

Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*

Hans-Georg Beisel^{1,2},
Shun-ichiro Kawabata^{3,4}, Sadaaki Iwanaga³,
Robert Huber¹ and Wolfram Bode¹

¹Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, 82152 Martinsried, Germany, ³Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-82 and ⁴CREST, Japan Science and Technology Corporation, Japan

²Corresponding author
e-mail: beisel@biochem.mpg.de

Tachylectin-2, isolated from large granules of the hemocytes of the Japanese horseshoe crab (*Tachypleus tridentatus*), is a 236 amino acid protein belonging to the lectins. It binds specifically to *N*-acetylglucosamine and *N*-acetylgalactosamine and is a part of the innate immunity host defense system of the horseshoe crab. The X-ray structure of tachylectin-2 was solved at 2.0 Å resolution by the multiple isomorphous replacement method and this molecular model was employed to solve the X-ray structure of the complex with *N*-acetylglucosamine. Tachylectin-2 is the first protein displaying a five-bladed β -propeller structure. Five four-stranded antiparallel β -sheets of W-like topology are arranged around a central water-filled tunnel, with the water molecules arranged as a pentagonal dodecahedron. Tachylectin-2 exhibits five virtually identical binding sites, one in each β -sheet. The binding sites are located between adjacent β -sheets and are made by a large loop between the outermost strands of the β -sheets and the connecting segment from the previous β -sheet. The high number of five binding sites within the single polypeptide chain strongly suggests the recognition of carbohydrate surface structures of pathogens with a fairly high ligand density. Thus, tachylectin-2 employs strict specificity for certain *N*-acetyl sugars as well as the surface ligand density for self/non-self recognition.

Keywords: animal lectin/crystal structure/horseshoe crab/*N*-acetylglucosamine/ β -propeller

Introduction

Phylogenetically ancient host defense systems are based on innate immunity, with germline-encoded defense molecules for the recognition of microbial pathogens. This kind of immunity is found in all multicellular organisms. Although vertebrate organisms additionally employ adaptive immunity with random specificity, this adaptive part of the host defense requires sufficient information about the pathogen. This information is provided by the ability of the innate immunity host defense system to detect specific carbohydrate patterns that are unique to and also essential for infectious microorganisms (Fearon

and Locksley, 1996; Medzhitov and Janeway, 1997). Lectin-like effector proteins are essential parts of both invertebrate and vertebrate innate immunity and, therefore, innate immunity is not an ancient relic, but an important first line defense mechanism providing a helping hand to acquired immunity. Studies on innate immunity defense molecules found in invertebrate as well as in vertebrate organisms will elucidate further the basic mechanisms used by animals to distinguish self and non-self materials.

The invertebrate *Tachypleus tridentatus* (Japanese horseshoe crab), an arthropod, relies completely on innate immunity, employing a unique and very efficient host defense system (reviewed in Muta and Iwanaga, 1996; Iwanaga *et al.*, 1998). The hemolymph of horseshoe crabs contains soluble defense proteins and one type of cells, the hemocytes, which also contain defense proteins including a hemolymph clotting cascade. The plasma exhibits hemagglutination and cytolytic activities. Some 99% of the horseshoe crab hemocytes contain large (L) and small (S) granules (Toh *et al.*, 1991). The hemocytes undergo degranulation upon contact with pathogens, releasing various host defense molecules into the hemolymph. The S-granules contain antimicrobial peptides and proteins. The L-granules also contain antimicrobial peptides and, additionally, several different lectins, serpin-like proteinase inhibitors, α_2 -macroglobulin and a hemolymph clotting cascade. This clotting cascade converts the soluble L-granule protein coagulogen (Nakamura *et al.*, 1976; Bergner *et al.*, 1996) into a coagulin gel via activation of the protease zymogen proclotting enzyme. The gel generated immobilizes invading microorganisms and prevents local hemolymph leakage (Nakamura *et al.*, 1988; Muta *et al.*, 1995). The proclotting enzyme is activated via two possible pathways, triggered by bacterial lipopolysaccharide (LPS) and fungal (1,3)- β -D-glucan, respectively. This branched clotting cascade is extremely sensitive and, therefore, has been employed as the 'limulus test' for detection of endotoxin in clinical chemistry (Tanaka and Iwanaga, 1993).

The tachylectins 1–4 stored in horseshoe crab *T. tridentatus* L-granules of hemocytes are lectins comprising various specificities. Tachylectin-1 binds to the core region of LPS and to lipoteichoic acid (LTA), which are specific for Gram-negative and Gram-positive bacteria, respectively. It agglutinates Gram-negative and Gram-positive bacteria and exhibits antibacterial activity towards Gram-negative bacteria (Saito *et al.*, 1995). Tachylectin-3 binds to LPS O-antigen and to LTA (Inamori *et al.*, 1999). Tachylectin-4 also recognizes S-type LPS O-antigen and displays a more potent hemagglutination activity against human A-type erythrocytes than tachylectin-2 (Saito *et al.*, 1997). There is no sequence similarity among the known tachylectins and, additionally, the sequence of tachylectin-2 shows no significant similarity to any other known protein.

Tachylectin-2 shows specific activity for *N*-acetyl-D-glucosamine (GlcNAc, association constant $K_a = 1.95 \times 10^4/\text{M}$) and *N*-acetyl-D-galactosamine (GalNAc, $K_a = 1.11 \times 10^3/\text{M}$), but no affinity toward glucose, glucosamine, galactose or galactosamine. It binds to *N*-acetylallo-lactosamine (β -D-Gal-[1,6]-D-GlcNAc, $K_a = 1.55 \times 10^4/\text{M}$), but not to *N*-acetylactosamine (β -D-Gal-[1,4]-D-GlcNAc). It also binds to *N,N'*-diacetylchitobiose with a minimum concentration of 3.13 mM for inhibition of hemagglutinating activity, compared with 0.097 mM for GlcNAc, and binds *N,N',N''*-triacetylchitotriose with a minimum concentration of 6.25 mM for inhibition (Okino *et al.*, 1995). These data suggest the requirement for an acetamido group and a free 4-OH group for a carbohydrate epitope to be recognized by tachylectin-2. The configuration of C4 with a 4-OH group *trans* to CH₂OH (4-OH equatorial in the bioactive conformation) of GlcNAc leads to an association constant 17-fold higher than that for GalNAc, which has a *cis* 4-OH group (axial 4-OH). On the cellular level, tachylectin-2 is known for the agglutination of *Staphylococcus saprophyticus* KD and human A-type erythrocytes. *Staphylococcus saprophyticus* is reported to carry α (1-2)-linked GlcNAc as an additional substituent of its LTA poly(glycerophosphate) backbone (Ruhland and Fiedler, 1990). Tachylectin-2 does not bind to other bacteria such as *Staphylococcus aureus* 209 P, *Staphylococcus epidermidis* K3, *Micrococcus luteus*, *Enterococcus hirae* and *Escherichia coli* strain B (Okino *et al.*, 1995).

The X-ray structure of tachylectin-2 in complex with GlcNAc gives a structure-based explanation for the binding specificity and the mechanism employed for self/non-self distinction.

