

2.2 Å Resolution Structure Analysis of Two Refined *N*-acetylneuraminyl-lactose–Wheat Germ Agglutinin Isolectin Complexes

Christine S. Wright

Department of Biochemistry and Molecular Biophysics
MCV/VCU, Box 614 MCV Station
Richmond, VA 23298-0001, U.S.A.

(Received 19 March 1990; accepted 25 June 1990)

The crystal structures of complexes of isolectins 1 and 2 of wheat germ agglutinin (WGA1 and WGA2) with *N*-acetylneuraminyl-lactose (NeuNAc- α (2-3)-Gal- β (1-4)-Glc) have been refined on the basis of data in the 8 to 2.2 Å resolution range to final crystallographic *R*-factors of 17.2% and 15.3% ($F_o > 1\sigma$), respectively. Specific binding interactions and water association, as well as changes in conformation and mobility of the structure upon ligand binding, were compared in the two complexes. The temperature factors ($B = 16.3 \text{ Å}^2$ and 18.4 Å^2) were found to be much lower compared with those of their respective native structures (19 to 22 Å^2). Residues involved in sugar binding, dimerization and in lattice contacts exhibit the largest decreases in *B*-value, suggesting that sugar binding reduces the overall mobility of the protein molecules in the crystal lattice.

The binding mode of this sialyl-trisaccharide, an important cell receptor analogue, has been compared in the two isolectins. Only one of the two unique binding sites (4 per dimer), located in the subunit/subunit interface, is occupied in the crystals. This site, termed the “primary” binding site, contains one of the five amino acid substitutions that differentiate WGA1 and WGA2. Superposition of the refined models in each of the independent crystallographic environments indicates a close match only of the terminal non-reducing NeuNAc residue (root-mean-square Δr of 0.5 to 0.6 Å). The Gal-Glc portion was found to superimpose poorly, lack electron density, and possess high atomic thermal factors. In both complexes NeuNAc is stabilized through contact with six amino acid side-chains (Ser114 and Glu115 of subunit 1 and Ser62, Tyr64, Tyr(His)66 and Tyr73 of subunit 2), involving all NeuNAc ring substituents. Refinement has allowed accurate assessment of the contact distances for four hydrogen bonds, a strong buried non-polar contact with the acetamido CH_3 group and a large number of van der Waals’ interactions with the three aromatic side-chains. The higher affinity of *N*-acetylneuraminyl-lactose observed by nuclear magnetic resonance studies for WGA1 can be explained by the more favorable binding interactions that occur when residue 66 is a Tyr. The tyrosyl side-chain provides a larger surface for van der Waals’ stacking against the NeuNAc pyranose ring than His66 and a hydrogen bond contact with Gal (C2-OH), not possible in WGA2.

Saccharide-induced conformational changes in the side-chains involved in sugar binding were found to be small but comparable in all crystallographically different complexes, and they correlate with the degree of immobilization reflected by the temperature factors.

1. Introduction

Protein–carbohydrate interactions have been described for a collection of proteins of diverse

† Abbreviations used: WGA, wheat germ agglutinin; WGA1, wheat germ agglutinin isolectin 1; WGA2, wheat germ agglutinin isolectin 2; P-site, primary sugar binding site; S-site, secondary sugar binding site; n.m.r., nuclear magnetic resonance; NeuLac, *N*-acetylneuraminyl-lactose; NLI1, *N*-acetylneuraminyl-lactose–WGA1 complex; NLI2, *N*-acetylneuraminyl-lactose–WGA2 complex; r.m.s., root-mean-square; HA, hemagglutinin.

origin and function, based on crystal structure data (Quirocho, 1986; Reeke & Becker, 1988; Weis *et al.*, 1988; Rini, 1986; Delbaere *et al.*, 1989). Oligomeric plant lectins, which display stringent sugar specificities, constitute a special class of these proteins. Although their physiological function in the plant remains obscure, they have served as important tools *in vitro* for targeting carbohydrate receptors in eukaryotic cell systems (Goldstein & Poretz, 1986).

As a representative of the highly conserved group of lectins of the Gramineae family, wheat germ agglutinin (WGA†) possesses several properties

distinct from other plant lectins: (1) high stability due to the presence of 16 disulfide bridges distributed over four homologous isostructural domains, A, B, C and D; (2) sugar specificity for two different types of saccharide, *N*-acetyl-D-glucosamine (GlcNAc) and *N*-acetylneuraminic acid (NeuNAc) (α -anomer) (Goldstein & Hayes, 1978; Peters *et al.*, 1979; Bhavanandan & Katlic, 1979; Monsigny *et al.*, 1980), in contrast to most lectins that display specificity for a single type of sugar only; (3) requirement of both subunits for sugar binding (Wright, 1980b); (4) the presence of two independent, non-co-operative binding sites per subunit (4 per dimer: Nagata & Burger, 1974; Privat *et al.*, 1974a; Kronis & Carver, 1985a), located in domain contact regions (A/D, B/C) of opposite subunits and related by local 2-fold symmetry (Wright, 1984).

Spectroscopic measurements probing the environment of bound saccharide have not been able to separate effects specific to each of these two independent sites (Jordan *et al.*, 1977; Thomas *et al.*, 1977; Privat *et al.*, 1974b; Kronis & Carver, 1985a). However, it has been possible to characterize these sites by crystal structure analysis of the WGA-(GlcNAc)₂ complex (Wright, 1980b), and a remarkable degree of homology in the amino acids essential for binding has been noted (Wright, 1984). Although crystals are unstable in the presence of GlcNAc and its β (1-4)-linked oligomers, mild glutaraldehyde treatment allowed identification of the two unique sites at 2.8 Å (1 Å = 0.1 nm) resolution. Only one site was found to be highly occupied and was termed the "primary" or P-site. At the other, low occupancy site ("secondary" or S-site), sugar binding competes with lattice interactions between neighboring molecules. NeuNAc sugars are excluded from this second site due to charge repulsion of its carboxylate group.

Three closely related variants of WGA, called isolectins, have been characterized in hexaploid wheat (Allen *et al.*, 1973; Rice & Etzler, 1975; Etzler, 1985; Wright & Raikhel, 1989). The two major isolectins (WGA1 and WGA2), which differ in five sequence positions, have been the subject of careful crystallographic studies (Wright, 1987, 1989) and equilibrium sugar binding studies using nuclear magnetic resonance (n.m.r.) techniques (Kronis & Carver, 1982, 1985a,b). The affinities of the receptor analogue *N*-acetylneuraminyl-lactose (NeuNAc- α (2-3)Gal- β (1-4)Glc: NeuLac) for WGA1 and WGA2 have been compared by n.m.r. techniques, indicating higher affinity by a factor of 2 to 5 for WGA1 (Kronis & Carver, 1982). Since a sequence difference (His66 to Tyr66) is observed in only one of the two unique sites (P-site), this difference in affinity has been attributed to binding at this site.

WGA exhibits specificity only for acetylated sialic acid sugars. This has been explained in structural terms (Wright, 1980b) in that NeuNAc satisfies the stereochemical requirement for an equatorial *N*-acetyl group and an adjacent equatorial OH group, the specificity determining factors for GlcNAc sugars. Sialic acid sugars play an important

role as mediators of biological recognition. They are commonly found in exposed terminal positions of cell surface carbohydrate chains and serve as receptors for viruses, hormones, lectins, toxins, mycoplasma and protozoa. Thus, structural studies of sialyloligosaccharide-protein complexes on the atomic level should provide some general insights into the structural requirements for sialic acid interaction with external protein ligands.

Due to the selective binding of NeuNAc to the P-site in WGA crystals, sialyloligosaccharide crystal complexes are stable and suitable for high resolution studies. Thus, 2.2 Å resolution data were measured on the NeuLac complexes of both WGA1 and WGA2, and the structures crystallographically refined.

This paper presents the results of the comparison of the two refined complexes in terms of specific interactions of bound saccharide with protein groups and ordered water molecules to explain the difference in the binding affinity of NeuLac for WGA1 and WGA2. The two complexes are also examined both in independent crystallographic environments and for saccharide-induced conformational changes and differences in mobility with respect to their native structures. Comparison of the refined native structures at 2.0 Å resolution indicated a close match of all residues that interact with saccharide in this binding site (Wright, 1989).

2. Experimental Procedures

Native WGA1 and WGA2 crystals, which belong to space group C2, were grown as reported (Wright, 1977). The sugar complexes were prepared by soaking for 1 to 2 week periods in solutions containing 3 mM-neuramin-lactose (85% (w/v) *N*-acetylneuraminyl- α (2-3)galactose- β (1-4)glucose, Sigma Inc.), 10 mM-acetate buffer (pH 6.0), 10 mM-NaCl, 0.02% (w/v) sodium azide. Intensity data were collected on a CAD4 Nonius diffractometer mounted on a sealed Cu fine-focus X-ray tube as described (Wright, 1987). Five crystals were needed for the complete 2.2 Å datasets of each of the isolectin complexes. In the case of the WGA2 complex, a 4 Å dataset collected earlier using a rotating anode source (Wright, 1980b), was merged with this more recent high-resolution dataset. The intensity data from the multiple crystals were merged and scaled using the computer library ROCKS (G. N. Reeke Jr) on an IBM 3081 computer.

The final datasets scaled with *R*-factors of 8.2% (6327 overlaps, 22,663 measurements) for the WGA1 complex (NLI1) and 8.7% (8902 overlaps, 23,032 measurements) for the WGA2 complex (NLI2). Of a total of 17,266 unique reflections 10,893 are observed (63% >1 σ (I)) in the case of NLI1 and 11,925 of a total of 17,516 are observed (68% >1 σ (I)) in the case of NLI2. The effective resolution for each of the datasets, computed using the program of Swanson (1988), is 2.46 Å (NLI1) and 2.40 Å (NLI2).

Approximate atomic co-ordinates for the trisaccharide were obtained from low-resolution work (Wright, 1980b) by molecular modeling in an optical comparator (Richards, 1968). For initial comparisons of the 2 complexes, the electron density was averaged over the 2 non-crystallographically related monomers in the asymmetric unit and maps with coefficients $[2F(\text{complex}) - F(\text{native})] \exp(i\phi_c)$ and $[F(\text{complex}) - F(\text{native})]$

Table 1
Crystal parameters for refined WGA structures

	WGA1	NLI1	WGA2	NLI2
A. Unit cell parameters				
<i>a</i> (Å)	51.22	51.00	51.34	51.44
<i>b</i> (Å)	73.60	73.64	73.53	73.35
<i>c</i> (Å)	91.42	91.41	91.54	91.68
α (°)	90.00	90.00	90.00	90.00
β (°)	98.00	97.91	97.59	97.52
γ (°)	90.00	90.00	90.00	90.00
B. Position of local 2-fold axis				
<i>x</i> -translation (Å)	6.30	6.23	6.055	5.99
<i>y</i> -position (Å)	-0.028	-0.038	-0.036	-0.040
<i>z</i> -position (Å)	22.69	22.66	22.68	22.71

$\exp(i\phi_c)$ were examined by computer graphics (Evans and Sutherland, PS330). Unaveraged maps were found to have a higher noise level, but did not reveal any additional features.

