

Comparison of the Refined Crystal Structures of Two Wheat Germ Isolectins

Christine Schubert Wright

Department of Biochemistry and Molecular Biophysics

MCV/VCU

Box 614 MCV Station

Richmond, VA 23298-0001, U.S.A.

(Received 3 January 1989, and in revised form 10 May 1989)

The crystal structures of two closely related members of the multigene family of wheat lectins (isoelectins 1 and 2) have been compared. These isoelectins differ at five sequence positions, one being located in the saccharide binding site modulating ligand affinity. Crystals of the two isoelectins are closely isomorphous (space group $C2$). The atomic models are based on structure refinement at 1.8 Å resolution in the case of isoelectin 2 (WGA2) and 2.0 Å resolution in the case of isoelectin 1 (WGA1). Refinement results for WGA1, recently completed with a crystallographic R -factor of 16.5% ($F_o > 3\sigma(F_o)$), are presented.

Examination of a difference Fourier map, $[F_{WGA2} - F_{WGA1}]$, at 2.0 Å resolution and direct superposition of the two models indicated an overall close match of the two structures. Local differences are observed in the region of residues 44 to 69, where three sequence differences occur, and at highly mobile external residues on the surface. The average positional discrepancy (root-mean-square Δr) for corresponding protein atoms in the two crystal structures is 0.64 Å for independent protomer I and 0.61 Å for protomer II (0.29 Å and 0.30 Å for main-chain atoms). The mean atomic temperature factors are very similar 20.9 versus 22.0 Å². Regions of high flexibility coincide in the two isoelectin structures. Of the 210 water sites identified in WGA1, 144 have corresponding positions in WGA2. A set of 51 well-ordered sites was found to be identical in the two independent environments in both structures, and was considered to be important for structure stabilization.

Both of the unique sugar binding sites superimpose very closely, exhibiting root-mean-square positional differences ranging from 0.29 Å to 0.42 Å. The side-chains of the critical tyrosine residues, Tyr73 (P-site) and Tyr159 (S-site), superimpose best, while other highly flexible aromatic groups (Tyr64 and Trp150) and several water sites display large differences in position (0.5 to 1.0 Å) and high temperature factors. The aromatic side-chains of Tyr66 in WGA1 and His66 in WGA2 are oriented similarly.

1. Introduction

Wheat germ agglutinin (WGA†) belongs to a superfamily of evolutionarily conserved lectins that share highly stable, disulfide-rich structures and sugar specificity for *N*-acetyl-D-glucosamine. The majority of these lectins characterized to date belong to the Gramineae family (Stinissen *et al.*, 1983a; Strossberg *et al.*, 1986). Three genetic variants of WGA (isoelectins) are known to exist in commercial wheat germ preparations (Allen *et al.*, 1973; Ewart, 1975; Rice & Etzler, 1975). This heterogeneity is attributed to the presence of three

separate genomes in hexaploid wheat (*Triticum aestivum*) (Rice, 1976; Peumans *et al.*, 1982; Stinissen *et al.*, 1983b). The biological and molecular properties of these wheat lectins have been extensively documented (for reviews, see Goldstein & Hayes, 1978; Etzler, 1985; Lis & Sharon, 1986).

The primary structures of all three isoelectins have been determined. WGA1 and WGA2 were sequenced as mature proteins of 171 amino acid residues (Wright *et al.*, 1984; Wright & Olafsdottir, 1986). Sequence information for isoelectin 3 (WGA3) was derived from a cDNA clone coding for a 186 residue precursor molecule (Raikhel & Wilkins, 1987). This pro-WGA, which contains a glycosylated 15 residue extension at the carboxyl terminus, exists in the endoplasmic reticulum and has sugar binding capability. Post-translational processing

† Abbreviations used: WGA, wheat germ agglutinin; r.m.s., root-mean-square; P-site, primary site; S-site, secondary site.

removes this extension and yields the mature 171 residue molecule (Mansfield *et al.*, 1988).

In a comparison of the three isolectin sequences (Wright & Raikhel, 1989), the most striking feature of WGA1 is its total lack of histidine residues. The two histidine residues present in WGA2 and WGA3 (nos 59 and 66) play a structural and functional role. A total of nine variable sequence positions has been identified among the three sequences. The sequence of WGA3 differs in a larger number of positions from those of WGA1 and WGA2 (8 and 7 positions, respectively) than do the latter two from one another (5 sequence positions). In addition, these sequences can be subdivided into four very similar 43 residue segments (Wright *et al.*, 1984) that correspond to the discretely folded isostructural regions, observed in earlier crystal structure studies and termed domains A, B, C and D (Wright, 1977, 1980a). On the basis of this internal redundancy it was proposed that these highly homologous wheat lectins evolved from a common single domain ancestor (Wright *et al.*, 1985).

Crystallographic studies were carried out initially on crystals containing all three isolectins (Wright, 1977). Later, the isolectins were crystallized separately and structural differences determined between the two most abundant isolectins (WGA1, WGA2) (Wright, 1981; Wright & Olafsdottir, 1986), because they display distinctly different sugar binding affinities for *N*-acetyl-neuraminyllactose (Kronis & Carver, 1982) and electrophoretic mobilities (Allen *et al.*, 1973; Rice & Etzler, 1975). However, their ability to undergo subunit interchange, forming stable hybrid dimers (Rice & Etzler, 1975; Peumans *et al.*, 1982), suggests very similar three-dimensional structures.

The atomic structure of WGA2 was the first to be crystallographically refined and analyzed in detail (Wright, 1987). To be able to pinpoint small structural differences between these two isolectins, refinement was carried out on WGA1 using data to resolution limits (2.0 Å (1 Å = 0.1 nm) resolution). The resulting model was analyzed for its overall similarity to the refined WGA2 model, the location of associated ordered water sites, and for differences in specific regions where mutations have occurred and sugar ligands bind.

2. Experimental Procedures

The WGA isolectins were isolated and crystallized as described (Bassett, 1975; Lacelle, 1979; Wright, 1981). The crystals are closely isomorphous and belong to space group *C*2 (see Table 1). Intensity data for WGA1 were extended from 2.4 to 2.0 Å on a Nonius CAD-4 diffractometer equipped with a Philips Cu-fine-focus X-ray tube and interfaced with a PDP-11/23 computer. Reflections were scanned and absorption curves applied as described for WGA2 (Wright, 1987). Data reduction was carried out on an IBM 3081 using the 'ROCKS' program library (G.N. Reeke Jr). These new data were merged with an existing 2.4 Å data set collected earlier on a rotating anode X-ray source (Wright, 1981). The final merged 2.0 Å data set (50 to 2.0 Å) consisted of 29,266 total

Table 1
Crystal parameters for WGA1 and WGA2

	WGA1	WGA2
<i>A. Unit cell parameters</i>		
<i>a</i> (Å)	51.22	51.34
<i>b</i> (Å)	73.60	73.53
<i>c</i> (Å)	91.42	91.54
α (°)	90.00	90.00
β (°)	98.00	97.59
γ (°)	90.00	90.00
<i>B. Position of local 2-fold axis</i>		
<i>x</i> -translation (Å)	6.30	6.055
<i>y</i> -position (Å)	−0.028	−0.036
<i>z</i> -position (Å)	22.69	22.68

measurements for 22,549 unique lattice points and scaled with an *R*-factor of 8.2%, based on 7831 overlapping intensities from 6 crystals. Overall only 61% of all reflections were found to be observed ($I > 1\sigma(I)$) at 2.0 Å and 80% at 2.5 Å resolution. This is due to a sharp decrease in intensity in the outer resolution range (2.3 to 2.0 Å) to below 50% (see Table 2). In the case of WGA2 the situation was somewhat better, as a larger number and better-quality crystals were used yielding a larger number of overlapping data (35,505) (Wright, 1987). The completeness ratios here are 64% and 86% at 2.0 Å and 2.5 Å resolution, respectively.

