ELSEVIER

Knowledge-based modeling of a legume lectin and docking of the carbohydrate ligand: The *Ulex europaeus* lectin I and its interaction with fucose

Arnaud Gohier,* Juan Felix Espinosa,† Jesus Jimenez-Barbero,† Pierre-Alain Carrupt,‡ Serge Pérez,* and Anne Imberty*

**CERMAV-CNRS, [1]Grenoble, France*
†Instituto de Quemica Organica, CSIC, Madrid, Spain
‡Institut de Chimie Thérapeutique, BEP, Université de Lausanne, Lausanne-Dorigny, Switzerland

Ulex europaeus *isolectin I is specific for fucose-containing oligosaccharide such as H type 2 trisaccharide* α-L-*Fuc (1→2)* β-D-*Gal (1→4)* β-D-*GlcNAc. Several legume lectins have been crystallized and modeled, but no structural data are available concerning such fucose-binding lectin. The three-dimensional structure of* Ulex europaeus *isolectin I has been constructed using seven legume lectins for which high-resolution crystal structures were available. Some conserved water molecules, as well as the structural cations, were taken into account for building the model. In the predicted binding site, the most probable locations of the secondary hydroxyl groups were determined using the GRID method. Several possible orientations could be determined for a fucose residue.*

All of the four possible conformations compatible with energy calculations display several hydrogen bonds with Asp-87 and Ser-132 and a stacking interaction with Tyr-220 and Phe-136. In two orientations, the O-3 and O-4 hydroxyl groups of fucose are the most buried ones, whereas two other, the O-2 and O-3 hydroxyl groups are at the bottom of the site. Possible docking modes are also studied by analysis of the hydrophobic and hydrophilic surfaces for both the ligand and the protein. The SCORE method allows for a quantitative evaluation of the complementarity of these surfaces, on the basis of molecular lipophilicity calculations. The predictions presented here are compared with known biochemical data. © *1996 by Elsevier Science Inc.*

Keywords: Molecular modeling, docking, fucose, Ulex europaeus

Color Plate for this article are on page 363–364.

Address reprint requests to: Dr. Anne Imberty, CERMAV-CNRS, BP 53, F-38041 Grenoble Cedex 09, France

Paper submitted as part of the Electronic Conference of the Molecular Graphics and Modelling Society, October 1996.

[1]Affiliated with Université Joseph Fourier, Grenoble, France.

INTRODUCTION

Specific recognition of carbohydrates by protein receptors is of growing interest in biology. In addition to immunoglobulins, carriers, and toxins, carbohydrates can interact with lectins, a class of multivalent and ubiquitous carbohydrate-binding proteins.[1] These proteins can be purified easily and their specificities make them useful for labeling and separating complex oligosaccharides such as the histo-blood group determinants. The *Ulex europaeus* isolectin I (UEL 1) is specific for fucose and fucose-containing oligosaccharides such as H type 2 trisaccharide α-L-Fuc (1→2) β-D-Gal (1→4) β-D-GlcNAc.[2] At the present time, the primary sequence is known[3] but there is no available X-ray structure for this lectin.

Since several crystal structures are available for legume lectins, building the three-dimensional structure of the protein by homology methods can be envisaged. The *Dolichos biflorus* lectin, which is specific for *N*-acetylgalactosamine and can therefore agglutinate A-type red blood cells, has already been modeled.[4] The lectins with known tridimensional structures belong to the mannose-specific family or to the galactose (or galactose-containing oligosaccharide)–

Journal of Molecular Graphics 14:322–327, 1996

655 Avenue of the Americas, New York, NY 10010

0263-7855/97/$15.00
PII S0263-7855(97)00010-6

specific family. Since the lectin studied here is specific for fucose, the major problem is the docking of the carbohydrate moiety in the lectin-binding site. Our group previously described methods for such docking studies[5] along with energy parameters appropriate for protein–carbohydrate interactions.[6]

We describe here different approaches that can be used to model the orientation of α-L-fucopyranose in the *Ulex europaeus* lectin-binding site. The aim is to take advantage of the presence of both hydrophilic and lipophilic faces on the carbohydrate. The GRID method[7] is used on the protein surface to predict the anchoring points for the fucose hydroxyl groups, whereas calculation of molecular properties such as the lipophilicity[8] could help in selecting the most probable docking mode.