Results and discussion

Overall tachylectin-2 structure and topology

Tachylectin-2 is the first example of a 5-fold β -propeller structure. Tachylectin-2 exhibits the shape of a rather flat pentagonal torus, with an approximate height of 25 Å and a length of an edge of the molecular pentagon of 28 Å (corresponding to a 'diameter' of 45–48 Å). The single polypeptide chain of tachylectin-2 is organized in five β -sheets, which are arranged in consecutive order and with 5-fold pseudosymmetry around a central tunnel. Each twisted β -sheet is built up by four antiparallel β -strands with a W-like topology. The first strand of each β -sheet is always the innermost strand, bounding the central tunnel of the molecule. From there on, the strands are counted as 2 and 3 (middle strands) and finally strand 4, which is the outermost strand with the largest distance from the central tunnel. All five innermost strands enter the propeller at a common side (its 'entrance' side) and run almost parallel to one another along the 5-fold propeller axis towards the 'exit' side, forming the central tunnel. One of the β -sheets (called propeller blade I) is formed by the C-terminal segment representing the innermost strand and the N-terminal polypeptide chain forming strands 2, 3 and 4 (Figure 1). Strands 1 and 2 (except in the β -sheet with the chain termini) are connected by short loops of five residues, while strands 2 and 3 are connected by β -turns of two residues. Strands 3 and 4 are connected by the large 3→4 loop, starting with nine residues followed by

a four-residue helical segment and connected to strand 4 by a further residue (Figures 1 and 2). After strand 4, the chain turns through the 'connecting segment' towards the central tunnel to begin the innermost strand of the next β -sheet. Optimal superposition of the five β -sheets based on C $_{\alpha}$ coordinates (of all residues excluding the connecting segments between the β -sheets) shows their virtual equivalence, with a mean r.m.s. distance of 0.46 Å.

The most striking feature of the tachylectin-2 amino acid sequence are the five 47 residue tandem repeats with a high internal sequence identity of 49–68% (Figure 2), reflected by the topological symmetry of the structural model. These conserved residues are located at equivalent positions in all five β -sheets, including the 3→4 loop in each β -sheet. A remarkable feature of the tachylectin-2 X-ray structure is a ring of conserved phenylalanine residues anchored in the innermost strand of each β -sheet (Figure 3). In other β -propellers, in contrast, small side chains normally are found in the first strand close to the central tunnel, while larger side chains are employed in the outer strands to fill the wider inter-blade gaps.

The cyclic arrangement of β -sheets found in β -propeller structures gives rise to a central tunnel running through the toroid-shaped molecule. The diameter of the central tunnel in tachylectin-2 is 7–8 Å, neglecting the water molecules. Most of the known β -propeller structures use their central tunnel or its entrance for coordinating a ligand or binding a substrate to carry out an enzymatic reaction (Baker *et al.*, 1997). In tachylectin-2, the central tunnel is not a ligand-binding site but is filled with water molecules arranged in a closed cage of pentagonal dodecahedral symmetry (Figure 3). This structural element is also found in clathrate hydrates (Mak and Zhou, 1997). The water molecules of the pentagonal dodecahedron in tachylectin-2 exploit their full hydrogen-bonding capacity. The water cage, located near the entrance side of the central tunnel, reduces the tunnel diameter to ~1.8 Å. Therefore, the water cage in tachylectin-2 is empty, as no guest species fits in the tunnel. The water dodecahedron in tachylectin-2 reflects the 5-fold symmetry of the β -propeller and is framed by the polar main chain groups of the five innermost strands and by the benzyl side chains of the conserved phenylalanine residues mentioned above.

The innermost strand of blade I comprises the sequence 231-FRFLFF-236. All the phenylalanines are conserved within the five β -sheets, except Phe236 which is Leu48 in blade II; all these residues are located at equivalent positions in each β -sheet. Phe231 starts the innermost strand and points away from the water cage into the gap space between adjacent β -sheets. Phe233 stands edge-on to the water cage, while Phe235 is touching it with its face. Phe236 points away from the water cage and packs its side chain between two innermost strands. As this water cage is not affected by the crystal packing of tachylectin-2, it is expected to exist also in solution, and may contribute to the stabilization of the protein fold.

Comparison with other β -propeller structures

Protein structures with a 4- to 8-fold β -propeller geometry have been found, and only those with 5-fold symmetry have been lacking so far. Tachylectin-2 is closing this gap for the first time, displaying 5-fold symmetry. As a consequence of the low *n*-fold symmetry, the angle by



Fig. 1. Stereo ribbon plot (top) and C_{α} representation (bottom) of the 5-fold tachylectin-2 β -propeller showing bound GlcNAc as a ball-and-stick model. The color coding is as in Figure 2. The β -sheet shown in red closes the circle, with the innermost β -strand being the C-terminus and the second β -strand beginning near the N-terminus. In the C_{α} representation, the connecting segments are drawn in green, and the white balls indicate the position of Trp134 and Trp169 and equivalents as shown in Figure 6. The plot was made with MOLSCRIPT (Kraulis, 1991) and rendered with POVray (1997).

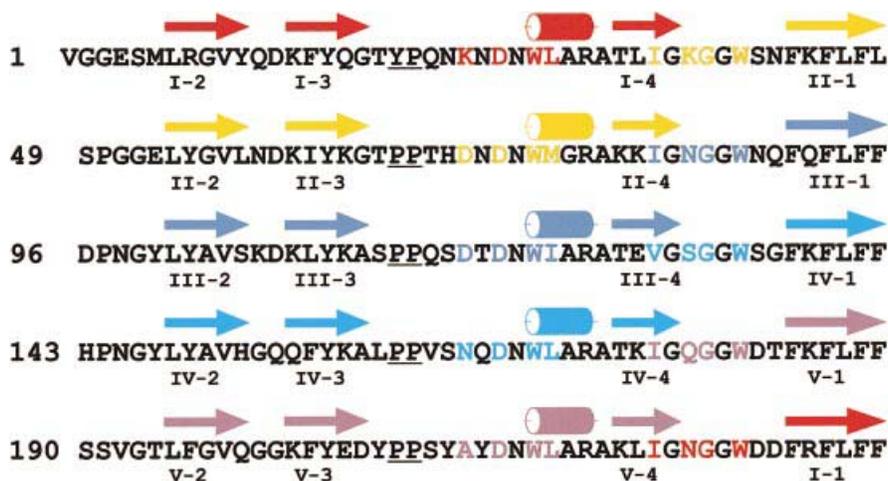


Fig. 2. Amino acid sequence of tachylectin-2 shown as five aligned tandem repeats (Okino *et al.*, 1995). The colors of secondary structure elements correspond to the ribbon plot shown in Figure 1. All residues belonging to one binding site are shown in the same color. The sequence XXXXDNL is part of the 3→4 loop of each β -sheet, starting with the underlined residues, whereas JGXGGW belongs to the connecting segment between two adjacent β -sheets. The β -sheets are named I–V, and individual β -strands in each β -sheet 1–4.

