α- and β-linked glucosyl polymers there are polar groups comprising the hydroxyl (or acetamido group in the case of lysozyme substrates) at position 2 and hydroxyl groups at positions 3 and 6. These polar groups together with the ring oxygen and the C1–O4 bond are located approximately in the plane of the glucopyranose ring, and from a peripheral polar surface. The glycosidic linkage is more nearly normal to the sugar ring for α-linked and equatorial for β-linked glucopyranose sugars. When viewed in a direction normal to the plane of the glucopyranose rings, the surfaces are nonpolar and have different characteristics for α-linked and β-linked sugars. Let us define these surfaces as the A face and the B face for instances in which the numbering of the carbon atoms increases in a clockwise and anticlockwise direction, respectively.

For α-(1–4) linked sugars the preferred geometry (discussed in Sect. 3.4) generates a helical structure in which there is a twist of about $-55°$ between adjacent sugars and a hydrogen bond between the O2 hydroxyl of one sugar and the O3′ hydroxyl of an adjacent sugar (Fig. 1). The CH bonds at positions 1, 2, and 4 are located on the A face. The C1–H, C2–H and C4′-H atoms form a nonpolar cluster (shown hatched in Fig. 1). The C–H atoms at positions 3 and 5 protrude from the B face, but the nonpolar character of this face

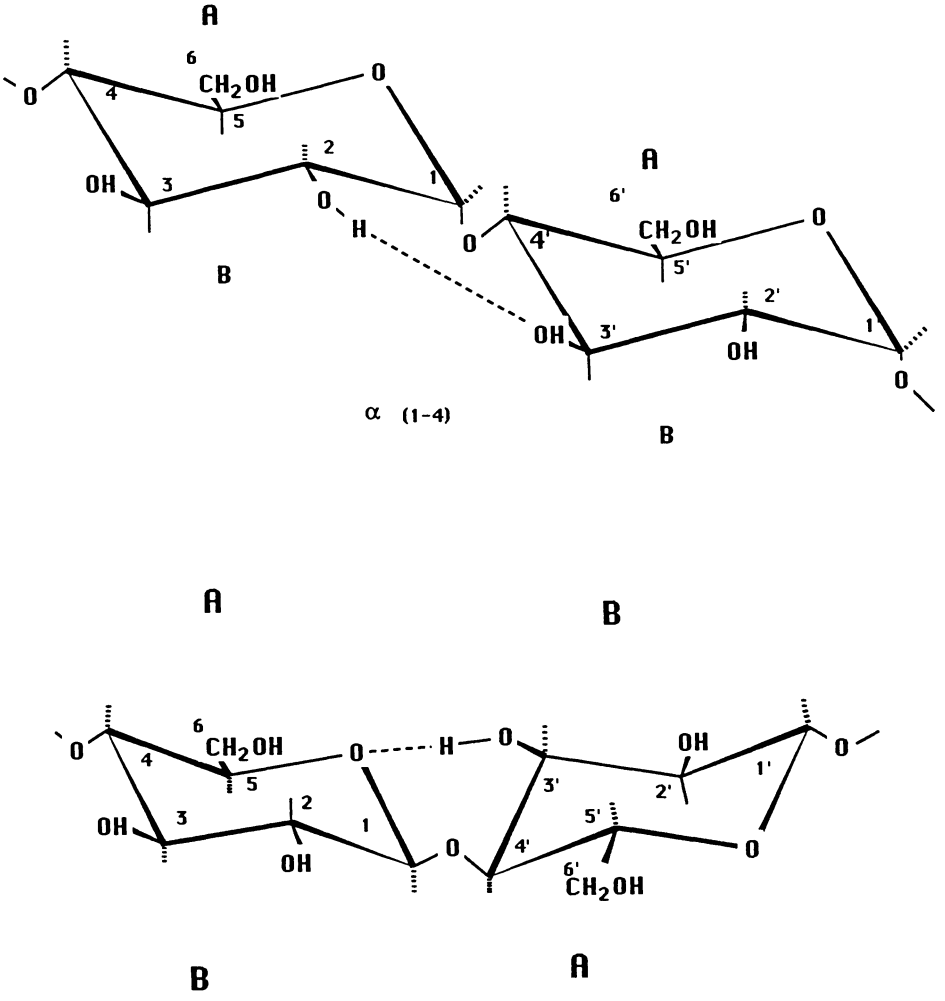


Fig. 1. The preferred conformation for α-(1–4) and β-(1–4) linked glycosyl polymers. The A and B faces are defined as those for which the numbering system is clockwise and anticlockwise, respectively. The CH bonds define the nonpolar faces (*hatched*)

is interrupted by the glycosidic oxygen. Thus α-(1–4) linked glucosyl polymers present a nonpolar A face to the protein, which repeats about every 5 Å with a twist of about −55° between adjacent sugars.

For β-(1–4) linked glucosyl polymers the preferred geometry generates a ribbonlike molecule with a twist of about 180° between adjacent sugars and a hydrogen bond between the ring oxygen of one sugar and the O3 hydroxyl of the adjacent sugar (Fig. 1). Thus, the oligosaccharide presents alternately

an A face and a B face to the surface. The CH bonds at positions 1, 3, and 5 are located on the B face. The β-glycoside oxygen is directed away from this face. The 180° twist of the adjacent sugar allows the CH bond at position 4′ from the A face of the adjacent sugar to contribute to the nonpolar surface already generated by the B face. The disposition of the ring oxygen and the glycosidic oxygen means that the remainder of the A face has a more polar character. Thus, β-(1–4) linked glucosyl polymers present a nonpolar B face that repeats about every 5 Å with a twist of 180° between adjacent sugars.

These polar and nonpolar surfaces are important for oligosaccharide recognition.

2 Lysozyme

2.1 Introduction

The crystallographic studies on the hen egg-white form of lysozyme (HEWL; FLEMING 1922), begun at the Royal Institution in the early 1960s, culminated in the first determination of the three-dimensional structure of an enzyme (BLAKE et al. 1965, 1967a). When combined with the results of similar investigations of several of the inhibitor complexes of the enzyme (JOHNSON and PHILLIPS 1965; BLAKE et al. 1967b), most importantly with the β-(1–4) linked trisaccharide of *N*-acetylglucosamine [$(GlcNAc)_3$] such studies provided the first real insight into the nature of the active site of an enzyme. In the 22 years since publication of this early work the enzyme has proved a model system for the study of many aspects of protein structure and function. The ease of crystallization of the native protein in the tetragonal from (STEINRAUF 1959) and the relatively robust nature of the resultant crystals with respect to irradiation by X rays and to the diffusion of inhibitor or substrate molecules have resulted in a large number of binding studies to characterize the nature of the interactions between the protein and ligand oligosaccharides (BLAKE et al. 1967b; FORD et al. 1974; KELLY et al. 1979, CHEETHAM et al. 1987) and to examine the possible mechanistic pathway for the action of the enzyme (PHILLIPS 1966, 1986). These studies have been carried out in both the solid and solution phases, the former involving X-ray crystallographic techniques while the latter has seen the application of many spectroscopic techniques, e.g., nuclear magnetic resonance (DOBSON 1975; CASSELS 1979; CHEETHAM 1986), ultraviolet spectroscopy (HOLLER et al. 1975), and fluorescence spectroscopy (LEHRER and FASMAN 1966).

In this review we focus on the results obtained from crystallographic studies on a number of different types of lysozyme in association with a variety of oligosaccharide substrates, ranging from the simple monosaccharide *N*-acetylglucosamine (GlcNAc) (JOHNSON and PHILLIPS 1965), on the one hand, to the tetramer transition-state analogue tetra-NAG lactone (FORD et al. 1974), on the other. In particular, the detailed nature of the protein–carbohydrate interactions, occurring within these enzyme–inhibitor complexes, is examined in the

light of recent experimental results from a study at high resolution (1.75 Å) of the trisaccharide complex of HEWL (CHEETHAM et al. 1987).

The large amount of structural data from studies at lower resolution of the saccharide complexes of lysozymes derived from other sources, namely human lysozyme (HuL; BANYARD 1973), tortoise egg-white lysozyme (TEWL; GALLAY 1984), goose egg-white lysozyme (GEWL; WEAVER and MATTHEWS, unpublished results), and T4 bacteriophage lysozyme (T4L; GRUTTER and MATTHEWS 1982), allows one to compare the modes of binding in the different systems and, in particular, to examine the features within the active-site cleft that have been conserved among the different species during the course of their evolution and may hence be regarded als crucial to their function.

2.1.1 Lysozyme: Definition and Substrate Specificity

The name lysozyme is given to a wide variety of different enzymes distributed throughout the animal and plant kingdoms, occurring in the tissues and secretions of many vertebrates and invertebrates as well as in bacteria and bacteriophages (JOLLES and JOLLES 1984). These enzymes are defined as 1,4-β-*N*-acetylmuramidases and act by cleaving the mucopolysaccharide chains of the cell walls of Gram-positive bacteria (SALTON 1952). The long copolymer chains of the cell walls consist of alternating *N*-acetylmuramic (MurNAc) and GlcNAc units (JEANLOZ et al. 1963), linked by a β-1,4 glycosidic bond to produce a ribbonlike arrangement of the saccharide units; the preferred β conformation of these (1–4) polymers was recognized from the crystal structure determination of cellobiose (JACOBSON et al. 1961). The lysozyme acts by cleaving the copolymer chains specifically between the 1,4-β linkages of adjacent MurNAc-GlcNAc residues (SALTON and GHUYSEN 1959, 1960).

The most extensively studied of all the different lysozymes are those derived from avian egg-white. There are two distinct types of enzyme from this source (PRAGER et al. 1974): c type (e.g., HEWL) and g type (e.g., GEWL). The c-type lysozymes are normally distinguished by their lower molecular weights. Arguably, the greatest wealth of structural information on these enzymes has come from crystallographic studies on HEWL.

2.1.2 Structure of Hen Egg-White Lysozyme

The study of BLAKE et al. (1965) (2 Å) showed the lysozyme molecule to be roughly ellipsoidal in shape. ($45 \times 30 \times 30$ Å^3), with a deep cleft (subsequently identified as the active site) running across the surface and dividing the molecule into two distinct lobes within which both α-helices and antiparallel β-pleated sheet were identified (BLAKE et al. 1967a). The picture of the molecule that emerged from these studies has been still further refined over the years. Application of the restrained least-squares refinement procedure of HENDRICKSON and KONNERT (1980) has now reduced the R factor for the model to 17% (HANDOLL et al. 1987) and, most significantly, allowed a detailed analysis of hydrogen-bonding patterns with the protein, description of the water structure, and assignment of individual isotropic temperature factors to all the atoms of each residue.

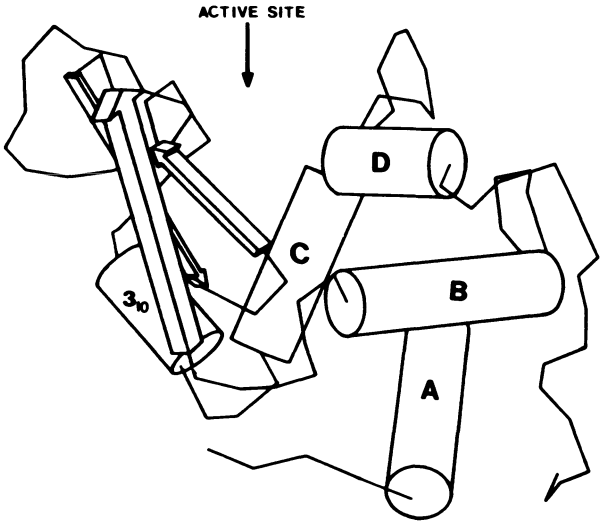


Fig. 2. Schematic representation of the native HEWL molecule. The α- and 3_{10}-type helices are shown as *cylinders*; β-strands are represented by *arrows*. The position of the active-site cleft, in which oligosaccharides are observed to bind to the enzyme, is indicated

Concurrent with this work a crystallographic study of a high-pH form of HEWL, also tetragonal, has been made at 1.6-Å resolution (HANDOLL 1985). This study has shown the low-pH (4.7) and high-pH (7.5) forms to be structurally very similar; the study at 1.6 Å thus provides the most detailed and complete description of the HEWL molecule in the tetragonal crystal form. The structure of triclinic lysozyme has been determined at 1.4 Å (MASON et al. 1984).

The main-chain conformation of the molecule (Fig. 2) is little changed from that described in 1967. The molecule comprises five helices four α-helices (labeled A–D) formed by residues 4–16, 24–34, 88–96, and 108–113 and one 3_{10} helix from residues 80–85. The A–C helices surround a hydrophobic core of the large aromatic side chains of Trp 28, 108, and 111, Tyr 23, and Met 105; the A and C helices lie on the protein surface, partially exposed to solvent, while the B helix is completely buried. Residues 41–79 form a smaller, mainly hydrophilic region containing a three-stranded, antiparallel β-sheet (41–60) connected to the hydrophobic core by a long coiled loop running from residues 60 to 79. Four disulfide bridges connect residues 6 and 127, 30 and 115, 64 and 80, and 76 and 94. The analysis of the isotropic temperature factors broadly concurs with that in the earlier work at 2.0 Å on the low-pH form, giving the highest B values to residues at the lips of the active-site region and, in addition, allows a more precise description of the mobility of different regions of the protein than was possible at 2.0 Å.

2.1.3 The Active Site and the Binding of Oligosaccharides

The active site of the protein contains six subsites, A to F, each of which can accommodate a single saccharide unit. Within this region two carboxylic acids, glutamic acid 35 and aspartic acid 52 (separated by less than 6 Å), were identified as the catalytic residues, and from their environments in the crystal

Table 1. Crystallographic studies of saccharide binding to lysozyme

Lysozyme type (crystal form)	Saccharide ligand	Binding site(s)	Resolution of study	References
Hen egg white (tetragonal)	NAG (α and β anomers)	C	6.0 Å and 2.0 Å	1, 2, 3, 4, 5
	NAM	C	6.0 Å	2, 4
	6-Iodo α-methyl NAG	C	6.0 Å	1, 2, 4
	α-Benzyl NAM	Not bound in active site	6.0 Å	1
	Glucosamine	Nonspecific	6.0 Å	1
	Muramic acid	Nonspecific	6.0 Å	1
	Di-NAG (α) (terminal NAG in α-form)	Terminal NAG residue bound as α-NAG at C; second sugar lies out of run of main cleft	6.0 Å	1, 2
	Di-NAG (β) (terminal NAG in β-form)	B and C	6.0 Å	1, 2
	NAG-NAM	One sugar ring binds at C, the other lies out of main run of cleft, as for di-NAG (α)	6.0 Å	2, 4
	(NAG-NAG-ø) phenyl-β-D-chitobioside	B and C	2.0 Å	3, 4
	NAG-glucose	C and D (second mode also observed in which second sugar moiety no longer occupies D, but lies out of cleft)	2.0 Å	3, 4
	NAG-xylose	C and D	2.0 Å	3, 4
	Tri-NAG	A, B, and C	6.0 Å; 2.0 Å and 1.75 Å	2, 4, 6, 7, 8
	NAM-NAG-NAM	B, C, and D	2.5 Å	9
	Tetra-NAG	A, B, C, and D	2.5 Å	10, 11
	Tetra-NAG lactone	A, B, C, and D	2.5 Å	11
	Glucono-(1,4)--lactone	D	2.0 Å	3, 4, 12
	N-Acetylglucos--amino-(1,5)--lactone	C	6.0 Å	3, 4, 12
Hen egg white (triclinic)	β-NAG	C	2.0 Å	13
	Di-NAG	B and C	2.0 Å	13
Iodine-inactivated hen egg white	Tri-NAG	B, C, and D	2.5 Å	14
Human	Tri-NAG	B, C, and D	2.5 Å	15, 16
	Tetra-NAG	A, B, C, and D	6.0 Å	16

Table 1. (continued)

Lysozyme type (crystal form)	Saccharide ligand	Binding site(s)	Resolution of study	References
Tortoise egg white	NAG	C	6.0 Å	17
	Tri-NAG	A, B, and C	6.0 Å	17
	$(NAG\text{-}NAM)_2$	A, B, C, and D	2.5 Å	18
Turkey egg white	NAG-NAM	C and D; E and F	2.8 Å	19
Goose egg white	Tri-NAG	Sites analogous to B, C and D in HEWL	2.8 Å	20
Bacteriophage T4	β-NAG	Site analogous to C in HEWL	2.4 Å	21
	α-NAG	No binding	2.4 Å	21
	NAM	Site analogous to C in HEWL	2.4 Å	21
	Tri-NAG	Sites analogous to A, B, and C in HEWL	2.4 Å	21

NAG, *N*-acetyl glucosamine (GlcNAc); NAM, *N*-acetyl muramic acid (MurNAc)

References: [1] JOHNSON and PHILLIPS 1965; [2] BLAKE et al. 1967b; [3] BEDDELL et al. 1970; [4] IMOTO et al. 1972; [5] PERKINS et al. 1979; [6] JOHNSON et al. 1968; [7] CHEETHAM et al. 1987; [8] Present review; [9] KELLY et al. 1979; [10] GRACE 1980; [11] FORD et al. 1974; [12] BEDDELL et al. 1970; [13] KURACHI et al. 1976; [14] BEDDELL et al. 1970; [15] BANYARD 1973; [16] OATLEY 1973; [17] LIDDINGTON 1981; [18] GALLAY 1984; [19] SARMA and BOTT 1977; [20] WEAVER and MATTHEWS, unpublished results reported in WEAVER et al. 1985; [21] ANDERSON et al. 1981

structure different ionic structures were proposed for each. The carboxyl group of Asp 52 lies in a generally polar environment and is likely to be ionized below pH 4. Glu 35, on the other hand, is in a predominantly nonpolar hydrophobic pocket and is likely to remain in the protonated state at higher pH. These observations formed the basis for the proposed hydrolysis of the β-(1–4) glycosidic linkage between the fourth (D) and fifth (E) saccharides in the substrate (PHILLIPS 1966). The mechanism accounts for the catalytic rate enhancement by the protein as a result of distortion associated with binding of the saccharide ring at site D, which promotes the formation of a carbonium ion, Glu 35, acting as a general acid catalyst and donating a proton to the glycosidic oxygen. The results are breaking of the C(1)–O glycosidic bond and development of a carbonium ion at C(1), which is stabilized in situ by Asp 52. The positive charge on the carbonium ion is shared to some extent with the neighboring ring oxygen O(5), leading to the formation of a carbonium-oxonium ion with a partial double bond between C(1) and O(5). This reaction scheme accounts for the specificity of enzymic hydrolysis (for the bond between the fourth and fifth sugar units from the nonreducing end, and the glycosidic linkage lying between sites D and E is the one cleaved. Ample biochemical information has substantiated the mechanism proposed by PHILLIPS (RUPLEY et al. 1967; IMOTO et al. 1972; BANERJEE et al. 1975), and the ionization states of the two carboxylic acid residues confirmed by the measurement of their macroscopic pKs (KURA-

MITSU et al. 1977). Recently it has proven possible to observe the ionization states of these two residues directly by the use of neutron diffraction (MASON et al. 1984).

At the time that the above mechanism for the action of lysozyme was elucidated structural information on the nature of the sites in the upper half of the binding cleft (A–C) had been obtained principally from the results of crystallographic studies with the competitive inhibitor $(GlcNAc)_3$ (BLAKE et al. 1967b; JOHNSON et al. 1968; IMOTO et al. 1972). Table 1 shows all the lysozyme–oligosaccharide complexes subjected to study by X-ray crystallographic methods. To data, no direct experimental observations of binding in sites E and F have been made on the HEWL form. With the tetragonal crystals, failure to observe binding in these lower sites is a direct result of the unfavorable packing arrangement of neighboring molecules in the crystal lattice, blocking sites E and F. Attempts to occupy sites E and F in the other known crystal forms of HEWL (triclinic, monoclinic, and orthorhombic), have also resulted in failure (MOULT et al. 1976; KURACHI et al. 1976; HOGLE et al. 1981; ARTYMIUK and RICE 1980). However, saccharide binding in the lower sites of the binding cleft has been observed in the turkey form of the enzyme (SARMA and BOTT 1977).

2.2 Crystallographic Studies of Oligosaccharide Binding to HEWL at Low and Medium Resolutions

2.2.1 The Binding of Monosaccharides to HEWL

In mucopolysaccharide cell walls the GlcNAc units are firmly joined by β-(1–4) linkages, but isolated molecules of GlcNAc in solution mutarotate, in the usual way, to produce an equilibrium mixture of the α and β forms. Initial observations at 6.0 Å suggested that the binding of GlcNAc [and $(GlcNAc)_2$] may be nonspecific (JOHNSON and PHILLIPS 1965), but structural studies at 2.0 Å (BLAKE et al. 1967b) showed the binding to be specific and also dependent upon the anomeric form of the sugar. The most specific interactions between the GlcNAc molecules and the enzyme are hydrogen bonds involving the NH and carbonyl oxygen groups of the *N*-acetyl side chain (which is essential for binding) and the main-chain CO and NH groups of protein residues 107 and 59, respectively. These interactions are essentially the same for both the α and β forms of the sugar. The remaining interactions, however, are different depending on which configuration of the GlcNAc molecule is bound. In the case of β-GlcNAc, hydrogen bonds are formed between the O(6) and O(3) oxygens of the sugar and the NH groups of the tryptophan side chains of residues 62 and 63. This mode of binding is almost identical to that observed for the sugar bound at site C in the trisaccharide complex (see Sect. 2.3) and in the structure with β-methyl *N*-acetylglucosaminide (PERKINS et al. 1979). α-GlcNAc, on the other hand, binds such that the O(1) oxygen of the sugar interacts with the main-chain NH of residue 109. In effect, the two modes of binding of the sugar are related by a 180° rotation about an axis perpendicular to the active-site cleft in the molecular surface. In both cases binding is enhanced by a number of favorable nonpolar contacts. Studies at low resolution with MurNAc (BLAKE et al. 1967b)

showed that this sugar also binds preferentially at site C, presumably because of the specificity of this site for *N*-acetyl groups. The orientation of the sugar, however, is different from that observed for GlcNAc because of the additional lactyl side chain at the O3 position.

2.2.2 Disaccharide Complexes of HEWL

The general features observed in the studies of monosaccharide binding to HEWL can be easily extrapolated to the disaccharide complexes of $(GlcNAc)_2$ and GlcNAc-MurNAc with HEWL. $(GlcNAc)_2$ was observed to bind in sites B and C, whereas the GlcNAc-MurNAc sugar bound at C and D. In addition, both these molecules appeared from the studies at 6.0 Å to bind at sites that were noncollinear with the other sites (B, C, and D). However, in the light of the results relating to the different modes of binding of anomeric sugars discussed in the previous section, this apparent anomaly can be explained. Since both of these disaccharides can exist with their terminal residues in the α and β forms, it is easy to see that when the terminal residue is in the α form and bound to the enzyme in the same way as α-GlcNAc, the second residue lies out of the main run of the active-site cleft.

2.3 The Study of Lysozyme–Oligosaccharide Binding at High Resolution

The study and refinement of an enzyme–inhibitor or enzyme–substrate complex at high resolution allow a much more detailed analysis of the structure than is possible at low or medium resolutions. In particular, the detailed nature of the interactions between enzyme and bound ligand are revealed, hydrogen bonds being well determined. An analysis of the thermal factors of both sugar and protein atoms also becomes feasible and, consequently, the identification of dynamic changes within the protein on substrate binding. In addition, the conformational changes within the protein accompanying the binding of the oligosaccharide can be clearly defined from a comparison of the refined coordinates of native and bound forms of the enzyme.

The crystals of the $(GlcNAc)_3$ bound complex were cocrystallized at pH 4.7 using the same method as that adopted in the earlier study at 2.0 Å. Details of the data collection and refinement methods, using the restrained least-squares procedure of HENDRICKSON and KONNERT (1980), are described elsewhere (CHEETHAM et al. 1987). The occupancy of $(GlcNAc)_3$ bound in the crystal was determined at 55%, this value producing the lowest overall R factor in the first stage of refinement. Rebuilding, a further round of refinement, and then a final rebuild resulted in a model for the HEWL–$(GlcNAc)_3$ complex with an R factor of 22.9%. This value, somewhat higher than R factors for native protein structures determined at comparable resolution (typically in the range of 18%–20%), is a reflector of the higher level of disorder within the crystal associated with the partial occupancy of the inhibitor.

The results of analyzing the refined structure of the enzyme–inhibitor complex are described in detail elsewhere. Here we attempt to outline the main structural features of the complex and to examine the nature of the interactions between the enzyme and trisaccharide molecules.

