Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Carbohydrate-binding properties of goat secretory glycoprotein (SPG-40) and its functional implications: structures of the native glycoprotein and its four complexes with chitin-like oligosaccharides

Janesh Kumar, Abdul S. Ethayathulla, Devendra B. Srivastava, Nagendra Singh, Sujata Sharma, Punit Kaur, Alagiri Srinivasan and Tej P. Singh*

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

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

Received 29 November 2006
Accepted 12 January 2007

PDB References: SPG-40, 2dsz, r2dszsf; SPG-40–$GlcNAc_3$, 2dt0, r2dt0sf; SPG-40–$GlcNAc_4$, 2dt1, r2dt1sf; SPG-40–$GlcNAc_5$, 2dt2, r2dt2sf; SPG-40–$GlcNAc_6$, 2dt3, r2dt3sf.

A 40 kDa glycoprotein (SPG-40) secreted during involution works as a protective signalling factor through its binding to viable cells. The crystal structure of the native protein has been determined at 2.3 Å resolution. This is the first report on the carbohydrate-binding properties of SPG-40; the structure determinations of the complexes of SPG-40 with four oligosaccharides of different lengths at resolutions ranging from 2.2 to 2.8 Å are described. Carbohydrate-binding studies with *N*-acetylglucosamines ($GlcNAc_n$, n = 3–6) using fluorescence spectroscopy revealed poor binding effects with $GlcNAc_3$ and $GlcNAc_4$, while $GlcNAc_5$ and $GlcNAc_6$ bound to SPG-40 with considerable strength; the dissociation constants (K_d) were estimated to be 260 ± 3 and 18 ± 4 μ*M*, respectively. SPG-40 was cocrystallized with $GlcNAc_3$, $GlcNAc_4$, $GlcNAc_5$ and $GlcNAc_6$. The overall structure of native SPG-40 was essentially similar to that reported previously at low resolution. The structures of its complexes with $GlcNAc_3$, $GlcNAc_4$, $GlcNAc_5$ and $GlcNAc_6$ revealed the positions of these oligosaccharides in the carbohydrate-binding groove and provided insights into the mechanism of binding of oligosaccharides to SPG-40, indicating that the preferred subsites in the carbohydrate-binding groove of SPG-40 were from −4 to −2. The structure of the protein remained unperturbed upon binding of $GlcNAc_3$ and $GlcNAc_4$, but the structure changed significantly upon binding of $GlcNAc_5$ and $GlcNAc_6$. Significant conformational variations were observed in the sugar-binding groove: Trp78 partially flipped out of the barrel in $GlcNAc_5$, while in the $GlcNAc_6$ complex a completely flipped-out Trp78 was observed along with several other conformational changes, including those of Asp186 and Arg242. Such changes upon binding to carbohydrates have not previously been observed in chitin-hydrolyzing chitinases and reflect less favourable binding of carbohydrates to SPG-40. As this appears to essentially be a binding protein, this loss of binding affinity might be compensated by other intermolecular interactions such as protein–protein interactions and also by the binding of its own glycan chain.

1. Introduction

Tissue homeostasis takes place as a result of coordinated regulation of the proliferation and elimination of cells. Developmentally regulated removal of cells is mostly controlled by programmed cell death (PCD; Strange *et al.*, 1992). Pregnancy induces a massive development of mammary alveolar structures, which at birth embody the differentiated secretory epithelium required for milk production. After the

cessation of lactation, a collapse of the lobulo-alveolar structures occurs, which is paralleled by a reductive remodelling of the gland. This process is termed involution and is characterized by a proteolytic degradation of the extracellular matrix and a loss of secretory epithelial cells, mainly by programmed cell death (Strange *et al.*, 1992). A 40 kDa glycoprotein is expressed in high concentrations during this period of involution, which is similar to a prominent protein in the whey secretions of nonlactating cows that has been reported to be an important marker protein for mammary function during involution (Rejman & Hurley, 1988; Kumar *et al.*, 2006). A similar 40 kDa glycoprotein has been isolated from the dry secretions of goat and is named SPG-40. Although its precise function is not known, it seems to act as a protective signalling factor during extensive tissue remodelling. SPG-40 is homologous (52% sequence identity) to the human macrophage chitinase (HCHT; Renkema *et al.*, 1995; Fusetti *et al.*, 2002) as well as to another mammalian chitinase-like protein (YM1; 47% sequence identity; Sun *et al.*, 2001; Tsai *et al.*, 2004). Several other animal glycoproteins, including a 39 kDa human cartilage glycoprotein (HCGP-39; Johansen *et al.*, 1993; Hakala *et al.*, 1993), bovine chondrocyte chitinase-like protein (CLP-1; GenBank accession No. AF011373), porcine heparin-binding glycoprotein (GP38k; Shackelton *et al.*, 1995), rat cartilage glycoprotein RCGP39 (GenBank accession No. AF062038) and a breast regression protein (BRP39; Morrison & Leder, 1994), are homologous to SPG-40 (identities varying from 93 to 69%) and constitute a single homologous family referred to in the following as SPX-40 proteins. To date, the crystal structures of only five native proteins from different sources, MGP-40 (PDB code 1ljy; Mohanty *et al.*, 2003), HCGP-39 (PDB code 1hjx, Houston *et al.*, 2003; PDB code 1nwr, Fusetti *et al.*, 2003), SPC-40 (PDB code 2esc; Kumar *et al.*, 2006) and SPS-40 (PDB code 2dpe, Srivastava *et al.*, 2006), as well as some complexes with carbohydrates such as those of HCGP-39 with various chitin-like oligosaccharide fragments (Houston *et al.*, 2003; Fusetti *et al.*, 2003), have been reported. The five native structures have been found to share an essentially similar folding, although the structures of the sugar-binding grooves differ appreciably. The complexes of HCGP-39 with sugars show that binding of both long and short oligosaccharides is possible, but their subsite preferences vary inconsistently (Houston *et al.*, 2003; Fusetti *et al.*, 2003). In sets of complexes with isolated (Houston *et al.*, 2003) and cloned (Fusetti *et al.*, 2003) HCGP-39, the sugar residues occupy different subsites for oligosaccharides of same length, indicating a total lack of consistency. In order to understand the mechanism of binding of oligosaccharides to SPX-40 proteins, we have determined the crystal structures of native SPG-40 and of several of its complexes with oligosaccharides of different lengths. As the oligosaccharides bind to SPG-40 with much lowered affinity, this gives rise to further questions. How specific is the recognition of carbohydrates in SPG-40? What is the optimum size of the carbohydrate chain? Which class of carbohydrates does it prefer? How many subsites are accessible in the groove? Which of the subsites are more preferred for binding? In order to answer these questions, we determined the crystal structures of the complexes of SPG-40 with various oligosaccharides as well as its native structure at high resolution. The results of these studies throw some light on the binding preferences of various oligosaccharides and reveal the accompanying conformational changes. Compared with the structure of chitinase, the accommodation of ligands by SPG-40 appears to require significant rearrangement of a number of residues in order to open the partly blocked sugar-binding site.

