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

Hans-Georg Beisel¹,², Shun-ichiro Kawabata³,⁴, Sadaaki Iwanaga³, Robert Huber¹ and Wolfram Bode¹

¹Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, 82152 Martinsried, Germany, ²Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-82 and ³CREST, Japan Science and Technology Corporation, Japan

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 invertebrates as well as in vertebrates will elucidate further the basic mechanisms used by animals to distinguish self and non-self materials.

The invertebrate *Tachylepus 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, α2-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 tachyllectin-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 $K_a = 1.95 \times 10^3/M$) and N-acetyl-d-galactosamine (GalNAc, $K_a = 1.11 \times 10^3/M$), but no affinity toward glucose, glucosamine, galactose or galactosamine. It binds to N-acetyllactosamine (β-D-Gal-[1,6]-D-GlcNAc, $K_a = 1.55 \times 10^4/M$), but not to N-acetyllactosamine (β-D-Gal-[1,4]-D-GlcNAc). It also binds to N,N'-diacetylatedibiose 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''-triacetylatedriotriose 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 CH$_2$OH (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 two (1–2)- linked GlcNAc as an additional substituent of its LTA polyglycosphate (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$_n$ 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 brings 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 Cα 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 Cα 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 XXXXDNWL is part of the 3–4 loop of each β-sheet, starting with the underlined residues, whereas GXXGWW belongs to the connecting segment between two adjacent β-sheets. The β-sheets are named 1–V, and individual β-strands in each β-sheet 1–4.
which the $\beta$-sheets are rotated around the central $n$-fold axis (ideally parallel to the inner $\beta$-strands; Murzin, 1992) to generate the cyclic arrangement of the protein fold is $72^\circ$ for $n = 5$ and even $90^\circ$ with $n = 4$. Thus, $\beta$-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 $\beta$-sheets. In tachylectin-2, this gap is filled, in particular, by the $3\to4$ loop of each $\beta$-sheet and the connecting segment between adjacent $\beta$-sheets. In tachylectin-2, this gap is filled, in particular, by the $3\to4$ loop of each $\beta$-sheet and the connecting segment between adjacent $\beta$-sheets, as well as by large side chains in the outer strands. In $\beta$-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 $\beta$-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, superimposed with the electron density map. The binding sites are all located between the connecting segment of adjacent $\beta$-sheets and the $3\to4$ loop of the following $\beta$-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$_\alpha$ 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 $3\to4$ loop of each $\beta$-sheet, while the second part is made by the connecting segment from the previous $\beta$-sheet (Figure 2). All $3\to4$ loops are located on the ‘exit’ side of the tachylectin-2 propeller, while all $\beta$-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 1. Binding sites of GlcNAc (NAG)

<table>
<thead>
<tr>
<th>Ligand atom</th>
<th>NAG 1</th>
<th>NAG 2</th>
<th>NAG 3</th>
<th>NAG 4</th>
<th>NAG 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>O7</td>
<td>M76 (N, 2.8)</td>
<td>L29 (N, 3.2)</td>
<td>L170 (N, 2.9)</td>
<td>I123 (N, 2.9)</td>
<td>L217 (N, 3.0)</td>
</tr>
<tr>
<td></td>
<td>W75 (N, 3.3)</td>
<td>W28 (N, 3.3)</td>
<td>W169 (N, 3.1)</td>
<td>W122 (N, 3.2)</td>
<td>W216 (N, 3.2)</td>
</tr>
<tr>
<td>C8</td>
<td>W40 (CE2, 3.6)</td>
<td>W228 (CE2, 3.5)</td>
<td>W134 (CE2, 3.6)</td>
<td>W87 (CE2, 3.6)</td>
<td>W181 (CE2, 3.6)</td>
</tr>
<tr>
<td>N2</td>
<td>G38 (O, 3.1)</td>
<td>G226 (O, 3.0)</td>
<td>G132 (O, 3.0)</td>
<td>G85 (O, 2.8)</td>
<td>G179 (O, 2.9)</td>
</tr>
<tr>
<td>O3</td>
<td>W40 (NE1, 3.0)</td>
<td>W228 (NE1, 2.8)</td>
<td>W134 (NE1, 2.8)</td>
<td>W87 (NE1, 2.8)</td>
<td>W181 (NE1, 2.9)</td>
</tr>
<tr>
<td></td>
<td>H2O (2.8)</td>
<td>H2O (2.8)</td>
<td>H2O (2.7)</td>
<td>H2O (2.6)</td>
<td>H2O (2.7)</td>
</tr>
<tr>
<td>O4</td>
<td>D71 (O, 2.5)</td>
<td>K24 (O, 2.6)</td>
<td>N165 (O, 2.6)</td>
<td>D118 (O, 2.6)</td>
<td>A212 (O, 2.7)</td>
</tr>
<tr>
<td></td>
<td>K37 (N, 3.0)</td>
<td>N225 (N, 3.1)</td>
<td>S131 (N, 3.1)</td>
<td>N84 (N, 3.1)</td>
<td>Q178 (N, 3.1)</td>
</tr>
<tr>
<td>H2O</td>
<td>D73 (O, 2.7)</td>
<td>D26 (O, 2.6)</td>
<td>D167 (O, 2.7)</td>
<td>D120 (O, 2.7)</td>
<td>D214 (O, 2.6)</td>
</tr>
<tr>
<td></td>
<td>I35 (O, 2.8)</td>
<td>I223 (O, 2.9)</td>
<td>V129 (O, 2.8)</td>
<td>I82 (O, 2.6)</td>
<td>I176 (O, 2.8)</td>
</tr>
</tbody>
</table>