2.3.1 Interactions Between Protein and Oligosaccharide

The location of the $(GlcNAc)_3$ molecule within the active site of HEWL together with the hydrogen-bonding interactions between the bound inhibitor molecule and protein atoms are shown in Fig. 3; the enzyme–saccharide contacts are given in Table 2. Electron density difference maps revealed the trimer to bond in the upper half (sites A–C) of the active-site cleft, with its free, reducing group positioned toward the middle of the cleft, pointing downward, and the terminal sugar making the same contacts with the enzyme as a single β-(GlcNAc) molecule. In sites A and B, the hydroxyl O6 oxygens of the pyranose rings

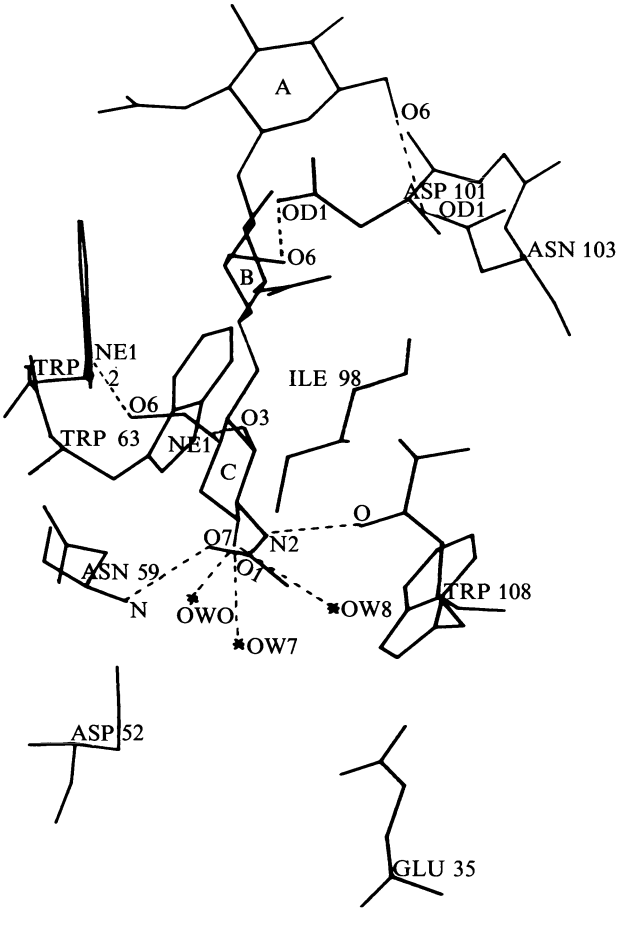


Fig. 3. $(GlcNAc)_3$ bound in sites A–C of the active-site cleft of HEWL. For clarity, only those residues involved in direct interaction with the bound trisaccharide are shown; the positions of Glu 35 and Asp 52 (the catalytic residues), which lie between sites D and E, are shown for reference. The hydrogen bonding contacts between $(GlcNAc)_3$ and HEWL are indicated by *dashed* lines; a cut-off distance of 3.3 Å between acceptor and donor atoms was used

Table 2. Protein–oligosaccharide interactions in the complex of $(GlcNAc)_3$ HEWL

Sugar site	Hydrogen Bonds			No. of VAN DER WAALS contacts <4.0 Å	Geometry of the glycosidic linkage
	Sugar atom	Protein Atom	Distance (A)		
A	O6	OD1 Asn 103	2.7	8	A to B: $\phi = -40°$; $\psi = -126°$ O5(A)–O3(B) = 3.2 Å
B	O6	OD1 Asp 101	2.4	10	
C	N2	CO Ala 107	2.8	25	B to C: $\phi = -90°$; $\psi = -117°$ O5(B)–O3(C) = 3.5 Å
	O7	NH Asn 59	2.8		
	O3	NE1 Trp 63	2.9		
	O6	NE1 Trp 62	2.9		
	O1	OWO Wat 142	3.2		
		OW7 Wat 141	2.6		
		OW8 Wat 141	3.3		

ϕ is the torsion angle about C1–O4′ defined by O5 C1 O4′ C4′; ψ is the torsion angle about O4′–C4′ defined by C1 O4′ C4′ C5′

interact with the OD1 of residues Asn 103 and Asp 101 of the protein, respectively. The involvement of Asn 103 in site A was not listed as one of the interactions in the original model for the complex at 2.0 Å. Neither of the *N*-acetyl side chains of the sugars in sites A and B is observed to be hydrogen-bonded to the protein, as was suggested in the earlier study in which a hydrogen bond between the nitrogen of the *N*-acetyl side chain of the sugar in site A and the OD1 of Asp 101 was shown (IMOTO et al. 1972). Both sugar side chains are, however, accessible to solvent (CHEETHAM, unpublished results); therefore, hydrogen bonding to solvent molecules, not well ordered within the protein structure and hence not defined in the electron density map, seems probable. The differences between the protein–oligosaccharide interactions in site A in the 2.0–Å and 1.75 Å models stems from the more precise location of the saccharide ring bound at this position in the study at higher resolution.

For the sugar bound in site C the number of hydrogen-bonding interactions between the $(GlcNAc)_3$ molecule and protein is noticeably greater than that in the other two sites. The hydroxyl oxygen (O6) and nitrogen (N2) atom of the *N*-acetyl side chain interact with the ring nitrogen of Trp 62 and the main-chain carbonyl oxygen of Ala 107, respectively. The oxygen (O7) of the *N*-acetyl side chain is hydrogen-bonded to the backbone NH of Asn 59. At the free, reducing end of the sugar the hydroxyl oxygen O(1), is hydrogen-bonded to three water molecules, 141 OW7 and OW8 and 142 OW0, which form part of a network of hydrogen-bonded waters bound in the middle region of the active site. The water 141 OW8 is one of six well-ordered waters bound within the active-site cleft of the native protein (HANDOLL 1985).

The three sugar rings are all bound at different depths within the active-site cleft. The saccharide ring at site A is the least tightly bound of the three GlcNAc units, lying almost on the surface of the binding cleft and relatively exposed to solvent (solvent accessible area, SA = 180 $Å^2$). The pyranose ring carrying

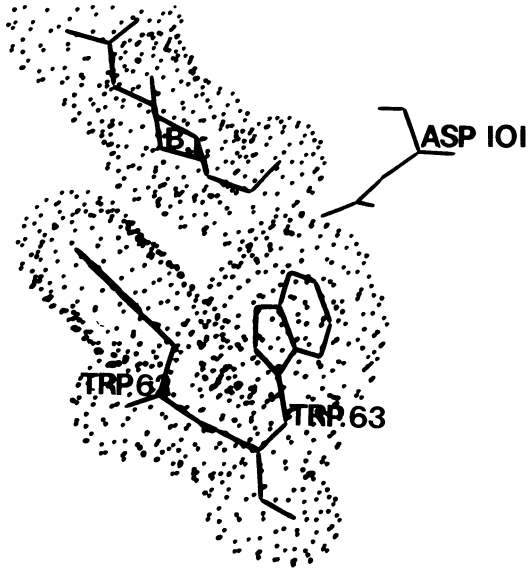


Fig. 4. Packing of the apolar face of the pyranose ring, bound at site B in the $(GlcNAc)_3$–HEWL complex, with the plane of the tryptophan ring of residue 62. *Dotted surface* represents the VAN DER WAAL radii of the atoms

the polar *N*-acetyl side-chain faces toward a mainly hydrophilic region of the protein surface, defined by residues Arg 73 and Asp 101. In site B the saccharide ring lies with its apolar face almost directly above the plane of the indole ring of Try 62 (Fig. 4), suggesting that the interaction between the sugar and the protein in this site is largely hydrophobic. The presence of this large hydrophobic interaction would suggest the binding energy for this site to be somewhat greater than that for site A, in which a single hydrogen bond is the only significant interaction between ligand and protein. This result conforms to the description of the relative free energies of substrate binding at sites A–F in HEWL given in IMOTO et al. (1972) and based on the experimental determination of the binding energies of a large number of different sugar molecules to the enzyme in the solution state (IMOTO et al. 1972, and references therein). The solvent accessibility of the sugar bound at this site (SA = 130 Å^2) is intermediate between that of those bound at sites A and C. The GlcNAc residue bound in site C lies well within the active-site cleft, at a depth of approximately 6 Å and relatively unexposed to solvent (SA = 77 Å^2). The position of Trp 63 is such that the plane of the tryptophan ring lies at right angles to the nonapolar face of the saccharide ring, allowing a hydrogen bond between the tryptophan ring nitrogen NE1 and the hydroxyl oxygen O3 of the sugar. The ring nitrogen of Trp 62 also lies toward the same face of the sugar ring, making a hydrogen bond with the hydroxyl O6. The nature of the interactions within this subsite are therefore of a much more specific nature than those within sites A and B. The environment of the apolar face of the sugar bound at this position is relatively hydrophobic, with residues Ile 98, Ala 107, Trp 108, and Val 109 all lying within 8.0 Å. None of the hydroxyl groups of the three pyranose rings act as both hydrogen-bond acceptors and donors in the complex, as had been observed in the study of monosaccharide binding

to L-arabinose (QUIOCHO and VYAS 1984) and subsequently suggested as a general feature of protein–carbohydrate interactions by QUIOCHO (1986; see Chapt. 5, this volume).

The hydrogen-bonding interactions within the binding subsites are directional by nature (BAKER and HUBBARD 1984) and determine the precise orientation in which the sugar is bound within the active site and thus the specificity of binding. The hydrophobic interactions, particularly that between the tryptophan ring of residue 62 and the pyranose ring bound in site B, also contribute to the binding energy but are much less specific.

2.3.2 Temperature Factors for the Trisaccharide Complex

One of the most important results to come out of the refinement of the HEWL–$(GlcNAc)_3$ complex at high resolution is a set of thermal factors for both the protein and carbohydrate atoms within the complex. As a consequence, one is able to monitor any changes in the mobility of regions of the protein as a result of binding the trisaccharide by comparison of the datda with temperature factors obtained from the refinement of native HEWL at 1.6-Å resolution (HANDOLL et al. 1987). In the native protein, the lips of the active site constitute one of the most mobile regions of the structure, a feature that is also brought out by molecular dynamics simulations (POST et al. 1986). This conformational flexibility presumably allows the binding site of the protein to accommodate readily an incoming substrate or inhibitor molecule.

The most noticeable mobility changes within the protein, on binding $(GlcNAc)_3$, were found for the atoms of those residues observed to interact directly with the bound inhibitor. These residues "stiffen" when the trisaccharide binds, resulting in lower thermal factors compared to those found in the native protein. For example, the average temperature factors for the side-chain atoms of Trp 62 in the bound and native protein environments are 21 and 47 $Å^2$, respectively. A further interesting point to emerge from an analysis of temperature factors was that the B factors of atoms in the $(GlcNAc)_3$ molecule that are observed to interact directly with the protein are of a similar magnitude to those of the protein atoms with which they interact. The results also clearly showed that the mobility of the GlcNAc molecule bound in site C (25 $Å^2$), which lies well buried within the active site cleft, is markedly lower than that of the sugar in site A (42 $Å^2$), which lies almost on the surface. The mobility of the sugar at site B is intermediate between the two (35 $Å^2$). Thus the mobility of the bound sugar molecules can be directly correlated with the binding affinity of the individual subsites in which they are located.

2.3.3 Conformational Changes Accompanying the Binding of $(GlcNAc)_3$ to HEWL

The conformational changes accompanying the binding of the trisaccharide to HEWL were determined by a comparison of the HEWL–$(GlcNAc)_3$ struc-

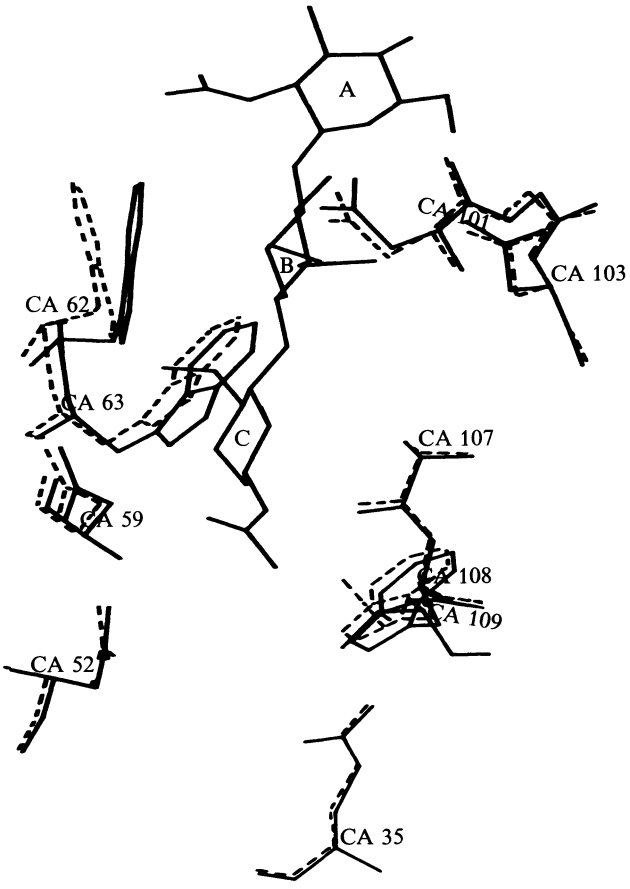


Fig. 5. Conformational changes within the active site of HEWL on binding $(GlcNAc)_3$. *Bold lines* indicate atomic positions in the bound complex, *dashed lines* those of the native protein

ture at 1.75 Å with the native structure at 1.6 Å. The major conformational changes between the bound and native forms of the protein are illustrated in Fig. 5. As observed in the study at 2.0 Å, the most striking movement is that of the tryptophan ring of residue 62, which rotates 17° about the CB position towards the bound inhibitor. This results in the face of the ring being brought close to the apolar face of the pyranose ring bound in site B, thus narrowing the active-site cleft and enhancing the interaction between protein and sugar at this site. Smaller conformational changes occur for Trp 62, Ala 107, Asp 101, and Asn 103, facilitating the formation of hydrogen bonds with the bound oligosaccharide. The accommodation of the $(GlcNAc)_3$ molecule within the active site of the protein also results in minor conformational changes for residues not directly involved in interaction with the bound molecule, but which lie in close proximity to the site of binding (e.g., Trp 108). In addition, small structural perturbations are observed in regions of the pro-

tein remote from the active site, suggesting that the conformational changes which occur on the binding of $(GlcNAc)_3$ to HEWL, although concentrated at the active site, are transmitted to some extent throughout the structure. This result is supported by evidence obtained from solution studies with nuclear magnetic resonance techniques (CHEETHAM 1986).

The conformation of the trisaccharide in the bound complex is somewhat distorted from the regular ribbonlike conformation seen in β-cellobiose. Table 2 gives the conformational parameters of the two β-(1–4) linkages between the A and B and the B and C sugar units. The glycosidic bridge between the saccharide rings bound in sites B and C is characterized by a twist of approximately 28°. This is intermediate between that observed for the disaccharide bound in triclinic lysozyme (twist of about 22°), but smaller than that observed between adjacent sugars in β-cellobiose (twist of about 33°) (see MO and JENSEN 1978, and references therein). An intramolecular hydrogen bond is formed between the ring oxygen O5 of sugar B and the hydroxyl O3′ of the C sugar. The twist between the sugars bound in sites A and B is approximately 70°, which is significantly larger than that between the saccharide units bound at sites B and C. As a consequence of the large twist between the adjacent saccharides the distance between O5 and O3′ is too large (3.5 Å) for an intramolecular hydrogen bond of the type O5(n-1)–O3′(n) to be formed. In chitobiose (MO and JENSEN 1978), where the degree of twist about the glycosidic bridge between β-(1–4) linked disaccharides units is similarly large (about 54°), this intramolecular hydrogen bond is also absent.

2.4 The Mode of Oligosaccharide Binding to Site D in HEWL

At the time that PHILLIPS proposed his mechanism for the action of lysozyme (PHILLIPS 1966) there was no direct experimental evidence for the mode of binding of sugars at this site. The proposal was based on judicious model-building studies, for which the available 2.0 Å structure of the trisaccharide complex of HEWL provided a starting point. Three additional GlcNAc residues were then built into the lower half of the binding cleft, in sites D–F, to produce a model for the hexasaccharide complex. It was argued that the observed binding in the trisaccharide complex may represent a part of the true enzyme–substrate complex that does not embrace the catalytic site. Careful model-building studies based upon this structure might, therefore, reveal the site and mechanism of catalysis.

The additional residues in sites E and F could be added to the model without difficulty, and a number of good interactions with the enzyme appeared possible. For the sugar residue at site D, however, it was found that the contacts of the C(6) and O(6) atoms with the main-chain CO of Asp 52, with Trp 108, and with the acetamido group of the sugar bound in site C were unacceptably close. To relieve this overcrowding, the pyranose ring of the D sugar was distorted away from the normal chair conformation toward a conformation in

which the C(6) lies in an axial position. Hydrogen-bonding interactions between the O(6) of the sugar and either CO of Asp 52 or Glu 35 were then feasible (BLAKE et al. 1967b; PHILLIPS 1966).

Theoretical calculations by LEVITT (1972, 1974), in which the binding of a hexasaccharide substrate to lysozyme was analyzed in terms of energy-minimization procedures, suggested that the hexamer could bind with all the sugar residues in the chair conformation. Distortion of the sugar in the D site did lead to better contacts with the protein but not to a significantly lower binding energy. LEVITT therefore suggested that the unfavorable binding energy associated with the D site arises from an increase in the electrostatic energy of Asp 52, as its solvation shell is displaced by the incoming sugar residue. He further proposed that the D subsite is able to accommodate a sugar residue in the distorted conformation, but that this conformation arises as the oxonium-carbonium ion intermediate develops rather than simply as a result of steric considerations, as suggested by PHILLIPS.

Transition-state theory suggests that an enzyme works by binding neither substrates nor products as tightly as possible but rather by binding the transition state of the reaction (PAULING 1946). This idea led to the proposal that one might study the mechanism of an enzyme reaction by designing stable molecules that resemble the transition states closely enough to bind to the enzymes in similar ways. Such analogues should themselves be very good inhibitors of the enzymes and should also bind tightly to them (WOLFENDEN 1969).

In the application of this idea to lysozyme, SECEMSKI and LIENHARD (1971) prepared the ∂-lactone (TACL), derived from tetra-*N*-acetylchitotetraose [$(GlcNAc)_4$], and found it to be a very effective inhibitor of the enzyme. A crystallographic study at 2.5 Å of the binding of this compond to HEWL (FORD et al. 1974) showed the tetrasaccharide lactone to bind in sites A, B, C, and D, with the sugar residues in the upper three sites located in similar positions to those observed in the tri-*N*-acetylchitotriose complex and analogous conformational changes within the active site region of the protein. A minor conformational change in Asp 52 was also indicated and appeared to be a special feature of the binding of TACL. Analysis of the electron density in site D indicated that the ∂-lactone was bound in a conformation close to a sofa, bringing the hydroxymethyl group C_6–O_6 axial. This study therefore seemed to provide experimental evidence to support the role of strain in the proposed mechanism of lysozyme catalysis. In a companion study on the binding of $(GlcNAc)_4$ to lysozyme (GRACE 1980) the tetrasaccharide was found to bind in sites A–D, although the occupancy of the sugar in site D was considerably less than that of the sugar in site C. The weak difference of electron density in site D, however, appeared similar to the stronger density in this region for TACL, in particular with respect to the lobe of density representing the C_6–O_6 bond. It was therefore suggested that the fourth sugar residue of the tetrasaccharide was bound in a similar conformation to the gluconolactone ring in the TACL–HEWL complex, but that the glucopyranose ring was less firmly localized than the lactone.

Some 5 years later KELLY et al. (1979) described a crystallographic study of the bacterial cell wall trisaccharide MurNAc-GlcNAC-MurNAC binding to

HEWL at 2.5 Å resolution. The trisaccharide bound in sites B, C, and D of the active site, with the residue in site D bound in the normal chair conformation. Structural changes accompanied binding of the oligosaccharide: the indole ring of Trp 62 moved toward site B and Trp 63 toward site C into a possible hydrogen-bonding position with the carbonyl oxygen of the 2-acetamido group of the GlcNAc residue. Also noted was concerted shift of the main-chain and side-chain atoms from Gly 71 to Leu 75, relayed from the active site region via a hydrogen-bonding network involving residues Ser 72, Ser 60, Asn 59, and Arg 61. The movement of the tryptophan rings was similar to that seen with the binding of both tri-*N*-acetylchitotriose and the tetrasaccharide lactone. In the Pro 70 region smaller conformational changes than those described by KELLY et al. have been reported on the binding of TACL and $(GlcNAc)_3$ to HEWL (FORD et al. 1974; JOHNSON et al. 1968); in the study of the tri-*N*-acetylchitotriose complex at 1.75 Å there is some evidence for movement in this area on binding the sugar, from a comparison with the structure at 2.0 Å of the low-pH form of the native enzyme. However, since the same region shows conformational flexibility between the low- and high-pH forms of the native enzyme (HANDOLL et al. 1987) and is synonymous with poorly defined regions of electron density, it proved difficult to determine the extent of the conformational change, if any, occurring in this region of the protein.

On comparison of their structure for the trisaccharide complex with that of the TACL complex determined by FORD et al. it was evident that the penetration of the trisaccharide into the active site was much less than that of the tetrasaccharide lactone; root mean-square deviation of the 43 common atoms between the two molecules in their respective binding modes was 1.23 Å (sites B, C, and D). On the basis of these results KELLY et al. suggested that the binding orientation which they observed for MurNAc-GlcNAc-MurNAc may be characteristic of an initial Michaelis complex. Thus, taken together, these two studies may be providing us with snapshots along the time course of the reaction. Arguably, the tetrasaccharide lactone study might offer a view of the transition state of the reaction, while the cell wall trisaccharide complex is more representative of either the product in the course of being released from the enzyme or of a part of the initial Michaelis complex with the substrate.

2.5 Comparative Studies of Oligosaccharide Binding to Different Forms of Lysozyme

2.5.1 Chicken-Type Lysozymes

Three-dimensional structure determinations have now been carried out on a large number of different c-type lysozymes isolated from humans (ARTYMIUK and BLAKE 1981) and from hen (HANDOLL et al. 1987), turkey (SARMA and BOTT 1977), and tortoise eggs (ASCHAFFENBURG et al. 1980; PULFORD 1982). They show that the conformation of the protein molecule is highly conserved among the different species and crystal structures. Binding studies with a number

Table 3. Protein–oligosaccharide interactions observed in the complexes of several c-type lysozymes with a number of different inhibitor molecules bound in the active-site cleft

Lysozyme	Inhibitor	Binding site(s)	Polar contacts			No. van der Waals contacts <4 Å	References
			Sugar atom	Protein Atom			
Hen	α-NAG	C	N2	CO	Ala 107		1
			O7	NH	Asn 59	–	
			O1	NH	Val 109		
	β-NAG	C	N2	CO	Ala 107		1
			O7	NH	Asn 59	–	
			O6	NE1	Trp 62		
			O3	NE1	Trp 63		
	Tri-NAG	A, B, C	O6(A)	OD1	Asn 103	8(A)	2
			O6(B)	OD1	Asp 101	10(B)	
			N2(C)	CO	Ala 107	25(C)	
			O7(C)	NH	Asn 59		
			O3(C)	NE1	Trp 63		
			O6(C)	NE1	Trp 62		
			O1(C)	OW0	Wat 142		
				OW7	Wat 141		
				OW8	Wat 141		
	TACL	A, B, C, D	O6(A)	No polar contacts		2(A)	3
			O6(B)	OD2	Asp 101	9(B)	
			N2(C)	CO	Ala 107	35(C)	
			O7(C)	NH	Asn 59		
			O3(C)	NE1	Trp 63		
			N2(D)	ND2	Asn 46	40(D)	
			O5(D)	OD2	Asp 52		
			O6(D)	NH	Val 109		
				OE1	Glu 35		
	NAM-NAG-NAM	B, C, D	O10(B)	OD1	Asn 103		4
			O10(B)	ND2	Asn 103		
			N2(C)	CO	Ala 107		
			O7(C)	NH	Asn 59		
			O3(C)	NE1	Trp 63		
			O5(D)	OD2	Asp 52		
			O6(D)	NH	Val 109		
				OE1	Glu 35		
			O10(D)	ND2	Asn 46		
Tortoise	$(NAG\text{-}NAM)_2$	A, B, C, D	Site A	No polar contacts		1(A)	5
			O12(B)	OH	Tyr 63	9(B)	
			O6(B)	OD2	Asp 102		
			O7(C)	CO	Gln 58	50(C)	
			N2(C)	CO	Ala 108		
			O3(C)	NE1	Trp 64		
			O10(D)	OD1	Asp 52	95(D)	
			O6(D)	OE2	Glu 35		
			O6(D)	NH	Val 110		

Table 3. (continued)

Lysozyme	Inhibitor	Binding site(s)	Polar contacts			No. van der Waals contacts <4 Å	References
			Sugar atom	Protein Atom			
Human	tri-NAG	B, C, D	Site B	No polar contacts		9(B)	6
			O3(C)	NE1	Trp 64	17(C)	
			O7(C)	NE1	Trp 64		
			O6(D)	OE2	Glu 35	11(D)	
			N2(D)	OD1	Asn 46		
Turkey	NAG-NAM	E, F′	O3(E)	CO	Asn 44		7
			N2(E)	CO	Phe 34		
			O7(E)	NH	Asn 44	–	
			O6(F′)	CO	Phe 34		
			O10(F′)	NH	Asn 113		

Bracketed letters, e.g. (A), refer to the binding subsite for which data are given.

