Acta Cryst. (1998). D**54**, 757–766

Structure of the C-Type Lectin Carbohydrate Recognition Domain of Human Tetranectin

JETTE SANDHOLM KASTRUP,[a]* BETTINA BRYDE NIELSEN,[a] HANNE RASMUSSEN,[a] THOR LAS HOLTET,[b] JONAS HEILSKOV GRAVERSEN,[b] MICHAEL ETZERODT,[b] HANS CHRISTIAN THØGERSEN[b] AND INGRID KJØLLER LARSEN[a]

[a]*Department of Medicinal Chemistry, Royal Danish School of Pharmacy, DK-2100 Copenhagen, Denmark, and* [b]*Laboratory of Gene Expression, Department of Molecular and Structural Biology, University of Aarhus, DK-8000 Aarhus C, Denmark. E-mail: kastrup@medchem.dfh.dk*

(*Received 1 September 1997; accepted 13 November 1997*)

Abstract

Tetranectin (TN) is a C-type lectin involved in fibrinolysis, being the only endogenous ligand known to bind specifically to the kringle 4 domain of plasminogen. TN was originally isolated from plasma, but shows a wide tissue distribution. Furthermore, TN has been found in the extracellular matrix of certain human carcinomas, whereas none or little is present in the corresponding normal tissue. The crystal structure of full-length trimeric TN (2.8 Å resolution) has recently been published [Nielsen *et al.* (1997). *FEBS Lett.* **412**, 388–396]. The crystal structure of the carbohydrate recognition domain (CRD) of human TN (TN3) has been determined separately at 2.0 Å resolution in order to obtain detailed information on the two calcium binding sites. This information is essential for the elucidation of the specificity of TN towards oligosaccharides. TN3 crystallizes as a dimer, whereas it appears as a monomer in solution. The overall fold of TN3 is similar to other known CRDs. Each monomer is built of two distinct regions, one region consisting of six β-strands and two α-helices, and the other region is composed of four loops harboring two calcium ions. The calcium ion at site 1 forms an eightfold coordinated complex and has Asp116, Glu120, Gly147, Glu150, Asn151, and one water molecule as ligands. The calcium ion at site 2, which is believed to be involved in recognition and binding of oligosaccharides, is sevenfold coordinated with ligands Gln143, Asp145, Glu150, Asp165, and two water molecules. One sulfate ion has been located at the surface of TN3, forming contacts to Glu120, Lys148, Asn106 of a symmetry-related molecule, and to an ethanol molecule.

1. Introduction

The animal C-type (calcium-dependent) lectins are, both in terms of function and protein architecture, a highly diverse family of carbohydrate-binding proteins. The C-type lectin superfamily has been divided into several different groups based on sequence and overall molecular architecture (Drickamer & Taylor, 1993; Day, 1994). C-type lectins are characterized by the carbohydrate recognition domain (CRD), exhibiting calcium-dependent binding of specific sugars. C-type lectins are mostly di- or trimeric molecules, in which each monomer contains one copy of the CRD (Drickamer & Taylor, 1993; Drickamer, 1995). The CRDs comprise approximately 120 amino-acid residues including either two (short form) or three (long form) pairs of cysteine residues, forming intrachain disulfide bridges, and share about 25 highly conserved amino-acid residues in terms of identity and relative position in the sequence (Day, 1994).

Human TN is a trimeric C-type lectin (Holtet *et al.*, 1997) primarily found in human plasma (Clemmensen *et al.*, 1986) but also present in various tissues and in the extracellular matrix of certain human carcinomas (Christensen & Clemmensen, 1989, 1991; Høgdall *et al.*, 1993; Wewer & Albrechtsen, 1992). In addition to the established binding of TN to the fourth kringle domain of plasminogen (Clemmensen *et al.*, 1986), the protein interacts in a calcium-dependent manner with complex sulfated polysaccharides (Clemmensen, 1989) and fibrin (Kluft, Los *et al.*, 1989), as well as with lipoprotein (a) (Kluft, Jie *et al.*, 1989). The biological functions of TN have not yet been clarified, but one function might be the targeting of plasminogen to fibrin or to extracellular matrix components. The fact that TN is part of the fibrinolytic system of proteins and, in addition, is abundant in plasma and in various tissues, reinforces the expectation that TN serves important functional roles in fundamental biological processes like tissue degradation, cell migration, and the spread of cancer by metastasis and invasion. TN may also play a potential role in mineralization during osteogenesis (Wewer *et al.*, 1994; Iba *et al.*, 1995).

In addition to human TN, the amino-acid sequences have been determined of TN from mouse (Ibaraki *et al.*, 1995; Sørensen *et al.*, 1995) and of the two tetranectin-like cartilage proteins, one bovine (Neame & Boynton, 1996) and one from reef shark (Neame *et al.*, 1992). The levels of sequence identity to human TN of these proteins are 81, 49 and 45%, respectively. The human TN polypeptide chain of 181 amino-acid residues with

ISSN 0907-4449

three intrachain disulfide bridges (Fuhlendorff *et al.*, 1987) is encoded in three exons (Berglund & Petersen, 1992). Exon 3 of the human TN gene encodes a single long-form CRD of 132 amino acids (residues 50–181).

Three-dimensional structures of C-type lectins containing short form CRDs have been determined for human E-selectin (Graves *et al.*, 1994), for human MBP (hMBP) (Sheriff *et al.*, 1994) and, in particular, for the rat mannose-binding proteins (rMBPs) from serum and liver, both with and without bound saccharides (Weis *et al.*, 1991, 1992; Ng *et al.*, 1996). In addition to full-length TN (Nielsen *et al.*, 1997), the structure of two other long-form CRDs have been reported recently: human lithostathine, which is a C-type lectin homologue without calcium binding sites (Bertrand *et al.*, 1996) and coagulation factors IX/X-binding protein, a lectin homologue containing two CRDs linked by a disulfide bond (Mizuno *et al.*, 1997).

Structural and biochemical analysis of wild-type and mutated CRDs have established that they share a common basic fold, including a conserved calcium ion binding site, although belonging to distinct groups of C-type lectins (Drickamer & Taylor, 1993; Day, 1994). The conserved calcium binding site appears to be a major determinant for specificity as well as for affinity in carbohydrate binding (Iobst & Drickamer, 1994; Iobst *et al.*, 1994; Kogan *et al.*, 1995; Kolatkar & Weis, 1996).

The X-ray structure of full-length TN was determined to 2.8 Å resolution (Nielsen *et al.*, 1997). We present here the X-ray structure of the CRD of recombinant human TN with an extension of five amino acids (TN3, residues 45–181) at 2.0 Å resolution, and give a detailed description of the structure of the CRD including the two calcium binding sites.

