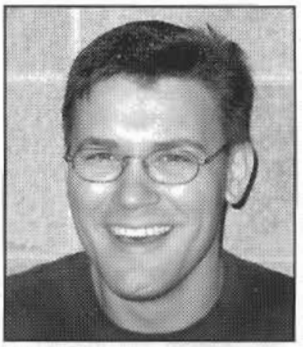
Mapping Protein: Carbohydrate Interactions

Gerald F. Audette[1]*, Louis T.J. Delbaere[2] and Jim Xiang[3†]

[1]*Department of Medical Microbiology & Immunology, University of Alberta, 1-15 Medical Sciences Building, Edmonton, AB, Canada, T6G 2H7;* [2]*Department of Biochemistry, University of Saskatchewan, Rm. A3.1, 107 Wiggins Road, Saskatoon, SK, Canada, S7N 5E5;* [3]*Departments of Microbiology, Oncology and Pathology, Saskatoon Cancer Centre, University of Saskatchewan, Saskatoon, SK, Canada, S7N 0W0*



Abstract: Many biologically important interactions occur between proteins and carbohydrates. The examination of these interactions at the atomic level is critical not only in understanding the nature of these interactions and their biological role, but also in the design of effective modulators of these interactions. While experimentally obtained structural information is preferred, quite often this information is unavailable. In order to address this, several methods have been developed to probe the interactions between protein and carbohydrate in the absence of structural data. These methods map the interactions between protein and carbohydrate, and identify the groups involved, both at the carbohydrate and protein level. Here, we review these developments, and examine the strengths, weaknesses, and pitfalls of these methods.

Keywords: Protein:carbohydrate interactions, alanine scanning, chemical mapping, molecular modeling, crystal structure

1. INTRODUCTION

Protein:carbohydrate interactions are involved in a wide range of biological processes. These processes include, among others, infection by invading microorganisms and the subsequent immune response, leukocytic trafficking and infiltration, and tumor metastasis [1-9]. Carbohydrates are uniquely suited for this role in molecular recognition as they possess the capacity to generate a wide array of structurally diverse moieties from a relatively small number of monosaccharide units. In order to decode the specific molecular signal presented by a carbohydrate epitope, carbohydrate-binding proteins must present a non-catalytic binding site complimentary to the carbohydrate. The immunoglobulins employ their complimentary-determining regions (CDR) for carbohydrate recognition during the immune response. Other carbohydrate-recognizing proteins are known as the lectins, a diverse group of unrelated proteins that employ a binding site referred to as the carbohydrate recognition domain (CRD).

The specificity of the protein:carbohydrate interactions presents an opportunity to design selective modulators, such as competitive inhibitors, of these interactions. These modulators are synthetic compounds that bind with high affinity and specificity to the carbohydrate-binding site. The search for such modulators of protein:carbohydrate interactions requires detailed information into the groups involved in binding, both at the protein and carbohydrate levels. Detailed experimental analysis of the protein, carbohydrate, and the resultant complex, obtained *via* X-ray crystallographic methods, is preferred in order to obtain a clear understanding of the interactions involved. However, the generation of high quality crystals suitable for X-ray diffraction experiments is often a significant hurdle in this process. Unbound carbohydrates are particularly difficult to crystallize, likely as a result of their inherent flexibility, and are therefore mainly examined *via* nuclear magnetic resonance (NMR) methods [10-13].

Several developments have enabled the examination of protein:carbohydrate interactions in the absence of experimentally obtained structural information of the protein:carbohydrate complex. Modern computing allows for the modeling of the protein:carbohydrate complex using the structure of the native protein as a template, which enables the investigator to identify the potential interactions that occur upon carbohydrate binding. Also, the ease of modern recombinant DNA technologies provides a robust means of mutational analysis of the amino acid residues involved in carbohydrate binding at the protein level. Finally, the synthesis of complex oligosaccharides and their congeners, modified derivatives of the parent carbohydrate, allows for the identification of the groups involved at the carbohydrate level. Computational methods into protein modeling, including sequence comparison, structure prediction and molecular docking, have been discussed elsewhere [14-19]. Here, we examine developments into the examination of protein:carbohydrate interactions in the absence of detailed structural information of the complex.

2. MAPPING THE PROTEIN

2.1 Directed Alanine Scanning

The assessment of the contribution of individual amino acids to carbohydrate binding is central to the identification

†Address correspondence to this author at the Departments of Microbiology, Oncology and Pathology, Saskatoon Cancer Centre, University of Saskatchewan, Saskatoon, SK, Canada, S7N 0W0; Tel: 306-655-2917; Fax: 306-655-2635; E-mail: jxiang@scf.sk.ca

of the functional epitope involved in carbohydrate binding. The use of alanine mutations in this evaluation is particularly useful in that the alanine mutation truncates the side chain at the β-carbon, enabling the investigation of the particular side-chain in carbohydrate binding. The alanine mutation is also useful in that an alanine does not introduce potential main-chain flexibility that can arise through replacement of a residue with a glycine. The systematic mutation of protein residues to alanine, known as alanine scanning, has been widely employed in the examination of protein-protein interactions [20-23]. As a *de novo* process, Ala scanning is a laborious endeavor, and as such other combinatorial methods of alanine mutant generation have been developed [24-25]. In the study of particular protein:carbohydrate interactions, a *de novo* examination of the protein *via* alanine scanning is unnecessary. Rather a directed examination of the residues believed to be involved in the formation of the complex, through the use of molecular modeling and similarities between carbohydrate binding proteins [26-31], provides clear indications into the roles specific amino acid residues play in the binding of the carbohydrate ligand.

2.1.1 Examples

2.1.1.1 CBM17 from Clostridium Cellulovorans Cel5A

Many glycoside hydrolases derived from aerobic microorganisms are modular proteins, which contain both catalytic and non-catalytic domains. The non-catalytic domain, often referred to as the carbohydrate binding module (CBM), has the primary function of associating the hydrolase with the substrate [32]. The CBMs are classified into ~20 families based on sequence similarities [33]. The family 17 CBM from *Clostridium cellulovorans* Cel5A (CBM17) was shown to bind non-crystalline cellulose in an extended binding site with an optimal affinity for cellohexaose [34]. Notenboom and colleagues have recently reported the examination of CBM17 binding using a combination of X-ray crystallography, alanine scanning and isothermal titration calorimetry [35]. This study examined both the uncomplexed CBM17 and CBM17 in complex with cellotetraose, and identified an extended, surface exposed binding site for CBM17 which provides an extensive hydrogen-bonding network for the carbohydrate. The cellotetraose complex served as a template for the mutagenesis and calorimetric studies. Mutation of several polar residues to alanine resulted in a significant decrease in binding affinity, and revealed that water may play an important role in binding. Two tryptophan residues, W88 and W135, were identified as being important in cellohexaose binding by CBM17, which is consistent with previous studies indicating two tryptophan residues interacting with two or three glucose units of cellohexaose [34]. Finally, this study provided a comparison between other CBMs and implications in binding specificity of crystalline *versus* non-crystalline forms of cellulose observed in different CBM families.