References: [1] IMOTO et al. 1972; [2] CHEETHAM et al. 1987; [3] FORD et al. 1974; [4] KELLY et al. 1979; [5] GALLAY and BLAKE unpublished results; [6] BANYARD 1973; [7] SARMA and BOTT

of small oligosaccharide inhibitors have been carried out for all the different enzyme forms mentioned above; Table 3 provides a comprehensive summary. The results for the human and tortoise (HuL and TEWL, respectively) are now considered in broad relation to those obtained from studies with the hen enzyme, which have been discussed in detail in the preceding sections. Initially, however, we review the studies on oligosaccharide binding to the turkey form of the enzyme (SARMA and BOTT 1977), which led to the first crystallographic evidence for the binding of sugars in the fifth and sixth subsites of the active-site cleft of lysozyme.

In the crystal structure of native turkey lysozyme, refined to 2.8 Å by SARMA and BOTT (1977), the packing of the protein molecules within the hexagonal cell is such that the entire length of the active-site cleft lies in the vicinity of the crystallographic sixfold axis and is not blocked by neighboring molecules, as it is in tetragonal HEWL (see Sect. 2.1.2). The turkey enzyme therefore seemed an ideal candidate with which to attempt to obtain direct crystallographic evidence of oligosaccharide binding at sites E and F in the active site of the enzyme. Difference electron density maps calculated between crystals soaked in the disaccharide GlcNAc-MurNAc and the native crystal showed four positive peaks, three strong ones in the vicinity of sites corresponding to subsites C, E, and F, as defined by BLAKE et al. for HEWL (1967b), and a weaker peak at the D subsite. No significant density was observed at sites corresponding to the A and B subsites, indicating that the binding, if any, of the disaccharide at these positions is extremely weak; this was not considered surprising since one of the residues, Asp 101, observed to interact with oligosaccharides in this region of the active site of the hen enzyme is substituted by a Gly in the turkey form. The strong/weak pair of peaks at subsites C and D was interpreted as binding of the disaccharide in a manner similar to that observed for the tetrasac-

charide $(GlcNAc)_4$ by GRACE (1980), with strong association at site C but somewhat weaker binding to site D. The saccharide bound at site C appears to make four hydrogen bonds to the protein, as found in the case of the hen enzyme (Fig. 3).

The observation of two strong peaks in the difference map beyond subsite D provided the first real opportunity to test the predictions of the model-building studies of BLAKE et al. (1967b) as to the nature of the interactions between the protein and bound sugar residues in these lower sites. The first of these peaks corresponded to a site close to the E subsite predicted for the hexasaccharide in HEWL. As in the predicted model, the sugar residue at this site makes three hydrogen bonds to the enzyme. Some of the residues involved in these interactions are, however, different from those predicted by the model; the saccharide ring in the turkey complex lies closer to Phe 34 than Gln 57, as in the modelled complex. The position of the second peak, however, does not correspond to the subsite termed F in the modelled complex. Instead, the MurNAc residue is bound at subsite F′, making two hydrogen bonds to the protein. One of these is between the carboxyl of the lactyl side chain of the MurNAc residue and the NH2 group of Asn 113; the second involves the O(6) oxygen of the saccharide and the main-chain CO group of Phe 34. Small conformational changes within the protein, on binding the disaccharide, were observed but not details reported.

PINCUS and SCHERAGE (1981) have predicted the three-dimensional structures of a number of complexes of lysozyme with cell wall substrates on the basis of theoretical calculations of the conformational energies of complexes of alternating GlcNAc-MurNAc copolymers with HEWL. They calculated the relative affinities of the disaccharide GlcNAc-MurNAc for all sequential pairs of sites A–F, including the E and F sites on both sides of the binding cleft. Earlier calculations with the hexasaccharide $(GlcNAc)_6$ had shown two possible low-energy structures for the binding of this molecule to HEWL. In one (the lowest energy structure) the substrate is bound with the saccharide in site D lying close to the surface of the binding cleft and the fifth and sixth sugar residues in E and F sites involving residues such as Arg 45, Asn 46, and Thr 47 on the so-called "left" side of the active site. The second structure (of somewhat higher energy) shows the residue at site D located deep into the binding cleft near Glu 35 and Asp 52 and the fifth and sixth residues lying on the "right" side of the active site, close to Phe 34 and Arg 114. Their calculations found sites C and D and the right-side sites E and F to be the highest affinity sites for the GlcNAc-MurNAc disaccharide. These results are therefore in good agreement with the experimental observations of SARMA and BOTT on the mode of binding of this disaccharide to turkey lysozyme. Tentative experimental identification of the theoretically predicted left-side binding mode for the $(GlcNAc)_6$ hexasaccharide has been reported by SMITH-GILL et al. (1984), using a monoclonal antibody, HyHEL-5, which binds specifically to an epitope including residues Arg 45, Asn 46, Thr 47, Asp 48, and Arg 68 on the far left side of HEWL.

Recent crystallographic studies on the binding of the tetrasaccharide $(GlcNAc\text{-}MurNAc)_2$ to the tortoise enzyme at 2.5 Å (BLAKE and GALLAY, unpublished results) have shown this sugar to bind to sites A, B, C, and D (Fig. 6).

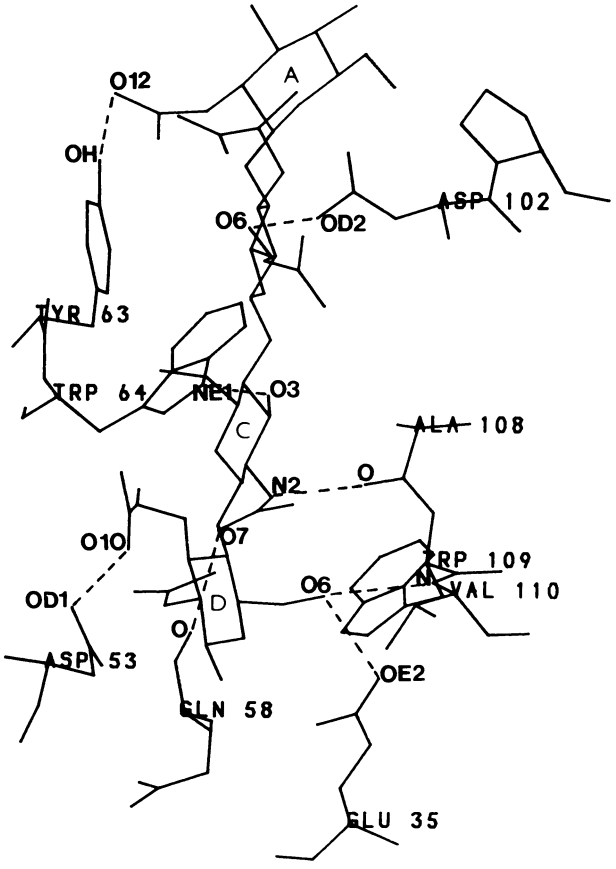


Fig. 6. Tetrasaccharide bound at the active site of TEWL in sites A, B, C, and D. The sugar $(GlcNAc\text{-}MurNAc)_2$ consists of alternating GlcNAc and MurNAc units. Only those residues in the immediate vicinity of the bound sugar molecule are shown. Hydrogen bonds between protein and sugar atoms are indicated by *dashed lines*; a cut-off distance of 3.3 Å between donor and acceptor atoms is used

The binding in sites A–C is analogous to that observed in HEWL, with similar types of conformational changes occurring within the active site. Tyr 63 moves into the active site, its hydroxyl oxygen forming a hydrogen bond to the O(12) of the MurNAc residue in site B and the plane of the tyrosine ring lying parallel to the apolar face of sugar residue B; the plane of the tryptophan ring of residue 62 in the HEWL complex lies in an almost identical position (see Fig. 3). In site C three hydrogen bonds are formed between the O3, O7, and N2 atoms of the sugar and the ring nitrogen (NE1) of Trp 64, the main-chain CO of Gln 58, and the main-chain CO of Ala 108, respectively; no waters have been included in the analysis of the model as yet. Trp 64 shows a small movement away from the active site, analogous to that observed for Trp 63 in the hen enzyme, to avoid close VAN DER WAALS contacts with the saccharide ring. The difference electron density in site D indicates that the sugar in this position is bound in the normal chair conformation, lying almost on the surface of the active-site cleft. This mode of binding is similar to that reported by KELLY et al. (1978) in their study of the MurNAc-GlcNAc-MurNAc complex of HEWL. Hydrogen bonding interactions are found between the O10 of the sugar

and OD1 of Asp 53 and between O6 and both the main-chain NH group of residue Val 110 and the carboxylate OE2 of residue 35. These interactions are analogous to those reported in the TACL complex by KELLY et al. There is no evidence for distortion of the saccharide ring as was observed by GRACE (1980) in the study of the $(GlcNAc)_4$–HEWL complex. One might suggest, therefore, that the complex with the tortoise enzyme provides a view of the Michaelis complex for the reaction, rather than the transition-state conformation of the enzyme complex.

All the studies of oligosaccharide binding to lysozyme discussed up to this point have suggested that the enzyme undergoes some conformational changes within the active-site region on binding a sugar molecule. The study of trisaccharide binding to HuL (BANYARD 1973) at 2.5 Å showed no significant movements of the protein when $(GlcNAc)_3$ bound at sites B, C, and D. Refinement of the data is, however, incomplete (P. ARTYMIUK, personal communication) one is therefore able to make only preliminary comments on the basis of the unrefined model. A more detailed analysis seems premature at this stage and may even be misleading. A comparison with the structure of the native protein determined at 1.5 Å (ARTYMIUK and BLAKE 1981) shows no evidence of conformational changes within the bound protein, supporting the earlier observations from the study at 2.5 Å. In the native crystal the protein molecules are very tightly packed; therefore, intermolecular interactions may restrict movement in the region of the active site. Furthermore, the dimensions of the active site in the human enzyme are close to those found for the hen enzyme with inhibitor bound; therefore, one might suggest that in the case of HuL, the incoming trisaccharide molecule can achieve as good a fit within the binding cleft as is observed for HEWL without necessitating conformational changes in the active-site region. The movements of the tryptophan rings in the vicinity of sites B and C in the hen enzyme on binding $(GlcNAc)_3$ have the effect of narrowing the active-site cleft and enhancing the interactions between protein and bound oligosaccharide. It may therefore be that the human enzyme crystallizes in a nonnative state, possibly due to the presence of a nitrate ion bound in site C, close to the position at which the *N*-acetyl group of a sugar residue might be expected to bind (ARTYMIUK 1979).

The binding studies with different members of the c-type lysozymes thus point to several features within the active site essential for the binding of oligosaccharides. The presence of a strong hydrophobic interaction between the residue at position 62 (or 63) and the sugar bound in site B seems critical to binding at this subsite. A more general requirement for the C and D subsites appears to be the presence of a large number of planar, polar residues, e.g., Asp and Asn, with which the hydroxyl oxygens of the saccharide rings can interact; such interactions were suggested as a general feature of protein–carbohydrate interactions by QUIOCHO (1986). The location of the two key catalytic residues identified by BLAKE et al. (1967b), Asp 52 and Glu 35, between the D and E subsites of the active-site cleft is also clearly of paramount importance. A comparison of the sequences of the hen, tortoise, turkey, and human lysozymes shows this conservation of particular residue types within individual subsites (Fig. 7).

	34	35	44	46	52	57	59	62	63	98	101	103	107	108	109	113	114
Hen	Phe	Glu	Asn	Asn	Asp	Gln	Asn	Trp	Trp	Ile	Asp	Asn	Ala	Trp	Val	Asn	Arg
Human	Trp	Glu	Asn	Asn	Asp	Gln	Asn	Tyr	Trp	Val	Asp	Gln	Ala	Trp	Val	Asn	His
Tortoise	His	Glu	Asn	Asn	Asp	Gln	Asn	Trp	Trp	Ile	Asp	Asn	Ala	Trp	Val	Asn	Arg
Turkey	Phe	Glu	Asn	Asn	Asp	Gln	Asn	Trp	Trp	Ile	Gly	Asp	Ala	Trp	Val	Asn	Arg

Fig. 7. Comparison of the amino acid sequences of hen, human, tortoise, and turkey lysozymes for residues within the active site cleft which are observed to participate in protein–oligosaccharide interactions. Numbering shown is that for the hen enzyme

2.5.2 Goose and Phage-Type Lysozymes

The three-dimensional structure of GEWL refined at 2.8 Å resolution by GRUTTER et al. (1983) shows similarities to the structures of both HEWL and bacteriophage T4L. The nature of the structural correspondence suggests that all three classes of lysozyme diverged from a common evolutionary precursor (WEAVER et al. 1985) although their amino acid sequences appear to be quite unrelated (GRUTTER et al. 1983). A comparison of the structures of the three lysozymes allows one to identify the changes in both amino acid sequence and three-dimensional structure have occurred within the active-site region of the enzyme during the course of evolution.

WEAVER et al. (1985) transformed the coordinates of the tetrasaccharide lactone (FORD et al. 1974) into the GEWL coordinate system to explore the overall relationship between the active site of GEWL and the active sites of HEWL and T4L. Their comparison of the three structures showed a number of similarities in the active sites, but also one striking difference. In the alignments of the backbones of the three structures Glu 73 (GEWL) corresponds spatially to Glu 35 (HEWL) and Glu 11 (T4L). Also, there are segments of the GEWL backbone in the active-site region that are similar to those seen in T4L and HEWL: e.g., HEWL backbone 57–59 and 107–108; T4L backbone 30–32 and 104–105; and GEWL 95–97 and 146, 147. However, although there are two aspartates in the GEWL active site, neither corresponds exactly to Asp 52 (HEWL) or Asp 20 (T4L). The discrepancy in the carboxyl group locations is about 10 Å for Asp 86 and 4 Å for Asp 97. WEAVER et al. suggested that the lack of structural correspondence in this region of the active site may, therefore, reflect some differences in the mechanism of action of the three lysozymes. For example, one or more of the lysozymes may undergo a substantial conformational change during catalysis, as a result of which the respective aspartates may be brought into comparable stereochemical positions; no experimental evidence for such a large conformational change, however, has been forthcoming from the large number of crystallographic studies of lysozyme–oligosaccharide complexes to data. A second, but less likely possibility, is that the catalytic

activity of GEWL does not require the presumptive "catalytic" aspartates seen in HEWL and T4L; although all three lysozymes cleave the same β-(1–4) glycosidic bond, they differ in their specifities towards saccharides with different substituents (ARNHEIM et al. 1973; KLEPPE et al. 1981).

2.6 The Extended Recognition Site for the Binding of Oligosaccharides to Lysozyme

In the case of phage lysozyme a great deal of information is potentially available from the studies of mutant enzymes that have modified activities. TSUGITA (1971) examined all the known mutants of the enzyme and divided the sites at which amino acid substitutions occur into two classes: essential and nonessential. An essential amino acid was defined as one for which any substitution substantially reduced the catalytic activity of the enzyme, and a nonessential amino acid was one for which at least one substitution did not impair enzymatic activity. On this basis TSUGITA designated Asp 20, Gln 105, Trp 138, Glu 22, Asn 140, and Glu 141 as essential residues. Examination of the three-dimensional structure of T4L showed that in each case the designated residue lies within the vicinity of the presumed active-site cleft (MATTHEWS and REMINGTON 1974).

Two mutants, CSE and 335D101, proved an exception to the above rule. The amino acid substitution in the CSE mutant was identified as a Glu to Lys change at position 128 of the amino acid sequence (TSUGITA and INOUYE 1968). In the native enzyme Glu 128 is located at the back of the C-terminal domain, 25 Å from the active-site cleft. Crystallographic analysis of the mutant enzyme has confirmed this location and has also shown that the replacement of the glutamic acid residue, which lies on the surface of the protein, by a lysine causes very little change in the three-dimensional structure of the enzyme (GRUTTER and MATTHEWS 1982); thermodynamic analysis showed the stability of the mutant enzyme to be comparable with that of the native enzyme (SCHELLMAN and HAWKES 1980; SCHELLMAN et al. 1981; GRUTTER et al., unpublished results). In the 335D101 mutant there are two amino acid substitutions within the C-terminal domain, with the substitution of Arg 125–Trp 126 by Thr–Leu (REMINGTON et al. 1978). Both mutant enzymes show only 4% of the activity of the wild type.

To rationalize the above results, GRUTTER and MATTHEWS proposed that the role of Arg 125 and Glu 128 is to participate in the recognition and binding of the cross-linking peptide that links adjacent oligosaccharide strands within the cell wall of *Escherichia coli* (GHUYSEN 1968; TSUGITA 1971); the involvement of Trp 126 was precluded on the basis of information derived from other mutants (REMINGTON et al. 1978). From the results of binding studies with a series of mono- and oligosaccharides and by analogy with HEWL, it was suggested that from the known orientation of the saccharide in site D, one can infer that the cross-linking peptide emanating from the 3-hydroxyl atom on the sugar must extend across the face of the upper, C-terminal lobe of the lysozyme molecule, passing in the immediate vicinity of Arg 125 and Glu 128 (ANDERSON et al. 1981). GRUTTER and MATTHEWS therefore postulated that this does indeed

occur, and that Glu 128 and Arg 125 participate in the binding and orientation of the peptide, these interactions helping in turn to position the saccharide part of the cell wall within the active-site cleft, in the optimal alignment for hydrolysis.

This hypothesis is consistent with a number of experimental findings. First, it is known that for the phage lysozyme at least part of the cross-linking peptide must be present for hydrolysis (BIENKOWSKA and TAYLOR 1979). Second, changes in the peptide linkage have been shown to modify the activity of the enzyme (TSUGITA 1971). Finally, the result is consistent with the known structural rerlationship between T4L and HEWL (REMINGTON et al. 1978). The C-terminal lobe of T4L has no analogous counterpart in the hen enzyme, suggesting that the function of the additional residues in the phage lysozyme is to enhance the activity of the enzyme towards *E. coli* cell walls. Conversely, HEWL can be seen as a muramidase of broader specificity but lower efficiency.

3 Glycogen Phosphorylase

3.1 Role in Glycogen Metabolism and Catalytic Properties

Glycogen is the principal storage form of carbohydrate in mammalian cells. It is used in skeletal muscle to provide ATP via glycolysis to sustain muscle contraction, in heart and brain to provide fuel during brief periods of anoxia, and in the liver in the maintenance of blood glucose levels. Glycogen consists of polysaccharide chains of α-D-glucose units linked by α-(1–4) glycosidic bonds with α-(1–6) branch points occurring at approximately very 10–12 sugar residues. Branching has the effect of creating a more open structure with increased solubility compared to linear chains. Rabbit muscle glycogen contains a covalently bound protein (molecular weight 37000) named glycogenin which is linked to the sugar through a tyrosyl residue (WHELAN 1986). The degradation and synthesis of glycogen are regulated both allosterically by changes in concentrations of metabolites and, in response to hormonal or neural signals, by an intricate classical enzyme cascade system involving protein kinases and protein phosphatases in what is one of the best understood processes in the whole of metabolic regulation (COHEN 1983).

Glycogen phosphorylase (α-1,4-glycan: orthophosphate glycosyl transferase, E.C. 2.4.1.1) catalyzes the first and rate-limiting step in the degradation of glycogen:

$$G_n + P_i \rightleftharpoons G_{n-1} + \text{Glc-1-P}$$

where G_n is an α-(1–4) linked glucosyl polymer of n sugars, P_i is inorganic phosphate and Glc-1-P is α-D-glucose 1-phosphate. The reversible phosphorylation of the α-(1–4) glycosidic linkage occurs at the nonreducing ends of the glycogen chains; in contrast to lysozyme, which hydrolyzes an internal β-(1–4)

glycosidic linkage, phosphorylase releases one sugar phosphate per enzymatic cycle. Phosphorylase digests to within four residues of a branch point. The synthesis of glycogen is catalyzed by glycogen synthese, which utilizes uridine-diphosphate-glucose to transfer a glucosyl unit to the nonreducing end of a glucosyl polymer having at least four sugar residues.

In resting muscle, phosphorylase exists in the *b* form and is inactive. Phosphorylase *b* is activated by increased levels of AMP (or analogues of AMP such as IMP) and is inhibited by glucose 6-phosphate and ATP. In response to nervous system or hormonal signals phosphorylase *b* is activated to phosphorylase *a* by reversible phosphorylation of a single serine residue, Ser 14. Phosphorylase *a* is active in the absence of AMP, although its activity may be augmented by AMP. The reverse reaction leading to inactivation of phosphorylase *a* is catalyzed by phosphorylase phosphatase. The control and catalytic properties of phosphorylase have been reviewed (GRAVES and WANG 1972; FLETTERICK and MADSEN 1980; DOMBRADI 1981; HELMREICH and KLEIN 1980; FLETTERICK and SPRANG 1982; JOHNSON et al. 1988).

Electron microscopy of striated muscle thin sections has shown that most of the glycogen occurs in particles (average diameter 400 Å). These are located in the sarcoplasm, close to the sarcoplasmic reticulum at the level of the I band in muscle (WANSON and DROCHMANS 1968). In a series of papers published in the 1970s (MEYER et al. 1970; HEILMEYER et al. 1970; HASCHKE et al. 1972; BUSBY and RADDA 1976; CAUDWELL et al. 1978), it was demonstrated that enzymes involved in glycogen metabolism were intimately associated with the glycogen particles, and that several of the control properties of the glycogen–protein complex correlated better with in vivo observations than those obtained from a purified mixture of the enzymes in solution. Over 95% of the protein attached to glycogen is accounted for by glycogen phosphorylase, glycogen synthase, debranching enzyme, phosphorylase kinase and protein phosphatase 1 (CAUDWELL et al.1978; COHEN 1978).

The discovery that phosphorylase is tightly bound to glycogen particles but can be activated and controlled, together with kinetic studies (WANG et al. 1965; METZGER et al. 1967), indicated that phosphorylase should have a separate glycogen binding site distinct from the catalytic site. Crystallographic studies showed that this is indeed the case. (WEBER et al. 1978; KASVINSKY et al. 1978). It is now known that glycogen synthase (LARNER et al. 1976) and phosphorylase phosphatase (PP-1; STRALFORS et al. 1985) have a glycogen binding site distinct from their catalytic sites, although no structural or sequence data are yet available to show whether they share common molecular features with the storage site of phosphorylase.

The function of the storage site in phosphorylase appears to be both to provide an anchor to localize the enzyme on the surface of the particle and to act as an additional control site. Preincubation of phosphorylase with glycogen enhances activity (WANG et al. 1965) and promotes dissociation of inactive tetramers to active dimers (METZGER et al. 1967). Kinetic studies (KASVINSKY et al. 1978) showed that the dissociation constant for maltoheptaose at the catalytic site is 20-fold greater than that for the storage site. The latter is similar to the dissociation constant for glycogen and has been estimated to be 1 m*M*

(KASVINSKY et al. 1978) or 2.6 m*M* (PHILIP et al. 1982). Prior occupation of the storage site appears to be an obligatory part of the kinetic mechanism.

Rabbit muscle phosphorylase can utilize either glycogen or oligosaccharide with similar catalytic rates at saturating substrate concentrations, but the relative affinity for glycogen is much higher than that for linear oligosaccharides (by about 400-fold HU and GOLD 1975). The discovery of the storage site suggested that immobilization of the enzyme on the glycogen particle produces an effective high concentration of substrate at the active site. Studies with a covalently bound phosphorylase–glycogen complex have shown that phosphorylase is active on its own carrier (SOTIROUDIS et al. 1978). However the storage site does not appear to play an obligatory role in determining the high specificity of phosphorylase for branched substrates. In studies in which an activated oligosaccharide was attached covalently to the storage site, PHILIP et al. (1982) found that the modified enzyme still had a higher affinity for glycogen than for maltoheptaose by nearly two orders of magnitude. However the initial rate of activity for the modified enzyme was 8–10 times that of the unmodified enzyme, indicating the contribution of the storage site to activation. It is interesting that the two nonregulatory phosphorylases from potato and *E. coli* do not have oligosaccharide storage sites, and that these enzymes have a preference for linear polysaccharide substrates (NAKANO et al. 1986; PALM et al. 1985; PHILIP et al. 1982).

