Acta Cryst. (1995). D**51**, 1013–1019

X-ray Structure of Wheat Germ Agglutinin Isolectin 3

BY K. HARATA,* H. NAGAHORA AND Y. JIGAMI

National Institute of Bioscience and Human Technology, 1-1 Higashi, Tsukuba, Ibaraki 305, Japan

(*Received 5 January* 1995; *accepted 20 March* 1995)

Abstract

Wheat germ agglutinin isolectin 3 (WGA3) was crystallized from 10 m*M* acetate buffer at pH 4.9 containing 6 m*M* $CaCl_2$ and 4%(*v*/*v*) ethanol. The crystal belongs to monoclinic space group $P2_1$ with unit-cell dimensions $a = 44.86$, $b = 91.02$, $c = 44.86$ Å, and $\beta = 110.22°$. The asymmetric unit contains two molecules ($V_m = 2.51$ Å^3 Da^{-1}). The crystal structure was solved by the molecular-replacement method and was refined by the simulated-annealing method. The conventional *R* value was 0.191 for 19 713 reflections [$|F_o| > 3\sigma(F)$] in the resolution range 8–1.9 Å. The r.m.s. deviations from the ideal bond distances and angles were 0.014 Å and 3.0°, respectively, and the estimated coordinate error was 0.2–0.25 Å. The two molecules in the asymmetric unit are related by the pseudo twofold symmetry and form a dimer structure. The backbone structures of the two subunits are nearly identical with the r.m.s. difference of 0.36 Å for the superposition of equivalent C^α atoms. The dimer structure is very similar to those of isolectins 1 and 2 with the r.m.s. difference of 0.35–0.39 Å for the C^α superposition. Since amino-acid residues which differ from those of isolectin 1 or 2 are not involved in the contact between the two subunits, the subunit–subunit interaction is not significantly affected by the replacement of these residues. As a result, the geometry of the sugar-binding sites which are located at the interface between the two subunit molecules is basically conserved among three isolectins.

Introduction

Lectins are a class of proteins which agglutinate cells or precipitate polysaccharides and glycoproteins. Wheat germ agglutinin (WGA) is a plant lectin which is specific to *N*-acetylglucosamine and *N*-acetylneuraminic acid (Sharon & Lis, 1989). WGA consists of three isolectins, WGA1 (35%), WGA2 (55%), and WGA3 (10%) (Rice & Etzler, 1975), which differ in 5–8 of their 171 amino-acid residues (Smith & Raikhel, 1989). To define the molecular basis of WGA interaction with carbohydrate, we have applied protein-engineering techniques and succeeded in the expression of recombinant WGA2 from yeast *Saccharomyces cerevisiae* (Nagahora, Ishikawa, Niwa, Muraki & Jigami, 1992). For the modification of WGA structure, it is essential to know the three-dimensional structure of three isolectins in relation to their sugar-binding activity.

Each isolectin consists of two equivalent subunits which have a fourfold repeat of a highly homologous domain structure. Crystallographic studies were carried out initially on crystals containing all three isolectins (Wright, Keith, Langridge, Nagata & Burger, 1974). Later, WGA1 and WGA2 were crystallized separately and their three-dimensional structures have been determined by X-ray analysis (Wright, 1987, 1989). WGA1 and WGA2 are so similar in their three-dimensional structure that they can form stable hybrid dimers by subunit interchange although a distinct difference in the sugar-binding affinity has been reported (Kronis & Carver, 1982). On the other hand, the binding affinity of WGA3 for *N*-trifluoroacetyl chito-oligosaccharides is more than ten times higher than that of WGA1 and WGA2 (Gatellier, Grivet & Delmotte, 1987). WGA3 differs in eight and seven amino-acid residues from WGA1 and WGA2, respectively, while the difference between WGA1 and WGA2 is found in five residues. The difference includes the 66th and 109th residues which are involved in the sugar-binding action (Wright, 1990; Wright & Jaeger, 1993). The crystallographic study of WGA3 has been carried out to elucidate how such amino-acid replacement affects the structure and the sugar-binding ability.