Structure refinement of WGA1 was carried out by the restrained parameter refinement method of Hendrickson & Konert (1980). As was the case with WGA2, the 2 independent protomers (I and II) in the asymmetric unit were refined separately. These independent subunits are related by a local 2-fold screw axis (see Fig. 1) and do not represent the physiologically active dimer molecule. The unique crystallographic 2-fold axis in the *C*2 space group coincides with the molecular dimer axis. Thus the crystal contains 2 types of dimers: those consisting of type I and those of type II protomers. Initial atomic positions for the 2 independent subunits in the asymmetric unit were those of the refined WGA2 structure. In this model 4 of the 5 mutated side-chains were substituted by model-building on an Evans & Sutherland graphics system using the program FRODO (Jones, 1978): Thr for Pro56, Gln for His59, Tyr for His66 and Ala for Ser93. The 5th mutation, Gly171, did not require any change, since the Ala171 in WGA2 was refined as Gly due to lack of electron density.

Reflections with $F_o < 3\sigma(F_o)$ and $d > 8$ Å were omitted from refinement (45% of the theoretical number of lattice points to 2.0 Å). In the 1st refinement stage only protein atoms were refined against diffraction data in the resolu-

Table 2
Distribution of observed data as a function of resolution

Resolution range (Å)	Total no. reflections	No. data $F_o \geq 1\sigma(F_o)$	No. data $F_o \geq 3\sigma(F_o)$
200.0–4.00	2901	2849(98%)	2849(98%)
4.00–3.00	3806	3383(89%)	3383(89%)
3.00–2.75	2043	1563(76%)	1562(76%)
2.75–2.50	2877	1572(55%)	1572(55%)
2.50–2.35	2362	1303(55%)	1062(45%)
2.35–2.20	3043	1397(46%)	701(23%)
2.20–2.00	5422	1712(32%)	1690(31%)

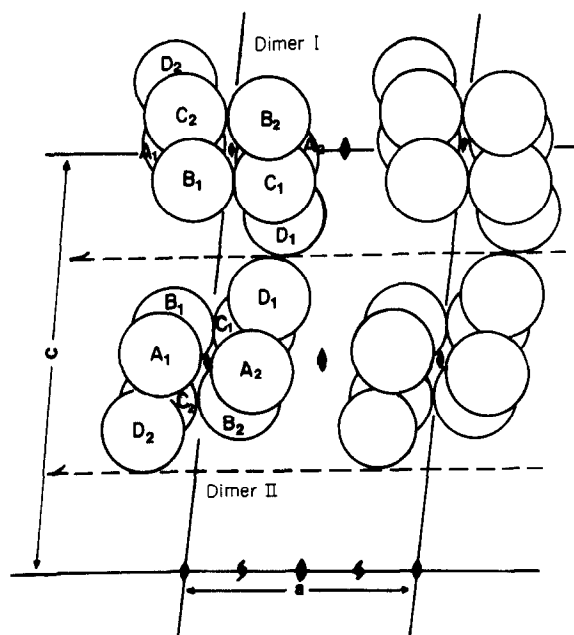


Figure 1. Schematic representation of the packing arrangement of WGA1 dimers in the a - c plane of the $C2$ lattice ($a = 51.22$ Å, $b = 73.60$ Å, $c = 91.42$, $\beta = 98^\circ$). The letters A, B, C, and D refer to the 4 isostructural domains in the polypeptide chain, and subscripts 1 and 2 denote the monomers that constitute each dimer. Crystallographically distinct dimers I and II are centered on 2-fold axes at 0, 0, 0 and 0, 0, 0.5, respectively. The non-crystallographic 2-fold screw axis relating these 2 types of dimers is shown as a broken arrow.

tion range 8.0 to 2.6 Å. The overall B -factor refined to a value of 11.0 Å². Individual atomic B -factors were allowed to vary as the resolution was increased from 2.6 to 2.0 Å. Poor conformations and close contacts were corrected by model-building using FRODO. A set of 61 water sites identical with sites in the WGA2 model was selected initially from the $[F_o - F_c]$ difference Fourier map and included as oxygen in the next refinement stage. During 10 subsequent refinement stages further water sites were sought without reference to the WGA2 structure, and 2 sequence corrections were incorporated. All atoms were refined with an occupancy of 1.0, with the exception of 6 water molecules located on the crystallographic 2-fold axis, which were assigned an occupancy of 0.5 and their positions fixed during refinement. The positions of 9 other atoms were fixed in later refinement cycles to prevent close contacts with nearby atoms that could not be remedied by model-building.

Electron density maps with coefficients $[(2F_o - F_c)\exp(i\alpha_c)]$ and $[(F_o - F_c)\exp(i\alpha_c)]$ were computed at the end of each refinement stage. These were based on only those reflections in the resolution range 8 to 2.0 Å or 10 to 2.0 Å that obeyed the criteria $F_o > 1\sigma$ and $|F_o - F_c| > 1.2(F_o + F_c)/2$ (Bode & Schwager, 1975).

The 2 refined isolectin structures were compared by direct superposition of the 2 electron density maps and by examination of a difference map at 2.0 Å resolution. This map was based on the final model phases of WGA2 and coefficients $[F_{WGA2} - F_{WGA1}]$.

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

Table 3
Refinement statistics of final cycle

Resolution range (Å)	8.0–2.0
No. of reflections $F_o > 3\sigma(F_o)$	12,442
No. of protein atoms	2313
No. of solvent atoms	210
Final r.m.s. co-ordinate shift (Å)	0.012
Final r.m.s. temp. factor shift (Å ²)	0.24
$F_o - F_c$	74.67
$A, B^\dagger$	40, –130
Average temp. factor (Å ²)	20.94
Final R -factor [‡]	16.5%

[†] Structure factor weights were determined from:
 $\text{sigma}(F_o) = A + B(\sin\theta/\lambda - 0.166)$.

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

3. Results and Discussion

(a) WGA1 refinement

The final refinement statistics and the restraint information are summarized in Tables 3 and 4. During the course of refinement (107 total cycles) the crystallographic R -factor dropped from 28.2% to 16.5% including 55.4% of all lattice points to 2.0 Å resolution ($F_o > 3\sigma$). The effective resolution of the electron density map is, however, estimated to be 2.3 Å (Swanson, 1988), because of the decrease in useful data in the higher resolution range (2.5 to 2.0 Å). Including the unobserved reflections in the final structure factor calculation raises the R -factor to 32.5%.

The model was readjusted in 11 model-building sessions and includes 210 solvent sites. Two sequence corrections were incorporated, at Phe109 (earlier Ser109) and Lys134 (earlier Gly134) (Raikhel & Wilkins, 1987; Wright & Raikhel, 1989).

Table 4
Restraint information

	Target value	r.m.s. deviation	No. of restraints
A. Distances (Å)			
Bond	0.023	0.018	2419
Angle	0.038	0.042	3284
Planar (1–4)	0.043	0.038	902
Planar groups	0.028	0.018	2220
Chiral volume (Å)	0.170	0.187	284
B. Non-bonded contacts (Å)			
Single torsion	0.50	0.201	816
Multiple torsion	0.50	0.219	744
Possible H-bond	0.50	0.267	233
C. Torsion angles (°)			
Planar (peptide)	3.0	2.6	348
Staggered	15.0	20.3	331
Orthonormal	20.0	37.8	27
D. Isotropic thermal factors			
Main-chain bond	2.0	1.83	1470
Main-chain angle	3.0	2.89	1912
Side-chain bond	2.0	1.99	939
Side-chain angle	3.0	3.21	1382