2.1.1.2 The B72.3 Fab

The mouse monoclonal antibody B72.3 recognizes the tumor-associated glycoprotein TAG72, and is employed in the detection of human malignancies [36]. B72.3 recognizes dimeric sTn-serine clusters of the *O*-linked carbohydrate NeuAc2-6αGalNacα1-O-Ser/Thr (sTn^2) [37-38]. While the crystal structure of the uncomplexed chimeric B72.3 Fab has been determined [39], no structural data on the B72.3:sTn^2 complex is available. Xiang and colleagues recently reported the examination of the B72.3:sTn^2 complex through the use of molecular modeling and directed alanine scanning [40]. The modeling, employing the coordinates of B72.3 Fab structure [39] as a template and a generated sTn^2 epitope, identified 17 CDR residues likely to be involved in interacting with the sTn^2 epitope, 11 heavy (H) chain CDR residues and 6 light (L) chain residues. Three of the 17 CDR residues, two H-chain and one L-chain, had been identified previously [41], and two other L-chain CDR residues were predicted to interact with the sTn^2 *via* main-chain atoms; these residues were not selected for mutation. Eleven of the remaining 12 predicted CDR residues (9 H-chain and 3 L-chain) were subjected to alanine mutagenesis and the affinity of binding to sTn^2 assessed; one residue, Ala33 of the H-chain, was mutated to Leu. Of the 12 predicted residues that were mutated in this study, eight exhibited a reduction in binding affinity upon mutation to alanine (Table **1**). The H33 Ala→Leu mutant exhibited an 84-fold decrease in binding affinity for the TAG72 antigen. The Ala was predicted to make several important van der Waals interactions with the sTn^2 epitope; mutation to Leu would result in the introduction of steric clashes. The role of hydrophobic interactions in binding was also supported with the H96 Tyr→Ala mutant, which exhibited a 40-fold decrease in binding affinity (Table **1**). These data, coupled with a previous observation that a Tyr→Phe mutant did not affect binding [41], suggesting that the hydroxyl group of Tyr96 plays less of a role than the aromatic moiety in antigen binding.

Table 1. Alanine Scanning of B72.3[1]

CDR Residue[2]	Mutation	Binding Affinity Reduction[3] (-fold)
H32	His → Ala	52
H33	Ala → Leu	84
H35	His → Ala	-
H50	Tyr → Ala	30
H52	Ser → Ala	17
H54	Asn → Ala	15
H56	Asp → Ala	8
H58	Lys → Ala	12
H96	Tyr → Ala	40
L32	Asn → Ala	-
L92	Trp → Ala	-
L94	Thr → Ala	-

[1]Adapted from Ref. [40].
[2]Residue numbering: H, heavy chain; L, light chain.
[3]Reduction in binding affinity is compared to the unmodified B72.3 Fab.

All three L-chain mutants exhibited no decrease in binding affinity; the authors felt that these residues, being at the terminal portion of the CDR, may not make significant contacts with the flexible sTn^2 epitope. The fourth mutant which exhibited no decrease in binding affinity of the TAG72 antigen was the H35 His→Ala mutant (Table **1**). This was unexpected due to the hydrogen-bonding capacity of His, which are commonly observed in CDR's of carbohydrate specific antibodies [26-27]. The authors felt that this could be rationalized in light of the observance that some contact residues exhibit dominance in the energetics of carbohydrate binding [42], or may be a reflection of the limitations of the modeling [40].

2.2 Prediction of Glycosylation Targets

The development of selective and directed inhibitors of glycoprotein synthesis is of significant pharmacological interest. Targeting the protein-carbohydrate linkage is an effective strategy, as this linkage is a distinguishing feature of glycoproteins [1]. The design of inhibitors of the glycosyltransferase reaction requires insight into both the epitope of the transferase itself as well as that of the target protein. The UDP-GalNAc:polypeptide *N*-acetylgalactosamine transferase (GalNAc-transferase), which catalyzes the transfer of a GalNAc moiety to a Ser or Thr residue on the target protein [43], is one glycosyltransferase that has recently come under such scrutiny.

The GalNAc-transferase has been shown to have an extended binding site which depends on the accessibility of the target sequence rather than its particular structure and sequence [44-45]. This extended binding site is comprised of a nonapeptide centered on the Ser/Thr residue (denoted R_0) to which the GalNAc will be transferred. The four N-terminal residues are labeled R_4-R_1, and the four C-terminal residues R_1-R_4. In an initial examination of 196 *O*-glycosylation sites extracted from the National Biomedical Research Foundation (NBRF) Protein Database [46], Elhammer and colleagues identified that, with the exception of the Ser/Thr at R_0, the GalNAc-transferase does not exhibit a readily identifiable amino acid sequence specificity across the extended nonapeptide [45]. Thus, development of a peptide-based competitive inhibitor for the GalNAc-transferase would require the examination of an enormous number of potential nonapeptides ($20^8 = 5.12 \times 10^{10}$). In order to streamline this process, Elhammer *et al.* applied a statistical model previously used to examine the HIV-1 protease [47] to probe the specificity of the GalNAc-transferase [45]. This method, referred to as the *h*-function method, predicts the probability that a nonapeptide will be glycosylated based on the frequency of a given amino acid residue at a particular subsite of the peptide (R_1,R_1', etc.) relative to a reference set of globular proteins [48]. The given peptide is predicted to be glycosylated if the value of the *h*-function is higher than a certain cutoff value (h_c). It was determined [45] that the optimal value of h_c was 0.19, this value being a compromise between over- and under-predicting glycosylation. When applied to the 196 glycosylated peptides from the NBRF, the *h*-function correctly predicted glycosylation events at a rate of 78.1% [45].

While the initial trial of the *h*-function method [45] in predicting glycosylation was generally successful, it was felt that the method might not yet be optimal. For example, the *h*-function did not account for differing glycosylation rates between Ser and Thr at R_0. Also, when an amino acid did not occur at a particular subsite, an arbitrary value was assigned in order to make the *h*-function non-zero. The *h*-function was therefore refined through use of the vector projection method [49], and sequence-coupling [50] in order to take these considerations into account. The resultant discriminant function, when applied to the training set from [45], correctly predicted glycosylation at a rate of 97.4% and 100% for the positive and negative training sets, respectively [50]. Applying this function to a test set of 30 oligopeptides, the success rate was improved to 90% from 70% for the *h*-function [50]. This method has more recently been employed in the context of artificial neural networks [51] and support vector machines [52]; this coupling should streamline the process of predicting oligopeptide inhibitors of GalNAc-transferases.

3. MAPPING THE CARBOHYDRATE

A separate approach to mapping protein:carbohydrate interactions is through examination of the carbohydrate itself. The development of methodologies to synthesize complex oligosaccharides has greatly enhanced the capability to identify critical groups of the carbohydrate recognized by the protein upon binding [53]. The ability to chemically synthesize complex oligosaccharides provides access to congeners such as monodeoxy and mono-*O*-alkyl (in particular, mono-*O*-methyl) derivatives of the parent carbohydrate. By identifying the effects of these slight structural changes on binding, a chemical map of the epitope, which groups interact between the two molecules upon binding, can be predicted [54]. Furthermore, coupling this chemical map of the carbohydrate with structural studies of the uncomplexed carbohydrate *via* NMR can provide significant insight into the tertiary structure of the carbohydrate epitope which is presented to the protein for binding.

