Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a bovine secretory signalling glycoprotein (SPC-40) at 2.1 Å resolution

Janesh Kumar,[a] Abdul S. Ethayathulla,[a] Devendra B. Srivastava,[a] Sujata Sharma,[a] S. Baskar Singh,[a] Alagiri Srinivasan,[a] Mahendra P. Yadav[b] and Tej P. Singh[a]*

[a]Department of Biophysics, All India Institute of Medical Sciences, New Delhi 110029, India, and [b]Indian Veterinary Research Institute, Izatnagar 243122, India

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

Received 11 April 2005
Accepted 30 May 2006

PDB Reference: SPC-40, 2esc, r2escsf.

A recently discovered new class of 40 kDa glycoproteins forms a major component of the secretory proteins in the dry secretions of non-lactating animals. These proteins are implicated as protective signalling factors that determine which cells are to survive during the processes of drastic tissue remodelling. In order to understand its role in the remodelling of mammary glands, the detailed three-dimensional structure of the bovine signalling glycoprotein (SPC-40) has been determined using X-ray crystallography. SPC-40 was purified from bovine dry secretions and crystallized using the hanging-drop vapour-diffusion method. The crystals belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 62.6$, $b = 67.4$, $c = 106.9$ Å. The protein was also cloned in order to determine its complete amino-acid sequence. Its three-dimensional structure has been determined using data to 2.1 Å resolution. The amino-acid sequence determination of SPC-40 reveals two potential N-glycosylation sites at Asn39 and Asn345, but electron density for a glycan chain was only present at Asn39. The protein adopts a conformation with the classical $(\beta/\alpha)_8$-barrel fold of triosephosphate isomerase (TIM barrel; residues 1–237 and 310–360) with the insertion of a small $\alpha+\beta$ domain (residues 240–307) similar to that observed in chitinases. However, the substitution of Leu for Glu in the consensus catalytic sequence in SPC-40 caused a loss of chitinase activity. Furthermore, the chitin-binding groove in SPC-40 is considerably distorted owing to unfavourable conformations of several residues, including Trp78, Tyr120, Asp186 and Arg242. Three surface loops, His188–His197, Phe202–Arg212 and Tyr244–Pro260, have exceptionally high B factors, suggesting large-scale flexibility. Fluorescence studies indicate that various sugars bind to SPC-40 with low affinities.

1. Introduction

The non-lactating or dry period is an important phase for optimizing milk production during subsequent lactation (Coppock *et al.*, 1974). The transition period between the lactating and non-lactating states of the mammary gland is a period of active involution during which the mammary gland undergoes extensive ultra-structural changes and the secretions contained in the gland undergo dramatic compositional changes (Hurley, 1989). A prominent protein in the whey secretions of non-lactating cows was reported to be an important marker protein for mammary function during that

period (Rejman & Hurley, 1988). It is a novel glycoprotein and seems to act as a protective signalling factor during extensive tissue remodelling. This signalling protein has a molecular weight of 40 kDa and is named SPC-40 (S, signalling; P, protein; C, cow; 40 is the molecular weight). It has a sequence identity of more than 75% to several 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, while the identical Cys20 is unpaired. Yet another similar protein, BRP39 (breast regression protein with MW 39 kDa), which also showed a sequence identity of 69% to SPC-40, was identified from a specific type of cancer of the mammary gland of mice (Morrison & Leder, 1994). Owing to their very similar amino-acid sequences and chemical properties, these mammalian proteins form a subclass of closely related mammalian glycoproteins (group I). These group I proteins are also homologous to chitinases, with an approximate sequence identity of 52% with human chitinase (Renkema *et al.*, 1995). The chitinases have a well defined carbohydrate-binding groove in which oligomers of *N*-acetylglucosamine (chitin polymers) bind preferentially. The active site of chitinases involves three acidic amino acids: Asp, Glu and Asp. A sequence comparison of SPC-40 with chitinases shows that the residue corresponding to Glu is changed to Leu. This abolishes the chitin-hydrolyzing capability of SPC-40. In spite of several similarities, chitinases are distinct functionally, with different properties to group I proteins. Therefore, they are classified here as group II proteins. There is yet another class of closely related proteins designated chitinase-like proteins. These proteins are also catalytically inactive owing to mutation of one of the catalytic residues, but are similar to chitinases in the folding of their polypeptide chain. They lack glycosylation sites and show mutations of several carbohydrate-binding residues. The prominent protein among them is a novel mammalian lectin, YM1 (Sun *et al.*, 2001). Although initially YM1 was reported to bind to carbohydrates (Sun *et al.*, 2001), it was subsequently described as not possessing an ideal carbohydrate-binding site (Tsai *et al.*, 2004). Overall, YM1 is homologous to both group I and group II proteins, but has several unique features that are distinct from members of both groups. The proteins of the YM1 subclass will be referred to hereafter as group III proteins. Because of the sequence and structural similarities between proteins of groups I, II and III, they essentially belong to a single superfamily, referred to hereafter as the SPX-40 superfamily for ease of discussion. In order to understand the function of SPC-40, we have determined its complete amino-acid sequence and its detailed three-dimensional structure at 2.1 Å resolution. The structure contains an eight-stranded large β/α-topoisomerase (TIM) barrel domain with a small $\alpha+\beta$ domain inserted in one of the barrel loops. Of the two potential glycosylation sites indicated by the amino-acid sequence, only one has been found to be glycosylated, with two GlcNAc and four Man residues. The structure also contains a unique conformationally flexible region consisting of three loops, His188–His197, Phe202–Arg212 and Tyr244–Pro260, which are presumably involved in receptor binding as they contain a number of free serine and arginine residues that protrude outward. As in other structures of the SPX-40 superfamily, three *cis*-peptides Ser36–Phe37, Leu119–Tyr120 and Trp331–Ala332 have been observed.

2. Experimental procedures

2.1. Isolation and purification

Fresh mammary secretions were collected from Holstein-Friesian cows 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 8.0. The cation exchanger CM-Sephadex C-50 (7 g l^{-1}) equilibrated in 50 m*M* Tris–HCl pH 8.0 was added and the mixture was stirred slowly for about 1 h with a mechanical stirrer. In order to remove the unbound proteins, the protein-bound gel was washed with an excess of 50 m*M* Tris–HCl pH 8.0. 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, which removed other impurities. The eluted protein solution was desalted and was again passed through a CM-Sephadex C-50 column (10 × 2.5 cm) which was pre-equilibrated with 50 m*M* Tris–HCl pH 8.0 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 concentrated samples were passed through a Sephadex G-100 column (100 × 2 cm) using 50 m*M* Tris–HCl pH 8.0 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 SDS–PAGE. The protein samples were blotted onto a polyvinylidene fluoride (PVDF) membrane. The sequence of the first 20 amino -acid residues from the N-terminus was determined using a PPSQ20 protein sequencer (Shimadzu, Kyoto, Japan) and clearly showed 100% identity with the already known N-terminal sequence of bovine signalling protein (Rejman & Hurley, 1988).

2.2. Complete sequence determination

Mammary-gland tissue was obtained from a non-lactating cow during the period of early involution and the complete cDNA sequence was determined. The total RNA was extracted by the phenol/chloroform method (Chomczynski & Sacchi, 1987). The reverse-transcription reaction took place

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 conserved nucleotide sequences from other proteins of the SPX-40 family (Johansen *et al.*, 1993; Hakala *et al.*, 1993; Shackelton *et al.*, 1995; Morrison & Leder, 1994; Mohanty *et al.*, 2003) and the N-terminal sequence of SPC-40 as obtained using Edman degradation were used for the design of primers. The sequences 5′-CTATCCTGTCGAGGCCAAAGGA-3′ and 5′-AATTTATTGGACCTTCTGGCC-3′ were used as forward and reverse primers, respectively. The PCRs were carried out with Taq polymerase (Promega, Madison, USA) using an MJ Research thermal cycler (model PTC-100). The 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 are given in Fig. 1.