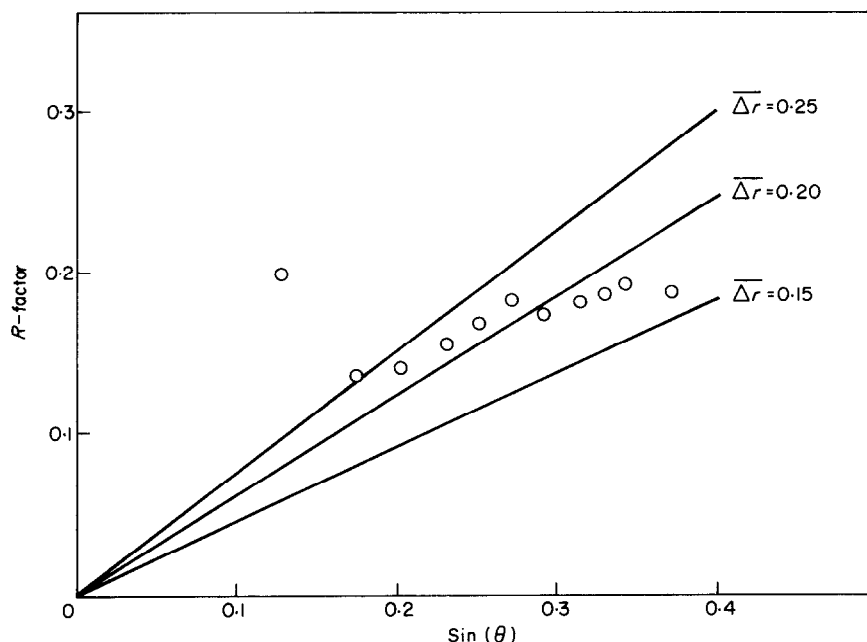


Figure 2. Observed variation of the R -factor as a function of the reciprocal of the resolution ($\sin\theta$). The average R -factor of each point is based only on reflections with $F_o > 3\sigma(F_o)$. The lines represent the theoretical variation of R as a function of different mean positional errors (in Å), according to Luzzati (1952).

Lys134 was refined as Ala, due to lack of side-chain density. The individual atomic B -values range from 2.0 Å^2 to 65.0 Å^2 . The final average isotropic B -factor of 20.94 Å^2 was only slightly lower than that obtained for WGA2 (22.0 Å^2), and dropped to 20.3 Å^2 when water was excluded. This value is high in comparison with mean temperature factors of many other well-refined protein structures, and may in part be due to the large proportion of missing high-resolution data (see Table 2). When refinement was restricted to 2.2 Å resolution in 26 trial cycles and all measurable data ($F_o > 1\sigma$, 71%) were included, an average B -value of 19.3 Å^2 and an R -factor of 16.8% were obtained.

The manner in which the R -factor varies with $\sin\theta$ is illustrated in Figure 2. Reference to the standard lines, which are based on co-ordinate errors only (Luzzati, 1952), indicates an upper limit of 0.2 to 0.25 Å in the mean co-ordinate error.

The final model conforms to standard stereochemistry with respect to bond distances and angles with standard deviations of 0.018 Å and 3.2° , respectively (see Table 5). Reminiscent of the refined WGA2 structure, strained conformations were observed for several peptide bonds in reverse turn regions involved in dimerization. For instance, the backbone dihedral angles of the invariant Asn-Asn pair sequences (-CysPro(Thr)AsnAsn-) in domains A, B and C, which assume a distorted β -turn conformation stabilized by a network of hydrogen bonds across the dimer interface (see Fig. 3), fall outside of region 3 in Figure 4. Moreover, Ala39, Pro82 and Ala125 exhibit ϕ and ψ values outside of region 1 (Fig. 4). These residues are located in identical position in the S-S linked C-terminal loops of domains A, B and C, and their

conformations are also stabilized by inter-subunit interactions (Wright, 1987). The ϕ angles of several other residues in regions 1 and 2 (Ser48, Cys141 in protomer I, and Ala71, Cys78, Lys96, Ala134 in protomer II) remain low ($< -40^\circ$) despite repeated model-building attempts.

Close torsion contacts could not be alleviated in several residues between side-chain and main-chain atoms: Ser19 (C to OG), Gln20 (C to CG), Trp41 (CA to CD1), Glu72 (N to OE1), Tyr145 (CA to CD1); among side-chain atoms: Gln79 (CB to OE1), Lys96 (CB to CE), Ile87 and Ile155 (CD1 to CG2); and between neighboring side-chains: Phe69 (O) to Tyr73 (CD2), Leu16 (CD1) to Met26 (CE), Lys88 (CG, CD) to Phe109 (CD1, CE1). Several close H-bonding contacts ($< 2.2 \text{ Å}$) with water also remained.

Table 5
Refined bond distances and angles

	Average distance	Ideal distance	r.m.s. deviation
A. Bond (Å)			
N-CA	1.463	1.470	0.018
CA-C'	1.525	1.530	0.016
C-N'	1.319	1.320	0.019
C-O	1.248	1.240	0.014
B. Angle (°)			
N-CA-C	110.92	109.60	3.97
C-CA-CB	110.57	109.60	3.51
CB-CA-N	110.34	109.60	3.80
CA-C-O	119.21	121.00	3.11
CA-C-N'	116.07	114.00	2.87
O-C-N'	124.55	124.0	2.41
C'-N-CA	122.67	123.0	3.32
ω	179.73	180.0	2.63

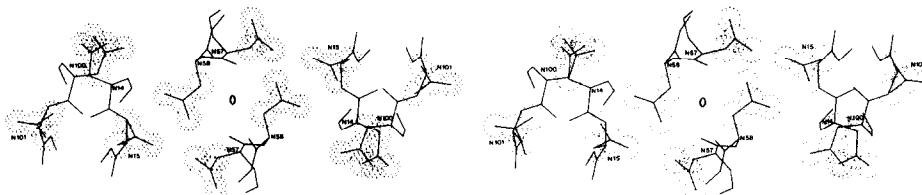


Figure 3. Stereoscopic view of the array of invariant Asn residues in the subunit contact region: Asn14-Asn15 (domain A), Asn57-Asn58 (domain B), and Asn100-Asn101 (domain C). The amide side-chains are emphasized by dot surfaces and the 2-fold axis is shown as a (•).

Superposition of the two independent protomers (I and II) was carried out using the transformation parameters shown in Table 1. The crystallographically different protomers superimposed with an r.m.s. difference of 0.59 Å over all atoms and 0.32 Å for main-chain atoms, excluding residue 171. The largest structural differences were found in the following side-chains: Lys33, Thr54, Arg45, Lys88, Lys96, Leu112, Lys130, Asp135 and Gly171. Of the four domains, domain D superimposed the worst, consistent with the larger degree of disorder observed in the electron density map.

Of the 210 solvent sites, 101 are associated with protomer I and 109 with protomer II. Forty-three

of these in protomer I and 48 in protomers II possess *B*-values above 30.0 Å² and 70 correlate in the two independent environments.

Residue 37 was originally refined as Asp, as determined by amino acid sequencing. Nucleotide sequence data, however, code for an Asn at this position (Raikhel & Wilkins, 1987). Thus, in later refinement cycles the amide form was refined, although it is not possible to distinguish between the amide and acid forms of aspartate on the basis of the electron density. The discrepancy in the sequence data may be a result of deamidation. Deamidation of asparagines when adjacent to Gly (Asn-Gly) is a common phenomenon in proteins and