METHODS

Homology modeling

Structure database Crystal structures are available for eight different legume lectins. For each lectin, the crystal structure selected has been the one complexed with a carbohydrate, if any, and with the highest resolution. Seven three-dimensional (3D) structures of lectins have been used in the present study: *Griffonia simplicifolia* isolectin IV (GSIV) complexed with blood group B determinant[9] (Protein Data Base [pdb] code 1LED), *Erythrina corallodendron* lectin (ECorL) complexed with lactose[10] (pdb code 1LTE), concanavalin A (ConA) complexed with mannose,[11] *Lathyrus ochrus* isolectin I (LOL) complexed with mannose[12] (pdb code 1LOB), lentil lectin (LCL) complexed with sucrose[13] (pdb code 1LES), pea lectin (PSL)[14] (pdb code 2LTN) and soybean lectin (SBL) complexed with oligosaccharide[15] (pdb code 1SBA). All of the structures were taken from the Protein Data Bank with the exception of the ConA structure, which is a gift from K. Hardmann.

Sequence alignment The sequence alignment between UEL I and the seven other legume lectins was performed with the graphical method HCA (hydrophobic cluster analysis).[16] This method is primarily based on the detection and comparison of hydrophobic clusters, which are indicative of the regular secondary structure elements constituting the hydrophobic core of globular proteins. In these plots, protein sequences are written on a duplicated-helical net and the clusters of contiguous hydrophobic residues (V, I, L, F, M, W, Y) are drawn. The analysis involves the visual comparison of hydrophobic cluster shapes, and their distribution, in order to find correspondences between the different plots.

In Figure 1, the 3D features of the top five lectins, (i.e., β strands and amino acids involved in the binding) are taken from an examination of their crystal structures. The sequence displayed here have about 40% identity. The LCL and PSL sequences are not shown since they are very similar to the LOL sequence. For UEL 1, the 3D features are deduced from those of the others, according to the shapes of the clusters.

Protein building The COMPOSER program,[17] within the SYBYL software package, was used for building the protein. The library of 3D crystallographic structures used in this study contains the seven legume lectins listed above. Structurally conserved regions mainly consist of the β strands (Figure 2).

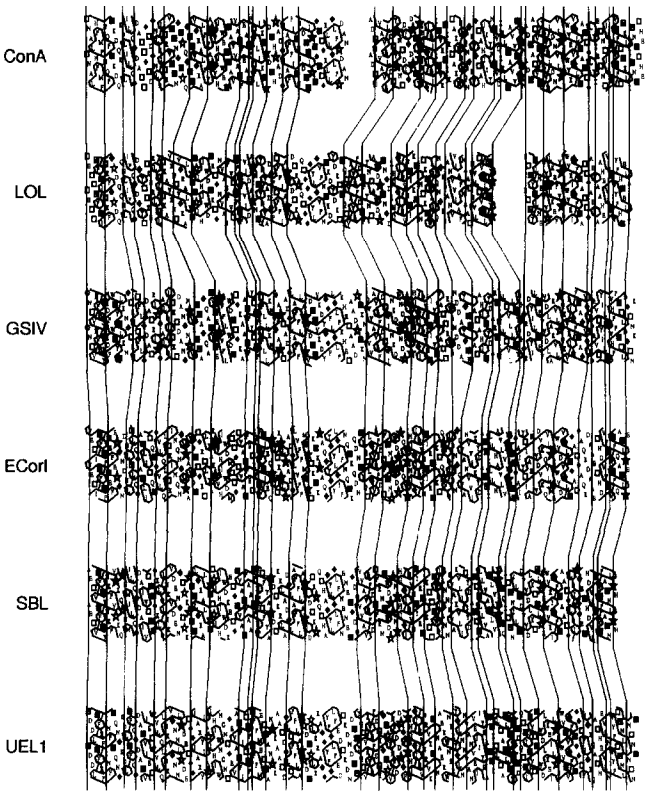

Figure 1. HCA plots of the legume lectin sequences. The limits of the β strands are indicated by vertical lines.