2. Experimental procedures

2.1. Purification

Fresh mammary secretions were collected from goats kept 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). In order to protect the protein from degradation by enzymes whose concentrations also increase during involution (Aslam & Hurley, 1997), protease inhibitors were added (Fang & Sandholm, 1995). The samples were pooled and skimmed and 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 the mixture was stirred slowly for about 1 h with a mechanical stirrer. The gel was allowed to settle and the solution was decanted. In order to remove 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 in a column (25 × 2.5 cm) and washed with the same buffer containing 0.2 *M* NaCl, which removed other impurities. The remaining proteins were eluted with the same buffer containing 0.5 *M* NaCl. The eluted protein solution was desalted and was again passed through a CM-Sephadex C-50 column (10 × 2.5 cm) which had been pre-equilibrated with 50 m*M* Tris–HCl pH 8.0 and was eluted with a linear gradient of 0.05–0.5 *M* NaCl in the same buffer. The elution profile showed three peaks. The peak corresponding to 0.3 *M* NaCl was pooled and was concentrated using an Amicon ultra-filtration 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 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. The sequence of the first 20 amino-acid residues from the N-terminus was determined using a PPSQ20 protein sequencer (Shimadzu, Kyoto, Japan) and showed 100% identity to the previously determined N-terminal sequence of goat signalling protein (Mohanty *et al.*, 2003). The complete nucleotide sequence is already known (Mohanty *et al.*, 2003); the Genebank accession No. is AY081150.

Table 1
Data-collection and refinement statistics for native and various complexes of SPG-40.

Values in parentheses are for the highest resolution shell.

	Native	$GlcNAc_3$	$GlcNAc_4$	$GlcNAc_5$	$GlcNAc_6$
PDB code	2dsz	2dt0	2dt1	2dt2	2dt3
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)					
a	63.1	62.7	62.7	62.5	62.3
b	66.2	66.6	66.5	66.4	66.4
c	108.0	107.4	107.6	107.3	106.6
Solvent content (%)	56.3	56.7	56.4	55.4	55.4
Unique reflections	18475	16835	27202	10410	20737
Resolution range (Å)	56.0–2.35	56.0–2.45	56.0–2.09	50.0–2.89	56.0–2.28
Completeness (%)	95.4 (99.0)	98.3 (95.0)	99.4 (99.4)	99.1 (99.0)	99.5 (96.5)
R_{sym} (%)	8.0 (55.0)	9.0 (35.0)	6.0 (49.0)	13.0 (60.0)	10.0 (49.0)
Mean $I/\sigma(I)$	12.0 (2.0)	7.0 (2.0)	17.0 (3.0)	8.0 (2.0)	8.0 (2.0)
R_{cryst}/R_{free} (%)	18.8/20.2	19.2/23.4	20.1/23.1	19.2/20.7	19.6/21.5
Protein atoms	2877	2877	2877	2877	2877
Glycan-chain atoms	61	39	39	28	61
Oligosaccharides	—	42	56	70	84
Water molecules	219	159	212	112	194
R.m.s. deviations					
Bond lengths (Å)	0.01	0.02	0.02	0.01	0.01
Bond angles (°)	1.5	1.9	2.2	1.9	2.1
Mean *B* factors (Å^2)	35.3	42.8	35.1	31.6	34.5
Protein atoms	37.0	36.0	34.6	29.4	32.2
Solvent atoms	55.2	62.4	56.5	41.5	46.0
Oligosaccharides	—	50.6	60.0	46.0	48.5
Ramachandran analysis					
Favoured regions (%)	90.9	90.2	91.5	89.3	90.5
Additionally allowed regions (%)	9.1	9.8	8.5	10.7	9.5

Table 2
Binding affinities of oligosaccharides to fungal chitinase (Fukamizo *et al.*, 2001), bacterial chitinase (Aronson *et al.*, 2003), HCGP-39A (Houston *et al.*, 2003) and SPG-40.