2.3. Fluorescence studies of protein–carbohydrate binding

In order to evaluate the binding characteristics of carbohydrates to SPC-40, various sugar compounds were used including monosaccharides such as glucose (Glc), *N*-acetylglucosamine (GlcNAc), glucosamine (GlcN), galactose (Gal), *N*-acetylgalactosamine (GalNAc) and mannose (Man), disaccharides such as $(GlcNAc)_2$, lactose and trehalose and chitooligosaccharides such as $(GlcNAc)_3$, $(GlcNAc)_4$, $(GlcNAc)_5$ and $(GlcNAc)_6$ (Sigma Chemical Co., St Louis, USA). As a positive control for the experiments, chitinase

```
   1 TAC AAG CTG ATC TGC TAC TAC ACC AGC TGG TCC CAG TAC CGG GAG GGT GAT GGG AGC TGC TTC CCA GAC GCC ATC GAC CCC    81
   1 Tyr Lys Leu Ile Cys Tyr Tyr Thr Ser Trp Ser Gln Tyr Arg Glu Gly Asp Gly Ser Cys Phe Pro Asp Ala Ile Asp Pro    27

  82 TTC CTG TGC ACC CAT GTC ATC TAC AGC TTT GCC AAC ATA AGC AAC AAT GAG ATC GAC ACC TGG GAG TGG AAT GAC GTG ACG   162
  28 Phe Leu Cys Thr His Val Ile Tyr Ser Phe Ala Asn Ile Ser Asn Asn Glu Ile Asp Thr Trp Glu Trp Asn Asp Val Thr    54

 163 CTC TAT GAC ACA CTG AAC ACA CTC AAG AAC AGG AAC CCC AAC CTG AAG ACC CTC CTA TCT GTT GGA GGA TGG AAC TTC GGT   243
  55 Leu Tyr Asp Thr Leu Asn Thr Leu Lys Asn Arg Asn Pro Asn Leu Lys Thr Leu Leu Ser Val Gly Gly Trp Asn Phe Gly    81

 244 TCT GAA AGA TTT TCC AAG ATA GCT TCC AAG ACC CAG AGT CGC AGG ACT TTC ATC AAG TCG GTG CCA CCA TTT CTG CGG ACC   324
  82 Ser Glu Arg Phe Ser Lys Ile Ala Ser Lys Thr Gln Ser Arg Arg Thr Phe Ile Lys Ser Val Pro Pro Phe Leu Arg Thr   108

 325 CAT GGC TTT GAT GGA CTG GAC CTA GCA TGG CTC TAC CCC GGG TGG AGA GAC AAG CGG CAT CTC ACC ACT CTG GTC AAG GAA   405
 109 His Gly Phe Asp Gly Leu Asp Leu Ala Trp Leu Tyr Pro Gly Trp Arg Asp Lys Arg His Leu Thr Thr Leu Val Lys Glu   135

 406 ATG AAG GCT GAG TTT GTA AGG GAA GCC CAA GCA GGC ACA GAG CAG CTT CTG CTC AGT GCA GCA GTA ACT GCA GGG AAG ATT   486
 136 Met Lys Ala Glu Phe Val Arg Glu Ala Gln Ala Gly Thr Glu Gln Leu Leu Leu Ser Ala Ala Val Thr Ala Gly Lys Ile   162

 487 GCT ATT GAC AGA GGC TAT GAC ATC GCC CAG ATA TCC CGA CAC CTG GAC TTC ATC AGC CTT TTG ACC TAT GAC TTT CAC GGA   567
 163 Ala Ile Asp Arg Gly Tyr Asp Ile Ala Gln Ile Ser Arg His Leu Asp Phe Ile Ser Leu Leu Thr Tyr Asp Phe His Gly   189

 568 GGC TGG CGC GGC ACA GTC GGA CAC CAC AGC CCC CTG TTT CGA GGC AAC AGC GAT GGA AGT TCT AGA TTC AGT AAC GCT GAC   648
 190 Gly Trp Arg Gly Thr Val Gly His His Ser Pro Leu Phe Arg Gly Asn Ser Asp Gly Ser Ser Arg Phe Ser Asn Ala Asp   216

 649 TAC GCT GTG AGC TAC ATG CTG AGG CTG GGG GCT CCA GCC AAT AAG CTG GTG ATG GGT ATC CCC ACT TTT GGG AGG AGC TAC   729
 217 Tyr Ala Val Ser Tyr Met Leu Arg Leu Gly Ala Pro Ala Asn Lys Leu Val Met Gly Ile Pro Thr Phe Gly Arg Ser Tyr   243

 730 ACT CTG GCC TCT TCC AGC ACA AGG GTG GGA GCC CCC ATC TCA GGG CCA GGA ATT CCA GGC CAG TTC ACC AAG GAG AAA GGG   810
 244 Thr Leu Ala Ser Ser Ser Thr Arg Val Gly Ala Pro Ile Ser Gly Pro Gly Ile Pro Gly Gln Phe Thr Lys Glu Lys Gly   270

 811 ATC CTT GCC TAT TAT GAG ATC TGT GAC TTC CTC CAC GGA GCC ACC ACC CAC AGA TTC CGG GAC CAG CAG GTC CCC TAT GCC   891
 271 Ile Leu Ala Tyr Tyr Glu Ile Cys Asp Phe Leu His Gly Ala Thr Thr His Arg Phe Arg Asp Gln Gln Val Pro Tyr Ala   297

 892 ACC AAG GGC AAC CAG TGG GTG GCG TAT GAC GAC CAG GAG AGT GTC AAA AAC AAG GCA CGG TAC CTG AAG AAC AGG CAG CTG   972
 298 Thr Lys Gly Asn Gln Trp Val Ala Tyr Asp Asp Gln Glu Ser Val Lys Asn Lys Ala Arg Tyr Leu Lys Asn Arg Gln Leu   324

 973 GCT GGC GCC ATG GTA TGG GCC CTG GAC CTG GAC GAC TTC CGG GGC ACC TTC TGT GGG CAG AAC CTG ACC TTT CCT CTC ACG  1053
 325 Ala Gly Ala Met Val Trp Ala Leu Asp Leu Asp Asp Phe Arg Gly Thr Phe Cys Gly Gln Asn Leu Thr Phe Pro Leu Thr   351

1054 AGT GCC ATC AAG GAT GTG CTT GCT AGG GTG TAG                                                              1086
 352 Ser Ala Ile Lys Asp Val Leu Ala Arg Val ***                                                               361
```

Figure 1
Nucleotide and deduced amino-acid sequences of signalling protein from cow (SPC-40). The amino acids are shown as three-letter codes. The stop codon is indicated by ***. The sequence has been deposited in GenBank with accession No. AY291312.

Table 1
Crystallographic data and refinement statistics.

The values given in parentheses correspond to the highest resolution shell.