Materials and methods

WGA was purchased from Seikagaku Kogyo Co. and WGA3 was purified according to the literature (Rice & Etzler, 1975; Nagahora *et al.*, 1992). Crystals were prepared under the same conditions as in the crystallization of WGA1 and WGA2 reported by Wright *et al.* (1974). 2.5 mg lyophilized WGA3 was dissolved in 0.2 ml 10 m*M* acetate buffer at pH 4.9 containing 6 m*M* $CaCl_2$ and 4%(*v*/*v*) ethanol. The sample solution was divided into a 50 µl portion in a concave well, which was allowed to stand at room temperature in a sealed small plastic container. Crystals, which showed a pseudo $C222_1$ symmetry, belong to monoclinic space group $P2_1$ and the unit cell contains four molecules ($V_m = 2.51$ Å^3 Da^{-1}). The cell dimensions were

* To whom correspondence should be addressed.

$a = 44.86$, $b = 91.02$, $c = 44.86$ Å, $\beta = 110.22°$. Diffraction data to 1.82 Å resolution were collected at 291 K with graphite-monochromated Cu $K\alpha$ radiation on an Enraf–Nonius FAST diffractometer with a GX21 generator (40 kV, 60 mA, focal spot size 0.3 × 3 mm). Two crystals were used for data collection. For each crystal, after the data collection for the resolution range of 6.8–1.82 Å, reflections in the resolution range of 24.5–2.2 Å were collected. A total of 83 180 reflections of the first crystal were merged to 25 173 unique reflections with an R_{merge} value of 0.081 and 59 108 reflections of the second crystal were merged to a set of 21 486 unique reflections with an R_{merge} value of 0.055. These two data sets were merged to produce a set of 26 197 unique reflections ($R_{merge} = 0.064$) which correspond to 82% of the unique reflections in the resolution range up to 1.82 Å. The R_{merge} value for the merge of two data sets was plotted against resolution in Fig. 1.

The structure was solved by the molecular-replacement method using a set of coordinates of WGA1 (Wright, 1989). Since the asymmetric unit contains two molecules, the dimeric structure related by the crystallographic twofold symmetry which was observed in the dimer structure of WGA1 was assumed in the initial model. Using the program *X-PLOR* (Brünger, Kuriyan & Karplus, 1987), the cross-rotation search was performed. Reflections in the resolution range from 15 to 4 Å were used for the calculation of the Patterson map and the Patterson vectors of 5–20 Å were used for the rotation search. Selected peaks from the cross-rotation map were subjected to Patterson correlation (PC) refinement (Brünger, 1990), which gave only one plausible solution. Successive translation searches using reflections in the resolution range 15–4 Å gave the correct position of the dimer unit in the unit cell.

The refinement of the structure was carried out by energy minimization combined with molecular-dynamics simulation. After molecular-dynamics simulation at 2000 K (1 ps), 1500 K (0.5 ps), 1000 K (0.5 ps) and 300 K (0.5 ps) followed by energy minimization for atomic parameters, eight amino-acid residues which differ between WGA1 and WGA3 were replaced. Then, molecular dynamics simulations at 2000 K (1 ps), 1500 K (0.5 ps), 1000 K (0.5 ps) and 300 K (0.5 ps) were performed. After the refinement of atomic parameters, positions of water molecules were obtained from $F_o - F_c$ and $2F_o - F_c$ maps. Electron-density peaks higher than 3σ in the $F_o - F_c$ map and with suitable interatomic contacts with the protein were taken into account, but the water atoms with the B value greater than 60 Å^2 were omitted during the refinement. The map calculation and refinement were repeated three times. At the final stage of the refinement, 358 fully occupied water molecules were included in the structure model. The refined atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven.* Except for *X-PLOR*, the computer programs used were those developed in the authors' laboratory.

