

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/227801036>

# Differences in hydropathic properties of ligand binding at four independent sites in wheat germ agglutinin-oligosaccharide crystal complexes

ARTICLE *in* PROTEIN SCIENCE · AUGUST 1996

Impact Factor: 2.85 · DOI: 10.1002/pro.5560050803

---

CITATIONS

45

---

READS

16

## 2 AUTHORS, INCLUDING:



Glen Kellogg

Virginia Commonwealth University

145 PUBLICATIONS 3,481 CITATIONS

[SEE PROFILE](#)

# Differences in hydropathic properties of ligand binding at four independent sites in wheat germ agglutinin–oligosaccharide crystal complexes

CHRISTINE SCHUBERT WRIGHT AND GLEN E. KELLOGG

Department of Biochemistry and Molecular Biophysics and Department of Medicinal Chemistry,  
Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298

(RECEIVED November 14, 1995; ACCEPTED May 21, 1996)

## Abstract

The binding interactions of *N*-acetyl-*D*-neuraminic acid and *N,N'* diacetyl-chitobiose (*GlcNAc-β-1,4-GlcNAc*), observed in crystal complexes of wheat germ agglutinin (WGA) at four independent sites/monomer, were analyzed and compared with the modeling program HINT (Hydropathic INTeractions). This empirical method allows assessment of relative ligand binding strength and is particularly applicable to cases of weak binding where experimental data is absent. Although the four WGA binding sites are interrelated by a fourfold sequence repeat (eight sites/dimer), similarity extends only to the presence of an aromatic amino acid-rich pocket and a conserved serine. Strong binding requires additional interactions from a contacting domain in the second subunit. Ligand positions were either derived from crystal structures and further optimized by modeling and molecular mechanics, or from comparative modeling. Analysis of the overall HINT binding scores for the two types of ligands are consistent with the presence of two high-affinity and two low-affinity sites per monomer. Identity of these sites correlates well with crystal structure occupancies. The high-affinity sites are roughly equivalent, as predicted from solution binding studies. Binding scores for the low-affinity sites are weaker by at least a factor of two. Quantitative estimates for polar, nonpolar, and ionic interactions revealed that H-bonding makes the largest contribution to complex stabilization in the seven bound configurations, consistent with published thermodynamic data. Although the observed nonpolar interactions are small, they may play a critical role in orienting the ligand optimally.

**Keywords:** HINT modeling; saccharide binding affinities; wheat germ agglutinin

Many biological recognition processes are mediated by specific interaction of carbohydrates with protein receptors. Cell surface oligosaccharides are complex structures that possess a high degree of conformational flexibility. Yet certain conformations are preferentially recognized by their specific receptors. In recent years considerable strides have been made toward a better understanding of factors that determine oligosaccharide recognition through X-ray structure determination of lectin–oligosaccharide complexes and conformational analysis of bound saccharides in terms of their potential energy surfaces (Imberty et al., 1993; Pérez et al., 1994).

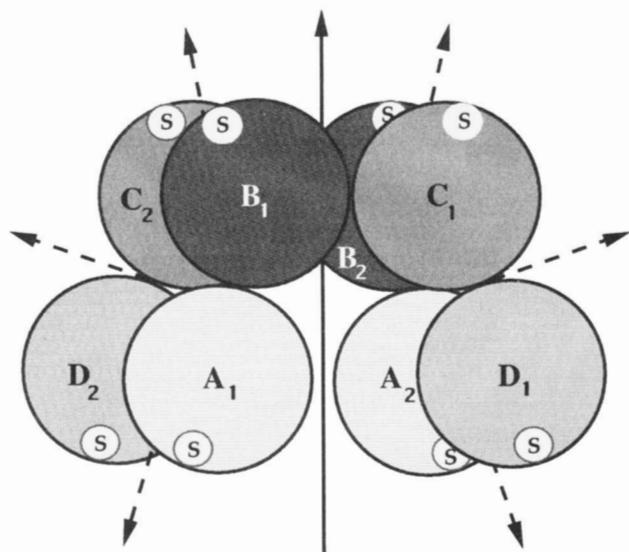
Lectins are a class of proteins well known for their selectivity in binding saccharides (Goldstein & Poretz, 1986; Sharon & Lis, 1989). They have been studied widely as tools to detect subtle changes in cell surface structure. For instance, wheat germ ag-

glutinin was the first lectin that was shown to bind more strongly to tumor cells than normal cells (Aub et al., 1963; Burger, 1969). This lectin, specific for terminal *N*-acetyl-*D*-neuraminic acid and *N*-acetyl-*D*-glucosamine, belongs to a highly conserved family of chitin-binding lectins (*Gramineae*), and is one of the most widely studied and best-characterized lectins (Lis & Sharon, 1981; Raikhel & Wilkins, 1984; Goldstein & Poretz, 1986; Wright, 1987). These lectins are unique compared with lectins from other families, because they possess multiple binding sites due to an internal fourfold structure repeat (Raikhel & Wilkins, 1984; Wright, 1987, 1992). The WGA monomer consists of four similar disulfide-rich 43-amino acid residue domains (A,B,C,D) arranged in tandem, and dimerizes in a “head to tail” fashion, forming an extensive monomer/monomer interface. This interface accommodates eight independent saccharide binding sites (four unique sites) between contacting domains (see Fig. 1), characterized by a cluster of 2–3 aromatic amino acids (see below).

Extensive efforts have been made using a range of solution techniques and X-ray crystallography to characterize the four types of binding sites structurally and thermodynamically. Results reported from several solution-binding studies have sug-

Reprint requests to: Christine Schubert Wright, Department of Biochemistry and Molecular Biophysics, Medical College of Virginia, Box 980614, Virginia Commonwealth University, Richmond, Virginia 23298.

Abbreviations: WGA, wheat germ agglutinin; GlcNAc, *N*-acetyl-*D*-glucosamine; NeuNAc, *N*-acetyl-*D*-neuraminic acid; NeuLac, *N*-acetylneuraminy lactose.



**Fig. 1.** Schematic representation of the WGA dimer. Domains are shown as large shaded circles and labeled A<sub>1</sub>, B<sub>1</sub>, C<sub>1</sub>, D<sub>1</sub>, etc. The position of the molecular twofold axis is indicated by an arrow. Dotted arrows represent the two types of pseudo-twofold axes generated in the dimer interface between domains of different dimers. "S" refers to the aromatic sugar-binding pocket.

gested the presence of two equivalent sites/monomer (four per dimer) (Privat et al., 1974b; Kronis & Carver, 1985; Baines et al., 1992). In contrast, combined evidence from crystal structures of several types of oligosaccharide complexes revealed that all four unique sites are functional (Wright, 1980, 1984, 1990, 1992). Presumably, two of these have affinities too weak to be detectable in solution. Simultaneous occupancy of all eight sites has not been observed in a single crystal structure because sugar binding competes with crystal packing in the different types of lattices. However, a recent NMR study in which chemically induced dynamic nuclear polarization (CIDNP) signals from the aromatic amino acids were monitored as a function of ligand binding (Siebert et al., 1996, in prep.), supports our premise that all eight sites are functional binding sites. These studies show almost complete signal suppression in the aromatic region at the highest concentration of (GlcNAc)<sub>4</sub> (1 mM). However, due to complexity of this region of the spectrum, it has not been possible to make definitive amino acid assignments to individual resonances.

Although it may be reasonable to assume that sites highly occupied in crystal structures correspond to those detectable in solution, the degree of site occupancy may not reflect binding strength, because crystal packing interactions are known to interfere with accessibility of several sites (Wright, 1992). Thus, at this time, it has neither been possible to identify unambiguously the high-affinity binding sites, nor to obtain a quantitative measure of binding affinities for the two weak sites. Low-affinity binding ( $K_a < 1 \text{ mM}$ ) of mono- or disaccharides is difficult to detect by any of the conventional solution techniques. However, their importance in biological recognition has been recognized widely. Moreover, it is believed that large affinity increases are possible in biological milieus where ligands or receptors are multivalent. For instance, multisite association between lectins and their receptors is a common phenomenon and constitutes a mechanism by which selective and strong bind-

ing can be achieved. This may involve crosslinkage. Structural evidence for such organized crosslinked networks has emerged recently from several lectin crystal structures complexed with multivalent glycopeptides (Weis et al., 1992; Wright, 1992; Bourne et al., 1994; Sharon, 1994; Dessen et al., 1995). In addition, there are examples of lectins with multiple weak, but conserved, binding sites (Hester et al., 1995). Binding affinities at each of these sites vary depending on the extent to which they are involved in intersubunit association.