3.1 The Chemical Mapping Method

The chemical mapping method was developed by Lemieux and colleagues in order to detect intermolecular hydrogen bonds between protein and oligosaccharide, in particular in the absence of crystallographic data of the protein:carbohydrate complex. The method was developed based on the analysis of the crystal structure of the fourth lectin from the legume *Griffonia simplicifolia* (GS-IV) in complex with the methyl glycoside of the Lewis b human blood-group determinant (Le^b-OMe), a tetrasaccharide [55], and was initially reported by Nikrad *et al.* [54]. The chemical mapping method produces congeners of the parent carbohydrate, and results in a series of deoxy and mono-*O*-methyl derivatives of the native carbohydrate. The effect of these alterations to the parent carbohydrate are evaluated in comparison to the parent carbohydrate using either ELISA [56] and/or radioimmunoassay [57] protocols, and retention of conformation for the carbohydrate congeners is monitored

Table 2. The Chemical Mapping Method

Effect of Congener[1]		Conclusions
Deoxy-	Mono-O-methyl	
Abolished Binding	Abolished Binding	Parent hydroxyl is deep within binding site and provides critical interactions
Greatly Diminished Binding	Little Change in Binding	Parent hydroxyl is located at the periphery of the binding site
Little Change in Binding	Little Change in Binding	Parent hydroxyl is solvent exposed in the complex

[1]The effect of the congener is based upon the binding affinity to the protein, with the unaltered carbohydrate at 100%.

via NMR [10-12]. Finally, recent advances in the synthesis of congeners with intramolecular tethers [58-61] enables the investigation of inter-saccharide flexibility/rigidity in carbohydrate binding. Coupling these data with hard sphere exo-anomeric effect (HSEA) calculations [62] results in the generation of a three dimensional model for the carbohydrate epitope recognized by the protein.

Several inferences can be made when examining the results of the binding studies into the congeners of the parent carbohydrate (Table **2**). When both the deoxygenation and the *O*-methylation of the H-type 2-OMe essentially abolish binding, the conclusion can be made that these hydroxyl groups provide critical hydrogen bonds with the protein. Also, when deoxygenation significantly decreases binding but the *O*-methylation causes little change in (and on occasion strengthens) binding, the indication is taken that the parent hydroxyl group resides at the periphery of the binding site in the complex where it will interact with peripheral amino acid residues. Finally, when small changes in binding are seen as a result of the structural modifications, it may be interpreted as arising from changes in hydration of polar groups which remain in the aqueous phase.

3.2 Examples

3.2.1 Aspergillus Niger ***Glucoamylase***

The chemical mapping protocol was applied to the amyloglucosidase (EC 3.2.1.3, glucoamylase) from *Aspergillus niger* [63], an industrially relevant enzyme that catalyzes the removal of β-D-glucose from the non-reducing ends of maltooligosaccharides [64-65]. This study was initiated due to the availability of structure of the native protein [66-67] and two inhibitor complexes [68-70]; the crystal structure of the enzyme:substrate complex was unavailable. The reference compound for the study was chosen to be methyl α–isomaltose (Fig. **1**) due to greater ease in derivative synthesis *versus* the corresponding maltoside. The effects of structural changes were monitored *via* alterations in the free energy of activation ($\Delta\Delta G^{\ddagger}$) of the enzyme with respect to the reference substrate [63]. The authors modeled transition state of the complex, revealing a network of hydrogen bonds between protein and carbohydrate; chemical mapping results of the carbohydrate were in agreement with the transition state model. Large increases in $\Delta\Delta G^{\ddagger}$ observed for deoxy and mono-*O*-methyl derivatives of OH-4 and OH-4' (Table **3**) indicate that these hydroxyls are in close contact with the protein; similar results were observed for OH-6, implicating it in catalysis. OH-2' and OH-3', which exhibit modest increases in $\Delta\Delta G^{\ddagger}$ in deoxy congeners and significant increases upon mono-*O*-methylation (Table **3**), are thought to be involved in substrate specificity. This study also indicated the role of ordered water molecules, not only in catalysis, but also in the maintenance of hydrogen bonds between carbohydrate and protein upon deoxygenation noting that this phenomenon has been discussed elsewhere [71-73]. The authors noted that the role of ordered waters would be best addressed through analysis of the crystal structure of glucoamylase in complex with either the 2-deoxy or 3-dexoy congener of the terminal cyclohexene unit of acarabose [63]; however, such a structure has yet to be reported.

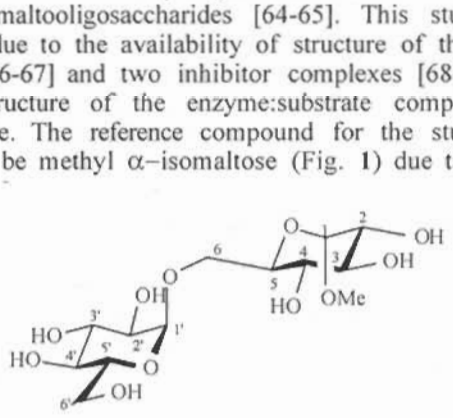


Fig. (1). Methyl α–isomaltose, used to probe the *A. niger* glucoamylase binding site. Atomic numbering is as in [63], and the effects of chemical mapping upon the activation energy of glucoamylase is shown in Table **2**.

3.2.2 Pseudomonas Aeruginosa PAK Pilin

Pseudomonas aeruginosa is an opportunistic pathogen that is of major concern to immunosuppressed or immunocompromised patients [74-75]. *P. aeruginosa* employs type IV pili, long fibres assembled from ~15kDa pilin monomers, as the dominant factor to initiate infection [76-77], and are thus promising therapeutic targets. While each pilin monomer contains a functional receptor-binding site, Lee and colleagues [78] demonstrated that the binding site is displayed only at the tip of the pilus structure. Pilin is known to bind to the glycosphingolipids asialo-GM_1 and asialo-GM_2 [78-79], and Sheth *et al.* [80] showed that the disaccharide β-D-GalNAc(1-4)β-D-Gal, present in both asialo-GM_1 and asialo-GM_2, is sufficient to inhibit pilin binding. In order to identify which of the seven functional groups (six hydroxyl and one *N*-acetyl) were involved in binding, the β-D-GalNAc(1-4)β-D-Gal disaccharide was subjected to chemical mapping and the inhibitory effectiveness of each congener was determined [81-82]. Surprisingly, all congeners of the disaccharide showed similar binding affinity to the parent compound and did not differ significantly as inhibitors of pilin binding [81]. However, one congener, the Gal-O2-propyl derivative,

Table 3. Chemical Mapping of Methyl α–Isomaltose and its effect on the Activation Energy ($\Delta\Delta G^{\ddagger}$) of *Aspergillus niger* Glucoamylase[1]