Special care was given to the construction of the loops in the variable region around the binding site. Some conserved water molecules, as well as the structural cations, were incorporated into the model. Finally, hydrogen atoms were added and charges were calculated. A validation of the model stereochemistry was performed using the PROCHECK program.[18]

Figure 2. Superimposition of the X-ray structures of the seven lectins of the database.

Figure 3. Ribbon representation of the model of Ulex europaeus *lectin I. The six water molecules included in the model are represented as sticks and the ions are represented by dotted surfaces.*

Docking of carbohydrate

Calculations of anchoring positions The GRID program[7] was used to predict the more favorable anchoring positions for the fucose residue. Both an OH (hydroxyl) probe and a CH_2 (methylene) probe were used in the calculations. The grid spacing was set to 0.5 Å and the calculations were limited to the binding site. In the results, the two lowest energy grid points were used for fitting two hydroxyls of fucose. Four solutions were therefore investigated, corresponding to the four pairs O-2/O-3, O-3/O-2, O-3/O-4, and O-4/O-3.

Complex optimization A search of the possible orientations for the fucose in the binding site was then undertaken, while taking the information from GRID as anchoring points. For each solution, a virtual bond was created between the two oxygen atoms and the sugar ring was oscillated about this bond to investigate its best position in the binding site. For all the possible complexes, several steps of minimization were performed using the TRIPOS force field with appropriate parameters for protein–carbohydrate complexes.[6]

Analysis of the models The four possible docking modes are analyzed in terms of energy of interaction, hydrogen-bonding schemes, and number of hydrophobic contacts. Also, the fits between the corresponding faces (hydrophilic and lipophilic surfaces) of the ligand and the protein are analyzed both graphically and quantitatively. The molecular lipophilicity potential (MLP) offers a quantitative 3D description of lipophilicity.[8] The best docking mode for the ligand in a protein-binding site should therefore be characterized by a maximal similarity between the MLP of the ligand (the intrinsic MLP) and that of the binding site (the perceived MLP). This similarity can be quantified by a SCORE function[19]: The more positive the SCORE function, the greater the similarity between intrinsic and perceived MLP.

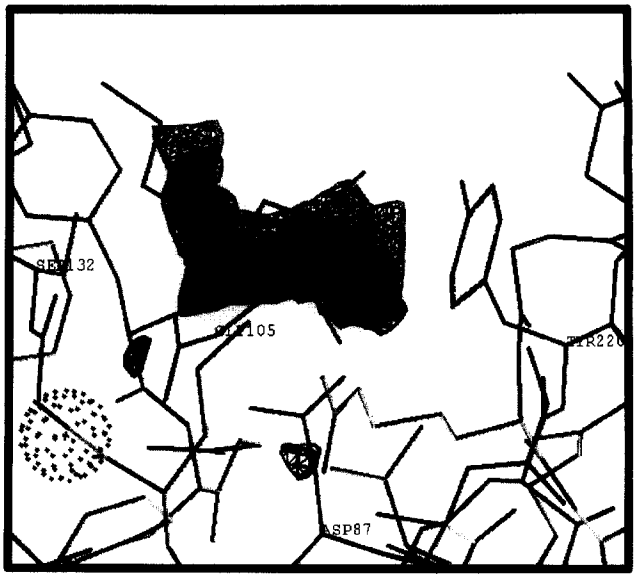

Figure 4. Energy isocontours displaying the favorable positions for an OH probe (black contours) and a CH2 probe (gray contours) as determined by the GRID program.

