Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of the buffalo secretory signalling glycoprotein at 2.8 Å resolution

Abdul S. Ethayathulla,‡ Devendra B. Srivastava,‡ Janesh Kumar, Kolandaivelu Saravanan, Sameeta Bilgrami, Sujata Sharma, Punit Kaur, Alagiri Srinivasan and Tej P. Singh*

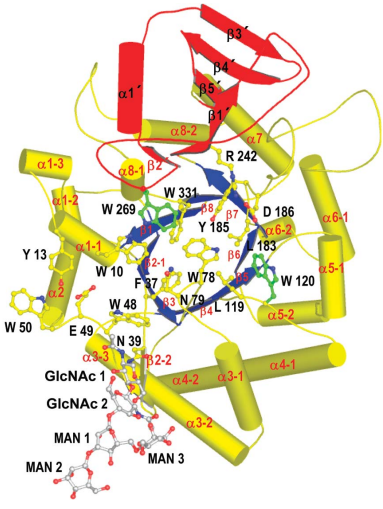
Department of Biophysics, All India Institute of Medical Sciences, New Delhi 110029, India

‡ These authors made equal contributions to this work.

Correspondence e-mail: tps@aiims.aiims.ac.in

Received 15 January 2007
Accepted 5 March 2007

PDB Reference: secretory signalling glycoprotein, 2o9o, r2o9osf.



The crystal structure of a 40 kDa signalling glycoprotein from buffalo (SPB-40) has been determined at 2.8 Å resolution. SPB-40 acts as a protective signalling factor by binding to viable cells during the early phase of involution, during which extensive tissue remodelling occurs. It was isolated from the dry secretions of Murrah buffalo. It was purified and crystallized using the hanging-drop vapour-diffusion method with 19% ethanol as the precipitant. The protein was also cloned and its complete nucleotide and amino-acid sequences were determined. When compared with the sequences of other members of the family, the sequence of SPB-40 revealed two very important mutations in the sugar-binding region, in which Tyr120 changed to Trp120 and Glu269 changed to Trp269. The structure showed a significant distortion in the shape of the sugar-binding groove. The water structure in the groove is also drastically altered. The folding of the protein chain in the flexible region comprising segments His188–His197, Phe202–Arg212 and Tyr244–Pro260 shows large variations when compared with other proteins of the family.

1. Introduction

The transition period between the lactating and nonlactating states of the mammary gland is a period of active involution during which the mammary gland undergoes extensive ultrastructural changes and the secretions contained in the gland undergo dramatic compositional changes (Hurley, 1989). During this period, a number of proteins are expressed. Among the proteins with high concentrations in the dry secretions of nonlactating cows, a new glycoprotein with a molecular weight of 40 kDa (SPX-40) was reported, the function of which is still unknown (Rejman & Hurley, 1988). The crystal structures of SPX-40 proteins from goat (Kumar *et al.*, 2007), cattle (Kumar *et al.*, 2006) and sheep (Srivastava *et al.*, 2006) have recently been reported. We have isolated the SPX-40 protein from buffalo dry secretions (SPB-40) and its complete amino-acid sequence has been determined. It shows an identity of 73% or above with other mammalian glycoproteins such as bovine chondrocyte chitinase-like protein (CLP-1; GenBank accession No. AF011373), human chondrocyte glycoprotein (YKL-40/HCGP-39; GenBank accession No. M80927; Johansen *et al.*, 1993; Hakala *et al.*, 1993), porcine heparin-binding glycoprotein (GP38k; GenBank accession No. U19900; Shackelton *et al.*, 1995) and rat cartilage glycoprotein (RATgp39; GenBank accession No. AF062038). All these proteins have identical chain lengths and similar glycosylation sites. They all contain five cysteine residues with two disulfide bridges, with Cys20 unpaired. SPB-40 also shows a striking similarity (68% sequence identity) to a 39 kDa breast regression protein (BRP39; Morrison & Leder, 1994). Because of their very similar amino-acid sequences and chemical properties and possible similar functions, these mammalian proteins form a subclass of mammalian glycoproteins that will be referred to in the following as group I proteins. Group I proteins also show significant sequence and structural similarities to chitinases (referred to here as group II proteins; Renkema *et al.*, 1995). Despite these similarities, these proteins differ in important structural and functional details, including the shape of the sugar-binding groove and affinity for sugars, differences in active-site residues *etc.* The chitinases have a

Table 1
Crystallographic data and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data-collection statistics	
PDB code	2o9o
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 63.1, b = 66.8, c = 108.5
No. of molecules in the ASU	1
V_M (Å^3 Da^{-1})	2.8
Solvent content (%)	57.0
Resolution range (Å)	20.0–2.80 (2.89–2.80)
No. of unique reflections	11349
Completeness (%)	97.1 (95.5)
R_{sym}†	12.0 (43.0)
Mean $I/\sigma(I)$	11.0 (2.7)
Refinement statistics	
R_{cryst}‡ (%)	18.6
R_{free}§ (%)	23.6
Protein atoms	2894
GlcNAc molecules/atoms	2/28
MAN molecules/atoms	3/33
Water molecules	117
R.m.s.d. in bond lengths (Å)	0.02
R.m.s.d. in bond angles (°)	1.8
R.m.s.d. in dihedral angles (°)	22.0
Average B factor from Wilson plot (Å^2)	43.0
Average B factor for all atoms (Å^2)	37.1
Residues in the most favourable regions (%)	87.4
Residues in the additionally allowed regions (%)	12.6

† $R_{sym} = \sum|I - \langle I\rangle|/\sum I$, where I is the observed intensity. ‡ $R_{cryst} = \sum||F_o(h)| - |F_c(h)||/\sum|F_o(h)|$, where $F_o(h)$ and $F_c(h)$ are the observed and calculated structure-factor amplitudes, respectively, for reflection h. § R_{free} was calculated against 5% of the complete data set excluded from refinement.

well defined carbohydrate-binding groove in which oligomers of *N*-acetylglucosamine (chitin polymers) bind preferentially and are hydrolyzed. The active site of chitinases involves three acidic amino acids, Asp, Glu and Asp. In group I proteins, Glu is replaced by Leu, resulting in a loss of chitin-hydrolyzing capability. Another class of closely related proteins are known as the chitinase-like proteins. These proteins are also catalytically inactive owing to the mutation of one of the catalytic residues, but have significant similarities to chitinases in the folding of their polypeptide chain. They lack glycosylation sites and show notable mutations in the carbohydrate-binding sequence. The prominent protein among them is a novel mammalian lectin, YM1 (Sun *et al.*, 2001). Although YM1 was initially reported to bind to carbohydrates (Sun *et al.*, 2001), it was subsequently described as having poor carbohydrate-binding properties (Tsai *et al.*, 2004). The proteins of the YM1 subclass will be referred to in the following as group III proteins. In order to understand the structural and functional properties of group I, II and III proteins, we report another structure of a group I protein from buffalo dry secretions (SPB-40).