Protein	Oligomer	Binding constant (μ*M*)
Fungal chitinase	$GlcNAc_4$	50
Bacterial chitinase	$GlcNAc_3$	22
	$GlcNAc_4$	3
	$GlcNAc_5$	0.06
	$GlcNAc_6$	0.07
HCGP-39A	$GlcNAc_4$	331
	$GlcNAc_6$	6.7
SPG-40	$GlcNAc_5$	260
	$GlcNAc_6$	18

2.2. Fluorescence studies of protein–carbohydrate binding

In order to evaluate the binding characteristics of carbohydrates to SPG-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, chitinase 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 using 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 400 μ*M*) of these ligands with 1 μ*M* SPG-40 in 25 m*M* Tris–HCl pH 8.0. Blank correction of the data was performed by titration without SPG-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 are the fluorescence intensity in the presence and absence of ligand, respectively, F_b is the maximum fluorescence signal of the SPG–ligand complex at saturation, L_o is the initial ligand concentration and K_d is the equilibrium dissociation constant.

2.3. Protein crystallization

The purified samples of SPG-40 were used for crystallization by the hanging-drop vapour-diffusion method. Oligosaccharides ($GlcNAc_3$, $GlcNAc_4$, $GlcNAc_5$ and $GlcNAc_6$) were dissolved in 25 m*M* Tris–HCl pH 7.8 containing 50 m*M* NaCl. The protein was dissolved in the above solution to a final concentration of 30 mg ml^{-1}. The final molar ratio of protein to oligosaccharide was 1:10. 10 μl drops of the above mixture were equilibrated against 50 m*M* NaCl, 25 m*M* Tris–HCl pH 7.8 containing 20%(*v*/*v*) ethanol at 298 K. Square-shaped colourless crystals with dimensions of up to 0.45 × 0.40 × 0.20 mm were obtained after one week.

2.4. X-ray intensity data collection and processing

Suitable crystals were used for data collection at 278 K using in-house facilities. 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 operated at 50 kV and 100 mA. Osmic Blue confocal optics was used to focus the Cu $K\alpha$ radiation. The data were indexed and scaled using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). The crystals belong to space group $P2_12_12_1$ (Table 1). The data for the native and the

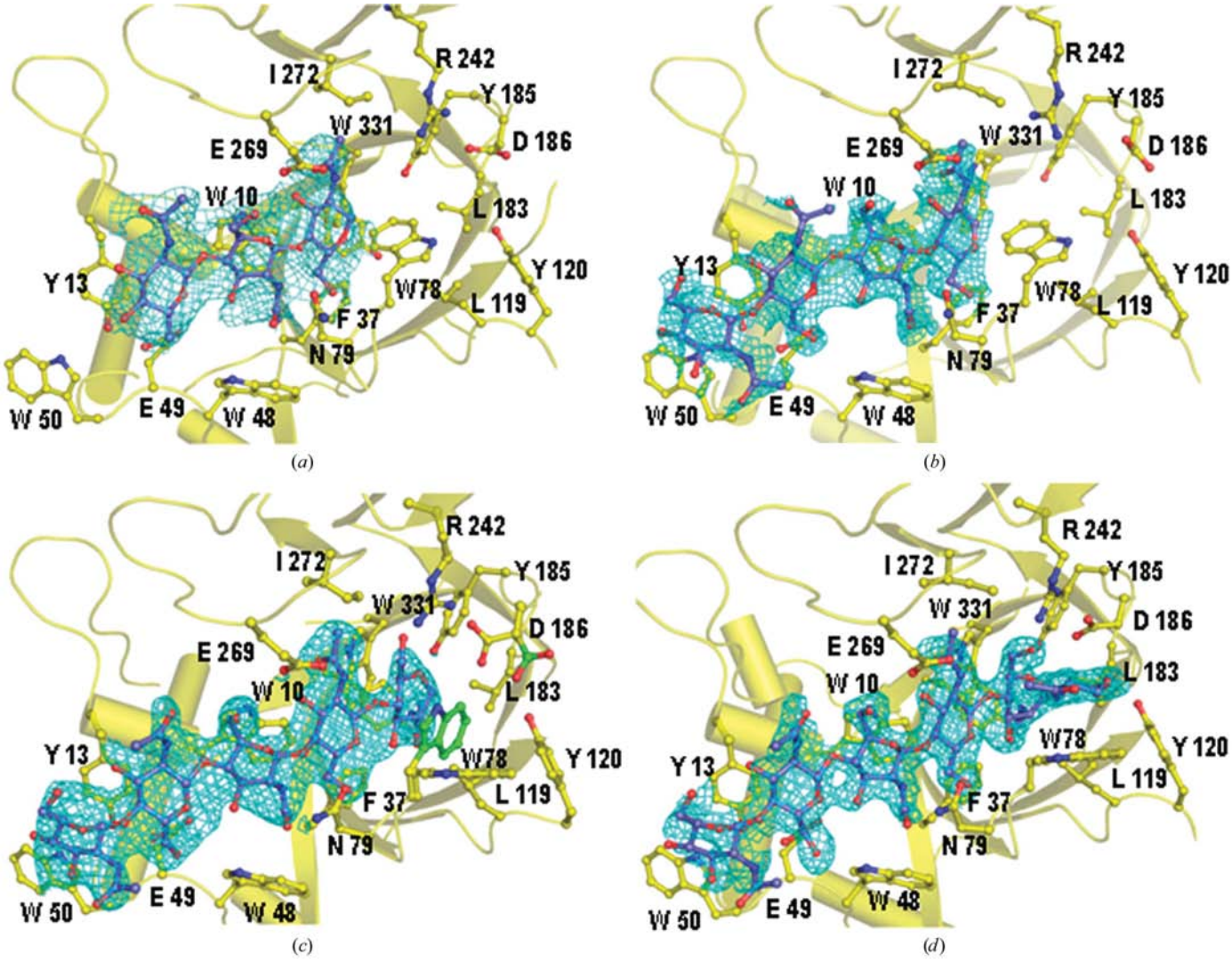

Figure 1
Difference $|F_o - F_c|$ electron densities for $GlcNAc_n$ at 2.0σ cutoff: (*a*) $GlcNAc_3$, (*b*) $GlcNAc_4$, (*c*) $GlcNAc_5$, (*d*) $GlcNAc_6$.

saccharide complexes were collected in the resolution range 2.2–2.8 Å. The unit cell contains four molecules. A summary of the data-collection statistics is given in Table 1.