RESULTS

Model of *Ulex europaeus* isolectin I

Overall architecture When looking at the overall architecture of the monomer (Figure 3), the lectin displays the same features as other legume lectins. The calcium and manganese ions are located near the carbohydrate-binding sites. The water molecules that were shown to be conserved

Table 1. Prediction of hydrogen bonds and hydrophobic contacts between lectin and fucose

	Hydrogen bonds		Hydrophobic contacts	
Solution	Fucose	Lectin	Fucose	Lectin
1	HO-3	Asp-87.OD1	C-1H	Tyr-220
	O-3	Ser-132.OH	C-2H	Tyr-220
	HO-4	Asp-87.OD2	C-3H	Phe-136
			C-4H	Phe-136
			C-5H	Phe-136
			Methyl	Tyr-220
2	HO-4	Asp-87.OD1	C-1H	Phe-136
	O-4	Ser-132.OH	C-2H	Phe-136
	HO-3	Asp87.OD2	C-3H	Tyr-220
			Methyl	Ile-130
3	HO-2	Asp-87.OD1	C-1H	Phe-136
	O-2	Gly-105.HN	C-3H	Tyr-220
	HO-3	Asp-87.OD2	C-4H	Tyr-220
			C-5H	Tyr-220
4	HO-3	Asp-87.OD1	C-1H	Tyr-220
	O-3	Ser-132.OH	C-3H	Tyr-220
	HO-2	Asp-87.OD2	C-4H	Phe-136
	O-2	Gly-105.HN	Methyl	Tyr-220

in all legume lectins[20] are incorporated in that model. Five of them play a role in the coordination of the cations, whereas the sixth one stabilizes a loop at the opposite side of the monomer.

Carbohydrate-binding site When comparing the binding site of UEL 1 with other lectins, some differences appear. The aspartate residue that is always observed in the bottom of the binding site is present here (Asp-87). However, the aspargine residue that usually gives one hydrogen bond to the carbohydrate ligand is replaced here by a serine residue (Ser-132). In general, legume lectins have an aromatic residue in the bottom of the site, which creates a stacking interaction with the hydrophobic face of the carbohydrate. In UEL 1, this position is occupied by a hydrophobic (Ile-130) amino acid. However, another aromatic residue, Tyr-220, is present in the binding site.

Analysis of the GRID results demonstrate a favorable region for OH groups above the carboxyl group of Asp87 (Figure 4). A low-energy isocontour for the CH probe is located above this OH-binding region. Such an arrangement agrees well with the stereochemistry of carbohydrate.

Docking of fucose

Description of the possible docking modes From the GRID study, it is possible to fit two hydroxyl groups above the carboxyl oxygens of Asp-87. Taking into account all of the combinations for locating O-2, O-3, and O-4 of fucose in these two possible positions, four different possibilities must be explored for docking the fucose in the binding site. For each possibility, the best orientation of the ring was then determined and all of the complexes were submitted to several energy minimization cycles. Color Plate 1 displays the four optimized complexes. In solutions 1 and 2, the O-3/O-4 pair of hydroxyl groups interacts with Asp-87, whereas in solutions 3 and 4 the O-2/O-3 pair plays this role.

As listed in Table 1, an additional hydrogen bond is given by Ser-132 (except in solution 3). This cooperative hydrogen-bonding scheme, involving two hydroxyl groups from the carbohydrate, and one carboxyl group and one hydroxyl group from the protein, is identical to what is observed in other legume lectins. The change from asparagine to serine does not disturb this network. When compared to other crystal structures or to the models of legume lectins, the number of hydrogen bonds predicted here is rather low. This