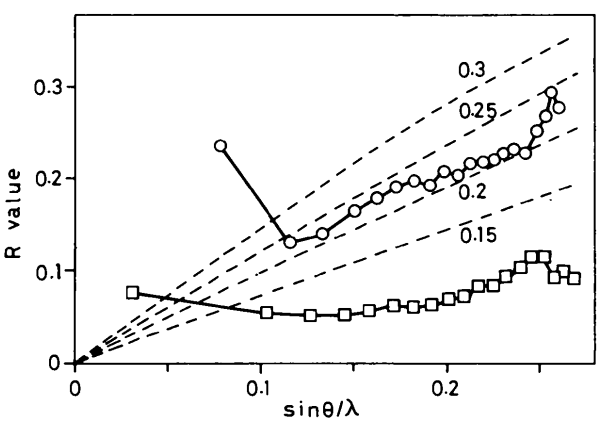

Fig. 1. Plot of R value (○) for the final model and R_{merge} value (□) for the merge of two data sets. Dashed lines indicate the upper estimate of the coordinate error (Luzzati, 1952).

Results and discussion

Crystal

In the earlier work of crystallization, it has been reported that WGA crystallizes in two crystal forms under the same conditions (Wright *et al.*, 1974). Crystals of WGA1 and WGA2 have a space group either $C2$ or $P2_1$. X-ray structure determination of WGA1 and WGA2 has been carried out for the $C2$ crystal (Wright, 1987, 1989). For WGA3, we have obtained a $P2_1$ crystal with the pseudo symmetry of $C222_1$ in the same conditions as those used in the crystallization of WGA1 and WGA2. The WGA3 crystal was isomorphous with the reported $P2_1$ crystals of other WGA's when their unit cell was transformed to a primitive one.

Crystals of WGA3 showed X-ray diffraction spots diffused in the direction of the b^* axis. This caused a problem in the estimation of a peak profile for the evaluation of intensity on a FAST diffractometer, especially, for reflections in low-resolution area. Such diffuse scattering resulted in less reliable measurement of intensity data as indicated from the plot of the R_{merge} values (Fig. 1).

Structure refinement

The structure was refined to an R value of 0.191 for 19 713 reflections with $|F_o| > 3\sigma(F)$ in the resolution range 8–1.9 Å. A total of 2706 atoms of two WGA3 molecules with 171 residues and 358 water molecules with full occupancy were included in the refinement.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1WGT). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AS0683).

Structures of the two WGA3 molecules were refined independently. The maximum values for the positive and negative residual electron densities in the $F_o - F_c$ map were 0.42 and $-0.41\,e\,Å^{-3}$, respectively, and the r.m.s. residual electron density was $0.087\,e\,Å^{-3}$. The r.m.s. deviations of bond distances and angles from their ideal values were 0.014 Å and 3.0°, respectively. In Fig. 1, the R value was plotted against the resolution. The coordinate error (Luzzati, 1953) was estimated as 0.2–0.25 Å. The Ramachandran plot (Ramakrishnan & Ramachandran, 1965) shown in Fig. 2 is characteristic with many glycine residues. The (φ, ψ) values of non-glycine residues fall in a normal range.

Subunit structure

In the DNA sequence (Smith & Raikhel, 1989), the N-terminal residue of WGA is Gln. On the other hand, Nagata (1985) has reported that pyroglutamic acid is the N-terminal residue. In the X-ray structures of WGA1 and WGA2, the N-terminal residue has been assigned as pyroglutamic acid (Wright, 1989). We carefully examined the $2F_o - F_c$ and omit maps and found that the pyroglutamyl group is better fitted to the electron density than Gln (Fig. 3). It has been reported that the C-terminal region is heterogeneous and WGA3 contains 88% of the protein with Gly171 as the C-terminal residue (Hagahora, *et al.*, 1992). In the present WGA3 structure, the C-terminal region is highly flexible and the poor electron density suggests the possibility of disorder.

The crystal contains two independent molecules which are related by pseudo twofold symmetry. WGA's consist of four domains designated *A* to *D*. Domains *A*, *B*, and *C* have 43 residues while the domain *D* consists of 42 residues. These domains have high homology in their sequence structure (Fig. 4) and the three-dimensional structure among them is very similar. For each molecule, equivalent C^α atoms of four domains are superimposed by the least-squares method. The r.m.s.d. value for the superposition of two domains is in the range 0.5–0.8 Å. The polypeptide chain of each domain is folded into a compact globule which has four disulfide linkages and five or six β-turns. The three-dimensional structure of each domain is maintained by the disulfide bonds and many NH···O hydrogen bonds linking peptide groups (Fig. 5). Most of contacts between side-chain groups are found within a domain. There are no significant contacts between *B* and *C* (Fig. 4). Between domains *A* and *B*, three pairs of residues, Glu11···Asn57, Asn15···His59 and Gly38···Asn57, are involved in the contacts less than 3.3 Å. On the other hand, domains *C* and *D* have contacts Asn101···Tyr145, Gly124···Asn143 and Thr128···Lys130.