2.5. Structure determination and refinement

The structures were 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 the original MGP-40 (Mohanty *et al.*, 2003; PDB code 1ljy) structure 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 *CNS* v.1.1 (Brünger *et al.*, 1998). The refinement calculations were interleaved with several rounds of model building using the program *O* (Jones *et al.*, 1991). Ligands were included only when they were well defined by unbiased $|F_o - F_c|$ maps (Fig. 1). The oligosaccharides were refined using bond-length and bond-angle parameters from idealized GlcNAc and MAN residues and glycosidic linkages (Jeffrey, 1990). Several further rounds of refinement with *CNS* v.1.1 interspersed with model building using $|2F_o - F_c|$ and $|F_o - F_c|$ maps were implemented. The positions of water molecules were determined from peak electron densities greater than 2.5σ in the $|F_o - F_c|$ maps and were only retained in the final model if they met the criteria of having peaks greater than 1.5σ in the $|2F_o - F_c|$ map, hydrogen-bond partners with appropriate distance and angle geometry and *B* values of less than 75 $Å^2$ in the final refinement cycle. The refinement statistics are summarized in Table 1.

3. Results and discussion

3.1. Fluorescence analysis of protein–carbohydrate binding

The binding studies of various sugars to SPG-40 have shown that the emission spectrum of SPG-40 has a maximum at

330 nm on 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 (Boraston *et al.*, 2000; Tabary & Frenoy, 1985; Eftink, 1997). None of the monosaccharides, disaccharides, trisaccharides and tetrasaccharides caused observable shifts in the emission maximum of 330 nm or significant changes in the fluorescence intensity at various concentrations. However, binding studies with $GlcNAc_5$ and $GlcNAc_6$ showed concentration-dependent increases in the fluorescence intensity, indicating notable binding of these sugars to protein, although no shift was observed in the emission maximum. The data showed saturable binding for chitopentaose and chitohexaose. Assuming binary interactions, dissociation constants of 18 ± 4 μM for $GlcNAc_6$ and 260 ± 3 μM for $GlcNAc_5$ were obtained (Table 2).

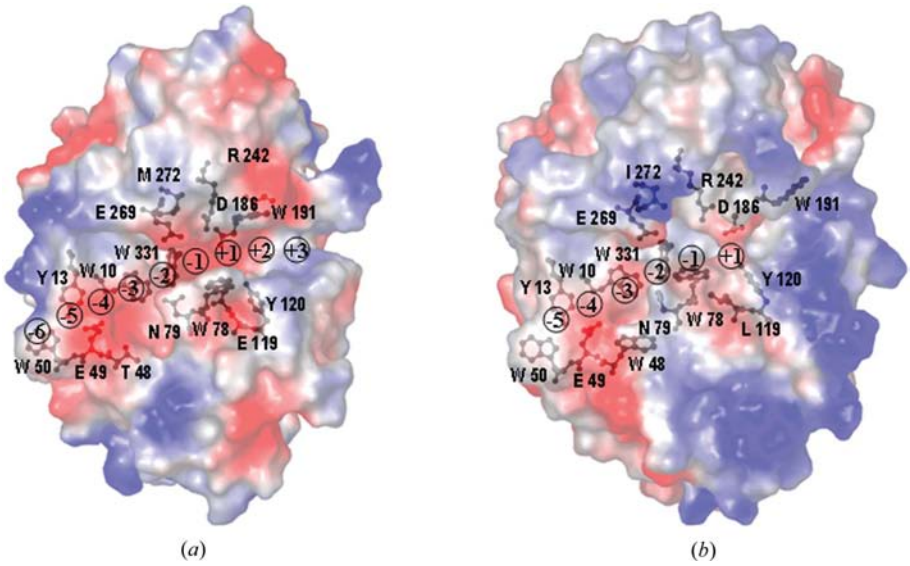

Figure 2
(*a*) *GRASP* (DeLano, 2002) view of a chitinase enzyme (Fusetti *et al.*, 2002) showing a schematic representation of the subsites in the carbohydrate-binding groove. The subsites −1 to +1 correspond to the scissile bond. A number of important residues involved in the interactions with the bound oligosaccharides are indicated in ball-and-stick representation. (*b*) *GRASP* (DeLano, 2002) drawing of SPG-40 showing subsites. The residues involved in carbohydrate binding are indicated as ball-and-stick models. SPG-40 is a far more positively charged protein. Owing to its positively charged nature as well as constraints from several aromatic residues, it does not seem to have a favourable structure beyond the +1 subsite.

