Examination of the structural basis for O(H) blood group specificity by *Ulex europaeus* Lectin I

Gerald F. Audette, Douglas J.H. Olson, Andrew R.S. Ross, J. Wilson Quail, and Louis T.J. Delbaere

Abstract: The structural basis for carbohydrate specificity of the first lectin from *Ulex europaeus* (UE-I) is reported. UE-I is a dimeric metalloglycoprotein that binds the H-type 2 human blood group determinant (α-L-Fucα(1→2)-β-D-Galβ(1→4)-β-D-GlcNAcα-), the blood group determinant present on the surface of O-type erythrocytes. The structural characteristics of UE-I involved in carbohydrate recognition have been examined using mass spectrometry (MS) and X-ray diffraction analysis. MS analysis allowed for discrimination between the different primary structures reported for UE-I. To examine the binding of the H-type 2 blood group determinant by UE-I, the methyl glycosides of the fucose monosaccharide (α-L-Fuc-OMe), known to exhibit primary binding specificity, and the H-type 2 trisaccharide (H-type 2-OMe) were, in two separate experiments, co-crystallized into the binding site of UE-I. The UE-I:α-L-Fuc-OMe complex crystallizes in the monoclinic space group $P2_1$, with unit cell dimensions a = 71.81, b = 69.00, and c = 119.02 Å, and β = 106.76°. Two UE-I dimers are observed to be present within the asymmetric unit, and the model has been refined to a R-value and R_{Free} of 0.202 and 0.289, respectively, to 2.3 Å resolution. The preliminary model of the UE-I:H-type 2-OMe complex has been refined at 3.0 Å resolution. The UE-I:H-type 2-OMe complex crystallizes in the orthorhombic space group $C222_1$, with unit cell dimensions a = 88.80, b = 164.75, and c = 77.42 Å, and a single UE-I dimer is present within the asymmetric unit. The carbohydrate recognition domain of UE-I has been identified to be comprised of residues Glu44, Thr86, Asp87, Arg102, Ala103, Gly104, Gly105, Tyr106, Ile129, Val133, Asn134, Trp136, Tyr219, and Arg222. Several critical protein–carbohydrate interactions have been identified, including the role of the hydrophobic interaction between the Thr86 side chain and C-5-CH_3 of the α-L-Fuc-OMe. The role of these interactions in carbohydrate recognition–binding by UE-I, as well as differences between the observed and previously modeled complexes, are discussed.

Key words: *Ulex europaeus* lectin I, H-type 2 human blood group determinant, protein–carbohydrate interactions, X-ray crystallography, chemical mapping.

Résumé : On a établi les bases structurales de la spécificité de l'hydrate de carbone de la première lectine provenant d'Ulex europaeus (EU-I). EU-I est une métalloglycoprotéine dimère qui se lie au déterminant du groupe sanguin humain H de type-2 (α-L-Fucα(1→2)-β-D-Galβ(1→4)-β-D-GlcNAcα-), le déterminant du groupe sanguin présent à la surface des érythrocytes de type-O. Les caractéristiques structurales de UE-I impliquées dans la reconnaissance du carbohydrate ont été examinées en faisant appel à la spectrométrie de masse (SM) et à la diffraction des rayons X. L'analyse par SM a permis de distinguer entre les différentes structures primaires proposées pour UE-I. Afin d'étudier le site de fixation du déterminant du groupe sanguin humain H de type-2 par UE-I, on a procédé à deux expériences séparées au cours desquelles on a procédé à la cocristallisation dans le site de fixation de UE-I des glycosides de méthyle d'une part du monosaccharide du fucose (α-L-Fuc-OMe), reconnu pour présenter une spécificité pour le site de fixation primaire, et d'autre part du trisaccharide H de type-2 (H de type-2-OMe). Le complexe UE-I:α-L-Fuc-OMe cristallise dans le groupe d'espace $P2_1$ avec a = 71,81, b = 69,00 et c = 119,02 Å et β = 106,76°. On a observé la présence de deux dimères de UE-I dans l'unité asymétrique et on a affiné le modèle jusqu'à des valeurs de R et R_{libre} de 0,0202 et 0,289 respectivement à une résolution de 2,3 Å. Le modèle préliminaire du complexe UE-I:H-type-2-OMe a été affiné jusqu'à une résolution de 3,0 Å. Ce complexe cristallise dans le groupe orthorhombique $C222_1$, avec a =

Received 10 September 2001. Published on the NRC Research Press Web site at http://canjchem.nrc.ca on 13 August 2002.

Dedicated to Professor R.U. Lemieux in memory of his friendship, his many valuable discussions, and scientific contributions.

G.F. Audette,[1] A.R.S. Ross,[2] and L.T.J. Delbaere.[3] Department of Biochemistry, University of Saskatchewan, 107 Wiggins Road, Saskatoon, SK S7N 5E5, Canada.
D.J.H. Olson. Plant Biotechnology Institute, National Research Council Canada, 110 Gymnasium Place, Saskatoon, SK S7N 0W9, Canada.
J.W. Quail. Department of Chemistry, University of Saskatchewan, 110 Science Place, Saskatoon, SK S7N 5C9, Canada.

[1]Current address: Department of Medical Microbiology & Immunology, 1–15 Medical Sciences Building, University of Alberta, Edmonton, AB T6G 2H7, Canada.
[2]Current address: Plant Biotechnology Institute, National Research Council Canada, 110 Gymnasium Place, Saskatoon, SK S7N 0W9, Canada.
[3]Corresponding author (e-mail: louis.delbaere@usask.ca).

 DOI: 10.1139/V02-134

88,80, *b* = 164,75 et *c* = 77,42 Å et on n'observe la présence que d'un seul dimère UE-I dans l'unité asymétrique. On a établi que le domaine de reconnaissance du carbohydrate du UE-I comprend les résidus Glu44, Thr86, Asp87, Arg102, Ala103, Gly104, Gly105, Tyr106, Ile129, Val133, Asn134, Trp136, Tyr219 et Arg222. On a identifié plusieurs interactions protéine–carbohydrate critiques, y compris le rôle de l'interaction hydrophobe entre la chaîne latérale du Thr86 et le C-5-CH_3 du α-L-Fuc-OMe. On discute du rôle de ces interactions dans la reconnaissance/fixation du carbohydrate par UE-I ainsi que des différences entre les complexes modélisés antérieurement et les complexes observés.

Mots clés : lectine I d'*Ulex europaeus*, déterminant du groupe sanguin humain H de type-2, interactions protéine–carbohydrate, diffraction des rayons X, carte chimique.

[Traduit par la Rédaction]

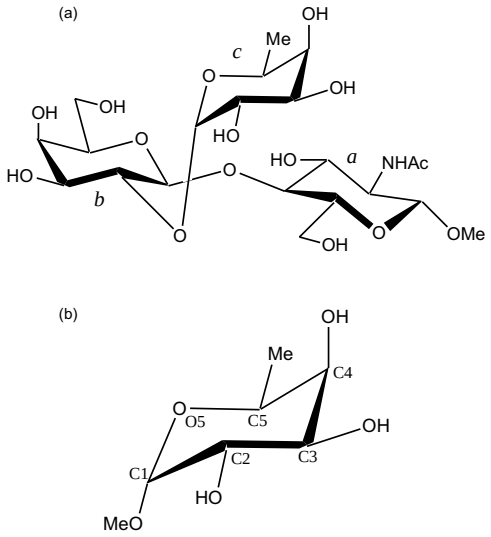
Introduction

Protein–carbohydrate interactions are involved in a wide range of important biological processes. These processes include, but are not limited to, cellular infection by invading microorganisms, the subsequent cellular defense and immune response to the invading pathogen, fertilization, leukocyte trafficking and infiltration, and cancer metastasis (1–3). The lectins are a class of proteins that exhibit a polyvalent affinity for carbohydrates; this polyvalency arises from the association of monomeric subunits into mainly dimeric and tetrameric molecules (3). A wide variety of different lectins are present in living cells and tissues. Their carbohydrate specificity is akin to antigenic recognition by antibodies (4) and enzymatic substrate specificity (5), yet lectins are neither products of the immune system nor do they exhibit enzyme-like catalytic properties. The role of lectins appears limited to specific cross-linking interactions, which allow lectins to serve as recognition molecules for various biological events (1, 2).