Temperature factors of molecule 1 are plotted against the residue number (Fig. 6). The profile of the temperature-factor plot is similar among domain *A*, *B*, and *C* while the temperature factor of domain *D* is higher than that of the other domains. A relatively high temperature factor is observed in the first ten residues of each domain. Therefore, four domains have similar characteristics not only in their three-dimensional structure but also in their thermal motion.

Dimer structure

As shown in Fig. 7, two subunit molecules are related by the pseudo twofold axis to form a dimer structure. The structure of the main-chain folding is nearly identical between the two subunit molecules. The equivalent C^α atoms between the two molecules are superimposed by the least-squares method. The transformation that optimizes the superposition of the two molecules is as

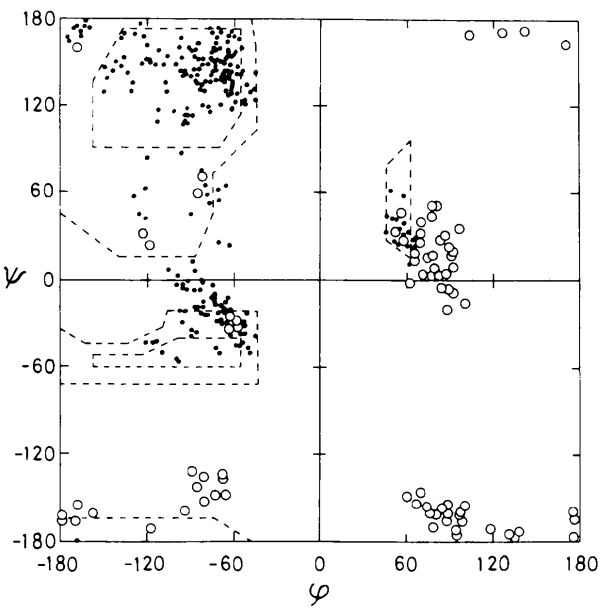

Fig. 2. Ramachandran plot for molecules 1 and 2 (φ, ψ angles in °). Glycine residues are denoted by open circles

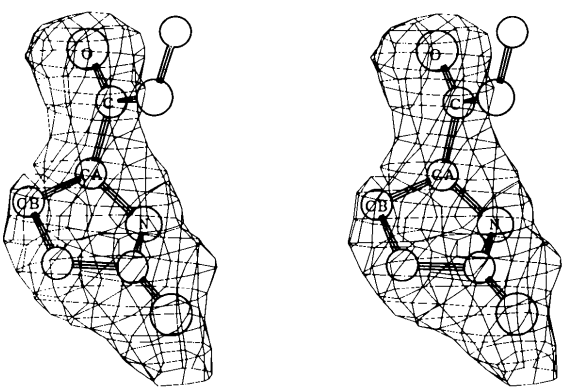

Fig. 3. Stereoview of the $F_o - F_c$ omit map for the N-terminal residue (Pgl1) of molecule 1. Contours are plotted at 2σ level.

follows,

$$X' = 0.3434X + 0.0025Y - 0.9392Z - 3.5963$$
$$Y' = 0.0024X - 1.0000Y - 0.0018Z + 90.4423$$
$$Z' = -0.9392X - 0.0016Y - 0.3434Z - 4.8551.$$

The psuedo twofold axis is nearly parallel to the *ac* plane and makes an angle of 35.0° with the *a* axis. The center of the dimer is 3.0 Å shifted from the *b* axis. The r.m.s. difference is 0.36 Å for C^{α} atoms and 2.72 Å for all atoms. Except for the two residues at the C-terminus, the positional difference for each C^{α} atom is less than 1.0 Å.