To this end, we have attempted using molecular cloning techniques in an effort to delineate binding affinities among the four WGA sites (Rice, 1994). However, isolated single domains were found to exist only as monomers, eliminating the possibility of recreating high-affinity hybrid sites. In the present study, a theoretical modeling approach was taken to determine the relationship among the four structurally similar WGA binding sites. The computer program HINT (Hydropathic INTeractions) (Kellogg et al., 1992) was used to obtain empirical estimates of ligand binding strength and to analyze the specific interactions (polar, hydrophobic, or ionic) that differentiate and stabilize bound ligand in each of the four binding environments. This program was developed to investigate the nature of protein/ligand (or drug) interactions in macromolecules of known three-dimensional structure (Wireko et al., 1991) and carefully modeled complexes. The procedure involves optimization of X-ray determined ligand positions through additional modeling and inclusion of hydrogen atoms. Calculations were made on the two specific ligands, NeuNAc and GlcNAc- $\beta$ 1,4-GlcNAc (GlcNAc-2), for which atomic positions are available from crystal structures (Wright, 1980, 1984, 1990; Wright & Jäger, 1993). In interpreting the binding scores for individual interactions and for each isolated ligand complex, two questions were addressed: (1) Which of the four sites are the two high-affinity sites? (2) Can HINT provide reasonable quantitative estimates for binding strength of weakly bound ligands?

#### Description of the WGA binding sites

WGA contrasts with other well characterized lectins (*leguminosae*) (Sharon, 1993) in that it possesses more than one carbohydrate binding site per monomer. Each of the four domains that constitute the monomer is a carbohydrate-recognition domain (CRD). Their binding sites are composed of two binding regions, a shallow surface pocket characterized by three quasi-conserved aromatic amino acids, and one conserved serine on one domain and a nonconserved region that consists of one or two polar residues, on the contacting domain of the second monomer. In the nomenclature adopted here, each unique site is referred to by the domain that contributes the aromatic pocket. Because twofold related sites are equivalent, there are two of each type of site present in the dimer: A<sub>1</sub> or A<sub>2</sub>, B<sub>1C2</sub> or B<sub>2C1</sub>, C<sub>1B2</sub> or C<sub>2B1</sub>, D<sub>1A2</sub> or D<sub>2A1</sub>. The subscripts designate the domain that contributes the polar region on the opposing monomer. Because the D-domain has no polar region (see Table 1), the binding site of domain A<sub>1</sub> or A<sub>2</sub> consists of an aromatic pocket alone.

The aromatic pocket forms the main portion of the binding site and may suffice for minimal binding. Rice (1994) has used isothermal titration calorimetry techniques to measure binding constants for the association of  $\beta$ 1,4-linked oligomers of GlcNAc ((GlcNAc)<sub>3</sub> and (GlcNAc)<sub>4</sub>) to a single domain of

**Table 1.** Residues in WGA binding pockets<sup>a</sup>

Site (domains involved)	Aromatic pocket Residue positions (Mon 1)				Polar site Residue positions (Mon 2)	
	19	21	23	30	28	29
A1-D2	Ser	Tyr	Tyr	Tyr	Pro	Gly
B1-C2	Ser	Tyr	Tyr/His	Tyr	Ser	Glu
C1-B2	Ser	Trp	Phe	Phe	Ala	Glu
D1-A2	Ser	Trp	Ser	Tyr	Gly	Asp

<sup>a</sup> Amino acids shown in bold type indicate their involvement in hydrogen bonding.

WGA (domB), prepared by recombinant techniques. The values obtained were roughly 10 times lower compared with published binding constants, which are in the range of  $1.2\text{--}8.3 \times 10^{-5}$  M for the four-domain WGA molecule (Lotan & Sharon, 1973; Nagata & Burger, 1974; Privat et al., 1974a; Baines et al., 1992). Clearly, high-affinity binding requires auxiliary contacts from across the dimer interface, particularly in the case of *N*-acetylated sialosides (Wright, 1992).

#### The HINT model

The genesis of HINT is based on the suggestion that hydrophobic fragment constants, reduced to atomic values with inherent bond, ring, chain, branching, and proximity factors, could be used to evaluate interactions between small molecules and large molecules (Abraham & Leo, 1987). The HINT model for biomolecular interactions includes all experimentally accepted non-covalent hydrophobic and polar interactions collectively referred to as hydropathy. All of these are related to solvent partitioning phenomena because the dissolution of a "ligand" in a mixed solvent system involves the same fundamental processes and atom-atom interactions as biomolecular interactions within or between proteins and ligands. Because polar species are clearly differentiated from hydrophobic species, the nature of interactions between an agent and its receptor can be postulated, as can the general environment of the receptor site. This empirical approach represents an important difference from models based on molecular mechanics approaches, because hydrophobicity, which is a free energy-like property, includes a contribution from entropy usually ignored by most models. Likewise, derivation of hydrophobic atom constants for each atom in a drug or protein provides localized parameterized thermodynamic information. These constants, which are the key parameters for the HINT model, thus can be information-rich thermodynamic parameters, and also encode solvent effects, because hydrophobicity/hydropathy is defined in terms of solubilities.

In recent studies, HINT has been used to examine in detail a number of biomolecular systems where experimental information is incomplete. For cases where a variety of ligands are proposed to bind at the same site, HINT results have been shown to correlate well with solution binding measurements. In particular, studies of hemoglobin allosteric modifiers (Wireko et al., 1991; Abraham & Kellogg, 1994) and nonnucleoside HIV-1 reverse transcriptase inhibitors (Gussio et al., 1996) have been

reported. Although this type of simple correlation is valuable, the insight obtained by examining the specific interactions leading to the ligand binding is of particular value. HINT examination of the noncovalent interactions at hemoglobin subunit interfaces (Abraham et al., 1996, in prep.) has clarified the role of hydrophobic-hydrophobic interactions in the allosteric mechanism and has indicated good agreement between HINT subunit association scores and experimental measurements of  $\Delta G$ . In an extensive study of the nucleotide sequence selectivity of adriamycin intercalation, where 64-base pair quadruplets were modeled individually and analyzed with HINT (Fornari et al., 1996), a series of selectivity rules for optimized binding could be derived. This latter case is particularly cogent for the current study. There is no practical experimental manner to prepare samples and analyze the binding for such a large number of "receptors" within a single macromolecule. The use of the HINT model in this study enabled a rational description of this system. Although the results were found to agree with the limited available experimental data, this study has led to a much deeper understanding of the factors influencing drug intercalation into nucleotides, not possible previously.