3.2 Crystallographic Studies

The crystall structures of rabbit muscle glycogen phosphorylase *b* (WEBER et al. 1978; SANSOM et al. 1985; JOHNSON et al. 1988), and phosphorylase *a* (SPRANG and FLETTERICK 1979) have been solved. The crystallographic interpretations were greatly aided by knowledge of the complete amino acid sequence (TITANI et al. 1977). The structure of phosphorylase *b* has been refined at 1.9-Å resolution and the current R factor is 0.185 (K.R. ACHARYA et al., manuscript in preparation). The structure of phosphorylase *a* has been refined at 2.1-Å resolution, and the current R factor is 0.16 (R. FLETTERICK, private communication). The refinements have indicated an extra Ile residue at position 307; hence the numbering of residues from 308 onwards has been increased by 1 from that given in TITANI et al. (1977). The c-DNA sequence has indicated seven further minor changes (NAKANO et al. 1986). Both crystals have the enzyme in the T state, in the terminology of MONOD et al. (1965). The phosphorylase *b* crystals are obtained in the presence of 2 m*M* IMP. IMP is a weak activator that confers activity but does not result in the concomitant increase in affinity for substrate which is the remarkable feature of AMP activation (BLACK and WANG 1968). The phosphorylase *a* crystals are obtained in the presence of glucose, which is an inhibitor that favors the T state. (HELMREICH et al. 1967; KASVINSKY et al. 1978; WITHERS et al. 1979).

A schematic diagram of the phosphorylase *b* dimer is shown in Fig. 8. Each subunit (total molecular weight 97434; 842 amino acids) is approximately ellipsoidal. The fold is relatively compact, but there are many indentations and small protrusions from the surface that increase the accessible surface area.

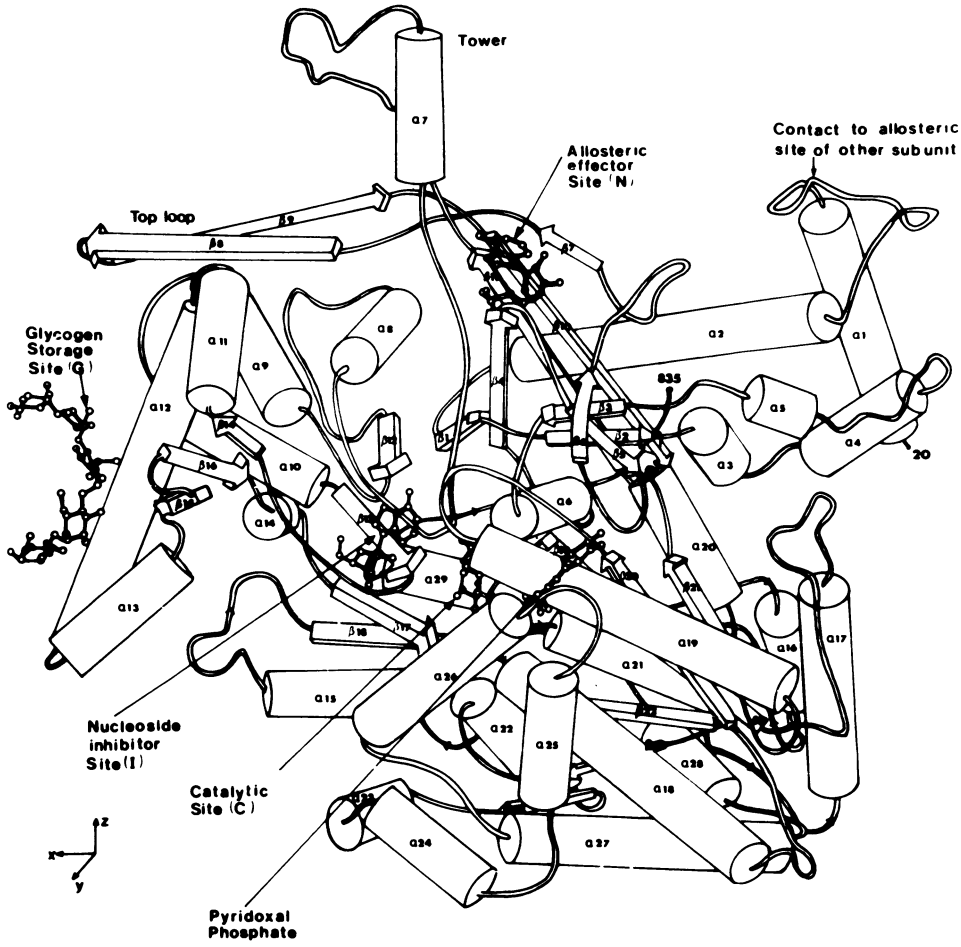


Fig. 8. Schematic diagram of the phosphorylase *b* monomer viewed down the crystallographic *y* axis. α-Helices and β-strands are represented by *arrows* and *cylinders*, respectively. Glucose 1-phosphate and pyridoxal phosphate (partially obscured by the structural elements of the protein) are shown bound at the catalytic site C, AMP at the allosteric effector site N, and adenosine at the nucleoside inhibitor site I. The five sugars of maltoheptaose which occupy the major site of the glycogen storage site. The minor site which is just below the β8–β9 turn is not shown

There are two main excursions away from the body of the subunit comprising the cap residues 36–45 and the tower residues 260–276. These make substantial contacts with the other subunit. In phosphorylase *b* the N-terminal 18 residues and the last residue at the C terminus are disordered. In phosphorylase *a* the seryl phosphate at position 14 leads to a substantial increase in the order of most N-terminal residues.

Phosphorylase is an α/β protein with 53% of amino acids in α-helix and 19% in β-sheets. The large polypeptide chain may be divided into domain 1 (residues 19–484) and domain 2 (residues 485–841). These regions exhibit self-contained contiguous folds, each centered on a core of parallel β-sheets (nine-

stranded for domain 1 and six-stranded for domain 2). The two domains interact with one another through numerous VAN DER WAALS interactions and hydrogen bonds and several ionic interactions. Domain 2 contains an internal region (residues 562–711) whose topology is identical to the nucleotide binding domain of lactate dehydrogenases and other enzymes. Phosphorylase contains an essential cofactor, pyridoxal phosphate, which is linked via a Schiff base to Lys 680. This lysine residue is located on α-21, which is equivalent to the α-E helix of the nucleotide binding domain, in the terminology of ROSSMANN et al. (1974). The pyridoxal phosphate is buried in the center of the molecule where domains 1 and 2 come together.

From binding studies in the crystal, four metabolite binding sites have been identified. The catalytic site (site C) is adjacent to the cofactor and is buried some 12 Å from the surface with access provided by a narrow channel. The allosteric effector site (site N) is located in domain 1 and is some 32 Å from the catalytic site at the subunit–subunit interface. The glycogen storage site (site G) is located entirely in a subdomain of domain 1 (residues 286–485) and is on the surface of the enzyme some 30 Å from the catalytic site and 39 Å from the allosteric site. Finally, a nucleoside or nucleotide inhibitor site (site I) is located at the entrance to the active site channel and comprises residues from both domains 1 and 2.

The distribution of these binding sites on the phosphorylase dimer is such that the catalytic, inhibitory, and glycogen storage sites are on one side (that nearest the viewer in Fig. 8), and the allosteric site is on the far side of the molecule. Thus, attachment of phosphorylase to the glycogen particle helps to localize the substrate but leaves the allosteric effector site free to interact with nucleotides or glucose 6-phosphate.

Phosphorylase has been shown to be active in the crystal with oligosaccharide substrates (KASVINSKY and MADSEN 1976), although the apparent K_m for maltoheptaose was extremely high (170 m*M*). In the crystal no equilibrium binding of oligosaccharide has been observed at the catalytic site either in phosphorylase *a* or phosphorylase *b*; despite the fact that both from the observations of KASVINSKY and MADSEN and from our own of catalysis in the crystal (HAJDU et al. 1987), it is known that oligosaccharide must visit the catalytic site. The transient binding is readily explained by the observation that access to the active-site channel is partially blocked by loops of chain comprising residues 281–288 and 133–136. Although monosaccharides bind well in the crystal, model-building studies show that movements of these loops are required to accommodate oligosaccharides at the catalytic site. These movements are also part of the allosteric response. Phosphorylase digests to within four sugar residues of a branch point (GRAVES and WANG 1972). The length of the active-site channel is such that it could accommodate a tetrasaccharide (after a conformational change) and place the α-(1–6) branch point on the surface.

Thus, although in principle the crystal studies on phosphorylase should allow the description of two distinct oligosaccharide binding sites, in fact only the glycogen storage site has been elucidated in detail. Nevertheless a study of the binding of monosaccharides at the catalytic site has provided useful information on protein–carbohydrate interactions, and a brief summary is given here.

3.3 Protein–Carbohydrate Interactions at the Catalytic Site

The catalytic site is located at the center of the molecule and includes residues from the C-terminal end of the "nucleotide binding domain" β-sheet (corresponding to ends of βA and βD strands) and the C-terminal end of strands 7 and 9 of the nine-stranded sheet at the core of domain 1. In this respect the catalytic site exemplifies features noted by BRANDEN (1980, 1986) namely that in α/β proteins the catalytic site is located at the carboxyl edge of a parallel β-sheet with active-site residues found in the loop regions. In other enzymes introns are frequently found in these loop regions, thus suggesting an evolutionary mechanism for functional variations at the edge of the sheet with a common structural motif.

Studies of phosphorylase *b* binding have led to descriptions of the following compounds of site C: glucose; deoxynorjirimycin; glucose 1-phosphate; glucosamine 1-phosphate; glucose 1,2-cyclic phosphate; heptulose 2-phosphate, uridine-diphosphate-glucose (UDPG); and two compounds with trigonal geometry at C1: heptenitol and D-gluconhydroximo-1,5-lactone-*N*-phenylurethane (JOHNSON et al. 1980; JENKINS et al. 1981; MCLAUGHLIN et al. 1984; HAJDU et al. 1987; and unpublished results). In phosphorylase *a* the binding of glucose and glucose 1,2-cyclic phosphate have been studied in detail (SPRANG et al. 1982; WITHERS et al. 1982).

For phosphorylase *b* the most informative studies have been those with heptenitol and heptulose 2-phosphate (MCLAUGHLIN et al. 1984; HAJDU et al. 1987). Heptenitol (HEHRE et al. 1980) has been introduced as a substrate for phosphorylase by KLEIN et al. (1984, 1986) as part of a study with glycosylic substrates in which the potential anomeric carbon atom is linked via an electron-rich bond. Phosphorylase catalyzes the utilization of 2,6-anhydro-1-deoxy-D-gluco-hept-1-enitol (heptenitol) in the presence of phosphate to form 1-deoxy-α-D-gluco-heptulose 2-phosphate (heptulose 2-phosphate). Heptenitol is used exclusively as a substrate for the degradative pathway, and the reaction itself does not require oligosaccharide primer. Heptulose 2-phosphate is a dead-end product and a potent inhibitor of the enzyme ($K_i = 14$ μM for the AMP-activated form of the enzyme). It is the strongest competitive inhibitor known for rabbit muscle phosphorylase *b*. These characteristics suggest that the compound has certain features of a transition-state analogue. In the crystals of phosphorylase *b*, heptulose 2-phosphate was formed by incubation of heptenitol and phosphate in the presence of AMP and oligosaccharide (as activators) in a series of time-resolved experiments (HAJDU et al. 1987).

The structure of the phosphorylase–heptulose 2-phosphate complex is shown in Fig. 9. Table 4 shows a comparison of hydrogen-bond interactions between phosphorylase *a* and the T-state inhibitor glucose, between phosphorylase *b* and the substrate heptenitol, which has a half-chair conformation, and phosphorylase *b* and the tight binding product heptulose 2-phosphate. The results show that, as expected, hydrogen bonds to the polar groups are important for monosaccharide binding at the buried catalytic site. The hydrogen bond potential of almost every polar group is satisfied. The glucosyl pyranose moieties bind in a similar fashion to both forms of the enzyme with slightly different

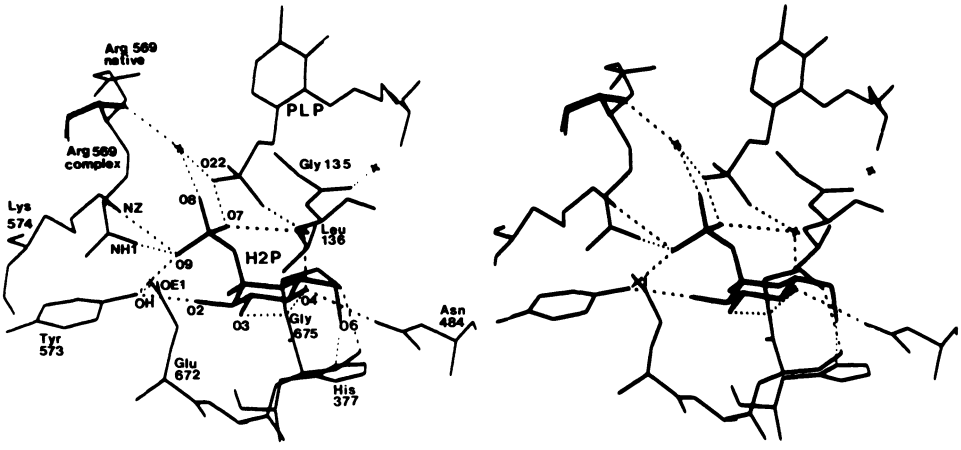


Fig. 9. A stereodiagram of the major contacts between phosphorylase *b* and heptulose 2-phosphate at the catalytic site of phosphorylase *b*. The positions of Arg 569 in the native and in the complex are shown

Table 4. Hydrogen bonds[a] to sugars at the catalytic site of phosphorylase

Atom number	Glucose[b]	Heptenitol[c]	Heptulose 2-phosphate[c]
α-O1	Leu 136 N (long)	–	Leu 136 N (long)
O2	Asn 284 ND2 Tyr 573 OH	Asn 284 ND2 Tyr 573 OH Glu 672 OE1	Tyr 573 OH Glu 672 OE1 (long)
O3	Glu 672 OE2 Ser 674 N	Glu 672 OE1 Ser 674 N (long) Gly 675 N	Glu 672 OE1 (long) Ser 674 N (long) Gly 675 N
O4	Asn 484 OD1 Gly 675 N	Asn 484 OD1 Gly 675 N Water 897.2	Asn 484 OD1 Gly 675 N
O5	His 377 ND1	Leu 136 N (long) His 377 ND1	Leu 136 N (long)
O6	His 377 ND1 Asn 484 OD1	His 377 ND1	His 377 ND1 Asn 484 OD1
O7			PLP 680 022 Water 897.2
O8			Water 879.1 Gly 135 N
O9			Arg 569 NH1 Tyr 573 OH Lys 574 NZ

[a] Hydrogen bond distances <3.3 Å; "long" distance, between 3.3 Å and 3.8 Å

[b] Glucose bound to phosphorylase *a* (from STREET et al. 1987; SPRANG et al. 1982)

[c] Heptenitol and heptulose 2-phosphate bound to phosphorylase *b* (HAJDU et al. 1987)

orientations but, nevertheless, a very similar hydrogen-bonding pattern. Comparison of the positions of heptenitol and heptulose 2-phosphate shows small differences in positions of the glucopyranose ring (separation of C1 atoms is 0.7 Å), which result mostly from the half-chair geometry of heptenitol. The major difference is the displacement of the loop of chain-carrying residues Asp 283 and Asn 284 by the side chain of Arg 569 in the heptulose 2-phosphate complex (HAJDU et al. 1987). Both His 377 and Glu 672 also make small but obvious movements away from this ligand to optimize contact distances that would otherwise be too short. Surprisingly, in the complex with heptulose 2-phosphate there are only weak interactions to the anomeric oxygen O1 and the ring oxygen O5. It is just at these positions that groups involved in catalysis and transition state stabilization might be expected. The lack of direct interactions supports the notion that catalysis is effected though indirect interactions with the cofactor phosphate, substrate phosphate, and other groups (JOHNSON et al. 1988). QUIOCHO (1986) has noted the high specificity contributed hy hydrogen bonds to carbohydrate-binding sites. He observes that the peripheral hydroxyls nearly always participate both as hydrogen bond acceptors and donors of the type

$$(NH)_n \rightarrow OH \rightarrow O$$

where NH and O are hydrogen bond donor and acceptor groups, respectively, and OH is a nonanomeric sugar hydroxyl and $n=1$ or 2. Table 5 shows that all the hydroxyls participate in such hydrogen bonds (if we assume OH Tyr 573 participates as an acceptor). As also observed by QUICHO for arabinose-binding protein, these polar residues are involved in extensive hydrogen bond networks to other residues of the protein and water molecules, and that they nearly all involve planar groups of atoms (His, Asn, Glu, or main-chain NH).

The sugars make VAN DER WAALS interactions (<4 Å) to residues: Gly 135, Leu 136, Leu 139, Asn 284, His 377, Val 455, Asn 484, Tyr 573, Glu 672, Ala 673, Ser 674, Gly 675. Each sugar atom makes between one and five VAN DER WAALS contacts, with atoms C1 and C2 making the fewest contacts. This is in contrast to the arabinose-binding protein, which exhibits a very high affinity for arabinose ($K_i = 4.1 \times 10^{-7}$ *M*) and has three to seven VAN DER WAALS contacts per atom (QUIOCHO 1986). The impression received for the phosphorylase sugar binding site is of a cavity at the center of the molecule contributing both polar and nonpolar interactions that match the corresponding components on the sugar. These interactions give some degree of flexibility to allow the accommodation of a glucopyranose ring with half-chair geometry (e.g., heptenitol), chair geometry (e.g., glucose), or chair geometry with a methyl group in the β configuration in addition to the α-linked phosphate ester (e.g., heptulose 2-phosphate).

STREET et al. (1986) have carried out binding studies in solution for a number of deoxy and fluoro-substituted glucose analogues. They show that α-D-glucose exhibits a three fold lower K_i than β-D-glucose ($K_i = 1$ m*M* and 3 m*M*, respectively) suggesting a preference for, but not exclusive binding of, the α anomer. Loss of either the 1 OH or 2 OH results in a loss of 1.5 kcal/mol in binding energy, but replacement of these hydroxyls by fluorine causes no change, indicat-

Table 5. Hydrogen bonds to oligosaccharides at the glycogen storage site of phosphorylase *b*

Site	Atom	Protein contacts
Major site		
S3		None
S4	O3	Ser 429 OG Val 431 O Glu 433 OE2
S5	O2	Glu 433 OE1 Lys 437 NZ
	O5	Asn 407 ND2 Water 901.2
	O6	Tyr 404 0
S6	O4	Gln 408 ND2
	O6	Water 908.6 – Gln 401 O
S7	O4	Water 862.1 – Gln 408 ND2
	O6	Glu 405 OE2
Minor Site		
S8	O3	Arg 398 NH1, NH2
	O4	Water 908.8
	O6	Water 861.9 – Arg 409 NH2 Water 908.7 – Arg 409 NH1
S9	O2	Water 848.8 – Ala 213 N Water 848.9 – His 208 ND1
	O3	Water 848.8 – Ala 213 N Water 849.0 – Gln 211 O Water 890.6 – Leu 359 O
	O5	Water 908.8
	O6	Water 861.9 – Arg 409 NH2

ing that these groups are acting as acceptors. Deoxygenation at C3, C4, or C6 leads to a large loss of affinity, which is not surprising in view of the interactions detailed in Table 6. In addition, our binding studies on phosphorylase *b* have shown that variations in relative affinity can be rationalized on the basis of contacts observed. For example, deoxynorjirimicin (1-deoxy-5-amino-D-glucose) is a poor inhibitor (ARIKI and FUKUI 1977). In the crystal, deoxynorjirimicin binds weakly, presumably because of some degree of unfavorable contacts between the main-chain NH Leu 136 and ND1 His 377 with the ring nitrogen of deoxynorjirimycin. On the other hand glucosamine 1-phosphate, which is a substrate for phosphorylase (ROMERO et al. 1980) and can be released from liver glycogen (KIRKMAN and WHELAN 1986), binds tightly to phosphorylase *b* in the crystal (NAJMUDIN 1984), probably because of the proximity of the 2-amino group to the acidic side chain of Glu 672.

Glucose 1-phosphate, glucose 1,2-cyclic phosphate, and heptulose 2-phosphate bind to phosphorylase *b* with their sugars in approximately the same

Table 6. Torsion angles for oligosaccharides at the glycogen storage site of phosphorylase *b*

Disaccharide	ϕ	ψ	O2(n)→O3(n-1) Å	References
Major site				
S4→S3	174	157	2.9	
S5→S4	135	154	4.0	
S6→S5	167	159	2.5	
S7→S6	159	171	3.0	
Minor site				
S9→S8	172	185	3.2	
Maltose	181	193		1
β-Methyl maltoside	169	190		2
β-Maltose	180	197		3
α-maltose	172	182		4
α-Phenyl maltoside (A)	170	160		5
α-Phenyl maltoside (B)	168	160		5
6-Iodo α-phenyl maltoside	128	146		5

ϕ is the torsion angle (°) about C1–O1 defined by C4′ O1 C1 O4; ψ is the torsion angle (°) about O1–C4′ defined by C1 O1 C4′ O1′. The position (0, 0) is when O4, C1, O1, C4′ and O1′ are in the same plane

References: [1] QUIGLEY et al. (1970); [2] CHU and JEFFREY (1967); [3] GRESS and JEFFREY (1977); [4] TAKUSAGAWA and JACOBSON (1978); [5] TANAKA et al. (1976)

positions but with different phosphate binding positions. In heptulose 2-phosphate, because of the methyl group in the β configuration, the torsion angle O5-C1-O1-P, is approximately −136°, which places the phosphate in direct contact with the pyrixoal phosphate 5′-phosphate group (P–P separation 5.1 Å; MCLAUGHLIN et al. 1984; HAJDU et al. 1987; Table 6). In glucose 1-phosphate the O5-C1-O1-P torsion angle is 117°, a value similar to that found in the single crystal structure (BEEVERS and MACONOCHIE 1965; NARENDRA et al. 1984). The phosphorus is thus placed more than 7 Å from the cofactor phosphorus. In glucose 1,2-cyclic phosphate the phosphorus is constrained to a position nearer to that observed in heptulose 2-phosphate than in glucose 1-phosphate and the P–P separation to the cofactor phosphorus is likewise intermediate (JENKINS et al. 1981). The phosphate recognition site is therefore much less well defined in the T-state structure of phosphorylase *b* and allows phosphate groups to be placed in positions that differ by as much as 2.6 Å. The changes observed with heptulose 2-phosphate represent the initial stages of the allosteric conformational response that leads to an increase in affinity for the phosphate moiety through the creation of the phosphate recognition site and the simultaneous opening of the active-site cleft to allow access of the oligosaccharide (HAJDU et al. 1987). The proximity of the heptulose 2-phosphate phosphate to the cofactor phosphate is of significance for the catalytic mechanism, but a detailed discussion of the mechanistic proposals (see MADSEN and WITHERS 1986; JOHNSON et al. 1988) is outside the scope of this review.

Despite its internal location the catalytic site contains a number of well-localized water molecules; in particular, the cofactor 5′-phosphate is fully solvated. In the phosphorylase *b*–heptulose 2-phosphate complex two water molecules are displaced, one by the O3 atom and one by the phosphate group. Although the resolution (3 Å) is insufficient to determine the new water structure, the difference map suggests that, apart from the above waters, there is very little perturbation. Some of these waters participate in VAN DER WAALS interactions.

The phosphorylase catalytic site for the terminal glucosyl residue differs from the carbohydrate binding site of other proteins and enzymes in that no aromatic group is involved in significant VAN DER WAALS intreractions. His 377 and Tyr 573 participate almost exclusively as hydrogen bond donors or acceptors. There is no stacking of the carbohydrate nonpolar face against an aromatic residue, as is observed in lysozyme and arabinose-binding protein, for example. However, if an oligosaccharide is extended along the active site towards the exterior (which requires some significant conformational movements of the protein), then the fourth or fifth sugar could interact with Tyr 613 at the entrance to the cleft. In the T state, Tyr 613 is stacked close to Phe 285, and together these two aromatic residues form the nucleoside inhibitor site.

3.4 Protein–Carbohydrate Interactions at the Glycogen Storage Site

The most detailed structural studies have resulted from binding experiments in the crystal with maltoheptaose. A preliminary account of oligosaccharide binding to phosphorylase *b* has been given by JOHNSON et al. (1983), but the detailed description given below has awaited the high-resolution refinement of the native protein and the maltoheptaose complex. GOLDSMITH et al. (1982) have described a maltoheptaose conformation bound to phosphorylase *a* and used the results to propose a model for glycogen. Further details of the interactions at an intermediate stage in the refinement have also been given (GOLDSMITH and FLETTERICK 1983).