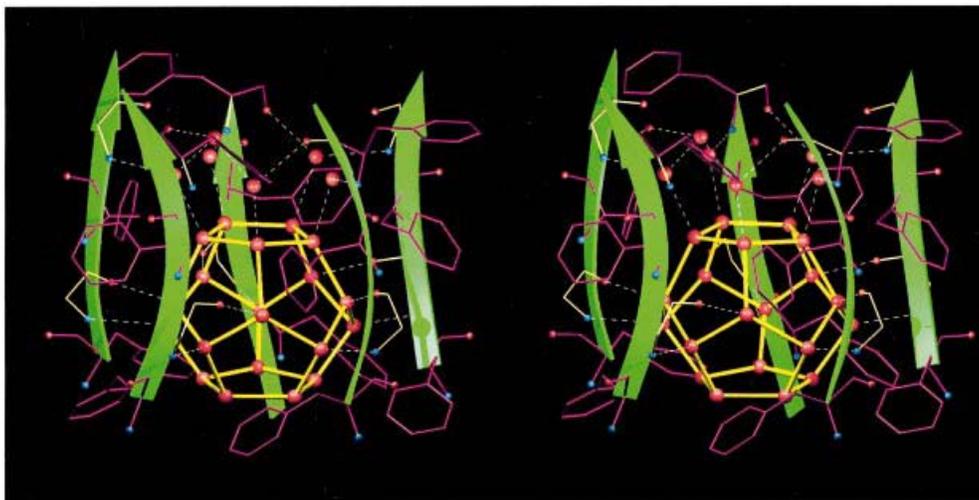


Fig. 3. Stereo plot showing the innermost strand of each β -sheet with the conserved phenylalanine residues (see Figure 2) and the pentagonal dodecahedral water structure filling the central tunnel of the tachylectin-2 β -propeller. The plot was made with MOLSCRIPT (Kraulis, 1991) and rendered with POVray (1997).

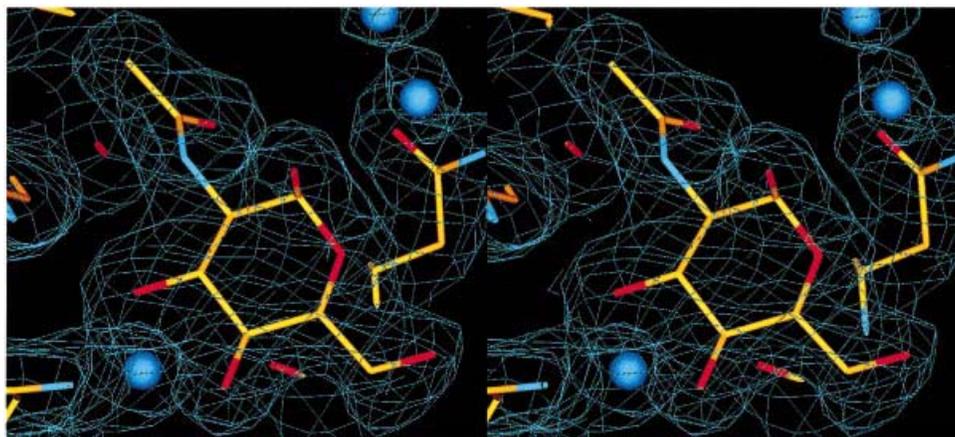


Fig. 4. Final 2.0 Å resolution $2mF_o - DF_c$ electron density map calculated with REFMAC (Murshudov *et al.*, 1997). The density shown corresponds to the carbohydrate-binding site illustrated in Figure 6; the map is normalized and contoured at 1.0 σ . The non-polar face of the sugar molecule packs against a conserved asparagine (Asn27 and equivalents). The figure was prepared with BOBSCRIPT (Kraulis, 1991).

which the β -sheets are rotated around the central n -fold axis (ideally parallel to the inner β -strands; Murzin, 1992) to generate the cyclic arrangement of the protein fold is 72° for $n = 5$ and even 90° with $n = 4$. Thus, β -propeller structures with 4-fold symmetry [hemopexin (Faber *et al.*, 1995) and hemopexin-like domains (Li *et al.*, 1995)] and 5-fold symmetry (such as tachylectin-2) necessarily have a large spatial gap between adjacent β -sheets. In tachylectin-2, this gap is filled, in particular, by the 3 \rightarrow 4 loop of each β -sheet and the connecting segment between adjacent β -sheets, as well as by large side chains in the outer strands. In β -propellers of higher symmetry, such as galactose oxidase ($n = 7$; Ito *et al.*, 1991) or methanol dehydrogenase ($n = 8$; Xia *et al.*, 1992), there is considerably less of a gap, which can be filled effectively by side chains from adjacent β -sheets.

Tachylectin-2 in complex with N-acetylglucosamine

The X-ray structure of tachylectin-2 in complex with GlcNAc reveals five quasi-equivalent binding sites, with virtually identical occupancy and geometry in the crystal. Figure 4 displays the ligand bound at blade IV, super-

imposed with the electron density map. The binding sites are all located between the connecting segment of adjacent β -sheets and the 3 \rightarrow 4 loop of the following β -sheet, perfectly reflecting the 5-fold symmetry of tachylectin-2 (Figure 1). Tachylectin-2 shows virtually no change in main or side chain conformation upon binding of GlcNAc. The r.m.s. deviation between the apo- and the liganded structure is 0.18 Å (calculated from all C_α positions).

Since all five carbohydrate-binding sites are equivalent, only the one illustrated in Figure 6 (blade IV) will be described in detail. Additional structural data on all binding sites is given in Table I.

Each of the five binding sites comprises two parts. The first part is built up by the 3 \rightarrow 4 loop of each β -sheet, while the second part is made by the connecting segment from the previous β -sheet (Figure 2). All 3 \rightarrow 4 loops are located on the 'exit' side of the tachylectin-2 propeller, while all β -sheet-connecting segments are located on the 'entrance' side. The binding pocket (Figures 5 and 6) is bounded by Phe137 and Trp169 as the back. Phe137 is located at the beginning of the innermost strand and it is strictly conserved, as mentioned above. The bottom of the pocket is made by the main chain atoms of Asp167–

Table I. Binding sites of GlcNAc (NAG)

Ligand atom	NAG 1	NAG 2	NAG 3	NAG 4	NAG 5
O7	M76 (N, 2.8) W75 (N, 3.3)	L29 (N, 3.2) W28 (N, 3.3)	L170 (N, 2.9) W169 (N, 3.1)	I123 (N, 2.9) W122 (N, 3.2)	L217 (N, 3.0) W216 (N, 3.2)
C8	W40 (CE2, 3.6)	W228 (CE2, 3.5)	W134 (CE2, 3.6)	W87 (CE2, 3.6)	W181 (CE2, 3.6)
N2	G38 (O, 3.1)	G226 (O, 3.0)	G132 (O, 3.0)	G85 (O, 2.8)	G179 (O, 2.9)
O3	W40 (NE1, 3.0) (K37) H ₂ O (2.8)	W228 (NE1, 2.8) N225 (N, 3.0) H ₂ O (2.8)	W134 (NE1, 2.8) S131 (N, 3.1) H ₂ O (2.7)	W87 (NE1, 2.8) N84 (N, 3.2) H ₂ O (2.6)	W181 (NE1, 2.9) Q178 (N, 3.1) H ₂ O (2.7)
O4	D71 (O, 2.5) K37 (N, 3.0)	K24 (O, 2.6) N225 (N, 3.1)	N165 (O, 2.6) S131 (N, 3.1)	D118 (O, 2.6) N84 (N, 3.1)	A212 (O, 2.7) Q178 (N, 3.3)
H ₂ O	D73 (O, 2.7) I35 (O, 2.8)	D26 (O, 2.6) I223 (O, 2.9)	D167 (O, 2.7) V129 (O, 2.8)	D120 (O, 2.7) I82 (O, 2.6)	D214 (O, 2.6) I176 (O, 2.8)

In the first column, the ligand atom or mediating water molecule is given. The other columns show to which protein atoms they bind in all five binding sites (NAG 1–NAG 5). H₂O denotes the mediating water molecule as shown in Figure 7. For example, O7 M76 (N, 2.8) reads as: O7 of GlcNAc forms a hydrogen bond to backbone nitrogen (N) of Met76 with a distance of 2.8 Å (O denotes a backbone carbonyl oxygen).