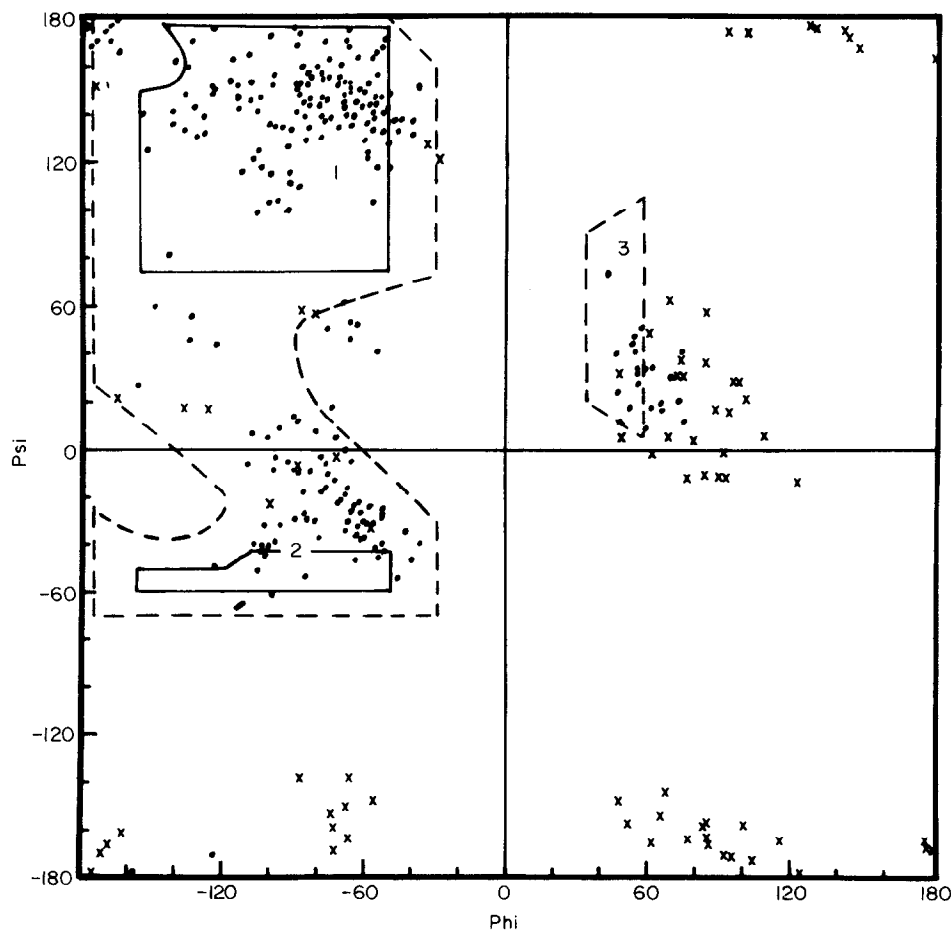


Figure 4. Ramachandran plot of the main-chain dihedral angles ϕ and ψ in protomers I and II of WGA1 (Ramakrishnan & Ramachandran, 1965). The 3 distinctive regions of stable conformation are labeled 1, 2, 3. Glycines are represented by a cross (x).

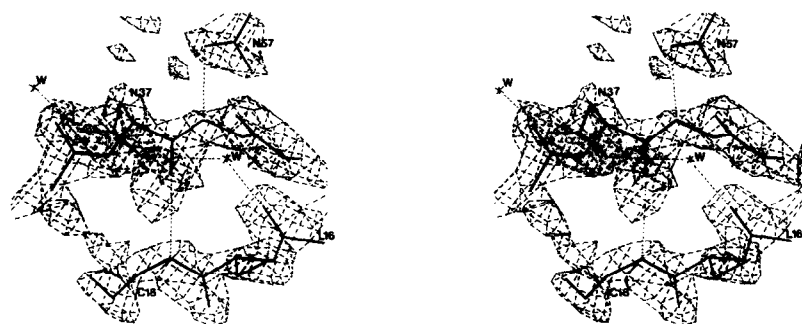


Figure 5. Stereo illustration of the region of the electron density map ($2F_o - F_c$) at Asn37. Possible hydrogen bonds are shown as broken lines.

peptides (Bornstein & Balian, 1977; Meinwald *et al.*, 1986; Geiger & Clarke, 1987). However, as can be seen in Figure 5, structural considerations exclude the likelihood that deamidation could have occurred in the folded protein. The conformation of Asn37-Gly38 is such that a nucleophilic attack by the peptide NH group of Gly38 on the Asn side-chain is not possible. Moreover, the electron density is extremely clear in this region. The conformation of both the peptide bond (Asn-Gly) and the side-chain amide group are immobilized by H-bonds and the atomic B -values are very low (2.0 and 3.0 Å² for ND2 in protomers I and II). For these reasons it appears more likely, that deamidation of Asn37 was a consequence of the acidic conditions used in protein sequencing, which are favorable for deamidation.

Fourier maps and difference Fourier maps were examined at contour levels as low as ± 0.15 electrons in highly disordered regions of the structure. The largest disorder is seen at the C-terminal ends (Asp170-Gly171) in both crystallographically independent protomers. In protomer II, however, where an ionic interaction appears to be possible between the free carboxylate group of Gly171 and the free amino group of Lys96 of protomer I, atomic B -values for Gly171 are not as high (40 to 50 Å²) as those in protomer I (50 to 65 Å²). Flexibility about the CA-C bond of Asp170 allows a number of different orientations for Gly171. Although trial refinements were carried out on several of these in the two independent protomers to determine the most highly occupied position, negative difference density (contour level -0.15 electrons) still remained for the whole residue at the end of the refinement. The same problem was encountered during refinement of WGA2 (Wright, 1987). Since amino acid sequence data were inconclusive at that time, it was speculated that the polypeptide chain might end with 170 and not 171. The recently published nucleotide sequence for WGA3 (Raikhel & Wilkins, 1987), however, is in agreement with the WGA1 sequence, suggesting Gly for residue 171. C-terminal analysis by hydrazinolysis (Wright & Raikhel, 1989) confirmed the presence of a Gly residue in WGA1 and WGA3, and an Ala residue in WGA2.

Electron density is completely lacking for the side-chain atoms of Lys134, earlier believed to be a Gly residue. The side-chain of Lys149 and the indole ring of Trp150, initially also misidentified as Gly (Wright *et al.*, 1984), lack electron density in protomer I. In protomer II low density is observed for both, although in the case of Trp150 the difference map shows negative differences at the position of the indole ring, indicating low occupancy of the refined position. This is unexpected, since in most protein structures tryptophan residues are located in less exposed regions of the molecule, are well ordered and often have structural and functional significance.

Examples of clearly resolved electron density are the inter-subunit regions where saccharide binds. As observed with WGA2, the aromatic side-chains of the two tyrosine residues, Tyr73 and Tyr159, which provide the most important contact for sugar ligands, are well resolved and have low atomic B -values in both the independently refined protomers (6.5 and 9.2 Å² for Tyr73; 19.6 and 12.8 Å² for Tyr159). In the case of Tyr159, the side-chain B -values are the lowest of all residues in the flexible D-domain (residues 130 to 171). Several other residues (Tyr21, Trp41, Glu11) involved in intermolecular contacts, also display lower B -values for their side-chain atoms as compared with main-chain atoms. The side-chain amide nitrogen atom of Asn58, which forms one of the critical contacts around the dimer axis (see Fig. 3), refined with the lowest allowed B -value in both protomers (≤ 2.0 Å²). This region, which includes mutations at residues 56 and 59, assumes a strained reverse-turn conformation, and is represented by strong, clearly defined electron density.

Electron density connections were observed between several side-chains that are in van der Waals' contact. For instance, in the case of the N-terminal residue (pyroglutamic acid) a strong density connection exists across the dimer axis, and the cyclized side-chains of the N termini of the two subunits are in van der Waals' contact (see Fig. 6). To prevent unreasonably close contact between these side-chains, the carbonyl atoms (CD, OE) were fixed in later refinement cycles. Similarly, the side-chain atoms of Lys88 (CD, CG) in protomer II were

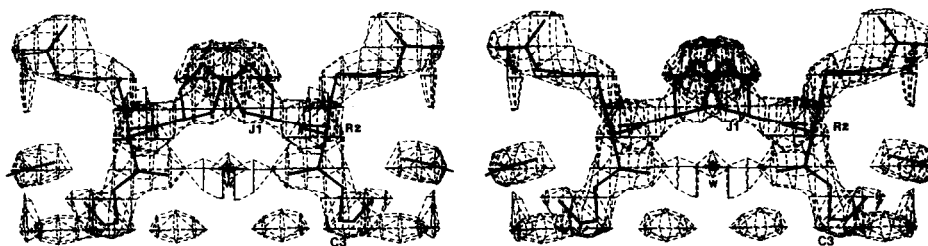


Figure 6. Stereoscopic view of the electron density representing the blocked N termini of 2 WGA1 subunits. Stacking interactions are possible between the cyclized pyroglutamate side-chains across the dimer axis.

fixed to prevent close contact with Phe109. The side-chains of Tyr23 and Arg45 were also found to approach closely in protomer II, where an electron density connection is seen.