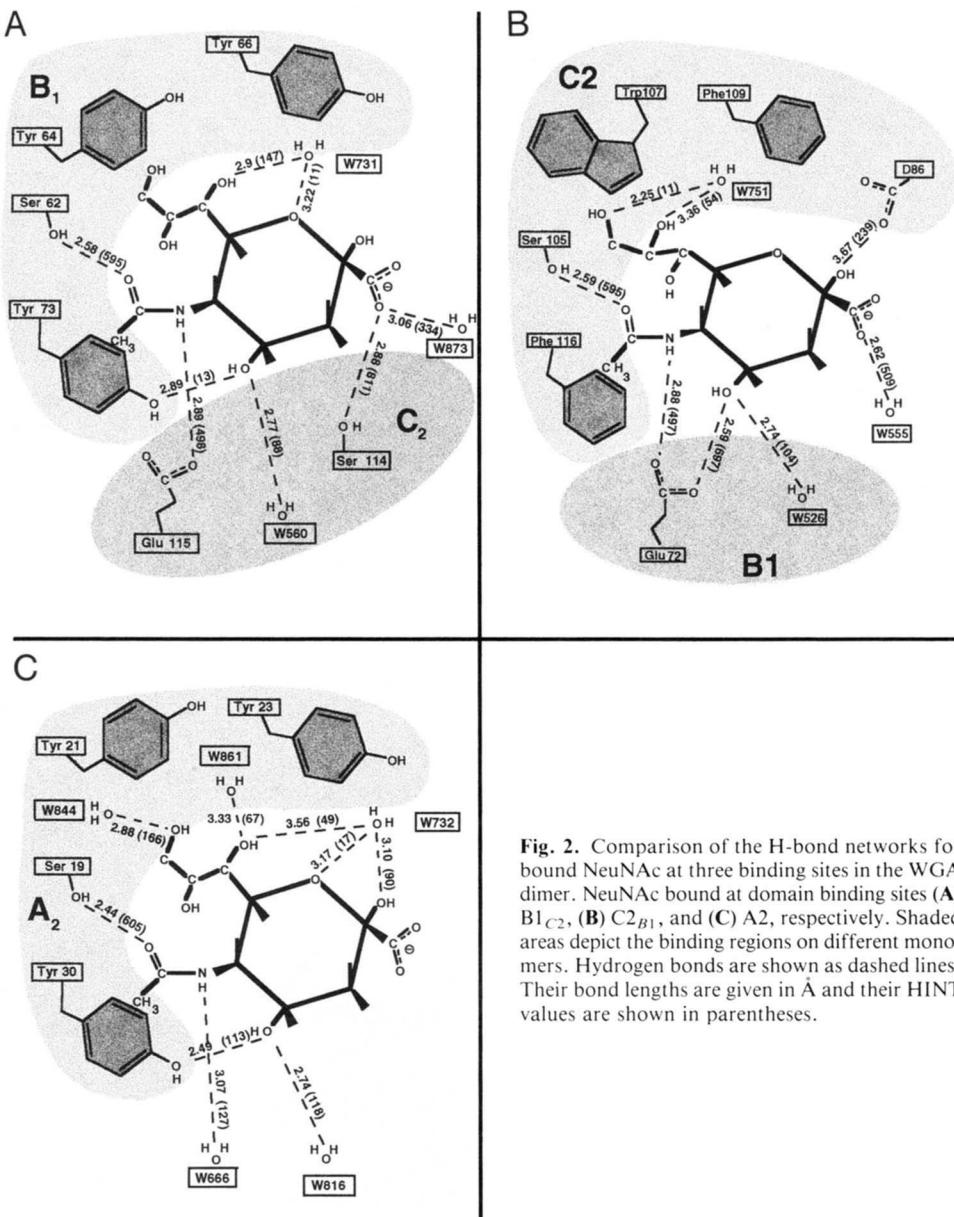
#### Results and discussion

The binding interactions observed in crystal complexes of WGA for the terminal nonreducing NeuNAc residues of *N*-acetyl-neuramyl lactose and the T5 sialoglycopeptide of glycophorin A (NeuNAc- $\alpha$ 2,3-Gal- $\beta$ 1,3-( $\alpha$ 2,6-NeuNAc)GalNAc) (Wright, 1990, 1992) were analyzed at three different sites, and those for bound GlcNAc-2 (Wright, 1980) at four different sites. Hydrogen bonding patterns for these binding configurations examined are illustrated schematically in Figures 2 and 3. The magnitudes of the HINT scores shown for the most important H-bonds in Figures 2 and 3 are very sensitive to interatomic distances and chemical properties of the interacting atoms (or groups of atoms). In those cases involving water, the HINT value listed reflects the net score between all water atoms and, for instance, those of a ligand -OH- or -NH- group. Table 2 compares the crucial H-bond contact distances between ligand and amino acid side chains in the four different binding environments. This ver-

**Table 2.** H-bond lengths ( $\text{\AA}$ )

Ligand bound	Site	Ligand functional groups				
		C=O	NH	3-OH	4-OH	COO <sup>-</sup>
<b>WGA crystal soaks (native C2 crystals)</b>						
(GlcNAc) <sub>2</sub>	B1 <sub>C2</sub>	2.68	2.53	2.80	3.43	
	C2 <sub>B1</sub> <sup>a</sup>	2.62	2.66		3.20	
	D1 <sub>A2</sub> <sup>a</sup>	2.71	3.18	2.82/2.97		
	A2 <sup>a</sup>	2.56		2.49		
NeuNAc (NeuLAC)	B1 <sub>C2</sub>	2.85	2.66		3.11	2.77
<b>WGA1-T5 co-crystals (P2<sub>1</sub>2<sub>1</sub>2 complex)</b>						
NeuNAc	B1 <sub>C2</sub>	2.58	2.89		2.89	2.88
	C2 <sub>B1</sub>	2.59	2.88		2.59	
	A2	2.44			2.49	

<sup>a</sup> Ligand was modeled.

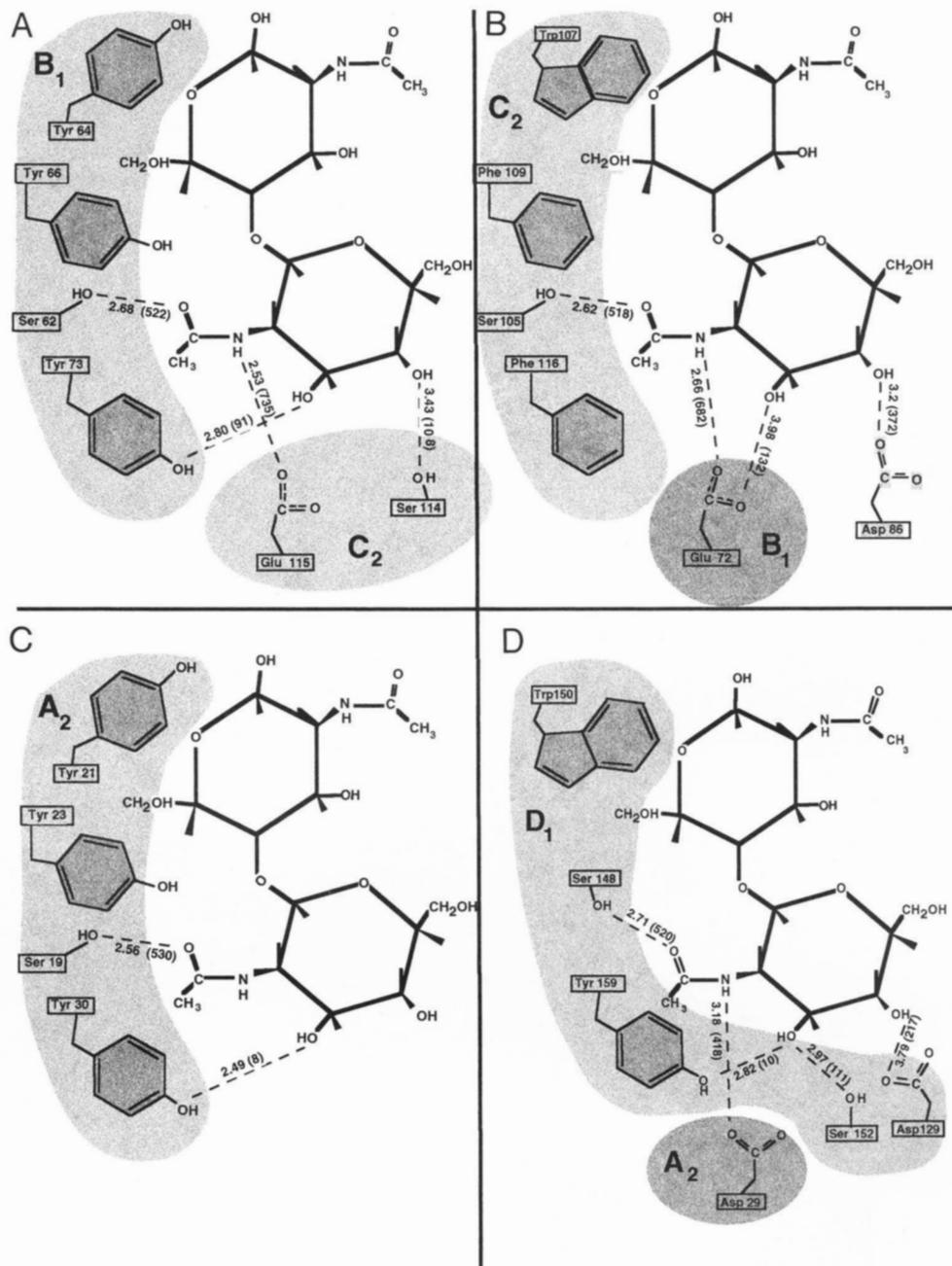


**Fig. 2.** Comparison of the H-bond networks for bound NeuAc at three binding sites in the WGA dimer. NeuAc bound at domain binding sites (**A**)  $B1_{C2}$ , (**B**)  $C2_{B1}$ , and (**C**)  $A2$ , respectively. Shaded areas depict the binding regions on different monomers. Hydrogen bonds are shown as dashed lines. Their bond lengths are given in Å and their HINT values are given in parentheses.

ifies that, in those cases in which the saccharide was modeled (GlcNAc-2 bound at sites  $C2_{B1}$  and  $A2$ ), the orientation of bound ligand is closely comparable to those derived from crystallographic data. As apparent, analogous hydrogen bond lengths agree to within 0.3 Å (see Table 2).