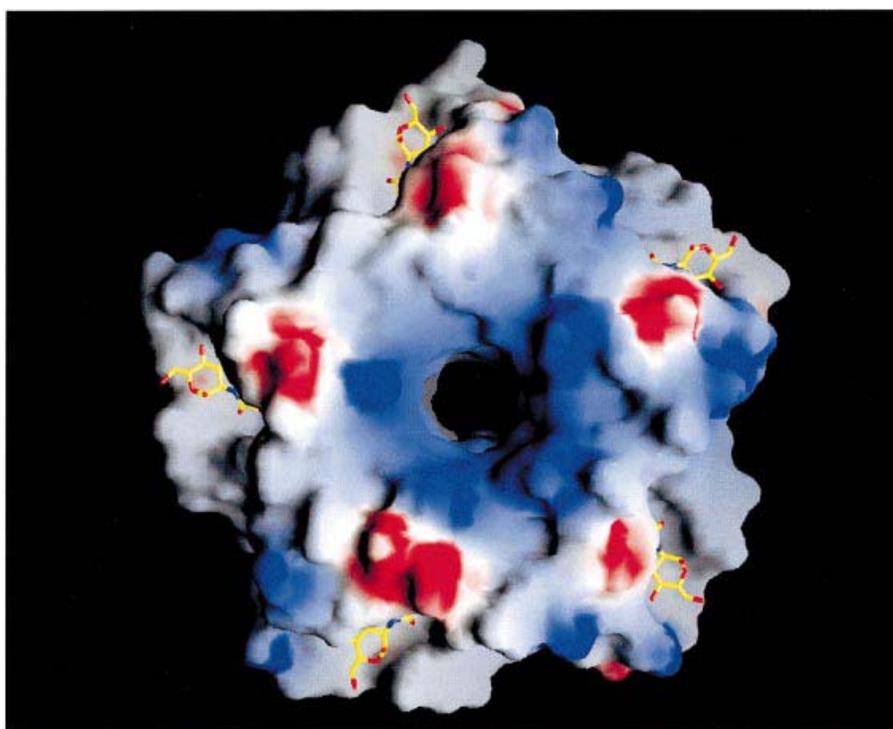


Fig. 5. Space-filling model of tachylectin-2 showing a surface potential view with the five binding sites. The GlcNAc molecules are shown as ball-and-stick models. The acetamido group is pointing inside. The water molecules in the central tunnel are not shown. Positively charged areas are displayed in blue, negative charge is shown in red. The plot was made with GRASP (Nicholls *et al.*, 1991).

Trp169. The right side of the binding pocket is bounded by the side chains of Asn168 and Leu170. On the left side, the pocket boundary is built up by the main chain atoms of Gly130–Trp134 and the side chain of Trp134. Thus, the back of the binding pocket is hydrophobic (Trp134/Phe137/Leu170), and the left wall is more polar than the right wall. The GlcNAc is bound with the acetamido group pointing inside the pocket and the 6-OH group pointing outside into the solvent region. Of the sugar functional groups, the acetamido group and the 3-OH and 4-OH groups are used for ligand recognition.

Most polar protein–carbohydrate interactions are accomplished by protein backbone carbonyl oxygens and NH groups or via a water molecule in contact with protein

groups. In spite of the apparent unimportance of their side chain, many of the amino acid residues involved in carbohydrate recognition and binding are conserved between all five tandem repeats of the protein sequence (Figure 2). Some of the conserved residues have common functions, such as Leu7 anchored in a hydrophobic region, or Tyr16 where a bulky surface residue is needed in the protein fold. Residue Ala30 is located in the helical segment of the 3→4 loop, and Ala32 is located at the end of this segment. Of special interest are those conserved residues building up the binding site. Of the residues shown in Figure 6, both tryptophans, Gly132, Phe137 and Asp167 are strictly conserved in all binding sites. Val129 is an isoleucine in all other binding sites, whereas Ser131

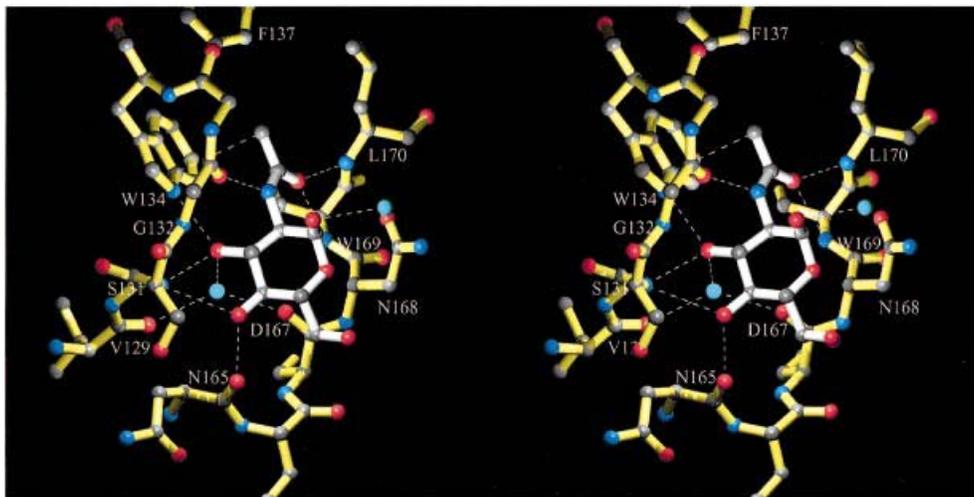


Fig. 6. Stereo plot showing one binding site geometry (blade IV) of tachylectin-2 in complex with GlcNAc. All five binding sites are equivalent and details are given in Table I. Except for Trp134, all other polar protein–sugar interactions are made with the protein backbone. The mediating water (cyan colored spheres) molecule coordinating to the sugar 3-OH group is conserved in all five binding sites, whereas the non-binding sugar oxygen atoms occasionally are coordinated to water molecules in the crystal structure. The plot was made with MOLSCRIPT (Kraulis, 1991) and rendered with POVRay (1997).

and Asn165 are not conserved. Leu170 of the hydrophobic backside of the binding pocket is not strictly conserved, but functionally duplicated by Met76 in blade II and by Ile123 in blade III. The strictly conserved Gly36 is obligatory because any side chain in this position would occlude the sugar-binding site in each β -sheet. The highly conserved sequence providing the carbohydrate-binding site seems to be necessary for creating five identical binding sites and, therefore, five equivalent environments on the protein surface utilizing the 5-fold β -propeller symmetry.