2. Experimental

2.1. *Data collection*

TN3 crystallizes in space group $P4_22_12$ with one monomer per asymmetric unit as previously described (Kastrup *et al.*, 1997). A high-resolution data set has been collected to 2.0 Å (97% complete) at the EMBL-Hamburg X11 beamline (λ = 0.927 Å) equipped with a MAR detector. The data collection was performed at 277 K with a crystal-to-detector distance of 150.0 mm and with 70° rotation in steps of 1.5° oscillations. Autoindexing and data processing were performed with *DENZO* (Otwinowski, 1993) and the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994). $R_{\rm merge}$ is 6.2% for all data between 15.0 and 2.0 Å and 22.1% in the outermost shell (2.05–2.0 Å). The data set has an average $I/\sigma(I)$ of 11.4 and of 3.4 in the outer shell. The data in the resolution range 6.0–2.0 Å were scaled to a 25.0–2.7 Å data set collected on a Rigaku RU-200 rotating-anode generator with an R-AXIS II imaging-plate detector as previously described (R-AXIS data) (Kastrup *et al.*, 1997), using only the data in the resolution range 25.0–3.0 Å. The two data sets scaled with an $R_{\rm merge}$ of 5.3% for all data between 6.0 and 3.0 Å. The resulting data set is 98% complete in the resolution range 25.0–2.0 Å.

2.2. *Structure determination and refinement*

The structure of TN3 was solved by molecular replacement using the structure of the CRD of rMBP from serum as search model (Kastrup *et al.*, 1997). Three different electron-density maps (25.0–2.7 Å resolution, R-AXIS data) were calculated using phases from (1) a polyalanine structure of rMBP (Weis *et al.*, 1992), (2) a model of TN3, and (3) a polyalanine model of TN3. The TN3 model was built using the structure of the CRD of rMBP and the program *SYBYL* (Tripos Associates, Inc.). Insertions and deletions in loop areas were constructed using the loop-search command. This model of TN3 comprises residues 58–180. Based on these three electron-density maps an initial structure of TN3 was obtained.

The structure was subjected to a slow-cool refinement protocol in *X-PLOR* (Brünger *et al.*, 1987) using the R-AXIS diffraction data in the resolution range 10.0–2.7 Å, alternating with graphical sessions using the program *O* (Jones *et al.*, 1991). During these refinements the R value dropped from 49.9 to 34.7% [$R_{\rm free}(10\%)$ = 45.3%]. However, at this point refinement did not progress further, and several surface loops remained poorly defined. Only diffuse electron density was observed for 14 N-terminal and two C-terminal residues. Inclusion of the synchrotron data did not improve refinement.

At this stage, least-squares refinements were applied, using the *TNT* refinement program package (Tronrud *et al.*, 1987; Tronrud, 1992). The electron-density map was improved significantly and the N-terminal residues in the first β-strand ($\beta 0'$) were readily traced. As the refinements progressed the electron density of previously poorly defined surface loops was clearly improved, and it was possible to assign the remaining N- and C-terminal residues. The positions of two calcium ions were located in a difference Fourier map (4σ) as well as one sulfate ion and one ethanol molecule. The occupancy of the sulfate ion was restrained to 0.5. In addition, the occupancies of the atoms within the side chains of Thr48, Val49 and Met58, as well as of Thr137–Ala142, were restrained to 0.5, due to two distinct conformations. The final R value is 21.8% [$R_{\rm free}(5\%)$ = 26.7%].

Refinements and map calculations in *TNT* were initially performed using the R-AXIS data set in the resolution range 25.0–2.7 Å. The resolution range was then extended using the merged data set between 25.0 and 2.0 Å. Initially, the individual B values were restrained to 20.0 Å^2 and were refined only in the last

Table 1. *Refinement statistics for the final structure of TN3*

Resolution range (Å)	25.0–2.0
R value (%)	21.8
R free (5%)	26.7
Number of atoms	1123
Protein atoms	1055
Calcium ions	2
Sulfate ions	1
Water molecules	58
Ethanol molecules	1
R.m.s. deviations from ideal geometry	
Bond length (Å)	0.016
Bond angle (°)	2.6
Torsion angle (°)	17.7
Trigonal non-planarity (Å)	0.017
Planarity (Å)	0.016
Non-bonded contacts (Å)	0.026
Average *B* values (Å^2)	
All atoms	33
All protein atoms	33
Main-chain atoms	29
Side-chain atoms	39
Calcium ion 1	28
Calcium ion 2	42
Sulfate ion	39
Water molecules	45
Ethanol molecule	37
R.m.s. ΔB	2.0

cycles of refinement. The *PEAKMAX* subroutine from *CCP*4 as well as *PEKPIK* within *TNT* were used to define peaks in the difference maps (3σ cut-off level), for locating water molecules. A water molecule was accepted, if the identified peak correlated with a separate peak in the corresponding $2F_o - F_c$ electron-density map, and if one or more hydrogen bonds (3.3–2.3 Å) could be identified. One ethanol molecule was identified in the electron-density maps. Initially, a water molecule was inserted but additional density was observed after refinement, corresponding to an ethyl moiety. Ethanol was used as additive in the crystallization of TN3 (Kastrup *et al.*, 1997). The final refinement statistics and average *B* values are tabulated in Table 1.†

3. Results and discussion

3.1. *The quality of the structure*

The structure of TN3, which comprises the CRD of TN (residues 50–181) and additional five residues preceding the CRD, has been determined at 2.0 Å resolution and refined to an *R* value of 21.8% (Table 1). All amino-acid residues of TN3 (residues 45–181), as well as two calcium ions, one sulfate ion, one ethanol molecule, and 58 water molecules, have been located in the final electron-density map. The fit of the structure in its electron density has been evaluated by the real-space *R* values per residue (Fig. 1) and rotamer side-chain fits (Jones *et al.*, 1991). The average real-space *R* values are 9.1% for the main-chain atoms and 13.5% for side chains. The highest values are observed within two surface loops (residues 53–55 and 138–141) and in the C-terminal region. Only three residue side chains (Leu51, Glu137, Trp163) have r.m.s. deviations greater than 2.1 Å from the closest rotamer. However, the electron density for these residues is well defined. All residues are located within allowed regions in the Ramachandran plot (Laskowski *et al.*, 1993), and one *cis*-peptide bond (Pro144) has been identified. Two different side-chain conformations have been identified for Thr48, Val49 and Met58. In addition, the amide moieties linking Thr138 and Glu139, and Thr141 and Ala142, respectively, are observed in two distinct conformations. The distribution of the *B* factors are illustrated in Fig. 1. The parts of the structure with the highest main-chain *B* values correlate with regions of highest real-space *R* values. The side chains of ten residues (Lys55, Gln66, Arg101, Glu107, Arg130, Lys134, Glu139, Arg169, Ile180, Val181) are less well defined.