2. Experimental procedures

2.1. Isolation and purification

Fresh mammary secretions were collected from Murrah buffalo maintained at the Indian Veterinary Research Institute, Izatnagar, India. Secretions were collected on days 4, 8, 12, 16 and 20 after the onset of the involution period (day 0 being the last milking day). The samples were diluted twice with 50 m*M* Tris–HCl pH 7.8. Cation-exchanger CM-Sephadex (7 g l^{-1}) equilibrated in 50 m*M* Tris–HCl pH 7.8 was added and stirred slowly for about 1 h using a mechanical stirrer. The washed protein-bound gel was packed into a column (25 × 2.5 cm) and washed with the same buffer containing 0.2 *M* NaCl to remove other impurities. The remaining proteins were eluted with the same buffer containing 0.5 *M* NaCl. The eluted protein solution was desalted and again passed through a CM-Sephadex C-50 column (10 × 2.5 cm) pre-equilibrated with 50 m*M* Tris–HCl pH 7.8 and eluted with a linear gradient of 0.05–0.5 *M* NaCl in the same buffer. The elution profile contained three peaks. The peak corresponding to 0.3 *M* NaCl was collected. The protein solution from this peak was concentrated using an Amicon ultrafiltration cell (Bedford, USA). The concentrated samples were passed through a Sephadex G-100 column (100 × 2 cm) using 50 m*M* Tris–HCl pH 7.8 containing 0.5 *M* NaCl. The elution profile showed the presence of two peaks. The second peak in this final chromatographic step corresponded to a molecular weight of 40 kDa as indicated by matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI–TOF; Kratos-Shimadzu, Kyoto, Japan) and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). The protein samples were blotted onto a polyvinylidene fluoride (PVDF) membrane and the sequence of the first 20 N-terminal amino-acid residues was determined using a PPSQ20 protein sequencer (Shimadzu, Kyoto, Japan), which confirmed the protein.

2.2. Complete amino-acid sequence determination

In order to determine the complete amino-acid sequence of SPB-40, mammary-gland tissue was obtained from a nonlactating buffalo. The total RNA was extracted using the phenol/chloroform method (Chomczynski & Sacchi, 1987). The poly(A$^+$) mRNAs were isolated from the total RNA using an oligo(dT)-cellulose column (Amersham Pharmacia Biotech, Piscataway, USA). The small syringe column packed with oligo(dT)-cellulose was washed with 10 ml high-salt buffer (Salt 1; 1 *M* NaCl, 1 m*M* Na_2 EDTA, 40 m*M* Tris–HCl pH 7.4). The total RNA was mixed with an equal volume of Salt 1 buffer, warmed to 338 K and cooled immediately by placing it on ice. The chilled RNA was passed through the column packed with oligo(dT)-cellulose. The column was washed with 3 ml low-salt buffer (Salt 2; 0.1 *M* NaCl, 1 m*M* Na_2 EDTA). The reverse-transcription reaction was carried out with Moloney murine leukaemia virus (MMLV) reverse transcriptase polymerase using oligo(dT) primers. A portion (2 µl) of the reverse transcriptase polymerase chain reaction (RT-PCR) was used for PCR amplification of the gene. The sequences 5′-CTATCCTGTCGAGGCCAAAGGA-3′ and 5′-AATTTATTGGA-CCTTCTGGCC-3′ were used as forward and reverse primers, respectively. Nucleotide sequencing was carried out on the cloned double-stranded DNA (pGEM-T) using an automatic sequencer (model ABI-377). The complete nucleotide and derived amino-acid sequences have been deposited in the GenBank with accession No. AY295929.

2.3. Protein crystallization

Lyophilized samples of SPB-40 were redissolved to a final concentration of 30 mg ml^{-1} in a buffer containing 25 m*M* Tris–HCl pH 7.8, 0.2 m*M* NaCl. Crystallization took place using the hanging-drop vapour-diffusion method in Linbro plates at 298 K. 10 µl droplets of protein solution were equilibrated against 1 ml reservoir solution consisting of the protein buffer with 19%(*v*/*v*) ethanol pH 7.8. Thin square-shaped colourless crystals with dimensions of up to 0.45 × 0.40 × 0.15 mm were obtained after a week.

2.4. X-ray intensity data collection and processing

The crystals of SPB-40 were stable in the X-ray beam. One crystal of dimensions 0.45 × 0.40 × 0.15 mm was used for data collection at 278 K. The crystal was mounted in a glass capillary of 1 mm diameter

and both ends were sealed with beeswax. The capillary was then fixed in the goniometer head. The intensities were measured using a 345 mm diameter MAR Research dtb imaging-plate scanner mounted on a Rigaku RU-300 rotating-anode X-ray generator operating at 50 kV and 100 mA. Osmic Blue confocal optics were used to focus the Cu $K\alpha$ radiation. The data were indexed and scaled using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). A summary of the data-collection statistics is given in Table 1. The value of 12% for $R_{\rm sym}$ (Table 1) is a little high, but this is as expected considering the large number of weak intensities.

2.5. Structure determination and refinement

The structure was determined by the molecular-replacement method using the program *AMoRe* (Navaza, 1994) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The coordinates of SPC-40 (Kumar *et al.*, 2006; PDB code 2esc) were used as the search model. The rotation function was calculated using diffraction data in the resolution range 10.0–4.0 Å with a Patterson radius of 14 Å. Both rotation and translation searches resulted in unique solutions. Further positional and *B*-factor refinements were performed with *REFMAC*5.0 (Murshudov *et al.*, 1997). The refinement calculations were interleaved with several rounds of model building using *O* (Jones *et al.*, 1991). The electron densities for three segments, His188–His197, Phe202–Arg212 and Tyr244–Pro260, indicated discontinuities that caused difficulties in chain tracing. OMIT maps were calculated for these segments and the protein chains were adjusted into electron density with a lower cutoff (0.7σ). The structure was refined further, which reduced the $R_{\rm cryst}$ and $R_{\rm free}$ factors to 0.264 and 0.312, respectively. The Fourier $|2F_{\rm o} - F_{\rm c}|$ map indicated an