Four domains in a molecule are arranged to form a deep crevice between *A* and *B* domains and *C* and *D* domains. Two domains, *A* and *B*, are located in the crevice of the other molecule (Fig. 7) and domain *A* is placed in between domains *A* and *D* of the other molecule. Although each domain has only few inter-domain contacts within a molecule (Fig. 4), many short contacts are found between the two subunit molecules. The fact that most of the intermolecular contacts are found between main-chain peptide groups indicates the close contact of the two subunit molecules. Intermolecular contacts less than 3.3 Å involving side-chain groups are only those of Asn15···Asn58, Tyr73···Glu115 and Arg84···Asp86. Water molecules also contribute to stabilize the dimer structure and 20 water molecules are involved in a water-mediated hydrogen-bond bridge between the two molecules.

The average temperature factor of molecule 1 is 16.1 $Å^2$ for the peptide group and 17.1 $Å^2$ for all atoms while the corresponding values for molecule 2 are 16.2 and 17.5 $Å^2$. The temperature-factor difference for each residue is mostly less than 10 $Å^2$ and the r.m.s. difference is 5.0 $Å^2$.

Comparison with WGA 1 and WGA 2

WGA1 and WGA2 crystallize in two forms in space group $C2$ or $P2_1$ and their X-ray structures have been determined for the $C2$ crystal (Wright, 1987, 1989). In these structures, the asymmetric unit contains two molecules and each molecule forms a dimer structure with a symmetry-related mate on the twofold axis. Therefore, the $C2$ crystal contains two independent dimer units. On the other hand, in the present WGA3 structure, the asymmetric unit contains one dimer unit which has pseudo twofold symmetry. To compare with the structure of WGA1 and WGA2, the dimer unit of WGA3 was superimposed on each dimer unit of those isolectins. The r.m.s.d. values for equivalent C^{α} atoms are 0.37 and 0.39 Å for WGA1 and 0.38 and 0.35 Å for WGA2. These values, which are comparable with the difference between WGA1 and WGA2 (0.29 and 0.30 Å, respectively), indicate that in spite of the different crystal symmetry the dimer structure of WGA3 is nearly identical to those of WGA1 and WGA2.

WGA2 differs in eight residues from WGA1 and in seven residues from WGA2. Gly9 and Lys53 in WGA3 are Asn and Ala, respectively, in WGA1 and WGA2. These residues in three WGA's have no contacts with other residues. In WGA3, water molecules are hydrogen bonded to O(Gly9) of molecule 1, O(Gly9), N(Gly9), and NZ(Lys53) of molecule 2. The hydroxyl group of Thr56 in WGA1 has contacts with NZ(Lys44) and OD1(Asn57). In contrast, the 56th residue in WGA2 and WGA3 is Pro, of which a carbonyl O atom forms a hydrogen bond with a water molecule. The side-chain group of His59 in WGA2 and WGA3 forms a contact with OD1 of Asn15 while NE2 of Gln59 in WGA1 is hydrogen bonded to O(Pro56). The Tyr66 residue of WGA1 is replaced by His in WGA2 and WGA3. In