3.2. *The overall structure*

TN3 crystallizes as a dimer whereas it had been found to be a monomer in solution (Holtet *et al.*, 1997). The structure of TN3 (Fig. 2) is very similar to that of the CRD of full-length TN [r.m.s. deviation of 0.50 Å for 124 Cα atoms (residues 58–181)], and two calcium ions have been located at the same positions and with identical protein ligands (Nielsen *et al.*, 1997). Each monomer is built of two regions. One region is composed of two α-

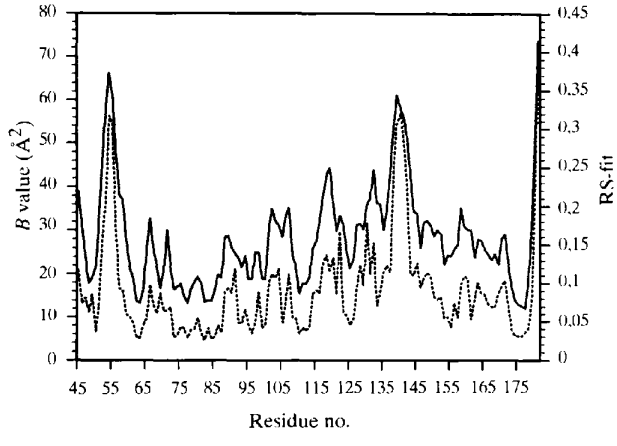

Fig. 1. The real-space *R* value (in %) for the main-chain atoms (dashed line) and the average *B* values (in Å^2) of the main-chain atoms (continuous line) as a function of residue number.

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1TN3).

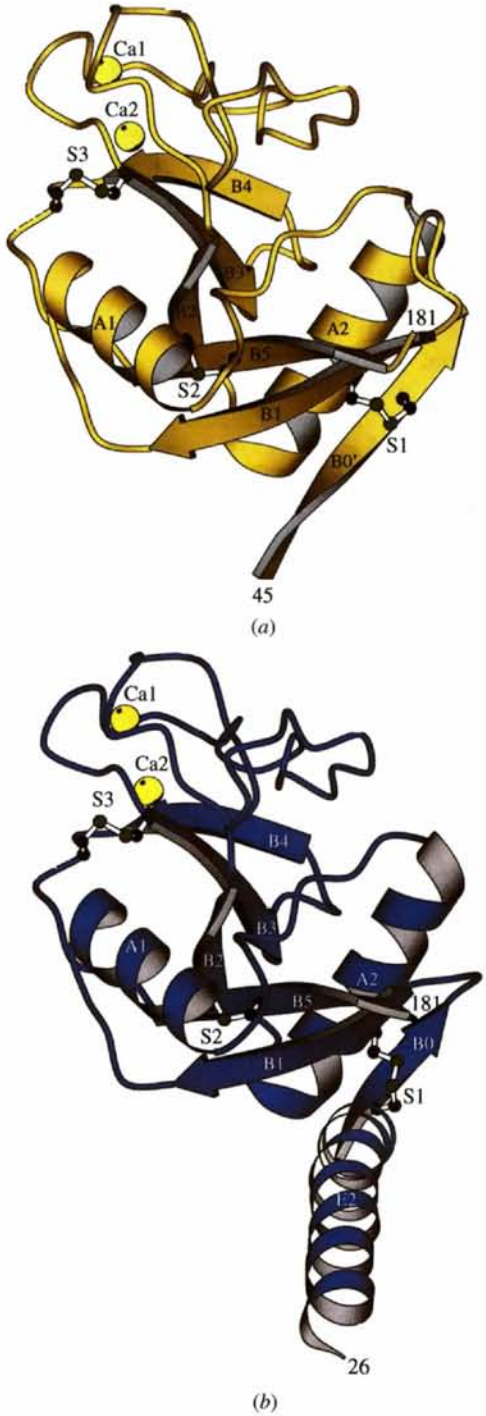

Fig. 2. (*a*) Structure of the TN3 monomer. The protein consists of two α-helices (α1 70–80 and α2 90–104), six β-strands (β0′ 45–52, β1 58–65, β2 108–112, β3 152–157, β4 162–166, and β5 173–180), and a characteristic loop region (L1 116–123, L2 125–129, L3 135–143, and L4 145–151). The two calcium ions are shown as yellow spheres and the three disulfide bridges are labeled S1–S3. (*b*) The structure of the full-length TN monomer. Major structural differences are observed only in the N-terminal parts of TN3 and the CRD of full-length TN. The figure was displayed using *MOLSCRIPT* (Kraulis, 1991).