The 3→4 loop of each β -sheet starts with ProPro (except Tyr20/Pro21) and continues with the sequence **XXXXDNWLARA** (strictly conserved residues in bold, italics for residues conserved in at least three of five positions, Figure 2). This is the second half of the loop towards the C-terminus carrying the conserved stabilizing features. The Asp167 (blade IV) forms a salt bridge with Arg172, which itself is hydrogen-bonded to the backbone carbonyl oxygen of the non-conserved residue Leu160. The Leu160 carbonyl oxygen additionally participates in a hydrogen bond to the side chain nitrogen of the conserved tryptophan residue Trp169. Thus, of the sequence **XXXXDNWLARA**, the residues underlined are involved in ligand interactions only via their backbone oxygen and nitrogen atoms, while the conserved side chains are essential for building and stabilizing the carbohydrate recognition sites in all five β -sheets, made in a virtually identical manner (Figure 7). This first half of the binding site (right side in Figure 6) hydrogen-bonds to the acetamido carbonyl group, the water molecule mediating between the sugar 3-OH group and the carbonyl oxygen of Asp167, and the sugar 4-OH group via the Asn165 carbonyl oxygen (Figure 6).

The opposite wall of the binding pocket (left side in Figure 6), recognizing the nitrogen and the methyl group of the acetamido group and additionally the 3-OH and the 4-OH group of GlcNAc, is made up of the sequence **IGXGGW** (strictly conserved residues in bold, italics for residues conserved in at least three of five positions;

Figure 2). It is formed by the connecting segment from the previous β -sheet (Figure 7). The starting isoleucine or valine (Val129) carbonyl oxygen is used together with the conserved Asp167 backbone carbonyl oxygen for fixing the water molecule which binds to the sugar 3-OH group. The non-conserved X in **IGXGGW** uses its backbone NH for a hydrogen bond to the sugar 4-OH group together with the X in **XXXXDNWLARA** described above. The 3-OH group of GlcNAc is coordinated by the tryptophan side chain nitrogen and also by the backbone NH of X in **IGXGGW** and the mediating water molecule.

The requirement for a free 4-OH group of the binding sugar molecule is evident from Figure 6. The 4-OH group is used for carbohydrate binding, with hydrogen bonding to two protein backbone atoms explaining the affinity of tachylectin-2 for β -D-Gal-[1,6]-D-GlcNAc. This also explains the lack of affinity for β -D-Gal-[1,4]-D-GlcNAc with a bulky substituent in the 4-position. The 17-fold higher affinity for GlcNAc compared with GalNAc is in agreement with the interaction of the 4-OH group, which in the bioactive conformation exhibits the less favorable axial position in GalNAc (Figure 6). The five equivalent carbohydrate-binding sites in tachylectin-2 reside in a plane on one face of the water-soluble molecule (Figure 5).

Besides the free 3-OH and 4-OH groups, the acetamido group of GlcNAc is essential for binding to tachylectin-2 (Okino *et al.*, 1995). Its NH group is a hydrogen donor for the carbonyl oxygen of G in **IGXGGW**, while the acyl oxygen is hydrogen-bonded to both backbone N atoms of WL in **XXXXDNWLARA**. The acetamido methyl group is in van der Waals contact with Trp134 and is pointing into the hydrophobic pocket made by the side chains Phe137, Trp134 and Leu170. Among the proteins for which structural information about the binding of *N*-acetyl sugars is available are wheat germ agglutinin (WGA) and influenza virus hemagglutinin (HA). In the structure of WGA in complex with *N*-acetylneuraminyl-lactose (Wright, 1990), the aromatic side chain of a tyrosine is in van der Waals contact with the acetamido methyl group of the sugar, additionally forming a hydrogen

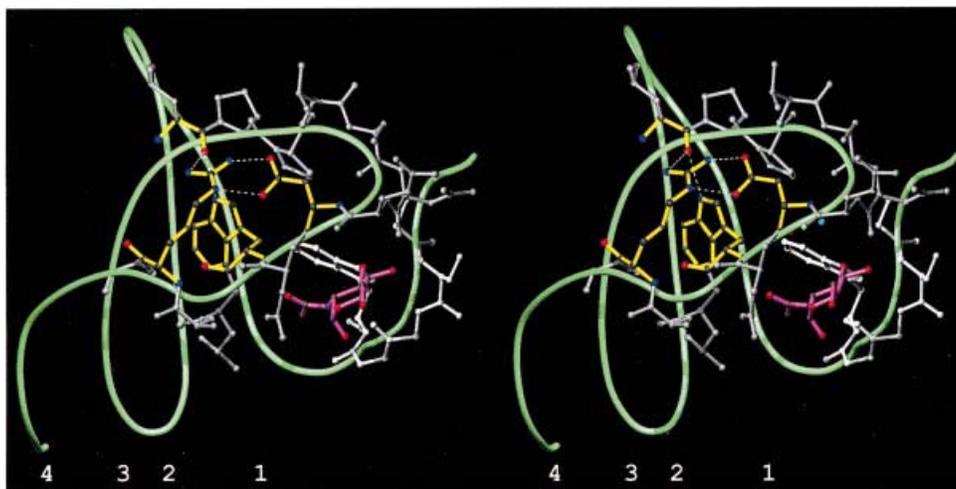


Fig. 7. Stereo plot of the tachylectin-2 binding site in one complete β -sheet (blade IV). The GlcNAc molecule is shown with pink colored bonds. The 3 \rightarrow 4 loop of the β -sheet and the connecting segment from the previous β -sheet are drawn in light gray and white, respectively. The binding loop-stabilizing features (Asp, Arg, Trp and backbone C=O) are colored with yellow bonds. The plot was made with MOLSCRIPT (Kraulis, 1991) and rendered with POVRay (1997). The β -strands of the β -sheet shown are numbered 1–4 as defined in Figure 2.

Table II. Crystal parameters

Crystal	Unit cell parameters	
	$a = b$ (Å)	c (Å)
NAT11	89.42	73.38
NAT12	90.61	71.43
Pb(OAc) ₂	90.84	71.58
Me ₃ PbCl	90.71	71.69
Double derivative	90.82	71.54
GlcNAc	89.46	73.62

All crystals belong to space group P3₁21.

bond from the aromatic hydroxyl group to an OH of the neuraminy part of the ligand; the acetamido NH is coordinated to a glutamic acid COO⁻, while the acetamido carbonyl group is not involved in binding the *N*-acetyl group. In the structure of influenza virus HA in complex with α (2,6) sialyllactose (Weis *et al.*, 1988), the sialic acid acetamido methyl group is in van der Waals contact with a tryptophan, while the acetamido NH group forms a hydrogen bond to a glycine C=O, and the sugar acetamido C=O forms a hydrogen bond to a leucine side chain. In the case of chicken hepatic lectin (CHL), which selectively binds to GlcNAc, there is no experimental structure available, but a model of the binding site–ligand complex is based on the X-ray structure of the liver mannose-binding protein (MBP-C) with α -methyl-*N*-acetyl-D-glucosamine. The results strongly suggest that GlcNAc binds to CHL in the same way as to MBP-C, but selectivity is achieved by non-polar contacts of two additional residues (Val191/Tyr195) with the 2-acetamido substituent of the bound ligand (Burrows *et al.*, 1997). In all the cases, including tachylectin-2, a van der Waals contact of the *N*-acetyl methyl group to an aromatic side chain is found. This implies the importance of non-polar interactions of the acetamido group for the *N*-acetyl recognition, since the presence of the *N*-acetyl group is also essential for carbohydrate recognition by tachylectin-2.