(b) *Comparison of the refined models of isolectins 1 and 2*

(i) *Overall structures*

The two isolectin structures could be compared directly, since the sites of mutation are not involved in lattice contacts. The unit cell parameters are very similar, deviating only slightly in the monoclinic β angle (see Table 1). The atomic co-ordinates of WGA2 had been refined with an R -factor of 17.9% (Wright, 1987). However, the model used here was further refined to incorporate the sequence corrections of Phe109, Lys134 (refined as Ala), and Trp150 (refined as Ala in protomer II), and 12 additional water sites.

Each of the independent protomers of isolectin 2 was superimposed with the corresponding one of isolectin 1. In the case of protomer I (1147 protein atoms) this could be done directly, while in the case of protomer II (1143 protein atoms) a translational adjustment to correct for the difference in the translational parameter relating the two independent protomers (see Table 1) was necessary: $X_{WGA1} = X_{WGA2} + 0.247$ Å. The two independent subunits are related by a local screw axis, which runs parallel to the crystal a -dimension, intersects the c -axis at $1/4c$ and the b -axis at -0.03 Å (see Table 1). The results of the separate comparisons are shown in Table 6. Positional displacements (r.m.s. Δr) of 0.64 Å and 0.61 Å were obtained superimposing all atoms, and 0.29 Å and 0.30 Å for main-chain atoms only. These values are closely comparable in magnitude with the ones obtained when the two protomers in the crystallographically different environments were superimposed in the two isolectin structures (0.59 Å and 0.71 Å, see Table 6) and in several other well-refined crystal structures (Finzel *et al.*, 1985; Wang *et al.*, 1985; Weber *et al.*, 1987). Residues with the largest deviation are listed in Table 7. Their positional displacement correlates closely with the degree of flexibility (thermal value). This is illustrated in Figure 7, where the mean B -values for corresponding atoms in WGA1 and WGA2 are plotted against their r.m.s. difference in position (Δr).

Examination of the backbone dihedral angles and disulfide bridge conformation confirms that the overall structures of the two isolectins are identical. The 16 backbone hydrogen bonds that stabilize the domain fold are also closely comparable in their lengths in all four domains. Contact around the non-crystallographic 2-fold axis between residues in domain C of independent protomer I (Gly94, Gly95, Lys96) and domain D of protomer II (Cys164, Gln165, Cys169, Gly171) is longer by 0.4 to 0.6 Å in WGA1. This is consistent with the longer screw translation observed between these protomers in the WGA1 asymmetric unit. Contact between the D-domains of protomers I and II (see Fig. 1) is roughly comparable in the two isolectins (Val140 (I) with Gly137 (II) and Gly138 (II)). The strong lattice interactions observed along the crystal b -axis between dimers related by the c -center of symmetry are the same in WGA1, involving one ion pair (Arg2-Glu72), several strong H-bonds (Tyr21 (OH) to Asp86; Tyr23 (OH) to Glu72), and numerous van der Waals' contacts.

Correlation of water sites in the two refined structures is presented in Table 8. Overall, there are fewer water sites associated with protomer I as

Table 6
Positional differences in isolectin structures

Structures superimposed†	r.m.s. displacement (Å)	
	Main-chain atoms	All atoms
A. WGA1/WGA2		
Protomer I (res. 1–171)	0.29	0.64
B-domain region (233 atoms)		0.63
P-site region (145 atoms)		0.42
S-site region (125 atoms)		0.30
Protomer II (res. 1–171)	0.30	0.61
B-domain region (233 atoms)		0.60
P-site region (145 atoms)		0.35
S-site region (125 atoms)		0.29
B. Protomer I/protomer II		
WGA1	0.32	0.59
WGA2	0.28	0.71
P-site (WGA1)		0.47
S-site (WGA1)		0.45

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

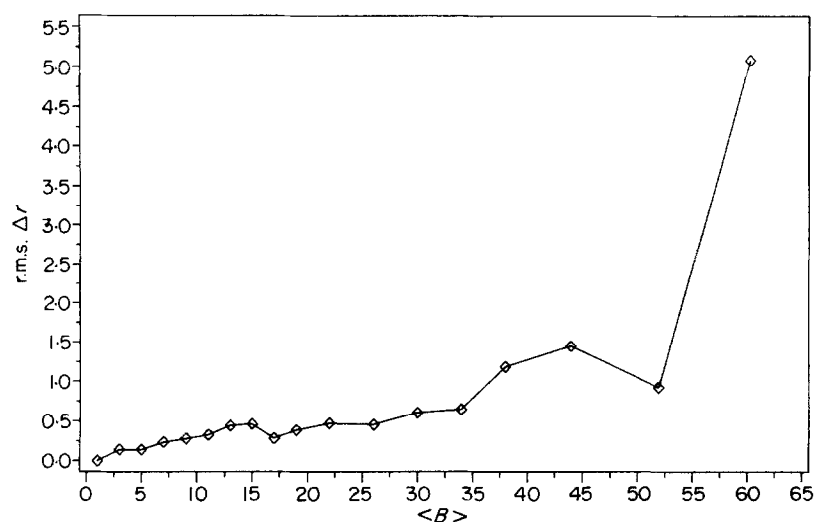


Figure 7. Correlation of thermal parameters with r.m.s. positional displacement of corresponding atoms in the refined WGA1 and WGA2 models.

Table 7
Residues with displacements of more than 1.0 Å

Residue	Atom (Δr)†	Deviating torsion angles
A. Protomer I‡		
Glu5	OE1(1.17), OE2(1.9)	χ^3
Ser8	OG(1.9)	χ^1
Lys33	CE(2.9), NZ(4.2)	χ^3 and χ^4
Lys44	CE(1.6), NZ(3.8)	χ^2 and χ^4
Arg45	NE(1.4), NH2(1.4), NH1(2.6)	χ^4
Thr54	CG2(1.2)	χ^1
Tyr64	CD2(1.05), CE2(1.03)	χ^2 , plane tilt
Lys88	CD(1.8), CE(1.4)	χ^2 , χ^3 and χ^4
Lys96	CG(1.85), CE(1.44), NZ(1.25)	All side-chain angles
Leu97	CD1(2.29), CD2(2.73)	χ^1 and χ^2
Pro99	CG(1.8)	χ^1 and χ^2
Asn100	OD1(1.35), ND2(1.29)	χ^2
Leu102	CD1(2.38), CD2(2.65)	χ^2
Phe109	CD2(1.16), CE2(1.46)	χ^2 , plane tilt
Leu112	CG(1.28), CD1(2.34), CD2(2.93)	χ^1 and χ^2
Ser114	OG(2.04)	χ^1
Lys130	CD(1.14), CE(2.16), NZ(1.41)	χ^2 and χ^4
Asp135	CB(1.34), CG(1.05), OD2(1.25)	χ^2
Asp170	OD2(1.25), C(1.03), O(1.21)	ψ , χ^1
Gly171	N(3.48), CA(4.8), C(6.8), O(8.1), OT(7.6)	Highly disordered
B. Protomer II		
Glu5	OE1(1.2), OE2(1.1)	χ^2
Met10	CE(1.04)	χ^3
Leu16	CD1(2.36), CD2(2.56)	χ^2
Met26	CE(1.39)	χ^3
Lys33	CG(2.87), CD(4.96), CE(5.67), NZ(8.53)	χ^1 , χ^2 and χ^3
Lys44	CG(1.47), CE(2.34), NZ(3.88)	χ^3 and χ^4
Ser48	OG(1.82)	χ^1
Gly51	O(1.29)	ψ
Thr54	O(1.03)	ψ
Arg84	NH2(1.22), NH1(1.14)	χ^4
Ile87	CD1(1.51)	χ^2
Lys88	CG(3.02), CD(3.21), CE(4.21), NZ(5.27)	All side-chain angles
Asn100	ND2(1.77)	χ^1 and χ^2
Leu102	CD1(2.21), CD2(2.66)	χ^1 and χ^2
Leu112	CD1(2.52), CD2(2.49)	χ^1 and χ^2
Lys130	NZ(1.31)	χ^4
Asp135	OD2(1.44)	χ^2
Arg139	CG(1.24)	All side-chain angles
Val140	CG2(1.25)	χ^1
Lys149	CE(1.04), NZ(1.78)	χ^3
Gly171	O(2.66), OT(1.67)	Disordered

† Positional difference between WGA1 and WGA2 in Å.