Extracts from the seeds of *Ulex europaeus* (common gorse), a leguminous shrub native of Western Europe, have been shown to agglutinate type O erythrocytes (6); the first lectin present in these extracts, *U. europaeus* lectin I (UE-I), was the agglutinating agent, with a carbohydrate specificity for α-L-fucose (7). The classical ABO blood-group in humans was initially identified by Landsteiner (8) at the turn of the 20th century. Landsteiner identified three classes of blood, based upon the agglutination of erythrocytes from one individual when mixed with the serum of another. He labeled these blood-groups in terms of antigenic substances as groups A and B, based on this agglutinating activity. The third group, O, was thus labeled for its lack of agglutinating activity (9). The ABO blood group determinants were subsequently identified as being carbohydrate in nature (10). The O blood group determinant, identified as the H-type 2 trisaccharide (α-L-Fucα(1→2)-β-D-Galβ(1→4)-β-D-GlcNAcα-) (Fig. 1*a*), was also the precursor of the A and B blood group determinants, which are tetrasaccharides (10). It is this trisaccharide that is recognized by UE-I. Though classified as an α-L-fucose-specific legume lectin, UE-I binds α-L-fucose with an affinity that is approximately 100-fold less than the binding of the complete H-type 2 determinant (11).

In the absence of crystallographic information, a method has been developed to probe the interactions between protein and carbohydrate. This method, known as chemical mapping (12), was developed based on the interactions between the Lewis b blood group determinant and the fourth lectin from *Griffonia simplicifolia* (GS-IV) observed in the crystal structure (13). The chemical mapping method employs the generation of monodeoxy and mono-*O*-methyl derivatives of the carbohydrate to generate a chemical map of the carbohydrate epitope, which is recognized by the protein. Application of this method to the binding of the H-type 2 blood group determinant by UE-I has been reported (11, 14–16). These inhibition studies into the binding of the H-type 2 determinant by UE-I have revealed several key interactions between protein and carbohydrate (Table 1) (16), and allowed for the generation of the chemical map for the epitope of the H-type 2 trisaccharide recognized by UE-I (Fig. 1*a*).

Fig. 1. The human H-type 2 blood group determinant. (*a*) The β-methyl glycoside of the H-type 2 trisaccharide (α-L-Fucα(1→2)-β-D-Galβ(1→4)-β-D-GlcNAcα-) (H-type 2-OMe). Structural units *a*, *b*, and *c* are β-D-GlcNAc-OMe, β-D-Gal, and α-L-Fuc, respectively. (*b*) The methyl glycoside of α-L-fucose (α-L-Fuc-OMe) shown in a perspective representation with atomic numbering for clarity.



The lectins from leguminous plants have contributed significantly to our understanding of the structural nature of carbohydrate specificity exhibited by proteins (17–20) — the majority of knowledge about lectins and their interactions with carbohydrate is from crystallographic analysis of the Glc/Man, Gal/GalNAc and complex specific legume lectins (20). Very little structural information is available regarding the carbohydrate specificity of the α-L-fucose-specific lectins, though the structure of the native UE-I has been recently reported to 2.2 Å resolution (21). This study revealed

Fig. 2. MALDI-TOF MS spectra of trypsin-digested UE-I. Several peaks were found to correspond to peptide fragments of an in silico digestion of UE-I, and are indicated in the spectra and in Table 2. The horizontal axis represents the observed *m*/*z* ratios for the UE-I peptide fragments, and the vertical axes represent the percent relative intensity (left) and measured peak height of the most intense or base peak (right). The peak at *m*/*z* 1889.01 confirmed the presence of Thr in the X86-Asp87 *cis*-peptide, and the peak at *m*/*z* 2097.13 allowed for discrimination between the Konami et al. (22) and Audette et al. (21) primary structures of the UE-I peptide Glu204–Arg222. The peak at *m*/*z* 842.51 is due to the autolysis of trypsin.

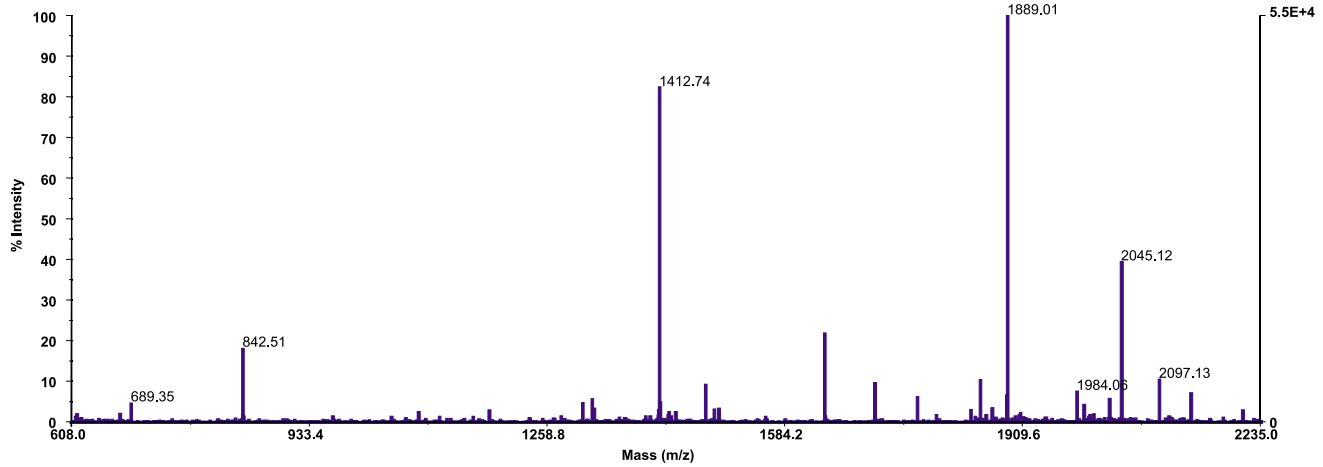


Table 1. Chemical mapping of the epitopes of H-type 2-OMe recognized by UE-I.

	Relative potencies[b]	
Position altered[a]	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*)[14]	3	—

Note: For source, see ref. 16.
[a]Structural unit designations are as described in Fig. 1*a*.
[b]By radioimmunoassay, setting the potency of H-type 2-OMe at 100 (11).

several differences between other legume lectins, and included the modeling of the β-methyl glycoside of the H-type 2 trisaccharide (H-type 2-OMe) within the carbohydrate binding site to predict the interactions between protein and carbohydrate in correlation with the available chemical mapping data (21).

The present study examines the structural nature of the recognition of the H-type 2 blood group determinant by UE-I through mass spectrometry and X-ray diffraction analysis of the complex of UE-I and the methyl glycoside of α-L-fucose (α-L-Fuc-OMe; Fig. 1*b*) at 2.3 Å resolution. Also, a preliminary analysis of a low resolution structure of the UE-I:H-type 2-OMe complex allows for the identification of the interactions between protein and carbohydrate at the periphery of the binding site, which confer extended carbohydrate specificity to the lectin. The interactions observed between protein and carbohydrate are in correlation with the observed chemical mapping data; however, these observed interactions differ from those predicted in the previous modeling studies. The observed interactions, carbohydrate specificity of UE-I, and differences from the modeling studies are discussed.

Results and discussion

Mass spectrometry

Electrospray ionization mass spectrometric (ESI-MS) analysis of UE-I indicates that the molecular weights of the two monomeric UE-I subunits are 30 049.0 and 31 243.0 Da (data not shown). These values correlate well with observations made during SDS-PAGE analysis of UE-I, in which the dimer separates as two distinct bands of approximately 30 kDa molecular weight (data not shown). It was noted, however, that the molecular weight of the UE-I monomer, based on the primary structure observed in the electron density is 26 606 Da (21). This apparent difference in molecular weight between UE-I monomers may be attributed to differential glycosylation states of each subunit.

Examination of the crystal structure of the native UE-I (21) revealed several differences between the reported primary structure of UE-I (22) and that observed in the electron density. To verify the crystallographically observed amino acid sequence of UE-I (21) vs. that of the published primary structure (22), matrix-assisted laser desorption/ionization-time of flight mass spectrometric (MALDI-TOF MS) analysis was performed on a tryptic digest of UE-I. Both the upper and lower bands of SDS-PAGE separated UE-I were subjected to identical digestion and MALDI-TOF MS protocols; a representative mass spectrum is shown in Fig. 2. The *m*/*z* ratios for each peptide peak were compared with those predicted for a theoretical (in silico) tryptic digest. Peaks at *m*/*z* 842.51, 1045.56, 2211.10, and 2383.93 can be attributed to autolysis of trypsin, and were thus excluded from analysis.

Several peaks were found to correspond to those predicted from the theoretical digest (Table 2). Of particular interest was the peak observed at *m*/*z* 2097.13. This peak corresponds to the residues between Glu204 and Arg222, and allowed for discrimination between the differing primary

Table 2. MALDI-TOF analysis of trypsinized UE-I.