The non-polar patch, formed by the aliphatic protons and carbons at the various epimeric centers of a sugar

molecule, is usually packed against the face of aromatic side chains (Weis and Drickamer, 1996). In tachylectin-2, this patch is not stacked to any aromatic side chain, but packed against a strictly conserved asparagine (Asn168, Figures 2 and 4). The non-polar face of the sugar ring in the WGA complex (isolectin 2, Wright, 1990) as well as in the influenza virus HA complex (Weis *et al.*, 1988) is also not stacked to an aromatic side chain, but packed against histidine or serine, respectively. The lack of interaction of the non-polar sugar face with an aromatic side chain seems to be important for the discrimination between GlcNAc/GalNAc and non-acetylated ligands, since the interactions of the sugar acetamido group compensate for it.

The 1-OH and 6-OH groups as well as the pyranose ring oxygen are not used for recognition and, therefore, are not coordinated by protein groups. In the crystal structure of the complex, some of these oxygen atoms are hydrogen-bonded to water molecules. The 6-OH group of GlcNAc points into the solvent region, and substitution at this position should not affect ligand binding to tachylectin-2. Since *N,N'*-diacetylchitobiose (2-acetamido-2-deoxy-4-*O*-[2-acetamido-2-deoxy- β -D-glucopyranosyl]-D-glucopyranose) binds to tachylectin-2 with a minimum concentration for inhibition of hemagglutinating activity 32-fold higher than for GlcNAc, and *N,N,N'*-triacetylchitotriose with a 64-fold higher concentration for inhibition, 1-OH substitution in the β -configuration does not prevent binding to tachylectin-2 as long as the substituent exhibits no strong interference with the strictly conserved Asn168 (and equivalents, Figure 4).

Conclusions

The nature of the binding pocket explains the strict specificity of tachylectin-2 for GlcNAc and GalNAc residues with a free 4-OH group and the preference for GlcNAc. All data presented here strongly suggest tachylectin-2 to be a host defense molecule recognizing GlcNAc or GalNAc units on microbial surfaces. Since β -linked *N,N'*-diacetylchitobiose as well as *N,N,N'*-triacetylchitotriose, which are essentially very short

Table III. Data collection and MIR phasing statistics

	Data sets					
	NAT11	NAT12	Pb(OAc) ₂	Me ₃ PbCl	Double derivative	GlcNAc
Resolution (Å)	28.4–2.0	20.3–2.2	20.4–2.4	20.9–2.8	24.6–2.2	15.2–2.0
Measurements	80 354	50 142	45 687	33 985	68 282	83 799
Unique measurements	24 455	16 641	12 986	8625	17 329	20 896
% Complete (last shell/Å)	98.0 (90.2) (2.02–1.95)	94.7 (71.8) (2.27–2.19)	95.1 (72.2) (2.48–2.40)	99.6 (98.3) (2.90–2.80)	98.5 (95.8) (2.28–2.20)	93.7 (92.2) (2.03–2.43)
Mean <i>I</i> /Sig <i>I</i>	11.9	16.8	15.0	12.5	21.8	12.9
<i>R</i> _{sym} (%) last shell	7.2 (31.5)	8.5 (37.0)	9.9 (36.7)	10.6 (32.7)	7.6 (36.2)	7.4 (17.1)
<i>R</i> _{iso} (%)	–	–	15.4	9.0	12.9	–
<i>R</i> _{Cullis-iso}	–	–	0.57	0.74	0.43	–
<i>P</i> _{iso}	–	–	2.92	1.36	3.95	–
Residues for phase calculation (Å)	–	20.0–2.2	20.0–2.5	20.0–3.0	20.0–2.2	–
Mean FOM	–	0.57	–	–	–	–

$$R_{\text{sym}} = \frac{\sum |I(h)_i - \langle I(h) \rangle|}{\sum \langle I(h) \rangle}$$

$$R_{\text{iso}} = \frac{\sum |F_{\text{PH}} - F_{\text{P}}|}{\sum F_{\text{P}}}$$

*R*_{Cullis-iso} = r.m.s. lack of closure/r.m.s isomorphous difference.

*P*_{iso} (phasing power) = $\langle |F_{\text{H}}| \rangle$ /r.m.s. lack of closure for all reflections.

Mean FOM, mean figure of merit; NAT11, native data set used for final refinement of tachylectin-2 model; NAT12, native data set used for MIR phasing and initial model building.

Table IV. Refinement statistics

	Tachylectin-2	Tachylectin-2–GlcNAc
Resolution range (Å)	15.0–2.0	15.0–2.0
Reflections in working set	20 560	18 820
Reflections in test set	2297	2072
<i>R</i> _{cryst} (%)	18.6	16.2
<i>R</i> _{free} (%)	22.9	20.2
Protein atoms (non-H)	1901	1905
Solvent atoms (non-H)	129	158
Ligand atoms (non-H)	–	75
Average <i>B</i> -factor (Å ²)	27.1	27.8
R.m.s. Δ <i>B</i> (2 Å cut-off)	3.1	2.7
Deviations from ideality (r.m.s.)		
Bond lengths (Å)	0.009	0.009
Bond angles (°)	1.6	1.6

$$R_{\text{cryst}} = \frac{\sum_{hkl \in W} ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum_{hkl \in W} |F_{\text{obs}}|}$$

$$R_{\text{free}} = \frac{\sum_{hkl \in W} ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum_{hkl \in W} |F_{\text{obs}}|}$$

fragments of, in horseshoe crabs, the ubiquitous polysaccharide chitin, also bind (weakly) to tachylectin-2, self-recognition of chitin by tachylectin-2 seems to be prevented mainly because there is no free 4-OH group in the β-1,4-linked D-GlcNAc units of chitin, except in distant terminal units.

Multiple binding of the five binding sites to repetitive structures will generate very tight interactions, despite the comparatively weak affinity of a single binding site. There seems to be no cooperativity, as the protein structure does not change at all upon sugar binding, but there is additivity, as binding of a further ligand of a multidomainial target is more favorable if one or more binding sites are already occupied. The tight binding of sufficiently separated

GlcNAc/GalNAc units with free 4-OH groups of the horseshoe crab itself will therefore be prevented. This mechanism of self/non-self distinction is reinforced by a distance of two individual binding sites of 25 and 40 Å, respectively (according to the pentagonal geometry), and seems to prevent multiple binding to distant terminal GlcNAc units of chitin. Thus, tachylectin-2 recognizes surface structures with a high density of GlcNAc/GalNAc groups. In vertebrates, mannose-binding protein (MBP) belonging to the collectins (reviewed in Turner, 1996) features carbohydrate surface recognition employing this principle. Its three-lectin by domain cluster has three sugar-binding sites separated 53 Å (Weis and Drickamer, 1994), making the simultaneous binding of several ligands less probable for carbohydrate structures with a low ligand density.