‡ I and II refer to the 2 independent subunits in the asymmetric unit.

Table 8
Correlation of solvent sites in WGA1 and WGA2

Type of comparison	Number of sites	Average B -value ($\text{\AA}^2$)
A. Protomers I and II†		
In both WGA1 and WGA2	51	24.2
In WGA1 only	8	36.8
In WGA2 only	4	22.9
In WGA1 and either protomer of WGA2	11	36.2
In WGA2 and either protomer of WGA1	8	26.4
B. Protomer I		
In WGA1 and WGA2	9	32.5
In WGA1 only	18	33.1
In WGA2 only	15	36.1
C. Protomer II		
In WGA1 and WGA2	14	36.1
In WGA1 only	21	35.6
In WGA2 only	19	38.6

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

compared with protomer II in both crystal structures, 101 *versus* 109 in WGA1; and 91 *versus* 107 in WGA2. The pattern of water association with any one of the protomers in the four possible crystallographic environments was analyzed by pairwise superposition of the models using molecular graphics. A common set of 51 sites could be correlated in all four environments. The majority of these sites are considered to be well determined, since their average temperature factor is relatively low ($24.2 \text{ \AA}^2$). With the exception of only one, all these waters solvate specific protein groups, and thus play a role in stabilizing protein conformation. Appreciably fewer matches are observed in all the other categories listed in Table 8. For instance, a total of 19 water sites could be correlated in any three of the four environments and numbers ranging from nine to 21 were found to be associated with only one of the two types of protomers in each crystal structure. Even fewer are unique to both protomers in only one of the crystal structures, eight sites in WGA1 and five in WGA2. Most of these are, however, located in regions of sequence differences. In general, those sites completely unique to only one environment in either structure possess a high degree of positional disorder, as expected.

Regions of high flexibility coincide, not surprisingly, in the two structures. These are typically found at exposed reverse turns and at the C-terminal end. Comparing the mean B -values of the individual domains in WGA1 to those in WGA2 (see Table 9), a similar pattern is observed. The domain B -values decrease in the C- to N-terminal direction, indicating a larger degree of thermal stability, which can be correlated with involvement in dimerization. Domain D has the highest average B -value and participates the least in dimer interactions. This is more pronounced in protomer I, where this

Table 9
Temperature factor comparison

	Average B -value ($\text{\AA}^2$)	
	WGA1	WGA2
A. Protomer I†		
Domain A	15.58	17.35
Domain B	17.53	19.00
Domain C	17.30	18.56
Domain D	32.42	32.85
B. Protomer II		
Domain A	16.17	17.21
Domain B	19.59	19.50
Domain C	21.11	22.90
Domain D	24.48	25.08
All protein atoms (protomers I and II)	20.30	21.45

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

terminal domain is less constrained by lattice contacts than in protomer II.

(ii) *Specific regions*

The $2 \text{ \AA}$ difference map was essentially found to be in agreement with the earlier calculated $2.2 \text{ \AA}$ map, which had been averaged over the two independent protomers in the asymmetric unit (Wright & Olafsdottir, 1986). Significant features in this map (>0.2 electrons) are confined to the region surrounding the three closely spaced mutations in domain B, residues 56, 59 and 66, and are consistent in the two independently refined protomers. No difference density is observed at residues 93 (Ser *versus* Ala) and 171 (Ala *versus* Gly), located at disordered regions at the surface. Because of the higher resolution and availability of refined atomic positions of both isolectins, it has now been possible to interpret small features in the difference map and quantify structural shifts more precisely than in an earlier study. Readjustments of several side-chains appear to be a direct result of the nature of the mutations, preserving van der Waals' and hydrophobic interactions in the domain/domain contact region that stabilize the dimer structure.

The region of the largest difference density is depicted in Figure 8. Superposition of all of the atoms in this region (residues 15, 38–39, 43 to 69, 100 to 102, and 112–113) yields a r.m.s. displacement of $0.63 \text{ \AA}$ in protomer I (133 atoms) and $0.60 \text{ \AA}$ in protomer II (140 atoms). These values are as large as the overall r.m.s. deviation calculated for each independent protomer, and for comparisons of independently refined protomers within each crystal structure (see Table 6). Thus, this indicates a substantial shift, considering that this region of the structure is well ordered. Overall, a general displacement of the region Ser43 to Phe69 in WGA1 is observed in the direction away from the dimer axis (by about 0.2 to $0.3 \text{ \AA}$). van der Waals' contacts between several self-interacting residues around the

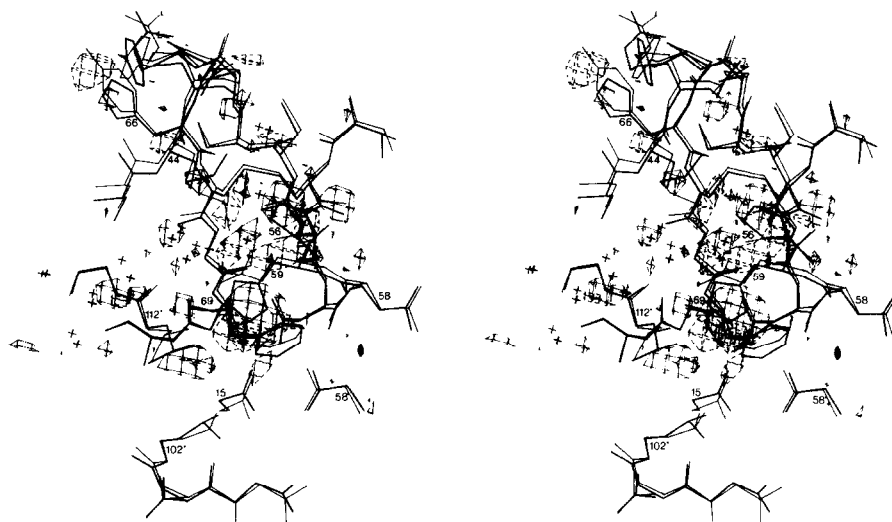


Figure 8. Stereoscopic view of the superimposed models of WGA1 (thin bonds) and WGA2 (thick bonds) displaying the structural shifts in the B-domain region that are results of 3 sequence substitutions: Thr to Pro at residue 56, Gln to His at residue 59, and Tyr to His at residue 66. The difference density ($F_{WGA2} - F_{WGA1}$) is shown superimposed. Broken contours represent negative density. The position of the dimer axis is indicated by (•). Residue numbers denoted with a superscript * belong to the dimer-related subunit.

dimer axis are slightly longer in WGA1 (Pro82, CD-CD) contact, Phe69, CZ-CZ contact). However, a very strong H-bond interaction involving the amide side-chains of Asn15 and Asn58 of opposite subunits is preserved (see Fig. 9). The most important structural differences can be summarized as follows:

(1) The largest difference is seen at residue 59, where Gln replaces His in WGA1. Although the χ_1 and χ_2 angles of the two types of side-chain differ by only 10 to 15°, the imidazole in WGA2 is oriented very differently compared with the amide group of Gln59 in WGA1 (see Fig. 9). The side-chains of both residues are stabilized by hydrogen bonds. In the case of His59, both imidazole nitrogens engage in H-bond interactions: NE2 with Asn15 (OD1) and ND1 with two water molecules (W178, W188). These ordered water molecules are not present in WGA1, because they are displaced by the amide group of Gln59. No direct interaction with Asn15 is observed in WGA1, but a low occupancy water molecule (W241) appears to mediate the contact between OE1 of Gln59 and OD1 of Asn15. The NE2 atom of Gln59 is at a suitable distance from the

carbonyl group of Thr56 (2.75 Å) for H-bond formation (see Fig. 9).