3.2. Structural features

SPX-40 proteins are chemically and structurally homologous to chitinases (Fusetti *et al.*, 2002). Chitinases hydrolyze chitin, a polymer of *N*-acetylglucosamine, whereas SPX-40 proteins do not possess catalytic activity owing to the mutation of a catalytic Glu residue to Leu and a partially blocked catalytic site. Similar to proteases, the substrate-binding site in chitinases can be divided into subsites as shown in Fig. 2(*a*). In order to understand the mode of carbohydrate binding to SPG-40, the carbohydrate-binding groove was similarly divided into accessible subsites (Fig. 2*b*). As seen in Fig. 2(*a*), the carbohydrate-binding groove in chitinases can be easily divided into nine subsites across the $(\beta/\alpha)_8$ barrel. As per current convention (Davies *et al.*, 1997), they are numbered from −6 to +3, −6 being at the nonreducing end. In chitinases, the scissile glycosidic bond is observed between the −1 and +1 subsites. When compared with that of chitinases (Fusetti *et al.*, 2002; Aronson *et al.*, 2003), the carbohydrate-binding groove in SPG-40 appears to be considerably distorted and blocked. Thus, carbohydrate binding would require significant changes in the conformation of the amino acids involved in the formation of the groove. In order to determine the carbohydrate-binding properties of SPG-40 and the accompanying conformational changes, the structures of SPG-40 in its native state (structure 1) and of its complexes with $GlcNAc_3$ (structure 2), $GlcNAc_4$ (structure 3), $GlcNAc_5$ (structure 4) and $GlcNAc_6$ (structure 5) were determined. The structures were determined by molecular replacement and refined extensively. The data-collection and refinement statistics are given in Table 1. The final model comprises 2877 protein atoms from 361 residues in each of the five structures. 42, 56, 70 and 84 oligosaccharide atoms were observed in

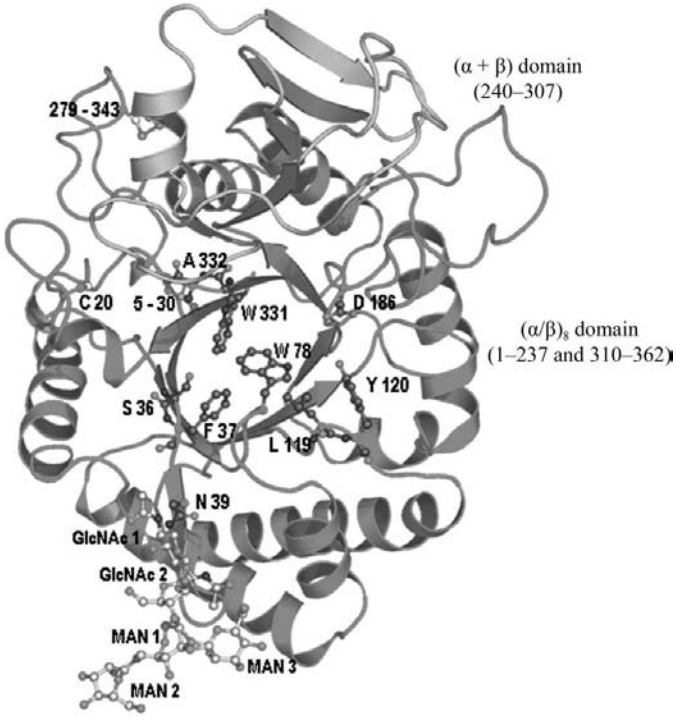

Figure 3
Overall structure of SPG-40 at 2.35 Å resolution. The $(\alpha/\beta)_8$ TIM-barrel domain and $(\alpha + \beta)$ small domain are shown. The residues involved in the formation of the *cis*-peptide and the glycan chain containing two GlcNAc residues and three MAN residues are shown. The locations of Cys residues are also indicated.

structures 2, 3, 4 and 5, respectively. The positions of 219, 159, 212, 112 and 194 water molecules were determined for structures 1, 2, 3, 4 and 5, respectively. The overall protein structure in structures 2, 3, 4 and 5 is essentially similar to the native structure, with r.m.s. (root-mean-square) deviations of 0.21, 0.20, 0.23 and 0.26 Å, respectively, on C^{α} atoms. The Ramachandran plots (Ramachandran & Sasisekharan, 1968) reveal that 90.9, 90.2, 91.5, 87.5% and 90.5% of the nonglycine and nonproline residues of structures 1, 2, 3, 4 and 5, respectively, were located in the most favoured regions of the map, while 9.1, 9.8, 8.5, 12.6 and 9.6% lie in additionally allowed regions.

The overall structures of the native protein (Fig. 3) and its complexes with $GlcNAc_n$ are shown in Fig. 4. The protein structure is broadly divided into two globular domains: a large $(\beta/\alpha)_8$ triose phosphate isomerase (TIM) barrel (Banner *et al.*, 1975) domain (1–237 and 310–362) and a small ($\alpha + \beta$) domain (residues 240–307). 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 (Fig. 3). Three *cis*-peptide bonds, Ser36-Phe37, Leu119-Tyr120 and Trp331-Ala332, have identical values of the torsion angle ω in all the five structures,