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 a isoleucine in all other binding sites, whereas Ser131...
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 conserved sequence providing the carbohydrate-binding sites in tachylectin-2 reside in a 5-fold β-propeller symmetry.

The 3→4 loop of each β-sheet starts with ProPro (except Tyr20/Pro21) and continues with the sequence XXXXDNWL/RA (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 XXXXDNWL/RA, 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 /GXGGW (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 /GXGGW uses its backbone NH for a hydrogen bond to the sugar 4-OH group together with the X in XXXXDNWL/RA 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 /GXGGW 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 /GXGGW, while the acyl oxygen is hydrogen-bonded to both backbone N atoms of WL in XXXXDNWL/RA. 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-acetyleneuraminyl-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
bond from the aromatic hydroxyl group to an OH of the neuraminyl 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 mannosé-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 (isolecitin 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-4-0-[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, which are essentially very short...
The prevention of self-recognition of chitin by tachylectin-2 seems to be mainly because there is no free 4-OH group in the -1,4-linked D-GlcNAc units of chitin, except in distant terminal units. Thus, tachylectin-2 recognizes GlcNAc 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, 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 CryxChem plates by the vapor diffusion method using 100 mM sodium acetate, pH 4.6, 2.0 M sodium formate as precipitant against 800 μl of precipitant in the reservoir and sitting drops set up by 2.2 μl of protein solution and 1.5 μl of precipitant buffer. Within 2 days, crystals (Table II) appeared at 20°C up to a final size of ~600×200×150 μm³. They belong to the trapezoidal space group P3₁21 and have unit cell dimensions a = b = 89.42 Å, c = 73.38 Å (NATI1), one molecule per asymmetric unit and V_M = 3.14 Å³ 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 Å, c = 71.43 Å (Table II). Therefore, a second native data set (NATI2) was measured in order to analyze heavy atom derivative diffraction data. Interpretable derivatives were obtained with Pb(AcO)₂ (10 mM, 10 days), Me₃PbCl (Table II). Therefore, a second native data set (NATI2) was measured in order to improve the native data set.

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 mono-chromatized CuKα 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 NATI2, first using isomorphous difference Patterson maps. Heavy atom locations were confirmed by difference Fourier methods with appropriate 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 Bricegno, 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 Å 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 FRODO and REFMAC (Murshudov, 1993) 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 Å) data set NATI1 for refinement, the initial model built in the NATI1 cell was placed with X-PLOR rigid body refinement in the NATI2 cell, and used for all subsequent refinement cycles of tachylectin-2. The initial R-factor for the first refinement cycle of the unreined model including all side chains was 31.9% (R_free = 35.6%). It dropped to 18.6% (R_free = 22.9%, resolution range 15.0–2.0 Å) 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 2mFo – DFo and mFo – DFo 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 AR Refinement. The R-factor of the final model including 158 water molecules was 16.2% (R_free = 20.2%, resolution range 15.0–2.0 Å). 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


POVRay program, version 3.02 (1997) URL: http://www.povray.org


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