The 2 individual monomers in the asymmetric unit, which will here be referred to as monomers I and II (see Fig. 1 of Wright, 1989), were refined by the restrained parameter refinement method (PROLSQ) of Hendrickson & Konnert (1980). All observed diffraction data ($F_o > 1\sigma$) in the resolution range 8 to 2.2 Å were included in the refinement. The observed structure factors were pre-scaled by a constant to be approximately on absolute scale. Rescaling was periodically necessary during refinement to maintain a value close to 1.0 for K_c in the expression $F_c = K_c f_c \exp(-Bs^2)$. The C2 lattice dimensions used in each refinement are the average over the parameters determined for the 5 crystals used for data collection on each of the 2 crystal complexes. As can be discerned from Table 1, they are closely isomorphous with those of their native structures. The trisaccharide was idealized as a special group against a standard conformation derived by model building using published atomic co-ordinates from the crystal structures of *N*-acetylneuraminic acid (Flippen, 1973) and *N*-acetyl-lactosamine (Longchambon *et al.*, 1981). During refinement, bond distances and angles of the trisaccharide group were restrained as well as the planar conformation of the acetamido and carboxylate substituents of NeuNAc. Atomic positions for protein atoms and water (refined as oxygen) were adopted from the refined native structures (Wright, 1987, 1989). The 2 water molecules in the sugar binding site were removed from the list. All atoms were refined with an occupancy of 1.0 except for several water-oxygen sites situated on the 2-fold axis, which were refined with an occupancy of 0.5. After 10 to 11 initial cycles of refinement at 3.0 Å and 2.5 Å resolution, using an overall temperature factor, high-resolution data were included in 6 additional cycles refining individual atomic *B*-values. This reduced the initial *R*-factors of about 25% to 17% and 18% in the 2 refinements. All water sites were carefully examined, and weak sites deleted. Additional water molecules were added during subsequent refinement cycles. A set of 15 atoms, consisting either of water molecules located on the dimer axis or of atoms that caused persistent close contacts, was kept fixed during later refinement cycles. Due to the presence of negative difference density for a portion of the trisaccharide (Gal-Glc) the atomic occupancies were lowered to values that eliminated the negative difference density [$F_o - F_c$]. Since this had the effect of diminishing the

B-values substantially, the occupancies were reset to 1.0 in 4 additional cycles at the end of the refinement. At this stage difference density was practically absent.

For comparison purposes, refinements were carried out on both native structures, WGA1 and WGA2, in the same resolution range (8 to 2.2 Å) and using all observed data ($F_o > 1\sigma$). The initial models used were those obtained from the refinements at higher resolution (2.0 and 1.8 Å; Wright, 1987, 1989). These models were first refined with an overall temperature factor at 2.5 Å in 6 to 7 cycles, and then at 2.2 Å in an additional 11 to 12 cycles of temperature factor refinement. Their final co-ordinates agree with r.m.s. differences of 0.16 Å (WGA1) and 0.12 Å (WGA2), including water, with those obtained in the earlier refinements at higher resolution.

Electron density maps of the type $[2F_o - F_c]$ were computed at intermittent refinement stages using all terms in the resolution range 10 to 2.2 Å. These maps were completely examined after the first 15 to 16 cycles of refinement by computer graphics using the program FRODO (Jones, 1978). Difference maps $[F_o - F_c]$ were examined for additional water sites and misplaced atoms.

Structural comparisons were carried out with each of the independent subunit molecules separately, applying a translational adjustment in the case of molecule 2 to correct for the difference in the non-crystallographic relationships of the 2 isoelectin complexes (see Table 1).

The refined atomic co-ordinates for NLI1 and NLI2 have been deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).

3. Comparison of the Isoelectin Complexes

(a) Refinement

Progress in the refinement of the two complexes differed markedly. Some of the important parameters of the final cycle of refinement are compared in Table 2. As can be judged from the low *R*-factor, the NLI2 complex refined very well. Over 64 cycles of least-squares refinement the *R*-factor decreased steadily from an initial value of 24.9% to a final value of 15.3% using all observed data ($> 1\sigma$) in the 8 to 2.2 Å resolution range. Up to cycle 45 only reflections with F_o values greater than 3σ were used. As an indication of the quality of the weaker data, it is interesting to note that inclusion of all observed

Table 2
Refinement statistics for final cycle

	NLI1	NLI2
Resolution range of data (Å)	8.0-2.2	8.0-2.2
No. of reflections $F_o > 1\sigma(F_o)$	10,522	11,537
Total no. of atoms	2626	2632
No. of solvent atoms	214	235
Final r.m.s. co-ordinate shift (Å)	0.013	0.009
Final r.m.s. temp. factor shift (Å ²)	0.36	0.36
Average temperature factor (Å ²)	16.41	18.30
$F_o - F_c$	96.60	79.35
<i>A</i> , <i>B</i> †	52, -140	42, -135
Final <i>R</i> -factor‡	17.2%	15.3%

† Structure factor weights were determined from:

$$\sigma(F_o) = A + B(\sin \theta/\lambda - 0.166).$$

‡ $R = \sum |F_o| - |F_c| / \sum |F_o|$.

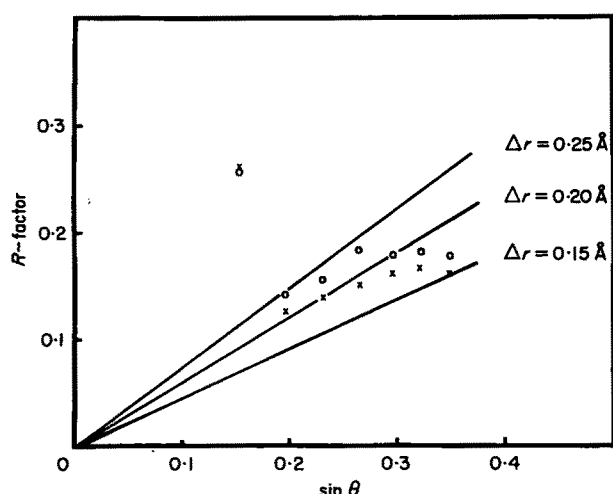


Figure 1. Variation of the R -factor for the final cycles of the NLI1 (○) and NLI2 (×) refinements as a function of resolution ($\sin \theta$). The average R -factors in each resolution range are based on all observed data ($F_o > 1\sigma$). The superimposed Luzzati (1952) lines provide an upper estimate for the co-ordinate errors.

data in subsequent cycles (80 with $F_o < 3\sigma$) did not increase the R -factor (15.6% at this stage). In contrast, a significant increase in the R -factor was observed during refinement of the NLI1 complex (from 16.6% to 17.7%) due to inclusion of a much larger number of weak reflections (999 with $F_o < 3\sigma$). NLI1 refined with a significantly higher final R -factor (17.2%) than NLI2 (15.3%), although its average atomic B -factor is significantly lower (16.4 Å² versus 18.3 Å²). The large discrepancy in the R -factors cannot be explained only on the basis of the larger number of solvent molecules included in the NLI2 refinement (see Table 2), but is most likely due to the fact that the NLI2 dataset consists of a larger number of observed data in the high-resolution range (1100 more between 2.6 to 2.2 Å). While both intensity datasets scaled with similar merge R -factors (8.2% and 8.7%), the average intensity for the NLI2 dataset was higher by a factor of 2.0 over that of NLI1. This would also suggest a higher degree of accuracy of the high-resolution data and explain the significantly lower value for the final average residual [$F_o - F_c$] for NLI2 (see Table 2). The R -factor variation with respect to resolution is shown in Figure 1, indicating an approximate mean co-ordinate error of 0.2 to 0.25 Å (Luzzati, 1952).

Refinement of both native structures at 2.2 Å yielded very similar final R -factors (see Table 3). In these refinements a larger number of water sites were included in the models than in the earlier high-resolution refinements (33 additional sites for WGA2 and 7 for WGA1). These were water molecules located in the respective complexes for which electron density was found to be present also in the native maps. Statistics for the final cycles of the four refinements at 2.2 Å are shown in Table 3. Significant differences are seen in the average

Table 3
Comparison of refinement parameters

	NLI1	WGA1	NLI2	WGA2
A. Co-ordinate refinement				
No. of cycles	11	6	10	7
Resolution range (Å)	8–2.5	10–2.5	8–2.5	8–2.5
No. of reflections $F_o > 1\sigma$	7943†	8990	8538†	9780
R -factor‡	19.0%	18.9%	19.3%	19.6%
Overall B -factor (Å ²)	9.5	15.2	10.6	16.1
B. Co-ordinate and B-factor refinement				
Resolution range of data (Å)	8–2.2	8–2.2	8–2.2	8–2.2
Effective resolution§ (Å)	2.48	2.45	2.40	2.43
No. of reflections $F_o > 1\sigma(F_o)$	10,522	11,671	11,537	12,475
Reflections in 2.4–2.2 Å shell	1436	1821	2083	1725
Total no. of atoms	2624	2530	2634	2545
Average temperature factor (Å ²)	16.44	19.40	18.30	22.00
$F_o - F_c$	96.58	76.00	79.35	72.35
Final R -factor‡	17.2%	16.5%	15.3%	16.4%

† $F_o > 3\sigma(F_o)$.

‡ $R = \sum \|F_o\| - \|F_c\| / \sum \|F_o\|$.

§ As defined by Swanson (1988).

atomic B -factors and the average values of the residuals [$F_o - F_c$]. Despite the fact that WGA2 was refined against a larger number of reflections than WGA1 and the two complexes, its final R -factor was not the lowest. It is apparent from Table 3 that the magnitude of the final R -factor correlates more satisfactorily with the completeness of the data in the high-resolution shell (2.4 to 2.2 Å), which is highest for NLI2.

Refinement caused distortion of the trisaccharide geometry, although bond distances, angles, two planar groups and 1–3 contacts were restrained. In particular, the conformation of the hexose rings tended to distort to a more planar conformation. This effect was most serious in the case of the glucose residue, probably due to the lack of electron density. The proper conformation was restored by model building at each stage of the refinement.

(b) Electron density maps

Comparing the 2.2 Å electron density maps of the two complexes, poorer definition is observed for the NLI1 complex. This made it more difficult to eliminate close contacts during refinement and manual intervention was necessary to a larger extent to restore geometry. Model readjustment increased the R -factor at each new stage by 1 to 1.5%, whereas an increase of only 0.5 to 0.8% was observed with NLI2.