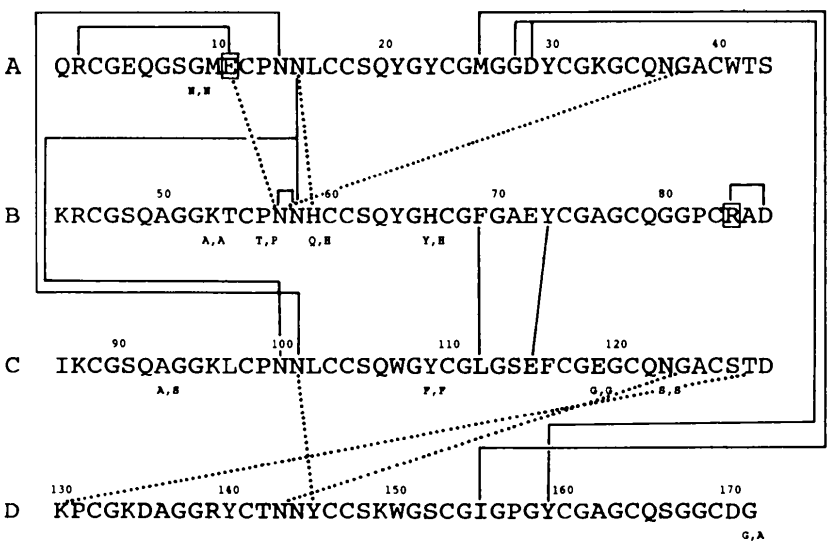

Fig. 4 Amino-acid sequence of WGA3. Amino-acid residues of WGA1 and WGA2, which differs from WGA3, are given below the corresponding residues. Each row corresponds to a domain designated from *A* to *D*. Interaction between residues in adjacent domains within each molecule are represented by dotted lines. Residues involved in subunit–subunit interactions at the dimer axis (self-interaction) are shown as boxed in, and those that make contact in the subunit–subunit interface are interconnected by solid lines.

WGA3, the imidazolyl group of His66 forms hydrogen bonds with water molecules. The hydroxyl group of Ser93 in WGA2 is hydrogen bonded to O(Gln92) while the corresponding residue is Ala in WGA1 and WGA3. In WGA1 and WGA2, Gly119 and Ser123 have no contacts with other residues. On the other hand, these residues are Glu119 and Asn123 in WGA3. The carboxyl group of Glu119 forms hydrogen bonds with Gln106 and water molecules. The side-chain group of Asn123 is in hydrogen-bonding contact with O(Lys96), N(Cys98), and water molecules. The carboxyl terminal residue is Ala in WGA2 and Gly in WGA1 and WGA3. This residue is highly flexible and has no significant contact with other residues.

The comparison among three isolectins reveals the change in interresidue contact caused by the amino-acid replacement. It is noteworthy that the amino-acid residues involved in the replacement have contacts with other residues only within each molecule and not with residues in the other molecule of the dimer unit. This is the reason that the dimer structure does not significantly differ among three isolectins. Since the sugar-binding sites are located on the interface of subunits, the change of the subunit–subunit interaction may have a disastrous effect on the sugar-binding activity even if the structure of each subunit molecule does not change.

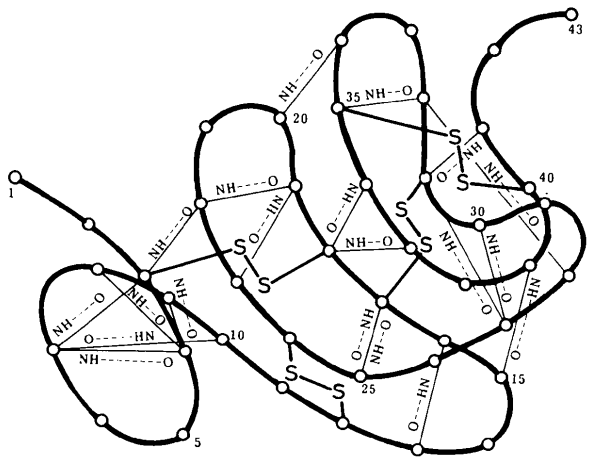

Fig. 5. Schematic drawing of the backbone structure of domain *A*.

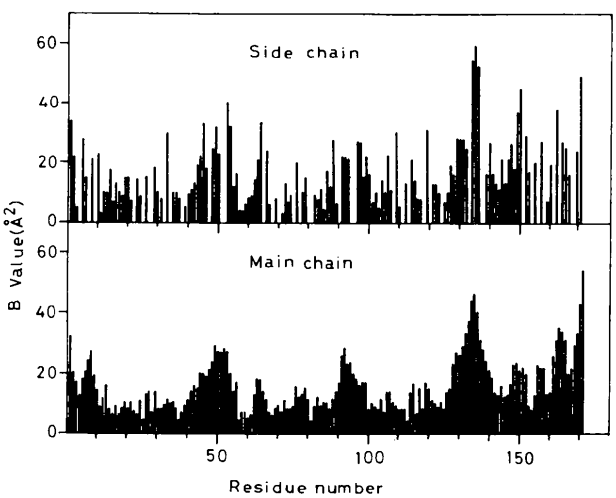

Fig. 6. Plot of temperature factors averaged for main-chain peptide and side-chain groups in molecule 1 against residue number.