Data-collection statistics	
PDB code	2esc
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	62.6
b	67.4
c	106.9
V_M (Å^3 Da^{-1})	2.8
Solvent content (%)	56.4
Resolution range (Å)	20.0–2.1
No. of unique reflections	26514
Completeness (%)	97.7 (97.8)
R_{sym} (%)	7.7 (44.7)
Mean $I/\sigma(I)$	8.8 (2.0)
Refinement statistics	
R_{cryst} (%)	17.3
R_{free} (%)	22.0
Protein atoms	2872
Sugar atoms (two GlcNAc and four Man)	72
Water molecules	247
R.m.s. deviation in bond lengths (Å)	0.01
R.m.s. deviation in bond angles (°)	2.0
R.m.s. deviation in dihedral angles (°)	21.3
Average *B* factor from Wilson plot (Å^2)	36.0
Average *B* factor for all atoms (Å^2)	41.3
Residues in most preferred regions (%)	91.7
Residues in additionally allowed regions (%)	8.3

from *Penicillium chrysogenum* was titrated with $(GlcNAc)_4$. Solute-quenching experiments were also performed using KI with chitinase and SPC-40 in the presence and absence of sugars (Boraston *et al.*, 2000). The binding of these sugars to protein was monitored by measuring the tryptophan fluorescence (Tabary & Frenoy, 1985; Eftink, 1997). All fluorescence experiments were performed on a Hitachi F-4500 fluorescence spectrophotometer (Tokyo, Japan). The excitation wavelength was fixed at 295 nm. Emission intensities were collected over the wavelength range 315–380 nm. The excitation and emission slit widths were kept at 5 nm. Fluorescence emission scans were performed at room temperature by titrating several concentrations (25, 50, 100, 150, 200, 250, 300 and 350 µ*M*) of these ligands with 1 µ*M* SPC-40 in 25 m*M* Tris–HCl pH 8.0. The data were corrected using titration without SPC-40 at the corresponding ligand concentrations in 25 m*M* Tris–HCl pH 8.0. The equilibrium dissociation constants were obtained by fitting the fluorescence intensity data to the following single-site binding equation using nonlinear regression analysis (*GraphPad Prism* v.4.03 for Windows, GraphPad Software, California, USA): $F - F_o = (F_b - F_o) \times [L_o/(K_d + L_o)]$, where F and F_o refer to the fluorescence intensity in the presence and absence of ligand, respectively, F_b refers to the maximum fluorescence signal of the SPC–ligand complex at saturation, L_o is the initial ligand concentration and K_d is the equilibrium dissociation constant.

2.4. Protein crystallization

The purified samples of SPC-40 were used for crystallization by the hanging-drop vapour-diffusion method. 30 mg ml^{-1} protein solution in 25 m*M* Tris–HCl, 50 m*M* NaCl pH 7.8 was equilibrated against the same buffer containing 19%(*v*/*v*) ethanol at 298 K. Thin square-shaped colourless crystals with dimensions of up to 0.40 × 0.35 × 0.15 mm were obtained within a week.

2.5. X-ray intensity data collection and processing

A suitable crystal with dimensions of 0.30 × 0.25 × 0.15 mm was used for data collection at 278 K. 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. The Osmic Blue confocal optics was used for focusing the Cu *Kα* radiation. The data were indexed and scaled using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). The crystals belong to space group $P2_12_12_1$, with unit-cell parameters a = 62.6, b = 67.4, c = 106.9 Å. The unit cell contains four molecules, with a single monomer in the asymmetric unit. The final data show an overall completeness of 98%, with an R_{sym} of 7.7% to 2.1 Å resolution. A summary of the data-collection statistics is given in Table 1.

2.6. Structure determination and refinement

The structure was determined by the molecular-replacement method using the program *AMoRe* from the *CCP*4 suite (Navaza, 1994). The coordinates of MGP-40, which has a sequence identity of 96% to SPC-40 (Mohanty *et al.*, 2003; PDB code 1ljy), 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 well above the noise levels. Further positional and *B*-factor refinements were performed with *REFMAC*5 (Murshudov *et al.*, 1997).

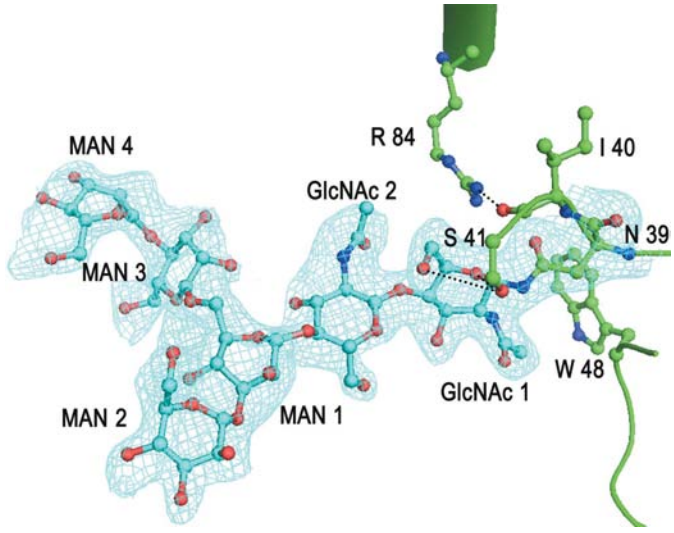

Figure 2
Final $|2F_o - F_c|$ electron-density map at 1.2σ cutoff for a glycan chain (cyan) consisting of two GlcNAc and four Man residues linked to Asn39 (green). The N-linked glycan chain (cyan) makes two hydrogen bonds to Ser41 O$^{\gamma}$. Arg84 interacts with Ile40.

The refinement calculations were interleaved with several rounds of model building with *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 densities with lower cutoff (0.7σ). The structure was refined further, reducing the $R_{\rm cryst}$ and $R_{\rm free}$ factors to 0.264 and 0.312, respectively. The difference electron-density $|F_{\rm o} - F_{\rm c}|$ map computed at this stage indicated the presence of an Asn39-linked carbohydrate chain containing two residues of GlcNAc and four residues of 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* (Collaborative Computational Project, Number 4, 1994). Several further rounds of refinement with *REFMAC*5 interspersed with model building using $|2F_{\rm o} - F_{\rm c}|$ and $|F_{\rm o} - F_{\rm c}|$ Fourier maps caused the refinement to converge to $R_{\rm cryst}$ and $R_{\rm free}$ factors of 0.173 and 0.220, respectively. The positions of 247 water molecules were determined with peak electron densities greater than 2.5σ in the $|F_{\rm o} - F_{\rm c}|$ map and were retained in the final model only if they met the criteria of having peaks greater than 1.5σ in the $|2F_{\rm o} - F_{\rm c}|$ 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 acids (Fig. 1). The complete nucleotide and amino-acid sequences have been deposited in GenBank with accession No. AY291312. The sequence contains five cysteines, four of which are involved in the formation of two disulfide bridges, Cys5–Cys30 and Cys279–Cys343, while the side chain of Cys20 is free. The sequence revealed two potential N-linked glycosylation sites with Asn39-Ile40-Ser41 and Asn345-Leu346-Thr347 motifs. The sequence identity of SPC-40 ranges from 95 to 69% to the other 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). SPC-40 also shows significant sequence identities with proteins of groups II and III such as human chitinase (HCHT; Fusetti *et al.*, 2002) and chitinase-like protein YM1 (Sun *et al.*, 2001) (Fig. 3). It may be mentioned here that chitinase enzymes (group II) hydrolyze chitin polymers and their catalytic site includes the three acidic residues Asp, Glu and Asp. In this regard, group I proteins are catalytically inactive because of the mutation of the catalytic Glu to Leu. The members of group III are also unable to hydrolyze chitin polymers because of the mutation of one of the active-site residues. The aromatic residues that are lined up along the walls of the sugar-binding groove and are implicated in sugar binding are generally conserved in these proteins (Fig. 3). As seen from Fig. 3, there are notable mutations in the sugar-binding residues of YM1.