Inspection of disordered regions indicates that a higher level of electron density compared to that of the native structure is present for the side-chains of Lys149 and Trp150 (refined as Ala in native WGA1) in independent monomer I in the case of NLI1. This has allowed refinement of the complete side-chains of Lys149 and Trp150 in the complex. Lys134 was refined as Ala in both monomers. This residue was

Table 4
Comparison of average B-values ($\text{\AA}^2$)

	WGA1	NLI1	WGA2	NLI2
A. Monomer I†				
Domain A	14.92	11.82	16.89	13.68
Domain B	15.72	11.31	18.13	14.46
Domain C	16.22	11.52	17.72	14.57
Domain D	27.22	25.72	31.85	26.79
B. Monomer II				
Domain A	15.30	10.96	16.84	13.54
Domain B	17.72	12.91	19.04	14.71
Domain C	19.11	15.80	21.92	19.49
Domain D	22.55	18.73	24.80	20.15
C. Monomers I and II				
All protein atoms	18.94	15.81	20.93	17.80
All water molecules	27.33	23.23	30.81	25.49
Trisaccharide (NeuLac)		40.35		39.31

† I and II refer to the 2 independent monomers in the asymmetric unit.

formerly misidentified as Gly (Wright & Raikhel, 1989) due to complete lack of side-chain density. However, in the NLI1 map weak density is

observed for this side-chain in monomer II. As in native WGA2, in NLI2 residues Lys134 and Lys149 in monomer I and Trp150 and Lys134 in monomer II were refined as Ala, although slightly more density is seen for these side-chains. In contrast, the flexible C-terminal Ala (refined as Gly) possesses more density in monomer I of native WGA2 than in the sugar complex, while the opposite is true for monomer II, where the terminal carboxylate group can form an ion pair with Lys96 of monomer I. This terminal domain has a much higher overall B-factor (see Table 4) in monomer I than in monomer II (32 versus 25 $\text{\AA}^2$). These observations would support the notion, suggested from the difference in the overall temperature factors, that the protein structure is more ordered in the complexed as compared with the native state.

Examination of the electron density maps in the region of the trisaccharide indicates that good electron density is present only for the terminal NeuNAc residue. This is in agreement with the difference density observed in difference maps of the type $[F(\text{complex}) - F(\text{native})] \exp(i\phi_c)$. Comparing the two complexes shows marginally more density for the NeuNAc-Gal portion in NLI1 than in NLI2.

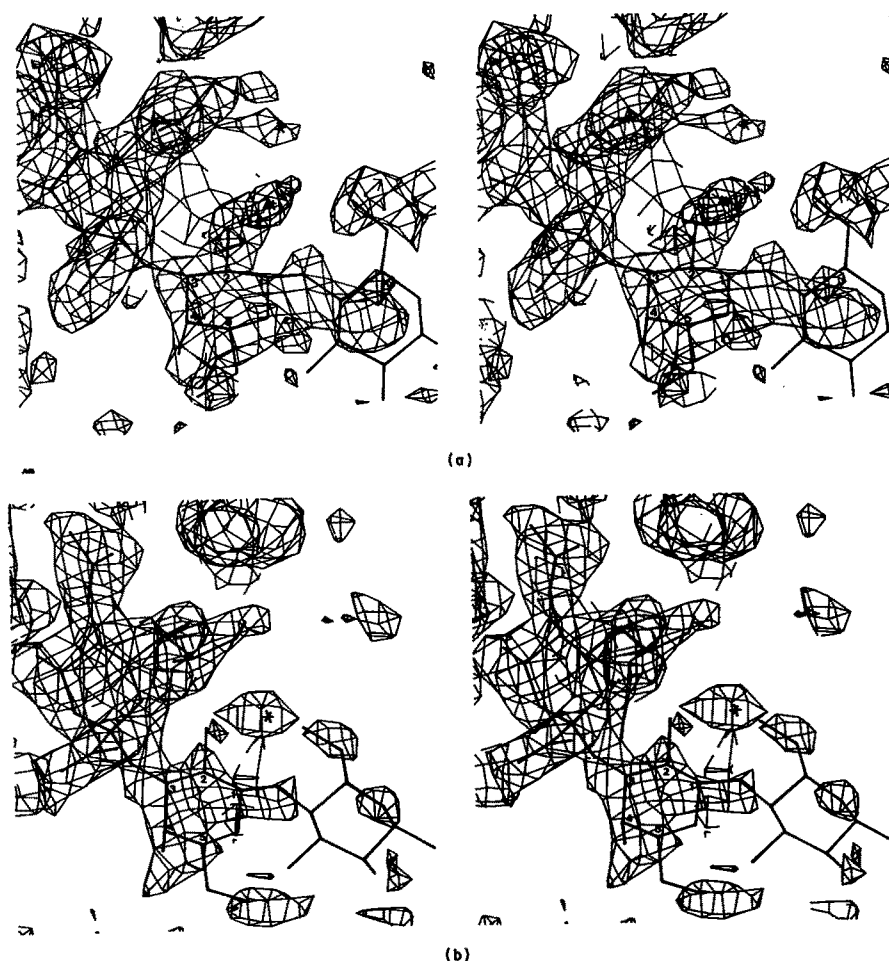
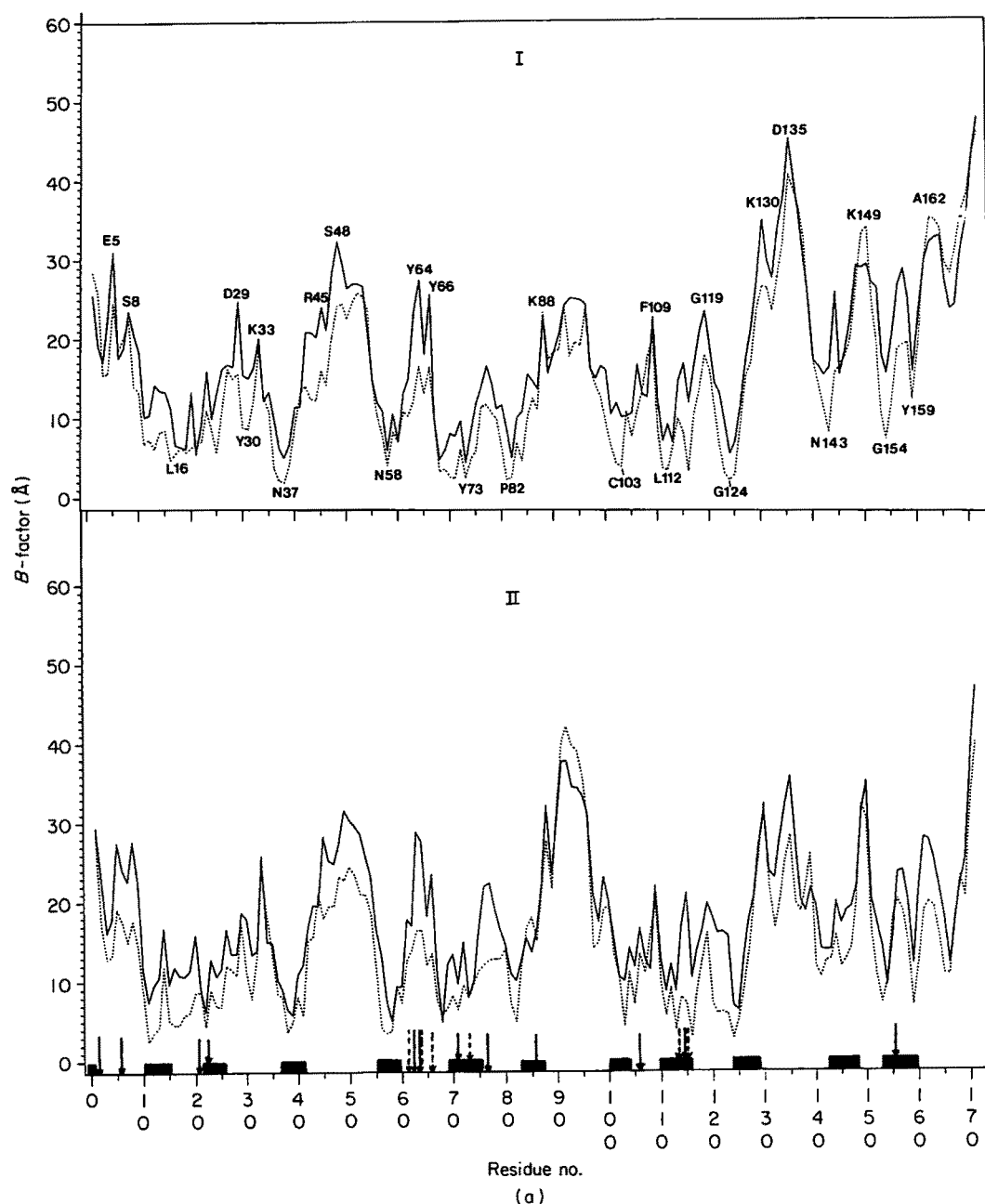


Figure 2. (a) Stereo view of the region of the NLI1 electron density map ($F_o - F_c$) representing the bound NeuLac trisaccharide. The refined positions of water molecules are shown by asterisks. (b) Stereo view of the region of the NLI2 electron density map ($2F_o - F_c$) representing the bound NeuLac trisaccharide. The contour level is identical with that shown in (a).



In particular, the glycerol and carboxylate of NeuNAc and the C2-OH portion of Gal have better density in NLII, as shown in Figure 2(a) and (b) for independent monomer I. The glucose residues lack density almost completely in both complexes, indicating disorder. This is consistent with the high atomic B -values obtained for this residue ($>50 \text{ \AA}^2$).

(c) Temperature factors

A dramatic drop in the atomic B -values, which had initially been adopted from the refined native structures, was immediately observed when B -factors were included in the second stage of refinement. The final values obtained for both the overall isotropic B -factors and for the average value of the refined atomic B -factors of the two complexes

are lower by 30% and 15%, respectively, compared with those of the native structures (see Tables 3 and 4). In Table 4, the average values are also compared for each of the four domains in the four refinements. The difference between the values for native *versus* complexed crystals is much larger for protein atoms than water sites. Inspecting B -values of groups of atoms within side-chains and comparing corresponding residues in the two independent environments and in both isolectin structures, shows reasonable agreement. Figure 3(a) and (b) compare the distribution of the averaged B -values for each amino acid along the polypeptide chain in the trisaccharide complexes and native structures. Although the general appearance of the two curves is similar, overall lower values are observed for the complex. When analyzing this on a percentage