Congener	$\Delta\Delta G^{\ddagger}$ (kcal/mol)[2]	Congener	$\Delta\Delta G^{\ddagger}$ (kcal/mol)[2]
2-Deoxy	0.26	2-O-Me	-0.047
3-Deoxy	2.1	3-O-Me	1.9
4-Deoxy	3.9	4-O-Me	4.9
2'-Deoxy	0.96	2'-O-Me	4.0
3'-Deoxy	0.52	3'-O-Me	4.5
4'-Deoxy	5.5	4'-O-Me	6.2
6'-Deoxy	5.8	6'-O-Me	4.9

[1]Adapted from [63].
[2]$\Delta\Delta G^{\ddagger}$ measured in comparison to methyl α-isomaltose, Fig (**1**), set at $\Delta\Delta G^{\ddagger} = 0$ kcal/mol.

exhibited almost an order of magnitude greater inhibition of pilin binding compared to the parent compound [81]. From these results, the authors inferred that the binding of β-D-GalNAc(1-4)β-D-Gal by the PAK pilin was mainly influenced by hydrophobic forces. This is unusual in that interactions between protein and carbohydrate generally contain hydrogen bonding in addition to hydrophilic interactions [8, 28-31, 71, 83-85]. However, an earlier NMR study of a PAK pilin peptide in complex with a receptor analog suggests that the pilin CRD may interact with the apolar surface of the carbohydrate, and that this interaction dominates the overall free energy of binding [86].

Subsequent structural studies of truncated forms of *P. aeruginosa* pilins [87-88] do not support the model of a hydrophobic mediated receptor binding. The crystal structure of the truncated PAK pilin [87] indicated that receptor binding may be mediated mainly through interactions with main-chain atoms which could provide hydrogen bond donors and acceptors, but did not exhibit an extended hydrophobic patch which could be explained by the chemical mapping studies. The authors also indicated that while some modeling of the pilin:carbohydrate complex was performed, no single significantly more plausible conformation was observed [87]. Truncation of the pilin removes the first 28 N-terminal residues, which form a N-terminal hydrophobic helix; however, the truncated pilins retain the biological characteristics of the intact pilin. Indeed, the truncated K122-4 pilin retains binding affinity to asialo-GM_1, both in binding assays and in *in vivo* studies [88], as does the PAK pilin (R.T. Irvin, personal communication). It has been suggested (B. Hazes, personal communication) that the chemical mapping results were masked due to hydrophobic interactions between the N-terminal helix of the pilin and the hydrophobic linker of the disaccharide employed in the mapping studies [81]. This may indeed be the case due to the observance that the addition of a terminal methyl ester to the hydrophobic linker of the β-D-GalNAc(1-4)β-D-Gal disaccharide resulted in an almost threefold decrease in pilin binding inhibition [81]. It is possible that the pilin CRD provides the hydrogen bonding requirements for carbohydrate binding and that the hydrophobic N-terminal helix of the intact pilin adds increased binding affinity through hydrophobic interactions. It is also plausible that the pilin CRD interacts with the apolar surface of the carbohydrate upon binding as suggested from the NMR studies [86]. However, crystallographic analysis of the pilin in complex with the β-D-GalNAc(1-4)β-D-Gal disaccharide will be required to address this.

3.2.3 The Human H-type 2 Blood-Group Determinant

The H-type 2 human blood-group determinant, the trisaccharide α-L-Fucα(1→2)-β-D-Galβ(1→4)-β-D-GlcNAcα-R shown in (Fig. **2**), plays a central role in the ABO human blood-group system, the dominant determinant system for transplant and transfusion considerations [89]. The H-type 2 trisaccharide not only is the determinant of O-type erythrocytes, but is also the precursor of both the A and B blood group determinants. The first lectin from seeds of the legume *Ulex europaeus* (UE-I) was identified as the agglutinating agent of O-type erythrocytes by Matsumoto and Osawa [90], following which studies by Watkins [91] and others [88, 92-95] revealed that the H-type 2 trisaccharide was indeed the carbohydrate recognized by UE-I in the agglutination reaction.

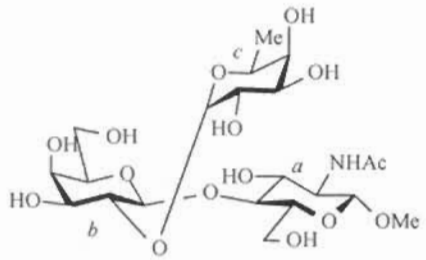


Fig. (2). The methyl glycoside of the H-type 2 blood-group determinant [α-L-Fucα(1→2)-β-D-Galβ(1→4)-β-D-GlcNAcα-OMe] (H-type 2-OMe). Structural units *a*, *b* and *c* are β-D-GlcNAc-OMe, β-D-Gal and α-L-Fuc, respectively.

The application of the chemical mapping method to the H-type 2 trisaccharide revealed several key interactions between the lectin and carbohydrate (Table **4**) [93, 96]. Chemical mapping indicates that the three hydroxyls (O2*c*, O3*c*, O4*c*) of the α-L-Fuc unit provide critical hydrogen bonds to UE-I, the 3*b*-hydroxyl group of the β-D-Gal unit is interacting at the periphery of the UE-I binding site, and that the 3*a*-hydroxyl group of the β-D-GlcNAc unit remains within the aqueous phase. In addition to these polar interactions, replacement of the 6*a*-OH group of the β-D-GlcNAc moiety by a hydrogen atom more than doubled the binding potency, indicating a hydrophobic interaction. Also, the hydrophobic *C*-5-CH_3 of the fucose is close enough to be a part of this hydrophobic region, and its replacement by hydrogen has a significant deleterious effect on binding (Table **4**). The authors felt that the replacement of the methyl group by a hydrogen atom may have a profound effect on the conformational preference of a building unit that provides three key hydroxyl groups to the binding region [97]. Finally, the results of the chemical mapping suggested a role

Table 4. The Chemical Mapping of the H-type 2-OMe Recognized by UE-I[1]

Position Altered[2]	Relative Potencies[3]	
	Monodeoxy	Mono-O-Methyl
3a	46	40
6a	250	150
3b	3	70
4b	37	44
6b	60	45
2c	0.6	Inactive
3c	0.11	Inactive
4c	Inactive	Inactive
5c (*nor*)[4]	3	-

[1]Adapted from Ref. [96].
[2]Structural unit designations are as described in Fig. (**2**).
[3]By radioimmunoassy, setting the potency of H-type 2-OMe at 100 [97].
[4]Ref. [93].

for an ordered water structure in binding of the carbohydrate by UE-I [73, 96].

The lectins from leguminous plants have, in recent years, come under significant investigation as paradigms for protein:carbohydrate interactions. The legume lectins differ widely in their carbohydrate specificity and quaternary structure; however, the tertiary structure of the legume lectin monomer is relatively constant [28-29, 83-84]. Indeed, the most widely observed lectin fold is that of the legume lectin [31]. Thus it has been possible to model the protein:carbohydrate complex of a legume lectin in the absence of crystallographic data for the uncomplexed protein. Gohier and colleagues [98] reported the modeling of the UE-I monomer using the published primary sequence of UE-I [99]. This study included the modeling of α-L-fucose, the monosaccharide which exhibits primary binding specificity for UE-I, into the UE-I CRD and predicted four separate possible binding modes. The predicted binding of α-L-fucose could correlate with the chemical mapping studies [96], though the authors did note that the number of hydrogen bonds between protein and carbohydrate were low in comparison to other complexes [98].