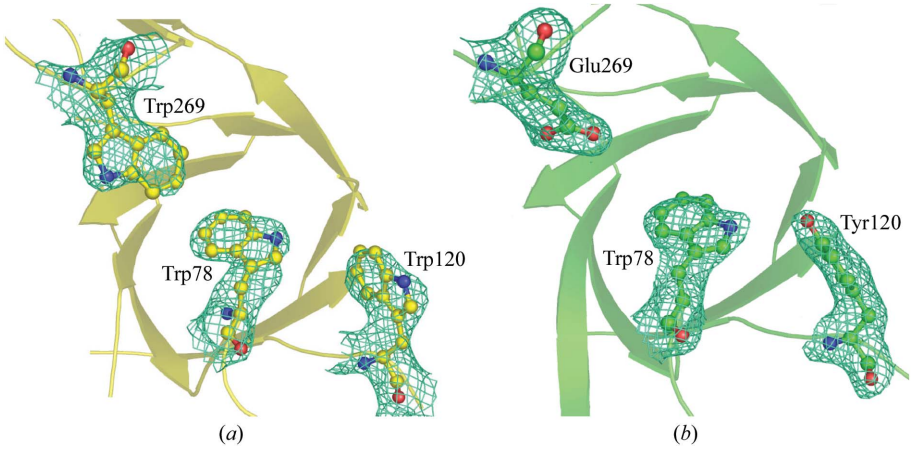


Figure 1
(*a*) $|F_{\rm o} - F_{\rm c}|$ map of Trp120 and Trp269 in SPB-40 (yellow). (*b*) Characteristic density for Tyr120 and Glu269 in SPC-40 (green). The electron densities show a clear difference in the side chains of residues 120 and 269 in SPB-40 and SPC-40.

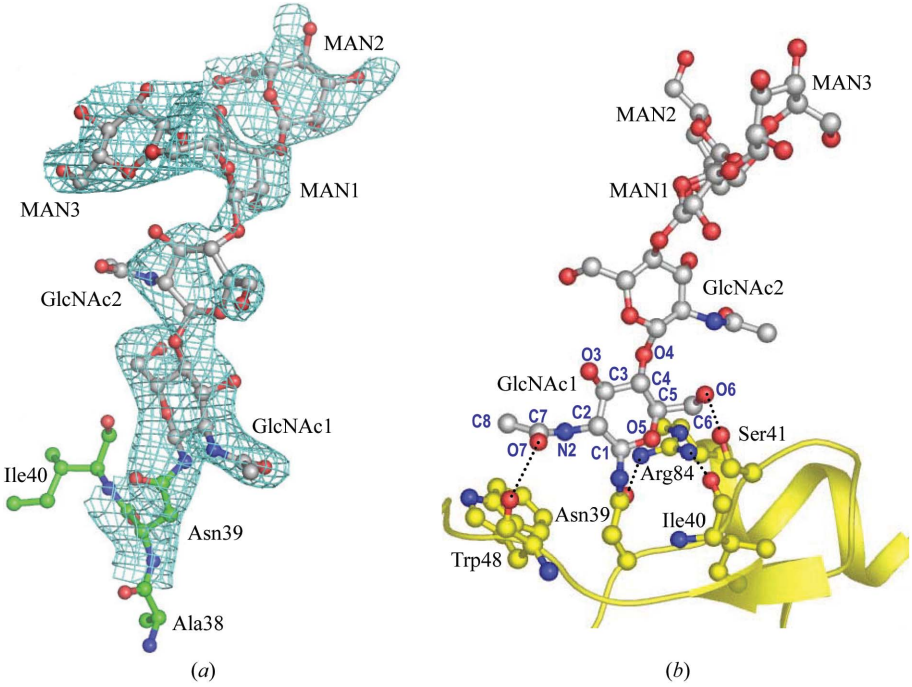


Figure 2
(*a*) $|F_{\rm o} - F_{\rm c}|$ electron-density map contoured at 2.5σ cutoff for a glycan chain consisting of two GlcNAc and three MAN residues linked to Asn39. (*b*) The interactions of the glycan chain with the protein chain are also shown.

ambiguity for Tyr120 and Glu269. Therefore, an OMIT map was calculated without these two residues, which showed characteristic density for a Trp residue at both positions. As seen from Fig. 1, Trp120 and Trp269 fit well into the density at these positions. These residues were included in subsequent refinement cycles. After further refinement and manual model rebuilding, difference $|F_o - F_c|$ and Fourier $|2F_o - F_c|$ maps were calculated which indicated the presence of an Asn39-linked carbohydrate chain containing two residues of *N*-acetylglucosamine (GlcNAc) and three residues of mannose (MAN) (Fig. 2). The oligosaccharide chain was refined using bond-length and bond-angle parameters from idealized GlcNAc and MAN residues and glycosidic linkages (Jeffrey, 1990). Water molecules were added using the program *ARP*/*wARP* (Perrakis *et al.*, 1999). Several rounds of refinement with *REFMAC*5.0 (Murshudov *et al.*, 1997)