The overall HINT binding scores for each bound configuration are listed in Table 3. These represent net values averaged over the positive and negative contributions of all interactions between sugar ligand and its binding site. An experimental error of  $\pm 200$  was estimated based on the uncertainty in positioning the ligand optimally (0.2–0.3 Å). Comparison of the different types of summed contributions indicates that hydrogen bonding dominates in all cases, in good agreement with the characteristic negative values reported for the enthalpy and entropy changes upon binding of these two types of ligands to WGA by Kronis and Carver (1984) and Baines et al. (1992). The

net scores shown in Table 3 consist of summed positive and negative contributions from all hydrogen bond contacts (some of which are not shown in Figs. 2, 3), and include those involving water in the case of the highly refined sialoglycopeptide complex. The magnitudes of the HINT scores for the most important H-bonds, as shown in Figures 2 and 3, confirm that the *N*-acetyl group is responsible for most of the binding, as it is involved in two strong hydrogen bonds. In particular, it is worth noting that the HINT score for the H-bond between its -CO-group and the invariant Ser-OH at domain position 19 is closely comparable in all seven binding configurations, and may thus represent the strongest hydrogen bond. The HINT data also show that the presence of a carboxylate group in the binding site is powerful by virtue of its H-bonding and ionic properties. Thus, the -NH- group contributes a strong ligand contact in sites  $B1_{C2}$  and  $C2_{B1}$  through interaction with a glutamic acid residue



**Fig. 3.** H-bond networks for bound GlcNAc-2 at the four unique WGA binding sites. GlcNAc-2 bound at sites (**A**) B<sub>1C2</sub>, (**B**) C<sub>2B1</sub>, (**C**) A<sub>2</sub>, and (**D**) D<sub>1A2</sub>, respectively. Shaded areas depict the binding regions on different monomers. Hydrogen bonds are shown as dashed lines. Their bond lengths are given in Å and their HINT values are shown in parentheses.

at domain position 29, Glu 115, and Glu 72, respectively. Small inconsistencies in its HINT score in these two environments may be attributed to involvement of this carboxylate group in other contacts (see Figs. 2B, 3B). In the weakly occupied sites (A<sub>2</sub> and D<sub>1A2</sub>), this strong polar contact is either weak or absent. In the case of the A-domain site, which completely lacks a polar region, binding affinities should resemble those measured by Rice (1994) for single-domain proteins. All hydrophobic contributions are small, but strongest in the C-domain site (C<sub>2B1</sub>), where the aromatic side chains have a more nonpolar character.

#### NeuNAc binding

Comparison of the overall HINT binding scores for the three different bound configurations of NeuNAc revealed that under all conditions (with and without water), sites B<sub>1C2</sub> and C<sub>2B1</sub> yielded the largest values (1,273 and 1,346), suggesting that these sites are the high-affinity sites of WGA. Both sites were observed to be highly occupied in the asymmetric WGA1/T5 crystal complex, where they participate in crosslinkage of the bivalent tetrasaccharide. Although the C-domain site had earlier been

**Table 3.** HINT binding parameters

Ligand bound	Site	Overall HINT-score <sup>a</sup>	Contributions from		
			H-bonds <sup>b</sup>	Hydrophob.	Acid/base
<b>WGA crystal soaks</b>					
(GlcNAc) <sub>2</sub>	B1 <sub>C2</sub>	1,136	1,689 (3)	442	774
	C2 <sub>B1</sub> <sup>c</sup>	1,297	1,645 (3)	521	953
	D1 <sub>A2</sub> <sup>c</sup>	725	1,058 (4)	315	757
	A2 <sup>c</sup>	657	614 (2)	403	708
NeuNAc (NeuLAc)	B1 <sub>C2</sub>	1,516	1,999 (4)	402	894
<b>WGA1-T5 co-crystals</b>					
NeuNAc	B1 <sub>C2</sub>	1,273	1,938 (4)	398	869
	C2 <sub>B1</sub>	1,346	2,149 (3)	548	1,272
	A2	623	715 (2)	454	568
NeuNAc with H <sub>2</sub> O	B1 <sub>C2</sub>	677	2,493 (7)	390	983
	C2 <sub>B1</sub>	1,012	2,828 (6)	555	1,063
	A2	-20	1,349 (7)	457	752

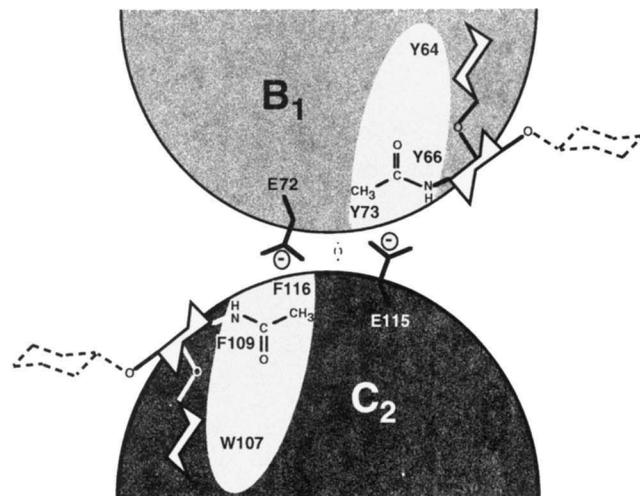
<sup>a</sup> Estimated error  $\pm 200$  (see text).<sup>b</sup> Numbers in parentheses indicate the total number of well-oriented H-bonds ( $<3.4 \text{ \AA}$ ), including waters with HINT scores  $>50$ .<sup>c</sup> Ligand was modeled.

considered nonfunctional due to its uniquely different aromatic character (Wright, 1984), site B1<sub>C2</sub> is the best characterized site crystallographically. It is the only site that is completely accessible in native WGA crystals. Both types of sugars (GlcNAc and NeuNAc) were observed to bind here in all three crystal complexes examined. This site has a highly favorable binding environment, allowing ligand stabilization through 3–4 hydrogen bond interactions (see Table 2; Figs. 2A, 3A). Two of these come from the polar region on domain C2. The discrepancy in the two NeuNAc HINT scores for site B1<sub>C2</sub> of the two different types of NeuNAc-oligosaccharide complexes (NeuLac and WGA1/T5 co-crystals) (see Tables 2, 3) may be a result of linkage specificity. The larger value for the NeuLac (NeuNAc- $\alpha$ 2,3-Gal- $\beta$ 1,4-Glc) complex (1,516) suggests that  $\alpha$ -2,3 linked NeuNAc binds more favorably at this site than NeuNAc linked to Gal with an  $\alpha$ 2,6-linkage. This rationale would also explain the difference between binding scores of sites B1<sub>C2</sub> and C2<sub>B1</sub> in the WGA1/T5 complex (1,273 versus 1,346), where the  $\alpha$ 2,3-linked NeuNAc sits in site C2<sub>B1</sub> and the  $\alpha$ 2,6 linked NeuNAc in site B1<sub>C2</sub>. Although these discrepancies are within what we believe to be the uncertainty of HINT data, they agree with earlier binding data (Kronis & Carver, 1982), and with the NeuLac/WGA1 crystal structure revealing that stronger binding of  $\alpha$ 2,3-NeuLac is due to additional contacts involving Tyr 66 (Wright, 1990).