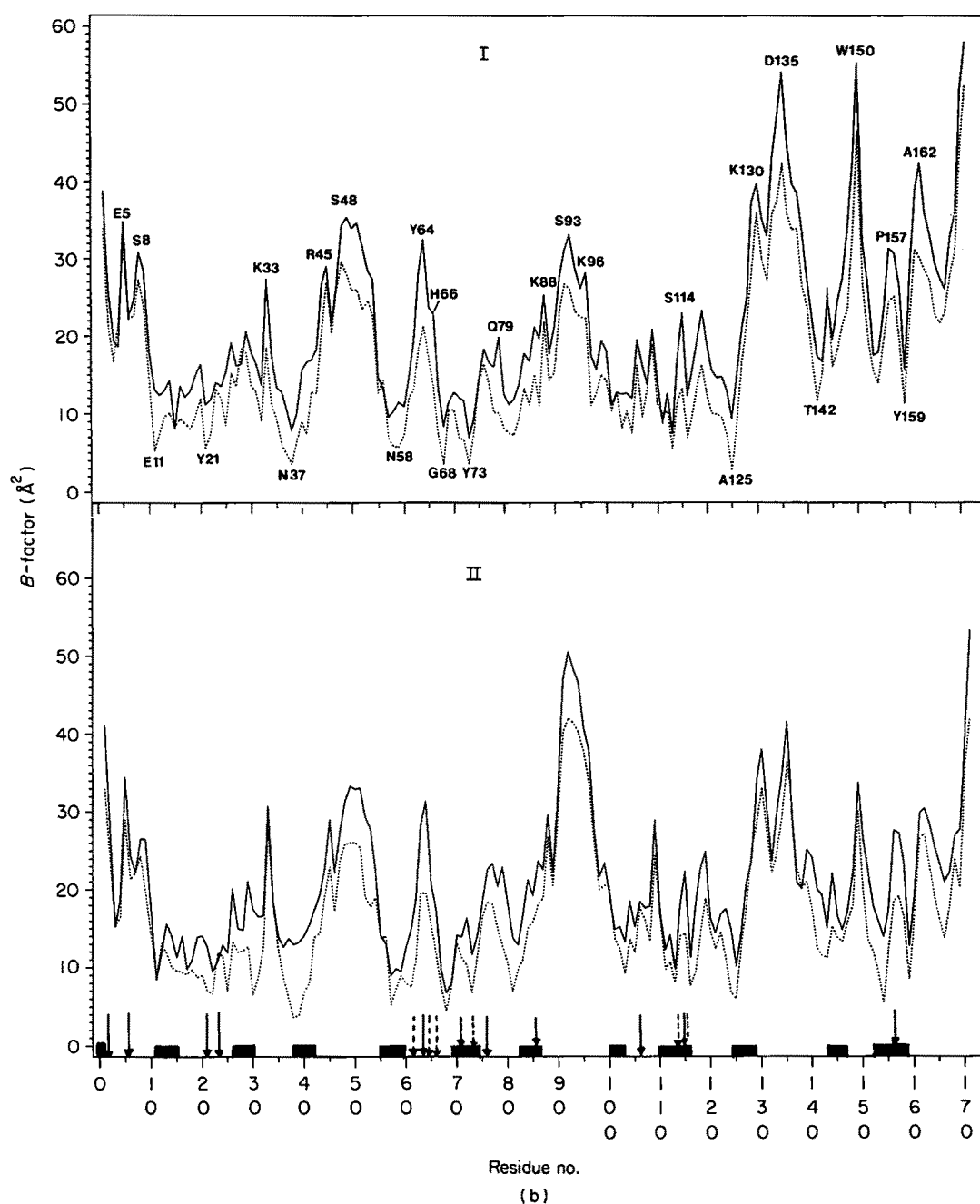


Figure 3. (a) Comparison of the variation of the refined average amino acid B -values along the polypeptide chain between native WGA1 (continuous line) and its NeuLac complex (broken line). Values for independent monomer I are shown on top and those for monomer II on the bottom. Regions of the polypeptide chain marked by filled bars indicate involvement in subunit-subunit interactions. Residues flagged by filled arrows are involved in intermolecular lattice contacts and residues flagged by broken arrows interact with bound saccharide. (b) Comparison of the variation of the refined average amino acid B -values along the polypeptide chain between native WGA2 (continuous line) and its NeuLac complex (broken line). I and II refer to independent monomers I and II. Specific regions of the polypeptide chain are marked as described in (a).

basis, it is clear that some regions are more affected than others. For instance, a much larger decrease is seen in residues that are involved in intermolecular interactions in the subunit/subunit interface and in lattice interactions, as marked in Figure 3(a) and (b). Residues exposed in flexible surface loop regions, on the other hand, are not affected as much. However, there is some disagreement in the two

lattice environments. For instance, a much higher degree of disorder is observed for the D-domain in monomer I than in monomer II. The reduction in the average B -value for these domains is, however, much more pronounced in monomer I (see Table 4). This is probably due to the fact that the D-domain in monomer II is involved to a higher degree in lattice contacts.

Table 5
Restraint information

	NLI1			NLI2		
	Target value	r.m.s. deviation	No. of restraints	Target value	r.m.s. deviation	No. of restraints
Distances (Å)						
Bond	0.025	0.018	2524	0.022	0.017	2513
Angle	0.040	0.041	3438	0.036	0.039	3424
Planar	0.045	0.039	914	0.042	0.038	900
Planar groups	0.035	0.021	2248	0.028	0.018	2240
Chiral volume (Å)	0.200	0.212	284	0.180	0.193	282
Non-bonded contacts (Å)						
Single torsion	0.50	0.214	847	0.5	0.203	851
Multiple torsion	0.50	0.249	809	0.5	0.231	726
Possible H-bond	0.50	0.297	212	0.5	0.257	233
Torsion angles (°)						
Planar (peptide)	3.0	2.8	348	3.0	2.6	348
Staggered	15.0	20.3	336	15.0	21.2	329
Orthonormal	20.0	41.5	28	20.0	42.8	29
Isotropic thermal factors						
Main-chain bond	2.0	1.57	1560	2.0	1.76	1568
Main-chain angle	3.0	2.58	2046	3.0	2.78	2060
Side-chain bond	2.0	1.55	954	2.0	1.84	935
Side-chain angle	3.0	2.48	1402	3.0	2.94	1374

The average atomic *B*-factors for the trisaccharide were found to vary with distance of the sugar residue from the binding site. The values obtained in the two different crystal environments for NeuNAc vary from 29 to 37 Å², for Gal from 44 to 46 Å², and for Glc from 51 to 57 Å².

(d) *Superposition of the refined models*

The refined atomic models of NLI1 and NLI2 had final values of 0.018 Å and 0.017 Å, and 2.1° and 2.0° for the r.m.s. deviation from ideal bond distances and angles, respectively. The overall quality of the models is comparable, as can be discerned from Table 5. In the case of NLI1, however, a larger number of residues remained with poor conformations (ϕ, ψ) and close non-covalent interatomic contacts in the final model.

Different types of superposition were carried out to assess the extent of structure variability in different lattice environments, the amount of saccharide-induced conformational change and any structural difference in the isoelectin binding sites that would explain the higher affinity of NeuLac for WGA1. The results are summarized in Tables 6 and 7. Comparison of the isoelectin complexes (all atoms) to one another in both crystallographic environments, yields values of r.m.s. Δr of the same order as is observed for the corresponding native structures (0.6 to 0.7 Å, see Table 6), and for superposition of molecules in different crystallographic environments (0.71 Å (NLI1) and 0.69 Å (NLI2), see Table 7). An important observation made in these comparisons is the higher degree of structure conservation in the sugar binding region (P-site; r.m.s. Δr = 0.29 to 0.55 Å) than is observed overall. Residues involved in sugar binding in this region in the two independent molecules are shown superimposed in Figure

Table 6
Comparison of isoelectin structures

Structures superimposed	r.m.s. displacement (Å)	
	Monomer I†	Monomer II†
A. NLI1/NLI2		
All atoms (res. 1–171 + NeuLac)	0.69	0.63
Main-chain atoms only	0.54	0.40
P-site region (151 atoms)	0.55	0.54
NeuNAc-Gal-Glc (43 atoms)	0.95	0.98
NeuNAc (21 atoms)	0.50	0.57
B. WGA1/WGA2‡		
All atoms (res. 1–171)	0.64	0.61
Main-chain atoms only	0.29	0.30
P-site region (145 atoms)	0.42	0.35

† Monomer I and II refer to the 2 independent molecules in the asymmetric unit.

‡ Values taken from Wright (1989).

Table 7
Positional differences in isoelectin structures

Structures superimposed	r.m.s. displacement (Å)	
	NLI1	NLI2
A. NeuLac-complex/native		
Monomer I (res. 1–171)†	0.34	0.36
P-site (169 atoms)	0.30	0.47
Monomer II (res. 1–171)	0.37	0.33
P-site region (169 atoms)	0.50	0.28
B. Monomer I/Monomer II		
All atoms (res. 1–171 + NeuLac)	0.71	0.69
P-site region (157 atoms)	0.42	0.44
NeuNAc-Gal-Glc (43 atoms)	0.89	0.81
NeuNAc (21 atoms)	0.61	0.37

† I and II refer to the 2 independent molecules in the asymmetric unit.

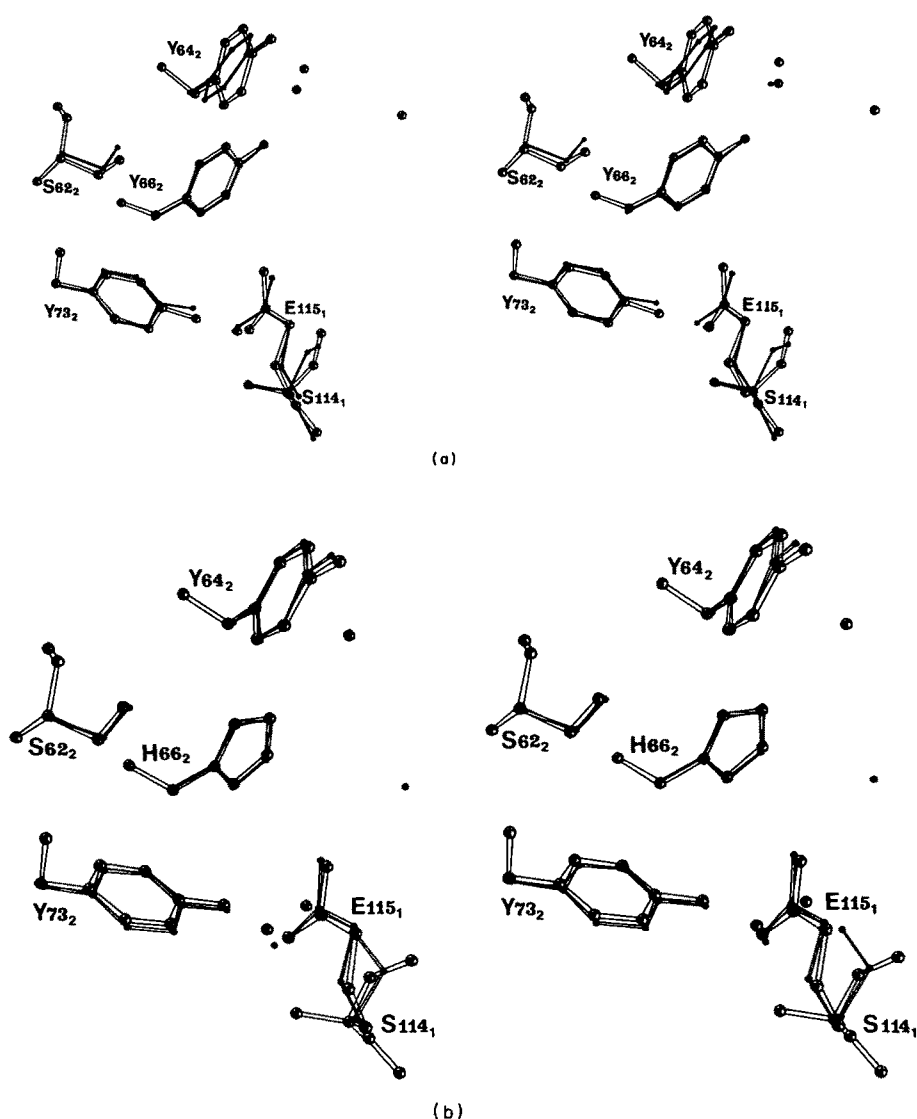


Figure 4. (a) Superposition of the residues in the "primary" sugar binding sites of independent molecules I (open bonds) and II (filled bonds) in the WGA1-NeuLac complex (NLI1). Water molecules are represented by large circles for dimer I and small circles for dimer II. (b) Superposition of the residues in the "primary" sugar binding sites of independent molecules I (wide bonds) and II (narrow bonds) in the WGA2-NeuLac complex (NLI2). Water molecules are represented as described for (a).