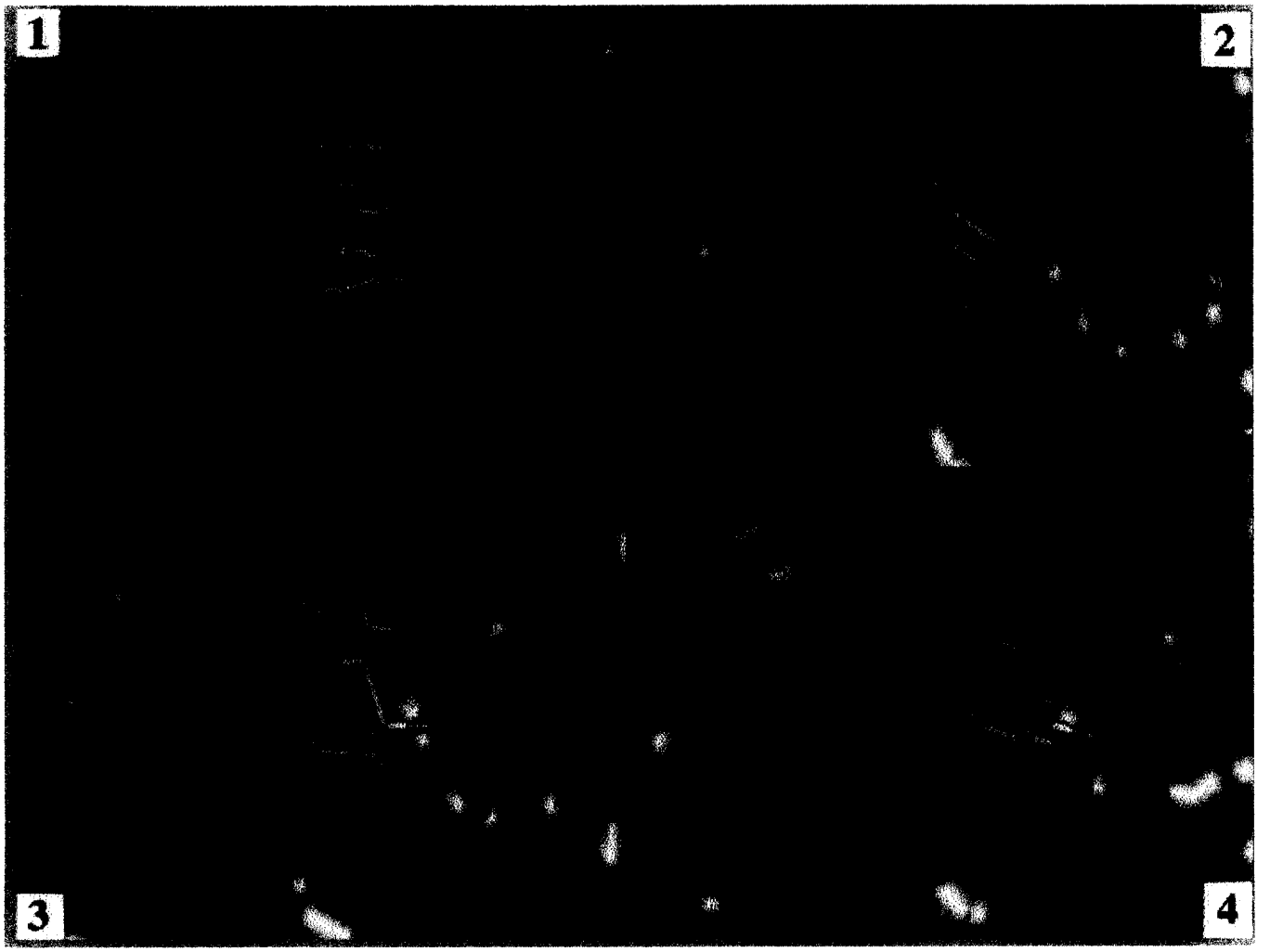

Figure 5. The four possible orientations of fucose in the binding site of Ulex *lectin. The Connolly surface has been represented for the protein-binding site.*

could be due to the hydrophobic characteristics of fucose. In all four docking modes displayed in Color Plate 1, stacking interactions are observed between the fucose and both Phe-136 and Tyr-220 amino acids. As for the calculated energy of interaction, no significant differences appeared between the four displayed solutions.