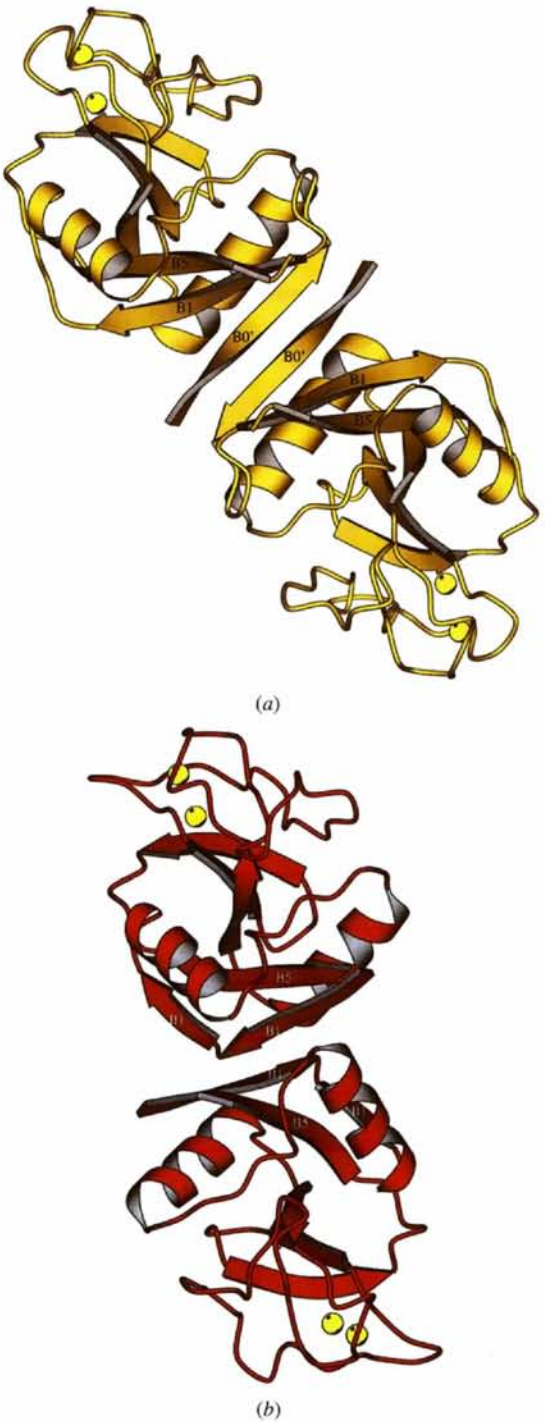

Fig. 3. (*a*) The TN3 dimer formed of two crystallographically related monomers, viewed along the twofold axis of the dimer. A β-sheet of six strands is formed upon dimerization. (*b*) The CRD of rMBP from serum, which also forms a dimeric structure by pairing of two β-strands. In this case, a four-stranded β-sheet is built, as the β0′ strand is not present in this short form CRD. The figure was displayed using *MOLSCRIPT* (Kraulis, 1991).

Table 2. *Ligands of the two calcium binding sites of TN3 and the corresponding calcium ion ligands from the CRDs of hMBP and rMBP from serum*

TN3 Residue	Distance (Å) to calcium ion	hMBP Residue	rMBP Residue
Calcium site 1			
Asp116 OD1	2.7	Asp168 OD1	Asp161 OD1
Asp116 OD2	2.7	Asp168 OD2	Asp161 OD2
Glu120 OE1	2.6	Glu172 OE1	Glu165 OE1
Glu120 OE2	2.6	Glu172 OE2	Glu165 OE2
Gly147 O	2.4	Asn195 OD1	Asp188 OD1
Glu150 O	2.5	Glu200 O	Glu193 O
Asn151 OD1	2.0	Asp201 OD1	Asp194 OD1
Water	2.4	Water	Water
Calcium site 2			
Gln143 OE1	2.5	Glu192 OE1	Glu185 OE1
Asp145 OD1	3.1	Asn194 OD1	Asn187 OD1
Glu150 OE1	2.4	Glu200 OE1	Glu193 OE1
Asp165 O	2.8	Asp213 O	Asp206 O
Asp165 OD1	2.4	Asp213 OD1	Asp206 OD1
Water	2.5		
Water	3.2		
		Asn212 OD1†	Asn205 OD1†
		Two sulfate O atoms†	Two mannose O atoms†

† Additional ligands at calcium site 2 observed in the structure of the MBPs.

helices (α1 and α2) and six β-strands (β0′, β1, β2, β3, β4, and β5). The six β-strands form two antiparallel three-stranded β-sheets, comprising β0′, β1, β5, and β2, β3, β4, respectively. The other region consists of four loops (L1–L4). The two regions are separated by the β2-strand and each contain a distinct hydrophobic core.

However, major structural differences are observed in the N-terminal part of TN3 as compared to the CRD of full-length TN (Fig. 2). In TN3, one long β-strand is formed (β0′, residues 45–52), which is followed by a loop and the β1-strand (residue 58–65). In full-length TN, a long α-helix (E2) proceeds to residue 52, and is then followed by a short β0-strand (residues 53–56), which is connected to the β1-strand (residues 59–65) by a β-turn. As a consequence, the N-terminal disulfide bridge S1 (Cys50–Cys60) is linking two β-strands in TN3, whereas it connects an α-helix and a β-strand in TN. These differences are obviously due to the different possibilities for molecular interaction in the two structures.

3.3. *The dimerization site*

Upon dimerization of TN3 in the crystal, a large antiparallel six-stranded β-sheet is formed as a consequence of the interactions between the long N-terminal β-strands (β0′) of two crystallographically related monomers (see Fig. 3*a*). The N-terminal region of TN3 includes five amino acids preceding the CRD of TN. These amino acids add to the number of apolar residues of the β0′-strand (Ala45, Leu46, Val49, Cys50 and Leu51). Upon dimerization, the hydrophobic residues of each monomer are buried within the dimeric protein. The main-chain N and O atoms of two residues from each strand (Thr48 and Cys50) are involved in the formation of four hydrogen bonds between the monomers. The hydrophobic residues included in the dimerization area of TN3 are, in full-length TN, involved in the formation of a triple α-helical coiled coil as well as in E2 helix–CRD interactions (Nielsen *et al.*, 1997). The structural differences show how elegantly the demand for burying hydrophobic residues may be obeyed.

Dimerization of the CRD of rMBP from serum is accomplished in a similar way as in TN3 (Weis *et al.*, 1992). However, the dimerization area is formed by the pairing of two β1-strands, leading to the formation of a four-stranded β-sheet (Fig. 3*b*). In TN, as well as in MBPs, the region preceding the CRD is a long α-helix involved in a trimeric coiled coil. The fact that dimeric structures are formed from components of trimeric molecules in each case is quite remarkable.

3.4. *The calcium binding sites*

The CRD of TN binds two calcium ions (calcium sites 1 and 2) by eightfold and sevenfold coordinated complexation, respectively, involving both protein ligands and water molecules. The protein ligands are located in the loop region (L1, L3 and L4), except for one ligand located in β4. The distance between the two calcium ions (at sites 1 and 2) is 8.6 Å, which is similar to the distances found in the MBPs (Sheriff *et al.*, 1994; Weis *et al.*, 1991, 1992; Ng *et al.*, 1996).

The eightfold coordinated calcium ion at site 1 coordinates to five amino-acid residues (Asp116, Glu120, Asn151, Gly147 and Glu150) and to one water molecule (Table 2, Fig. 4*a*). The protein-derived ligands are