The *in vivo* mechanism of tachylectin-2 action in *T.tridentatus* remains to be elucidated. Probably, it agglutinates oceanic microorganisms coming into contact with horseshoe crab hemolymph in the same way that it agglutinates *S.saprophyticus* KD *in vitro* (Okino *et al.*, 1995). Furthermore, tachylectin-2 may function in an opsonin-like way by marking the surface of invading microorganisms as a signal for killing them. Yet another possible mechanism of tachylectin-2 action is blocking of microbial surfaces, thus preventing their adhesion and invasion of host cells.

Tachylectin-2 seems to be designed for recognition of quite specific surface carbohydrate groups. The only known bacteria to which tachylectin-2 binds are Gram-positive *Staphylococcus* species, which have a special GlcNAc-substituted LTA. In the absence of GlcNAc substitution, no binding occurs. Tachylectin-1, in contrast, recognizes Gram-positive bacteria in a more promiscuous way, binding to unsubstituted LTA, as well as Gram-negative bacteria, employing ligands from the LPS core region.

Thus, tachylectin-2 complements the broad specificity of host defense proteins such as tachylectin-1 with the specialized feature of GlcNAc/GalNAc recognition.

Materials and methods

Protein purification

Tachylectin-2 can be isolated in at least three isoforms (a, b and c). Compared with isoforms a and b, isoform c exhibits the amino acid exchanges I129V and H213Y. Tachylectin-2 (isoform b) was purified according to the procedure described in Okino *et al.* (1995). The protein solution used for crystallization was 10 mg/ml in 10 mM MOPS, pH 7.1.

Crystallization, data collection and derivatization

Crystallization was carried out in CrysChem plates by the vapor diffusion method using 100 mM sodium acetate, pH 4.6, 2.0 M sodium formate as precipitant against 800 μ l of precipitant in the reservoir and sitting drops set up by 2.2 μ l of protein solution and 1.5 μ l of precipitation buffer. Within 2 days, crystals (Table II) appeared at 20°C up to a final size of $\sim 600 \times 200 \times 150 \mu\text{m}^3$. They belong to the trigonal space group P3₁21 and have unit cell dimensions $a = b = 89.42 \text{ \AA}$, $c = 73.38 \text{ \AA}$ (NAT11), one molecule per asymmetric unit and $V_M = 3.14 \text{ \AA}^3/\text{Da}$ corresponding to a 61% solvent content (Matthews, 1968).

To prepare tachylectin-2 heavy atom derivatives, the crystals were soaked in a solution of the heavy atom compound in freshly prepared precipitation buffer. Harvesting the crystals in fresh precipitant solution changed the unit cell dimensions to $a = b = 90.61 \text{ \AA}$, $c = 71.43 \text{ \AA}$ (Table II). Therefore, a second native data set (NAT12) was measured in order to analyze the heavy atom derivative diffraction data. Interpretable derivatives were obtained with Pb(AcO)₂ (10 mM, 10 days), Me₃PbCl (10 mg/ml, 2 days) and the double derivative Pb(AcO)₂/Me₃PbCl (5 mg/ml each, 4 days).

The complex of tachylectin-2 with GlcNAc was prepared by co-crystallizing the protein using the conditions described above, but with 5 mM GlcNAc (Sigma) added to the precipitation buffer. These crystals belong to the same crystal form as the native ones (Table II).

All diffraction data were collected using a 300 mm MAR Research (Hamburg, Germany) image plate detector mounted on a Rigaku (Tokyo, Japan) RU200 rotating anode X-ray generator with graphite monochromatized CuK α radiation. All image plate data were processed with MOSFLM (Leslie, 1991) and the CCP4 program suite (Collaborative Computational Project, Number 4, 1994).

Phase calculation, model building and refinement

The structure of tachylectin-2 was solved by the multiple isomorphous replacement (MIR) method using the three heavy atom derivatives described above. All derivative data were analyzed with the native data set NAT12, first using isomorphous difference Patterson maps. Heavy atom locations were confirmed by difference Fourier methods with appropriate initial single isomorphous replacement phases using CCP4 programs, as well as by Patterson vector superposition methods implemented in SHELX-97 (Sheldrick, 1991). The refinement of heavy atom parameters and calculation of MIR phases were done with SHARP (La Fortelle and Bricogne, 1997). The final parameters are given in Table III. The initial MIR phases were improved with SOLOMON (Abrahams and Leslie, 1996), resulting in a 2.2 \AA electron density map that was interpretable in terms of a protein structure. All model building was done with FRODO (Jones, 1978). Refinement was performed by conjugate gradient and simulated annealing protocols as implemented in X-PLOR 3.851 (Brunger *et al.*, 1987). The final cycle of refinement was performed with REFMAC (Murshudov *et al.*, 1997) using the conjugate direction method with a maximum likelihood residual and a bulk solvent model. All protocols included refinement of individual isotropic B -factors. In order to use the higher resolution (2.0 \AA) data set NAT11 for refinement, the initial model built in the NAT12 cell was placed with X-PLOR rigid body refinement in the NAT11 cell, and used for all subsequent refinement cycles of tachylectin-2. The initial R -factor for the first refinement cycle of the unrefined model including all side chains was 31.0% ($R_{\text{free}} = 35.6\%$). It dropped to 18.6% ($R_{\text{free}} = 22.9\%$, resolution range 15.0–2.0 \AA) for the final model including 129 water molecules. The water model was calculated using ARP (Lamzin and Wilson, 1993) and verified by visual inspection. The final refinement statistics are shown in Table IV.

The X-ray structure of tachylectin-2 in complex with GlcNAc was solved using the model of the native protein excluding water molecules and calculating $2mF_o - DF_c$ and $mF_o - DF_c$ style electron density maps after five cycles of refinement with REFMAC. After adding the sugar molecules, the model was refined further by X-PLOR and REFMAC using the same protocols as described above. The water model was calculated and refined with ARP. The R -factor of the final model

including 158 water molecules was 16.2% ($R_{\text{free}} = 20.2\%$, resolution range 15.0–2.0 \AA). The final refinement statistics are shown in Table IV. In a Ramachandran plot of tachylectin-2 in complex with GlcNAc calculated with PROCHECK (Laskowski *et al.*, 1993), 91.1% of the residues lie in the most favored regions, 8.9% in additionally allowed regions.

The atomic coordinates of tachylectin-2 in complex with GlcNAc have been deposited in the Protein Data Bank with entry code 1tl2.