Analysis of the surfaces To discriminate between the different docking modes, the agreement between the shape of the fucose residue and the binding site surface was visualized (Figure 5). In all four docking modes, the fucose appears to fit between two hydrophobic walls created by Tyr-220 and Phe-136.

The hydrophilic contacts are established at the bottom of the well. Since the balance between hydrophilic and hydrophobic contacts seems to be of first importance in the binding mode of fucose, we used another approach to quantify the quality of the docking modes. Color Plate 2 displays the molecular lipophilicity potentials of fucose and UEL 1 in both docking modes. The agreement between the intrinsic MLP and the perceived MLP is quantified by the SCORE function. In all solutions, the hydroxyl groups in the bottom of the site (top of each drawing in Color Plate 2) create an excellent agreement between the hydrophilic area of the fucose and the binding site. There are more variations in the fitting of the lipophilic surfaces, but the agreement is good for the four solutions. This is reflected by the high value of the SCORE function. This function is slightly higher for solution 3, but the differences could not be considered as significant.

DISCUSSION AND CONCLUSION

The modeling results presented here can be compared with previously published binding data. Several derivatives of fucose have been tested by R.U. Lemieux et al.[21] The results indicate that removing either O-4 or O-3 hydroxyl groups yields totally inactive compounds whereas removing the O-2 group only lowers the binding ability. The same experiments were conducted on the fucose moiety of H type 2 trisaccharide, which is the best ligand for UEL 1.[22] In this case, the O-4 hydroxyl is also absolutely necessary for binding. Removal of the O-3 group yields an almost inactive compound whereas removal of the O-2 group has a less dramatic effect. From these studies, it has been concluded that the three hydroxyl groups are buried in the binding site, and that O-3 and O-4 are more directly involved in hydrogen bonding with the host protein.[23] These experimental studies would be in agreement with solution 1 or 2 of our modeling study.

We present here an attempt to construct the three-dimensional model of *Ulex europaeus* lectin I solely on the basis of knowledge of its primary structure. The most difficult part was to model the orientation of the fucose ligand in the binding site, since the stereochemistry of this carbohydrate differs from that observed in crystal structures of lectin complexes. An extensive docking study yielded four possible docking modes for the ligand. The results show the complementary forces that emanate from hydrogen bonding and van der Waals forces, including hydrophobic interactions. At the present time, and in the absence of further experimental evidence, it is not possible to conclude which of the four possible docking modes is most probable. Further studies, involving larger oligosaccharides and using transferred NOESY experiments, are in progress in order to obtain additional information on the binding.

ACKNOWLEDGMENT

The collaboration between France and Spain was supported by a PICASSO bilateral program.