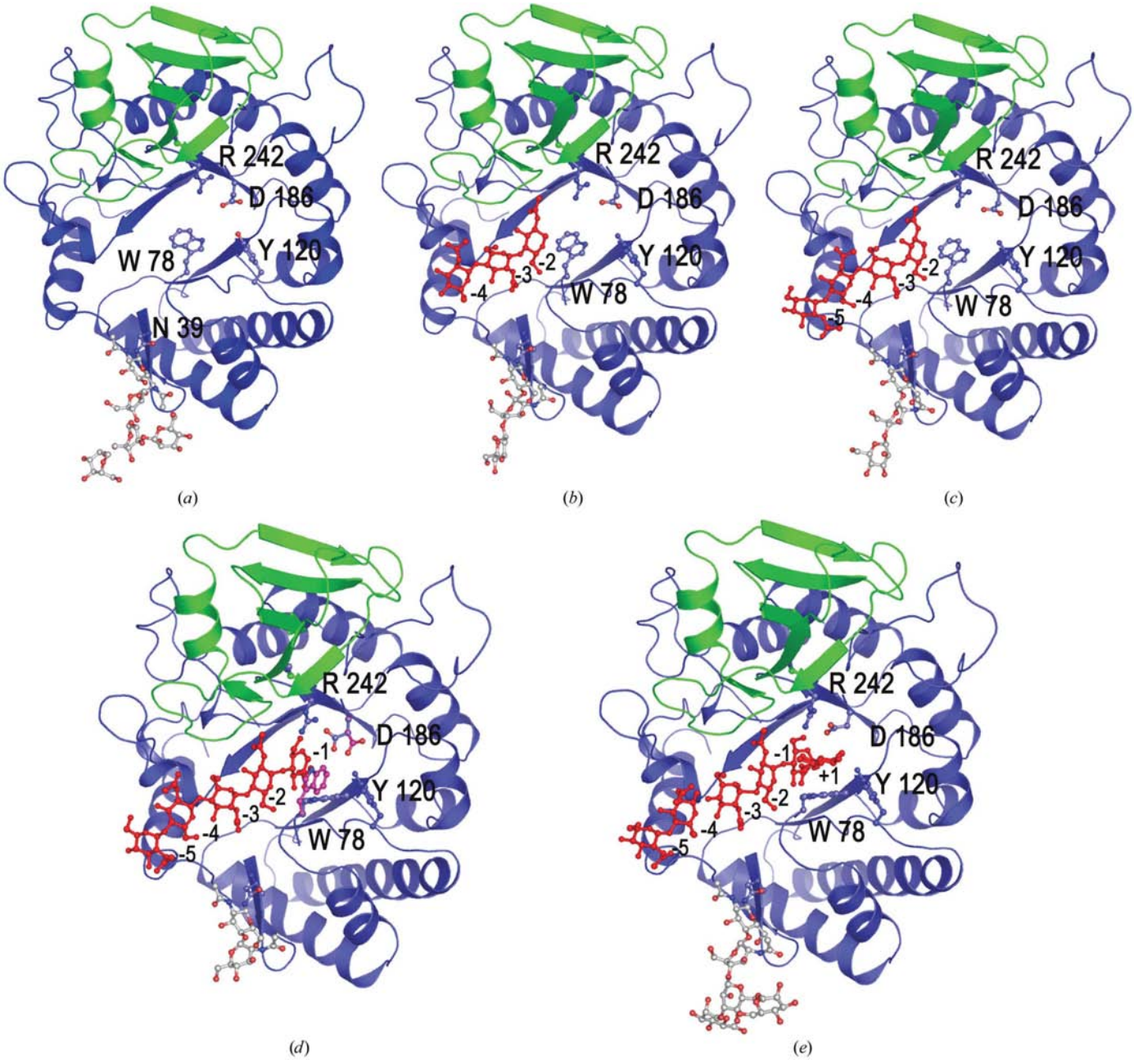

Figure 4
Structures of native SPG-40 (*a*) and of its complexes with oligosaccharides, showing the locations of carbohydrates in the carbohydrate-binding groove: (*b*) $GlcNAc_3$ at subsites −4 to −2, (*c*) $GlcNAc_4$ at subsites −5 to −2, (*d*) $GlcNAc_5$ at subsites −5 to −1 and (*e*) $GlcNAc_6$ at subsites −5 to +1. The protein conformation on binding to $GlcNAc_3$ and $GlcNAc_4$ is unperturbed, while it changes partially on binding to $GlcNAc_5$ and is perturbed considerably on binding to $GlcNAc_6$.

Table 3
Subsites occupied by various oligosaccharides in bacterial chitinase (Aronson *et al.*, 2003), human chitinase (HCHT; Fusetti *et al.*, 2002), HCGP-39A (Houston *et al.*, 2003), HCGP-39B (Fusetti *et al.*, 2003) and SPG-40.

Occupied subsites are indicated by asterisks.

		Sugar-binding subsites									
Protein	Saccharide	−7	−6	−5	−4	−3	−2	−1	+1	+2	+3
Chitinase	$GlcNAc_7$			*	*	*	*	*	*	*	
	$GlcNAc_8$		*	*	*	*	*	*	*	*	
HCHT	$GlcNAc_2$						*	*			
	Allosamidin					*	*	*			
HCGP-39A	$GlcNAc_3$	*	*	*							
	$GlcNAc_4$						*	*	*	*	
	$GlcNAc_8$				*	*	*	*	*	*	
HCGP-39B	$GlcNAc_2$		*	*							
	$GlcNAc_4$						*	*	*	*	
	$GlcNAc_5$					*	*	*	*	*	
	$GlcNAc_6$					*	*	*	*	*	*
SPG-40	$GlcNAc_3$				*	*	*				
	$GlcNAc_4$			*	*	*	*				
	$GlcNAc_5$			*	*	*	*	*			
	$GlcNAc_6$			*	*	*	*	*	*		

indicating a stringent requirement for *cis* bonds in the SPG structure (Fig. 3). The structure determinations also indicated the presence of an N-linked glycan chain at Asn39 (Fig. 4).