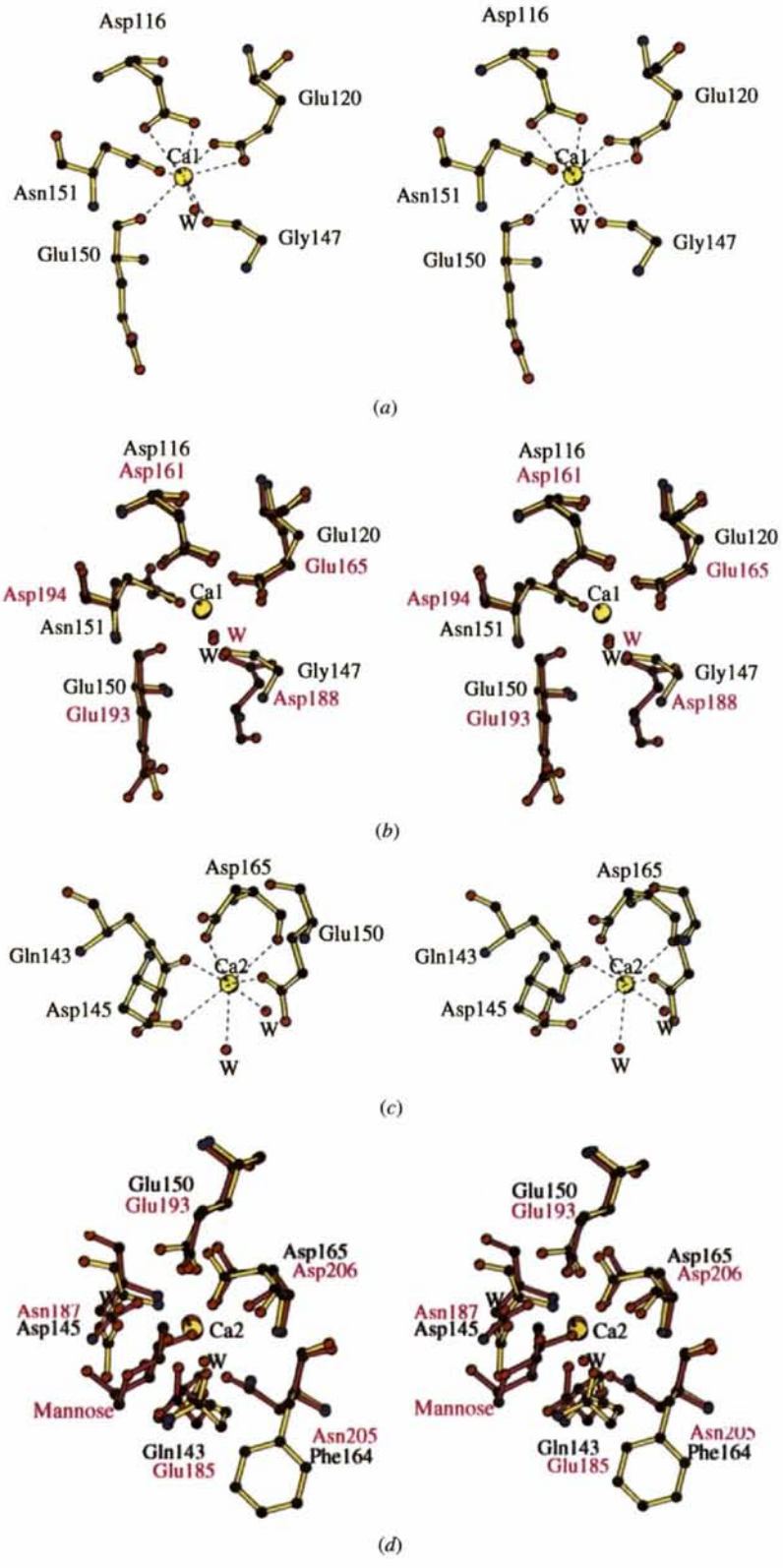

Fig. 4. The two calcium binding sites of TN3 and comparison with rMBP. (*a*) Site 1 of TN3, showing in stereo the eightfold coordination of calcium by five protein ligands and one water molecule. (*b*) Comparison of site 1 with site 1 in rMBP from serum. (*c*) Site 2 of TN3, showing the sevenfold coordination of calcium by four protein and two water ligands. (*d*) Comparison of site 2 with site 2 in rMBP from serum. TN is colored according to atom type and rMBP is colored in magenta. The figure was generated using *MOLSCRIPT* (Kraulis, 1991).

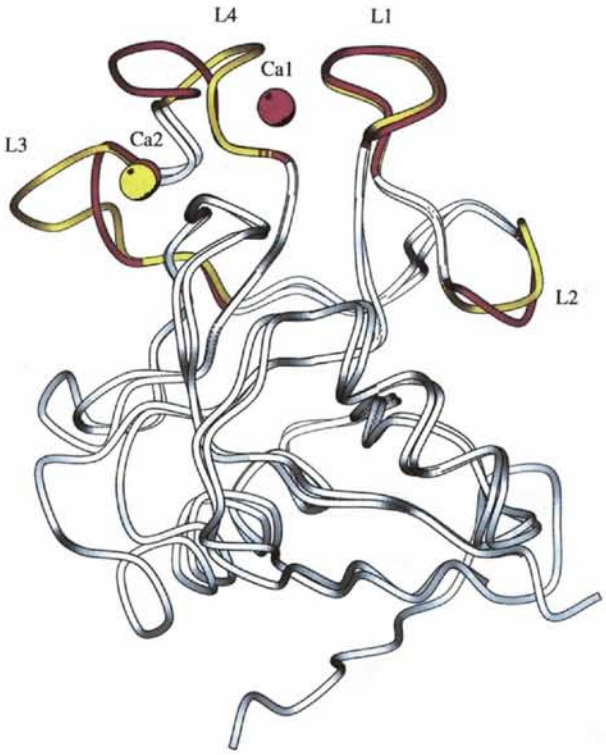

Fig. 5. A tube representation of TN3 and the CRD of rMBP from serum (r.m.s. of 0.96 Å for 108 Cα atoms). The four loops (L1–L4) of TN3 are shown in yellow and of rMBP in magenta. The remaining residues of the two structures are uncolored. The calcium ions are shown as yellow and magenta spheres, respectively. The position of the calcium ions at site 1 is almost identical in the two structures. The figure was generated using *MOLSCRIPT* (Kraulis, 1991).

located in L1 and L4. The complexation of calcium at binding site 1 in TN3 resembles that of the MBPs with a similar coordination geometry of the calcium ion (Table 2, Fig. 4*b*). However, Gly147 in TN3 corresponds to Asp188 in rMBP and Asn195 in hMBP, the main-chain O atom of Gly147 thereby replaces the side-chain O atoms in the MBPs as a coordinating ligand. The water molecule that ligates to the calcium ion at site 1 is structurally conserved among the proteins (Table 2). Apart from this particular water molecule only very few water molecules are conserved among the structures.