References

- Abrahams, J.P. and Leslie, A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Crystallogr.*, **D52**, 30–42.
- Baker, S.C., Saunders, N.F.W., Willis, A.C., Ferguson, S.J., Hajdu, J. and Fulop, V. (1997) Cytochrome *cd*₁ structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to β -propeller folds. *J. Mol. Biol.*, **269**, 440–455.
- Bergner, A., Oganessyan, V., Muta, T., Iwanaga, S., Typke, D., Huber, R. and Bode, W. (1996) Crystal structure of coagulogen, the clotting protein from horseshoe crab: a structural homologue of nerve growth factor. *EMBO J.*, **15**, 6789–6797.
- Brunger, A.T., Kuriyan, J. and Karplus, M. (1987) Crystallographic R factor refinement by molecular dynamics. *Science*, **235**, 458–460.
- Burrows, L., Iobst, S.T. and Drickamer, K. (1997) Selective binding of *N*-acetylglucosamine to chicken hepatic lectin. *Biochem. J.*, **324**, 673–680.
- Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Faber, H.R., Groom, C.R., Baker, H.M., Morgan, W.T., Smith, A. and Baker, E.N. (1995) 1.8 \AA crystal structure of the C-terminal domain of rabbit serum haemopexin. *Structure*, **3**, 551–559.
- Fearon, D.T. and Locksley, R.M. (1996) The instructive role of innate immunity in the acquired immune response. *Science*, **272**, 50–54.
- Inamori, K., Saito, T., Iwaki, D., Nagira, T., Iwanaga, S., Arisaka, F. and Kawabata, S. (1999) A newly identified horseshoe crab lectin with specificity for blood group A antigen recognizes specific O-antigens of bacterial lipopolysaccharides. *J. Biol. Chem.*, **274**, 3272–3278.
- Ito, N., Phillips, E.V., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D.S. and Knowles, P.F. (1991) Novel thioether bond revealed by a 1.7 \AA crystal structure of galactose oxidase. *Nature*, **350**, 87–90.
- Iwanaga, S., Kawabata, S. and Muta, T. (1998) New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. *J. Biochem.*, **123**, 1–15.
- Jones, T.A. (1978) A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.*, **11**, 268–272.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- La Fortelle, E. de and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.
- Lamzin, V.S. and Wilson, K.S. (1993) Automated refinement of protein models. *Acta Crystallogr.*, **D49**, 129–147.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Leslie, A.G.W. (1991) Molecular data processing. In Moras, D., Podjarny, A.D. and Thierry, J.C. (eds), *Crystallographic Computing 5*. Oxford University Press, Oxford, pp. 50–61.
- Li, J. *et al.* (1995) Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed β -propeller. *Structure*, **3**, 541–549.
- Mak, T.C.W. and Zhou, G. (1997) *Crystallography in Modern Chemistry*. John Wiley, Chichester, pp. 1174.
- Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
- Medzhitov, R. and Janeway, C.A. (1997) Innate immunity: the virtues of a nonclonal system of recognition. *Cell*, **91**, 295–298.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **D53**, 240–255.
- Murzin, A.G. (1992) Structural principles for the propeller assembly of β -sheets: the preference for seven-fold symmetry. *Proteins*, **14**, 191–201.

- Muta,T., Seki,N., Takaki,Y., Hashimoto,R., Oda,T., Iwanaga,A., Tokunaga,F. and Iwanaga,S. (1995) Purified horseshoe crab factor G. Reconstitution and characterization of the (1→3)-β-D-glucan-sensitive serine protease cascade. *J. Biol. Chem.*, **270**, 892–897.
- Muta,T. and Iwanaga,S. (1996) Clotting and immune defense in Limulidae. In Rinkevich,B. and Muller,W.E.G. (eds), *Progress in Molecular and Subcellular Biology. Vol. 15, Invertebrate Immunology*. Springer-Verlag, Berlin, pp. 154–189.
- Nakamura,S., Iwanaga,S., Harada,T. and Niwa,M. (1976) A clottable protein (coagulogen) from amoebocyte lysate of japanese horseshoe crab (*Tachypleus tridentatus*). Its isolation and biochemical properties. *J. Biochem.*, **80**, 1011–1021.
- Nakamura,T., Tokunaga,F., Morita,T., Iwanaga,S., Kusumoto,S., Shiba,T., Kobayashi,T. and Inoue,K. (1988) Intracellular serine-protease zymogen, factor C, from horseshoe crab hemocytes. Its activation by synthetic lipid A analogues and acidic phospholipids. *Eur. J. Biochem.*, **176**, 89–94.
- Nicholls,A., Sharp,K. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.*, **11**, 281–296.
- Okino,N., Kawabata,S., Saito,T., Hirata,M., Takagi,T. and Iwanaga,S. (1995) Purification, characterization and cDNA cloning of a 27-kDa lectin (L10) from horseshoe crab hemocytes. *J. Biol. Chem.*, **270**, 31008–31015.
- POVRay program, version 3.02 (1997) URL: <http://www.povray.org>
- Ruhland,G.J. and Fiedler,F. (1990) Occurrence and structure of lipoteichoic acids in the genus *Staphylococcus*. *Arch. Microbiol.*, **154**, 375–379.
- Saito,T., Kawabata,S., Hirata,M. and Iwanaga,S. (1995) A novel type of limulus lectin—L6. *J. Biol. Chem.*, **270**, 14493–14499.
- Saito,T., Kawabata,S., Hirata,M. and Iwanaga,S. (1997) A newly identified horseshoe crab lectin with binding specificity to O-antigen of bacterial lipopolysaccharides. *J. Biol. Chem.*, **272**, 30703–30708.
- Sheldrick,G. (1991) Tutorial on automated Patterson interpretation to find heavy atoms. In Moras,D., Podjarny,A.D. and Thierry,J.C. (eds), *Crystallographic Computing 5*. Oxford University Press, Oxford, pp. 145–157.
- Tanaka,S. and Iwanaga,S. (1993) Limulus test for detecting bacterial endotoxins. *Methods Enzymol.*, **223**, 358–364.
- Toh,Y., Mizutani,A., Tokunaga,F. and Iwanaga,S. (1991) Morphology of the granular hemocytes of the japanese horseshoe crab *Tachypleus tridentatus* and immunocytochemical localization of clotting factors and antimicrobial substances. *Cell Tissue Res.*, **266**, 137–147.
- Turner,M.W. (1996) Mannose-binding lectin: the pluripotent molecule of the innate immune system. *Immunol. Today*, **11**, 532–540.
- Weis,W., Brown,J.H., Cusack,S., Paulson,J.C., Skehel,J.J. and Wiley,D.C. (1988) Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature*, **333**, 426–431.
- Weis,W.I. and Drickamer,K. (1994) Trimeric structure of a C-type mannose-binding protein. *Structure*, **2**, 1227–1240.
- Weis,W.I. and Drickamer,K. (1996) Structural basis of the lectin–carbohydrate recognition. *Annu. Rev. Biochem.*, **65**, 441–473.
- Wright,C.S. (1990) 2.2 Å resolution analysis of two refined N-acetylneuraminyl-lactose–wheat germ agglutinin isolectin complexes. *J. Mol. Biol.*, **215**, 635–651.
- Xia,Z., Dai,W., Xiong,J. and Hao,Z. (1992) The three-dimensional structures of methanol dehydrogenases from two methylotrophic bacteria at 2.6 Å resolution. *J. Biol. Chem.*, **267**, 22289–22297.

Received December 12, 1998; revised and accepted March 2, 1999