$[M + H]^+$ (*m/z*)		Residue range
Predicted	Observed	
689.36	689.35	Ser152–Arg157
1412.71	1412.74	Val37–Arg50
1888.98	1889.01	Ala84–Arg101
2045.08	2045.12	Ala84–Arg102
1984.00	1984.06	Tyr162–Lys179
2097.07	2097.13	Glu204–Arg222
2140.10	2140.17	Arg161–Lys179

sequences. Within this region, a Gly → Ser change was required to account for the electron density observed at residue 217 (21). The corresponding $[M + H]^+$ ion for this region (with the Gly → Ser change) has a predicted *m*/*z* of 2097.07, which matches closely the observed MS peak. The peptide fragment with a Gly at position 217 would have an *m*/*z* of 2067.03; no corresponding peak was observed in the mass spectrum.

No other peaks were observed that would correspond to peptides containing expected differences in the UE-I sequence. It is possible that glycosylation of the protein may prevent correlation between observed and predicted *m*/*z* values for glycosylated peptides. Also, non-specific cleavage by trypsin may result in peptide fragments that will not correlate with an in silico trypsin digest of the target protein. Nevertheless, the available MS data suggests that the primary structure determined via X-ray analysis (21) is indeed correct. This, together with the fact that Asn116 was not observed in the structure of native UE-I, indicates that continuous numbering of UE-I residues can be applied throughout the sequence (Ser1–Ser242). For native UE-I (21), however, discontinuous residue numbering (Ser1–Cys115 and Asp117–Ser243) was retained to coincide with the published amino acid sequence of Konami and co-workers (22).

Quality of the crystallographic models of the UE-I–carbohydrate complexes

The final model of the UE-I:α-L-Fuc-OMe complex consists of two UE-I dimers within the asymmetric unit, related by non-crystallographic symmetry (NCS). The structure was refined to 2.3 Å resolution (diffraction data summarized in Table 3). The final model consists of 240 amino acid residues in three of the four UE-I monomers; the fourth UE-I monomer contains 239 amino acid residues. No electron density was observed for the final two C-terminal residues in UE-I subunits A, B, and C, nor for the final three C-terminal residues for UE-I subunit D. Four calcium and four manganese ions are present in the model (one Ca^{2+} and one Mn^{2+} per subunit). There are four bound α-L-Fuc-OMe molecules in the final model, with one α-L-Fuc-OMe molecule located in the predicted carbohydrate recognition domain (CRD) of each UE-I monomer. Electron density was observed at Asn23 and Asn111 of all UE-I monomers, which correspond to *N*-linked oligosaccharide at these positions; however, the observed density was not sufficiently resolved to allow for modeling of the *N*-linked carbohydrate. There are 266 solvent (water) molecules present in the final model, and a single MPD molecule was observed adjacent to Cys60 of the third UE-I monomer.

The stereochemical quality of the final model was analyzed using the program PROCHECK (23). A total of 80.6% of the amino acid residues are within the most favored regions of the Ramachandran plot (24); 18.7% of the residues are located in the additional allowed regions, and six residues (0.7%) are located in generously allowed regions. The mean coordinate error from Luzzatti plot analysis (25) is 0.15 Å. The final refinement statistics are summarized in Table 3.

The model of the UE-I:H-type 2-OMe complex consists of a single UE-I dimer within the asymmetric unit refined to 3.0 Å resolution (Table 3). The model consists of 240 amino acid residues in both UE-I subunits, two calcium and two manganese ions (one Ca^{2+} and Mn^{2+} per subunit), and a single bound H-type 2-OMe is present in the final model, located in the predicted CRD of the UE-I subunit A. Electron density was observed at Asn23 and Asn111 of both UE-I monomers, which correspond to *N*-linked oligosaccharide at these positions; however, the observed density was not sufficiently resolved to allow for modeling of the *N*-linked carbohydrate. Because of the resolution of the data, no solvent (water) molecules were included in the model. The stereochemical quality of the model was analyzed using the program PROCHECK (23). A total of 68.4% of the amino acid residues are within the most favored regions of the Ramachandran plot (24); 27.8% of the residues are located in the additional allowed regions, and the remaining residues (3.8%) are located in generously allowed regions. The mean coordinate error from Luzzatti plot analysis (25) is 0.20 Å; refinement statistics are summarized in Table 3.

Overall structure of UE-I

The tertiary and quaternary structure of UE-I have been discussed previously (21). Both the UE-I:α-L-Fuc-OMe and UE-I:H-type 2-OMe complexes of UE-I exhibit a similar overall structure to that of the native protein (Fig. 3). A comparison between the UE-I molecules indicates only slight differences; the average root mean square differences (rmsd) between C^α atoms of the native UE-I and the first and second dimers of the UE-I:α-L-Fuc-OMe complex are 0.63 and 0.68 Å, respectively. The average rmsd between the native UE-I and the UE-I:H-type 2-OMe complex is 0.96 Å, and between the UE-I:H-type 2-OMe complex and the first and second dimers of the UE-I:α-L-Fuc-OMe complex are 0.93 and 0.87 Å, respectively.

Examination of the monomeric subunits of UE-I in both complexes also indicate that differences between subunits are minimal. In the UE-I:α-L-Fuc-OMe complex, the overall average rmsd between C^α atoms is 0.46 Å. The B-factors are also similar between the UE-I subunits, with average B-factors (main-chain/side-chain) of 20.4 $Å^2$/20.5 $Å^2$, 22.6 $Å^2$/22.2 $Å^2$, 22.1 $Å^2$/22.3$Å^2$, and 19.4 $Å^2$/19.8 $Å^2$ for UE-I subunits A–D, respectively. Side chains of several residues were observed to have little electron density: Lys83 of subunit A, Asn40 of subunit B, Thr240 of subunit C, and Asn39, Lys83, and Asn239 of subunit D. These residues were replaced with alanines during refinement. In the UE-I:H-type 2-OMe complex, the average rmsd between C^α atoms of subunit A and B is 0.20 Å, and the average B-factors

Table 3. Data collection and refinement statistics for the UE-I complexes.

	α-L-Fuc-OMe	H-Type 2-OMe
Diffraction data		
Resolution range (Å)	40–2.30	40–3.00
(Last shell) (Å)	(2.32–2.30)	(3.03–3.00)
No. unique reflections	44 883 (944)	9 793 (244)
Completeness (%)	87.9 (76.8)	89.1 (77.7)
Mean $I/\sigma(I)$	14.93 (2.62)	11.29 (2.47)
R_{Symm}[a]	0.075 (0.457)	0.157 (0.489)
Crystal mosaicity (°)	1.49	1.61
Refinement statistics		
Diffraction agreement		
Resolution range (Å)	40–2.30	30–3.0
R-value[b]	0.202	0.284
R_{Free}[c]	0.289	0.421
No. of unique reflections	28,417	8,376
Quality of structure		
Model parameters		
Total protein non-H atoms	7449	3738
Solvent molecules	266	—
B-factor model	Individual isotropic	Individual isotropic
Mean B-factors (Å^2)[d]		
Main-chain	21.0	19.4
Side-chain	21.2	19.2
Bound carbohydrate	27.8	20.5
Solvent molecules	23.2	—
rmsd from stereochemical ideality		
Bonds (Å)	0.01	0.01
Angles (°)	1.67	1.99
Dihedrals (°)	26.14	26.01
Impropers (°)	1.30	1.55

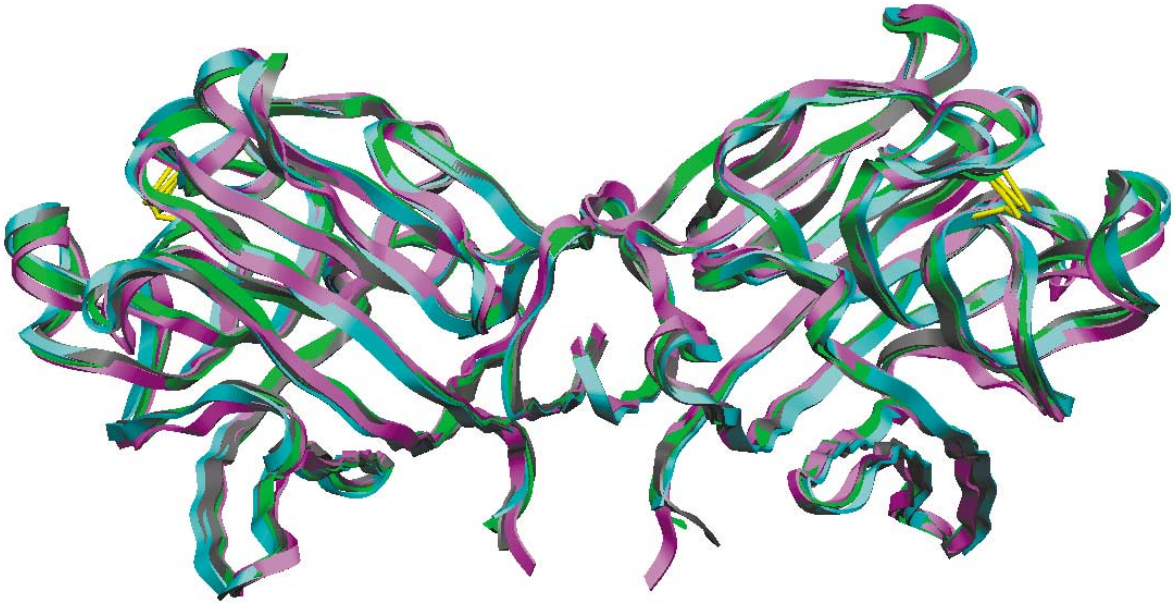
[a] $R_{Symm} = \Sigma_j|I_j - \langle I\rangle|/\Sigma_j I_j$ where j is a set of observations of equivalent reflections and $\langle I\rangle$ is the average intensity of the reflection.
[b] *R*-value = $\Sigma_{hkl}|F_o - F_c|/\Sigma_{hkl}|F_o|$. All data, final refinement cycle.
[c] Randomly chosen test set of 5% of unique reflections (1419 reflections).
[d] Mean B-factors are averaged over both UE-I dimers in the UE-I:α-L-Fuc-OMe complex, and over both monomeric subunits of the UE-I:H-type 2-OMe complex.