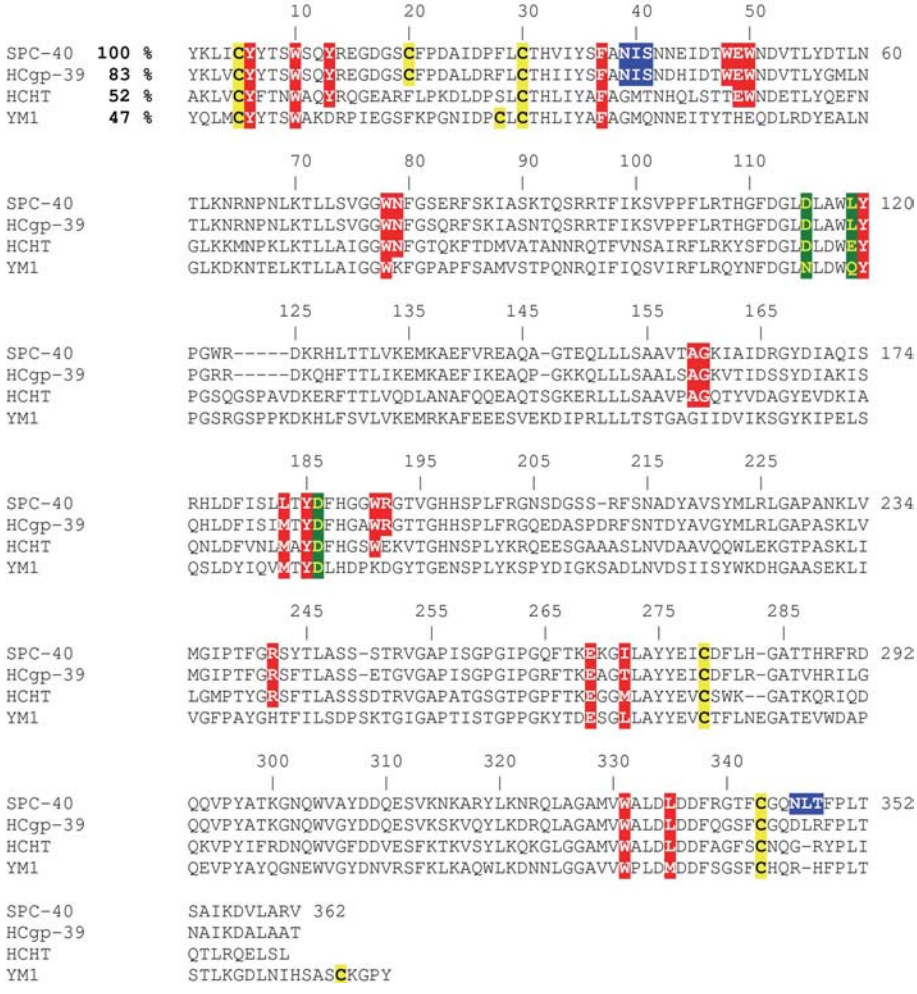

Figure 3
Multiple sequence alignment involving SPC-40, HCgp-39, HCHT and YM1. Catalytic residues (Asp115, Leu119 and Asp186) are shaded in green and residues that make the walls of the barrel and are attributed as being involved in sugar binding are shaded in red. Cysteine residues are shaded yellow and N-linked glycosylation sites are shown in blue.

3.2. Fluorescence analysis of protein–carbohydrate binding

Binding studies of various sugars with SPC-40 have shown that the emission spectrum of SPC-40 has a maximum at 330 nm with an excitation at 295 nm, which is characteristic of a tryptophan residue located in a hydrophobic environment. The shift in emission maximum and changes in the emission intensity on titration with ligands are indicative of the binding/stacking of ligands against a tryptophan residue (Tabary & Frenoy, 1985; Eftink, 1997; Boraston *et al.*, 2000). None of the monosaccharides, disaccharides, trisaccharides and tetrasaccharides caused observable shifts in the emission maximum of 330 nm (Fig. 4*a*) or significant changes in the fluorescence intensity at various concentrations (Fig. 4*b*). This may be attributed to the poor stacking of sugar molecules at the subsites where these short carbohydrate molecules bind to the sugar-binding groove in SPC-40. However, binding studies with $(GlcNAc)_5$ and $(GlcNAc)_6$ showed concentration-dependent increases in the fluorescence intensity, indicating notable bindings of these sugars to protein (Fig. 4*a*), although no shift was observed in the emission maximum. The data showed saturable binding for chitopentaose and chitohexaose. Assuming binary interactions, dissociation constants of 20 $\pm$ 2 μM for $(GlcNAc)_6$ and 271 $\pm$ 2 μM for $(GlcNAc)_5$ were obtained (Fig. 4*b*).

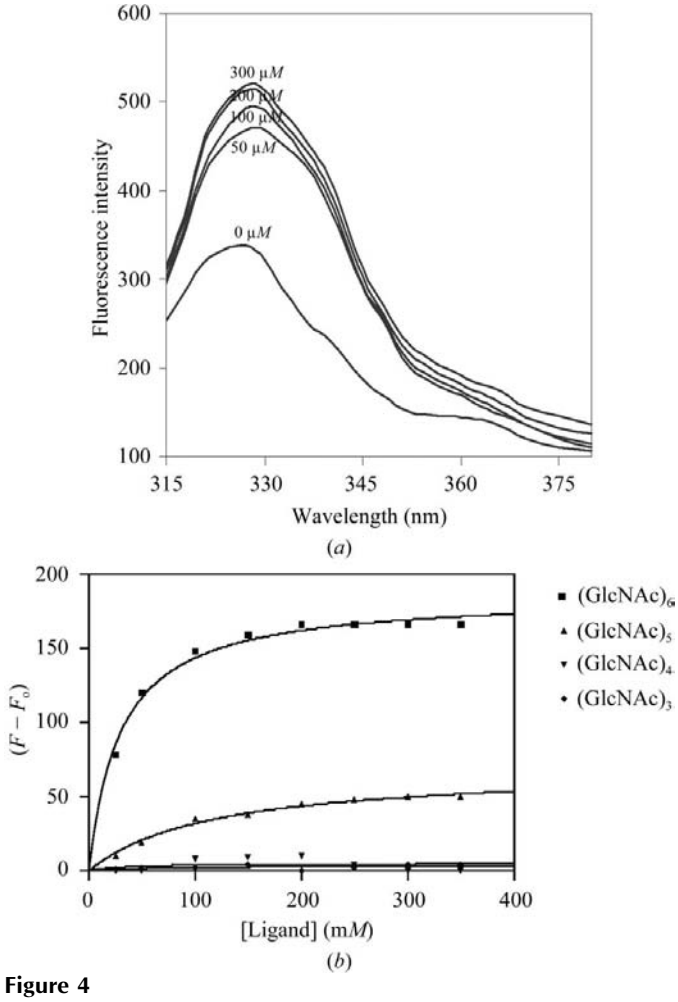

Figure 4
(*a*) Effects of $(GlcNAc)_6$ on the intrinsic fluorescence of SPC-40. Increasing amounts of $(GlcNAc)_6$ were added to 1 μM SPC-40 in 25 mM Tris–HCl pH 8.0 and the emission spectra were recorded from 315 to 380 nm upon excitation at 295 nm. No shift in λ_{max} from 330 nm was observed. (*b*) The binding curves determined for $(GlcNAc)_6$, $(GlcNAc)_5$, $(GlcNAc)_4$ and $(GlcNAc)_3$. Relative fluorescence intensity data $(F - F_o)$ at 330 nm were fitted to a binary interaction model (see §2). The curves correspond to $(GlcNAc)_6$ (squares), $(GlcNAc)_5$ (triangles), $(GlcNAc)_4$ (inverted triangles) and $(GlcNAc)_3$ (diamonds).