3.3. Carbohydrate binding

The most characteristic features of SPG-40 structure are (i) the two-domain organization of the protein with a $(\beta/\alpha)_8$ TIM-barrel domain and a small $(\alpha + \beta)$ domain, (ii) glycosylation at Asn39, (iii) formation of three *cis*-peptide bonds, (iv) a relatively less ordered surface domain which is unique to SPG-40 and (v) a long well formed carbohydrate-binding groove. The carbohydrate-binding studies were carried out using fluorescence spectroscopy. $GlcNAc_3$ and $GlcNAc_4$ did not show appreciable binding affinities; hence, their binding constants could not be estimated. In the case of $GlcNAc_5$ and $GlcNAc_6$, these binding constants were found to be 260 ± 3 and 18 ± 4 μ*M*, respectively, by fluorescence. The binding constants for various carbohydrates to other proteins such as HCGP-39 (Houston *et al.*, 2003), bacterial chitinase (Aronson *et al.*,

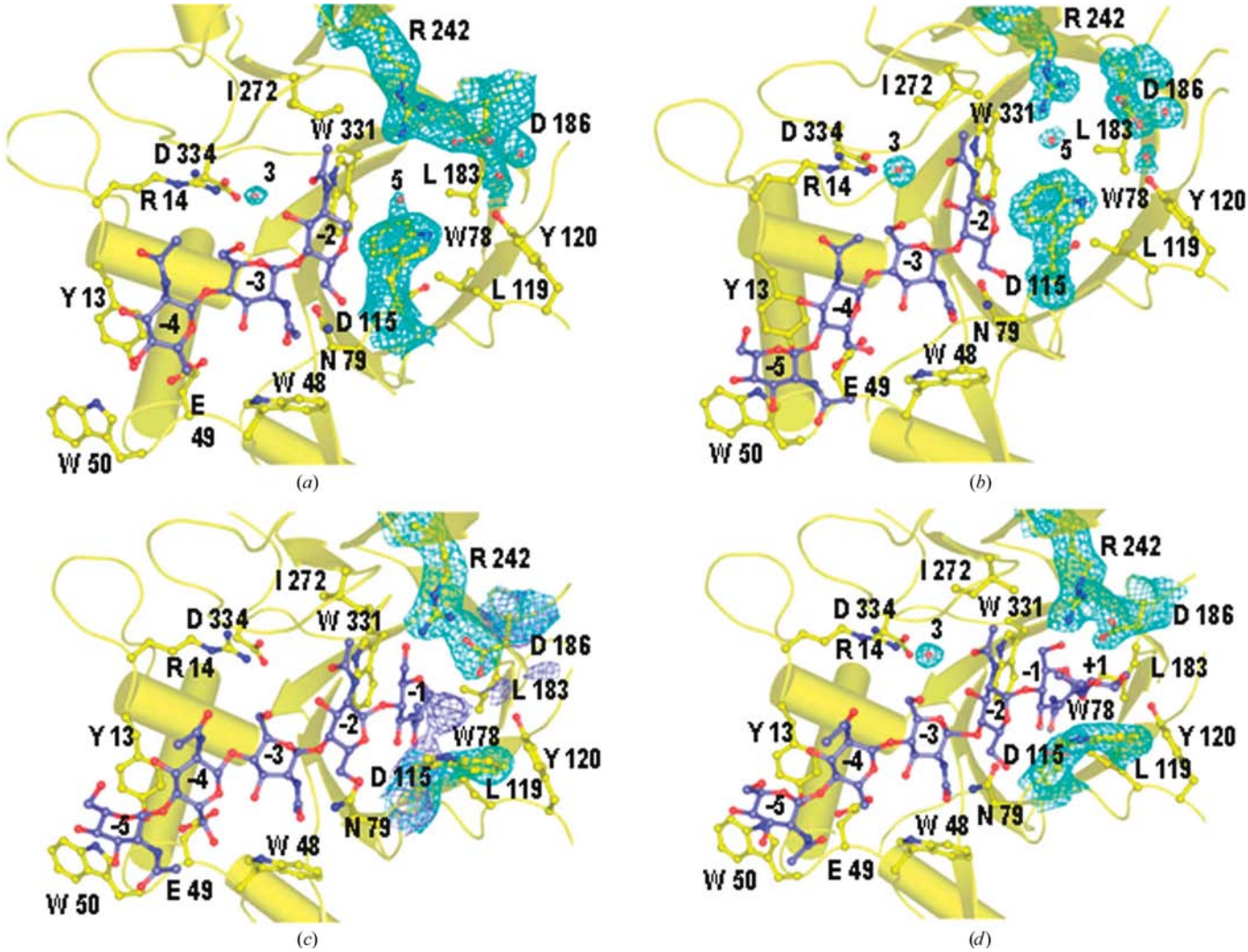

Figure 5
$|2F_o - F_c|$ electron densities for Trp78, Asp186 and Arg242 with bound (*a*) $GlcNAc_3$, (*b*) $GlcNAc_4$, (*c*) $GlcNAc_5$ and (*d*) $GlcNAc_6$. In (*c*) partial densities for Trp78 and Asp186 are also observed (blue colour).

2003) and fungal chitinase (Fukamizo *et al.*, 2001) are also listed in Table 2. The values of the binding constants for these proteins are much higher then those observed for SPG-40, indicating a less favourable environment in the carbohydrate-binding groove of SPG-40. The groove is formed by aromatic residues which provide stacking interactions with the hydrophobic portions of the bound sugar rings. $GlcNAc_3$ is held at subsites −4 to −2 (Fig. 5*a*), $GlcNAc_4$ occupies subsites −5 to −2 (Fig. 5*b*) and $GlcNAc_5$ is located at subsites −5 to −1 (Fig. 5*c*), while $GlcNAc_6$ is placed at subsites −5 to +1 (Fig. 5*d*). These observations indicate that the most frequently occupied subsites in SPG-40 are −4 to −2 (Table 3).