In the report of the uncomplexed UE-I [100], the binding of the H-type 2 trisaccharide was predicted through the docking of the carbohydrate epitope described by chemical mapping (Fig. **3a**). The model of the UE-I:H-type 2-OMe complex predicted several potential hydrogen bonds which could explain the chemical mapping studies. For example, all three hydroxyls of the fucopyranose moiety were predicted to make hydrogen bonds with residues at the bottom of the CRD; disruption of these interactions *via* deoxygenation or mono-*O*-methylation would lead to greatly diminished or abolished binding. The model also allowed for the contribution of an ordered water structure as suggested by Lemieux [73], and the conformation of the trisaccharide was similar to that predicted in recent tethering studies of the H-type 2 trisaccharide [61]. It was noted that the UE-I CRD differed from that predicted in the Gohier *et al*. [98] study; the differences were attributed to anomalies between the published sequence [99] and the observed electron density [100].

The recent examination of two complexes of UE-I, UE-I in complex with the methyl glycosides of α-L-fucose (α-L-Fuc-OMe) and the H-type 2 trisaccharide (H-type 2-OMe), has provided an experimental evaluation of the chemical mapping method [101]. The bound α-L-Fuc-OMe in the UE-I:α-L-Fuc-OMe complex (Fig. **3b**) makes several hydrophilic and hydrophobic interactions with amino acids of the CRD. These interactions correlate well with the chemical mapping data (Table **4**), including the role of the C-5-CH_3 group of the fucose moiety in carbohydrate binding. This group, shown to produce an inactive inhibitor of UE-I binding when replaced by a hydrogen (Table **4**) [93], was observed to be in a close van der Waals interaction with $C^{\gamma 2}$ of the Thr86 side-chain and several van der Waals contacts with hydrophobic residues of the UE-I CRD. The presence of a Thr at position 86 of UE-I is a unique feature of the Fuc-specific legume lectins; the residue at this position is generally Ala [28-29]. This study also examined a low resolution (3.0 Å) structure of the UE-I:H-type 2-OMe complex, which provided insight into the conformation of the complete epitope required for binding. A comparison of the crystallographically observed and predicted models of the carbohydrate bound by UE-I revealed that while both studies explain the chemical mapping studies, the carbohydrate adopts different orientations between predicted [100] and observed [101] complexes (Fig **3c**). The differing orientations were felt to arise from the participation of residues unexpected from examination of the native structure, such as the interaction between Thr86 C^{γ} and C-5-CH_3 of the α-L-Fuc-OMe molecule [101], which led to an inaccurate starting model of the complex which correlated with the chemical mapping data [100].

4. DISCUSSION & OUTLOOK

Directed alanine scanning has proven to be effective in the identification of the requirements in protein-protein interactions, and has recently been extended to protein:carbohydrate interactions. The examination of CBM17 has shown that this method may successfully identify residues involved in substrate binding through directed alanine scanning based on the structure of a related complex [35]. Also, the directed alanine scanning of the B72.3 Fab [40] has identified several critical residues important in antigen recognition, and may lead to an effective humanization of the chimeric antibody for therapeutic use.

The recent advances in the prediction of *O*-linked glycosylation targets is an important step towards the design of peptide-based inhibitors against GalNAc-transferases. The successes of the discriminant function [50] in the context of neural networks [51] and support vector machines [52] in predicting glycosylated nonapeptides are indeed promising. However, the method thus far has been only tested on a