3.3. Overall structure

The amino-acid sequence determination of SPC-40 shows that it consists of a single polypeptide chain of 361 amino-acid residues (GenBank accession No. AY291312). The crystal structure of SPC-40 has been determined using crystallographic methods and refined using data to 2.1 Å resolution. Structural evaluation of the final model of the protein using *PROCHECK* (Laskowski *et al.*, 1993) indicate that 91.7% of the residues are in the most favoured regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968). The refined model includes all 361 residues, six glycan (two GlcNAc and four Man) residues and 247 water molecules, yielding R_{cryst} and R_{free} values of 0.173 and 0.220, respectively. The r.m.s. deviations in bond lengths and angles are 0.01 Å and 2.0°, respectively. The overall folding of the protein chain is shown in Fig. 5. The structure is broadly divided into two globular domains: a large $(\beta/\alpha)_8$ topoisomerase (TIM) barrel (Banner *et al.*, 1975) domain and a small $(\alpha+\beta)$ domain. The TIM-barrel domain contains both the N- and C-termini and is made up of two polypeptide segments, 1–237 and 310–360 (Fig. 5). The polypeptide chain crosses over to form a small domain consisting of residues 240–307 (Fig. 5) containing $\beta1'$ (240–246), $\beta2'$ (272–274), $\alpha1'$ (275–281), $\beta3'$ (286–289), $\beta4'$ (296–300) and $\beta5'$ (303–307). The chain returns to merge with the TIM-barrel domain for residues 310–360. The

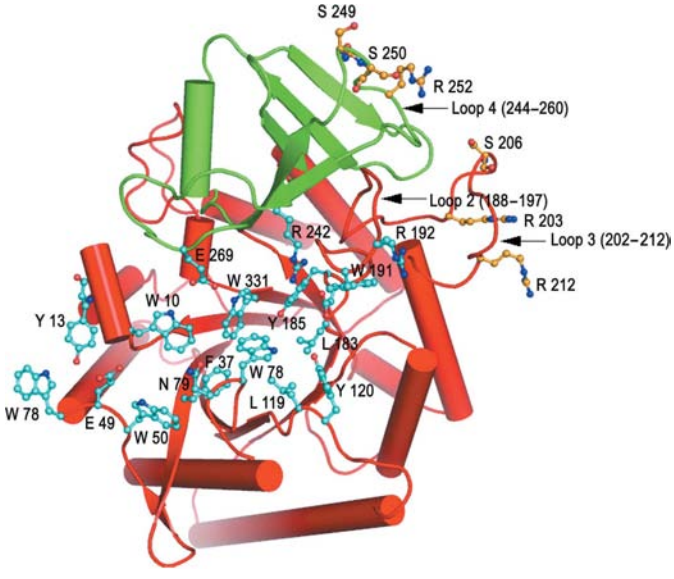

Figure 5
Ribbon diagrams (DeLano, 2002) of SPC-40. (*a*) Top-view orientation showing residues involved in carbohydrate binding (cyan) and some of those assumed to be important in receptor recognition (gold).

Table 2
Organization of secondary-structure elements in SPC-40.

TIM: triose phosphate isomerase barrel β/α-domain (residues 2–237 and 310–360); SD, small $\alpha+\beta$ domain (residues 240–307).

Secondary structure	Residues	Domain
β1	2–8	TIM
α1-1	9–12	TIM
α1-2	16–18	TIM
α1-3	22–24	TIM
β2-1	32–41	TIM
β2-2	44–46	TIM
α2	52–60	TIM
β3	70–76	TIM
α3-1	82–90	TIM
α3-2	92–108	TIM
β4	113–117	TIM
α4	126–146	TIM
β5	152–157	TIM
α5-1	161–167	TIM
α5-2	170–174	TIM
β6	180–182	TIM
α6-1	216–225	TIM
α6-2	230–232	TIM
β7	233–237	TIM
$\beta1'$	240–246	SD
$\beta2'$	272–274	SD
$\alpha1'$	275–281	SD
$\beta3'$	286–289	SD
$\beta4'$	296–300	SD
$\beta5'$	303–307	SD
α7	310–322	TIM
β8	327–330	TIM
α8-1	333–335	TIM
α8-2	350–360	TIM

secondary-structure elements are listed in Table 2. The eight-stranded parallel β-sheet structure forms the core of the protein molecule, while eight pieces of α-helix surround it, covering at least three-quarters of the barrel from outside. The interior of the barrel is filled predominantly with hydrophobic residues. There are two optimally formed disulfide bonds between residues 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 SPC-40 which is located in a tightly packed hydrophobic pocket containing residues Tyr7, Ala24, Ile25, Phe338 and Phe349. Hence, it is practically inaccessible to solvent molecules. The functional significance of Cys20 is not clear. Three *cis*-peptide bonds (residues Ser36–Phe37, Leu119–Tyr120 and Trp331–Ala332) have been observed in the structure. Notably, two of them, Leu119–Tyr120 and Trp331–Ala332, are located in the saccharide-binding cleft. *Cis*-peptide bonds have also been reported at corresponding positions in the structures of other members of the superfamily (Sun *et al.*, 2001; Fusetti *et al.*, 2002, 2003; Houston *et al.*, 2003; Tsai *et al.*, 2004). The amino-acid sequence of SPC-40 has revealed two potential N-glycosylation sites with Asn-*X*-Ser/Thr sequence motifs; these are at Asn39 and Asn345. Our crystallographic analysis of SPC-40 shows an attachment only at Asn39, with remarkably good-quality electron density for six glycan residues (Fig. 2). It contains a mannose-rich hexasaccharide with two GlcNAc and four Man residues (Fig. 2). In contrast, only a disaccharide of two GlcNAc residues was observed in HCgp-39 (Houston *et al.*, 2003; Fusetti *et al.*, 2003). Two prominent protein–carbohydrate hydrogen bonds are formed between Ser41 and the first GlcNAc residue in addition to van der Waals contacts with the glycan chain (Fig. 2). The hydrogen-bonding interaction observed in SPC-40 is similar to that observed in HCgp-39. It is noteworthy that a second potential glycosylation site at Asn345 is present only in SPC-40 among all the members of group I proteins, although no glycosylation was observed in the structure.