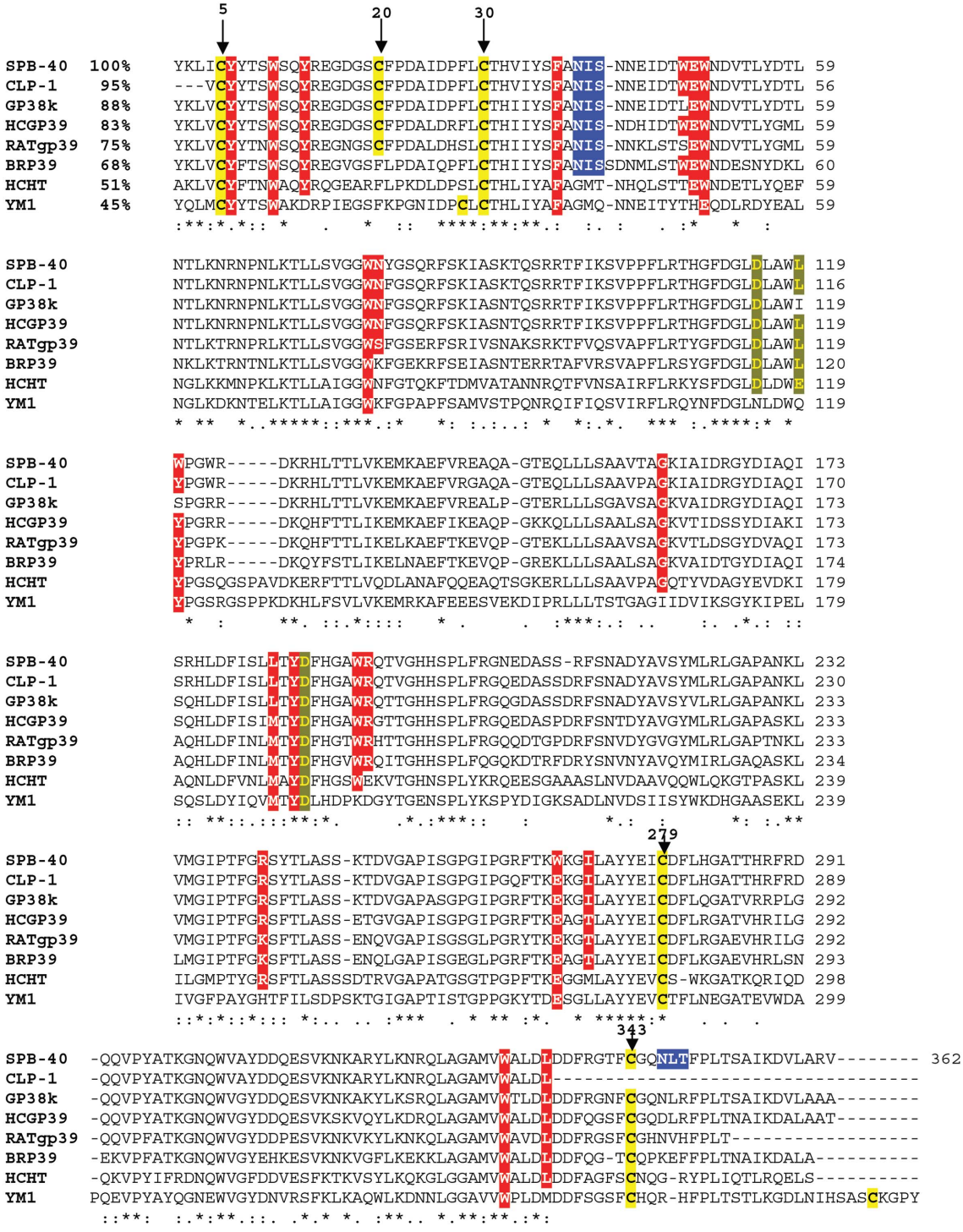


Figure 3
(*a*) Multiple sequence alignment of SPB-40 with CLP-1, GP38k, HCGP39, RATgp39, BRP39, HCHT and YM1. Catalytic residues (Asp115, Leu119 and Asp186) are indicated in green and residues making up the wall of the barrel and attributed as being involved in sugar binding are indicated in red. Cysteine residues are indicated in yellow and potential N-linked glycosylation sites are shown in blue.

interspersed with model building using $|2F_o - F_c|$ and $|F_o - F_c|$ Fourier maps caused the refinement to converge to R_{cryst} and R_{free} factors of 0.186 and 0.236, respectively. The positions of 117 water molecules were determined with peak electron densities greater than 2.5σ in the $|F_o - F_c|$ map and were retained in the final model only if they met the criteria of having peaks greater than 2.0σ in the $|2F_o - F_c|$ map, hydrogen-bonding partners with appropriate distance and angle geometry and B values less than 75 Å^2 in the final refinement cycle. The refinement statistics are summarized in Table 1.

3. Results and discussion

3.1. Sequence analysis

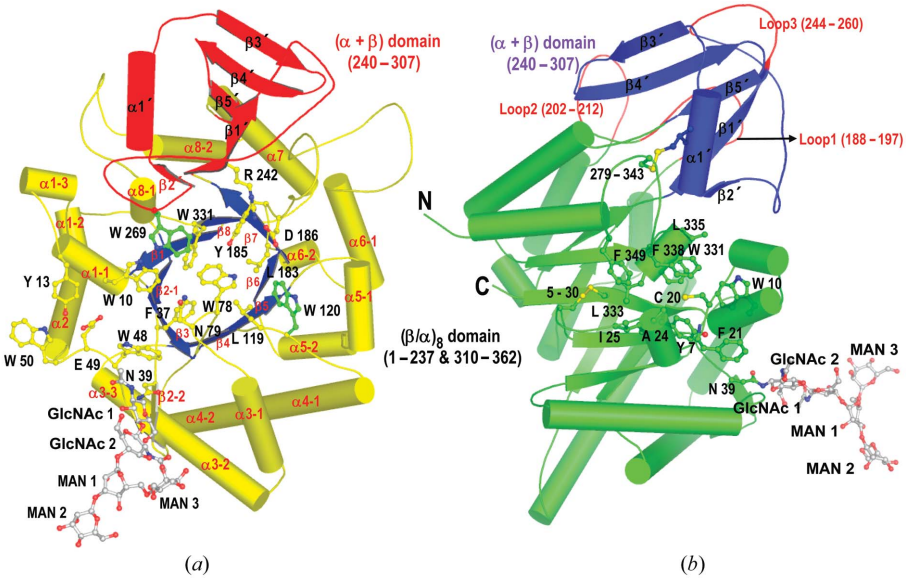
The complete sequence determination of the mature protein shows the presence of 361 amino-acid residues. The nucleotide and amino-acid sequences have been deposited in GenBank with accession No. AY295929. The sequence contains five cysteines, four of which are involved in the formation of two disulfide bridges, Cys5–Cys30 and Cys279–Cys343. The side chain of Cys20 is free. The sequence revealed two potential N-linked glycosylation sites with Asn39-Ile40-Ser41 and Asn346-Leu347-Thr348 motifs. The sequence identity of SPB-40 ranges between 96 and 68% with the group I proteins CLP-1 (GenBank accession No. AF011373), GP38k (Shackelton *et al.*, 1995; GenBank accession No. U19900), HCGP-39 (Hakala *et al.*, 1993; GenBank accession No. M80927), RATgp39 (GenBank accession No. AF062038) and BRP39 (Morrison & Leder, 1994; GenBank accession No. X93035) (Fig. 3). Its sequence identity with the group II protein HCHT is 52% (Fusetti *et al.*, 2002) and it shows an identity of 46% with the group III protein YM1 (Sun *et al.*, 2001) (Fig. 3).