REFERENCES

1 Sharon, N. and Lis, H. Lectins as cell recognition molecules. *Science* 1989, **246,** 227–246
2 Cazal, P. and Lalaurie, M. Recherches sur quelques phyto-agglutinines specifiques des groupes sanguins ABO. *Acta Haematol.* 1952, **8,** 73–82
3 Konami, Y., Yamamoto, K., and Osawa, T. The primary structures of two types of the *Ulex europeus* seed lectin. *J. Biochem.* 1991, **109,** 650–658
4 Imberty, A., Casset, F., Gegg, C.V., Etzler, M.E., and Perez, S. Molecular modeling of the *Dolichos biflorus* seed lectin and its specific interaction with carbohydrates: α-D-*N*-Acetyl-galactosamine, Forssman disaccharide and blood group A trisaccharide *Glycoconj. J.* 1994, **11,** 400–413
5 Imberty, A., Hardman, K.D., Carver, J.P., and Pérez, S. Molecular modeling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A. *Glycobiology* 1991, **1,** 456–484
6 Pérez, S., Meyer, C., and Imberty, A. Practical tools for accurate modeling of complex carbohydrates and their interactions with proteins. In *Modelling of Biomolecular Structures and Mechanisms* (A. Pullman, J. Jortner, and B. Pullman, eds.). Kluwer Academic Publishers, Dordrecht, 1995, pp. 425–454
7 Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important molecules. *J. Med. Chem.* 1985, **28,** 849–857
8 Testa, B., Carrupt, P.-A., Gaillard, P., Billois, F., and Weber, P. Lipophilicity in molecular modeling. *Pharmaceut. Res.* 1996, **13,** 335–343
9 Delbaere, L.T., Vandonselaar, M., Prasad, L., Quail, J.W., Wilson, K.S., and Dauter, Z. Structures of lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution. *J. Mol. Biol.* 1993, **230,** 950–965
10 Shaanan, B., Lis, H., and Sharon, N. Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science* 1991, **254,** 862–866
11 Hardman, K.D., Agarwal, R.C., and Freiser, M.J. Manganese and calcium binding sites of concanavalin A. *J. Biol. Chem.* 1982, **157,** 69–86
12 Bourne, Y., Roussel, A., Frey, M., Rougé, P., Fontecilla-Camps, J.C., and Cambillau, C. Three-dimensional structure of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: Fine specificity of the monosaccharide binding site. *Proteins Struct. Funct. Genet.* 1990, **8,** 365–376
13 Casset, F., Hamelryck, T., Loris, R., Brisson, J.R., Tellier, C., Dao-Thi, M.H., Wyns, L., Poortmans, F.,

Pérez, S., and Imberty, A. NMR, molecular modeling, and crystallographic studies of lentil lectin–sucrose interaction. *J. Biol. Chem.* 1995, **270,** 25619–25628

14 Prasthofer, T., Phillips, S.R., Suddath, F.L., and Engler, J.A. Design, expression, and crystallization of recombinant lectin from the garden pea (*Pisum sativum*) *J. Biol. Chem.* 1989, **264,** 6793–6796

15 Dessen, A., Gupta, D., Sabesan, S., Brewer, C.F., and Sacchettini, J.C. X-Ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. *Biochemistry* 1995, **34,** 4933–4942

16 Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J.P. Hydrophobic cluster analysis: An efficient new way to compare and analyse amino acid sequences. *FEBS Lett.* 1987, **224,** 149–155

17 Blundell, T.L., Carnay, D.P., Gardner, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J.P., Singh, D.A., Sibanda, B.L., and Sutcliffe, M.J. Knowledge-based protein modelling and design. *Eur. J. Biochem.* 1988, **172,** 513–520

18 Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 1993, **26,** 283–291

19 Carrupt, P.A., Gaillard, P., Billois, F., Weber, P., Testa, B., Meyer, C., and Pérez, S The molecular lipophilicity potential (MLP): A new tool for log *P* calculations and docking, and in comparative molecular field analysis (CoMFA). In *Lipophilicity in Drug Action and Toxicology* (V. Pliska, B. Testa, and H. van de Waterbeemd, eds.). VCH, Weinheim, 1996, pp. 195–217

20 Loris, R., Stas, P.P.G., and Wyns, L. Conserved waters in legume lectin crystal structures. *J. Biol. Chem.* 1994, **269,** 26722–26733

21 Hindsgaul, O., Khare, D.P., Bach, M., and Lemieux, R.U. Molecular recognition III. The binding of the H-type 2 human blood group determinant by the lectin I of *Ulex europaeus. Can. J. Chem.* 1985, **63,** 2653–2658

22 Hindsgaul, O., Norberg, T., Le Pendu, J., and Lemieux, R.U. Synthesis of type 2 human blood-group antigen determinants. The H, X, and Y haptens and variations of the H type 2 determinant as probes for the combining site of the lectin I of *Ulex europaeus. Carbohydr. Res.* 1982, **109,** 109–142

23 Lemieux, R.U. The chemical mapping of protein–carbohydrate complexes. In *Complex Carbohydrates in Drug Research* (K. Bock and H. Clausen, eds.). Munskgaard, Copenhagen, 1994, pp. 188–197