are (main-chain/side-chain) 19.1 Å^2/17.9 Å^2 and 19.7 Å^2/20.4 Å^2 for subunits A and B, respectively. In all subunits of both UE-I complexes no electron density was observed for residues 241 and 242, sequenced as Asn and Ser by Konami and co-workers (22).

The carbohydrate binding site of UE-I

Carbohydrate specificity in the legume lectins is based upon the monosaccharide that exhibits the highest binding affinity, even though monosaccharide binding is weaker than the binding of the complete oligosaccharide, which the lectin recognizes (20). Examination of the native UE-I identified the UE-I CRD as a small depression at the surface of the lectin (22), comprised of residues Glu44, Thr86, Asp87, Gly104-Gly105, Ile129, Val133, Asn134, Trp136, Tyr219, and Arg222 (21). A molecule of α-L-Fuc-OMe was observed in the CRD of each UE-I subunit (Fig. 4*a*). Each α-L-Fuc-OMe molecule fit the observed electron density well; the average B-factor for the four α-L-Fuc-OMe molecules is 27.8 Å^2. The α-L-Fuc-OMe molecules make several hydrophilic and hydrophobic interactions with UE-I, which are summarized in Table 4 and illustrated in Fig. 4*b*. Of particular interest is the hydrophobic interaction between C6 of the bound α-L-Fuc-OMe and $C^{\gamma 2}$ of Thr86, with an average interatomic distance of 3.4 Å. This interaction likely plays a significant role in the carbohydrate specificity of UE-I, and will be discussed in greater detail in the following section. A stacking interaction, as is commonly observed in the carbohydrate complexes of legume lectins (20), is provided by Tyr219 to the hydrophobic back face of the bound α-L-Fuc-OMe. Also, the residues involved in carbohydrate binding by UE-I must be updated to include Arg102, due to an observed hydrogen bond between the N^{H2} and O2 of α-L-Fuc-OMe, as well as Ala103 and Tyr106, which make van der Waals contacts. Indeed, it was observed that the Arg102 side chain moved an average of 1.1 Å (measured via the C^{ζ} atom of the side chain) towards the carbohydrate upon binding to make the observed hydrogen bond with the O2 of α-L-Fuc-OMe (Fig. 5). Arg222 also exhibited movement towards the CRD upon carbohydrate binding, with an average movement of 0.7 Å (measured via the C^{ζ} atom of the side chain); no other residues of the CRD exhibited appreciable movement upon carbohydrate binding (Fig. 5). The UE-I CRD is, therefore, comprised of Glu44, Thr86, Asp87, Arg102, Ala103,

Fig. 3. The UE-I dimer. Shown above is the ribbon representation of the superposition of the C^{α} trace of the various UE-I dimers. The first and second UE-I dimer of the UE-I:α-L-Fuc-OMe complex are shown in gray and green, respectively, the native UE-I (21) is in cyan, and the UE-I:H-type 2-OMe complex is in magenta. Figures 3 through 5 were produced using the program SETOR (53).



Gly104, Gly105, Tyr106, Ile129, Val133, Asn134, Trp136, Tyr219, and Arg222.

A single H-type 2-OMe molecule was found in the UE-I:H-type 2-OMe CRD, and is present in the CRD of subunit A; the electron density in the CRD of subunit B was insufficiently resolved to allow for modeling of the bound carbohydrate. The galactose and *N*-acetylglucosamine moieties of the H-type 2-OMe fit the observed electron density, however, the fucose moiety of the trisaccharide was not as well resolved. It is thought that the fucose moiety of the bound H-type 2-OMe will indeed appear as is observed in the UE-I:α-L-Fuc-OMe complex due to similar positions of the C1, O1, and C7 atoms of α-L-Fuc-OMe and the C1*c*, O2*b*, and C2*b* atoms of H-type 2-OMe (structural unit designations as in Fig. 1*a*); this shall be confirmed upon examination of the UE-I:H-type 2-OMe complex at higher resolution.

Carbohydrate specificity of UE-I

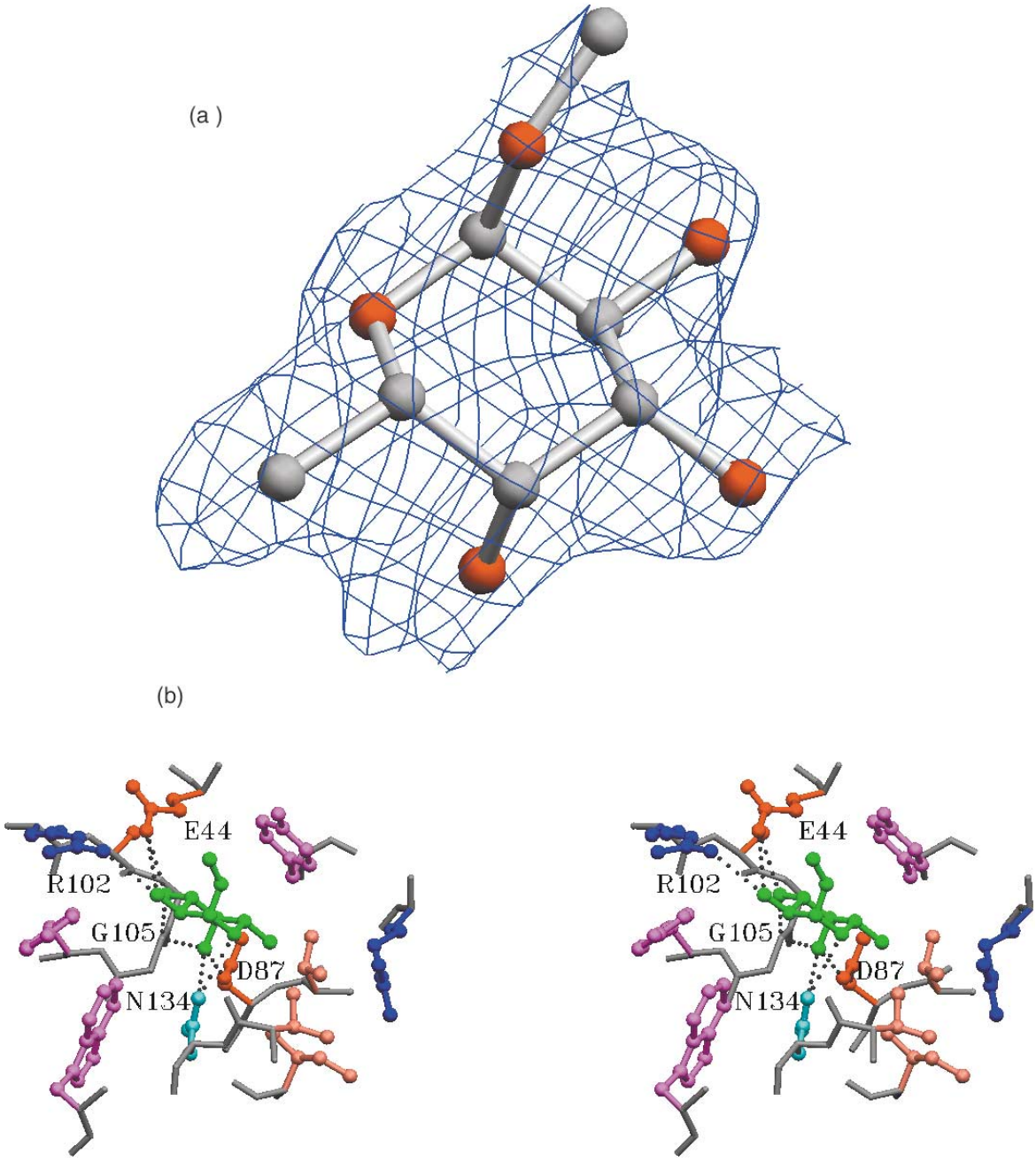
The development of methodologies for complex oligosaccharide synthesis have greatly aided the study of protein–carbohydrate interactions (26); these methodologies are particularly useful when crystallographic data of the complex is unavailable. Indeed, the identification of the effects of slight structural changes to the carbohydrate on the efficacy of binding allows for the prediction of the chemical map of the carbohydrate epitope (12).