The structure of α-(1–4) linked polymers of D-glucose can be described by the torsion angles about the glycosidic linkages C1–O4′ and O4′–C4′ where superscript prime refers to the adjacent sugar. Energy calculations (e.g., SUNDANARAJAN and RAO 1969; REES and SMITH 1975) and observations on the single crystal structures of maltose (e.g., CHU and JEFFREY 1967; QUIGLEY et al. 1970; GRESS and JEFFREY 1977; TAKUSAGAWA and JACOBSON 1978) show that the favored conformation is one in which the O2 hydroxyl is hydrogen bonded to the O3′ hydroxyl of the adjacent sugar. This gives rise to a variety of helical conformations for amylose polymers (e.g., WU and SARKO 1978). Significant perturbations from preferred geometry may occur, however, in the presence of additional constraints, as observed in the single crystal structures of 6-iodo phenylmaltoside (TANAKA et al. 1976) and cyclohexamylose hexahydrate (MANOR and SAENGER 1974; SAENGER 1979).

The glycogen storage site is situted on the surface of the molecule, well removed from other sites and the subunit–subunit interface (Fig. 10). The site is composed of a major site and a minor site. At the major site the interactions

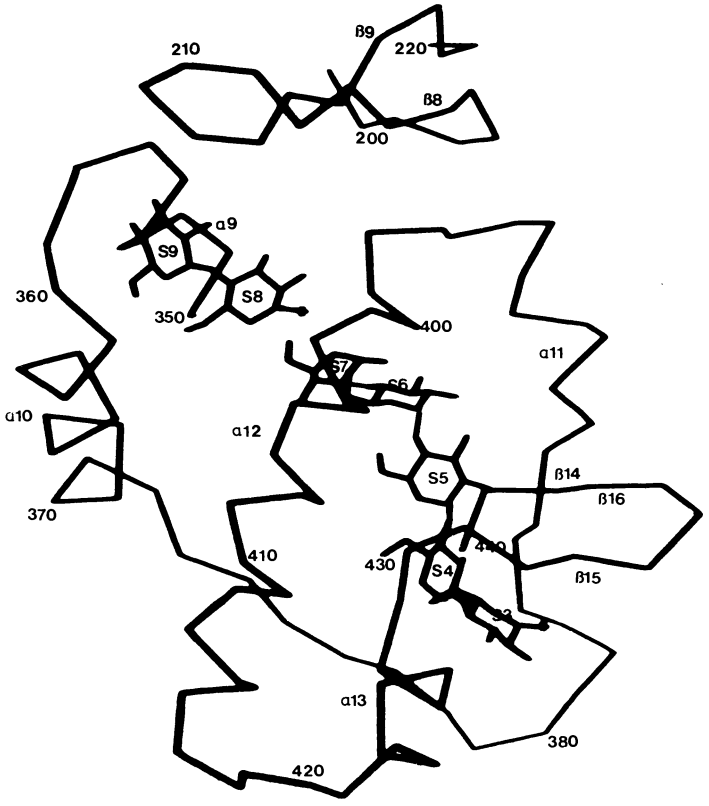


Fig. 10. Maltoheptaose bound at the glycogen storage site of phosphorylase *b*. The view shows residues 199–220 and residues 350–440 Cα atoms only to demonstrate the relationship between the major and minor sites and the secondary structural elements

are to residues of the α12-helix and the spur formed by the small antiparallel sheet (β15–β16). Thus, all the contacts to the oligosaccharide are contained in the local stretch of chain residues 398–437 (α12–α13–β15–β16). It will be interesting to see if these local regions of secondary structure form an archetypal glycogen binding site for other enzymes. In phosphorylase *b* these residues make extensive side chain–side chain contacts to other secondary structural elements. The minor site consists of only two glucosyl residues, and it lies above the nonreducing end of the major site, making contacts to the top of the α12-helix, the loop of antiparallel β-sheet from β8–β9, and one contact to a residue from α9. Thus, the elements of this site are composed of residues from different parts of the polypeptide chain. Both sites utilize residues from a turn between two antiparallel β-strands, but otherwise they have no significant features in common.

The crystal structure of the phosphorylase *b*–maltoheptaose complex at 2.5 Å resolution has been refined by restrained least-squares crystallographic refinement methods (McLaughlin et al. manuscript in preparation). The coordinates

for α-D-glucose were taken from ARNOTT and SCOTT (1972). The resulting crystallographic R value is 0.146, with root mean-square deviation from ideal bond lengths of 0.025 Å. Only five glucosyl units are identified at the major site, labeled S3–S4–S5–S6–S7, where the reducing end of the oligosaccharide is in site S3, and this site has the lowest crystallographic z axis. The numbering system follows that of GOLDSMITH and FLETTERICK (1983) where in phosphorylase *a* two additional sites were located at the reducing end but were not in contact with the protein.

The phosphorylase *b*–maltoheptaose complex is shown in Fig. 11. Hydrogen bonds are listed in Table 5. The sugar in S3 makes no contacts (<4 Å) with the protein but participates in a good intramolecular hydrogen bond (O3–O2′) to the sugar in S4. The sugar in S4 makes specific hydrogen bonds from its O3 hydroxyl to the side chain of Ser 429, Glu 433, and the main-chain carbonyl of Val 431 and is further stabilized by VAN DER WAALS interactions with Leu 411,

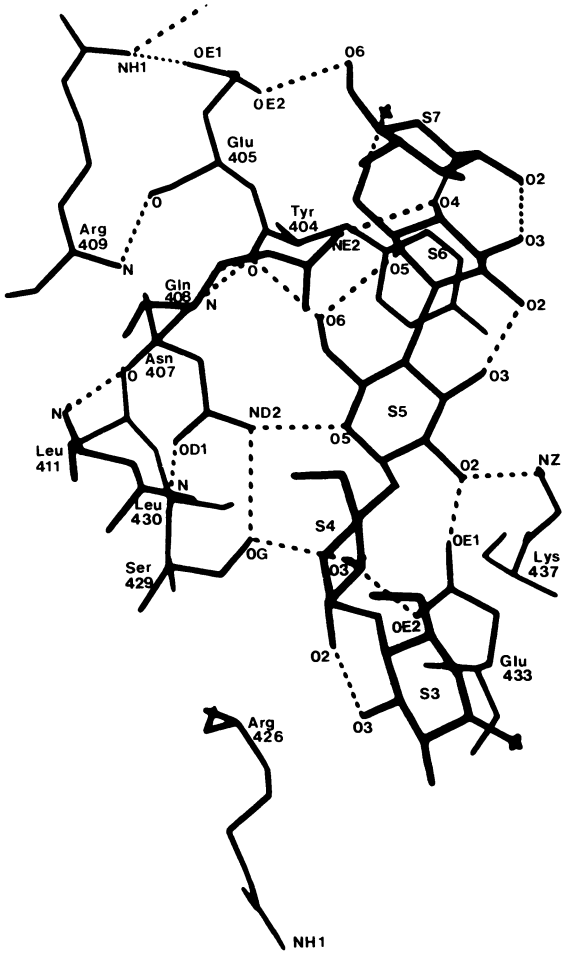


Fig. 11. Interactions between maltoheptaose and phosphorylase *b* at the major site. Val 431 has been omitted for clarity. Only five glucosyl sites are occupied. Glu 433 spans sites S4 and S5; Tyr 404 spans sites S5 and S6. Residues Glu 405, Asn 407, and Arg 409 are involved in hydrogen-bond networks through side-chain interactions in addition to their main-chain α-helical interactions

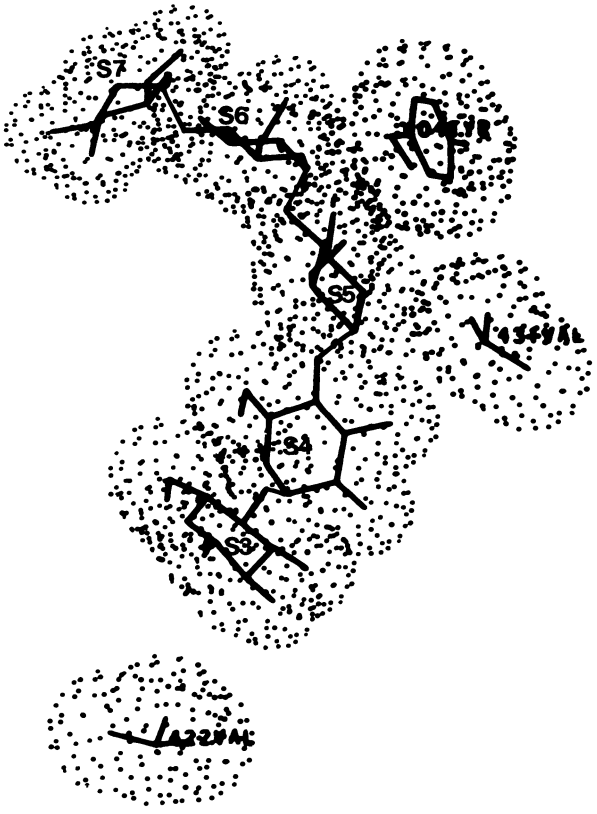


Fig. 12. The VAN DER WAALS interactions of three important nonpolar residues with maltoheptaose at the major glycogen storage site. Tyr 404 stacks between S5 and S6, Val 431 to S5 and partly S4. Val 422 is shielded by the oligosaccharide but is not in direct VAN DER WAALS contact

Arg 426, Ser 429, and Glu 433. There is no intramolecular hydrogen bond between S4 and S5. The O2 hydroxyl of the sugar in S5 is hydrogen-bonded to Glu 433 and Lys 437. It is clear from Fig. 12 and Table 6 that S5 is the strongest site. Every available functional group on the sugar in this site is involved in hydrogen bonds, and there are good VAN DER WAALS contacts with Tyr 404, Asn 407, Gln 408, Val 431, Glu 433, and Lys 437. The sugar is almost inaccessible to solvent when bound at site S5. The geometry between site S5 and S6 is standard as is that between sites S6 and S7, with good intramolecular hydrogen bonds. The sugars in sites S6 and S7 make two hydrogen bonds directly to the protein and two further interactions through water molecules.

Tyr 404 makes the most significant contacts with the sugar in S6 and the glycosidic linkage between S5 and S6. The aromatic residue fits into the groove formed by the glycosidic linkage with the lone pair of electrons on the oxygen directed away from the tyrosine. The ring oxygen atom O5 and the hydrogens of C1 and C2 of the sugar in site S6 and C4 of the sugar in site S5 are directed towards the aromatic ring.

Overall, the left-handed oligosaccharide curls in toward the protein and then away, making the most extensive interactions in sites S4, S5, and S6, with the helical structure (like a roll of cellotape) gripped between a thumb (β15 to β16 sheet) and a forefinger (α12). Analysis of the solvent accessible surface area of the protein shows that the oligosaccharide covers twice as much surface accessible area associated with apolar atoms than with polar atoms, although the environment of the sugar binding site is not significantly more apolar than the rest of the protein. Val 422, Leu 425, and other residues form a nonpolar patch that is not in direct contact with the oligosaccharide but is made inaccessible by it (Fig. 12). This patch contributes some 16.5% of the total area lost in ligand binding.

Small side-chain movements are observed on binding oligosaccharide for residues from the α12-helix for Gln 401, Gln 408, and Arg 409. Tyr 404 makes no movement and appears to be in the right orientation to accept the oligosaccharide. More extensive movements are observed for the whole of the loop of chain 425–437. These changes do not significantly alter the major contacts to the protein. For example, the ion pairs between Glu 432 and Arg 386, Asp 470 and Arg 427, and Arg 398 and Asp 355 are preserved. The movements, especially of the external residues Glu 433 and Lys 437, allow optimal contact with the oligosaccharide and have been observed in all oligosaccharide–phosphorylase complexes.

In general, the contacts observed for maltoheptaose binding to the major storage site of phosphorylase *b* are similar but not identical to those described for phosphorylase *a* (e.g., GOLDSMITH et al. 1987). However, the conformational changes in phosphorylase *b* are less extensive than in phosphorylase *a,* where it was necessary to stabilize the crystals with glucose and caffeine; for example, movement of Tyr 404 was observed.

The torsion angles for the oligosaccharide are shown in Table 6 and compared with standard values. It is seen that the oligosaccharide has the preferred helical conformation except for the linkage between sugars S4 and S5. Here the conformation is similar to that of 6-iodo phenyl maltoside (TANAKA et al. 1976), apparently because of the strong hydrogen bonds made to Glu 433, Lys 437, and Ser 429. It is significant that, in this instance, strong contacts with the protein have caused a deviation from preferred geometry.

An outstanding feature of the glycogen storage site is that, although it has high affinity for α-(1–4) linked oligosaccharides, it has no specificity for glucose or glucose 1-phosphate (JOHNSON et al. unpublished results). The pattern of substitution observed from binding studies is shown in Table 7. From the frequent involvement of aromatic residues in sugar binding it might have been anticipated that maltose would bind in sites S5–S6. Instead, it binds in sites S4–S5, showing the importance of the hydrogen bonds to the ionizable residues Glu 433 and Lys 437 in determining the specificity.

Maltotriose binds to subsites S4–S5–S6, confirming the higher affinity of these internal sites compared with the peripheral sites. Maltotetraose binds S3–S4–S5–S6, preferring to shield to some extent Val 422, Leu 425, and the aliphatic components of Glu 433 and Arg 426 rather than to occupy subsite G7, which requires movement of Gln 408. This preference is easily perturbed, however,

for acarbose, a modified maltotetraose in which the two sugars at the nonreducing terminus are replaced by a hydroxymethyl conduritol residue linked α-(1–4) to a 4-amino-4,6-dideoxyglucose residue (TRUSCHUIT et al. 1981), binds to sites S4–S5–S6–S7 (Fig. 14). This demonstrates the importance of the hydrogen bond between O6 of site S5 and the main-chain carbonyl oxygen of Tyr 404. If acarbose were to bind like maltotetraose, it would place the 6-deoxy sugar in site S5; it would not only be unable to make the hydrogen bond, but it would

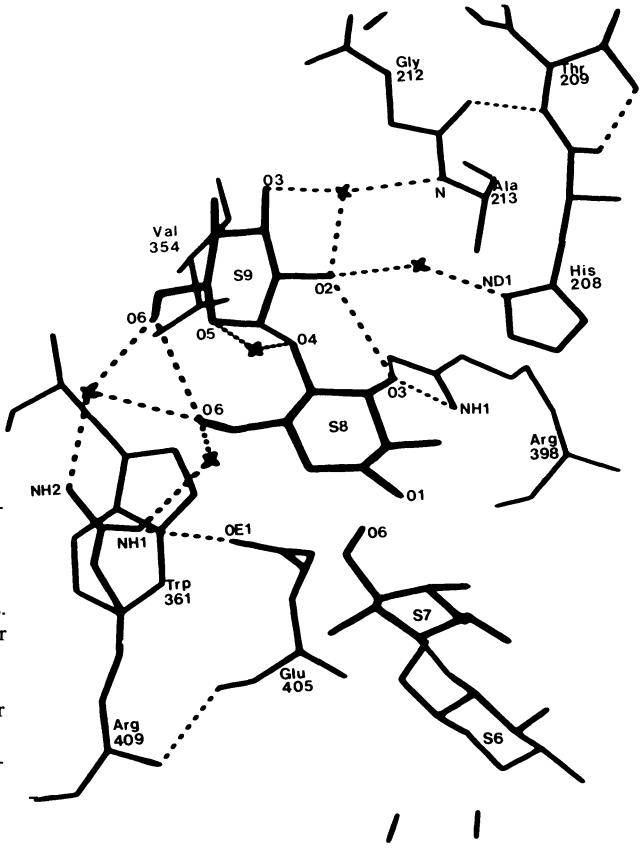


Fig. 13. Interactions between maltoheptaose and phosphorylase *b* at the minor glycogen storage site. Only two sites, S8 and S9, are occupied. Water molecules are shown as crosses. Sites S6 and S7 from the major storage site are also shown. His 208, Arg 398, and Arg 409 make the most important polar interactions and Val 354 the most important nonpolar interaction. Trp 361 is partially shielded from the sugar by Arg 409

Table 7. Occupation of subsites at the major glycogen storage site of phosphorylase *b*

	S3	S4	S5	S6	S7
Maltoheptaose	V	V	V	V	V
Maltopentaose	V	V	V	V	V
Maltotetraose	V	V	V	V	
Maltotriose		V	V	V	
Maltose		V	V		
Acarbose		V	V	V	V (conduritol ring)

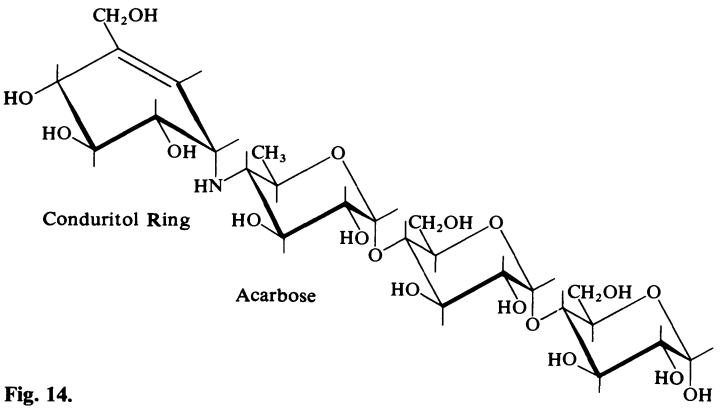


Fig. 14.

also leave a hole. Acarbose binding to phosphorylase *a* has been described recently (GOLDSMITH et al. 1987), and a comparison with solution studies reported.

Binding at the minor site (Fig. 13) is slightly weaker than at the major site, and, regardless of which oligosaccharides are used in the binding study, only two sites are occupied. The interactions of this site have not been described previously. The site comprise the loop from $\beta 8$ to $\beta 9$ from above, with Trp 361, Val 354, Glu 405, and Arg 409 from below. The sugar in S8 makes hydrogen bonds between O3 and the side chain of Arg 398 and between O6 through a water to Arg 409. The sugar in S9 makes a hydrogen bond from O2 through a water to the main-chain nitrogen of Ala 213 and through another water to His 208, O6 hydrogen bonds through a water to Arg 409. Although Trp 361 forms part of the site and is coplanar with the sugar in S8, it is partially shielded from direct interaction by the side chains of Glu 405 and Arg 409 and only makes VAN DER WAALS contact with C6. Val 354 makes good contact with C2 of the sugar in site S9. The disaccharide has the preferred conformation (Table 6), and there are good intramolecular hydrogen bonds between O3 of S8 and O2 of S9 and between the O6 hydroxyls of the two sugars. It is noteworthy that both the arginine side chains at this site are held in place by charge or polar contacts (Arg 409 to Asp 362 and Glu 405; Arg 398 to Asp 355 and Gln 401).

The C1 of the sugar in site S8 is approximately 3 Å from the O6 hydroxyl at the top of the major site. Model-building studies with molecular graphics have shown that an α-(1–6) branching glucosyl residue could bridge sites S7 and S8. The minor site is sensitive to the occupation of other sites on the protein and does not appear to be filled when AMP is bound at the allosteric site (JOHNSON et al. 1983), nor is it filled in the substrate complexes with heptulose 2-phosphate (HAJDU et al. 1987).

The catalytic site and the storage site exhibit the common feature of extensive hydrogen bonding between the protein and the glucopyranose OH groups of the most tightly bound sugars. In almost every other feature the catalytic site and the storage site are different. The catalytic site is buried; the storage site is exposed. The catalytic site involves residues from the carboxyl terminal ends

of parallel β-sheet regions; the storage site involves residues from α-helices and the loop connecting two antiparallel strands. The catalytic site has no nonpolar interactions with an aromatic group; an aromatic group (Tyr 404) provides an important interaction at the storage site. However, these differences are hardly surprising in view of the functional differences of the two sites. The role of the catalytic site is to bring oligosaccharide and phosphate together so that phosphorylysis and not hydrolysis occurs. This involves a buried site removed from the bulk solvent and probably some distortion at the site of action of the α-(1–4) linked polysaccharide from its preferred geometry. The storage site, in contrast, must be exposed to locate the enzyme on the surface of glycogen and is designed specifically to exclude glucose by demanding contacts to O3 and O2′ hydroxyls of adjacent sugars.

4 Amylases

Starch and related polymers are universal sources of dietary carbon throughout the plant and animal kingdoms. The amylases which digest these polysaccharides are widely distributed in living organisms and form a diverse collection of enzymes with differing physical properties, specificities, and mechanisms (THOMA et al. 1971; TAKAGI et al. 1971). The starch industries depend on the use of amylases, and these enzymes have therefore been widely studied not only for their academic interest but also for their industrial applications. The crystal structures of two amylases have been solved: Taka-amylase A from *Aspergillus oryzae* (478 amino acids) at 3 Å resolution (MATSUURA et al. 1984) and α-amylase from porcine pancreas molecular weight approximately 53000) at 5 Å resolution (PAYAN et al. 1980). These preliminary studies already provide an interesting comparison with existing oligosaccharide-binding proteins, and more data should be available in the near future.

Taka-amylase A (E.C. 3.2.1.1, α-1,4-glucan 4-glucohydroloase) catalyzes the endohydrolysis of starch. Thus the reaction is similar to that with lysozyme in that an endohydrolytic reaction is catalyzed. On the other hand, the substrate α-(1–4) linked glucopyranoses is similar to that of phosphorylase. The 478 amino acid chain is folded into two domains. Domain A (residues 1–380) is composed of a $(\beta\alpha)_8$ structure similar to that of triosephosphate isomerase, pyruvate kinase, glycolate oxidase, aldolase, and other enzymes. Domain B (residues 381–478) is an eight-stranded antiparallel β-sandwich structure. A covalently bound carbohydrate group that is somewhat disordered in structure has been located at Asn 197. The catalytic site has been identified as a large cleft at the carboxyl end of the parallel β-barrel of domain A. An essential calcium ion is buried near the cleft and possibly plays a role in stabilizing the cleft architecture.

The active site has been located in binding studies with maltotriose. Only a disaccharide maltose was observed in the difference density map indicating that some hydrolysis might have taken place during the 6 days of the experiment. The sugar at the reducing end of maltose (site 4) makes several important hydrogen bonds. The O6 interacts with His 122, an interaction that may be essential

for catalysis since 6-substituted phenyl-α-maltosides are not hydrolyzed. The O2 and O3 hydrogen bond to Asp 297 and His 296, respectively. The reducing hydroxyl is directed toward Glu 230, and because of its environment (nonpolar) and position MATSUURA et al. (1984) propose that this glutamic acid residue could play the same role as a general acid in the catalytic mechanism as Glu 35 does in lysozyme. A second acidic residue, Asp 297, which is in a hydrophilic environment, could play the role of an electrostatic stabilising group similar to that played by Asp 52 in lysozyme.

Model-building studies have shown that the cleft can accommodate seven sugars with the observed maltose binding site extended by three sugars at the reducing end and two sugars at the nonreducing end in a model described as a curved binding model characteristic of the helical structure of α-(1–4) linked glucopyranoses. This model places the cleavage point between the fourth and fifth subsites as indicated from kinetic studies.

No details are avaliable for VAN DER WAALS interactions as yet, but the active site does include Tyr 82, which is in contact with the α-(1–4) glucosidic bond of the maltose (Y. MATSUURA, private communication), an interaction that is reminiscent of the interaction seen at the glycogen storage site of phosphorylase. The indole ring of Trp 83 lies close to the sugar in site 3 possibly making both VAN DER WAALS contact and a hydrogen bond between O6 of the glucose and NE of the tryptophan, an interaction similar to that observed in the lysozyme–chitotriose complex.

The low-resolution crystal structure of porcine amylase (PAYAN et al. 1980) shows the enzyme to contain a deep cleft that runs for 30 Å on one side of the molecule and separates two globular units which are very different in size. Substrate binding studies with maltotriose show two binding sites per molecule. One of these is situated in the deep cleft and has been assigned as the active site. The second site is on the surface. The observation of two binding sites is in agreement with deductions from solution studies (LOYTER and SCHRAMM 1966). It is too early to draw comparisons between the two amylase structures or to compare the surface binding site to the glycogen storage site of phosphorylase.

The structure of an α-amylase inhibitor, Hoe-467A, a 74 amino acid protein, has recently been solved (PFLUGRATH et al. 1986), and cocrystals with amylase have been reported. This should form an interesting system for future studies.

5 Summary and Conclusions

The high-resolution structural studies on oligosaccharide binding to lysozyme, phosphorylase, and amylase allow some general comments on the nature and diversity of the recognition sites. Although the data base is very small, some unifying principles emerge, and it will be interesting to see whether these are more universally applicable when more structures become available.