The HINT binding score for site A2 is roughly 50% lower compared with those for the high-affinity sites, consistent with the weak occupancy observed in the crystal structure. However, this binding mode of the tetrasaccharide in the crystal is further stabilized through a lattice contact with a symmetry mate of a twofold-related neighboring molecule. Although site D1<sub>A2</sub> (or D2<sub>A1</sub>) appears to be a functional site (Wright, 1984), lack of specificity for NeuNAc may be due to an unfavorable charge interaction, as suggested by model-building (Wright, 1992).

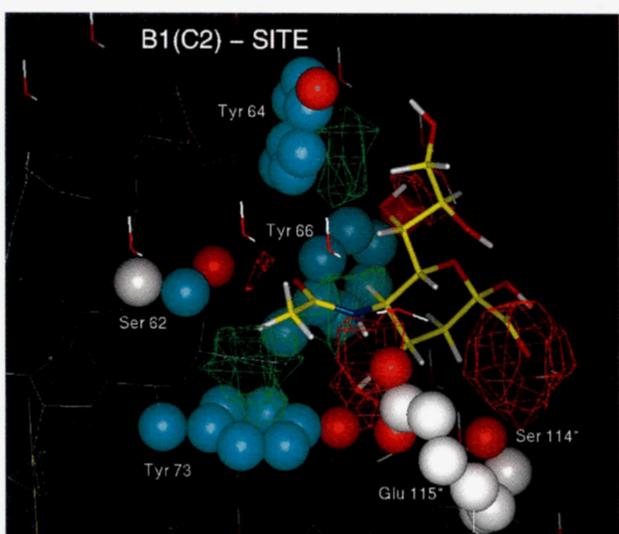
As illustrated in Figure 4, the two high-affinity sites (B1<sub>C2</sub> and C2<sub>B1</sub>) are quasi-twofold related. Their polar and nonpolar interaction contours generated for bound NeuNAc by HINT are

visualized in Figures 5A and B. The polar interactions, shown in red in Figures 5A and B, coincide primarily with regions of H-bonding, and are shaped similarly in these two sites. In both sites, NeuNAc is stabilized through four hydrogen bond contacts. This is consistent with the large HINT scores for H-bonding, which is somewhat higher for site C2<sub>B1</sub> (see Table 3) due to interactions with two carboxylate side chains. It is noteworthy that the beneficial binding contribution from Asp 86 had not been recognized earlier. In contrast, the lower polar interaction values obtained for binding at site A2 (see Fig. 5C; Table 3) are consistent with the presence of only two hydrogen bonds. As ex-

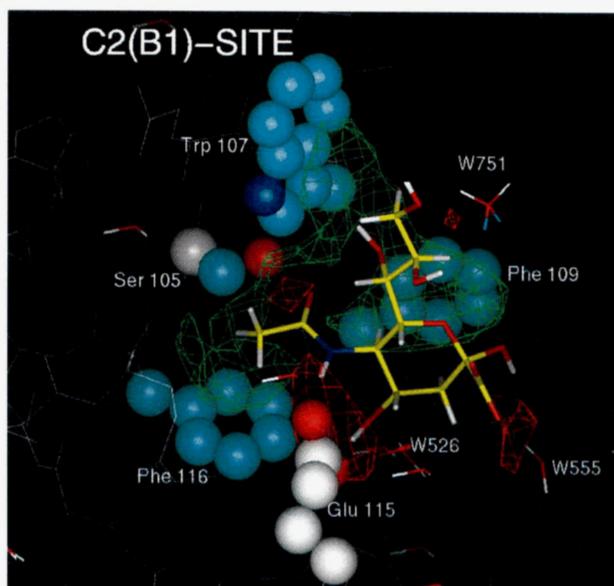


**Fig. 4.** Schematic diagram of the two pseudo-twofold related high-affinity binding sites (B1<sub>C2</sub> and C2<sub>B1</sub>) located between domains B1 and C2. Two orientations for oligosaccharide binding are illustrated. The sugar ring outlined by dashed lines represents the binding mode of sialyloligosaccharides.

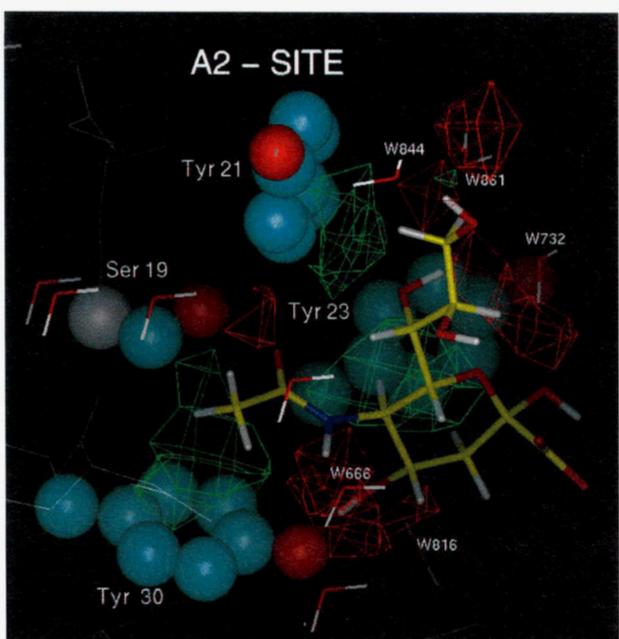
A



B



C



**Fig. 5.** Comparison of HINT interaction contours for bound NeuNAc at three WGA binding sites (figure prepared with Insight II). NeuNAc is shown in yellow. Side chains that interact with NeuNAc are modeled as 0.6 CPK spheres and color-coded according to their locations in the aromatic binding site (cyan) or at the polar region in the opposing domain (white). Hydrophobic and polar interactions were contoured on arbitrary scale and are shown in green and red, respectively. **A:** Site  $B1_{C2}$  (contour levels 20 and 800). **B:** Site  $C2_{B1}$  (contour levels 50 and 800). **C:** Site A2 (contour levels 20 and 700).

pected, hydrophobic interactions (green contours) are localized in areas of ligand contact with the aromatic side chains, for instance, the *N*-acetyl methyl group and carbon atoms of the pyranoside rings. These interactions are strongest in site  $C2_{B1}$ , where two Phe and one Trp are present. As expected, closely similar patterns are observed for the nonpolar contours (green) in sites A2 and  $B1_{C2}$ , where the aromatic residues are identical.

#### GlcNAc-2 binding

The HINT binding scores for the four binding environments with GlcNAc-2 bound, show a clear division into high- and low-affinity sites. The largest values are observed for sites  $B1_{C2}$  and  $C2_{B1}$  (1,136 and 1,261), in agreement with the above described

NeuNAc scores (see Table 3). However, difficulties were encountered in modeling GlcNAc-2 at site  $C2_{B1}$  due to steric crowding between the C4-OH and Ala 71 on domain B1. A good H-bond interaction between its 3-OH and the carboxyl group of Glu 72 cannot exist as observed in the NeuNAc complex (see Figs. 2B, 3B). NeuNAc lacks an OH on the corresponding carbon atom and thus can bind closer to this subsidiary polar region on domain B1 (Ala 71–Glu 72). Thus, the two sugars are not perfectly superimposable in their bound configurations at this site. Nevertheless, the net HINT scores for the modeled GlcNAc-2 exhibit strong binding, which is contributed partially by the proximity of Asp 86. This Asp is well-determined crystallographically, because it plays a critical role in intersubunit contact through an ion pair interaction with Arg 84.

The values obtained for sites A2 and D1<sub>A2</sub> are significantly lower, suggesting low-affinity binding. The A-domain site appears to be the weakest site due to absence of binding interactions from domain D of monomer 2 (see Table 1; Fig. 3C). The reduced binding score for site D1<sub>A2</sub> had been attributed to the absence of a third aromatic side chain at domain position 23 and to the presence of an Asp in place of the preferred Glu at domain position 29 of the opposing domain (A2). Because the crystallographic position of this disaccharide was not optimal (due to weak electron density), the importance of residue Ser 152, corresponding to Tyr 23, Tyr 66, and Phe 109 in sites A2, B1<sub>C2</sub>, and C2<sub>B1</sub>, respectively, was not realized earlier. Optimization of the positions of bound disaccharide and this serine hydroxyl group, and subsequent HINT analysis has now revealed several favorable interactions that contribute significantly to overall binding of GlcNAc-2 at this site. Ser 152 forms a relatively strong H-bond to the C3-OH of GlcNAc and an additional polar contact was discovered between the C4-OH and a second carboxylate group (Asp 129) (see Fig. 3D). This interaction is analogous to that described above in site C2<sub>B1</sub>. Moreover, HINT demonstrated that Asp 29 of subunit A2 provides a fairly strong H-bond contact to the NH-group of GlcNAc despite its longer distance (3.18 Å) (see Figs. 3D, 6D).