The amino-acid residues involved in ligand binding at calcium site 2 (Table 2, Fig. 4*c*) are located in L3 (Gln143), in L4 (Asp145 and Glu150), and in β4 (Asp165). Glu150 is a ligand for both calcium ions, through the main-chain carbonyl O atom at calcium site 1 and a side-chain O atom at calcium site 2. The protein ligand, Asp165, coordinates with both the side-chain and the main-chain O atoms to the calcium ion at site 2. In addition, two water molecules are ligated to the calcium ion. Contrary to calcium site 1, marked differences in the coordination geometry at calcium site 2 are observed among TN3 and the MBPs (Table 2, Fig. 4*d*). The calcium ion is sevenfold coordinated in TN3, whereas it is eightfold coordinated in the MBPs. In TN3 the coordination complex is a distorted pentagonal bipyramid as in the MBPs (Weis *et al.*, 1992). However, the ligands of the pentagon and the ligands occupying the apical positions are different in TN3 and in the MBPs.

Despite the similarities among the calcium ligands of TN3 and the MBPs, large differences occur in the loops (L3 and L4) harboring the calcium ions (Fig. 5). TN3 has a deletion of one amino-acid residue in L4. This loop

```
          50                  60                  70                  80
hTN       C L K G T K V H M K C F L A F T Q T K T F H E A S E D C I S R G G T L S T P Q
mTN       C L K G T K V N L K C L L A F T Q P K T F H E A S E D C I S Q G G T L G T P Q
bCP       A L E G T K F G K K C Y L A A E G L K H F H E A N E D C I S K G G T L V V P R
rsCP      C L K G T K I H K K C Y L A S R G S K S Y H A A N E D C I A Q G G T L S I P
Identity    *   * * *     * *   * *       *     *   *   * * * *     * * * *     *

          90                  100                 110                 120
hTN       T G S E N D A L Y E Y L R Q S V G N E A E I W L G L N D M A A E G T W V D M T G
mTN       S E L E N E A L F E Y A R H S V G N D A N I W L G L N D M A A E G A W V D M T G
bCP       S A D E I N A L R D Y G K R S L P G V N D F W L G I N D M V A E G K F V D I N G
rsCP      S S D E G N S L R S Y A K K S L V G A R D F W I G V N D M T T E G K F V D V N G
Identity        *       *     *       *                   *   *   * * *     * *     * *     *
Ca ligands                                                          1       1

          130                 140                 150                 160
hTN       A R I A Y K N W E T E I T A Q P D G G K T E N C A V L S G A A N G K W F D K R C
mTN       G L L A Y K N W E T E I T T Q P D G G K A E N C A A L S G A A N G K W F D K R C
bCP       L A I S F L N W D Q - - - A Q P N G G K R E N C A L F S Q S A Q G K W S D E A C
rsCP      L P I T Y F N W D R - - - S K P V G G T R E N C V A A S T S G Q G K W S D D V C
Identity              * *             *   * *     * * *       *         * * *   *     *
Ca ligands                        2   2   1     12 1                              2

          170                 180
hTN       R D Q L P Y I C Q F G I V
mTN       R D Q L P Y I C Q F A I V
bCP       H S S K R Y I C E F T I P
rsCP      R S E K R Y I C E Y L I P
Identity            * * *       *
```

Fig. 6. Alignment of amino-acid sequences of the CRDs of human TN (hTN), mouse TN (mTN), bovine cartilage protein (bCP), and reef shark cartilage protein (rsCP). The numbering shown is for TN. Strictly conserved amino acids among the four sequences are marked by an asterisk. The protein ligands for the calcium sites 1 and 2 are indicated by the numbers 1 and 2, respectively. The alignment was performed using the program *Cameleon* (Oxford Molecular Ltd.).

points towards L1, whereas it points towards L3 in the MBPs. In addition, an insertion (Glu137–Thr141) in L3 increases the loop size in TN3. The loops L1 and L2 occur in comparable conformations in TN3 and in the MBPs. However, neither the insertion in L3 nor the deletion in L4 of TN3 perturb the structural positions of the amino-acid residues involved in binding of calcium ions as compared with the MBPs.

3.5. *Comparison of the calcium binding sites of tetranectin-like proteins*

The alignment of amino-acid sequences of the CRDs of human TN, mouse TN, bovine cartilage protein, and reef shark cartilage protein is shown in Fig. 6. 50 amino acids are strictly conserved among the four proteins, including all protein ligands at calcium site 1, which appears to be fully preserved.

All four protein ligands (Gln143, Asp145, Glu150 and Asp165) at calcium site 2 are conserved between human and mouse TN, suggesting this site to be preserved in the closely related mouse TN. However, differences in protein ligands at site 2 of the more distinct cartilage proteins, especially from reef shark, are observed. In bovine cartilage protein all four residues but one are strictly conserved, *i.e.* Asp145 of human TN is an Asn. As Asn is still capable of coordinating a calcium ion, calcium site 2 is probably also preserved in bovine cartilage protein. In the cartilage protein from reef shark only two residues (Glu150 and Asp165 in human TN)