The chemical mapping method was applied to the systematic investigation of the interaction between UE-I and the H-type 2 blood group determinant (15), which resulted in the generation of a chemical map of the H-type 2 epitope (Fig. 1) (16). Several conclusions were made upon examination of the results of inhibition studies into the binding of H-type 2-OMe by UE-I (Table 1). In particular, when both deoxygenation and *O*-methylation of hydroxyl groups of the epitope abolish binding, these hydroxyls are identified as being deep within the CRD making critical hydrogen bonds with the protein. When deoxygenation significantly decreases but *O*-methylation exhibits little effect on binding, these hydroxyl groups are interpreted as being closer to the periphery of the binding site. Finally, when neither deoxygenation nor *O*-methylation significantly affect binding, these groups are identified as remaining in the aqueous phase upon complexation.

The observed interactions that occur between UE-I and the bound carbohydrate in the UE-I:α-L-Fuc-OMe complex correlate well with the observed chemical mapping data. In particular, O2, O3, and O4 of α-L-Fuc-OMe make several hydrogen bonds to residues of the UE-I CRD (Fig. 4*b*, Table 4). These interactions occur deep within the CRD. For example, O4 of the α-L-Fuc-OMe was observed to make hydrogen bonds with Asp87 $O^{\delta 2}$ (2.8 Å), Gly105 N (3.0 Å), and Asn134 $N^{\delta 2}$ (3.0 Å). Both Asp87 and Gly105 are located at the bottom of the UE-I CRD, and Asn134 is located close to the bottom of the CRD (Fig. 4*b*). Deoxygenation of the group would result in removal of the hydrogen bonding capacity of that group, while *O*-methylation would prevent interaction through steric interference between the carbohydrate and protein. Disruption of one or all of these interactions would prevent stabilization of the fucose moiety within the CRD as observed in inhibition studies (Table 1) (14).

The presence of a Thr at the X86–Asp87 position is a unique feature of the α-L-Fuc-specific legume lectins; the residue at this position is generally Ala (20). Chemical mapping of the H-type 2-OMe indicated that alteration of the C-5-CH_3 group of the fucose moiety results in significantly reduced binding (Table 1). The authors noted, however, that the effect of replacing this group with a hydrogen atom may have an effect on the conformational preference of the carbohydrate (14). The examination of the UE-I:α-L-Fuc-OMe complex reveals that the C-5-CH_3 group of the α-L-Fuc-OMe makes a significant hydrophobic interaction with $C^{\gamma 2}$ of the Thr86 side-chain as well as several van der Waals

Fig. 4. The bound carbohydrate within the UE-I CRD. (*a*) The bound α-L-Fuc-OMe. Shown is the α-L-Fuc-OMe found in the CRD of the UE-I subunit A of the UE-I:α-L-Fuc-OMe complex in the electron density from an $\boldsymbol{F}_o$-$\boldsymbol{F}_c$ omit electron density map of the final refined structure contoured at the 1σ contour level. The α-L-Fuc-OMe is coloured according to standard element colours (as per SETOR; 53), and the electron density is shown in blue. (*b*) Stereographic representation of the bound α-L-Fuc-OMe within the CRD of the first UE-I subunit in the UE-I:α-L-Fuc-OMe complex. Backbone atoms are in gray, side-chains are in standard colours, and bound carbohydrate is in green. Interactions between the bound α-L-Fuc-OMe and UE-I are listed in Table 4, and hydrogen bonds between the protein and carbohydrate are highlighted.



contacts with hydrophobic residues of the UE-I CRD (Table 4; Fig. 4*b*). These hydrophobic interactions provide stability to the UE-I:α-L-Fuc-OMe complex, and arise from the orientation of the α-L-fucose within the UE-I CRD, in which the hydroxyl groups of the fucose moiety make the most favorable hydrogen bonds with the protein. Thus, disruption of this interaction through replacement with a hydrogen atom (14) would result in significantly reduced binding, as observed in the decreased potency of the congener during inhibition studies (Table 1) (16).

While carbohydrate specificity of the legume lectins is classified based upon which monosaccharide exhibits greatest binding affinity (20), the true specificity for the lectins is for oligomeric structures; this extended specificity arises from interactions between protein and carbohydrate at the periphery of the CRD. Two possible interactions were

Fig. 5. Comparison of the CRD observed in the native UE-I (21) and in the UE-I:α-L-Fuc-OMe complex. Shown is the stereo view of the superposition of the CRD from subunit A of the native UE-I onto the CRD observed from subunit A of the UE-I:α-L-Fuc-OMe complex. The Arg102 side chain was observed to move an average of 1.1 Å towards the carbohydrate upon binding to hydrogen bond with O2 of the α-L-Fuc-OMe molecule. Arg222 was also observed to move towards the CRD upon carbohydrate binding (0.7 Å); however, no other residues exhibited appreciable movement towards the CRD upon binding the α-L-Fuc-OMe. Main chain atoms are in gray, side chain atoms of the native UE-I are in green, side chain atoms of UE-I in the complex are in magenta, and the bound α-L-Fuc-OMe molecule (represented in ball and stick form) is in cyan.

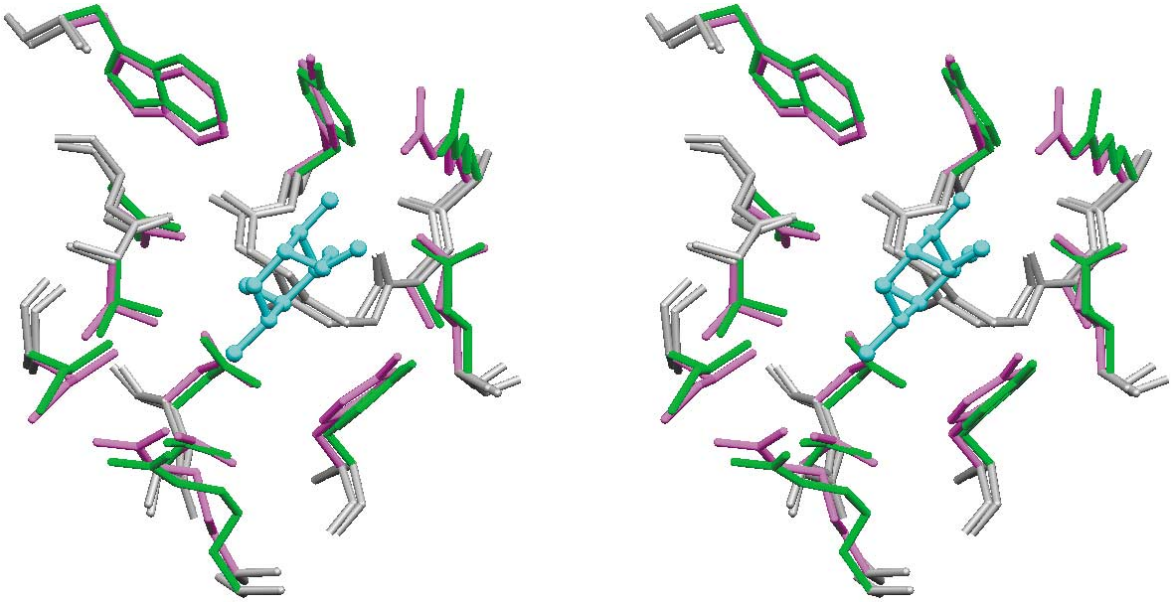


Table 4. Interactions between the bound α-L-Fuc-OMe and UE-I.

α-L-Fuc-OMe	UE-I	Distance (Å)[a]
Hydrogen bonds		
O2	Glu44-$O^{\delta1/\varepsilon2}$	3.5
	Arg102-N^{H2}	3.6
O3	Glu44-$O^{\varepsilon2}$	2.8
	Gly105-N	2.9
O4	Asp87-$O^{\delta2}$	2.8
	Gly105-N	3.0
	Asn134-$N^{\delta2}$	3.0
O5	Asn134-$N^{\delta2}$	3.4
van der Waals contacts[b]		
C1	Val133, Asn134	
C2	Glu44, Tyr106, Asn134	
C3	Glu44, Gly105, Tyr219	
C4	Asp87, Gly105, Asn134, Tyr219	
C5	V133, N134, Y219	
C6	T86, D87, I129, V133, Y219	
C7	Tyr219	
O1	Tyr219	
O2	Glu44, Arg102, Tyr106	
O3	Glu44, Ala103, Gly104, Gly105	
O4	Asp87, Gly104, Gly105, Asn134	
O5	Val133, Asn134	

[a]Hydrogen bond distances are averaged over all four bound α-L-Fuc-OMe molecules.