5.1 Topology of Oligosaccharide Recognition Sites

There appear to be very limited similarities in the topology of the protein chain at the oligosaccharide recognition sites. In lysozyme the oligosaccharide binding site is composed of residues from the antiparallel sheet, irregular loop regions, and the N and C termini of α-helices (Fig. 2). In phosphorylase at the catalytic site, the residues come from loops between parallel β-strands and α-helices (Fig. 8 and Fig. 9); in phosphorylase at the glycogen storage major site, the residues come from an α-helix and an antiparallel sheet, and at the glycogen storage minor site residues also come from an antiparallel sheet and α-helices but with different dispositions to those at the major site (Fig. 10). In amylase the site is located at the carboxyl end of a sheet that is part of an eightfold (β-α) barrel. The only common feature among these five sites is the utilization of residues from loop regions associated with either reverse turns or more extended irregular conformations. Lysozyme and phosphorylase (glycogen storage site: major site) both use residues from two adjacent strands of an antiparallel sheet (in lysozyme Asp 52 and Asn 59 span sites D and C; in phosphorylase Glu 433 and Lys 437 span sites S4 and S5). The Cα atoms of these residues are separated by about 4.4 Å, which allows the groups to span an oligosaccharide site. At the phosphorylase glycogen storage site the involvement of the long, exposed α-helix is a dominant feature and residues from three turns of this helix make important interactions with the oligosaccharide helix. This feature is not observed for the other oligosaccharide-binding proteins.

5.2 Oligosaccharide Conformation

Lysozyme recognizes β-(1–4) linked *N*-acetyl-D-glucosyl polymers. β-(1–4) linked polysaccharides have ribbonlike conformations in which there is an approximately 180° twist between adjacent sugars generated by rotations about C1–O4′ and O4′–C4′ bonds of the glycosidic link. In cellobiose (JACOBSON et al. 1961) there is an intramolecular hydrogen bond between O5 of one sugar and O3′ hydroxyl of an adjacent sugar, but in chitobiose (MO and JENSEN 1978) this hydrogen bond is less favorable. $(GlcNAc)_3$ when bound to lysozyme exhibits the intramolecular hydrogen bond between the residues in B and C, but the length between residues A and B is rather long (3.5 Å) for a good hydrogen bond (Sect. 2.3; Table 2). In the lysozyme cell wall complex NAM residues are excluded from sites A and C because of the presence of the bulky lactyl group on the 3 position of the sugar, but NAM residues are allowed in site B where the 3 position is exposed. The presence of the lactyl group, however, interferes with the intramolecular hydrogen bond between GlcNAc and MurNAc residues in sites A and B, respectively. Thus, it is interesting that the homopolymer trisaccharide when already bound to the enzyme exhibits a perturbation towards a conformation required for the heteropolymer. In summary, the trisaccharide when bound to lysozyme shows perturbations from the "ideal" conformation. These perturbations are not very much larger than those already

observed in single crystal studies of, for example, chitobiose. They are essential if the oligosaccharide site is to accommodate NAM sugars.

Phosphorylase recognizes α-(1–4) linked glycosyl polymers. The polysaccharides have approximate left-handed helical conformations generated by rotations about the C1–O4′ and O4′–C4′ bonds of the glycosidic link (e.g., 6.6 residues per turn; rise per residue, −2.3 Å; twist per residue, −55°; GOLDSMITH and FLETTERICK 1983). In the preferred conformation there is an intramolecular hydrogen bond between the O2 hydroxyl of one sugar and the O3′ hydroxyl of an adjacent sugar. In the maltoheptaose binding studies the oligosaccharide is bound in the preferred conformation at both the major and the minor glycogen storage sites with the important exception of sites S4 and S5. These are the strongest binding sites, and the interactions with the protein cause significant perturbation of the preferred geometry while still permitting the oligosaccharide to adopt a stereochemically allowed conformation (Sect. 3.4; Table 6). Thus, in both oligosaccharide binding to lysozyme and to phosphorylase there are small perturbations from ideal geometry caused, presumably, by protein interactions.

5.3 Solvent Accessibility

The calculated solvent accessibility of the oligosaccharide subsites in lysozyme and phosphorylase correlates well with the experimentally observed binding patterns for oligosaccharides of different lengths and with the number of hydrogen bonds and VAN DER WAALS contacts made. Thus, in lysozyme the sugar in site C is almost totally buried (loss of 75% of free accessible surface area in making contact with the enzyme) and those in sites B and A progressively less so (Sect. 2.3). In phosphorylase (Sect. 3.4) the sugar in S5 is almost totally buried (loss of approximately 86% of free accessible surface area). The sugars in S4 and S6 are less buried (approximately 58% and 48%, respectively), whereas the two peripheral sites, S3 and S7, lose only 21% and 14%, respectively, of their free accessible surface area on forming the complex. These values can be compared with that for a sugar at the catalytic site of phosphorylase, where the change in solvent accessible area is 94%. At the minor site the change in solvent accessible area is 67% and 52% for the sugars in S8 and S9, respectively. As expected, the sugars at the end of an oligosaccharide chain have a greater accessible area both in the free and the bound forms, since these are shielded only on one side by an adjacent sugar. In absolute terms, the change in area shielded on forming the complex is less for the peripheral sites S3 and S7 than for the interior site S5 for the major site; for the minor site, the change in area for both S8 and S9 is comparable to that of S5.

If the comparison is restricted to the trisaccharide bound to lysozyme and the five sugars observed at the major storage site in phosphorylase, both oligosaccharide recognition sites are seen to be dominated by a single site (site C in lysozyme; site S5 in phosphorylase). In phosphorylase, site S5 although dominant does not provide sufficient specificity to recognize a monosaccharide, and disaccharide binding is achieved with sites S4 and S5. In lysozyme, both site

C and site E appear to play significant roles in recognition of the hexasaccharide substrate (IMOTO et al. 1972).

5.4 Polar Interactions

As noted by QUIOCHO (1986), the sugar recognition sites involve extensive hydrogen bonding to polar groups in the protein. In lysozyme at site C (Table 3) every peripheral hydroxyl group and the acetamido side chain are involved in hydrogen bonds to planar groups on the protein. The interaction of the acetamido group with the main-chain atoms is highly directional and plays a key role in specificity. For the trisaccharide only one ionizable group on the protein is involved (Asp 101 : site B), and each sugar polar group is involved in only one hydrogen bond. At the catalytic site in phosphorylase, in contrast, ionizable groups (His 377 and Glu 672) play a major role. Every hydroxyl group is involved as a donor and as an acceptor in hydrogen bonds to planar groups that form hydrogen bond networks to other groups on the protein (Table 4). At the storage site in phosphorylase ionizable groups (His 208, Arg 398, Arg 409, Glu 433, and Lys 437) also play a dominant role. Other groups involved include main-chain atoms, asparagine, serine, and tyrosine residues (Table 5). At the minor site of the glycogen storage site, all but one of the interactions are through water molecules. In amylase, ionizable groups (Asp, His, Glu) are important in recognition of the disaccharide (Sect. 4).

5.5 Nonpolar Interactions

Nonpolar contacts appear almost equally as important as polar groups in providing specificity for the sugar binding sites. In lysozyme the aromatic side chain of Trp 62 stacks against the nonpolar face of the sugar in site B, making VAN DER WAALS contacts with C1–H, C3–H, and C5–H atoms of the sugar (Sect. 2.3). In phosphorylase (Sect. 3.4) the aromatic side chain of Tyr 404 stacks across the glycosidic linkage between sites S5 and S6 and interacts with the C4–H atoms in site 5 and C1–H atoms in site 6. At the minor site Trp 361 makes some contribution. In amylase both tyrosine and trypotophan residues are involved (Sect. 4). These aromatic group – sugar interactions are important, but it is interesting that they are not the dominant features in specificity. In lysozyme monosaccharides bind preferentially in site C and do not utilize the stacking with Trp 62. In phosphorylase maltose binds in sites S4 and S5 and only partially uses the interaction with Tyr 404. At the catalytic site of phosphorylase there is no aromatic group that stacks against the sugar ring (Sect. 3.3).

Valine and isoleucine residues make important contributions. In lysozyme Val 109 and Ile 98 and in phosphorylase (storage site) Val 422 and Leu 425 are shielded from solvent by oligosaccharide binding. However, in phosphorylase (storage site) Leu 411, Val 431, and Val 354 make rather more interesting interactions with sugars in sites S4, S5, and S9, respectively. In each case a methyl group from the side chain packs neatly against the nonpolar A face

of the sugar containing the C–H hydrogens at positions 1, 2, and 4 on the pyranose ring. This nonpolar docking of α-(1–4) linked sugars against nonpolar aliphatic side chains appears to be an important feature for each of these sites. In addition, the aliphatic components of polar side chains such as arginine and glutamic acid make useful interactions.

Undoubtedly the shielding of these nonpolar groups on the protein by the complementary surface of the sugar makes a significant contribution to the binding energy. Yet comparison of the surfaces of both lysozyme and phosphorylase show that the surface is not significantly more apolar than the rest of the molecule. In the maltodextrin phosphorylase from *E. coli* (PALM et al. 1985), which exhibits 41% sequence homology with the rabbit muscle enzyme, there are several amino acid substitutions of critical residues that could explain the lack of glycogen storage site for this enzyme. Thus Tyr 404 → Asn, Leu 411 → Lys, Val 422 → Glu are all changes that would interfere with the nonpolar nature of the site, whereas the change Glu 433 → Asp would disrupt two of the key polar interactions.

The nonpolar shielding of the A and B faces for α- and β-linked glucosyl polymers, respectively, is in good agreement with those features of oligosaccharides noted in the introduction.

5.6 Disaccharide Specificity

Both lysozyme and phosphorylase (at the storage site) exhibit disaccharide specificity. This specificity is generated in different ways for β-(1–4) or α-(1–4) oligosaccharides. In lysozyme, action against the bacterial cell wall substrate leads to release of the disaccharide NAG–NAM as the smallest unit. NAM residues with the bulky lactyl group are excluded from site C and permitted in sites B and D (Sect. 2). In this case, the stereochemical contacts exclude a sugar from a particular site in the heteropolymer that leads to the disaccharide product. In phosphorylase, the high specificity for maltose and the lack of binding of glucose at the storage site are accounted for in a different way. Maltose binds in sites S4 and S5, making strong interactions through the O3 and O2 hydroxyls, respectively, to ionizable groups on the protein, and there are accompanying conformational changes in these residues (Sect. 3.4). It is likely that only when the bridge is made with the disaccharide can these changes take place. The preferred binding in sites S4–S5 for maltose rather than S5–S6, which would fully utilize the Tyr 404 contribution, may also be explained by the need to displace Gln 408 before a sugar can bind in S6. These features concerning an extended oligosaccharide recognition site involve favorable contacts made by bridging the span of a disaccharide as well as steric exlusion and may well be of general applicability to other enzymes, such as the amylases, which also exhibit disaccharide specificity.

Acknowledgements. We wish to thank Prof. G. Semenza (Zurich), Prof. A. Vasella (Zurich) and the Boehringer Chemical Company for gifts of compounds. We also wish to acknowledge Dr. C.C.F. Blake for permission to use unpublished data on human and tortoise lysozymes; Dr. P.J. Artymiuk for valuable help with refinements of ligand complexes both for lysozyme and phosphorylase; Dr.

D.I. Stuart, Dr. J. Hajdu, and Dr. N.G. Oikonomakos for their outstanding contributions to the structural studies on phosphorylase; and Dr. Y. Matsuura for permission to use unpublished data on Taka amylase.

Note added in proof. Since this article was submitted a high resolution structure of α-amylase has been published.

Buisson G, Duée E, Haser R, Payan E (1987) The three dimensional structure of porcine pancreatic α-amylase at 2.9 Å resolution. The role of calcium in structure and activity. EMBO J 6:3909–3916

References

Anderson WF, Grutter MG, Remington SJ, Matthews BW (1981) Crystallographic determination of the mode of binding of oligosaccharides to T4 bacteriophage lysozyme: implications for the mechanism of catalysis. J Mol Biol 147:523–543

Ariki M, Fukui T (1977) Affinity of glucose analogues for α-glucan phosphorylases from rabbit muscle and potato tubers. J Biochem (Tokyo) 81:1017–1024

Arnheim N, Inouye M, Low L, Laudin A (1973) Chemical studies on the enzymatic specificity of goose egg-white lysozyme. J Biol Chem 248:233–236

Arnott S, Scott WE (1972) Accurate x-ray diffraction analysis of fibrous polysaccharides containing pyranose rings. Part I. The linked atom approach. J Chem Soc Perkin II:324–335

Artymiuk PJ (1979) X-Ray Studies of Biological Macromolecules: The Refinement of Human Lysozyme at 1.5 Å Resolution. D. Phil. Thesis, University of Oxford

Artymiuk PJ, Blake CCF (1981) Human lysozyme at 1.5 Å resolution. J Mol Biol 152:737–762

Artymiuk PJ, Rice DW (1980–81) Progress report. Laboratory of molecular biophysics, University of Oxford, p 25

Aschaffenburg R, Blake CCF, Dickie HM, Gayen SK, Keegan R, Sen A (1980) The crystal structure of tortoise egg-white lysozyme at 6 Å resolution. Biochim Biophys Acta 625:64–71

Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 44:97–179

Banerjee SK, Holler E, Hess GP, Rupley JA (1975) Reaction of *N*-acetylglucosamine oligosaccharides with lysozyme. J Biol Chem 250:4355–4367

Banyard SE (1973) X-Ray Diffraction Studies of Enzymes. D Phil Thesis, University of Oxford

Beddell CR, Moult J, Phillips DC (1970) Crystallographic studies of the active site of lysozyme. In: Porter R, O'Connor M (eds) Ciba found symp on molecular properties of drug receptors. Churchill, London

Beevers CA, Maconochie GH (1965) The crystal structure of dipotassium glucose-1-phosphate dihydrate. Acta Crystallogr 18:232–236

Bienkowska K, Taylor A (1979) Low molecular weight substrate for the lysozyme of T4 bacteriophage. Eur J Biochem 96:581–584

Black WJ, Wang JH (1968) Studies on the allosteric activation of glycogen phosphorylase b by nucleotides. I. Activation of phosphorylase b by inosine monophosphate. J Biol Chem 243:5892–5898

Blake CCF, Koenig DF, Mair GA, North ACT, Phillips DC, Sarma VR (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2.0 Å resolution. Nature 206:757–761

Blake CCF, Mair GA, North ACT, Phillips DC, Sarma VR (1967a) On the conformation of the hen egg-white lysozyme molecule. Proc R Soc Lond [Biol] B 167:365–377

Blake CCF, Johnson LN, Mair GA, North ACT, Phillips DC, Sarma VR (1967b) Crystallographic studies of the activity of hen egg-white lysozyme. Proc R Soc Lond [Biol] B 167:378–388

Branden CI (1980) Relation between structure and function of α/β proteins. Q Rev Biophys 13:317–338

Branden CI (1986) The relation between protein structure in α/β domains and the intron-exon arrangement of the corresponding genes. Nobel symposium (in press)

Busby SJW, Radda GK (1976) Regulation of the glycogen phosphorylase system. Curr Top Cell Regul 10:89–160

Caudwell B, Antoniw JF, Cohen P (1978) Calsequestrin, myosin and the components of the protein-glycogen complex in rabbit skeletal muscle. Eur J Biochem 86:511–518

Cassels R (1979) Investigation of protein structure-lysozyme in the crystalline and solution states. Ph D thesis, University of Oxford

Cheetham JC (1986) Studies of enzyme-inhibitor interactions and protein dynamics. Ph D thesis, University of Oxford

Cheetham JC, Artymiuk PA, Phillips DC (1987) The refinement of an enzyme complex with inhibitor bound at partial occupancy: HEWL + triNAG at 1.75 Å resolution. (Manuscript in preparation)

Chu SJC, Jeffrey GA (1967) The crystal structure of methyl β-maltopyranoside. Acta Crystallogr 23:1038–1049

Cohen P (1978) Hormonal control of muscle glycogen metabolism. Curr Top Cell Reg 14:117–196

Cohen P (1983) Control of enzyme activity, 2nd edn. Chapman and Hall, London

Dobson CM (1975) The Confirmation of Lysozyme in Solution. D Phil Thesis, University of Oxford

Dombraidi V (1981) Structural aspects of the catalytic and regulatory function of glycogen phosphorylase. Int J Biochem 13:125–139

Fleming A (1922) On a remarkable bacteriolytic element found in tissues and secretions. Proc R Soc Lond (Biol) B93:306–317

Fletterick RJ, Madsen NB (1980) The structures and related functions of phosphorylase *a*. Annu Rev Biochem 49:31–61

Fletterick RJ, Sprang SR (1982) Glycogen phosphorylase structure and function. Acc Chem Res 15:361–369

Ford LO, Johnson LN, Machin PA, Phillips DC, Tjian RJ (1974) Crystal structure of a lysozyme-tetrasaccharide lactone complex. J Mol Biol 88:349–371

Gallay O (1984) X-ray studies on an enzyme: X-ray crystallography of the binding of a tetrasaccharide to tortoise egg-white lysosome. Part II. Thesis, University of Oxford

Ghuysen JM (1968) Use of bacteriolytic enzymes in determination of wall structure and their role in cell metabolism. Bacteriol Rev 32:425–464

Goldsmith E, Fletterick R (1983) Oligosaccharide conformation and protein saccharide interactions in solution. Pure Appl Chem 55:577–588

Goldsmith E, Sprang S, Fletterick R (1982) Structure of maltoheptose by difference Fourier methods and a model for glycogen. J Mol Biol 156:411–427

Goldsmith EJ, Fletterick RJ, Withers SG (1987) The three-dimensional structure of acarbose bound to glycogen phosphorylase. J Biol Chem 262:1449–1455

Grace DEP (1980) A refined analysis of the structure of lysozyme and its interactions with substrates. Phil thesis, University of Oxford

Graves DJ, Wang JH (1972) α-glucan phosphorylases-chemical and physical basis of catalysis and regulation. Boyer P (ed) The Enzymes, 3rd edn, vol 7. Academic, New York, pp 435–482

Gress ME, Jeffrey GA (1977) A neutron diffraction refinement of the crystal of β-maltose monohydrate. Acta Cryst B33:2490–2495

Grutter MG, Matthews BW (1982) Amino acid substitutions far from the active site of bacteriophage T4 lysozyme reduce catalytic activity and suggest that the C-terminal lobe of the enzyme participates in substrate binding. J Mol Biol 154:525–535

Grutter MG, Weaver LH, Matthews BW (1983) Goose lysozyme structure: an evolutionary link between hen and bacteriophage lysozymes? Nature 303:828–831

Hajdu J, Acharya KR, Stuart DI, McLaughlin PJ, Barford D, Oikonomakos NG, Klein H, Johnson LN (1987) Catalysis in the crystal: synchrotron radiation studies with glacogen phosphorylase *b*. EMBO J 6:539–546

Handoll HHG (1985) PhD thesis, University of Oxford

Handoll HHG, Artymiuk PJ, Grace DEP, Phillips DC (1987) A description and comparison of the refined structures of the original model of tetragonal hen egg-white lysozyme to 2.0 Å resolution and that of a closely related tetragonal form at 1.6 Å crystallised at higher pH. (Manuscript in preparation)

Haschke RM, Gratz KW, Heilmeyer LMG (1972) The role of phosphorylase activity in a muscle glycogen particle. J Biol Chem 247:5351–5356

Hehre EJ, Brewer CF, Uchiyama T, Schlesselmann P, Lehmann J (1980) Scope and mechanism of carbohydrase action. Stereospecific hydration of 2,6-anhydro-1-deoxy-D-gluco-hept-1-enitol catalysed by α- and β-glucosidases and an inverting exo-α-glucanase. Biochemistry 19:3557–3564

Heilmeyer LMG, Meyer F, Haschke RH, Fischer EH (1970) Control of phosphorylase activity in a muscle glycogen particle, *II*. Activation by calcium. J Biol Chem 245:6649–6656

Helmreich EJM, Klein H (1980) The role of pyridoxal phosphate in the catalysis of glycogen phosphorylases. Angew Chem Int 19:441–455

Helmreich EJM, Michaelides MC, Cori CF (1967) Effects of substrates and a substrate analog on the binding of 5′-adenylic acid to muscle phosphorylase *a*. Biochemistry 6:3695–3710

Hendrickson WA, Konnert JH (1980) Incorporation of stereochemical information into crystallographic refinement. In: Diamond R, Ramaseshan S, Venkatesan K (eds) Computing in crystallography. Indian Academy of Science, pp 13.01–13.23

Hogle J, Rao ST, Mallikarjunan M, Beddell C, McMullan RK, Sandaralingam M (1981) Studies of monoclinic hen egg-white lysozyme. I. Structure solution at 4 Å resolution and molecular packing comparisons with tetragonal and triclinic lysozymes. Acta Crystallogr B37:591–597

Holler E, Rupley JA, Hess GP (1975) Productive and unproductive lysozyme-chitosaccharide complexes. Equilibrium measurements. Biochemistry 14:1088–1094

Hu HY, Gold AM (1975) kinetics of glycogen phosphorylase *a* with a series of semisynthetic branched saccharides. Biochemistry 14:2224–2230

Imoto T, Johnson LN, North ACT, Phillips DC, Rupley JA (1972) Vertebrate lysozymes. In: Boyer PD (ed) The enzymes, 3rd edn, vol 7. Academic, New York, pp 665–868

Jacobson RA, Wunderlich JA, Lipscomb WA (1961) The crystal and molecular structure of cellobiose. Acta Crystallogr 14:598–607

Jeanloz RW, Sharon N, Flowers HM (1963) The chemical structure of a disaccharide isolated from *Micrococcus lysodeiktius* cell wall. Biochem Biophys res Commun 13:20–25

Jenkins JA, Johnson LN, Stuart DI, Stura EA, Wilson KS, Zanotti G (1981) Phosphorylase: control and activity. Philos Trans R Soc Lond [Biol] B293:23–41

Johnson LM, Phillips DC, Rupley JA (1968) The activity of lysozyme: An interim review of crystallographic and chemical evidence. Brookhaven Symp Biology 21:120–138

Johnson LN, Phillips DC (1965) Structure of some crystalline lysozyme-inhibitor complexes determined by x-ray analysis at 5 Å resolution. Nature 206:761–763

Johnson LN, Jenkins JA, Wilson KS, Stura EA, Zanotti G (1980) Proposals for the catalytic mechanism of glycogen phosphorylase *b*. J Mol Biol 140:565–580

Johnson LN, Stura EA, Sansom MSP, Babu YS (1983) Oligosaccharide binding to glycogen phosphorylase *b*. Biochem Soc Trans *11*: 142–144

Johnson LN, Hajdu J, Acharya KR, Stuart DI, McLaughlin PJ, Oikonomakos NG, Barford D (1988) Glycogen Phosphorylase *b*. In: Herve G (ed) Allosteric Enzymes. CRC Press, Florida (in press)

Jolles P, Jolles J (1984) Review of progress in lysozyme research. Mol Cell Biochem 63:165–189

Kasvinsky PJ, Madsen NB (1976) Activity of glycogen phosphorylase in the crystalline state. J Biol Chem 251:6852–6859

Kasvinsky PJ, Madsen NB, Fletterick RJ, Sygusch J (1978) X-ray crystallographic and kinetic studies of oligosaccharide binding to phosphorylase. J Biol Chem 253:1290–1296

Kelly JA, Sielecki AR, Sykes BD, James MNG, Phillips DC (1979) X-ray crystallography of the binding of the bacterial cell wall trisaccharide NAM–NAG–NAM to lysozyme. Nature 282:875–878

Kelly JA, Sielecki AR, Sykes BD, James MNG, Phillips DC (1979) X-ray crystallography of the binding of the bacterial cell wall trisaccharide NAM–NAG–NAM to lysozyme. Nature 282:875–878

Kirkman BR, Whelan WJ (1986) Glucosamine is a normal component of liver glycogen. FEBS Lett 194:6–11

Klein MW, Im MJ, Palm D, Helmreich EJM (1984) Does pyridoxal 5′-phosphate function in glycogen phosphorylase as an electrophilic or a general acid catalyst? Biochemistry 23:5853–5861

Klein MW, Im MJ, Palm DL (1986) Mechanism of the phosphorylase reaction: utilisation of D-glucohept-1-enitol in the absence of primer. Eur J Biochem 57:107–114

Kleppe G, Vasstrand E, Jensen HB (1981) The specificity requirements of bacteriophage T4 lysozyme. Involvement of the *N*-acetamido groups. Eur J Biochem 119:589–593

Kurachi K, Sieker LC, Jensen LH (1976) Structures of triclinic mono- and di-*N*-acetylglucosamine:lysozyme complexes – a crystallographic study. J Mol Biol 101:11–24

Kuramitsu S, Ikeda K, Hamaguchi K (1977) Effects of ionic strength and temperature on the ionization of the catalytic groups, Asp 52 and Glu 35, in hen lysozyme. J Biochem (Tokyo) 82:585–597