3.4. Comparison with other mammalian glycoproteins of the superfamily

Although the overall polypeptide fold of SPC-40 is essentially similar to the structure of MGP-40 reported previously, with a root-mean-square (r.m.s.) shift between the positions of C^{α} atoms of SPC-40 and MGP-40 of 0.54 Å (Mohanty *et al.*, 2003), the higher resolution of the current electron-density map has revealed several new structural features that were not reported in the previous low-resolution MGP-40 structure and has allowed more detailed characterization of a number of structural differences, particularly those involving the carbohydrate-binding groove and the region consisting of three flexible loops. These two aspects and other structural features were not discussed in the structure of MGP-40 (Mohanty *et al.*, 2003). The glycan chain in the present structure is observed with six residues compared with only two residues in the previous structure. Similarly, in another structure of a member of group I, HCgp-39 (Houston *et al.*, 2003; Fusetti *et al.*, 2003), only a disaccharide of two GlcNAc residues was observed. It is pertinent to note that the two structure reports of HCgp-39 are based on the protein obtained from (i) human chondrocytes (Houston *et al.*, 2003) and (ii) recombinant HCgp-39 as produced by a CHO transfectant (Fusetti *et al.*, 2003). Both these structures exist either as dimers (Fusetti *et al.*, 2003) or as tetramers (Houston *et al.*, 2003; Fusetti *et al.*, 2003), whereas SPC-40 repeatedly crystallizes as a monomer (PDB codes 1hjv, 1hjx, 1nwr and 2esc). In one of the HCgp-39 structures (Fusetti *et al.*, 2003), the most critical interaction that seems to be responsible for dimerization is formed between Lys148 N^{ζ} and Thr108 O. In SPC-40, the residue corresponding to Lys148 is Thr148 and hence is unable to form a similar interaction. In another structure of HCgp-39 (Houston *et al.*, 2003), the dimerization involves a new feature with three surface loops, 188–197, 202–212 and 244–260, that are highly flexible in SPC-40 and hence do not seem to support the dimerization. Although SPC-40 and HCgp-39 are closely related by sequence similarity, they display striking structural and functional differences. It may be mentioned here that SPC-40 is secreted by bovine mammary glands, while HCgp-39 is present in human chondrocytes (Houston *et al.*, 2003). Both native structures of HCgp-39, one of which was purified from human chondrocytes (Houston *et al.*, 2003) and the other of which was obtained from CHO transfectant as recombinant protein (Fusetti *et al.*, 2003), were identical, with r.m.s. shifts of 0.8 Å for their C^{α} traces. Since both structures are similar, the structure of the recombinant protein (Fusetti *et al.*, 2003)

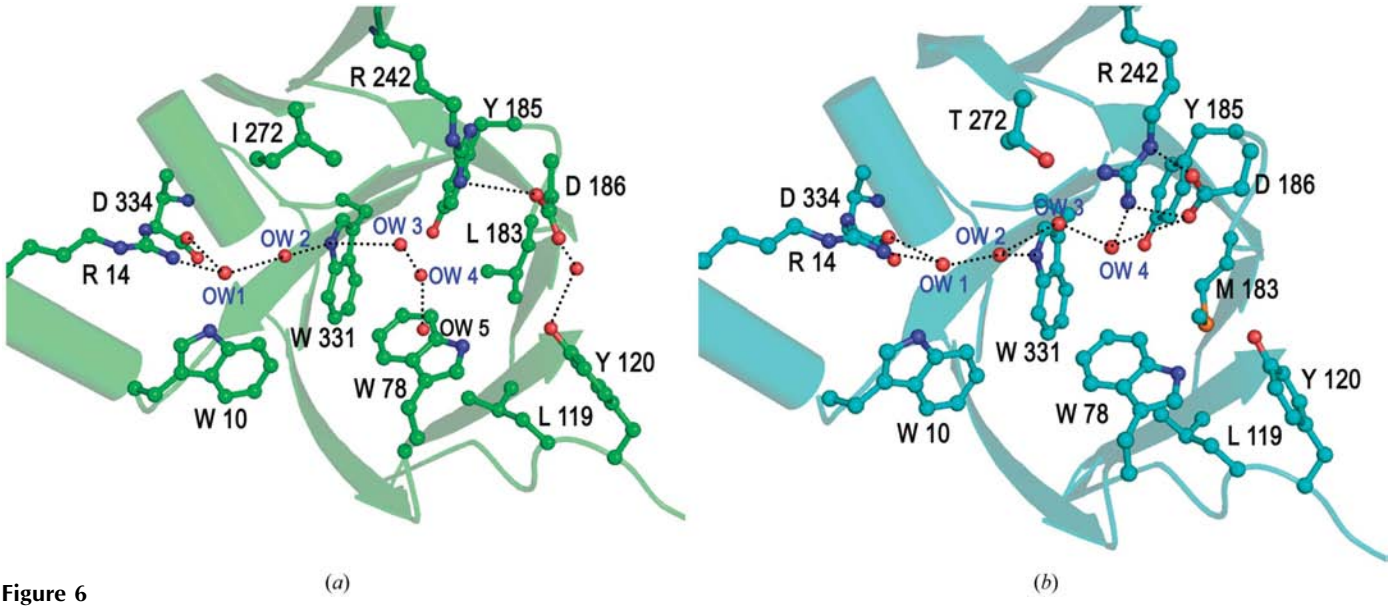

Figure 6
The relative environments near the β-barrel and Trp78 are shown for (*a*) SPC-40 and (*b*) HCgp-39. The important interactions involving residues Asp186, Tyr120, Leu183 (Met183 in HCgp-39), Arg242, Ile272 (Thr272 in HCgp-39) as well as solvent molecules are indicated.

which was reported later will be used in the subsequent discussion. The numbering scheme of SPC-40 will be used in the discussions. The least-squares superimposition of the C^{α} trace of SPC-40 on that of HCgp-39 (Fusetti *et al.*, 2003) indicates large-scale variations in the conformations of four segments consisting of residues 141–150 (loop 1), 188–197 (loop 2), 202–212 (loop 3) and 244–260 (loop 4). The average r.m.s. shift for the C^{α} atoms of these segments is 1.7 Å. Similarly, the values of the *B* factors for these loops in SPC-40 are also considerably higher (∼70.0 Å^2) than those found in HCgp-39 (∼25.0 Å^2). It is notable that the single flexible loop 1 (141–150) is at one end of the molecule, while the region consisting of three flexible loops 2, 3 and 4 (188–197, 202–212 and 244–260) is located at the other extremity of the molecule. The loop 1 in the HCgp-39 structure is found at the interface of the dimer and is involved in intermolecular interactions between the two monomers of HCgp-39. As a result, the conformation of loop 1 in HCgp-39 is well defined, while the corresponding loop in SPC-40 is significantly disordered and hence differs considerably from that of HCgp-39. The other three flexible loops 2, 3 and 4 constitute a single region that lies in close proximity of the $\alpha+\beta$ domain. Inspite of being placed closely, these loops show poor inter-loop interactions. On the other hand, several strong interactions have been observed involving the corresponding loops in HCgp-39. As a result, the three loops in HCgp-39 are packed tightly and are well ordered. The distance between the closest C^{α} positions of loops 3 and 4 in SPC-40 is 10.2 Å, while the corresponding separation in HCgp-39 is only 6.5 Å. Similarly, the shortest distance between loops 2 and 4 is 5.8 Å in SPC-40, whereas it is 4.7 Å in HCgp-39. A strong electrostatic interaction is observed between the side chains of Arg192 (loop 2) and Asp211 (loop 3) in HCgp-39 which is absent in SPC-40 because the residue corresponding to Asp211 (loop 3) in