(2) Rotational freedom about the NH-CA bond of residue Thr56 in WGA1 (Pro56 in WGA2) is responsible for shifts all along the chain back to Ser43. For example, positional differences in the adjacent carbonyl group and S-S bridge of Cys55 in the two isolectin structures are a direct consequence of this flexibility. As shown in Figure 9, the threonine side-chain is stabilized by three possible H-bonds involving the OG1 atom. One good H-bond is possible with the amide-OD1 atom of the neighbouring Asn57, and two others may exist involving NZ of Lys44 and a water molecule (W254), not present in WGA2. The Lys44 side-chain takes up a totally different orientation in the two isolectin structures. In WGA2 it is turned away from Pro56, while in WGA1 it sits close to OG1 of Thr56 (2.4 to 3.0 Å) and is further fixed by two possible interactions with ordered water (W177, W254) (see Fig. 9). This stabilization is reflected in the atomic B -values, which are lower by 10 to 15 Å² for the side-chain atoms in WGA1.

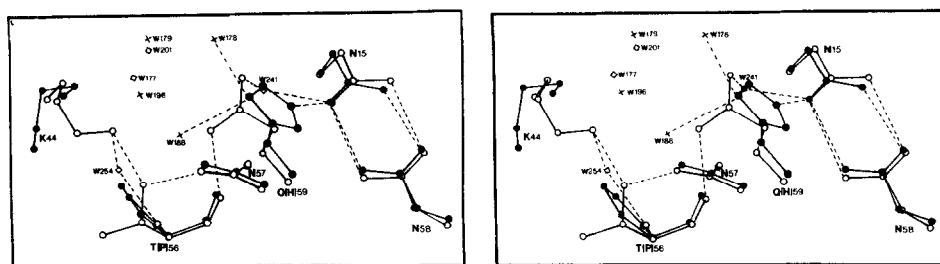


Figure 9. Stereo view of the immediate environment of residues 56 and 59 and their hydrogen bonding interactions in WGA1 (open circles) and WGA2 (filled circles). Water molecules are represented by crosses for WGA2 and squares for WGA1. Broken lines represent hydrogen bonds.

Moreover, the orientation of the Thr56 peptide bond is stabilized by hydrogen bonds. As discussed above, the -CO- group interacts with Gln59 and the -NH- group is solvated by the same water molecule (W254) that interacts with the side-chain hydroxyl group (OG1) and which is absent in WGA2. Several other water molecules in the vicinity also shift slightly in accord with their protein ligands (CO of Gly38, CO of Trp41, NH of Lys44, OG of Ser43).

(3) Notable shifts are observed for several hydrophobic residues in the dimer interface as a direct result of side-chain replacement at residue 59. The phenyl ring of Phe69 takes up different positions in the two structures to preserve van der Waals' contact with the differing side-chain orientations of residue 59 (see Fig. 8). The same is true in the dimer interface where the aromatic rings of residues 59 and 69 interact through hydrophobic contact with several residues of the dimer related C-domain (Leu102, Leu112, Phe116) in WGA2 (Wright, 1987; see Table 12). This important dimerization contact is preserved in WGA1, in that the side-chains of both Leu102 and Leu112 are shifted towards Gln59 (see Fig. 8).

(4) One of the five mutations that differentiate WGA1 from WGA2 (Tyr *versus* His66) is an integral part of the "primary" sugar binding site (or P-site). In this site side-chains from the B-domain of one protomer and the C-domain of the other contribute to sugar binding. Difference density is seen in the 2 Å difference map for the extra atoms present in Tyr66 (CZ-OH) (see Fig. 8). Although different orientations were observed for these side-chains in the two independent environments in both crystal

structures, suggesting a certain degree of flexibility, the superimposed aromatic rings of Tyr66 (WGA1) and His66 (WGA2) were found to match closely (see Fig. 10). Refined atomic temperature factors for these side-chains range from 20 to 30 Å² in the two isolectin structures. Their involvement in sugar binding has been suggested in earlier studies using simple oligosaccharides (Wright, 1980b, 1984). A more exact analysis of their role in stabilizing sugar ligands from refinement of these complexes will be presented elsewhere.

The side-chains of several other residues (Tyr64, Phe116, Ser62, Ser43) and of a number of water sites also undergo slight shifts. Atom by atom superposition in this region (145 atoms), excluding water, yields an r.m.s. difference of 0.42 Å for protomer I and of 0.35 Å for protomer II. These values are far below the overall r.m.s. deviations of the entire structure (see Table 6), indicating that precise conformations in this functionally important region of the molecule are conserved. The side-chain of Tyr73, considered to be the most critical contact for bound sugar ligands, superimposes best and possesses extremely low atomic thermal values ($B = 4$ to 12 Å²). Superposition of 125 atoms in the "secondary" binding site (S-site), located between A- and D-type domains of opposite protomers of the isolectin dimers, yielded an even lower r.m.s. difference, 0.30 Å and 0.29 Å for the two independent protomers (Fig. 11). The critical residues for sugar binding are, with the exception of one, conserved in comparison with the P-site (Wright, 1984). The residue corresponding to Tyr(His)66 (P-site) is Ser152, conserved in all three wheat isolectins

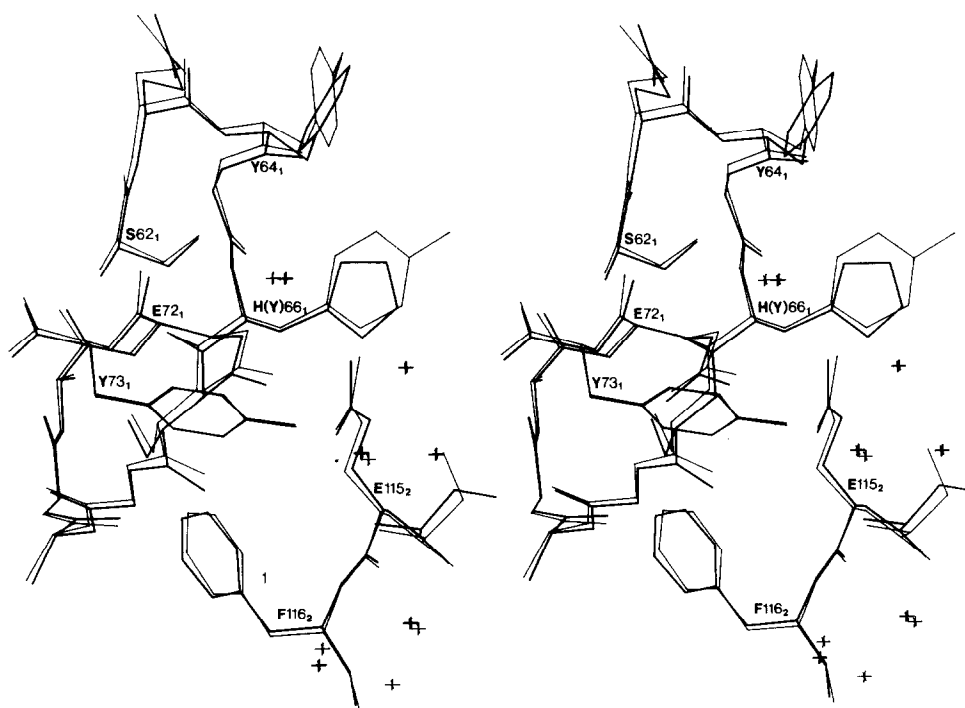


Figure 10. Stereo view of the superimposed models of WGA1 (weak lines) and WGA2 (thick lines) at the "primary" saccharide binding site in protomer I. Crosses denote water sites. Subscripts 1 and 2 refer to the dimer related subunits 1 and 2.