4(a) for NLI1 and (b) for NLI2. The values obtained for superposition of the trisaccharide are much higher (0.95 and 0.98 Å). As can be seen from Figure 5, the terminal NeuNAc residue superimposes best (r.m.s. $\Delta r = 0.5$ Å). The Gal-Glc portion superimposes poorly due to rotational freedom about the glycosidic bonds, yielding a discrepancy in the position of the glucose atoms of >1.0 Å in the crystallographically different environments.

Although there is a relatively large difference in the atomic temperature factors between the native structures and their NeuLac complexes, superposition of all the protein atoms indicates a very close match (0.33 to 0.36 Å, see Table 7). This is true for both independent molecules in both isolectins. When only atoms of the P-site were considered, some deviations are observed (0.28 to 0.50 Å, see

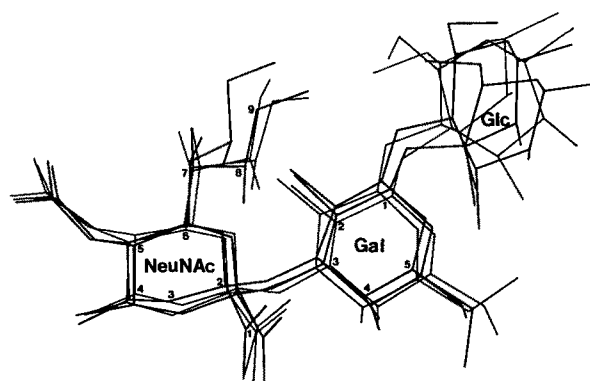


Figure 5. Superposition of the NeuLac trisaccharides in their bound configurations in the 2 independent molecules in NLI1 (thick bonds) and NLI2 (thin bonds).

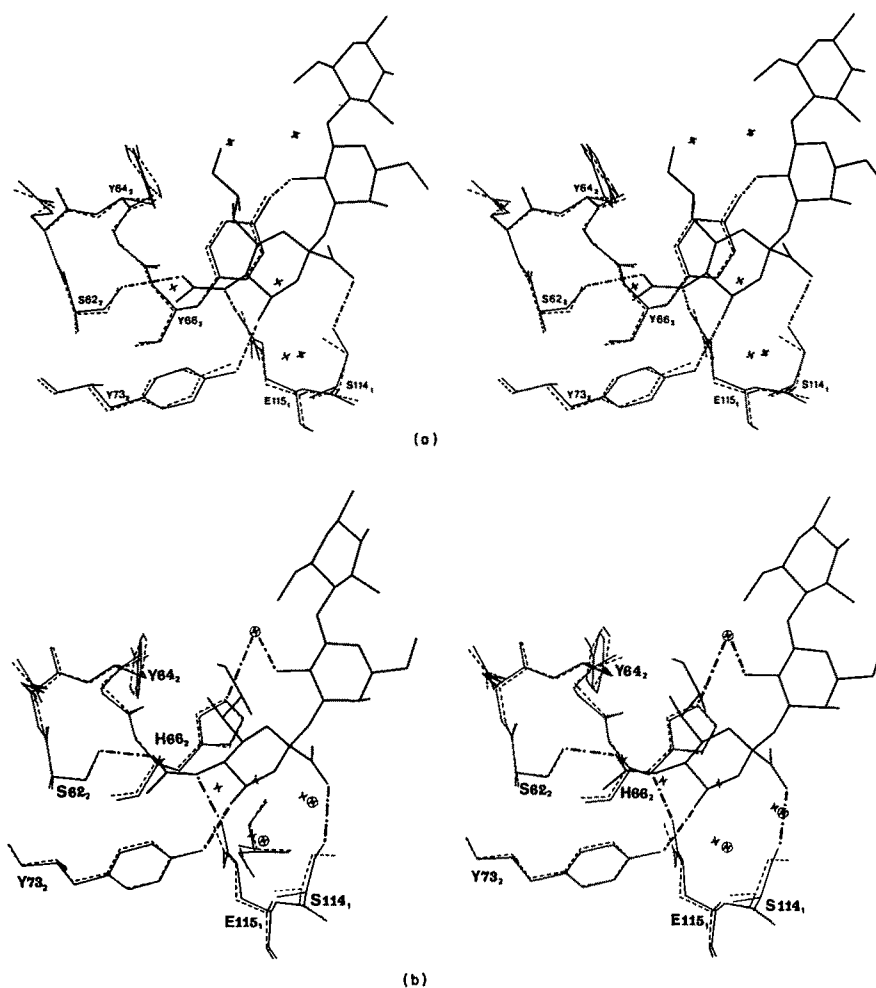


Figure 6. (a) Stereoscopic view of the "primary" binding site of WGA1 (independent molecule I) with NeuLac bound. The amino acids involved in sugar binding are depicted in their uncomplexed (broken bonds) and complexed conformations (filled bonds). The 5 hydrogen bonds that stabilize the complex are shown by lines of alternate dashes and dots. Water molecules present in the unoccupied binding site are depicted as 4-sided stars and those present in the complex as 6-sided stars. (b) Stereoscopic view of the primary binding site of WGA2 (independent molecule I) with NeuLac bound. Conditions are as described in (a) except that water molecules present in the complex are shown here as circles.

Table 7). The reason for this may be that sugar binding induces conformational changes in the critical side-chains to differing degrees in the two isolectins and in the two different environments. However, these Δr values are not out of line with those obtained for the superpositions of WGA1 and WGA2 and monomers I and II. The conformational changes observed are consistent in all environments and involve rotations only about side-chain bonds (χ_1) in Ser114, Ser62, Tyr64 and Tyr66 (see Fig. 6(a) and (b)). The largest changes are seen in Ser114 and Tyr64. Smaller, but real adjustments are observed for the side-chains of Ser62, Glu115 and the aromatic rings of Tyr73 and Tyr(His)66.

(e) Water association

All the water molecules observed in the native structure were included in the initial stage of refine-

ment of both complexes, refining an overall temperature factor at 3.0 Å and 2.5 Å resolution. The positions of all the water molecules were checked at this stage and in subsequent cycles difference electron density maps were examined for new sites. In the final list, 30 water molecules in NLI1 and 23 in NLI2 had been deleted and an additional 35 added in NLI1 and 60 in NLI2. The deleted water molecules tended to be weak and have high *B*-values, ranging when averaged for each independent environment from 37 to 42 Å² in NLI1 and NLI2. Five of these water molecules are located in the two sugar binding sites. A significantly larger number of water sites were located during the NLI2 refinement (235 total) as compared to that of NLI1 (214 total). Six of these new water molecules are associated with bound trisaccharide in NLI1 and four in NLI2. Examination of the regions of these new sites in the native electron density maps indi-

cated that electron density was present for a large number of them in WGA2 (33 of the 60) and a lesser number for WGA1 (7 of the 35). Several additional cycles of refinement were, therefore, carried out on the native models, including these additional water molecules. This reduced the final *R*-factor for WGA2 at 1.8 Å resolution from 17.9% to 17.5% ($F_o > 3\sigma$) and at 2.2 Å from 16.8% to 16.4%. In the case of WGA1, three to four refinement cycles carried out at 2.0 and 2.2 Å resolution ($F_o > 1\sigma$) yielded *R*-factors of 17.2% and 16.5%, respectively.

(g) *Protein-saccharide interactions*

Both polar and non-polar interactions are observed between NeuLac and six residues in the P-site. Table 8 provides a list of hydrogen bond lengths and van der Waals' contacts for the bound trisaccharide in the two crystallographic environments (Mol I and II). In Figure 7 these contacts are shown schematically using the WGA1 binding site as an example. There are four good hydrogen bonds formed between NeuNAc and four amino acid side-chains, and their distances are closely similar in the two environments and in both isolectins. The *N*-acetyl group makes the largest number of contacts contributing two H-bonds (CO to OH-Ser62 and NH to OE2-Glu115) and five to seven van der Waals' contacts with the phenyl ring of Tyr73, which provides a non-polar contact for the methyl group. The adjacent hydroxyl (C4-OH) is fixed by hydrogen bonds with the OH of Tyr73 and a well-ordered water molecule. The carboxylate is within

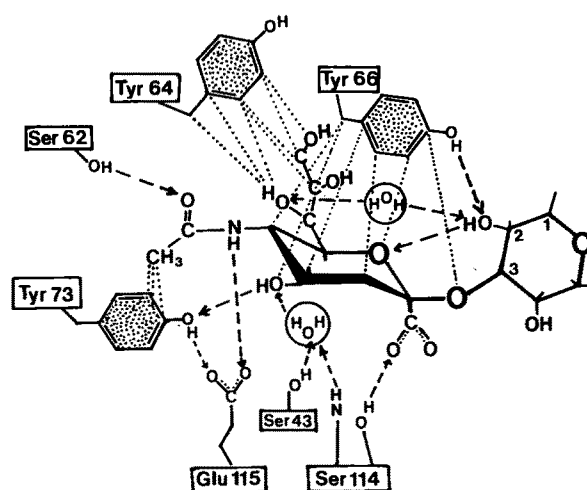


Figure 7. Schematic representation of the interactions between the NeuNAc-Gal portion of neuraminyl-lactose and specific amino acids in the WGA1 binding site. Hydrogen bonds are drawn as broken lines and potential van der Waals' contacts as dotted lines.

hydrogen-bonding distance of the OH of Ser114. This contact is very close in monomer II of NLI1, where a strong electron density connection is present.

In addition, numerous van der Waals' contacts stabilize the orientation of the sugar ring through non-polar stacking interactions with the aromatic side-chain of residue 66 (Tyr in WGA1 and His in WGA2). A third aromatic side-chain (Tyr64) is involved through non-polar contact with the glycerol tail of NeuNAc, which has different conformations and is solvated by ordered water to differing degrees in the different environments. This Tyr64 side-chain is flexible in each of the native structures and becomes immobilized in the presence of bound saccharide (see Table 9).

Only one water molecule, consistently found in both independent molecules in NLI1 and in NLI2, appears to play an important role in stabilizing the sugar complex. This water molecule solvates the hydroxyl group of Tyr73 in the unoccupied binding sites. It repositions slightly upon sugar binding (see

Table 8
Protein-ligand contacts in P-site

WGA residue	Ligand atom	WGA1 Mol I/II	WGA2 Mol I/II†
A. Hydrogen bonds (Å)			
NeuNAc			
Tyr73-OH	O4	3.12/2.49	2.57/2.64
Ser62-OH	CO (<i>N</i> -acetyl)	2.86/3.11	2.52/2.52
Glu115-COO ⁻	NH (<i>N</i> -acetyl)	2.68/2.70	2.92/3.19
Ser114-OH	O(COO ⁻)	2.67/2.45	2.67/2.52
Gal			
Tyr66	O2	2.70/2.64	
B. van der Waals' contacts (<4.5 Å)			
NeuNAc			
Tyr73 (ring)	CH ₃ -acetyl	5/5	7/7
Tyr64 (ring)	Glycolyl group	27/10	23/15
Tyr66 (ring)	Pyranoside ring	30-40	
His66 (ring)	Pyranoside ring		27-36
Gal			
Tyr66 (CZ, OH)	O3, O2	5	3
C. Water-mediated contacts (Å)			
Ser114 (NH)/			
Ser43(OH)/Wat-	O4(NeuNAc)	2.40/2.45	2.26/2.95
His66(NE2)/Wat-	O2(Gal)		2.53/—
Wat-	O2(Gal)	2.68/2.43	

† Mol I/II refers to the 2 independent monomers in the asymmetric unit.