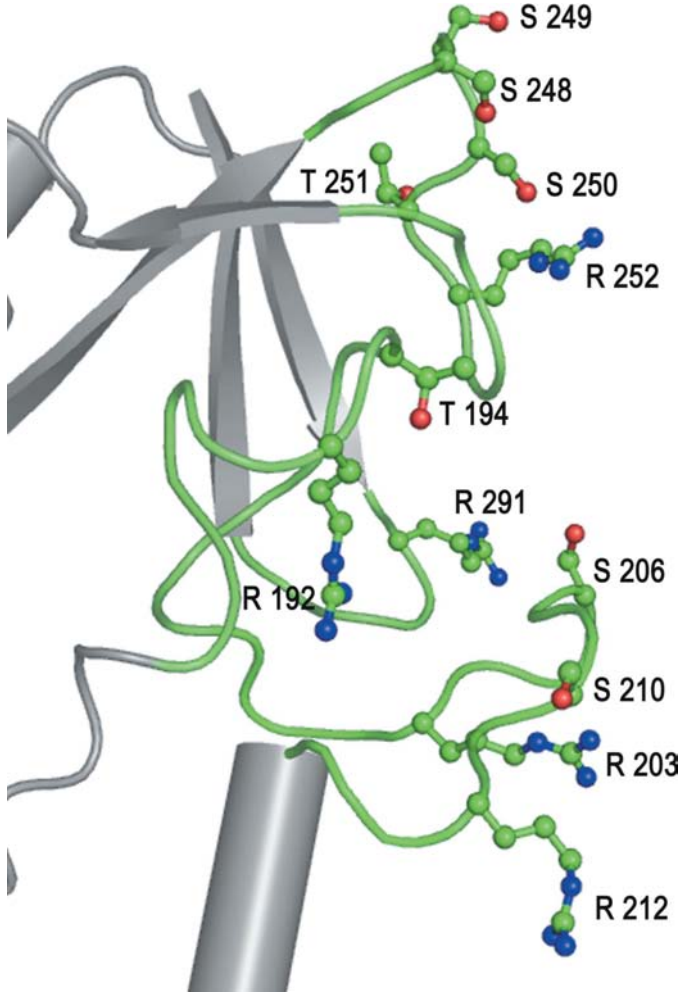

Figure 7
The residues (DeLano, 2002) of three flexible loops consisting of several serines and arginines are seen protruding outward.

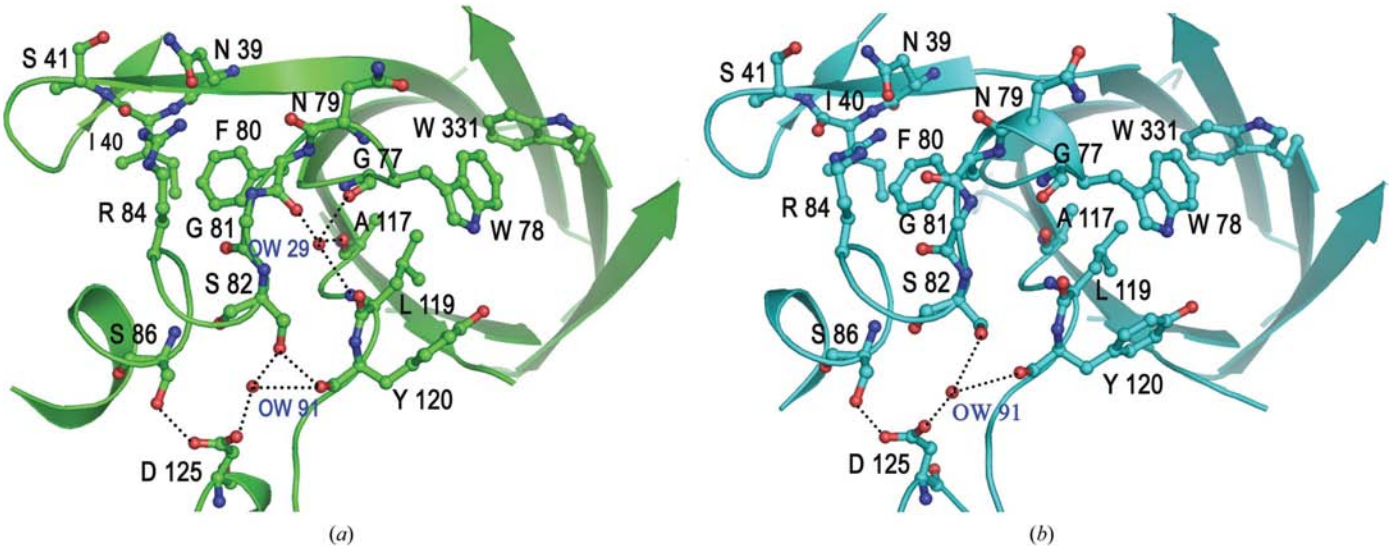

Figure 8
(*a*) The positions of water molecules OW29 and OW91 and their interactions with protein atoms and other hydrogen bonds in SPC-40. (*b*) The corresponding environment in HCgp-39. The conformations of Phe80-Gly81-Ser82 are different.

SPC-40 is absent (residue 211 is deleted in SPC-40). The distance between His188 $N^{\varepsilon 2}$ and Thr194 O within loop 2 is 2.54 Å in HCgp-39, while such an interaction is absent in SPC-40 owing to the very different conformations of the corresponding residues. In SPC-40, Asn205 $N^{\delta 2}$ (loop 3) forms a hydrogen bond with Asp292 $O^{\delta 2}$ at a distance of 2.62 Å, whereas this interaction is absent in HCgp-39 because the residue corresponding to Asp292 is replaced by Gly in HCgp-39. On the other hand, Arg203 O (loop 3) forms a hydrogen bond with Gln293 $N^{\varepsilon 2}$ (3.02 Å) in HCgp-39, while such an interaction is absent in SPC-40 owing to unfavorable orientations of both Arg203 and Gln293. Other critical interactions that differ in SPC-40 and HCgp-39 pertain to the involvement of Arg242 (loop 4): Arg242 NH1 in SPC-40 forms a hydrogen bond with Tyr185 O, while Arg242 NH2 interacts with Asp186 $O^{\delta 1}$ (Fig. 6*a*). In HCgp-39 Arg242 NH1 forms a hydrogen bond with Thr272 O^{γ}, while Arg242 NH2 interacts with Asp186 $O^{\delta 1}$. Thr272 O^{γ} in HCgp-39 forms yet another hydrogen bond to Glu269 $O^{\varepsilon 1}$ (Fig. 6*b*). The residue corresponding to Thr272 is Ile272 in SPC-40 and hence this interaction is absent in SPC-40. Phe290 is involved in extensive hydrophobic interactions in SPC-40, whereas the corresponding residue in HCgp-39 is Ile290 and interacts only poorly with its hydrophobic environment. The organization and flexibility of loops 188–197, 202–212 and 244–260 in SPC-40 allow the side chains of a number of serine and arginine residues to protrude outward, resulting in a strongly positively charged environment (Fig. 7). The combination of such a charged environment with highly flexible surface loops suggests it to be a potential receptor-binding site. The overall folding of this region together with the $\alpha+\beta$ domain represents a similar arrangement to that of the FK-binding protein (Itoh & Navia, 1995), which has been indicated as recognizing the type I TGFβ receptor (Huse *et al.*, 1999). In contrast, in HCgp-39 as well as in other proteins of the SPX-40 superfamily, the corresponding regions are well ordered with extensive intra-region interactions.