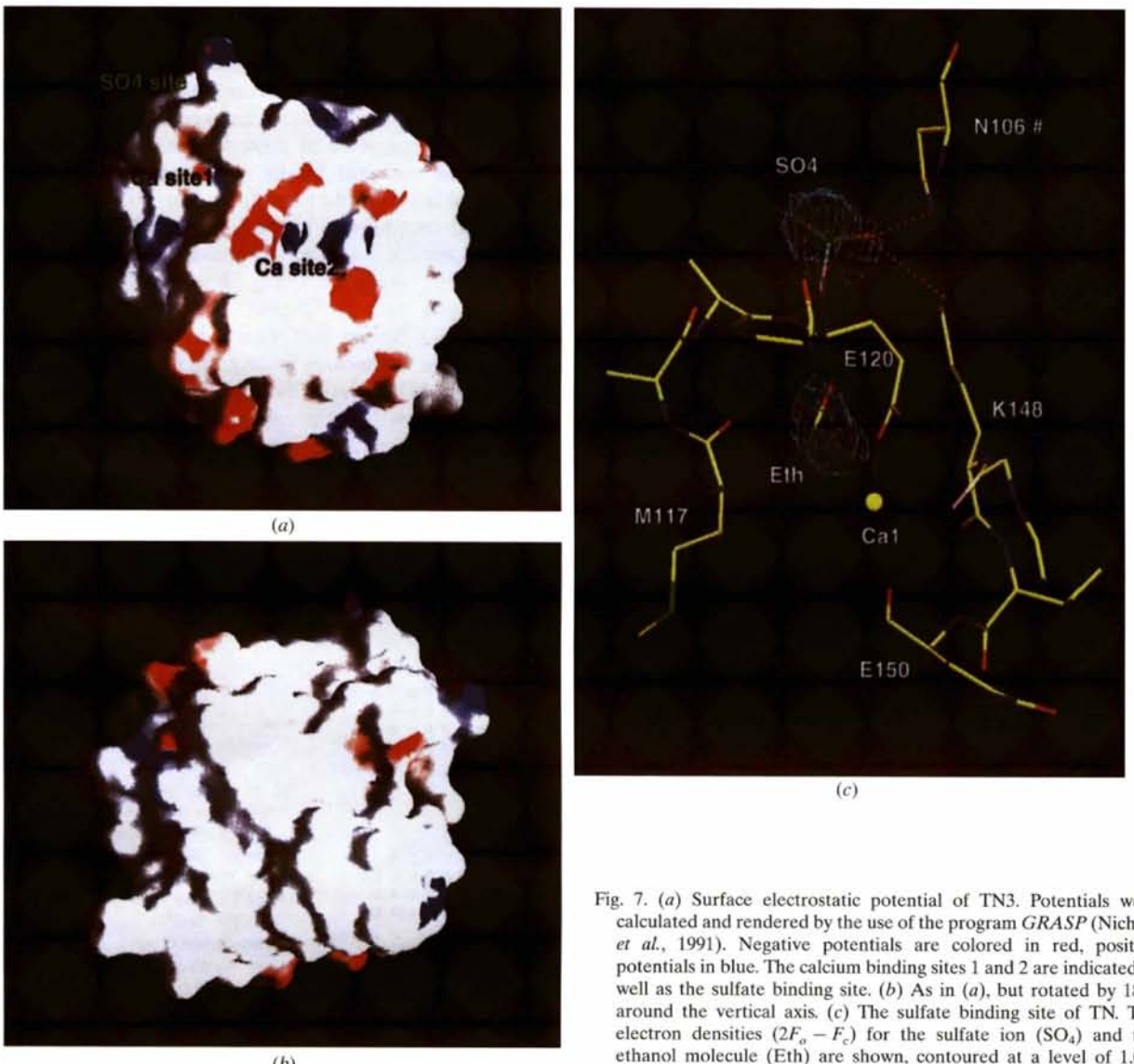

Fig. 7. (*a*) Surface electrostatic potential of TN3. Potentials were calculated and rendered by the use of the program *GRASP* (Nichols *et al.*, 1991). Negative potentials are colored in red, positive potentials in blue. The calcium binding sites 1 and 2 are indicated as well as the sulfate binding site. (*b*) As in (*a*), but rotated by 180° around the vertical axis. (*c*) The sulfate binding site of TN. The electron densities ($2F_o - F_c$) for the sulfate ion (SO_4) and the ethanol molecule (Eth) are shown, contoured at a level of 1.0σ. Hydrogen bonds are depicted as dotted lines. The figure was generated using the program *O* (Jones *et al.*, 1991).

are conserved, whereas Gln143 and Asp145 in human TN are a Lys and a Val, respectively. From these observations it might be concluded that calcium site 1, but not site 2, is preserved in the cartilage protein from reef shark.

3.6. *The putative carbohydrate binding site*

C-type lectins participate in a variety of cell-surface recognition events mediated by protein–calcium–carbohydrate interactions (Drickamer & Taylor, 1993; Day, 1994; Weis & Drickamer, 1996). So far, our understanding of the basis of carbohydrate recognition by TN is limited. It has been reported that TN interacts with sulfated oligosaccharides and the sulfated dye trypan blue (Clemmensen, 1989). In addition, TN binds to immobilized fucoidan (Graversen, J. H., unpublished work).

No predominant positively charged area, well suited for the binding of negatively charged molecules, is found on the surface of TN3 (see Figs. 7*a* and 7*b*). However, one sulfate ion has been located at the surface of TN3, forming contacts to the side-chain N^{ζ} atom of Lys148, to the main-chain N atom of Glu120, and to the O atom of the ethanol molecule, as well as to the main-chain N atom and the side-chain $N^{\delta 2}$ atom of Asn106 of a symmetry-related molecule (See Figs. 7*a* and 7*c*). This site, which is positioned about 9 Å from the calcium ion at site 1, might represent a binding site for sulfated saccharides. However, it appears to be a low-affinity sulfate binding site as the occupancy of the sulfate ion is 0.5 (average B value of 39 $Å^2$), despite a high concentration of ammonium sulfate in crystallization of TN3 (Kastrup *et al.*, 1997). In the structure of hMBP, a sulfate ion is ligated to the calcium ion at site 2 (Sheriff *et al.*, 1994). No sulfate ion has been observed at calcium site 2 in the present structure.

The structures of rMBPs or mutants of rMBP from serum complexed with oligosaccharides show that calcium site 2 is the carbohydrate binding site in the MBPs (Weis *et al.*, 1992; Ng *et al.*, 1996; Kolatkar & Weis, 1996; Ng & Weis, 1997). The C-type lectins might be divided into mannose-type and galactose-type lectins depending on sugar specificity. The sugar-binding of both types are similar, involving the 3- and 4-hydroxyl groups, which form hydrogen bonds to calcium ligands and coordinate to the calcium ion (Weis *et al.*, 1996). The conservation of this site among the CRDs suggests that calcium site 2 in TN is also involved in binding of carbohydrates.

Two of the calcium ligands (Glu185 and Asn187 in rMBP) are of importance for sugar specificity (Drickamer, 1992). Studies of rMBP mutants have shown that mutation of Glu185 and Asn187 into Gln185 and Asp187, respectively, along with a mutation of His189 to Trp and insertion of a glycine-rich loop, which is characteristic for the galactose binding asialoglycoproteins, converts the rMBP into a galactose-binding protein (Iobst *et al.*, 1994; Kolatkar & Weis, 1996). This loop is not present in TN3, although the ligands Gln143 and Asp145, which suggest that TN has specificity for galactose, are present. Structural conservation of the *cis*-Pro (Pro144 in TN3) among all known structures of animal C-type lectins indicate that the same spatial arrangement of these two ligands (Gln143 and Asp145 in TN3) is required for, and compatible with, binding of different sugars.

In rMBP from liver two water molecules are displaced upon binding of monosaccharides (Ng *et al.*, 1996). If the two water molecules at calcium site 2 in TN3 indicate the positions of two vicinal hydroxyl groups of a saccharide, the orientation of the saccharide will be different to that observed in the MBPs (Figs. 4*c*–4*d*). Phe164, which is at a position corresponding to that of the calcium ligand Asn205 in rMBP from serum, might be involved in stacking interactions with a saccharide ring and thereby be a determinant of the orientation of the bound saccharide. In the MBPs, two hydrophobic residues (Val194 and Val212 in rMBP from liver) have been suggested to be determinants of the orientation of mannose (Ng *et al.*, 1996). In TN3, the residue corresponding to Val212 is Lys166, which might be involved in similar hydrophobic interactions and/or in the formation of a salt bridge with a sulfate moiety of a sulfated oligosaccharide. Recent studies by Ng & Weis (1997) on complexes of a selectin-like mutant of rMBP from serum (residues 211–213 mutated to Lys-Lys-Lys) with oligosaccharides show the formation of a salt bridge between a sulfate moiety of a saccharide and Lys211. The side chain of Lys166 in TN3 is spatially close to that of Lys211 in the rMBP mutant. The hydrophobic interaction of mannose with Val194 in rMBP is not possible in TN3 due to structural differences in this region.