[b]van der Waals contacts were identified through visual inspection with a maximum distance of 4.0 Å.

observed between UE-I and the H-type 2-OMe — hydrogen bonds between O3 of Gal and Arg102 N^{H2} (3.1 Å) and between O6 of GlcNAc and N^{H1} of Arg222 (2.8 Å). Both interactions correlate with the chemical mapping data (Table 1); both interactions are located at the periphery of the binding site, and disruption of these interactions would result in a slight decrease in binding of the trisaccharide by the protein.

The presence of calcium and manganese ions in the legume lectin has been proposed to induce electron delocalization in the binding site and thus increase the strength of the hydrogen bonds between protein and bound carbohydrate (27). Asn134 both hydrogen bonds to O4 of the α-L-Fuc-OMe through $N^{\delta2}$ (3.0 Å) as well as coordinates the Ca^{2+} through $O^{\delta2}$. Thus, it is likely that the proposed electron delocalization occurs, which results in increasing the strength of the hydrogen bonds between protein and carbohydrate upon complexation.

Finally, an ordered water structure has been identified as playing a role in carbohydrate binding by legume lectins (20, 28–32). The identification of the high entropy barrier to binding of H-type 2-OMe by UE-I ($T\Delta S$ = –20.5 kcal mol^{-1}), as well as a high enthalpy of binding (–29 kcal mol^{-1}), led to the inference that an ordered water structure played a role in H-type 2-OMe binding by UE-I, and to the coining of the term "hydraphobic effect" (33) to describe this role of water in the binding of hydrophilic surfaces in the presence of large molar excesses of water as is found on the surface of proteins. No ordered water molecules were observed in the CRD of UE-I in either UE-I:carbohydrate complex. Indeed, the α-L-Fuc-OMe fits snugly within the UE-I CRD, making

numerous hydrophobic and hydrophilic interactions (Table 4; Fig. 4*b*). It can therefore be inferred that the role of ordered water molecules in carbohydrate binding is in the coordination of calcium and manganese ions, which results in stabilization of the protein–carbohydrate complex through the above noted electron delocalization of binding site residues. It is unlikely that ordered water molecules are present within the CRD after formation of the protein–carbohydrate complex, and the observed entropic and enthalpic effects are due, at least in part, to release of water from the CRD upon carbohydrate binding.

Comparison of modeling and crystallographic studies

The modeling of the H-type 2-OMe in the CRD of the native UE-I (21) predicted several interactions between protein and carbohydrate, which would correlate with the observed chemical mapping data (Table 1). The interactions observed between UE-I and the α-L-Fuc-OMe in the UE-I:α-L-Fuc-OMe complex (Fig. 4*b*; Table 4), and those between UE-I and the bound H-type 2-OMe in the UE-I:H-type 2-OMe complex, also correlate with the chemical mapping data. The observed interactions do not, however, correspond with those predicted from the modeling studies.

The differences between modeled and crystallographically observed binding of the H-type 2 epitope may be understood as follows. During the modeling studies, the roles of Thr86 and Arg102 in carbohydrate binding were not initially apparent. Also, an ordered water structure was thought to play a role in carbohydrate binding; this was found to not be the case upon examination of the CRD in the UE-I:α-L-Fuc-OMe complex (Fig. 4*b*). Finally, the starting orientation of the H-type 2-OMe during modeling was based upon the overlap of the carbohydrate on to the R-MPD observed within the CRD (21). These factors resulted in the minimization of a complex (21) that correlated well with the chemical mapping data but does not correlate with the bound carbohydrate observed in the CRD of UE-I.

Materials and methods

Protein and carbohydrate sources

UE-I was purchased from EY Laboratories (San Mateo, CA, U.S.A., lot No's 141216H0529–30, 150604H0808–30, and 150902J0514–30) in pure form; no further purification was performed. Both α-L-Fuc-OMe and H-type 2-OMe were prepared according to previously reported procedures (14, 16, 34) and provided by Professor R.U. Lemieux (University of Alberta, deceased) in pure form; no further purification was performed.

Mass spectrometric analysis of UE-I

Molecular mass analysis of the two monomeric UE-I subunits was performed by ESI-MS using a Quattro LC triple-quadrupole mass spectrometer (Micromass, Manchester, U.K.) operating in the positive ion mode. An aqueous solution of UE-I (2 μL at 9.4 μM) was introduced by flow injection at 15 μL min^{-1}, using water–acetonitrile (50:50) as the carrier. Electrospray conditions were as follows: capillary 3.78 V; cone 25 V; source temperature 85°C. The first quadrupole was scanned in the *m*/*z* range 600–2000 at 3.95 s per scan. Scans corresponding to the sample elution profile were combined and processed using the MaxEnt software algorithm (MassLynx, version 3.5, Micromass) to obtain molecular weight information for the two UE-I subunits.

Information on the primary structure of UE-I was obtained by MALDI-TOF MS analysis of tryptic digests from SDS-PAGE separated subunits. Approximately 2.5 μg protein were applied to each of three separate lanes on a 12% SDS-PAGE gel in the presence of β-mercaptoethanol as described by Laemmli (35). The separate UE-I bands were excised from the gel, cut into 1 mm squares, and placed in 1.5 mL microcentrifuge tubes for destaining, reduction, alkylation, and tryptic digestion.

Digestion and MS analysis of the SDS-PAGE purified subunits were performed at the National Research Council's Plant Biotechnology Institute in Saskatoon. Tryptic digestion of the excised bands proceeded according to reported protocols (36), after which the peptides were resuspended in 10 μL of 0.1% TFA and desalted by solid-phase extraction using disposable pipette tips ($ZipTip_{C18}$, Millipore) according to the manufacturer's instructions. Peptides were eluted in 1–2 μL of 0.1% TFA/75% acetonitrile directly onto a MALDI target plate. An equal volume of matrix solution (5 mg mL^{-1} α-cyano-4-hydrocinnamic acid in 0.1% TFA/50% acetonitrile) was applied to the sample and allowed to dry. A 1:1 mixture of matrix and mass calibrant was also placed beside the sample for close external calibration. MALDI-TOF MS analysis was performed on a Voyager-DE STR (Applied Biosystems, Framingham, MA) operating in the positive ion, delayed extraction, and reflectron modes and using the following parameters: accelerating voltage 20 000 V; grid voltage 72.5%; guide wire voltage 0.001%; delay 100 ns; laser power 2290. Spectra were obtained by combining 400 scans and applying mass calibration, baseline correction, noise filtering, and de-isotoping procedures using the spectrometer software (Data Explorer v. 4.0, Applied Biosystems).

Crystallization UE-I:carbohydrate complexes

Crystallization of the carbohydrate complexes of UE-I proceeded as follows. The UE-I complexes were produced by adding α-L-Fuc-OMe or H-type 2-OMe to UE-I with a final protein concentration of 10 mg mL^{-1} and carbohydrate concentrations of 0.34 or 0.64 mg mL^{-1} for α-L-Fuc-OMe or H-type 2-OMe, respectively, in 0.1 M sodium cacodylate buffer (pH 7.0), 0.1 mM calcium chloride, and 0.1 mM manganese chloride. These solutions were allowed to stand for at least 1 h on ice prior to crystallization. Crystals were grown by the vapour diffusion method from 4 μL drops containing 5 mg mL^{-1} UE-I:α-L-Fuc-OMe or UE-I:H-type 2-OMe complex, 0.05 M sodium cacodylate buffer (pH 4.0), 0.05 M sodium acetate buffer (pH 4.0), 5 mM calcium chloride, 5 mM manganese chloride, 25–30% (v/v) R,S-MPD, and 0.5% (v/v) PEG 550 MME ether were equilibrated over 1 mL reservoirs containing 0.1 M sodium acetate buffer (pH 4.0), 10 mM calcium chloride, 10 mM manganese chloride, 50–60% (v/v) R,S-MPD and 1% (v/v) PEG 550 MME at 287 or 293 K. These crystallization conditions are similar to those previously reported for the native UE-I (21, 37).