3.5. Carbohydrate recognition

It may be mentioned that chitin polymers bind to chitinase in a well formed carbohydrate-binding groove. The carbohydrate-binding groove in chitinase has a regular shape with a wide opening to allow easy diffusion of chitin polymers. Similarly, a well formed carbohydrate-binding groove is also observed in HCgp-39. In contrast, in YM1 the groove has an irregular shape and may be too narrow at certain sites to allow the diffusion of chitin-like polymers, while in SPC-40 it seems to have a visible obstruction midway, with a slight widening just before the obstruction site. The surface potentials indicate that YM1 is clearly an acidic protein (pI 5.3), while chitinase (pI 6.4) is slightly less acidic. In contrast, HCgp-39 (pI 8.3) and SPC-40 (pI 8.9) are basic in nature. Although the residues that form the inner walls of the carbohydrate-binding groove are generally conserved, there are important variations in the sequences that appear to alter the carbohydrate-accommodating capacities and binding capabilities of these proteins. In native SPC-40, Trp78 is located inside the barrel, thus reducing the inner space drastically. The corresponding Trp residue in native HCgp-39 also occupies a similar position, although less firmly than that in SPC-40. As seen from Fig. 8(*a*), an extra water molecule OW29 has been observed in the proximity of the loop containing Trp78. OW29 is hydrogen bonded to Gly77 O, Phe80 O, Ala117 O and Leu119 N. The water molecule corresponding to OW29 is missing in the structures of HCgp-39 (Houston *et al.*, 2003; Fusetti *et al.*, 2003). There is yet another important interaction in SPC-40 where Ser82 O^{γ} forms a strong hydrogen bond with Tyr120 O with a distance of 2.82 Å. The interaction corresponding to this is absent in HCgp-39 (Houston *et al.*, 2003; Fusetti *et al.*,

2003). Furthermore, another water molecule OW91 forms an extensive hydrogen-bonded network with Ser82 O, Ser82 O^{γ}, Tyr120 O and Asp125 $O^{\delta2}$. Although a water molecule corresponding to OW91 in SPC-40 is present in HCgp-39 (Fusetti *et al.*, 2003), it is only involved in two weak hydrogen bonds. The hydrogen-bonded network including water molecules OW29 and OW91 induce different conformations in the motif Phe80-Gly81-Ser82 in which the carbonyl O atom of Phe80 is hydrogen bonded to OW29 (Fig. 8*a*), while in HCgp-39 it is hydrogen bonded to Arg84 N^{ε} (Fig. 8*b*). The (φ, ψ) torsion angles of residues Phe80, Gly81 and Ser82 in SPC-40 are (−91.1, 129.0°), (−64.0, 119.3°) and (−48.2, −38.3°), respectively, while the corresponding values in HCgp-39 are (−80.0, −42.2°), (90.6, 96.5°) and (−45.7, −179.0°), respectively (Fig. 8*b*). In addition, in HCgp-39 Arg84 interacts firmly with Asn39 $O^{\delta2}$ and Ile40 O, while in SPC-40 it forms a hydrogen bond with Ile40 only. In a striking contrast to these observations, the corresponding Trp residues in both chitinases (HCHT) and chitinase-like proteins (YM1) are observed at an position distant from the barrel, indicating a large-scale conformational variation compared with those observed in group I proteins. In yet another critical difference, Asp186 in SPC-40 occupies a position with its side chain oriented towards the carbohydrate-binding groove and results in an interaction with Tyr120 through a conserved water molecule (Asp186 $O^{\delta1}\cdots$OW$\cdots$Tyr120 OH; Fig. 6*a*). In HCgp-39, the Asp186 side chain is oriented away from the carbohydrate-binding groove and hence it does not interact with Tyr120. Instead, it forms a hydrogen bond with a water molecule which is part of an ordered water network in the groove. The second oxygen $O^{\delta1}$ of Asp186 in HCgp-39 interacts with NH2 and N^{ε} of Arg242 (Fig. 6*b*), while Asp186 $O^{\delta1}$ in SPC-40 interacts only with NH2. Furthermore, the presence of Leu183 in SPC-40 does not hinder the interaction between Asp186 and Tyr120, while Leu183 is replaced by Met183 in HCgp-39 and seems to obstruct the solvent structure in the vicinity of Tyr120. Moving further along the wall of the carbohydrate-binding groove towards the direction of the carbohydrate entry point, Arg242 NH2 in HCgp-39 forms a hydrogen bond with Thr272 O^{γ}, generating a local hydrogen-bonded network. In contrast, in SPC-40 Thr272 is replaced by Ile272, eliminating the possibility of such an interaction. The absence of a critical interaction between Arg242 and Thr272 causes a void in SPC-40 as the side chains of both Ile272 and Arg242 are oriented in the opposite directions. This changes the shape of the carbohydrate-binding groove in SPC-40. In YM1, apart from the loss of several aromatic residues (Fig. 3) that provide an affinity for the binding of carbohydrates, there is an important mutation that reduces the width of the groove right at the entry point. The residue at position 79 is generally conserved in the proteins of the SPX-40 superfamily as aspargine, but in YM1 it is replaced by Lys79, which interacts with Glu269 through a water molecule. The closest distance between the side-chain atoms of Lys79 and Glu269 is 6.8 Å, whereas the corresponding distance in other structures that contain Asn79 is approximately 9.0 Å. It is important to mention that the orientation of the side chain of residue at position 79 is fixed as the neighbouring Trp78 tends to occupy two alternate positions. Therefore, an increase in the length of a residue at 79 is bound to protrude directly into the groove, thus reducing the width of the carbohydrate-binding groove considerably.

4. Conclusions

The structural studies of proteins of the SPX-40 superfamily reported to date (Sun *et al.*, 2001; Fusetti *et al.*, 2002, 2003; Mohanty *et al.*, 2003; Houston *et al.*, 2003; Tsai *et al.*, 2004) indicate several similar features in their overall structures, including the formation of a TIM barrel, an $\alpha+\beta$ domain and two identical disulfide bridges. On the other hand, several important differences have also been observed pertaining to the nature of glycosylation, the shape of the sugar-binding groove, the nature of the residues that are lined up along the walls of the groove and the structures of some important regions of these proteins. Other than chitinases, all these proteins are catalytically inactive; thus, the role of the carbohydrate-binding groove seems to have been altered. This is clearly evident from the differences in the shapes of carbohydrate-binding grooves. The different stereochemical features of these grooves suggest that they might be involved in the recognition of different types of oligosaccharides/glycans. Yet another unique feature in the structure of SPC-40 pertains to the flexible region consisting of three loops: 188–197, 202–212 and 244–260. This part of the protein has high temperature factors with a series of serine and arginine residues protruding outward from the surface of the protein. This indicates a strong potential for this site to form intermolecular interactions, presumably with receptors involving protein–protein recognition in addition to sugar binding. This possibility reflects an evolutionary progression from purely sugar binding as in chitinases to protein- and sugar-binding capacity as in the present molecule. However, the proof of this hypothesis must await further biochemical and structural studies to define the ligand-binding specificity.

The authors thank the Department of Science and Technology (DST), New Delhi for financial support. The liberal financial support from the Department of Biotechnology (DBT), New Delhi for establishing a proteomics facility is also acknowledged. The DST is also thanked for the support under FIST programme for the level II grant. JK, ASE and DBS 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.

Boraston, A. B., Tomme, P., Amandoron, E. A. & Kilburn, D. G. (2000). *Biochem. J.* **350**, 933–941.

Chomczynski, P. & Sacchi, N. (1987). *Anal. Biochem.* **162**, 156–159.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Coppock, C. E., Everett, R. W., Natzbe, R. P. & Ainsle, H. R. (1974). *J. Dairy Sci.* **57**, 712–718.
DeLano, W. L. (2002). *The PyMOL User's Manual.* DeLano Scientific, San Carlos, CA, USA.
Eftink, M. R. (1997). *Methods Enzymol.* **278**, 221–257.
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, D. R., 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., Massaque, J. & Kuriyan, J. (1999). *Cell*, **96**, 425–436.
Itoh, S. & Navia, M. A. (1995). *Protein Sci.* **4**, 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.
Laskowski, R., MacArthur, M., 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. & 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.
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.
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.
Tabary, F. & Frenoy, J. P. (1985). *Biochem. J.* **229**, 687–692.
Tsai, M. L., Liaw, S. H. & Chang, N. C. (2004). *J. Struct. Biol.* **148**, 290–296.