In conclusion, the structure determination of the CRD of TN, in addition to those of other C-type lectins, has revealed that, although the overall fold of the proteins is the same, important structural differences are observed at the two calcium sites and within the loops surrounding these sites. These differences are probably essential for the nature and specificity of carbohydrate binding in the individual proteins. In the longer term, the structure knowledge will provide a firm basis for the further exploration of the functional role of TN.

Dr Inge Thøger Christensen, The Royal Danish School of Pharmacy, is gratefully acknowledged for her help in construction of the model of TN3. The work was supported by grants from The Danish Biotechnology Programme, The Lundbeck Foundation, The Carlsberg Foundation, The Dansync Center for Synchrotron Radiation, and The EU Network Programme on 'Protein Crystallography', Contract Number CHRX-CT93-0143. We thank the European Union for support of the work at EMBL Hamburg through the HCMP

Access to Large Installations Project, Contract Number CHGE-CT93-0040.

References

Berglund, L. & Petersen, T. E. (1992). *FEBS Lett.* **309**, 15–19.

Bertrand, J. A., Pignol, D., Bernard, J.-P., Verdier, J.-M., Dagorn, J.-C. & Fontecilla-Camps, J. C. (1996). *EMBO J.* **15**, 2678–2684.

Brünger, T. A., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Christensen, L. & Clemmensen, I. (1989). *Histochemistry*, **92**, 29–35.

Christensen, L. & Clemmensen, I. (1991). *Histochemistry*, **95**, 427–433.

Clemmensen, I. (1989). *Scand. J. Clin. Lab. Invest.* **49**, 719–725.

Clemmensen, I., Petersen, L. C. & Kluft, C. (1986). *Eur. J. Biochem.* **156**, 327–333.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Day, A. J. (1994). *Biochem. Soc. Trans.* **22**, 83–87.

Drickamer, K. (1992). *Nature (London)*, **360**, 183–186.

Drickamer, K. (1995). *Nature Struct. Biol.* **2**, 437–439.

Drickamer, K. & Taylor, M. E. (1993). *Annu. Rev. Cell Biol.* **9**, 237–264.

Fuhlendorff, J., Clemmensen, I. & Magnusson, S. (1987). *Biochemistry*, **26**, 6757–6764.

Graves, B. J., Crowther, R. L., Chandran, C., Rumberger, J. M., Li, S., Huang, K.-S., Presky, D. H., Familletti, P. C., Wolitzky, B. A. & Burns, D. K. (1994). *Nature (London)*, **367**, 532–538.

Høgdall, C. K., Christensen, L. & Clemmensen, I. (1993). *Cancer*, **72**, 2415–2422.

Holtet, T. L., Graversen, J. H., Clemmensen, I., Thøgersen, H. C. & Etzerodt, M. (1997). *Protein Sci.* **6**, 1511–1515.

Iba, K., Sawada, N., Chiba, H., Wewer, U. M., Ishii, S. & Mori, M. (1995). *FEBS Lett.* **373**, 1–4.

Ibaraki, K., Kozak, C. A., Wewer, U. M., Albrechtsen, R. & Young, M. F. (1995). *Mamm. Genome*, **6**, 693–696.

Iobst, S. T. & Drickamer, K. (1994). *J. Biol. Chem.* **269**, 15512–15519.

Iobst, S. T., Wormald, M. R., Weis, W. I., Dwek, R. A. & Drickamer, K. (1994). *J. Biol. Chem.* **269**, 15505–15511.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kastrup, J. S., Nielsen, B. B., Rasmussen, H., Larsen, I. K., Holtet, T. L., Graversen, J. H., Elzerodt, M. & Thøgersen, H. C. (1997). *Acta Cryst.* D**53**, 108–111.

Kluft, C., Jie, A. F. H., Los, P., de Witt, E. & Havekes, L. (1989). *Biochem. Biophys. Res. Commun.* **161**, 427–433.

Kluft, C., Los, P. & Clemmensen, I. (1989). *Thomb. Res.* **55**, 233–238.

Kogan, T. P., Revelle, B. M., Tapp, S., Scott, D. & Beck, P. J. (1995). *J. Biol. Chem.* **270**, 14047–14055.

Kolatkar, A. R. & Weis, W. I. (1996). *J. Biol. Chem.* **271**, 6679–6685.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H., Atoda, H. & Morita, T. (1997), *Nature Struct. Biol.* **4**, 438–441.

Neame, P. J. & Boynton, R. E. (1996). *Techniques in Protein Chemistry VII*, pp. 401–407. New York: Academic Press.

Neame, P. J., Young, C. N. & Treep, J. T. (1992). *Protein Sci.* **1**, 161–168.

Ng, K. K.-S., Drickamer, K. & Weis, W. I. (1996). *J. Biol. Chem.* **271**, 663–674.

Ng, K. K.-S. & Weis, W. I. (1997). *Biochemistry*, **36**, 979–988.

Nichols, A., Sharp, K. A. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.

Nielsen, B. B., Kastrup, J. S., Rasmussen, H., Holtet, T. L., Graversen, J. H., Etzerodt, M., Thøgersen, H. C. & Larsen, I. K. (1997). *FEBS Lett.* **412**, 388–396.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Sheriff, S., Chang, C. Y. & Ezekowitz, R. A. B. (1994). *Nature Struct. Biol.* **1**, 789–794.

Sørensen, C. B., Berglund, L. & Petersen, T. E. (1995). *Gene*, **152**, 243–245.

Tronrud, D. E. (1992). *Acta Cryst.* A**48**, 912–916.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

Weis, W. I. & Drickamer, K. (1996). *Annu. Rev. Biochem.* **65**, 441–473.

Weis, W. I., Drickamer, K. & Hendrickson, W. A. (1992). *Nature (London)*, **260**, 127–134.

Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. (1991). *Science*, **254**, 1608–1615.

Wewer, U. M. & Albrechtsen, R. (1992). *Lab. Invest.* **67**, 253–262.

Wewer, U. M., Ibaraki, K., Schjørring, P., Durkin, M. E., Young, M. F. & Albrechtsen, R. (1994). *J. Cell. Biol.* **127**, 1767–1775.