Table 9
Average side-chain B-factors of residues in P-site

Amino acid	Side-chain temperature factor (Å ²)			
	WGA1 (I/II)	NLI1 (I/II)	WGA2 (I/II)	NLI2 (I/II)†
Ser62	13.5/16.4	10.9/13.3	16.7/16.0	14.2/11.3
Tyr64	31.3/30.7	17.7/16.8	33.0/34.7	22.8/20.7
Tyr(His)66	30.5/27.0	19.0/14.4	25.7/15.2	14.6/13.8
Tyr73	3.0/8.7	2.1/7.8	3.8/10.6	2.4/7.2
Ser114	18.3/21.4	11.5/10.0	19.9/20.3	14.4/15.9
Glu115	18.7/24.3	8.3/7.3	25.2/25.4	14.8/15.0

† I and II refer to the 2 independent molecules in the asymmetric unit.

Fig. 6(a) and (b)) and becomes immobilized, as suggested by the large drop in the B -values from 50.0 to 17.0 Å² in NLI1 and from 27.7 to 12.3 Å² in NLI2. In its new position it is tetrahedrally coordinated by hydrogen bonds with the C4-OH of NeuNAc, the backbone Ser114-NH, the Ser43-OH and either another water molecule or the Tyr73-OH.

As can be seen from Table 8, there is little contact between the galactose residue and the protein. However, the C2-OH group in both structures is involved in hydrogen bonding. In NLI1 it forms a hydrogen bond with the OH of Tyr66 and its orientation is further stabilized through a contact mediated by water with the glycerol C7-OH (see Fig. 6(a)). In NLI2 no direct hydrogen bond is possible with the imidazole side-chain of His66. However, in independent monomer I a water molecule is observed that could mediate contact with NE2 of His66 (see Fig. 6(b)).

4. Discussion

(a) Binding specificity

Refinement at high resolution supports the earlier low resolution results in that only the terminal non-reducing NeuNAc residue makes strong interactions with the lectin binding site. Although all of its ring substituents are in contact with protein side-chains, it is the *N*-acetyl group and the adjacent C4-OH that are the essential specificity determinants for the two types of sugar (GlcNAc and NeuNAc). They provide a cluster of three spatially close hydrogen bonds and a hydrophobic contact (acetamido-CH₃ with Tyr73 aromatic ring) in the least exposed part of the binding cavity, where the conformation of the protein is most stable (see Fig. 6(a) and (b)). The importance of such patches of interacting groups as stability-determining factors in protein-sugar complexes has been discussed by Lemieux (1989) for the Lewis B antigen (tetrasaccharide) interaction with *Griffonia simplicifolia* lectin. In the case of WGA, the most important amino acid residue for sugar binding, Tyr73, is constrained in the proper conformation in the native structure (B -factor

below 10 Å²) such that only a very slight B -value reduction is observed upon complex formation (see Table 9). Its side-chain is fixed by a hydrogen bond with the carboxylate of Glu115 (OE1) of the dimer-related subunit. This carboxylate group is drastically immobilized upon sugar binding (B -value reduction of 10 Å², see Table 9) when the plane of the carboxylate group becomes fixed by formation of a hydrogen bond between the other carboxylate oxygen (OE2) and the acetamido-NH- of NeuNAc.

Lemieux (1989) also maintains that more flexible residues, making contact with the saccharide at the periphery of the binding site, are important only for guiding the sugar moiety to the correct location in the initial binding step. Support for this proposal is found in the complexes studied here: both the carboxylate and glycerol groups of NeuNAc interact with the side-chains of two amino acid residues (Ser114 and Tyr64) that show comparatively large conformational changes, but are less important for binding. However, as in the case of influenza hemagglutinin (HA), the carboxylate group must be axial to the C2 ring carbon in order to interact with Ser114. This explains the anomeric specificity of WGA for α -NeuNAc in structural terms (Wright, 1980a). The importance for binding of the flexible hydrophilic glycerol tail, which is in van der Waals' contact with Tyr64, is questionable in the light of the modification studies of Peters *et al.* (1979). These authors demonstrated that binding of NeuNAc could be improved by trimming back the hydrophilic glycerol tail to the C7 atom. This finding contrasts with the situation in HA, where this glycerol group is necessary for stabilization of the NeuNAc-HA complex by providing two hydrogen bond contacts to protein side-chains (Weis *et al.*, 1988).

Comparing the NeuLac binding mode to that of GlcNAc-oligosaccharide (Wright, 1980b), the position of the glycerol group coincides with that of subsite 2 of GlcNAc- β (1-4)-GlcNAc. The GlcNAc residue bound in this site is oriented with its non-polar face roughly parallel to the plane of the Tyr64 side-chain (see Fig. 8). The larger number of van der Waals' contacts possible between the aromatic ring

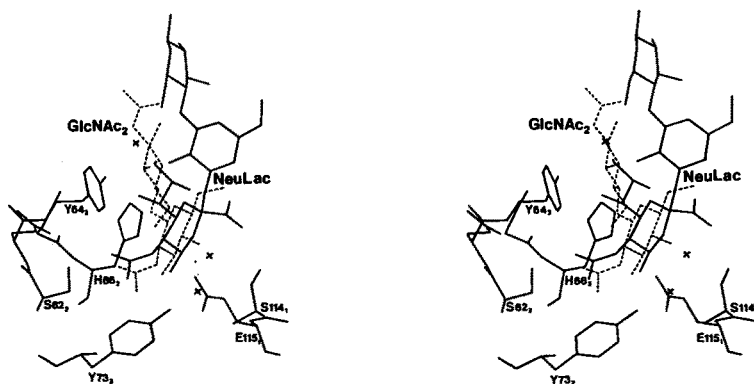


Figure 8. Stereoscopic view of the P-site of WGA2 with the NeuLac trisaccharide bound and the GlcNAc-GlcNAc disaccharide shown superimposed in broken outline. The position of GlcNAc-GlcNAc is as described by Wright (1984).

and the GlcNAc sugar ring compared to those of the glycerol group of NeuNAc explains the higher binding constant observed for GlcNAc oligomers. Subsite 2 in NeuNAc oligosaccharides is oriented away from the protein surface. On the other hand, GlcNAc in subsite 1 lacks a polar group in a position equivalent to the carboxylate of NeuNAc, and thus is stabilized by only three hydrogen bonds. Model building has shown that large substituents at C4 of GlcNAc prohibit binding in subsite 1 due to steric crowding. This confirms the specificity of WGA for terminal non-reducing residues of $\beta(1-4)$ -linked GlcNAc oligomers.

Binding studies carried out by Kronis & Carver (1982) using n.m.r. techniques demonstrated that both WGA isolectins are able to discriminate in their specificity for glycosidic linkages. WGA1 and WGA2 bind the $\alpha(2-3)$ isomer of NeuLac with a higher affinity than the $\alpha(2-6)$ isomer by factors of 7 and 3, respectively. Model building shows that in the complex with the $\alpha(2-6)$ -linked trisaccharide there is less contact possible between the Gal-Glc portion and the protein surface due to rotational freedom and extension of the distance between NeuNAc and Gal by an extra bond. Although the dissociation constants measured by n.m.r. for this isomer are very similar for WGA1 and WGA2 (9.5 and 10.6 mM; Kronis & Carver, 1982), the larger difference in affinity between the $\alpha(2-3)$ and $\alpha(2-6)$ isomers for WGA1 is probably due to loss of the hydrogen bond between Tyr66(OH) and Gal(C2-OH). A definitive understanding of the preference of WGA1 for either linkage will have to await results of the crystal structure determination of the complex of WGA1 and the glycoprotein tryptic sialoglycopeptide T5 (Wright & Kahane, 1987).

In the case of HA, discrimination between the two isomers depends on the identity of residue 226 (Leu or Gln). A higher specificity for $\alpha(2-6)$ over $\alpha(2-3)$ -linked sialic acid is observed in mutants containing Leu226, whereas the reverse is true for mutants with Gln226 (Rogers *et al.*, 1983; Sauter *et al.*, 1989). However, this does not have an obvious structural explanation (Weis *et al.*, 1988). In addition, side-group binding specificity of NeuNAc in HA differs in several other respects from that observed in WGA. Hydrogen bond contacts involving the acetamido CO- and the C4-OH groups, important in WGA, are not present. Stabilization of the acetamido group comes only from a non-polar contact of its methyl group with an aromatic side-chain (Trp153) as in WGA, and a hydrogen bond contact involving its amino group (Weis *et al.*, 1988; Sauter *et al.*, 1989).

In contrast to other sugar-protein complexes studied by X-ray diffraction, the sialic acid binding sites in WGA and HA are relatively shallow. For instance, the L-arabinose binding protein-saccharide complex, refined at 1.7 Å resolution (Quiocho, 1986; Quiocho & Vyas, 1984), shows the arabinose ligand bound in a deep pocket completely surrounded by amino acid ligands that interact with

all available hydroxyl groups. In the trisaccharide complex of lysozyme-(GlcNAc)₃, refined at 1.75 Å resolution, only one of the three subsites (site C) provides extensive binding, analogous to that seen in WGA. In this site, binding of the *N*-acetyl group plays an important role, but not all of the free hydroxyl groups of the sugar ring form hydrogen bonds (Johnson *et al.*, 1988).

Consideration of the energetics of ligand binding suggests that hydrogen bonding and van der Waals' stacking interactions of aromatic rings, as observed in the WGA-NeuLac complexes, should make negative contributions to ΔH and ΔS , as predicted by Ross & Subramanian (1981). This has been verified by n.m.r. analysis of the temperature dependence of binding of this trisaccharide to WGA1 (Kronis & Carver, 1985b). The structural results presented here suggest that the enthalpy and entropy values obtained by Kronis & Carver ($\Delta H^\circ = -13.3$ kcal/mol and $\Delta S^\circ = -31.9$ cal/mol K⁻¹) arise from contributions due to intrinsic binding of the ligand only, since no drastic conformational changes were observed (see below).

(b) Saccharide-induced conformational changes

Refinement has now allowed assessment of saccharide induced conformational changes. This question could not be addressed in earlier low-resolution work (Wright, 1980, 1984) due to inaccuracy in the atomic positions as well as the fact that the electron density map was averaged over the two independent molecules in the asymmetric unit. Since there are differences in the two environments in the side-chain conformations of some of the crucial residues (see Fig. 4(a) and (b)), which are of the same order as those observed upon sugar binding, each independent binding site (P-site) was examined separately.