Sugar-binding sites

WGA's have four sugar-binding sites, two primary sites and two secondary sites, located at the interface of two subunit molecules (Wright, 1990) The structure of a primary site of WGA1 is superimposed on that of WGA3. Fig. 8 shows that not only the main-chain structure but also the side-chain conformation at the sugar-binding site is well conserved between the two isolectins. The 66th residue is Tyr in WGA1 and His in WGA2 and WGA3 while the Tyr109 in WGA3 is Phe in WGA1 and WGA2. In the binding of sugars, that aromatic side-chain group of these residues faces the

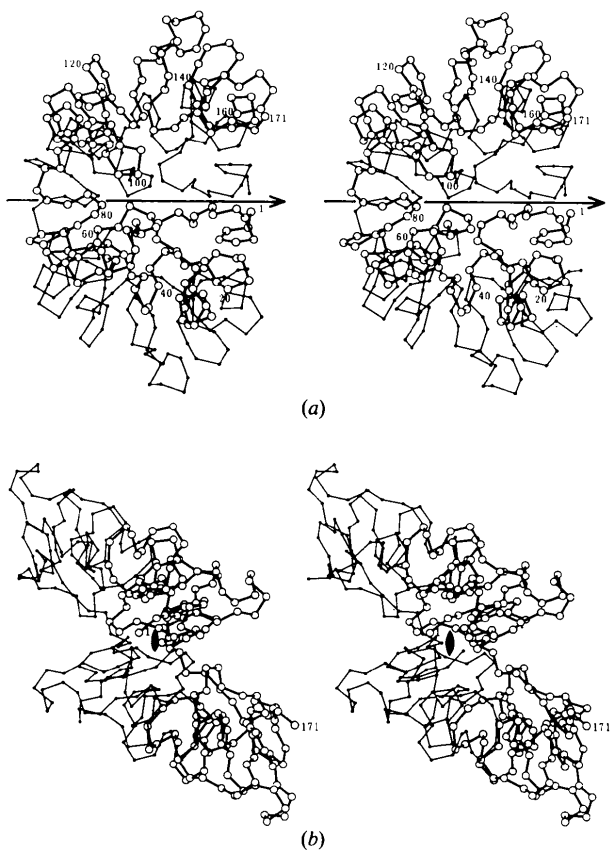

Fig. 7. Stereoviews of the dimer structure (*a*) and its side view (*b*). Molecule 2 is shown by thin lines. An arrow and a graphical symbol (◗) represent the pseudo twofold axis.

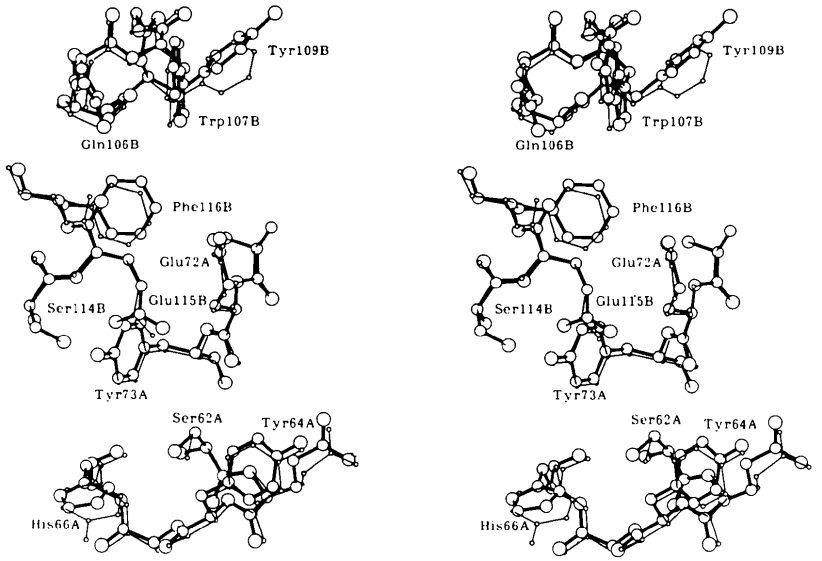

Fig. 8. Stereoview of the structure of the primary sugar-binding site. The corresponding region of WGA1 is superimposed and shown by thin lines. Amino-acid residues in molecules 1 and 2 are designated as *A* and *B*, respectively.

pyranose ring in van der Waals contact. Since the aromatic side-chain group of these residues is essential for sugar binding, the replacement may be possible only among aromatic amino-acid residues. The side-chain conformation is so flexible that the aromatic groups can move to have favourable contacts with sugar residues.