3.2. Overall structure

The final model of the protein was evaluated using *PROCHECK* (Laskowski *et al.*, 1993), which indicates the presence of 87.4% of residues in the most favoured regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968). The refined model includes all 361 protein residues, five glycan (two GlcNAc and three MAN) residues and 117 water molecules, yielding R_{cryst} and R_{free} factors of 0.186 and 0.236, respectively. The r.m.s. deviations in bond lengths and angles are 0.02 Å and 2.0°, respectively. The structure of SPB-40 is broadly divided into two globular domain: a large $(\beta/\alpha)_8$ triosephosphate isomerase (TIM) barrel (Banner *et al.*, 1975) domain and a small $(\alpha+\beta)$ domain (Fig. 4*a*). The TIM-barrel domain contains both the N- and the C-termini and is made up of two polypeptide segments, 1–237 and 310–362. The polypeptide chain crosses over to form a small domain consisting of residues 240–307. The eight-stranded parallel β-sheet structure forms the core of the protein molecule, while eight pieces of α-helices surround it, covering at least three-quarters of the barrel from the outside. The interior of the barrel is predominantly filled with hydrophobic residues. There are two optimally formed disulfide bonds Cys5–Cys30 and Cys279–Cys343. The latter disulfide bond is formed between the two domains and contributes to holding the two domains together. There is a free Cys20 in SPB-40, which is located in a tightly packed hydrophobic pocket containing residues Tyr7, Ala24, Ile25, Phe338 and Phe349 (Fig. 4*b*). Hence, it is practically inaccessible to solvent molecules. The functional significance of Cys20 is not yet clear. Three *cis*-peptide bonds (residues Ser36-Phe37, Leu119-Trp120 and Trp331-Ala332) have been observed in the structure. Two of them, Ser36-Phe37 and Trp331-Ala332, are similar to the *cis*-peptide bonds reported in the structures of other members of group I (Sun *et al.*, 2001; Tsai *et al.*, 2004; Kumar *et al.*, 2006; Fusetti *et al.*, 2002, 2003; Houston *et al.*, 2003; Srivastava *et al.*, 2007). The third *cis*-peptide Leu119-Tyr120 is mutated to Leu119-Trp120 in SPB-40, which is observed for the first time in this family.

3.3. Glycosylation site

The amino-acid sequence of SPB-40 indicates two potential N-glycosylation sites with Asn-*X*-Ser/Thr sequence motifs; these are at Asn39 and Asn346. The Asn39 site is present in β-strand β_{2-1} and Asn346 is present in the loop region between α-helices α_{8-1} and α_{8-2}.



(*a*) (*b*)

Figure 4
(*a*) The overall fold of SPB-40 in top-view orientation indicating the $(\beta/\alpha)_8$ domain (yellow) and the $(\alpha+\beta)$ subdomain (red). The sugar-binding residues are also indicated. The glycan chain containing two GlcNAc and three MAN residues is also shown (grey). (*b*) The overall fold of SPB-40 in side-view orientation indicating the $(\beta/\alpha)_8$ domain (green) and the $(\alpha+\beta)$ subdomain (blue). The positions of disulfide bridges and the hydrophobic environment of the free Cys20 are also shown. The flexible region, loop1 (188–197), loop2 (202–212) and loop3 (244–260) are also shown (red). The figure was drawn using *PyMOL* (DeLano, 2002).

Our crystallographic analysis of SPB-40 only shows attachment at Asn39, with a remarkably good quality of electron density for two GlcNAc and three mannose residues. The glycan chain is stabilized by a number of hydrogen-bonded interactions with the protein chain. Apart from the covalent linkage to Asn39, the GlcNAc1 residue of glycan forms two direct hydrogen bonds, with the O6 and O7 atoms of GlcNAc1 interacting with the Ser41 O^{γ} and carbonyl O atom of Trp48 (Fig. 2*b*). Apart from these strong hydrogen bonds, the glycan chains are also stabilized by hydrophobic interactions with Trp48 and Arg84.

3.4. Comparison of SPX-40 structures

The overall polypeptide fold of SPB-40 is essentially similar to the structures of MGP-40, SPC-40 and SPS-40 (Mohanty *et al.*, 2003; Kumar *et al.*, 2006; Srivastava *et al.*, 2006). The root-mean-square (r.m.s.) shifts between the positions of C^{α} atoms (358 residues) of SPB-40 with MGP-40, SPC-40 and SPS-40 are 0.5, 0.4 and 0.5 Å, respectively (Mohanty *et al.*, 2003; Kumar *et al.*, 2006; Srivastava *et al.*, 2006). Although the overall folding is similar, there are notable variations in the sugar-binding groove owing to the mutation of important residues such as Trp120 and Trp269. Large-scale variations have also been observed in the arrangement of the flexible segments His188–His197, Phe202–Arg212 and Tyr244–Pro260.

3.5. Sugar-binding groove

The carbohydrate-binding groove in chitinases is divided into subsites using numbering from −6 to +3 (Davies *et al.*, 1997), with −6 being at the nonreducing end. Using the same procedure for SPX-40, the numbers for the subsites were assigned as −5 to +1 as the sugar binding beyond subsite +1 is not well defined (Srivastava *et al.*, 2007; Kumar *et al.*, 2007). The groove consists of mainly aromatic residues that are important for stacking interactions in the binding of GlcNAc oligomers.

The major part of the secondary structure that is involved in sugar binding consists of β_1, α_{1-1}, α_{1-2}, β_{2-1}, α_2, β_3, β_4, β_6, β_8 and α_{8-1} from the TIM-barrel domain and $\beta_{1'}$ and $\beta_{2'}$ from the small ($\alpha+\beta$) domain. The carbohydrate binding can be understood from the interactions between protein and sugar residues. As observed in SPS-40 and SPG-40 (Srivastava *et al.*, 2007; Kumar *et al.*, 2007), at subsite −5 the sugar makes stacking interactions with Trp50 and Tyr13 and is also held in place by a hydrogen-bonding interaction with Glu49. At subsite −4, residues Trp10 and Trp48 are involved in stacking interactions with the saccharide, while at subsite −3 the saccharide is held in place by interactions with Trp10 and Asn79. It is noteworthy that subsites −6 to −3 contain the same sugar-binding residues as observed in other members of group 1 (Hakala *et al.*, 1993; Shackelton *et al.*, 1995; Morrison & Leder, 1994; Kumar *et al.*, 2006; Fusetti *et al.*, 2002, 2003; Houston *et al.*, 2003; Srivastava *et al.*, 2007; Kumar *et al.*, 2007). In