Superposition of the native and complex structures shows only small conformational changes in all the amino acid side-chains in contact with the trisaccharide (r.m.s. $\Delta r = 0.3$ to 0.5 Å, see Table 7). Although these conformational differences are of the order of errors expected in the co-ordinates, they may be considered as real on the basis of the consistency with which they are found in the different crystallographic environments. It is likely that these small changes optimize the fit in terms of good hydrogen bonding geometry and maximal van der Waals' contact to produce the energetically lowest complexed structure. The general trend is that side-chains with high *B*-factors, which are exposed in the native and become shielded by the ligand in the complex, exhibit larger conformational changes (Tyr64 and Ser114). In Figure 6(a) and (b), NeuLac is shown bound in the P-site of monomer I in the two complexes. A common feature observed between the two binding sites, and also in the second independent environment, is the twist of the Tyr64 aromatic ring, which can move either in the direction of the glycerol group or away from it, depending on the conformation of this hydrophilic

group, which differs in the two isolectin complexes. In addition, a substantial rotation of the Ser114 side-chain (as large as 90°) is observed in all four binding locations at the edge of the binding cavity. This adjustment results in a good hydrogen bond contact with the NeuNAc carboxylate, which is also flexible. The other residues in contact with saccharide undergo smaller movements. For instance, the aromatic ring of residue 66 moves in different directions in the two complexes. In WGA1 the Tyr side-chain is pushed back slightly by NeuNAc binding, while in WGA2 the His side-chain moves slightly closer. In contrast, the aromatic ring of Tyr73 moves slightly in a direction away from the saccharide in both complexes, and the plane of the carboxylate of Glu115 adjusts accordingly to maintain good hydrogen bond contact with the Tyr73-OH and to form a hydrogen bond with the acetamido -NH- of NeuNAc. Similar observations were made in the case of the lysozyme trisaccharide complex, where the most drastic conformational change is seen for Trp62, which moves toward the GlcNAc residue in subsite B and thereby becomes immobilized (Johnson *et al.*, 1988).

In the WGA-saccharide complexes no significant conformational changes are observed in regions remote from the P-site, nor are there any movements of chain segments, such as occur in lysozyme and possibly in L-arabinose binding protein. In the latter, a hinge bending motion between two domains has been suggested to control the opening and closing of the sugar binding cavity (Quirocho, 1988). However, this suggestion is not based on crystal structure data, since the native crystal structure is not available.

The superposition results described above are in agreement with interpretations made of difference density in the immediate vicinity of the ligand in $F(\text{complex}) - F(\text{native})$ maps. A large number of additional low-level difference features were observed in these maps regardless of the type of scaling used for the F_o values (1 or 2 parameter scaling). These noise features were not interpretable and coincided mostly with solvent regions

surrounding the trisaccharide and unoccupied spaces in the subunit/subunit interface.

(c) Comparison of binding modes in WGA1 and WGA2

To assess the differences in the binding mode of NeuLac in the WGA1 and WGA2 complexes, both crystallographic environments were examined in each of the NeuLac complexes. Superposition of the trisaccharides gave an estimated error of 0.5 to 0.6 Å in the atomic positions of the NeuNAc residue in the two distinct environments (see Table 6). Conformational flexibility of the trisaccharide is due to positional disorder and freedom of rotation about the glycosidic bonds. Nevertheless, comparison of hydrogen bond distances in the different environments shows good agreement (see Table 8). The binding sites of the two complexes are shown superimposed in Figure 9. In NLI2 the NeuNAc acetamido group binds slightly deeper into the binding site in both environments (see also Fig. 5), since no constraints are imposed by the side-chain of His64 as there are by Tyr66 (hydrogen bond to C2-OH of Gal) in NLI1. Binding constants (K_d) of this oligosaccharide measured by n.m.r. for WGA1 and WGA2 indicate a difference in affinity of approximately a factor of 5 (0.74 mM and 3.6 mM). The X-ray results show that replacement of Tyr66 in WGA1 by His in WGA2 is the major structural difference. This substitution leads to fewer van der Waals' contacts and one less hydrogen bond contact in the WGA2 complex (see Table 8), consistent with the reduced affinity observed by n.m.r. The Tyr66 side-chain, which adjusts slightly away from the NeuNAc sugar ring upon sugar binding in NLI1, also appears to be responsible for the conformational difference in the glycerol group. The C7-OH is pulled toward the Gal residue so that it can form a hydrogen bond with O2 through a water molecule.

Although the NLI2 complex refined much better by statistical criteria, the slightly lower atomic occupancies of the saccharide atoms and less strong density for the Gal residue suggest that binding in

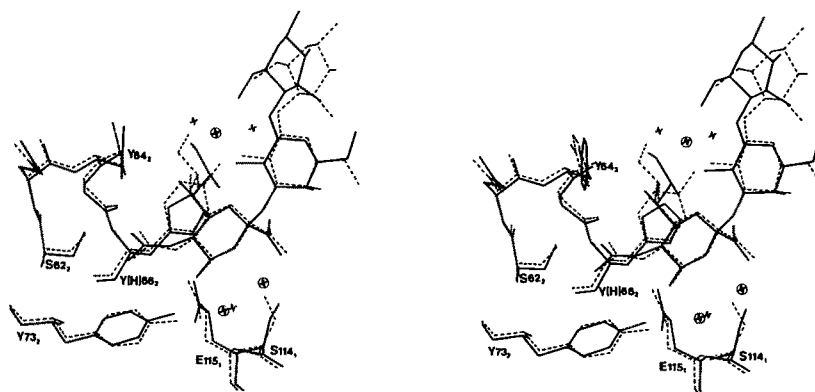


Figure 9. Superposition of the 2 isolectin trisaccharide complexes. The NeuLac trisaccharide is shown bound at the P-site in the NLI1 complex (broken bonds) and in NLI2 complex (continuous bonds). Water molecules present in the NLI1 complex are represented by crosses and those present in the NLI2 complex by circles.

the crystal is weaker than in WGA1. However, when occupancies are fixed at 1.0, atomic *B*-values for the trisaccharide of NL12 are comparable or even somewhat lower than those observed in the NL11 complex.

(d) *Importance of aromatic groups for stabilization of the protein-saccharide complexes*

The frequency with which aromatic amino acids are found in sugar binding sites suggests a significant function. The planar surfaces are able to provide patches of van der Waals' contact with non-polar surfaces or aliphatic groups of the sugar molecule. The larger the number of contacts possible, the greater is its effect on the stability of the complex. The kind of interaction most commonly observed is that of planar stacking with the pyranoside ring (see Table 10). This would serve to exclude water and completely shield one side of the sugar ring from surrounding solvent. In WGA five aromatic side-chains are directly involved in disaccharide binding and ten of the 13 aromatic amino acid residues of the molecule are located in the immediate vicinity of the two independent sugar binding sites (see Fig. 10). However, only two of the three aromatic amino acid residues in the P-site are also present in the inaccessible S-site (Tyr159 and Trp150; see

Table 10
Aromatic side-chains in sugar binding sites

Protein (sugar ligand)	Type of interaction	
	Planar contact	H-bond contact
WGA (GlcNAc, NeuNAc, NeuLAc)	Tyr64, Tyr(His)66, Tyr73, Tyr159, Trp150	Tyr66, Tyr73, Tyr159
Hemagglutinin (NeuNAc, NeuLAc)	Trp153	Tyr98
Pea lectin (Man, Glc)	Phe123	
ConA (Man, Glc)	Tyr123	
L-Arabinose binding protein (L-arabinose)	Trp16, Phe17	
Lysozyme (β 1,4(GlcNAc) ₃)	Trp62	Trp62, Trp63
Taka-amylase (maltoheptaose)		Tyr79, Trp83, His296, His122, His210, Tyr155

ConA, concanavalin A.

Fig. 10). The amino acid in the S-site corresponding to residue 66 (Tyr/His) is a serine (Ser152). This suggests that the binding affinity would not be as strong due to the smaller number of van der Waals' contacts of a serine hydroxyl group with the sugar

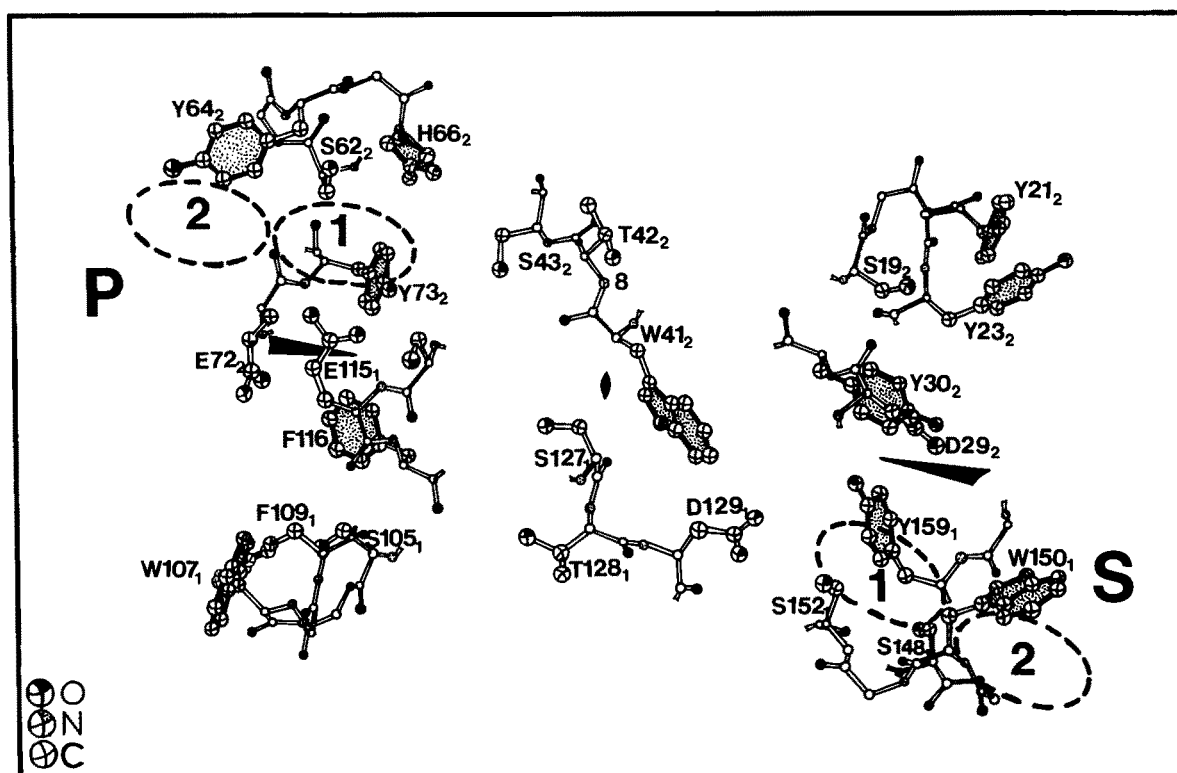


Figure 10. Distribution of selected amino acid residues spanning the 2 unique sugar binding sites (P-site and S-site) in the C_1D_1/A_2B_2 subunit/subunit contact region of WGA2. This Figure is similar to Fig. 6 of Wright (1984), but contains the amino acid corrections of 3 aromatic residues (W41, W150 and F109). Aromatic side-chains are emphasized by shading. The 2 quasi 2-fold axes relating sugar binding domains of opposite monomers are depicted by thick black arrows, and the quasi 2-fold axis relating half-monomers and the P-site to the S-site is marked in the center of the Figure. Subscripts 1 and 2 refer to 2 identical subunits of the dimer. Areas in broken outline indicate the approximate regions of subsites 1 and 2 in the GlcNAc-GlcNAc complex of WGA2.

ring compared to those of an aromatic group. Structure analysis of orthorhombic crystals of the WGA1-T5 complex (Wright & Kahane, 1987), currently in progress, should provide the remaining details of the binding mode at this site.