Crystals of both complexes took over 6 months to appear. Crystals of the UE-I:α-L-Fuc-OMe were generally clear plates of average dimensions 0.35 × 0.15 × 0.10 mm that were not suitable for X-ray diffraction analysis due to

disordered stacking of the plates during crystal growth. A single clear, rod-like crystal of α-L-Fuc-OMe complex suitable for X-ray diffraction analysis was, however, produced. The UE-I:α-L-Fuc-OMe complex crystallized in the monoclinic space group $P2_1$, with unit cell dimensions a = 71.81, b = 69.00, and c = 119.02 Å, and β = 106.76°. Two UE-I dimers are present within the asymmetric unit, with a Matthews coefficient (38) of 2.67 $Å^3$ Da^{-1} and an estimated 53.6% (v/v) solvent content. Crystals of the UE-I:H-type 2-OMe complex were generally clear plates of average dimensions 0.30 × 0.15 × 0.10 mm. The UE-I:H-type 2-OMe complex crystallizes in the orthorhombic space group $C222_1$, with unit cell dimensions a = 88.89, b = 164.75, and c = 77.42 Å. A single UE-I dimer is present within the asymmetric unit, with a Matthews coefficient (38) of 2.66 $Å^3$ Da^{-1} and an estimated 52.7% (v/v) solvent content. Crystals of both UE-I complexes were flash cooled for X-ray diffraction analysis in 0.35-mm cryoloops (Hampton Research, Laguna Niguel, CA) by direct immersion in liquid nitrogen. The crystals were then stored in liquid nitrogen prior to X-ray diffraction analysis.

Data collection of the UE-I:α-L-Fuc-OMe complex

X-ray diffraction data of the UE-I:α-L-Fuc-OMe complex were collected by Dr. M. Becker at beamline X12-C of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (Upton, NY) under the remote access (courier based) data collection program.[4] Data were collected in 180 images from a single crystal at 100 K on a Brandeis 1.2 CCD detector (39), using a wavelength of 1.0 Å, an exposure time of 50 s, and a 1° rotation per image. The diffraction data were indexed using the OneButton interface (40) for DENZO and scaled using the SCALEPACK software package (41). Diffraction data were collected to 2.00 Å resolution, however, the data were truncated to 2.30 Å resolution as it was observed that the mean $I/\sigma(I)$ of the data dropped below 2.0 at resolutions above 2.30 Å. A summary of the data collection statistics is shown in Table 3.

Structure solution and model refinement of the UE-I:α-L-Fuc-OMe complex

The molecular replacement method was used to solve the structure of the UE-I:α-L-Fuc-OMe complex using the program AMoRe of the CCP4 software package (42, 43). The search model was the native UE-I dimer (protein only; PDB accession code 1FX5) (21). To determine the initial positions of the two molecules within the asymmetric unit, the top six peaks from the rotation search were used in the translation search. The top peak from the initial translation search was interpreted as the first UE-I dimer, and was then fixed. With the first UE-I dimer fixed, the translation search was repeated to determine the position of the second UE-I dimer with respect to the first. After obtaining the initial coordinates for both molecules, rigid body refinement in AMoRe resulted in a correlation coefficient and R-value of 0.762 and 0.415, respectively.

The molecular replacement solution was subjected to an initial rigid-body, positional, and B-factor refinement followed by the generation of an electron density map using the X-PLOR software package (44, 45). It was observed that the two UE-I:α-L-Fuc-OMe complex molecules within the asymmetric unit were related by non-crystallographic symmetry (NCS) rotation and translation operations. During the initial rounds of structure refinement, the backbone atoms of both the UE-I:α-L-Fuc-OMe complex molecules were, therefore, restrained by the NCS operations. The R-value at the end of this initial round of refinement was 0.288 at 2.5 Å resolution.

Upon examination of the initial $2\boldsymbol{F}_o - \boldsymbol{F}_c$ electron density map using TURBO-FRODO (46), electron density was observed to correspond to a bound α-L-Fuc-OMe within the binding site of each UE-I subunit. The α-L-Fuc-OMe was placed into the electron density using coordinates obtained from hard sphere exo-anomeric (HSEA) calculations (47). Also, electron density was observed that would represent a calcium ion and a manganese ion in each UE-I monomer, known to be present adjacent to the binding sites of legume lectins (20); these ions were placed in the electron density during model building. Electron density was also observed near Cys60 of the third UE-I monomer, which would represent a molecule of the crystallization precipitant, R-MPD. Finally, electron density was observed at residues Asn23 and Asn111 of the UE-I monomers, which would correspond to *N*-linked oligosaccharide. This electron density was not sufficiently resolved, however, to allow for modeling of the *N*-linked carbohydrate.

Rounds of model building were followed by simulated annealing refinement, B-factor refinement, and new σ_A weighted (48, 49) electron density map generation using X-PLOR. Of the 44 883 unique reflections, 5% of the reflections were randomly selected to evaluate R_{Free} (50, 51); a total of 1419 reflections were chosen for the R_{Free} test set. All cycles of structure refinement, with exception of the final cycle, used a working set consisting of 95% of the unique reflections and a test set consisting of the 5% randomly chosen reflections. The final cycle of refinement used all data. The simulated annealing refinements consisted of a prepstage of 160 cycles of minimization followed by a slow-cooling stage starting at 3000 K and decreasing to 300 K in steps of 50 K and a time step of 0.5 fs (1 fs = 10^{-15} s). Slow-cooling was followed by 120 cycles of conjugate gradient minimization and 40 cycles of B-factor refinement. Engh and Huber (52) bond and angle parameters and a bulk solvent correction were employed throughout refinement; weighting was obtained from the CHECK procedure in X-PLOR. Water molecules were placed into the structure manually using a difference ($\boldsymbol{F}_o - \boldsymbol{F}_c$) electron density map. Upon refinement, some water molecules developed large B-factors (>40 $Å^2$). These were removed from the structure and new water molecules added as they were located in subsequent difference maps.

The space group of the UE-I:α-L-Fuc-OMe complex was determined to be $P2_1$, with two UE-I dimers within the

[4] Diffraction data for the UE-I:α-L-Fuc-OMe complex were collected at the single-crystal diffraction facility at beamline X12-C of the National Synchrotron Light Source at Brookhaven National Laboratory. This facility is supported by the United States Department of Energy Offices of Health and Environmental Research and of Basic Energy Sciences under prime contract DE-AC02–98CH10886, by the National Science Foundation, and by National Institutes of Health Grant 1P41 RR12408–01A1.

asymmetric unit. The two UE-I:α-L-Fuc-OMe complex molecules within the asymmetric unit were related by non-crystallographic symmetry (NCS) rotation and translation operations. The NCS rotation angle between the two molecules is 179.18°, and the NCS rotation axis is almost parallel to the *b*-axis. This relation between UE-I dimers is such that the crystal exhibits pseudo-C centering, which results in the intensities of the $h + k$ odd reflections being weak in comparison to the $h + k$ even reflections. This distribution of reflection intensities results in a large number of the reflections which are weak. It is known that pseudo-centering of a crystal, with this large number of weak reflections, results in abnormally high R and R_{Free} values. For this reason, the refinement of the UE-I:α-L-Fuc-OMe complex employed a low amplitude cut-off. The final R-value and R_{Free} of the UE-I:α-L-Fuc-OMe complex are 0.202 and 0.289, respectively.

Data collection of the UE-I:H-type 2-OMe complex

X-ray diffraction data for the UE-I:H-Type 2-OMe complex were collected at BioCARS beamline 14-BM-D of the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL).[5] Data were collected in 179 images from a single crystal at 100 K on a ADSC Quantum 4 CCD detector (Area Detector Systems Corporation, Poway, CA), using a wavelength of 1.00 Å, an exposure time of 15 s, a 1° rotation per image, and a crystal-to-detector distance of 168.5 mm. The diffraction data were indexed and scaled using the DENZO/SCALEPACK software package (41). A summary of the data collection statistics is shown in Table 3.

Structure solution and model refinement of the UE-I:H-type 2-OMe complex

The molecular replacement method was used to solve the structure of the UE-I:H-type 2-OMe complex using the program AMoRe of the CCP4 software suite (42, 43). The search model was the native UE-I dimer (PDB accession code 1FX5) (21). To determine the initial position of the UE-I dimer within the asymmetric unit, the top five peaks from the rotation search were used in the translation search. Rigid body refinement in AMoRe of the initial coordinates for the UE-I dimer resulted in a correlation coefficient and R-value of 0.561 and 0.447, respectively.

Upon examination of the initial $2\boldsymbol{F}_o - \boldsymbol{F}_c$ electron density map using TURBO-FRODO (46), density was observed to correspond to a bound H-Type 2-OMe within CRD of the first UE-I subunit. The electron density was not sufficiently resolved at this stage, however, to allow for fitting of the H-Type 2-OMe within the carbohydrate binding site. Also, electron density was observed that would represent a calcium and a manganese ion in each UE-I monomer, and electron density was observed at residues Asn23 and Asn111 of the UE-I monomers, which would correspond to *N*-linked oligosaccharide. The calcium and manganese ions and bound H-Type 2-OMe were added to the model in a similar fashion as that described for the UE-I:α-L-Fuc-OMe complex during subsequent rounds of model building; electron density was insufficiently resolved to allow for addition of *N*-linked oligosaccharide.