Larner J, Takeda Y, Hizukuri S (1976) The influence of chain size and molecular weight on the

kinetic constants for the span glucose to polysaccharide for rabbit muscle glycogen synthase. Mol Cell Biochem 12:131–136

Lehrer SS, Fasman GD (1966) The Fluorescence of Lysozyme and Lysozyme Substrate Complexes. Biochem Biophys Res Commun 23:133–139

Levitt M (1972) Conformation analysis of proteins. Ph D thesis, University of Cambridge

Levitt M (1974) On the nature of the binding of hexa-*N*-acetylglucosamine substrate to lysozyme. In: Blout ER, Bovrey FA, Goodman M, Lotan N (eds) Peptides, polypeptides and proteins. Wiley, New York, pp 99–113

Liddington R (1981) X-ray studies on lysozyme. Chemistry. Part II. Thesis, University of Oxford

Loyter A, Schramm M (1966) Multimolecular complexes of α-amylase with glycogen limit dextrin. J Biol Chem 241:2611–2617

Madsen NB, Withers SG (1986) Glycogen phosphorylase in coenzymes and cofactors. In: Dolphin D, Poulson R, Avramovic O (eds) Pyridoxal phosphate and derivatives. Wiley, New York

Manor PC, Saenger W (1974) Topography of cyclodextrin inclusion complexes. J Am Chem Soc 96:3630–3639

Mason SA, Bentley GA, McIntyre GJ (1984) Deuterium exchange in lysozyme at 1.4 Å resolution. In: Schoenborn P (ed) Neutrons in biology. Plenum, NY

Matsuura Y, Kusunoki M, Harada W, Kakudo M (1984) Structure and possible catalytic residues of Taka amylase A. J Biochem (Tokyo) 95:697–702

Matthews BW, Remington SJ (1974) The three dimensional structure of the lysozyme from bacteriophage T4. Proc Natl Acad Sci USA 71:4178–4182

McLaughlin PJ, Stuart DI, Klein HW, Oikonomakos NG, Johnson LN (1984) Substrate cofactor interactions for glycogen phosphorylase *b*. A binding study in the crystal with heptenitol and heptulose-2-phosphate. Biochemistry 23:5862–5873

Metzger B, Helmreich E, Glaser L (1967) The mechanism of activation of skeletal muscle phosphorylase *a* by glycogen. Proc Natl Acad Sci USA 57:994–1001

Meyer F, Heilmeyer LMG, Haschke RM, Fischer EH (1970) Control of phosphorylase activity in the glycogen particle. Isolation and characterisation of the protein glycogen complex. J Biol Chem 245:6642–6648

Mo F, Jensen LH (1978) The crystal structure of a β-(1–4) linked disaccharide, α-*N,N'*-Diacetylchitobiose monohydrate. Acta Crystallogr B34:1562–1569

Monod J, Changeux JP, Wyman J (1965) On the nature of allosteric transitims: a plausible model. J Mol Biol 12:88–118

Moult J, Yonath A, Traub W, Smilansky A, Podjarny A, Rabionvich P, Saya A (1976) The structure of triclinic lysozyme at 2.5 Å resolution. J Mol Biol 100:179–195

Najmudin S (1984) The control and activity of Phosphorylase: crystallographic study of some metabolite analogues (Phosphate, Glucosamine 1-Phosphate and Adenosine 5'-O-Monothio-phosphate) binding to Phosphorylase *b*. Biochemistry. Part II. Thesis, University of Oxford

Nakano K, Fukui T (1986) The complete amino acid sequence of potato α-glucan phosphorylase. J Biol Chem 261:8230–8236

Nakano K, Hwang PK, Fletterick RJ (1986) Complete cDNA sequence for rabbit muscle glycogen phosphorylase. FEBS Lett 204:283–287

Narendra N, Seshadri TP, Viswamittra MA (1984) The structure of the disodium salt of glucose-1-phosphate hydrate. Acta Crystallogr C40:1338–1340

Oatley SJO (1973) Structural studies of biological macromolecules. Chemistry. Part II. Thesis, University of Oxford

Palm D, Goerl R, Burger KJ (1985) Evolution of catalytic and regulatory sites in phosphorylases. Nature 313:500–502

Pauling L (1946) Molecular architecture and biological reactions. Chem Eng News 24:1375–1377

Payan F, Maser R, Pierrot M, Frey M, Astier JP (1980) The three-dimensional structure of α-amylase from porcine pancreas at 5 Å resolution. Acta Crystallogr B36:416–421

Perkins SJ, Johnson LN, Machin PA, Phillips DC (1979) Crystal structures of hen egg-white lysozyme complexes with Gd(III) and Gd(III)-*N*-acetylglucosamine. Biochem J 181:21–36

Pflugrath JW, Weigand G, Huber R (1986) Crystal structure determination, refinement and molecular model of α-amylase inhibitor Hoe-467A. J Mol Biol 189:383–386

Philip G, Gringel G, Palm D (1982) Rabbit muscle phosphorylase derivatives with oligosaccharides covalently bound to the glycogen storage sites. Biochemistry 21:3043–3050

Phillips DC (1966) The three-dimensional structure of an enzyme molecule. Sci Am 215:78–90
Phillips DC (1986) Protein structure and function. Fifty years of the Patterson function. Oxford University Press, Oxford
Pincus MR, Scheraga HA (1981) Prediction of the three-dimensional structures of complexes of lysozyme with cell wall substrates. Biochemistry 20:3960–3965
Post CB, Brooks BR, Karplus M, Dobson CM, Artymiuk PJ, Cheetham JC, Phillips DC (1986) Molecular dynamics simultaneous of native and substrate-bound lysozyme. A study of the average structures and atomic fluctuations. J Mol Biol 190:455–479
Prager EM, Wilson AC, Arnheim N (1974) Widespread distribution of lysozyme g in egg-white of birds. J Biol Chem 249:7295–7297
Pulford WCA (1982) X-ray studies of enzyme action. Ph D thesis, University of Oxford
Quigley GJ, Saiko A, Marchessault RM (1970) Crystal and molecular structure of maltose monohydrate. J Am Chem Soc 92:5834–5839
Quiocho FA (1986) Carbohydrate binding proteins: Tertiary structures and protein sugar interactions. Annu Rev Biochem 55:287–315
Quiocho FA, Vyas NK (1984) Novel stereospecificity of the L-arabinose binding protein. Nature 310:381–386
Rees DA, Smith PJC (1975) Polysaccharide conformation Part IX. Monte Carlo calculation of conformational energies for disaccharides and comparison with experiment. J Chem Soc II:836–840
Remington SJ, Anderson WF, Owen J, Ten Eyck LF, Grainger CT, Matthews BW (1978) Structure of the lysozyme from bacteriophage T4: an electron density map at 2.4 Å resolution. J Mol Biol 118:81–98
Romero PA, Smith EE, Whelan WJ (1980) Glucosamines as a substitute for glucose in glycogen metabolism. Biochem Int 1:1–9
Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of a nucleotide binding protein. Nature 250:194–199
Rupley JA, Butler L, Gerring M, Hartdegen FJ, Pacoraro R (1967) Studies on the enzymic activity of lysozyme, III. The binding of saccharides. Proc Natl Acad Sci USA 57:1088–1095
Saenger W (1979) Circular hydrogen bonds. Nature 279:343–344
Salton MRJ (1952) Cell wall of micrococcus lysodeikticus as the substrate of lysozyme. Nature 170:746–747
Salton MRJ, Ghuysen JM (1959) The structure of di- and tetra-saccharides released from cell walls by lysozyme and streptomyces F_1 enzyme and the $\beta(1-4)$ N-acetyl-hexosaminidase activity of these enzymes. Biochim Biophys Acta 36:552–554
Salton MRJ, Ghuysen JM (1960) Acetylhexosamine compounds enzymically released from Micrococcus Lysodeikticus cell walls. III. The structure of di- and tetra-saccharides released from cell walls by lysozyme and streptomyces F_1 enzyme. Biochim Biophys Acta 45:355–363
Sansom MSP, Stuart DI, Acharya KR, Hajdu J, Mclaughlin PJ, Johnson LN (1985) Glycogen phosphorylase *b* – the molecular anatomy of a large regulatory enzyme. J Mol Struct 123:3–25
Sarma R, Bott R (1977) Crystallographic study of turkey egg-white lysozyme and its complex with a disaccharide. J Mol Biol 113:555–565
Schellman JA, Hawkes RB (1980) Thermodynamic stability of T4 variant lysozymes. In: Jaenicke R (ed) Protein folding. North-Holland Biomedical, Amsterdam, pp 331–334
Schellman JA, Lindorfer M, Hawkes RB, Gruter MG (1981) Mutations and protein stability. Biopolymers 20:1989–1999
Secemski II, Lienhard GE (1971) The role of strain in catalysis by lysozyme. J Am Chem Soc 93:3549–3550
Smith-Gill SJ, Rupley JA, Pincus MR, Carty PR, Scheraga HA (1984) Experimental identification of a theoretically predicted "left-sided" binding mode for $(GlcNAc)_6$ in the active site of lysozyme. Biochemistry 23:993–997
Sotiroudis TG, Oikonomakos NG, Evangelopoulos AE (1978) Phosphorylase *b* covalently bound to glycogen. Eur J Biochem 88:573–581
Sprang S, Fletterick RJ (1979) The structure of glycogen phosphorylase *a* at 2.5 Å resolution. J Mol Biol 131:523–551
Sprang SR, Goldsmith EJ, Fletterick RG, Withers SG, Madsen NB (1982) Catalytic site of glycogen phosphorylase: structure of the T-state and specificity for α-D-glucose. Biochemistry 21:5364–5381

Steinrauf LK (1959) Preliminary x-ray data for some new crystalline forms of β-lactoglobulin and hen egg-white lysozyme. Acta Crystallogr 12:77

Stralfors P, Hiraga A, Cohen P (1985) The protein phosphatases involved in cellular regulation. Eur J Biochem 149:295–303

Street IP, Armstrong CR, Withers SG (1986) Hydrogen bonding and specificity. Fluorodeoxy sugars as probes of hydrogen bondong in the glycogen phosphorylase-glucose complex. Biochemistry 25:6021–6027

Sundanarajan PR, Rao VSR (1969) Studies on the helical conformations of amylose and possible interconversions. Biopolymers 8:313–323

Takagi T, Toda H, Isenura T (1971) Bacterial and mold amylases. In: Boyer PD (ed) The enzymes, 3rd edn, vol 5. Academic, New York, pp 235–271

Takusagawa F, Jacobson RA (1978) The crystal and molecular structure of α-maltose. Acta Crystallogr B34:213–218

Tanaka I, Tanaka N, Ashida T, Kakudo M (1976) The crystal and molecular structure of phenyl-α-maltoside. Acta Crystallogr B32:155–160

Thoma JA, Spradlin JE, Dygent S (1971) Plant and animal amylases. In: Boyer PD (ed) The enzymes. Academic, New York, pp 115–189

Titani K, Koide A, Hermann J, Ericsson LH, Kumar S, Wade RD, Walsh KA, Neurath M, Fischer EH (1977) Complete amino acid sequence of rabbit muscle glycogen phosphorylase. Proc Natl Acad Sci USA 74:4762–4777

Truschuit E, Frommer W, Junge B, Muller L, Schmidt D, Wingender W (1981) Chemistry and biochemistry of microbial α-glucosidase inhibitors. Angew Chem Int Ed Engl 20:744–761

Tsugita A (1971) Phage lysozyme and other lytic enzymes. In: Boyer PA (ed) The enzymes, vol 5. Academic, New York, pp 343–411

Wang JH, Shonka ML, Graves DJ (1965) Influence of carbohydrate on phosphorylase structure and activity I. Activation by preincubation with glycogen. Biochemistry 4:2295–2301

Wanson JC, Drochmans P (1968) Rabbit skeletal muscle glycogen. J Cell Biol 38:130–150

Weaver LH, Grutter MG, Remington SJ, Gray TM, Isaacs NW, Matthews BW (1985) Comparison of goose-type, chicken-type and phage-type lysozymes: changes in three-dimensional structure during evolution. J Mol Evol 21:97–111

Weber IT, Johnson LN, Wilson KS, Yeates DGR, Wild DL, Jenkins JA (1978) Crystallographic studies on the activity of glycogen phosphorylase *b*. Nature 274:433–437

Whelan WJ (1986) The initiation of glycogen synthesis. Bioessays 5:136–140

Withers SG, Sykes BD, Madsen NB, Kasvinsky PJ (1979) Identical structural changes induced in glycogen phosphorylase by two non-exclusive allosteric inhibitors. Biochemistry 18:5342–5348

Withers SG, Madsen NM, Sprang SR, Fletterick RJ (1982) Catalytic site of glycogen phosphorylase: structural changes during activation and mechanistic implications. Biochemistry 21:5372–5382

Wolfenden R (1969) Transition state analogues for enzyme catalysis. Nature 223:704–705

Wu HC, Sarko A (1978) The double helical structure of crystalline α-amylose. Carbohydr Res 61:27–40

Molecular Features and Basic Understanding of Protein–Carbohydrate Interactions: The Arabinose-Binding Protein–Sugar Complex

F.A. QUIOCHO

1 Introduction

Protein–carbohydrate interactions play an important role in a wide range of biological and biochemical processes. The carbohydrate-binding proteins and enzymes that have been extensively studied show wide diversity in size, subunit composition and three-dimensional structure (QUIOCHO 1986). Moreover, there is variation in the nature of the carbohydrate ligands, which range from simple monosaccharides to polysaccharides, either free, such as glycogen and starch, or conjugated polymers such as glycoproteins and glycolipids. Despite this diversity and variability the molecular basis of the mode of binding of sugar substrates shows common features (QUIOCHO 1986).

Our understanding of protein–carbohydrate interactions at the molecular level have come principally from the results of X-ray structure analysis. In

Howard Hughes Medical Institute and the Departments of Biochemistry and of Physiology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030, USA

Current Topics in Microbiology and Immunology, Vol. 139

terms of detail and accuracy, the 1.7-Å resolution, highly refined X-ray structure of the complex of L-arabinose-binding protein with its sugar substrate remains unsurpassed (QUIOCHO and VYAS 1984). Moreover, we believe that this complex not only exhibits essentially all of the molecular features of protein–carbohydrate interactions, but provides a useful framework to analyze other complexes and to formulate principles of such interactions. These basic features and principles are the primary focus of this essay; the brevity of this review reflects the simplicity and clarity of many of these features. A more comprehensive review of the three-dimensional structures of proteins and enzymes that bind carbohydrates and of sugar–protein interactions has recently appeared (QUICHO 1986).

2 L-Arabinose-Binding Protein: An Ideal Protein for Structure and Function Studies

L-Arabinose-binding protein is one of a large group of proteins located in the periplasm of Gram-negative bacteria. All these proteins serve as initial components of osmotic shock-sensitive active transport systems for a variety of carbohydrates, amino acids, and ions. Moreover, several of the sugar-binding proteins act as initial receptors in the simple behavioral response of bacterial chemotaxis. Structures of some of these proteins with their respective bound ligands have provided a wealth of new information on the association of sugar, anion (sulfate), or zwitterion (leucine). As indicated in Table 1, the arabinose-binding protein has been studied extensively by a variety of techniques, notably X-ray crystallography. Nuclear magnetic resonance technique and theoretical calculation have also been used to probe the liganded state of the arabinose-binding protein (CLARK et al. 1982; MAO et al. 1982). In addition to being well-characterized, the binding protein is ideal for structure and function studies for several other reasons: (a) the protein can be easily purified in gram quantities, and binding activity can be easily measured by equilibrium dialysis or by monitoring ligand-induced fluorescence change (PARSONS and HOGG 1974; MILLER et al. 1983); (b) kinetics of sugar binding are amenable to measurement using stopped-flow, rapid-mixing technique (MILLER et al. 1983); (c) excellent diffracting crystals can be easily obtained; (d) tertiary structure of the protein has been accurately determined at 1.7 Å resolution (QUIOCHO and VYAS 1984); and (e) the binding protein gene has been cloned (R.W. HOGG, private communication).

The L-arabinose-binding protein consists of a single polypeptide chain with a molecular weight of 33170 and a very high affinity sugar binding site (Table 1). Although the protein is utilized specifically for transport of L-arabinose, it also binds D-galactose with similar affinity and D-fucose 30-fold less tightly; variation in the affinity constants is attributed primarily to changes in k_{off} or dissociation rate contants (see Table 1 and MILLER et al. 1983). The binding of sugar substrates to the arabinose-binding protein and to other periplasmic sugar-binding proteins (MILLER et al. 1983) results in some of the tightest known protein–

Table 1. Physical and chemical properties of the L-arabinose-binding protein

Molecular weight: 33170

Amino acid composition for 306 residues (HOGG and HERMODSON 1977):
$Lys_{30}His_3Arg_8Asp_{21}Asn_{10}Thr_{16}Ser_{14}Glu_{21}Gln_{11}Pro_{15}$
$Gly_{29}Ala_{31}Val_{26}Met_{10}Ile_{16}Leu_{21}Tyr_6Phe_{12}Cys_1Trp_5$

Substrate binding:

A. Kinetic and equilibrium data (MILLER et al. 1983):

Substrate	$k_{on} \times 10^{-7}\ M^{-1}\ s^{-1}$	$k_{off}\ s^{-1}$	$k_{off}/k_{on} \times 10^7\ M$	$K_d \times 10^7\ M$
L-arabinose	2.4	1.5	0.6	0.98
D-galactose	0.8	1.8	2.2	2.30
D-fucose	1.2	37.0	31.0	38.0

B. Thermodynamic data (FUKUDA et al. 1983):

Thermodynamic parameters (25° C)		Substrate	
		L-arabinose	D-galactose
ΔG°	(kcal mol^{-1})	−8.7	−8.5
ΔH°	(kcal mol^{-1})	−15.3	−15.0
ΔS°	(cal K^{-1} mol^{-1})	−22.1	−21.8
ΔCp°	(cal K^{-1} mol^{-1})	−436	−376

Structure analysis:

A. Crystallography (QUIOCHO and VYAS 1984):
Space group: $P2_12_12_1$
Unit cell dimensions: $a = 55.46$ Å, $b = 71.82$ Å, $c = 77.84$ Å
Structure determination status: refinement of the protein structure with bound substrate at 1.7 Å resolution to an R factor of 13.7%

B. Low angle X-ray scattering (NEWCOMER et al. 1981b):
Radius of gyration of sugar-free protein = 21.22 Å
Radius of gyration of protein plus L-arabinose = 20.28 Å

carbohydrate complexes, if not the tightest. An explanation for this tight sugar binding is offered in a later section. Ligand affinity and the kinetics of substrate binding are fundamentally related to the functions of binding proteins in active transport and chemotaxis (MILLER et al. 1983; QUIOCHO and VYAS 1984; QUIOCHO 1986).

The crystal structure of the L-arabinose-binding protein with bound L-arabinose has been extensively refined at 1.7 Å resolution to an R factor of 13.7% (QUIOCHO and VYAS 1984). The structure is composed of two similar globular domains arranged in an ellipsoid (axial ratio of 2:1). A deep cleft occurs between the two domains wherein the sugar substrate is bound and becomes almost totally engulfed (NEWCOMER et al. 1981a, b; QUIOCHO and VYAS 1984). These structural and ligand-binding features are similarly found in other periplasmic binding proteins (VYAS et al. 1983; SAPER and QUIOCHO 1983; PFLUGRATH and QUIOCHO 1985).

3 A Ligand-Site Geometry for Binding Either Anomeric Sugar Substrate

X-ray analysis of the L-arabinose-binding protein has revealed a novel ligand-site geometry that can accomodate within the same exact site either the α or β anomeric form of the sugar substrate (QUIOCHO and VYAS 1984). All atoms of both sugar anomers interact with the protein via hydrogen bonds and van der Waals' contacts. Both sugar anomers are bound by the same number of hydrogen bonds, involving identical amino acid residues and two isolated water molecules within the binding site region.

It is important to note that the ability of the arabinose-binding protein to bind both sugar anomers is entirely consistent with the transport function of the binding protein: the open-chain aldehyde form derived from both anomers of the translocated L-arabinose is utilized by the bacterial cell in the biosynthesis of pentose phosphates (ENGELSBERG 1971). It is further noteworthy that many of the enzymes of glucose metabolism (e.g., hexokinase, glucokinase, glucose 6-phosphatase) can also act on both anomeric forms of their respective sugar substrates.

4 Hydrogen Bonds Stabilize the Protein–Sugar Complex

Hydrogen bonds are the major force in the stability of protein–sugar complexes. This is amply exemplified in the case of the binding protein–arabinose complex. As shown in Fig. 1, all hydrogen-bonding groups of each of the sugar anomers

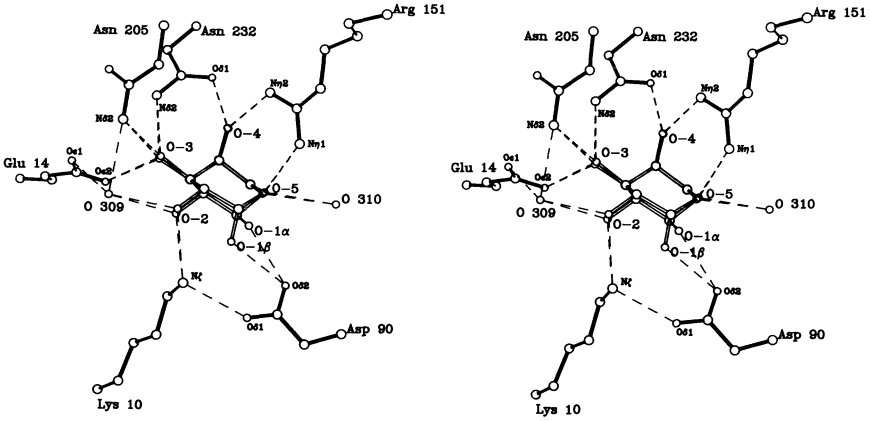


Fig. 1. Stereo drawing of the hydrogen bonds between L-arabinose-binding protein and α or β anomeric form of the L-arabinose substrate. The α-anomeric hydroxyl is labeled O-1α and the β-anomeric hydroxyl O-1β. Hydrogen bonds are represented by *dashed lines*. The residue identifiers are placed close to C-α carbons of the essential residues. The two isolated water molecules intimately involved in sugar binding are represented as O 309 and O 310. Arg 151 and Asn 232 from bidentate H bonds with the sugar. The sugar hydroxyls 2, 3, and 4 simultaneously donate and accept H bonds. Note that the refined positions of both sugar anomers are not exactly coincident (see QUIOCHO and VYAS 1984)

Table 2. Hydrogen bonds between L-arabinose-binding protein and α/β-L-arabinose

Donors (X)	Acceptors (Y)	α-L-arabinose		β-L-arabinose	
		X ··· Y (Å)	X-Ĥ ··· Y (°)	X ··· Y (Å)	X-Ĥ ··· Y (°)
O-1	Asp 90 Oδ2	2.77	159	2.74	166
O-2	Wat 309 O	2.61	155	2.60	156
O-3	Glu 14 Oε2	2.77	176	2.66	173
O-4	Asn 232 Oδ1	2.62	178	2.62	176
Lys 10 Nζ	O-2	2.73	146	2.86	151
Asn 205 Nδ2	O-3	3.03	169	3.09	167
Asn 232 Nδ2	O-3	2.97	162	2.98	159
Arg 151 Nη2	O-4	2.82	167	2.81	168
Arg 151 Nη1	O-5	3.05	161	2.99	160
Wat 310 O	O-5	2.80	–	2.74	–
	Overall mean =	2.82 (0.15)	164 (10)	2.81 (0.16)	164 (8)

are utilized in the formation of a total of ten hydrogen bonds. Hydrogen bond geometries, summarized in Table 2, are similar for both anomers. The hydrogen bonds are distributed equally between two types: five neutral-neutral and five neutral-charged hydrogen bonds.

Whereas the anomeric hydroxyl group of bound L-arabinose participates almost exclusively as a hydrogen bond donor, all other hydroxyls serve simultaneously as hydrogen bond donors and acceptors. These patterns of hydrogen-bonding of sugar hydroxyl groups are consistent with the following concepts (QUIOCHO 1986): (a) the anomeric hydroxyl, as a result of the "anomeric effect," is more acidic and thus a stronger-than-average hydrogen bond donor and weaker-than-average hydrogen bond acceptor; (b) because of the "cooperative effect," the involvement of hydroxyls (especially the nonanomeric hydroxyls) simultaneously as hydrogen bond donor and acceptor groups leads to stronger-than-average hydrogen bonds; and (c) in the low dielectric environment of the sugar binding site of proteins, hydrogen bond formation should not leave free the "active" hydrogen of the sugar hydroxyls. The last concept presupposes that the sugar hydroxyl is unlikely to serve solely as an H bond acceptor.