In lysozyme, Trp62 is observed to interact through van der Waals' stacking with the sugar ring of the GlcNAc residue in subsite B, this being the only contact of this residue with the protein surface. Two aromatic amino acid residues, Trp16 and Phe17, are observed to flank the saccharide in the L-arabinose binding protein complex (Quioco & Vyas, 1984). In the sialyl-lactose-hemagglutinin complex, the *N*-acetyl-methyl group is in contact with an aromatic side-chain (Trp153) (Weis *et al.*, 1988), analogous to WGA. The Man/Glc specific lectins concanavalin, pea and lentil lectin, also have an aromatic ring (Tyr or Phe) coplanar with the sugar ring of the non-reducing terminal Man of bound oligosaccharides (Rini, 1986).

(e) Mobility

The magnitude of atomic *B*-factors are an indication of how well determined the refined positions of the atoms in a crystal structure are. Since this factor includes contributions of lattice disorder, static disorder, conformational variability of side-chains in different molecules as well as mobility due to thermal vibrations, it is not straightforward to interpret the lower level of atomic *B*-factors observed here for the trisaccharide complexes. It seems unlikely that this decrease is an artifact of refinement or of the data, since, in the case of NLI2, refinement proceeded smoothly and this dataset was more complete than those of NLI1 and the native crystals (see Table 3). However, the 30% reduction of the overall isotropic *B*-factors of both complexes (see Table 3) indicates a higher degree of order in the complexed *versus* uncomplexed crystal lattices.

For proper behavior of the *B*-values, high-resolution data (<2.0 Å) is required. A large effect due to resolution can, however, be ruled out in these refinements, since in the case of the native structures no significant reduction in the average *B*-values was observed when the resolution was reduced from 1.8 Å to 2.2 Å. Since the largest effect is observed for parts of the structure stabilized either by protein-protein or protein-sugar interactions, this indicates a tightening of the protein in the presence of saccharide. Spectroscopic measurements on other macromolecules (Eftink & Biltonen, 1980) support this suggestion. For instance, techniques such as infrared and Raman spectroscopy, and n.m.r. relaxation studies suggest that ligand binding affects primarily low-frequency internal vibrational modes (torsional bond rotation and backbone bending). Such loss of rotational freedom may not occur only in residues in the ligand binding site but possibly throughout the molecule. Calorimetric measurements indicate that ligand binding causes a decrease in the heat capacity of macromolecules. Eftink & Biltonen (1980) have

discussed the possibility that ligand binding shifts the equilibrium distribution of the microstates of a macromolecule towards lower enthalpy microstates, which may represent "tighter" conformations. This interpretation would agree with the notion that "freezing out" of rotational modes of parts of a structure lowers their ability to absorb energy, thus lowering the heat capacity. Moreover, on the basis of hydrogen-deuterium exchange measurements, it has been suggested that enzymes become more rigid in the presence of inhibitor (Petsko & Ringe, 1984).

The possibility may also be considered that the ligand complex differs in its external vibrational modes with respect to its native structure. This would, for instance, involve differences in the concerted motion of closely associated molecules in the lattice and exhibit an overall effect on the atomic *B*-values. An analysis of the temperature factor pattern at 2.0 Å resolution in the case of lysozyme could be interpreted as rigid body motion of strongly interacting molecules (Sternberg *et al.*, 1979). However, lattice effects may be ruled out on the basis of another study with lysozyme that suggested that intrinsic vibrational characteristics of proteins remain the same in different crystal lattices. Comparing the refined structures of tetragonal and orthorhombic lysozyme, only atoms in regions involved in intermolecular lattice contacts showed significant mean-square displacements (Petsko & Ringe, 1984).

In summary, the significant reduction in the observed *B*-values of the complexed structures should be viewed in a qualitative sense only, since this may be the result of a combination of several effects. (1) Lack of data at sufficiently high resolution can lower *B*-values, even to the extent of becoming negative (Baker & Dodson, 1980). However, since *B*-values obtained for the native structures at identical resolution were much higher, this effect may not play a major role. (2) Replacement of the low ionic strength crystallization medium of native crystals (20 mM-sodium acetate, 5 mM-CaCl₂) with one containing trisaccharide (4 mM) may have a stabilizing effect on the lattice through non-specific interactions and may affect rigid body motion of the entire protein molecule (Sternberg *et al.*, 1979). (3) Saccharide binding in two equivalent binding sites, which lie between subunits (domains B1/C2 and B2/C1), tightens the dimer structure or reduces rotational freedom of residues involved in dimerization.

I thank H. Tonie Wright for critical reading of the manuscript. This work was supported by the National Institutes of Health research grant AI-17992.

References

- Allen, A. K., Neuberger, A. & Sharon, N. (1973). *Biochem. J.* **131**, 155-162.
- Baker, E. N. & Dodson, E. J. (1980). *Acta Crystallogr. sect. A*, **36**, 559-572.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rogers, J. R., Kennard, O.,

- Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
- Bhavanadan, V. P. & Katlic, A. W. (1979). *J. Biol. Chem.* **254**, 4000–4008.
- Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Spohr, U. & Lemieux, R. U. (1989). *Amer. Crystallogr. Assoc. Abstracts*, **17**, V5.
- Eftink, M. & Biltonen, R. (1980). In *Biological Microcalorimetry* (Beezer, A. E., ed.), pp. 343–411, Academic Press, London, New York, Toronto, Sydney and San Francisco.
- Etzler, M. E. (1985). *Annu. Rev. Plant Physiol.* **36**, 209–234.
- Flippin, J. L. (1973). *Acta Crystallogr. sect. B*, **29**, 1881–1886.
- Goldstein, I. J. & Hayes, C. E. (1978). In *Advances in Carbohydrate Chemistry and Biochemistry* (Tipson, R. S. & Horton, D., eds), vol. 35, pp. 127–340, Academic Press, New York, San Francisco and London.
- Goldstein, I. J. & Poretz, R. D. (1986). In *The Lectins: Properties, Function and Application in Biology and Medicine* (Liener, I. E., Sharon, N. & Goldstein, I. J., eds), pp. 43–247, Academic Press, Orlando, FL.
- Hendrickson, W. A. & Konnert, J. H. In *Biomolecular Structure, Conformation, Function, and Evolution* (Srinivasan, R., ed.), vol. 1, pp. 43–47, Pergamon Press, Oxford.
- Johnson, L. N., Cheetham, J., McLaughlin, P. J., Acharya, K. R., Barford, D. & Phillips, D. C. (1988). *Curr. Top. Microbiol. Immunol.* **139**, 81–128.
- Jones, T. A. (1978). *J. Appl. Crystallogr.* **11**, 268–272.
- Jordan, F., Bassett, E. & Redwood, W. R. (1977). *Biochem. Biophys. Res. Commun.* **75**, 1015–1021.
- Kronis, K. A. & Carver, J. P. (1982). *Biochemistry*, **21**, 3050–3057.
- Kronis, K. A. & Carver, J. P. (1985a). *Biochemistry*, **24**, 826–833.
- Kronis, K. A. & Carver, J. P. (1985b). *Biochemistry*, **24**, 834–840.
- Lemieux, R. U. (1989). *Chem. Soc. Rev.* **18**, 347–374.
- Longchambon, P. F., Ohanessian, J., Gillier-Pandraud, H., Duchet, D., Jacquinet, J. C. & Sinay, P. (1981). *Acta Crystallogr. sect. B*, **37**, 601–607.
- Luzzati, V. P. (1952). *Acta Crystallogr.* **5**, 802–810.
- Monsigny, M., Roche, A.-C., Sene, C., Magnet-Dana, R. & Delmotte, F. (1980). *Eur. J. Biochem.* **104**, 147–153.
- Nagata, Y. & Burger, M. M. (1974). *J. Mol. Chem.* **249**, 3116–3122.
- Peters, B. P., Ebisu, S., Goldstein, I. J. & Flashner, M. (1979). *Biochemistry*, **18**, 5505–5511.
- Petsko, G. A. & Ringe, D. (1984). *Annu. Rev. Biophys. Bioeng.* **13**, 331–371.
- Privat, J.-P., Delmotte, F. & Monsigny, M. (1974a). *FEBS Letters*, **46**, 224–228.
- Privat, J.-P., Delmotte, F., Mialonier, G., Bouchard, P. & Monsigny, M. (1974b). *Eur. J. Biochem.* **47**, 5–14.
- Quioco, F. (1986). *Annu. Rev. Biochem.* **55**, 287–315.
- Quioco, F. (1988). *Curr. Top. Microbiol. Immunol.* **139**, 135–148.
- Quioco, F. A. & Vyas, N. K. (1984). *Nature (London)*, **310**, 381–386.
- Reeke, G. N., Jr & Becker, J. W. (1988). *Curr. Top. Microbiol. Immunol.* **139**, 35–58.
- Rice, R. H. & Etzler, M. E. (1975). *Biochemistry*, **14**, 4093–4099.
- Richards, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.
- Rini, J. M. (1986). Ph.D. thesis, University of Toronto.
- Rogers, G. N., Paulson, J. C., Daniels, R. S., Skehel, J. J., Wilson, I. A. & Wiley, D. C. (1983). *Nature (London)*, **304**, 76–78.
- Ross, P. D. & Subramanian, S. (1981). *Biochemistry*, **20**, 3096–3102.
- Sauter, N. K., Bednarski, M. D., Wurzburg, B. A., Hanson, J. E., Whitesides, G. M., Skehel, J. J. & Wiley, D. C. (1989). *Biochemistry*, **28**, 8388–8396.
- Sternberg, M. J. E., Grace, D. E. P. & Phillips, D. C. (1979). *J. Mol. Biol.* **130**, 231–253.
- Swanson, S. M. (1988). *Acta Crystallogr. sect. A*, **44**, 437–442.
- Thomas, M. W., Walborg, E. F. & Jirgensons, B. (1977). *Arch. Biochem. Biophys.* **178**, 625–630.
- Weis, W., Brown, J. H., Cusack, S., Paulson, J. C., Skehel, J. J. & Wiley, D. C. (1988). *Nature (London)*, **333**, 426–431.
- Wright, C. S. (1977). *J. Mol. Biol.* **111**, 439–457.
- Wright, C. S. (1980a). *J. Mol. Biol.* **139**, 53–60.
- Wright, C. S. (1980b). *J. Mol. Biol.* **141**, 267–291.
- Wright, C. S. (1984). *J. Mol. Biol.* **178**, 91–104.
- Wright, C. S. (1987). *J. Mol. Biol.* **194**, 501–529.
- Wright, C. S. (1989). *J. Mol. Biol.* **209**, 475–487.
- Wright, C. S. & Kahane, I. (1987). *J. Mol. Biol.* **194**, 353–355.
- Wright, C. S. & Raikhel, N. (1989). *J. Mol. Evol.* **28**, 327–336.