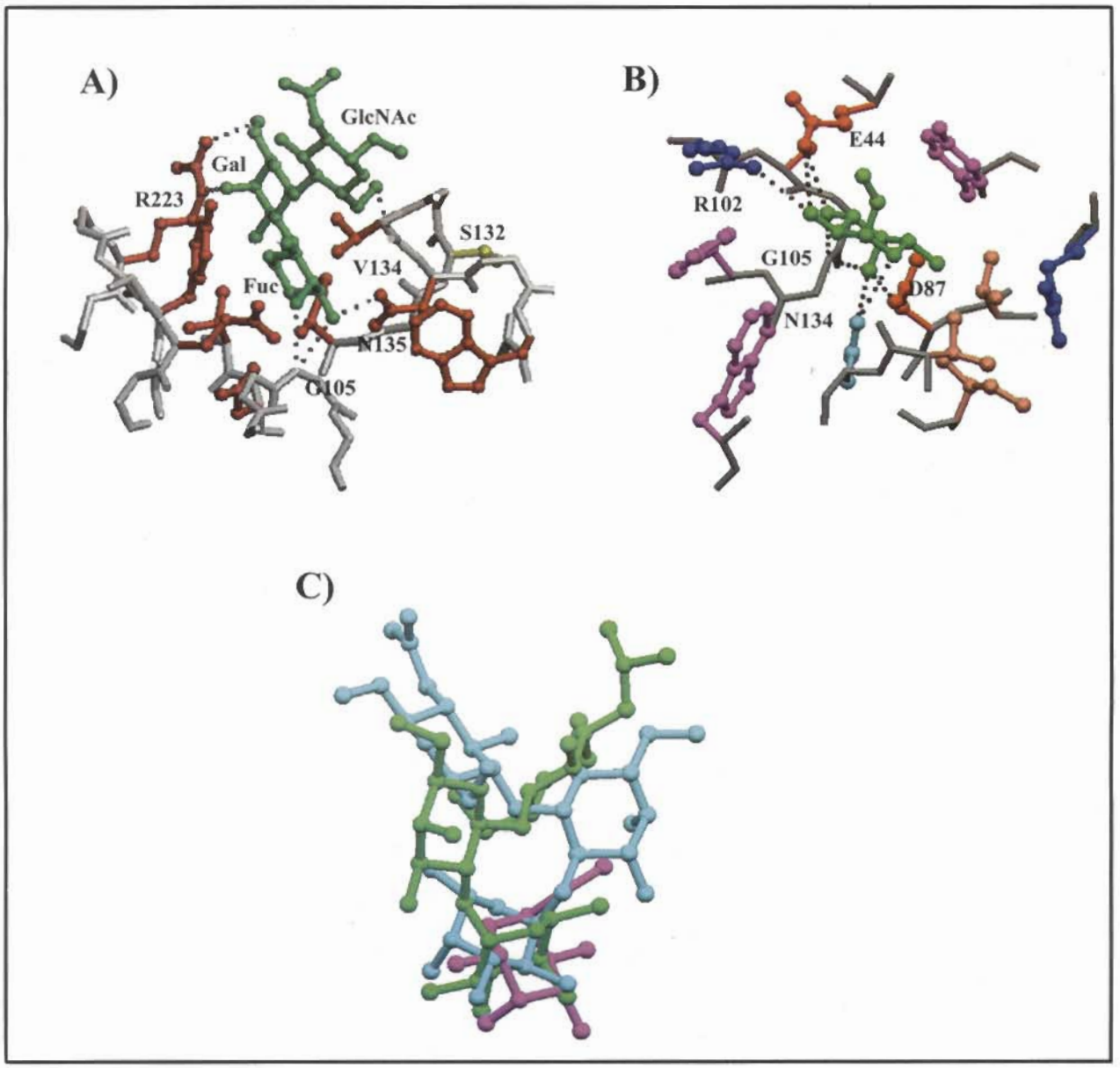


Fig. (3). The binding of H-type 2-OMe by UE-I. **a**) The docked H-type 2-OMe within the carbohydrate binding site of the native UE-I [100]. The docked carbohydrate is in green, main-chain atoms are in gray, and side-chains of the CRD residues are in red. Serine 132, thought to be involved in carbohydrate binding *via* earlier modeling studies [98], is in yellow. **b**) The binding of α-L-Fuc by UE-I, as observed in the crystal structure of the complex [101]. The bound α-L-Fuc-OMe is in green, main-chain atoms are in gray, and side-chains of the CRD residues are in standard colors (as per SETOR [103]). Both the modeled (**a**) and observed (**b**) of the UE-I:carbohydrate complex correlate with the chemical mapping data (Table **3**). **c**) The differing orientations of the carbohydrate in the UE-I CRD between the predicted model [100] and crystallographically observed [101] complexes. The modeled H-type 2-OMe is in green and the crystallographically observed α-L-Fuc-OMe and H-type 2-OMe molecules are in magenta and cyan, respectively. This figure was produced using the program SETOR [103].

limited test set (30 oligopeptides) [50]. A larger randomized test set should be examined in order to evaluate the robustness of the method and identify further potential refinements. Also, these studies should be coupled with binding and inhibition studies similar to those of Elhammer *et al.* [45] in order to asses the effect of particular amino acid residues in given subsites on binding/inhibitory power. The employment of this refined prediction protocol should provide an efficient starting point in the design of potential lead compounds.

Chemical mapping can provide a significant amount of insight into the carbohydrate epitope presented to the protein for binding. This method has an advantage in that the information obtained may be directly employed in the design of selective inhibitors. However, caution should be taken when interpreting the results of chemical mapping. As observed in the study of UE-I, while both the modeling [100] and crystallographically observed complexes [101] satisfied the chemical mapping data [96], the predicted and observed complexes differ in the orientation of the carbohydrate within the UE-I CRD. Also, during the investigation of the interaction between the PAK pilin and β-D-GalNAc(1-4)β-D-Gal [81], the mapping studies may have been masked by unexpected hydrophobic interactions with a highly hydrophobic surface feature subsequently proposed to reside near the putative binding site [87]. However, structure determination of the PAK:β-D-GalNAc(1-4)β-D-Gal complex is required in order to clarify the mode of binding.

The protein:carbohydrate interaction is central to numerous biological processes. Technologies such as oligosaccharide microarrays [102] have increased the rate of identification of new carbohydrate-binding proteins, presenting investigators with novel therapeutic targets. Recent developments in alanine scanning, glycosylation prediction and chemical mapping provide insight into protein:carbohydrate interactions in the absence of structural data of the complex, thereby providing a framework for the design of potential lead compounds. These methods however are labour intensive, requiring not only mutant/congener production, but also significant efforts in purification and assay optimization. Though the production of mutant proteins is less of a hurdle through modern recombinant DNA technologies, the generation of carbohydrate congeners is an expensive and time consuming process. Therefore, the full capacity of the chemical mapping method will require a significant streamlining of these processes in order to readily produce carbohydrate congeners for examination and testing.

ACKNOWLEDGEMENTS

We wish to thank Bart Hazes and Randy Irvin for helpful discussions and review of drafts of the manuscript. This work is funded by Canadian Institutes of Health Research grants MT-10162 to LTJD and ROP-15151 to JX.

ABBREVIATIONS

CDR	=	Complimentary-determining region
CRD	=	Carbohydrate recognition domain
CBM	=	Carbohydrate binding module
HSEA	=	Hard Sphere ExoAnomeric effect
GlcNAc	=	*N*-acetylglucosamine
Gal	=	Galactose
Fuc	=	Fucose
GalNAc	=	*N*-acetylgalactosamine
GalNAc-transferase	=	UDP- GalNAc:polypeptide *N*-acetylgalactosamine transferase

REFERENCES

[1] Sharon, H. & Lis, H. (**1989**) *Science, 246*, 227-234.

[2] Lasky, L.A. (**1992**) *Science, 258*, 964-969.

[3] Sastry, K. & Ezekowitz, R.A. (**1993**) *Curr. Opin. Immunol., 5*, 59-66.

[4] Barondes, S.H., Cooper, D.N.W., Gitt, M.A. & Leffler, H. (**1994**) *J. Biol. Chem., 269*, 20807-20810.

[5] Hoppe, H.-J. & Reid, K.B.M. (**1994**) *Prot. Sci., 3*, 1143-1158.

[6] Rosen, S.D. & Bertuzzi, C.R. (**1994**) *Curr. Opin. Cell Biol., 6*, 663-673.

[7] Sharon, N. & Lis, H. (**1995**) *Essays Biochem., 30*, 59-75.

[8] Lis, H. & Sharon, N. (**1998**) *Chem. Rev., 98*, 637-674.

[9] Karlsson, K.-A. (**1998**) *Mol. Microbiol., 29*, 1-11

[10] Poveda, A. & Jiménez-Barbero, J. (**1998**) *Chem. Soc. Rev., 27*, 133-143.

[11] Jiménez-Barbero, J., Asensio, J.L., Cañada, F.J. & Poveda, A. (**1999**) *Curr. Opin. Struct. Biol., 9*, 549-555.

[12] Bush, C.A., Martin-Pastor, M. & Imberty, A. (**1999**) *Annu. Rev. Biophys. Biomol. Struct., 28*, 269-293.

[13] Alonso-Plaza, J.M., Canales, M.A., Jiménez, A., Roldán, J.L., García-Herrero, A., Iturrino, L., Asensio, J.L., Cañada, F.J., Romero, A., Siebert, H.-C., André, S., Solís, D., Gabius, H.-J. & Jiménez-Barbero, J. (**2001**) *Biochim. Biophys. Acta, 1568*, 225-236.

[14] Xu, D., Xu, Y. & Uberbacher, E.C. (**2000**) *Curr. Prot. Pept. Sci., 1*, 1-21.

[15] Diller, D.J. & Merz, K.M., Jr. (**2001**) *Prot. Struct Funct. Genet., 43*, 113-124.

[16] Chen, Y.Z. & Zhi, D.G. (**2001**) *Prot. Struct. Funct. Genet., 43*, 217-226.

[17] Wang, Y., Anderson, J.B., Chen, J., Geer, L.Y., He, S., ffurwitz, D.I., Liebert, C.A., Madej, T., Marchler, G.H., Marchler-Baur, A., Panchenko, A.R., Shoemaker, B.A., Song, J.S., Thiessen, P.A., Yamashita, R.A. & Bryant, S.H. (**2001**) *Nuleic Acids Res., 30*, 249-252.

[18] Pieper, U., Eswar, N., Stuart, A.C., Ilyin, V.A. & Sali, A. (**2002**) *Nucleic Acids Res., 30*, 255-259.