4.1 Anomeric Sugar Hydroxyl Serves Preferentially as H Bond Donor

The key to the ability of the binding protein to recognize both anomers of L-arabinose while using the same essential residues in hydrogen bonding is the precise alignment of the atom Oδ2 of Asp 90 which enables it to accept a hydrogen bond from either the α (equatorial) or β (axial) anomeric hydroxyl (see Figs. 1, 2). Even though the receptor protein binds either anomeric form, the orientations of all essential residues remain unchanged, and the hydrogen bonds to both anomers show essentially the same geometries (Table 2). These are achieved by the slight divergence in positions of both anomers (see Fig. 1, and QUIOCHO and VYAS 1984).

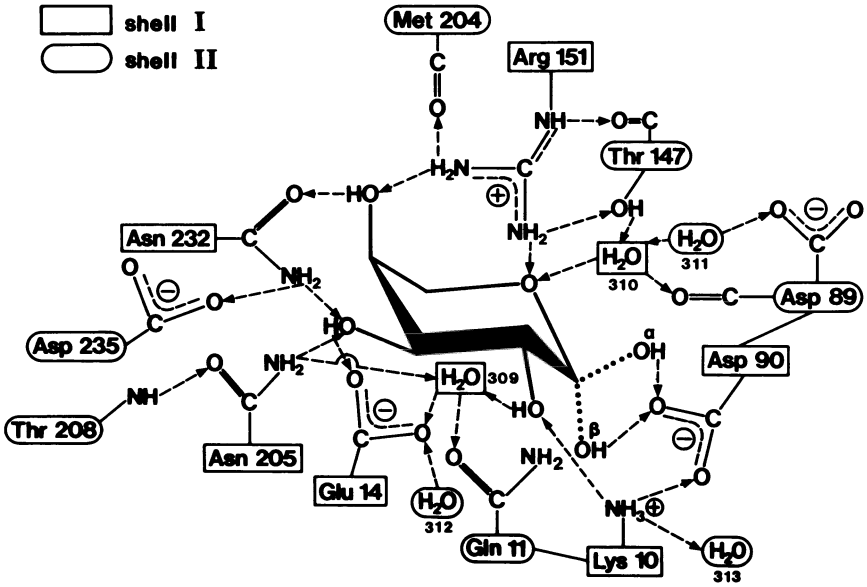


Fig. 2. Schematic diagram of the intricate networks of hydrogen bonds formed in the complex between the binding protein and L-arabinose substrate. Shell I includes the essential residues and isolated water molecules H-bonded to the sugar anomers and to the adjacent second shell of residues. Shell II represents residues H-bonded to the essential residues. The Shell II residue Gln 11 was incorrectly identified as Glu 11 in the original publication. (Adapted from QUIOCHO and VYAS 1984 with permission)

4.2 Nonanomeric Sugar Hydroxyls Act Simultaneously as Donors and Acceptors: "Cooperative" Hydrogen Bonding

Remarkably, the simultaneous participation of all of the nonanomeric hydroxyls as hydrogen bond donor and acceptor groups in the binding protein–arabinose complex can be described simply as:

$$(NH)_n \rightarrow OH \rightarrow O$$

where NH and O are hydrogen bond donor and acceptor groups in the protein binding site, respectively; OH is a nonanomeric sugar hydroxyl, and $n=1$ or 2. The NH donor groups emanate from the side chains of Lys 10, Arg 151, Asn 205, and Asn 232, whereas the O acceptor groups come from Glu 14 and Asn 232 side chains and the isolated water molecule 309.

The "cooperative" hydrogen bonds involving the nonanomeric hydroxyls are as follows (see Figs. 1, 2 and Table 2): OH-2 donates a hydrogen bond to water 309 and accepts an NH from the ζ-ammonium group of Lys 10. OH-3 is fully coordinated (including the C-O bond) as it donates an H bond to Glu 14 Oε2 and accepts two NH groups, one each from the Nδ2 side chain of Asn 205 and Asn 232. OH-4 provides an H bond to Asn 232 Oδ1 and accepts an NH group from Arg 151 Nη2.

4.3 Sugar Ring Oxygen Accepts Two H bonds

The arabinose is additionally bound to the receptor via hydrogen bonds with the sugar ring oxygen (O-5). In fact, by accepting H bonds from two donor groups, an NH from Arg 151 Nη1 and an OH from water 310, the hydrogen-bonding capacity of the ring oxygen is fully utilized.

4.4 Excellent Hydrogen Bond Geometry

Thus far, the hydrogen bond parameters listed in Table 2 are the only ones available from an extensively refined 1.7-Å resolution structure of a protein–sugar complex. Besides providing accurate data, these parameters serve as a yardstick with which to examine structures of other protein–carbohydrate complexes. These parameters are similar to those observed for hydrogen bonds in small molecule crystal structures of sugars and in highly refined high-resolution protein structures; the hydrogen bonds found in the latter two structures are, however, different from those formed in protein–sugar complexes.

The geometries of the cooperative hydrogen bonding shown in Equation 1 are those expected for maximal interactions (see Fig. 1). For example, the OH-3 is fully coordinated ($n=2$) in an arrangement that is essentially tetrahedral, including the sugar C-O bond. The hydrogen-bonding geometry of OH-4, which simultaneously donates and accepts one H bond, is also favorable: the atoms C-4 and O-4 of the sugar, Arg 151 Nη2, and Asn 232 Oδ1 are coplanar, and the H bond donated to or by OH-4 is oriented so that both lone pairs of electrons on either acceptor group (the sugar hydroxyl oxygen or the Oδ1 of Asn 232) are utilized in accepting the donor group. Similar analysis of the hydrogen-bonding system involving OH-2 leads to the same observation. On the other hand, as the sugar ring oxygen accepts two hydrogen bonds in an almost tetrahedral coordination, each of the lone pairs of electrons on the sugar ring oxygen is directed at a hydrogen bond donor group.

5 Polar Planar Side Chains Are Extensively Used in Sugar Binding: Bidentate and Networks of H Bonds

With the exception of Lys 10 all the residues utilized in binding the arabinose have planar polar side chains with two or more functional groups that are engaged in multiple hydrogen bonds (e.g., Asn, Glu, Asp, Arg). Furthermore, the H bond donor groups of these planar side chains, being NH groups in Arg and Asn, are in highly fixed orientations. Two factors account for these findings. First, two of the planar side chains (Arg 151 and Asn 232) make bidentate hydrogen bonds with the sugar (see Fig. 1). The geometry of these bidentate hydrogen bonds is such that the atoms Oδ1 and Nδ2 of Asn 232 and O-3 and O-4 of the sugar are coplanar, as are atoms Nη1 and Nη2 of Arg 151 and Ara O-5 and O-4. And, second, all planar side chains, together with the

sugar molecule, are involved in extensive networks of hydrogen bonds with other residues and water molecules in the binding site region. At least three shells of residues around the vicinity of the sugar participate in these networks; two of these shells are shown in Fig. 2. It is especially important to note that the formation of these networks leads to full utilization of all the functional groups of the essential side chain residues, the two water molecules, *every one* of the potential hydrogen bond donor groups of these residues (Arg 151, Asn 205 and Asn 232), and, as well, of the L-arabinose and two water molecules (see Fig. 2). This finding is the basis of concept (c) presented in Sect. 4.

Although Lys 10 differs from the rest of the essential residues, it is engaged in multiple interactions crucial to ligand binding. Its ammonium side chain is in an excellent position to donate a hydrogen bond to O-2, to fix via a salt-link the alignment of Asp 90 which is used in hydrogen-bonding the anomeric hydroxyl, and to make van der Waals' contacts or a very weak hydrogen bond with each of the anomeric hydroxyls.

The hydrogen-bonding scheme shown in Fig. 2, including the directions of the arrows, is entirely dictated by the result of the structure analysis of the complex. Remarkably, the scheme is also fully consistent. The networks of hydrogen bonds fix the essential residues and water molecules in favorable orientations for maximal sugar binding. Since some of the networks bridge the two domains, they further stabilize the "closed" conformation of the liganded protein. In this conformation, the two domains are in closer proximity to each other than in the postulated "open," unliganded conformation (NEWCOMER et al. 1981 b).

6 Roles of Isolated Bound Water in Protein–Sugar Interactions

Two water molecules (309 and 310, Fig. 1) are intimately involved in the binding of arabinose to the binding protein. Both water molecules are approximately tetrahedrally coordinated. Water 309 serves as an H bond acceptor of sugar OH-1 and further links the N-terminal domain (via Glu 14 and Gln 11) with the C-terminal domain (via Asn 205). Water 310 donates a hydrogen bond to the ring oxygen and is directly hydrogen-bonded to Asp 89 and Thr 147. Water molecules have the advantage of occupying a much smaller space than do the residues while still having unusual capacity to engage at least four H bonds. Moreover, water molecules can be more easily positioned to participate maximally in hydrogen bonding and in bridging binding site residues and the ligand.

7 Protein–Sugar Complexes Are Further Stabilized by Van der Waals' Contacts

Protein–sugar complexes are undoubtedly further stabilized by van der Waals forces. Since the structure of the protein–arabinose complex has been extensively refined, it has been possible to identify better and to describe some aspects

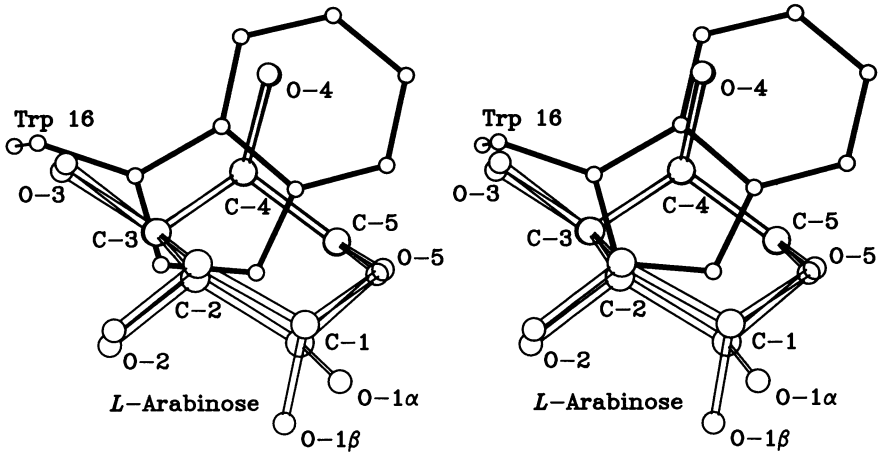


Fig. 3. Stereo drawing of the partial stacking of Trp 16 (*solid lines*) with the bound L-arabinose. A hydrophobic patch consisting of C-3, C-4, and C-5 is especially close to the Trp residue

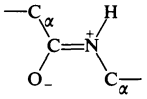
of these interactions (QUIOCHO 1986). All atoms of the arabinose substrates are utilized in these contacts, producing a total of about 45 van der Waals contacts in the range of 3.2–4.0 Å (QUIOCHO 1986). Of special significance is the finding that some of these contacts involve a cluster of nonpolar atoms within the L-arabinose. The disposition of both the equatorial OH-3 and the axial OH-4 to one side of the pyranose ring creates a hydrophobic patch composed of C-3, C-4, and C-5. As shown in Fig. 3, Trp 16 is partially stacked with the hydrophobic patch. This specific type of interaction (i.e., stacking of aromatic residues with a sugar ring) may be more common than therefore anticipated (QUIOCHO 1986) and should further confer substrate specificity by disallowing binding of epimers of sugar substrates because of steric hindrance or polarity of the hydroxyls. This has proven to be the case in the arabinose-binding protein since epimers of L-arabinose or D-galactose at positions C-3 and C-4 do not bind to the protein.

8 Bound Sugar, Essential Residues, and H Bonds Are Buried: Hydrogen Bonds and Uncompensated Charge Stabilization

Since most sugar binding sites are located in clefts, of which many form between domains, bound sugar substrates are frequently buried to differing extents (QUIOCHO 1986). For example, the solvent-accessible surface area of L-arabinose in the free state is reduced by 98% when bound in the cleft between the two domains (QUIOCHO and VYAS 1984). Moreover, almost all the hydrogen-bonding residues are also buried. The low dielectric constant environment resulting from enclosure of sugar substrate should increase the strength of hydrogen bonds in the complex.

Another significant discovery resulting from the structure analysis of the binding protein–arabinose complex is that two essential, charged residues (Glu 14 and Arg 151), which are buried and inaccessible to the bulk solvent, are uncompensated for. In particular there are no metal anions, negatively charged residues, or water molecules close to Arg 151, and although the residue resides on a helix, its guanidinium side chain is closer to the positive rather than to the negative end of the helix macrodipole. The fundamental question is, "How are the isolated charges of these sequestered residues, particularly Arg 151, stabilized?" We have proposed that the positive charge on the guanidinium side chain of Arg 151 is stabilized by way of the hydrogen bonds. Stabilization is achieved mainly via resonating peptide bonds, together with mutual polarization; two of the five hydrogen bonds to Arg 151 are formed with peptide bond carbonyl groups (see Fig. 2). The NH group of each of the two peptide bonds is in turn involved in a hydrogen bond array.

The existence of uncompensated buried charged residues may not be so unusual, since, as recently shown, the sulfate dianion that is completely sequestered deep in the ligand site cleft of the sulfate-binding protein is held in place mainly by seven hydrogen bonds; there are no countercharges within van der Waals' distance of the dianion (PFLUGRATH and QUIOCHO 1985). In this case it is evident that resonating peptide bonds, augmented by mutual polarization, are the major contributor to charge neutralization; five of the seven hydrogen bond donors to the sulfate are from peptide NH groups, and the CO groups of these five peptide bonds in turn initiate five different hydrogen bond arrays. Two of these arrays each have a positive charged residue as an H bond donor to the peptide CO group. Mutual polarization should stabilize the dipolar resonant structure I of the peptide involved and thus facilitate sulfate charge neutralization.



(I)

These experimental findings are prime direct evidence that isolated full charges buried in a solvent-inaccessible site within a protein can be stabilized by interactions via the mechanism outlined above for the uncompensated positively charged Arg 151 in the arabinose-binding protein and the dianion of the sulfate substrate buried in the sulfate-binding protein. These findings have far-reaching importance and implications in electrostatic interactions in biological macromolecules, ion-binding and -transport systems, enzyme catalysis, and other biochemical processes that require the stabilization of charged intermediates.

9 Sugar-Induced Protein Conformation Change

The finding that the arabinose bound to the receptor protein and many of the hydrogen-bonding residues are inaccessible to the bulk solvent provides strong evidence for a sugar-induced conformational change. Similar observa-

tions have been made in other binding protein–substrate complexes (VYAS et al. 1983; PFLUGRATH and QUIOCHO 1985). The nature of these structural changes is not fully understood, but we note that the peptide segments of these proteins that provide the groups directly hydrogen-bonded to the ligands are part of highly stable secondary and super-secondary structures in each domain. Therefore, access to and from the binding site is less likely to be achieved solely by widespread large concerted movements of these and other peptide segments and associated residues. Rather, a hinge-bending motion between the two domains could easily modulate the closing and opening of the cleft and further juxtapose for ligand-binding the essential residues poised in both domains. Results of low-angle X-ray scattering measurements (see Table 1) and theoretical calculations are consistent with this proposal (NEWCOMER et al. 1981 b; MAO et al. 1983). The decrease of 0.94 Å in the radius of gyration of the binding protein caused by arabinose binding (Table 1) can best be accounted for in terms of a substrate-induced cleft closure in which one domain rotates by 18° relative to the other domain about a hinge between the two domains (NEWCOMER et al. 1981 b).

10 Entropic Effects in Protein–Sugar Interactions

Because of several factors the contribution of the entropic effect to the formation of protein–sugar complexes is difficult to assess. Sugars are highly polar organic molecules; nevertheless, they have two types of solvent-accessible groups or surfaces at which water structures may differ: the polar hydroxyl groups that are engaged in hydrogen-bonding with some individual water molecules and the nonpolar groups that would have at their interface more locally ordered water structure. (Solvent-exposed binding site residues, which are also amphipathic, can be viewed in the same way.) In carbohydrates the polar surfaces generally exceed the hydrophobic surfaces. For example, in the free or unbound state of the L-arabinose, nonpolar groups constitute only about 25% of the solvent accessible area based on the RICHARDS (1977) criteria, the rest being confined to the polar hydroxyls and ring oxygen. Dehydration of both the sugar and the binding site accompanies protein–sugar complex formations, causing at least two effects on the structure of water adjacent to the two different surfaces. One of these effects, the hydrophobic effect, its contribution roughly proportional to the accessible area, leads to a disordering of the more highly organized water molecules at the interface of the nonpolar groups when these groups are buried in the complex. In the second effect, the hydrophilic effect, the polar hydroxyls and ring oxygen of the sugar substrate exchange their solvation shell of water for the polar groups of the amino acid residues that make hydrogen bond interactions in the binding site of proteins. This leads to an increase in the entropy of of the liberted individual water molecules that were originally hydrogen-bonded to the polar groups of the sugar and binding site residues. Since hydrogen bonds in protein–sugar complexes (especially those that are also part of networks) have fewer degrees of freedom than water,

they offer a more stable solvation shell for the bound sugar. There is immobilization of the sugar and protein as a result of complex formation, causing negative entropy due to loss of translation and rotation by virtue of the hydrogen bonds and van der Waals contacts formed in the complex.

From the foregoing discussions it is clear that an evaluation of the overall net effect of entropic forces (both from hydrophobic and hydrophilic effects) in protein–sugar interaction is fraught with difficulty. Nevertheless, thermodynamic characterizations of several different protein–carbohydrate associations provide important insight into the forces that stabilize protein–sugar interactions. In a recent review article, the present author has summarized the thermodynamic parameters for a total of eight complexes involving the extensively studied proteins lysozyme, arabinose-binding protein, wheat germ agglutinin, and hexokinase (QUIOCHO 1986). The complexes are: mono-, di-, tri-, and hexasaccharide of *N*-acetylglucosamine bound to the lysozyme; L-arabinose and D-galactose to the arabinose binding protein (data also shown in Table 1); α-(2-3)-neuraminyl-lactose to the wheat germ agglutinin; and D-glucose to hexokinase. In all cases, the values for $\Delta G°$ and $\Delta H°$ are all negative, favoring complex formation via processes that are enthalpically controlled. For all the complexes, with the exception of hexokinase, the values for $\Delta S°$ are also negative. It can be concluded from these thermodynamic data (see ROSS and SUBRAMANIAN 1981) that hydrogen bonding and van der Waals forces are the major factors in stabilizing protein–sugar complexes and not the hydrophobic effect. X-ray structures of several of these complexes, especially the binding protein–sugar complex, are consistent with this conclusion.

11 Concluding Remarks

We have discussed here many of the basic molecular features of protein–carbohydrate interactions, for the most part, as visualized at atomic resolution for the structure of binding protein–sugar complex. Several of these features, although not as firmly established, are also present in other protein– or enzyme–saccharide interactions (QUIOCHO 1986).

Since hydrogen bonds make a major contribution to the absolute energy or stability of interaction between protein and saccharide, several factors could affect or modulate this stability: (a) number of hydrogen bonds, (b) types of hydrogen bonds (neutral-neutral or neutral-charged), (c) "cooperativity," (d) geometry, (e) solvent accessibility, and (f) hydrogen bond networks. As clearly described above, not only do all these factors contribute to the stability of the binding protein–monopyranoside complex, but each is manifested nearly to the full extent. All the hydrogen-bonding groups of the L-arabinose are utilized, resulting in the formation of ten hydrogen bonds in the complex that are equally distributed between neutral-neutral and neutral-charged types of H bonds. With the exception of the C-1 hydroxyl, all sugar hydroxyls are involved in cooperative hydrogen bonding (i.e., hydroxyl participating simulta-

neously as donors and acceptors of hydrogen bonds). The parameters of the hydrogen bonds and the geometry of the cooperative hydrogen bonding indicate nearly excellent hydrogen bonds. The bound sugar and most of the hydrogen-bonding residues are buried and inaccessible to the bulk solvent. There are extensive networks of hydrogen bonds as a result of the formation of the protein–sugar complex. In view of these findings, it is not surprising that the binding of L-arabinose results in one of the tightest protein–sugar complexes.

Hydrogen bonds are of further importance in protein–sugar interactions. They are mainly responsible for conferring stereospecificity on the sugar binding site and ensuring correctness of fit for substrates. An additional essential feature of hydrogen bonds is that they are stable enough to provide significant overall ligand affinity but are of sufficiently low strength to allow rapid dissociation from the protein.

The binding protein–arabinose complex is further stabilized by van der Waals forces. Even considering this stabilization, the van der Waals contacts are numerous, especially for a bound monopyranoside substrate. Additional specificity is conferred by the partial stacking of aromatic residue with the non polar patch of sugars.

Because the binding protein–arabinose complex displays many, if not all, of the molecular features that might be expected of protein–carbohydrate interactions, it is an ideal system for further investigation. Moreover, the binding protein is very well characterized and, as stated above, many different techniques can be enlisted in these studies. For instance, we have used various deoxy sugars to probe the contribution of the interactions of each sugar hydroxyl to the overall affinity of the arabinose-binding protein–sugar complex (QUIOCHO et al. unpublished data). High-resolution structural analysis of other periplasmic sugar-binding proteins currently underway in our laboratory should provide further understanding of protein–carbohydrate interactions. In particular, study of the maltose-binding protein with bound oligosaccharides such as maltoheptaose offers the real possibility of not only unraveling the details of oligosaccharide binding but also obtaining accurate information on oligosaccharide conformations.

Binding sites of proteins have natural affinity and stereoselectivity toward substrates. The atomic structure of the binding protein–arabinose complex has enabled us to visualize clearly these two important properties.

Acknowledgements. I thank all my coworkers, particularly Dr. N. and K. Vyas, for their unstinting efforts and enthusiasm. I am especially grateful to Prof. Robert W. Hogg for providing the initial impetus for the study of binding proteins and for continued interest. The author's research was further supported by grants from the National Institutes of Health and the Welch Foundation.

References

Clark AF, Gerken TA, Hogg RW (1982) Proton nuclear magnetic resonance spectroscopy and ligand binding dynamics of the *Escherichia coli* L-arabinose binding protein. Biochemistry 21:2227–2233

Engelsberg E (1971) Regulation in the L-arabinose system. In: Vogel HJ (ed) Metabolic pathways: metabolic regulation 5. Academic, New York, pp 257–296

Fukuda H, Sturtevant JM, Quiocho FA (1983) Thermodynamics of the binding of L-arabinose and of L-galactose to the L-arabinose-binding protein of *Escherichia coli*. J Biol Chem 258:13193–13198

Mao B, Pear MR, McCammon JA, Quiocho FA (1982) Hinge-bending in L-Arabinose-binding protein. J Biol Chem 257:1131–1133

Hogg RW, Hermodson M (1977) Amino acid sequence of the L-binding protein from *Escherichia coli*. J Biol Chem 252:5135–5141

Miller DM, Olson JS, Pflugrath JW, Quiocho FA (1983) Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. J Biol Chem 258:13665–13672

Newcomer ME, Gilliland GL, Quiocho FA (1981a) L-Arabinose-binding protein–sugar complex at 2.4 Å resolution. Stereochemistry and evidence for a structural change. J Biol Chem 256:13213–13217

Newcomer ME, Lewis, BA, Quiocho FA (1981b) The radius of gyration of L-arabinose-binding protein decreases upon binding of ligand. J Biol Chem 256:13218–13222

Parsons RG, Hogg RW (1974) Crystallization and characterization of the L-arabinose binding protein of *Escherichia coli* B/r. J Biol Chem 249:3602–2607

Pflugrath JW, Quiocho FA (1985) Sulphate sequestered in the sulfate-binding protein is bound solely by hydrogen bonds. Nature 314:257–260

Quiocho FA (1986) Carbohydrate-binding proteins: tertiary structures and protein–sugar interactions. Annu Rev Biochem 55:287–315

Quiocho FA, Vyas NK (1984) Novel stereospecificity of the L-arabinose-binding protein. Nature 310:381–386

Richards FM (1977) Areas, volumes, packing, and protein structure. Annu Rev Biophys Bioeng 6:151–176

Ross PD, Subramanian S (1981) Thermodynamics of protein association reactions: forces contributing to stability. Biochem 20:3096–3102

Saper MA, Quiocho FA (1983) Leucine, isoleucine, valine-binding protein from *Escherichia coli*: structure at 3.0-Å resolution and location of the binding site. J Biol Chem 258:11057–11062

Vyas NK, Vyas MN, Quiocho FA (1983) The 3 Å resolution structure of a D-galactose-binding protein for transport and chemotaxis in *Escherichia coli*. Proc Natl Acad Sci USA 80:1782–1796

Subject Index