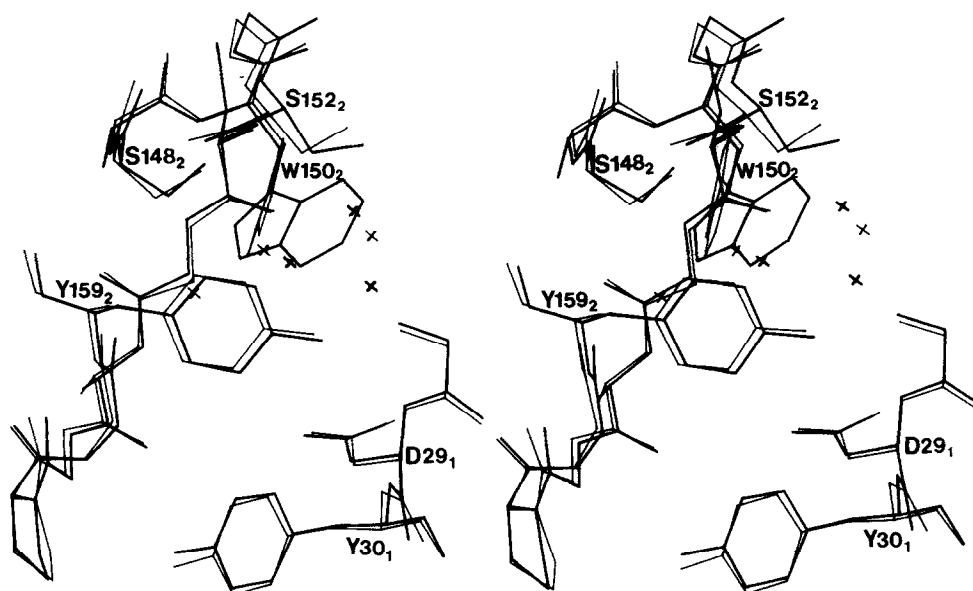


Figure 11. Stereo view of the superimposed models of WGA1 (thin lines) and WGA2 (thick lines) at the "secondary" saccharide binding site of protomer I. Crosses denote water sites. Subscripts 1 and 2 refer to the dimer related subunits 1 and 2.

(Wright & Raikhel, 1989), as well as in barley lectin (Lerner & Raikhel, 1989). It is interesting to note, that in rice lectin, which belongs to a different subtribe of Triticeae and differs from the wheat and barley lectins in 25% of its amino acid sequence, this Ser is replaced by a Tyr residue (Wilkins & Raikhel, 1989).

In summary, superposition of the two refined isolectin structures demonstrates clearly that despite local structural differences caused by the changes in sequence, dimerization interactions and the conformation in regions of functional importance are conserved to assure precise alignment of specific sugar ligands.

I thank Dr Natasha Raikhel for making available to me nucleotide sequence information of the lectins of barley and rice prior to publication. This research was supported by the National Institutes of Health grant AI-17992.

References

- Allen, A. K., Neuberger, A. & Sharon, N. (1973). *Biochem. J.* **131**, 155–162.
- Bassett, E. W. (1975). *Prep. Biochem.* **5**, 461–472.
- 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.
- Bode, W. & Schwager, P. (1975). *J. Mol. Biol.* **98**, 693–717.
- Bornstein, P. & Balian, G. (1977). *Methods Enzymol.* **47**, 132–145.
- Etzler, M. E. (1985). *Annu. Rev. Plant Phys.* **36**, 209–234.
- Ewart, J. A. D. (1975). *J. Sci. Food Agric.* **26**, 5–22.
- Finzel, B. C., Weber, P. C., Hardman, K. D. & Salemme, F. R. (1985). *J. Mol. Biol.* **186**, 627–643.
- Geiger, T. & Clarke, S. (1987). *J. Biol. Chem.* **262**, 785–794.
- 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.
- Hendrickson, W. A. & Konnert, J. H. (1980). In *Biomolecular Structure, Conformation, Function and Evolution* (Srinivasan, R., ed.), vol. 1, pp. 43–57. Pergamon Press, Oxford and New York.
- Jones, T. A. (1978). *J. Appl. Crystallogr.* **11**, 268–272.
- Kronis, A. K. & Carver, J. P. (1982). *Biochemistry*, **21**, 3050–3057.
- Lacelle, N. (1979). PhD thesis, Toronto University.
- Lerner, D. & Raikhel, N. V. (1989). *Plant Physiol.* **90**, in the press.
- Lis, H. & Sharon, N. (1986). *Annu. Rev. Biochem.* **55**, 35–67.
- Luzzati, V. P. (1952). *Acta Crystallogr.* **5**, 802–810.
- Mansfield, M. A., Peumans, W. J. & Raikhel, N. V. (1988). *Planta*, **173**, 482–489.
- Meinwald, Y. C., Stimson, E. R. & Scheraga, H. A. (1986). *Int. J. Peptide Protein Res.* **28**, 79–84.
- Peumans, W. J., Stinissen, H. M. & Carlier, A. R. (1982). *Planta*, **154**, 562–567.
- Raikhel, N. V. & Wilkins, T. A. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 6745–6749.
- Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
- Rice, R. H. (1976). *Biochim. Biophys. Acta*, **444**, 175–180.
- Rice, R. H. & Etzler, M. E. (1975). *Biochemistry*, **14**, 4093–4099.
- Stinissen, H. M., Peumans, W. J. & Carlier, A. R. (1983a). *Planta*, **159**, 105–111.
- Stinissen, H. M., Peumans, W. J., Law, C. N. & Payne, P. I. (1983b). *Theoret. Appl. Genet.* **67**, 53–58.
- Strossberg, A. D., Buffard, D., Kaminski, P. A., Chapot, M.-P., Rossow, P. W. & Foriers, A. (1986). In *Molecular Biology of Seed Storage Proteins and Lectins*, Proc. 9th Annual Symp. Plant Physiol., U. California, Riverside (Shannon, L. M. & Chrispeels,

- M. J., eds), pp. 1-16, The American Society of Plant Physiologists, Rockville Pike, MD.
- Swanson, S. M. (1988). *Acta Crystallogr. sect. A*, **44**, 437-442.
- Wang, D., Bode, W. & Huber, R. (1985). *J. Mol. Biol.* **185**, 595-624.
- Weber, I. T., Gilliland, G. L., Harman, J. G. & Peterkofsky, A. (1987). *J. Biol. Chem.* **262**, 5630-5636.
- Wilkins, T. A. & Raikhel, N. V. (1989). *Plant Cell*, **1**, 541-549.
- Wright, C. S. (1977). *J. Mol. Biol.* **111**, 439-457.
- Wright, C. S. (1980a). In *Biomolecular Structure, Conformation, Function and Evolution* (Srinivasan, R., ed.), vol. 1, pp. 9-17, Pergamon Press, Oxford and New York.
- Wright, C. S. (1980b). *J. Mol. Biol.* **141**, 267-291.
- Wright, C. S. (1981). *J. Mol. Biol.* **145**, 453-461.
- Wright, C. S. (1984). *J. Mol. Biol.* **178**, 91-104.
- Wright, C. S. (1987). *J. Mol. Biol.* **194**, 501-529.
- Wright, C. S. & Olafsdottir, S. (1986). *J. Biol. Chem.* **261**, 7191-7195.
- Wright, C. S. & Raikhel, N. (1989). *J. Mol. Evol.* **28**, 327-336.
- Wright, C. S., Gavilanes, F. & Peterson, D. L. (1984). *Biochemistry*, **23**, 280-287.
- Wright, H. T., Brooks, D. M. & Wright, C. S. (1985). *J. Mol. Evol.* **21**, 133-138.

Edited by W. Kendrickson