[19] Kilosanidze, G.T., Kutsenko, A.S., Espinova, N.G. & Tumanyan, V.G. (**2002**) *FEBS Lett., 510*, 13-16.

[20] Wells, J.A. (**1991**) *Meth. Enzymol., 202*, 390-411.

[21] Jin, L. & Wells, J.A. (**1994**) *Prot. Sci., 3*, 2351-2357.

[22] Morrison, K.L. & Weiss, G.A. (**2001**) *Curr. Opin. Chem. Biol., 5*, 302-307.

[23] Thorn, K.S. & Bogan, A.A. (**2001**) *Bioinformatics, 17*, 284-285.

[24] Gregoret, L.M. & Sauer, R.T. (**1993**) *Proc. Natl. Acad. Sci. USA, 90*, 4246-4250.

[25] Weiss, G.A., Watanabe, C.K., Zhong, A., Goddard, A. & Sidhu, S.S. (**2000**) *Proc. Natl. Acad. Sci. USA, 97*, 8950-8954.

[26] Padlan, A.E. (**1990**) *Proteins: Struct. Funct. Genet., 7*, 112-124.

[27] Cygler, M. (**1994**) *Res. Immunol., 145*, 36-40.

[28] Young,N.M. & Oomen, R.P. (**1992**) *J. Mol. Biol.*, 228, 924-934.

[29] Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (**1998**) *Biochim. Biophys. Acta, 1383*, 9-36.

[30] Delbaere, L.T.J. & Prasad, L. (**1999**) In *Carbohydrates. Structures, Syntheses and Dynamics* (Finch, P. ed.), pp. 319-329. Kluwer Academic Publishers, London.

[31] Vijayan, M. & Chandra, N. (**1999**) *Curr. Opin. Struct. Biol., 9*, 707-714.

[32] Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J.E., Boraston, A., Hazlewood, G.P. & Gilbert, H.J. (**1998**) *Biochem. J., 331*, 775-781.

[33] Tomme, P., Warren, R.A.J. & Gilkes, N.R. (**1995**) *Adv. Microb. Physiol., 37*, 1-81.

[34] Boraston, A.B., Chiu, P., Warren, A.J. & Kilburn, D.G. (**2000**) *Biochemistry, 38*, 11129-11136.

[35] Notenboom, V., Boraston, A.B., Chiu, P., Freelove, A.C.J., Kilburn, D.G. & Rose, D.R. (**2001**) *J. Mol. Biol., 314*, 797-806.

[36] Percivale, P., Bertoglio, S., Meszaros, P., Canavese, G., Cafiero, F., Gipponi, M., Campora, E., Gasco, M. & Badellino, F. (**1996**) *J. Clin. Oncol., 14*, 1599-1603.

[37] Kjeldson, T., Clausen, H., Hirohashi, S., Ogawa, T., Iijima, H. & Hakomori, S. (**1988**) *Cancer Res., 48*, 2214-2220.

[38] Reddish, M.A., Jackson, L., Koganty, R.R., Qiu, D., Hong, W. & Longenecker, B.M. (**1997**) *Glyconoj. J., 14*, 549-560.

[39] Brady, R.L., Edwards, D.J., Hubbard, R.E., Jiang, J.-S., Lange, G., Roberts, S.M., Todd, R.J., Adair, J.R., Emtage, J.S., King, D.J., & Low, D.C. (**1992**) *J. Mol. Biol., 227*, 253-264.

[40] Xiang, J., Srivamadan, M., Rajala, R. & Jia, Z. (**2000**) *Protein Eng., 13*, 339-344.

[41] Xiang, J., Sha, Y., Prasad, L. & Delbaere, L.T.J. (**1996**) *Protein Eng., 9*, 539-543.

[42] Davies, D.R., Padlan, E.A. & Sheriff, S. (**1990**) *Annu. Rev. Biochem., 59*, 439-473.

[43] Sadler, J.E. (**1984**) In *Biology of Carbohydrates* (Ginsburg, V. and Robbins, P., eds.) Vol. *2*, pp. 199-213, John Wiley & Sons, New York.

[44] Schechter, I. & Berger, A. (**1967**) *Biochem. Biophys. Res. Commun., 27*, 157-162.

[45] Elhammer, Å.P., Poorman, R.A., Brown, E., Maggiora, L.L., Hoogerheide, J.G. & Kézdy, F.J. (**1993**) *J. Biol. Chem., 268*, 10029-10038.

[46] NBRF Protein Database (**1993**) *Atlas of Protein and Genomic Sequences*. The National Biomedical Research Foundation, Compact Disk Data Storage, Release 35, January 1993, Washington, DC 20007.

[47] Poorman, R.A., Tomasselli, A.G., Heinrikson, R.L. & Kézdy, F.J. (**1991**) *J. Biol. Chem., 266*, 14554-14561.

[48] Nakashima, H., Nishikawa, K. & Ooi, T. (**1986**) *J. Biochem. (Tokyo), 99*, 153-162.

[49] Chou, K.-C., Zhang, C.-T., Kézdy, F.J. & Poorman, R.A. (**1995**) *Proteins Struct. Funct. Genet., 21*, 118-126.

[50] Chou, K.-C. (**1995**) *Prot. Sci., 4*, 1365-1383.

[51] Cai, Y.-D. & Chou, K.-C. (**1996**) *Anal. Biochem.*, 243, 284-285.

[52] Cai, Y.-D., Liu, X.-J., Xu, X.-B. & Chou, K.-C. (**2002**) *Peptides, 23*, 205-208.

[53] Lemieux, R.U. (**1978**) *Chem. Soc. Rev., 7*, 423-452.

[54] Nikrad, P.V., Beiberbeck, H. & Lemieux, R.U. (**1992**) *Can. J. Chem., 70*, 241-253.

[55] Delbaere, L.T.J., Vandonselaar, M., Prasad, L., Quail, J.W., Wilson, K.S. & Dauter, Z. (**1993**) *J. Mol. Biol., 230*, 950-965.

[56] Good, A.H., Yau, O., Lamontange, L.R. & Oriol, R. (**1992**) *Vox. Sang., 62*, 180-189.

[57] Spohr, U., Hindsgaul, O. & Lemieux, R.U. (**1985**) *Can. J. Chem., 63*, 2644-2652.

[58] Alibés, R. & Bundle, D.R. (**1998**) *J. Org. Chem., 63*, 6288-6301.

[59] Bundle., D.R., Alibés, R., Nilar, S., otter, A., Warwas, M. & Zhang, P. (**1998**) *J. Am. Chem. Soc., 120*, 5317-5318.

[60] Navarre, N., Amiot, N., van Oijen, A., Imberty, A., Poveda, A., Jiménez-Barbero, J., Cooper, A., Nutley, M.A. & Boons, G.-J. (**1999**) *Chem. Eur. J., 5*, 2281-2294.

[61] Wacowich-Sgarbi, S.A. & Bundle, D.R. (**1999**) *J. Org. Chem., 64*, 9080-9089.

[62] Lemieux, R.U., Bock, K., Delbaere, L.T.J., Koto, S. & Rao, V.S. (**1980**) *Can. J. Chem., 58*, 631-653.