The HINT binding interactions for the four bound configurations are visualized graphically in Figure 6A, B, C, and D. As expected, fewer polar and nonpolar interaction contours are present in sites D1<sub>A2</sub> and A2 compared with those of sites B1<sub>C2</sub> and C2<sub>B1</sub> (contoured at the same level). The HINT binding results are also in good agreement with the nonpolar character of site C2<sub>B1</sub> (see Table 3), as visualized in Figure 6B. The hydrophobic contours are more extensive at this site compared with all other sites.

#### *The role of water*

It is well known that ordered water plays a significant role in stabilizing bound saccharide. Our results, however, demonstrate that reasonable and informative molecular models of the oligosaccharide/WGA binding sites can be obtained without explicitly modeling water. This is because the source of the HINT constants from the experimental measurement of the water/octanol partition coefficient for small molecules suggests that to some extent solvation information is inherently included in the HINT model (Wireko et al., 1991).

As evident from Table 3, inclusion of explicit waters in the HINT analysis has not been completely satisfactory, causing a reduction in the overall binding scores. This is due in large part to negative terms of the acid/acid, base/base, and hydrophobic/polar type, arising from poorly oriented hydrogens. HINT requires all polar hydrogens to be optimally oriented such that the favorable positive scores exceed the negative (nominally O-O) repulsion terms. Our careful modeling at each site has, however, only been modestly effective. Moreover, no improvement was achieved using conventional molecular mechanics structure optimization to locate suitable minima. The difficulties remaining in properly treating water are related to the fact that each explicitly modeled water adds three rotational degrees of freedom to the model even when its crystallographic position is known. Moreover, the current HINT parameter set, consisting of a hydrophobic atom constant and solvent-accessible surface area for each atom, may be too limited to deal with all the

subtleties of water constrained in a small three-atom molecule. Thus, based on our experience, we recommend that only those waters should be included in the HINT model which are well-ordered and structurally important to ligand stabilization by forming hydrogen bonds across otherwise unfavorable interactions. Waters that show negative net HINT scores are either poorly positioned or may not be real, if they have high crystallographic *B*-factors. Such waters would cause unrealistic or false overall HINT scores for the complex and should be removed from the model. In our experimental system, this situation was encountered at the A2 site (see Table 3), where inclusion of 10 waters (*B*-factors > 39.0 Å<sup>2</sup>) reduced the overall binding score to a negative value.

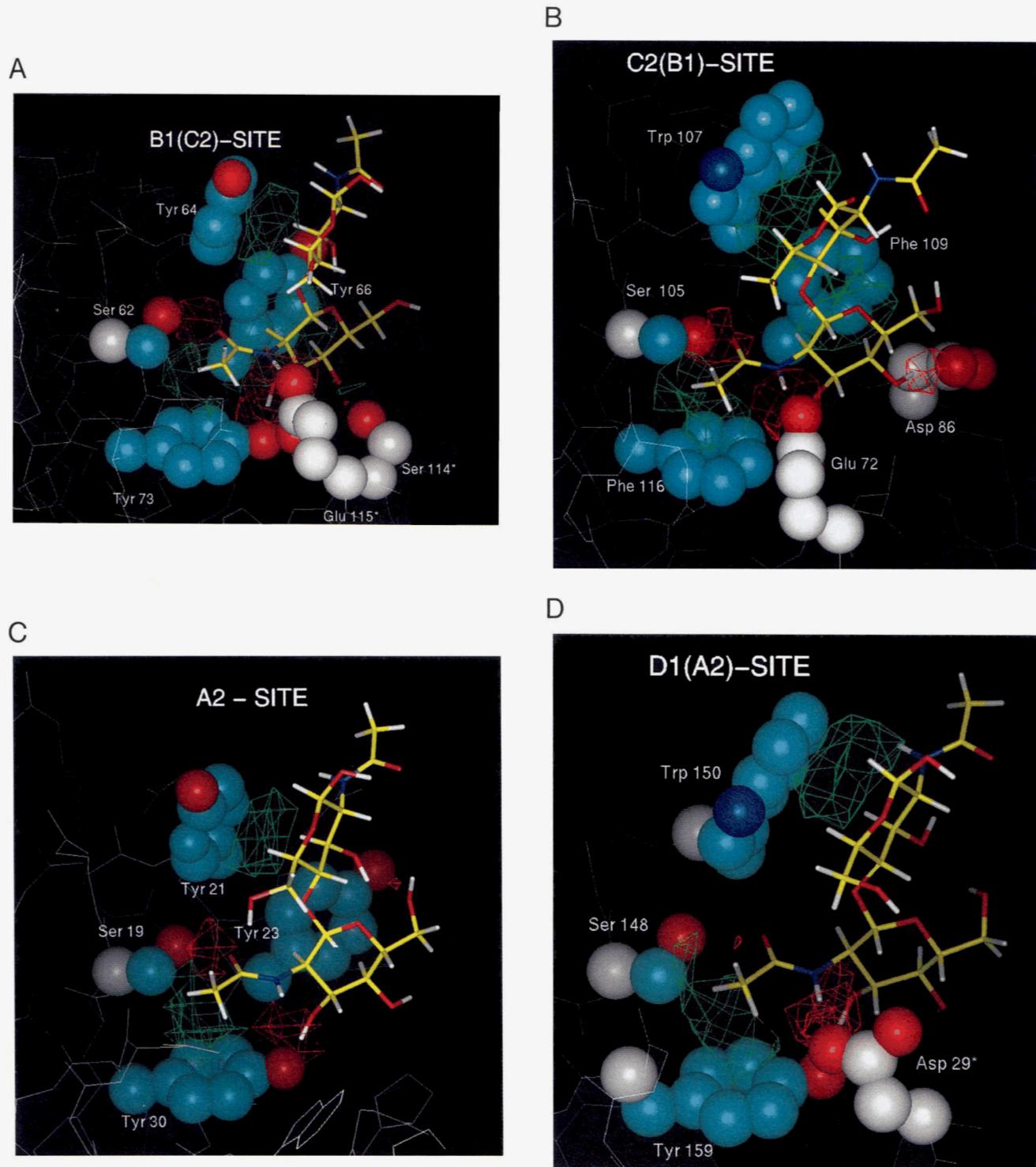
Several water molecules could be refined successfully in the two high resolution sialyloligosaccharide complexes (Wright, 1990; Wright & Jager, 1993). Although the overall patterns of H-bonded water molecules differ in the three sites (Fig. 2A,B,C), there is one well-ordered water molecule that solvates the C4-OH in all sites. This water molecule is well determined, with a low *B*-factor, and is tetrahedrally stabilized through other contacts with the protein. A number of other water molecules present in the two high-affinity sites (W873 and W555) (see Fig. 2A,B) also show strong HINT scores, suggesting that they may play important roles in complex stabilization.

#### *Conclusions*

HINT analysis of the four isolated saccharide-binding environments in WGA, using both crystallographically determined and modeled ligand positions, has provided reasonable answers to the questions set forth above. It firmly established the identity of the high-affinity sites and provided relative estimates for the low-affinity binding sites. In addition, it has allowed a deeper understanding of those binding interactions that are site-specific and specificity determining.

Although the net HINT binding scores are currently not directly translatable to solution binding constants ( $K_d$  or  $K_a$ ), they have yielded valuable comparative binding information not obtainable by any other experimental technique at this time. Considered within the specific framework of similar binding environments, the HINT data suggest that, indeed, two classes of binding sites exist in WGA. Sites B1<sub>C2</sub> and C2<sub>B1</sub> exhibit the highest and roughly equal HINT scores, which agrees well with solution binding data suggesting two equal affinity sites. Site B1<sub>C2</sub> was suspected to be one of the high-affinity sites based on high occupancy observed in all crystal complexes examined. HINT clearly identified site C2<sub>B1</sub> as the second high-affinity site for both ligands, despite the fact that GlcNAc-2 was modeled. Previous speculations had ruled functionality of this site out, in view of its markedly different aromatic character (Wright, 1984).