Crystal packing

Crystal structure viewed along the *b* axis is shown in Fig. 9. The dimer unit is located with its center near the *b* axis and is stacked along the twofold screw axis. The packing structure clearly explains the pseudo $C222_1$ symmetry of the crystal. The *a* and *c* axes form a lozenge and the pseudo twofold axis is parallel to its longer diagonal which is normal to the shorter one and the *b* axis. To fill the $C222_1$ symmetry, the pseudo twofold axis, which is 3.0 Å laterally shifted parallel to the *ac* plane, should cross with the *b* axis.

The molecular packing is similar to that of the *C*2 crystal of WGA1 and WGA2. In WGA3, intermolecular contacts between dimer units along the *b* axis are found between the domains *C* and *D*. Interdimer contacts, N(Gly95)···O(Gln165) and N(Val140)···O(Gly137), are also observed in the *C*2 crystal of WGA2. The hydroxyl group of Tyr21 and Tyr23 in molecule 1 forms a hydrogen bond with the carboxyl group of Asp86 of molecule 1 and Glu72 of molecule 2, respectively, of the next dimer unit along the *a* axis. The Tyr21 and Tyr23 residues in molecule 2 form similar hydrogen bonds with the adjacent dimer unit along the *c* axis. These hydrogen-bonding contacts are also observed in WGA2. Such similarities in the intermolecular contacts indicate that the packing structure of the $P2_1$ crystal is correlated with the packing of the *C*2 crystal by a small rotation and translation of the dimer unit, which maintains some of the interdimer hydrogen bonds.

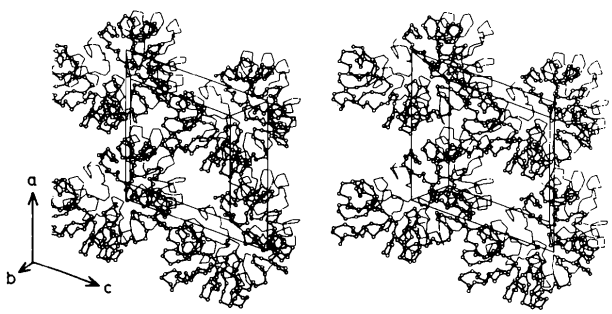

Fig. 9. Crystal structure viewed along the *b* axis. The C^{α} atoms of molecule 1 is shown by circles and molecule 2 is drawn with thin lines.

References

BRÜNGER, A. T. (1990). *Acta Cryst.* A**46**, 46–57.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.

GATELLIER, P., GRIVET, J. PH. & DELMOTTE, F. (1987). *Biochem. Biophys. Acta*, **916**, 428–436.

KRONIS, K. A. & CARVER, J. P. (1982). *Biochemistry*, **21**, 3050–3057.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

NAGAHORA, H., ISHIKAWA, K., NIWA, Y., MURAKI, M. & JIGAMI, Y. (1992). *Eur. J. Biochem.* **210**, 989–997.

NAGATA, Y. (1985). *Agric. Biol. Chem.* **49**, 535–536.

RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.

RICE, R. H. & ETZLER, M. E. (1975). *Biochemistry*, **14**, 4093–4099.

SHARON, N. & LIS, H. (1989). *Lectins*. New York: Chapman and Hall.

SMITH, J. J. & RAIKHEL, N. V. (1989). *Plant Mol. Biol.* **13**, 601–603.

WRIGHT, C. S. (1987). *J. Mol. Biol.* **194**, 501–529.
WRIGHT, C. S. (1989). *J. Mol. Biol.* **209**, 475–487.
WRIGHT, C. S. (1990). *J. Mol. Biol.* **215**, 635–651.
WRIGHT, C. S. & JAEGER, J. (1993). *J. Mol. Biol.* **232**, 620–638.
WRIGHT, C. S., KEITH, C., LANGRIDGE, R., NAGATA, Y. & BURGER, M. M. (1974). *J. Mol. Biol.* **87**, 843–846.