[63] Lemieux, R.U., Spohr, U., Bach, M., Cameron, D.R., Frandsen, T.P., Stoffer, B.B., Svensson, B. & Palcic, M.M. (**1996**) *Can. J. Chem., 74*, 319-335.

[64] Weill, C.E., Burch, R.J. & Van Dyk, J.W. (**1954**) Cereal *Chem., 31*, 150-158.

[65] Pazur, J.H. & Ando, T. (**1960**) *J. Biol. Chem., 235*, 297-302.

[66] Aleshin, A.E., Golubev, A., Firsov, L.M & Honzatko, R.B. (**1992**) *J. Biol. Chem., 267*, 19291-19298.

[67] Alsehin, A.E., Hoffman, C., Firsov, L.M. & Honzatko, R.B. (**1994**) *J. Mol. Biol., 238*, 575-591.

[68] Harris, E.M.S., Aleshin, A.E., Firsov, L.M. & Honzatko, R.B. (**1993**) *Biochemistry, 32*, 1618-1626.

[69] Aleshin, A.E., Firsov, L.M. & Honzatko, R.B. (**1994**) *J. Biol. Chem., 269*, 15631-15639.

[70] Stoffer, B., Aleshin, A.E., Firsov, L.M., Svensson, B. & Honzatko, R.B. (**1995**) *FEBS Lett., 358*, 57-61.

[71] Quiocho, F.A. (**1986**) *Annu. Rev. Biochem., 55*, 287-315.

[72] Quiocho, F.A., Wilson, D.K. & Vyas, N.K. (**1989**) *Nature, 340*, 404-407.

[73] Lemieux, R.U. (**1996**) *Acc. Chem. Res., 29*, 373-380.

[74] Rivera, M. & Nicorta, M.B. (**1982**) *Am. Rev. Respir. Dis., 126*, 833-836.

[75] Todd, T.R.J., Franklin, A., Mankinen-Irvin, P., Gurman, G. & Irvin, R.T. (**1989**) *Am. Rev. Respir. Dis., 140*, 1585-1589.

[76] Irvin, R.T., Doig, P.C., Sastry, P.A., Heller, B. & Paranchych, W. (**1990**) *Microbial. Ecol. Hlth. Dis., 3*, 39-47.

[77] Irvin, R.T. (**1993**) In *Pseudomonas aeruginosa* as an opportunistic pathogen (Campa, M., Bendinelli, M. & Friedman, H., eds.), pp. 19-42, Plenum Press, New York.

[78] Lee, K.K., Sheth, H.B., Wong, W.Y., Sherbourne, R., Paranchych, W., Hodges, R.S., Lingwood, C.A., Krivan, H. & Irvin, R.T. (**1994**) *Mol. Microbiol., 11*, 705-713.

[79] Yu, L., Leek K.K., Hodges, R.S., Paranchych, W. & Irvin, R.T. (**1994**) *Infect. Immun., 62*, 5213-5219.

[80] Sheth, H.B., Lee, K.K., Wong, W.Y., Srivastava, G., Hindsgaul, O., Hodges, R.S., Paranchych, W. & Irvin, R.T. (**1994**) *Mol. Microbiol., 11*, 715-723.

[81] Schweizer, F., Jiao, H., Hindsgaul, O., Wong, W.Y. & Irvin, R.T. (**1998**) *Can. J. Microbiol., 44*, 307-311.

[82] Jiao, H. & Hindsgaul, O. (**1999**) *J. Carbohydr. Chem., 18*, 499-513.

[83] Hamelryk, T., Loris, R., Bouckaert, J. & Wyns, L. (**1998**) *Trends Glycosci. Glycotech., 10*, 349-360.

[84] Bouckeart, J. Hamelryk, T., Wyns, L. & Loris, R. (**1999**) *Curr. Opin. Struct. Biol., 9*, 572-577.

[85] Rini, J.M. & Lobsanov, Y.D. (**1999**) *Curr. Opin. Struct. Biol., 9*, 578-584.

[86] Campbell, P.A., Wong, W.Y., Houston, M., Jr., Schweizer, F., Cachia, P.J., Irvin, R.T., Hindsgaul, O., Hodges, R.S. & Sykes, B.D. (**1997**) *J. Mol. Biol., 267*, 382-402.

[87] Hazes, B., Sastry, P.A., Hayakawa, K., Read, R.J. & Irvin, R.T. (**2000**) *J. Mol. Biol., 299*, 1005-1017.

[88] Keizer, D.W., Slupsky, C.M., Kalisiak, M., Campbell, A.P., Crump, M.P., Sastry, P.A., Hazes, B., Irvin, R.T. & Sykes, B.D. (**2001**) *J. Biol. Chem., 276*, 21486-24193.

[89] Daniels, G. (**1995**) Human Blood Groups. Blackwell Science, Oxford.

[90] Matsumoto, I. & Osawa, T. (**1969**) *Biochim. Biophys. Acta, 194*, 180-189.

[91] Watkins, W.M. (**1972**) In *Glycoproteins. Their Composition, Structure and Function* (Gottschalk, A., ed.), pp. 830-891, Elsevier, Amsterdam.

[92] Pereira, M.E.A., Kisailus, E.C., Gruezo, F. & Kabat, E.A. (**1978**) *Arch. Biochem Biophys., 185*, 108-115.

[93] Hindsgaul, O., Norberg, T., Le Pendu, J. & Lemieux, R.U. (**1982**) *Carbohydr. Res., 109*, 109-142.

[94] Le Pendu, J., Lemieux, R.U., Lambert, F., Dalix, A.-M. & Oriol, R. (**1982**) *Am. J. Hum. Genet., 34*, 402-415.

[95] Le Pendu, J., Cartron, J.P., Lemieux, R.U. & Oriol, R. (**1985**) *Am. J. Hum. Genet., 37*, 749-760.

[96] Du, M., Spohr, U. & Lemieux, R.U. (**1994**) *Glycoconj. J., 11*, 443-461.

[97] Hindsgaul, O., Khare, D.P., Bach, M. & Lemieux, R.U. (**1985**) *Can. J. Chem., 63*, 2653-2658.

[98] Gohier, A., Espinosa, J.F., Jimenez-Barbero, J., Carrupt, P., Perez, S. & Imberty, A. (**1996**) *J. Mol. Graph., 14*, 322-327.

[99] Konami, Y., Yamamoto, K. & Osawa, T. (**1991**) *J. Biochem., 109*, 650-658.

[100] Audette, G.F., Vandonselaar, M. & Delbaere, L.T.J. (**2000**) *J. Mol. Biol., 304*, 423-433.

[101] Audette, G.F., Olson, D.J.H., Ross, A.R.S, Quail, J.W. & Delbaere, L.T.J. (**2002**) *Can. J. Chem., 80*, 1010-1021.

[102] Fukui, S., Feizi, T., Galustian, C., Lawson, A.M. & Chai, W. (**2002**) *Nature Biotechnol., 20*, 1011-1017.

[103] Evans, S.V. (**1993**) *J. Mol. Graph., 11*, 134-138.