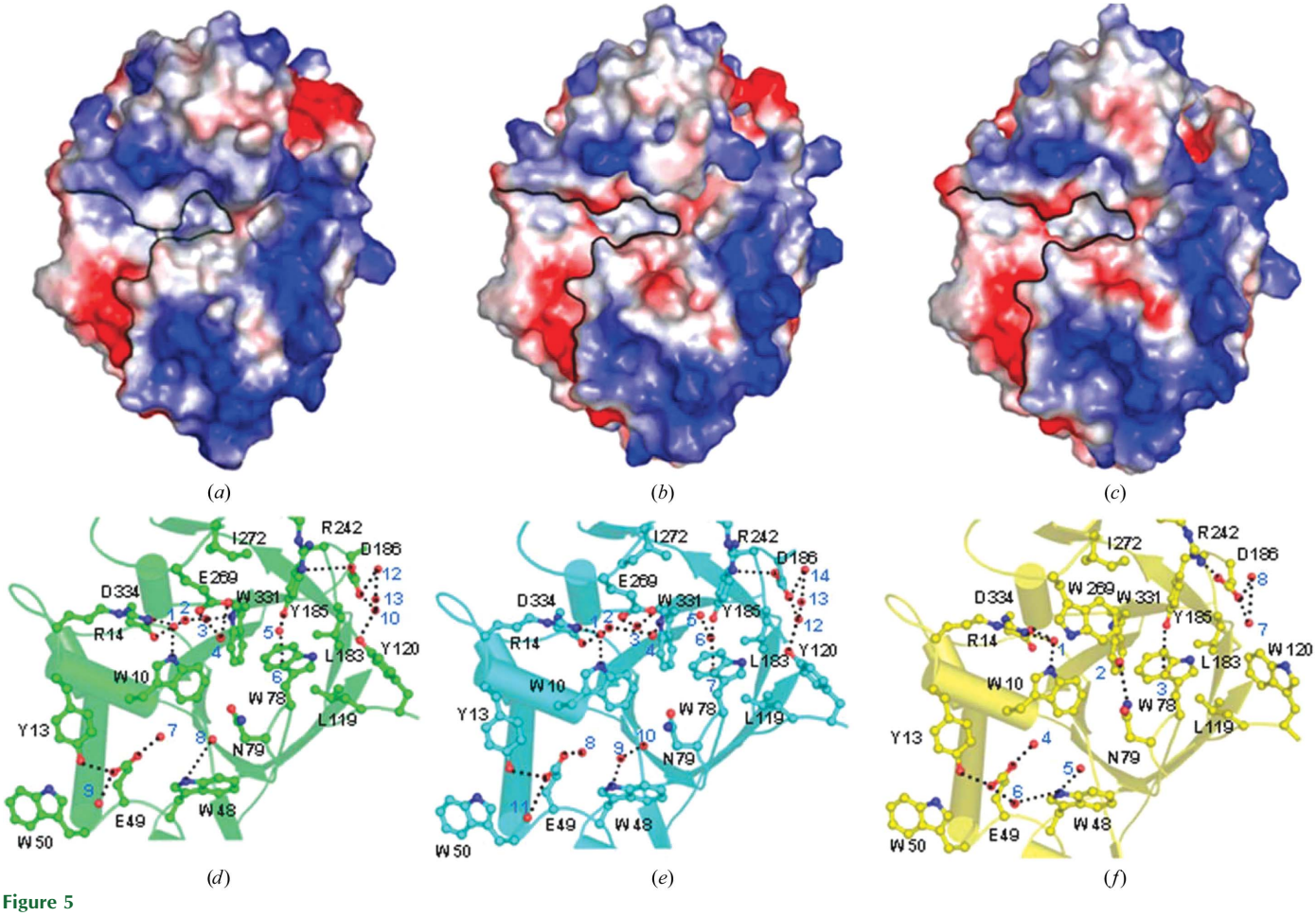


Figure 5
(*a*)–(*c*) A comparison of the sugar-binding groove of SPB-40 (*a*) with those of SPS-40 (*b*) and SPC-40 (*c*). The distortions in the shapes of the sugar-binding grooves are clearly visible. (*d*)–(*f*) A comparison of the water network in the sugar-binding groove of SPS-40 (*d*) with those in SPC-40 (*e*) and SPB-40 (*f*). The differences in the water network are a consequence of differences in the sequence.

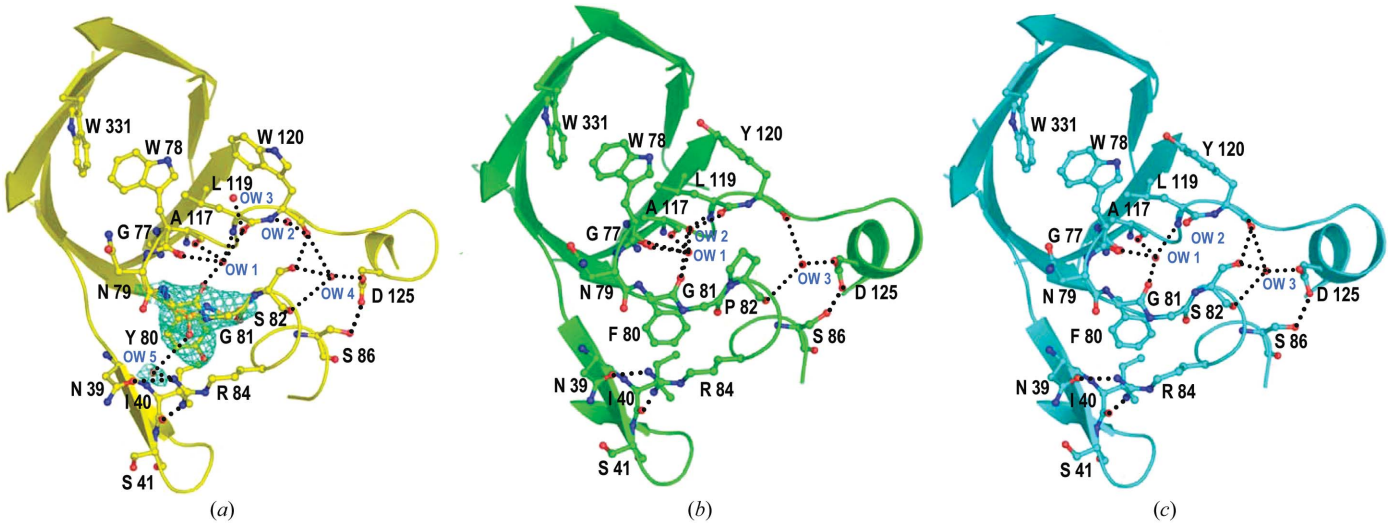


Figure 6
(*a*) As detemined by sequencing (GenBank accession No. AY295929) and indicated by a difference OMIT map, residue 80 in SPB-40 is Tyr. The corresponding residue in (*b*) SPS-40 and (*c*) SPC-40 is Phe. The hydrogen-bonding networks involving the Trp78 loop are also shown.

contrast, in the chitinase-like protein YM1 (group III; Sun *et al.*, 2001; Tsai *et al.*, 2004) the residues Tyr13, Glu49 and Trp50 are changed to the charged residues Asp13, His48 and Glu50, which cause additional distortion in the sugar-binding groove at subsites −5 and −4.