Refinement was performed as described for the UE-I:α-L-Fuc-OMe complex. During refinement, 5% of the unique reflections (493) were randomly chosen to evaluate R_{Free}, and the backbone atoms of the UE-I subunits were restrained by the NCS operators by which the subunits were related. The final R-value and R_{Free} of the structure are 0.284 and 0.421, respectively. At the resolution of the crystallographic data for the UE-I:H-Type 2-OMe complex (30–3.0 Å), it was not possible to locate solvent (water) molecules.

Protein data bank accession numbers

The coordinates and structure amplitudes for the UE-I:α-L-Fuc-OMe complex have been deposited in the RCSB Protein Data Bank (accession codes RCSB014305 and 1JXN). The coordinates and structure factor amplitudes for the UE-I:H-type 2-OMe complex have not been deposited in the PDB due to the low resolution of the crystallographic data; deposition shall be done upon refinement of the model against higher resolution diffraction data.

Acknowledgments

The authors wish to express their gratitude to Professor R.U. Lemieux, who provided both the carbohydrates for this study and insightful discussions prior to his death. The authors also wish to thank Dr. L. Prasad for helpful discussions, and Dr. M. Becker (NSLS) and Mr. G. Navrotsky (BioCARS) for their assistance during data collection. G.F.A. gratefully acknowledges financial assistance from the Colleges of Graduate Studies and Medicine from the University of Saskatchewan. This work was supported by a Canadian Institutes for Health Research operating grant to L.T.J.D. (MT-10162) and a Natural Sciences and Engineering Research Council of Canada (NSERC) operating grant to J.W.Q.

References

1. N. Sharon and H. Lis. Science (Washington, D.C.), **246**, 227 (1989).
2. N. Sharon and H. Lis. Essays Biochem. **30**, 59 (1995.).
3. N. Sharon and H. Lis. Chem. Rev. **98**, 637 (1998).
4. R.U. Lemieux, R. Cromer, and U. Spohr. Can. J. Chem. **66**, 3083 (1988).
5. R.U. Lemieux, U. Spohr, M. Bach, D.R. Cameron, T.P. Frandsen, B.B. Stoffer, B. Svensson, and M.M. Palcic. Can. J. Chem. **74**, 319 (1996).
6. W.C. Boyd and E. Shapleigh. Blood, **9**, 1195 (1954).
7. I. Matsumoto and T. Osawa. Biochim. Biophys. Acta, **194**, 180 (1969).
8. K. Landsteiner. Wien. Klin. Wschr. **14**, 1132 (1901).
9. K. Landsteiner. Biochem. Z. **104**, 280 (1920).

[5] Diffraction data for the UE-I:H-type 2-OMe complex were collected at beamline 14BM-D of the BioCARS Sector 14 at the Advanced Photon Source. Use of the Advanced Photon Source was supported by the United States Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31–109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707.

10. W.M. Watkins. *In* Glycoproteins. Their composition, structure and function. *Edited by* A. Gottschalk. Elsevier, Amsterdam. 1972. pp. 830–891.
11. O. Hindsgaul, D.P. Khare, M. Bach, and R.U. Lemieux. Can. J. Chem. **63**, 2653 (1985).
12. P.V. Nikrad, H. Beiberbeck, and R.U. Lemieux. Can. J. Chem. **70**, 241 (1992).
13. L.T.J. Delbaere, M. Vandonselaar, L. Prasad, J.W. Quail, K.S. Wilson, and Z. Dauter. J. Mol. Biol. **230**, 950 (1993).
14. O. Hindsgaul, T. Norberg, J. Le Pendu, and R.U. Lemieux. Carbohydr. Res. **109**, 109 (1982).
15. U. Spohr, E. Paszkiewicz-Hnatiw, N. Morishima, and R.U. Lemieux. Can. J. Chem. **70**, 254 (1992).
16. M. Du, U. Sphor, and R.U. Lemieux. Glycoconjugate J. **11**, 443 (1994).
17. N.K. Vyas. Curr. Opin. Struct. Biol. **1**, 732 (1991).
18. N.M. Young and R.P. Oomen. J. Mol. Biol. **228**, 924 (1992).
19. V. Sharma and A. Surolia. J. Mol. Biol. **267**, 433 (1997).
20. R. Loris, T. Hamelryck, J. Bouckaert, and L. Wyns. Biochim. Biophys. Acta, **1383**, 9 (1998).
21. G.F. Audette, M. Vandonselaar, and L.T.J. Delbaere. J. Mol. Biol. **304**, 423 (2000).
22. Y. Konami, K. Yamamoto, and T. Osawa. J. Biochem. **109**, 650 (1991).
23. R.A. Laskowski, M.W. MacArthur, D.S. Moss, and J.M. Thornton. J. Appl. Cryst. **26**, 283 (1993).
24. C. Ramakrishnan and G.N. Ramachandran. Biophys. J. **5**, 909 (1965).
25. V. Luzzatti. Acta Cryst. **A5**, 802 (1952).
26. R.U. Lemieux. Chem. Soc. Rev. **7**, 423 (1978).
27. L.T.J. Delbaere, M. Vandonselaar, L. Prasad, J.W. Quail, J.R. Pearlstone, M.R. Carpenter, L.B. Smillie, P.V. Nikrad, U. Spohr, and R.U. Lemieux. Can. J. Chem. **68**, 1116 (1990).
28. Y. Bourne, P. Rouge, and C. Cambillau. J. Biol. Chem. **265**, 18161 (1990).
29. R. Loris, P.P.G. Stas, and L. Wyns. J. Biol. Chem. **269**, 26722 (1994).
30. R. Ravishankar, K. Suguna, A. Surolia, and M. Vijayan. Acta Crystallog. Sect. D, **55**, 1375 (1999).
31. M. Vijayan and N. Chandra. Curr Opin. Struct. Biol. **9**, 707 (1999).
32. J. Habash, J. Raftery, R. Nuttall, H.J. Price, C. Wilkinson, A.J. Kalb, and J.R. Helliwell. Acta Crystallog. Sect. D, **56**, 541 (2000).
33. R.U. Lemieux. Acc. Chem. Res. **29**, 373 (1996).
34. M. Dejter-Juszynski and H.M. Flowers. Carbohydr. Res. **41**, 308 (1975).
35. L. Laemmli. Nature (London), **227**, 680 (1970).
36. O.N. Jensen, M. Wilm, A. Shevchenko, and M. Mann. *In* 2-D proteome analysis protocols. *Edited by* A.J. Link. Humana Press, Totowa, NJ. 1999. pp. 513–530.
37. M. Vandonselaar and L.T.J. Delbaere. J. Mol. Biol. **243**, 345 (1994).
38. B.W. Matthews. J. Mol. Biol. **33**, 491 (1968).
39. W.C. Phillips, M. Stanton, A. Stewart, H. Qian, C. Ingersoll, and R.M. Sweet. J. Appl. Crystallog. **33**, 243 (2000).
40. R.M. Sweet and J.M. Skinner. *In* Proceedings from the macromolecular crystallography computing school. *Edited by* P.E. Bourne and K. Watenpaugh. Western Washington University, WA, U.S.A. August 17–22, 1996.
41. Z. Otwinowski and W. Minor. Methods Enzymol. **276**, 307 (1997).
42. CCP4. The SERC (U.K.) collaborative computing project No. 4: A suite of programs for protein crystallography. Acta Crystallog. Sect. D, **50**, 760 (1994).
43. J. Navaza. AMoRe: an automated package for molecular replacement. Acta Crystallog. Sect. A, **50**, 157 (1994).
44. A.T. Brünger, J. Kuriyan, and M. Karplus. Science (Washington, D.C.), **235**, 458 (1987).
45. A.T. Brünger. X-PLOR. Version 3.1. A system for X-ray crystallography and NMR. Yale University Press, New Haven and London. 1993.
46. A. Roussel and C. Cambillau. TURBO-FRODO. Biographics, Marseilles, France. 1992.
47. R.U. Lemieux, K. Bock, L.T.J. Delbaere, S. Koto, and V.S. Rao. Can. J. Chem. **58**, 631 (1980).
48. R.J. Read. Acta Crystallog. Sect. A, **42**, 140 (1986).
49. R.J. Read. Acta Crystallog. Sect. A, **46**, 900 (1990).
50. A.T. Brünger. Nature (London), **355**, 472 (1992).
51. A.T. Brünger. Acta Crystallog. Sect. D, **49**, 24 (1993).
52. R.A. Engh and R. Huber. Acta Crystallog. Sect. A, **47**, 392 (1991).
53. S.V. Evans. J. Mol. Graph. **11**, 134 (1993).