On the scale of the HINT scoring system, sites A1 (or A2) and D1<sub>A2</sub> (or D2<sub>A1</sub>) are weaker by at least a factor of two. However, the structural differences that exist between these sites, would suggest their nonequivalence in solution. The D1<sub>A2</sub> site, which was the only other site occupied in the WGA/GlcNAc-2 crystal complex, is missing one aromatic residue, and therefore has weaker hydrophobic contributions to overall binding. The HINT scores for the A-site suggest that it is the weakest of the four WGA sites, consistent with the fact that it is missing a subsidiary binding region. It thus acts as a monomeric domain



**Fig. 6.** Comparison of the HINT binding interactions of bound GlcNAc-2 at four WGA sites (figure prepared with Insight II). Atomic models are color-coded as described in Figure 5. HINT interaction contours at sites (A)  $B1_{C2}$ , (B)  $C2_{B1}$ , (C) A2, and (D)  $D1_{A2}$ , respectively. Contour levels are 100 (hydrophobic, green) and 500 (polar, red).

shown to have reduced  $K_a$ 's (Rice, 1994). Its net score can be considered to represent that portion of overall binding contributed by the aromatic pocket alone. Based on this estimate for minimum binding, it becomes possible to evaluate and quanti-

tate the effect on binding strength from the additional polar contacts coming from monomer 2 in all other sites.

HINT analysis of two carefully modeled complexes has also clarified the role of several polar residues that provide new fa-

vorous interactions not realized from knowledge of the crystal structure alone. These involve hydrogen bond contacts between (1) a conserved aspartic acid side chain with the C4-OH of GlcNAc-2 in sites D1<sub>A2</sub> (Asp 129) and C2<sub>B1</sub> (Asp 86), and (2) the -OH- of Ser 152 and the C3-OH in site D1<sub>A2</sub>.

In terms of reliability of the method, our studies have demonstrated that HINT consistently recognizes hydrogen bonds, ionic interactions, and interactions involving hydrophobic groups. In all four binding configurations, hydrogen bonding and polar van der Waals interactions constitute the largest contribution to overall binding, consistent with thermodynamic data. Subtle differences in their distribution may be considered as site-specific determinants of fine-specificity.

## Materials and methods

### Modeling rationale

The three crystal complexes examined with HINT are (1) the monoclinic 3.0-Å structures of WGA/GlcNAc-β1,4-GlcNAc (Wright, 1984) with two unique sites occupied per subunit (B1<sub>C2</sub> and D1<sub>A2</sub>) termed "primary" and "secondary" binding sites; (2) the monoclinic 2.2-Å complex WGA1/NeuLac (Wright, 1990) with one site occupied per subunit (B1<sub>C2</sub> or B2<sub>C1</sub>); and (3) the crosslinked orthorhombic 2-Å structure of WGA1 complexed with the tryptic octa-sialoglycopeptide T5 of glycophorin A with three of the eight domain sites in the dimer occupied (dimer in asymmetric unit) (Wright, 1992; Wright & Jäger, 1993). The bivalent tetrasaccharide crosslinks neighboring molecules by binding with its terminal α2,6-linked NeuNAc to site B1<sub>C2</sub> of one molecule and with its α2,3-linked NeuNAc to site C2<sub>B1</sub> of a crystallographically related molecule. Site A2 is occupied by the α2,6-linked NeuNAc of a second T5 molecule. Superposition experiments indicate that the binding mode of NeuNAc is similar at all three sites (RMSA = 0.5–0.75 Å). Four of the remaining sites (B2<sub>C1</sub>, C1<sub>B2</sub>, A1, and D2<sub>A1</sub>) are inaccessible in this crystal form, whereas site D1<sub>A2</sub> is vacant and does not bind NeuNAc, as demonstrated by soaking experiments (Wright & Jäger, 1993).

Modeling was conducted with the programs FRODO (Jones, 1978) and Insight II (Biosym Technologies, San Diego, California). In addition, inspection of the tabulated HINT values of each interatomic contact served as an important guide in making fine conformational adjustments in each round of modeling with the goal to optimize the balance between favorable (positive values) and unfavorable (negative values) interactions. It was particularly important to orient hydrogen atoms of methyl and hydroxyl groups and water molecules properly. In cases where refined crystallographic coordinates were available (Bernstein et al., 1977), only slight adjustments were necessary. Because the GlcNAc-2 positions were deduced from unrefined difference Fourier maps, more extensive modeling was required for this disaccharide, especially for subsite 2. In modeling the binding of GlcNAc-2 at sites A1 (or A2) and C1<sub>B2</sub> (or C2<sub>B1</sub>), which are inaccessible in the monoclinic WGA crystals, the coordinates of NeuNAc were used as guide coordinates. The goal was to position the disaccharide with its nonreducing terminal residue roughly superimposable with NeuNAc as observed bound in the WGA1/T5 complex. Three crucial contact distances were preserved (see Table 2): (1) van der Waals' distance (3.8–4.5 Å) between the N-acetyl methyl group and the phenyl ring at domain

position 30 (Tyr 30 or Phe 116); (2) H-bonding distance (2.3–2.8 Å) between the acetamido carbonyl oxygen and the Ser-OH at domain position 19 (Ser 19 or Ser 105); and (3) van der Waals' stacking contact (3.8–4.5 Å) of the pyranose ring atoms of GlcNAc with the aromatic side chain in subsite 1 at domain position 23 (Tyr 23 or Phe 109), and in subsite 2 at domain position 21 (Tyr 21 or Trp 107). The glycosidic bond angles ( $\phi$  and  $\psi$ ) were adjusted to maximize van der Waals' contact with the aromatic side chains and to allow a good intermolecular H-bond contact between the 3-OH of the reducing GlcNAc residue and the ring oxygen of the nonreducing residue (2.45–2.57 Å).

### HINT calculations

HINT calculations were performed using a version of the program integrated into the Insight II modeling environment (HINT 2.0 I) (EduSoft, LC, Ashland, Virginia). An interaction score ( $b_{ij}$ ) is computed for each interacting atom pair ( $i,j$ ), where  $i$  is an atom on the ligand and  $j$  is an atom on the protein. It is defined as the product of the atom's solvent-accessible surface area and hydrophobic atom constants, and a parameter that is a function of distance between the interacting atoms (Kellogg et al., 1992).

$$b_{ij} = \Sigma_i \Sigma_j [s_i a_i s_j a_j R_{ij} + r_{ij}],$$

where  $s$  and  $a$  are the solvent-accessible surface area and the hydrophobic atom constants, respectively, and  $R_{ij}$  and  $r_{ij}$  are functions of distance between interacting atoms. The general protocol used for each binding study was as follows. (1) The protein molecule, bound water, and ligand were displayed in Insight II from separate PDB coordinate files (Bernstein et al., 1977). (2) Hydrogen atoms were added to all molecules and optimized by model-building. (3) Subsequently the water molecules were merged with the protein data file. (4) Forcefield potential types were assigned to all ligand and protein atoms. (5) The protein (including water) was "partitioned" by HINT to assign the hydrophobic atom constants and solvent-accessible surface area constants to each atom. Partitioning was accomplished using the "dictionary" option, which uses a look-up table of previously calculated amino acid residue values. All ionizable side chains and capping groups were treated as if at neutral pH, thus assuming that acids are ionized and bases protonated. Parameters for polar hydrogen atoms were explicitly included, whereas non-polar hydrogens were implicitly included in united atoms. (6) The ligand was partitioned by HINT using the "calculate" option, which uses the method of Hansch and Leo (1979; Abraham & Leo, 1987) and the chemical structure/connectivity of the ligand to estimate its log  $P$ . Only polar hydrogens were calculated explicitly. (7) The distance behavior between interacting atoms was modeled with the default HINT function (Kellogg et al., 1992) ( $R_{ij} = e^{-r}$ ;  $r_{ij} = 50.0 \epsilon_{ij} [(v r_{vdw}/r)^{12} - 2(v r_{vdw}/r)^6]$ ), where  $r$  is the interatomic distance,  $r_{vdw}$  is the sum of the van der Waals' radii for the interacting atoms, and  $\epsilon_{ij}$  is the Lennard-Jones interaction parameter (Levitt, 1983; Levitt & Perutz, 1988). A fractional value,  $v$ , of 0.9 modified the summed van der Waals' radii to reduce the influence of steric repulsion on the HINT score. (8) A table listing binding scores for each interacting atom pair was generated and inspected after each round of modeling. Suspiciously large negative values arising from overly close contacts of the acid/acid, base/base, and hy-