The structures of complexes of SPS-40 and SPG-40 (Srivastava *et al.*, 2007; Kumar *et al.*, 2007) show that at subsite −2 the residues Phe37, Asn79, Trp78, Trp331 and Glu269 are involved in extensive interactions with saccharides. Therefore, subsite −2 appears to be the most preferred subsite in these proteins (Srivastava *et al.*, 2007; Kumar *et al.*, 2007). At positions −1 and +1, the residues Trp78, Trp331, Arg242, Asp186, Tyr185, Leu183, Ala156 and Tyr120 are suitably aligned to interact with saccharides. Two important sugar-binding groove residues, Tyr120 and Glu269 in SPS-40 (Srivastava *et al.*, 2007) and SPG-40 (Kumar *et al.*, 2007), are changed to Trp120 and Trp269 in SPB-40. As these are part of the wall of the sugar-binding groove, this indicates a variation in the dimensions of the groove. As seen in Figs. 5(*a*)–5(*c*), the overall shape of the sugar-binding groove in SPB-40 differs considerably from those of SPS-40 and SPC-40. Furthermore, the network of water molecules in the sugar-binding groove of SPB-40 also differs from those observed in native SPS-40 (Srivastava *et al.*, 2006) and native SPC-40 (Kumar *et al.*, 2006). As seen in Figs. 5(*d*) and 5(*e*), the residues Tyr120 and Glu269 are involved in hydrogen bonding to the water network in the sugar-binding grooves of both SPS-40 (Srivastava *et al.*, 2006) and SPC-40 (Kumar *et al.*, 2006). Similar hydrogen bonding is absent in SPB-40 (Fig. 5*f*) owing to a lack of water molecules in this region. The relatively low number of water molecules in the SPB-40 sugar-binding groove is likely to primarly be a consequence of the different character of Trp120 and Trp269, the residues that replace Tyr120 and Glu269 in this variant, although the 2.8 Å resolution of the SPB-40 structure, which is relatively low, makes the modelling of water difficult. This water network has to be displaced, along with conformational changes of some of the important sugar-binding residues, Trp78, Asp186, Arg242 and Trp269, for binding of saccharides. It is noteworthy that these sugar-binding residues are located in the loop regions, indicating the possibility that they may undergo conformational changes.

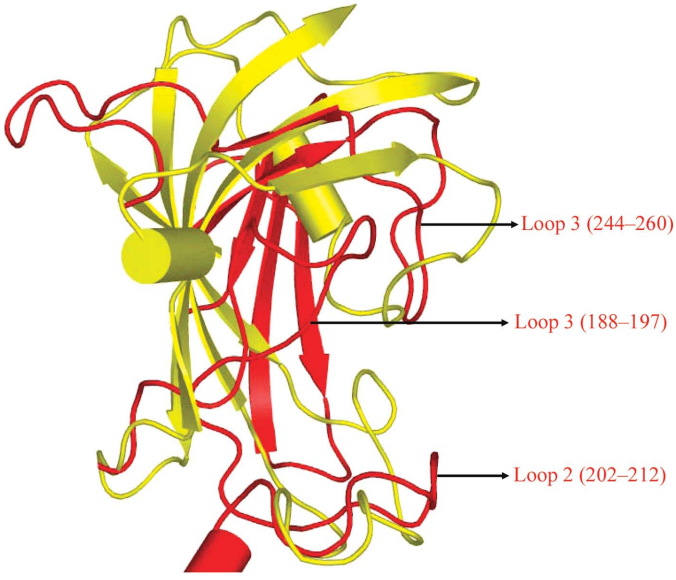


Figure 7
Superimposition of the small ($\alpha+\beta$) domain, including the three flexible loops, of SPB-40 (red) onto the FK-binding protein (yellow).

3.6. Trp78 and its conformation

The most striking structural feature that influences the saccharide binding to the groove in SPX-40 is the conformation of Trp78 and its manoeuverability upon the introduction of carbohydrate molecules. The structure of SPB-40 shows that, as observed in other SPX-40 proteins (Mohanty *et al.*, 2003; Kumar *et al.*, 2006; Srivastava *et al.*, 2006), the side chain of Trp78 is oriented into the barrel and the loop supporting it is stabilized by a network of hydrogen bonds. In the structure of SPS-40 (Srivastava *et al.*, 2006) Gly77, Asn79, Phe80, Gly81, Pro82, Arg84, Ala117, Leu119 and Tyr120 and three water molecules OW1, OW2 and OW3 (Fig. 6*b*) are part of the hydrogen-bonded network, while in SPC-40 (Kumar *et al.*, 2006) Gly77, Asn79, Phe80, Gly81, Ser82, Arg84, Ala117, Leu119 and Tyr120 and two water molecules OW1 and OW2 (Fig. 6*c*) are involved in the corresponding network of hydrogen bonds. In SPB-40, two important residues, Phe80 and Tyr120 of the supporting loop, are replaced by Tyr80 and Trp120 (Fig. 6*a*). As a result, Tyr80 and Gly81 adopt two

alternate conformations with (φ, ψ) values of (−123.5, 129.0°)/(−86.8, −66.2°) and (−66.6, 132.9°)/(124.8, 100.0°), respectively. As seen in Fig. 6(*a*), the carbonyl O atom of Tyr80 occupies two positions and both are involved in hydrogen bonds to water molecules OW1 and OW5. The residues involved in the hydrogen-bonded network in the loop are Gly77, Asn79, Tyr80, Gly81, Ser82, Arg84, Ala117, Leu119 and Trp120, together with four water molecules OW1, OW2, OW3 and OW4. Since these interactions stabilize the conformation of the loop that lends support to the orientation of Trp78, a change in them will directly affect the orientation of the side chain of Trp78.

3.7. Flexible region

The least-squares superposition of the C^{α} trace of SPB-40 on those of SPC-40 and SPS-40 (Kumar *et al.*, 2006; Srivastava *et al.*, 2006) indicates large-scale variations for three segments consisting of residues His188–His197 (loop 1), Phe202–Arg212 (loop 2) and Tyr244–Pro260 (loop 3). The average r.m.s. shift for the C^{α} atoms of these three segments is 1.8 Å, while the individual side chains adopt substantially different orientations. The average values of the *B* factor for the atoms of these loops are ∼70 Å^2. This value is more than twice the average value for the rest of the molecule. Despite being in close proximity, the three loops interact poorly. Another notable feature is reflected by an array of charged side chains that protrude outward. The overall folding of this region together with the $\alpha+\beta$ domain (Fig. 7) represents a similar arrangement to that observed for the FK-binding protein (Itoh & Navia, 1995). It may be mentioned here that the FK-binding protein binds to the type I TGFβ receptor (Huse *et al.*, 1999). On the basis of structural similarity, it may be speculated that the flexible region including the $\alpha+\beta$ domain of SPB-40 protein may also be involved in interactions with the TGFβ type of receptors.

4. Conclusions

The structurally similar proteins of group I (Mohanty *et al.*, 2003; Kumar *et al.*, 2006; Srivastava *et al.*, 2006) contain similar structural arrangements, including the formation of a TIM barrel, an (a+β) domain and two identical disulfide bridges. The mutation of one of the conserved *cis*-peptide residues in Leu119–Tyr120 to Leu119–Trp120 in SPB-40 is a notable variation. It is directly related to the shape of the sugar-binding groove. The mutations of two important sugar-binding residues, Tyr120 and Glu269, to Trp120 and Trp269, respectively, appear to have notable effects on the shape of the sugar-binding groove. Another observation pertains to the flexible region consisting of three loops, His188–His197, Phe202–Arg212 and Tyr244–Pro260, which shows the presence of an FK-binding protein-like domain, with the residues having high temperature factors and charged residues such as serines, aspartates, glutamates and arginines protruding outward from the surface of the protein, suggesting the possibility of this site forming intermolecular interactions, presumably with receptors, involving protein–protein recognition in addition to sugar binding.

The authors thank the Department of Science and Technology (DST), New Delhi for financial support. A liberal grant from Department of Biotechnology (DBT), New Delhi for establishing the Clinical Proteomics facility is gratefully acknowledged. The DST is also thanked for support under the FIST programme for a level II grant. ASE, DBS and JK thank the Council of Scientific and Industrial Research (CSIR), New Delhi for the award of fellowships.

References

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. (1975). *Nature (London)*, **255**, 609–614.
Chomczynski, P. & Sacchi, N. (1987). *Anal. Biochem.* **162**, 156–159.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davies, G. J., Wilson, S. K. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.
DeLano, W. L. (2002). *The PyMOL User's Manual.* DeLano Scientific, San Carlos, CA, USA.
Fusetti, F., Pijning, T., Kalk, K. H., Bos, E. & Dijkstra, B. W. (2003). *J. Biol. Chem.* **278**, 37753–37760.
Fusetti, F., von Moeller, H., Houston, D., Rozeboom, H. J., Dijkstra, B. W., Boot, R. G., Aerts, J. M. & van Aalten, D. M. (2002). *J. Biol. Chem.* **277**, 25537–25544.
Hakala, B. E., White, C. & Recklies, A. D. (1993). *J. Biol. Chem.* **268**, 25803–25810.
Houston, R. D., Recklies, A. D., Krupa, J. C. & van Aalten, D. M. (2003). *J. Biol. Chem.* **278**, 30206–30212.
Hurley, W. L. (1989). *J. Dairy Sci.* **72**, 1637–1646.
Huse, M., Chen, Y. G., Massague, J. & Kuriyan, J. (1999). *Cell*, **96**, 425–436.
Itoh, S. & Navia, M. A. (1995). *Protein Sci.* **11**, 2261–2268.
Jeffrey, G. A. (1990). *Acta Cryst.* B**46**, 89–103.
Johansen, J. S., Jensen, H. S. & Price, P. A. (1993). *Br. J. Rheumatol.* **32**, 949–955.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kumar, J., Ethayathulla, A. S., Srivastava, D. B., Sharma, S., Singh, S. B., Srinivasan, A., Yadav, M. P. & Singh, T. P. (2006). *Acta Cryst.* D**62**, 953–963.
Kumar, J., Ethayathulla, A. S., Srivastava, D. B., Singh, N., Sharma, S., Kaur, P., Srinivasan, A. & Singh, T. P. (2007). *Acta Cryst.* D**63**. In the press.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Mohanty, A. K., Singh, G., Paramasivam, M., Saravanan, K., Jabeen, T., Sharma, S., Yadav, S., Kaur, P., Kumar, P., Srinivasan, A. & Singh, T. P. (2003). *J. Biol. Chem.* **278**, 14451–14460.
Morrison, B. W. & Leder, P. (1994). *Oncogene*, **9**, 3417–3426.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.
Rejman, J. J. & Hurley, W. L. (1988). *Biochem. Biophys. Res. Commun.* **150**, 329–334.
Renkema, G. H., Boot, R. G., Muijsers, A. O., Donker-Koopman, W. E. & Aerts, J. M. F. G. (1995). *J. Biol. Chem.* **270**, 2198–2202.
Shackelton, L. M., Mann, D. M. & Millis, A. J. (1995). *J. Biol. Chem.* **270**, 13076–13083.
Srivastava, D. B., Ethayathulla, A. S., Kumar, J., Singh, N., Das, U., Sharma, S., Srinivasan, A. & Singh, T. P. (2006). *J. Struct. Biol.* **156**, 505–516.
Srivastava, D. B., Ethayathulla, A. S., Kumar, J., Somvanshi, R. K., Sharma, S., Dey, S. & Singh, T. P. (2007). In the press.
Sun, Y. J., Chang, N. C., Hung, S. I., Chang, A. C., Chou, C. C. & Hsiao, C. D. (2001). *J. Biol. Chem.* **276**, 17507–17514.
Tsai, M. L., Liaw, S. H. & Chang, N. C. (2004). *J. Struct. Biol.* **148**, 290–296.