drophobic/polar type, were investigated further. These could, in most cases, be eliminated through model building, raising the value of the overall binding score. However, in cases of poorly oriented water molecules, several rounds of modeling were necessary. Orienting these waters through molecular mechanics energy minimization was found not to be as effective. (9) To locate and visualize graphically the different types of interactions (favorable or unfavorable) in the atomic model, HINT InterMolecular interaction maps were calculated between the protein and ligand atoms, using the same parameters and functions as were used to generate the table. In this type of map calculation, a fine grid (1.0 Å) of "observer atoms" are superimposed over the interaction region. These observer atoms record the types and intensities of intermolecular interactions occurring at their grid coordinates in order to produce a contourable interaction map.

### Acknowledgments

This research was supported by the National Institutes of Health, grant AI-17992.

### References

- Abraham DJ, Kellogg GE. 1994. The effect of physical organic properties on hydrophobic fields. *J Comput-Aided Mol Design* 8:41-49.
- Abraham DJ, Leo AJ. 1987. Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins Struct Funct Genet* 2:130-152.
- Aub JC, Tieslau C, Lankester A. 1963. Reactions of normal and tumor cell surfaces to enzymes. I. Wheat-germ lipase and associated mucopolysaccharides. *Proc Natl Acad Sci USA* 50:613-619.
- Baines G, Lee RT, Lee YC, Freire E. 1992. Microcalorimetric study of wheat germ agglutinin binding to N-acetylglucosamine and its oligomers. *Biochemistry* 31:12624-12628.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rogers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535-542.
- Bourne Y, Bolgiano B, Liao D, Strecker G, Cantau P, Herzberg O, Feizi T, Cambillau C. 1994. Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides. *Nature Struct Biol* 1:863-869.
- Burger M. 1969. A difference in the architecture of the surface membrane of normal and virally transformed cells. *Biochemistry* 62:994-1001.
- Dessen A, Gupta D, Subramaniam S, Brewer CF, Sacchettini JC. 1995. X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. *Biochemistry* 34:4933-4942.
- Goldstein IJ, Poretz RD. 1986. Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. In: Liener IE, Sharon N, Goldstein IJ, eds. *The lectins: Properties, function and application in biology and medicine*. Orlando, Florida: Academic Press. pp 43-247.
- Gussio R, Pattabiraman N, Zaharevitz DW, Kellogg GE, Topol I, Rice WG, Schaeffer CA, Burt SK, Erickson JW. 1996. All atom models for the non-nucleoside binding site of HIV-1 reverse transcriptase. A molecular model complexed with inhibitors: A 3D QSAR approach. *J Med Chem* 39:1645-1650.
- Hansch C, Leo AJ. 1979. *Substituent constants for correlation analysis in chemistry and biology*. New York: J. Wiley and Sons. pp 1-339.
- Hester G, Kaku H, Goldstein IJ, Wright CS. 1995. Structure of mannose-specific snowdrop (*Galanathus nivalis*) lectin is representative of a new plant family. *Nature Struct Biol* 2:472-479.
- Imbert A, Bourne Y, Cambillau C, Rougé P, Pérez S. 1993. Oligosaccharide conformation in protein/carbohydrate complexes. *Adv Biophysical Chem* 3:71-117.
- Jones TA. 1978. A graphics model building and refinement system for macromolecules. *J Appl Crystallogr* 11:268-272.
- Kellogg GE, Joshi GS, Abraham DE. 1992. New tools for modeling and understanding hydrophobicity and hydrophobic interactions. *Med Chem Res* 1:444-453.
- Kronis KA, Carver JP. 1982. Specificity of isolectins of wheat germ agglutinin for sialyloligosaccharides: A 360-MHz proton nuclear magnetic resonance binding study. *Biochemistry* 21:30050-3057.
- Kronis KA, Carver JP. 1984. Thermodynamics of wheat germ agglutinin-sialyloligosaccharide interactions by protein nuclear magnetic resonance. *Biochemistry* 24:834-840.
- Kronis KA, Carver JP. 1985. Wheat germ agglutinin dimers bind sialyloligosaccharides at four sites in solution: Proton nuclear magnetic resonance temperature studies at 360 mhz. *Biochemistry* 24:826-833.
- Levitt M. 1983. Molecular dynamics of native protein. I. Computer simulation of trajectories. *J Mol Biol* 168:595-620.
- Levitt M, Perutz MF. 1988. Aromatic rings act as hydrogen bond acceptors. *J Mol Biol* 210:751-754.
- Lis H, Sharon N. 1981. Lectins in higher plants. In: Marcus A, ed. *The biochemistry of plants: A comprehensive treatise, proteins and nucleic acids*. Vol 6. New York: Academic Press. pp 371-447.
- Lotan R, Sharon N. 1973. The fluorescence of wheat germ agglutinin and its complexes with saccharide. *Biochem Biophys Res Commun* 55:1340-1350.
- Nagata Y, Burger MM. 1974. Wheat germ agglutinin: Molecular characteristics and specificity for sugar binding. *J Biol Chem* 249:3116-3122.
- Pérez S, Imbert A, Carver JP. 1994. Molecular modeling: An essential component in the structure determination of oligosaccharides and polysaccharides. *Adv Computational Biol* 1:147-202.
- Privat JP, Delmotte F, Miallonier G, Bouchard P, Monsigny M. 1974a. Fluorescence studies of saccharide binding to wheat germ agglutinin (lectin). *Eur J Biochem* 47:5-14.
- Privat JP, Delmotte F, Monsigny M. 1974b. Protein-sugar interactions. Association of  $\beta$  (1-4) linked N-acetyl-D-glucosamine oligomer derivatives with wheat germ agglutinin (lectin). *FEBS Lett* 46:224-227.
- Raihel NV, Wilkins TA. 1984. Isolation and characterization of a cDNA clone encoding wheat germ agglutinin. *Proc Natl Acad Sci USA* 81:6745-6749.
- Rice A. 1994. Cloning, expression, purification and characterization of domain B of wheat germ agglutinin [thesis]. Richmond, Virginia: Virginia Commonwealth University.
- Sharon N. 1993. Lectin-carbohydrate complexes of plants and animals: An atomic view. *Trends Biochem Sci* 18:221-226.
- Sharon N. 1994. When lectin meets oligosaccharide. *Nature Struct Biol* 1:843-845.
- Sharon N, Lis H. 1989. Lectins as cell recognition molecules. *Science* 246:862-866.
- Weis WI, Drickamer K, Hendrickson WA. 1992. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* 360:127-134.
- Wireko FC, Kellogg GE, Abraham DJ. 1991. Allosteric modifiers of hemoglobin. 2. Crystallographically determined binding sites and hydrophobic binding/interaction analysis of novel hemoglobin oxygen effectors. *J Med Chem* 34:758-767.
- Wright CS. 1980. Crystallographic elucidation of the saccharide binding mode in wheat germ agglutinin and its biological significance. *J Mol Biol* 141:267-291.
- Wright CS. 1984. Structural comparison of two distinct sugar binding sites in wheat germ agglutinin isolectin 2. *J Mol Biol* 178:91-104.
- Wright CS. 1987. Refinement of the crystal structure of wheat germ agglutinin isolectin 2 at 1.8 Å resolution. *J Mol Biol* 194:501-529.
- Wright CS. 1990. 2.2 Å resolution structure analysis of two refined N-acetylneuraminyllactose wheat germ agglutinin isolectin complexes. *J Mol Biol* 215:635-651.
- Wright CS. 1992. Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. Structural basis for cooperative lectin-cell binding. *J Biol Chem* 267:14345-14352.
- Wright CS, Jäger J. 1993. Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J Mol Biol* 232:620-638.