3.4. Conformational changes associated with saccharide binding

As seen from Table 3, $GlcNAc_3$ occupies subsites −4, −3 and −2, while $GlcNAc_4$ saturates subsites −5, −4, −3 and −2. Binding to the protein occurs without any perturbation to the protein structure (Figs. 4*b* and 4*c*). The subsites in contact with $GlcNAc_5$ are −5, −4, −3, −2 and −1. The binding of $GlcNAc_5$ induces notable conformational changes in the protein. The side chain of Trp78 is partially rotated from $\chi_1 = -51^\circ$ in the native protein to $\chi_1 = -173^\circ$ in the complex (Figs. 4*d* and 5*c*). Other residues that are affected by pentasaccharide binding are Asp186 and Arg242. On further increasing the length of polysaccharide to $GlcNAc_6$, subsites −5, −4, −3, −2, −1 and +1 are filled, resulting in large-scale conformational changes in the protein (Figs. 4*e* and 5*d*).

The electron density for water molecules and some of the important residues involved in carbohydrate binding in the groove in the native structure of SPG-40 are indicated in Fig. 6. In the process of binding, the carbohydrate displaces several water molecules which are present in the native SPG-40 structure (Fig. 6) and this is accompanied by large-scale conformational changes in the residues of the carbohydrate-binding groove (Figs. 5*c* and 5*d*). The electron densities for Trp78 and Asp186 indicated positions similar to those in the native structure (Fig. 6) in both the $GlcNAc_3$ and $GlcNAc_4$ complexes (Figs. 5*a* and 5*b*), while in the complex with $GlcNAc_5$ two occupancies were observed for Trp78 and Asp186, indicating a perturbation in the positions of these residues in order to accommodate $GlcNAc_5$ (Fig. 5*c*) and indicating that a pentasaccharide is probably the minimum length required to induce opening of the carbohydrate-binding groove by large-scale conformational changes in Trp78 and Asp186 (Figs. 4*d* and 5*c*). In the complex with $GlcNAc_6$, only single conformations were found, but these were completely different from those of the native structure (Figs. 5*d*).

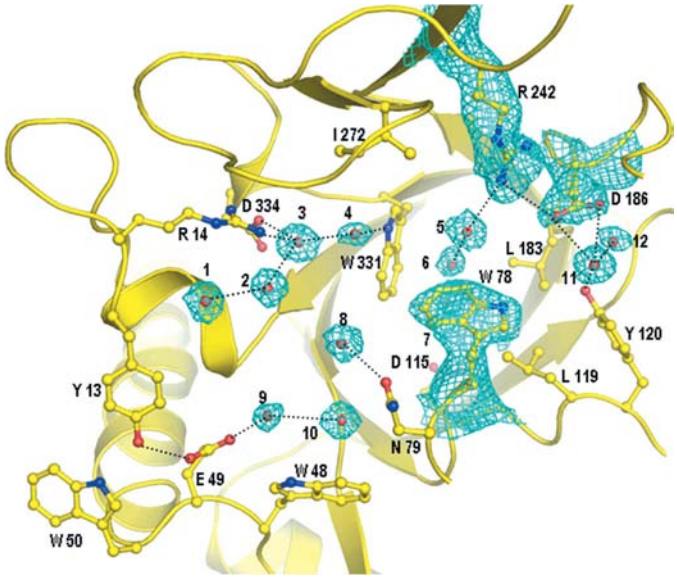

Figure 6
Water molecules with electron densities $|2F_o - F_c|$ at 1.5σ in the carbohydrate-binding groove in the native SPG-40 structure. The electron densities for Trp78, Asp186 and Arg242 are also shown.

Table 4
Hydrogen-bonding interactions of various oligosaccharides with SPG-40.

The interactions that are common in all four oligosaccharides are shown in bold.

Residues involved			Distance (Å)			
Binding subsite	Sugar	Protein	$GlcNAc_3$	$GlcNAc_4$	$GlcNAc_5$	$GlcNAc_6$
+1	GlcNAc N2	Trp78 $N^{\varepsilon1}$				3.23
	GlcNAc O6	Tyr120 OH				2.73
−1	GlcNAc O3	Trp78 N			3.10	2.98
	GlcNAc O5	Tyr185 OH			3.47	3.80
	GlcNAc O^6	Asp186 $O^{\delta1}$			3.23	2.80
		Arg242 NH1			3.48	
		Arg242 NH2			3.47	3.45
−2	GlcNAc O3	Glu269 $O^{\varepsilon1}$	2.53	2.72	2.83	2.78
	GlcNAc O4	Asn79 $O^{\delta1}$	2.51	2.64	2.63	2.69
	GlcNAc O6	Asn79 $O^{\delta1}$		3.13		3.20
		Trp78 N	2.75	2.68	2.88	2.89
		Asn79 N	2.83	2.70	2.75	2.75
		Wat O		2.78	2.54	2.52
		Wat O	2.93		3.47	3.44
	GlcNAc O7	Asn79 $O^{\delta1}$		2.78		
		Trp331 $N^{\varepsilon1}$	3.08	3.30	3.21	3.32
		Wat O	2.88	3.49		
−3	GlcNAc N2	Asn79 $O^{\delta1}$	2.77	2.82	2.78	
	GlcNAc O6	Glu269 $O^{\varepsilon1}$	3.01	3.40	3.41	3.11
	GlcNAc O7	Asn79 $O^{\delta1}$			2.66	
−4	GlcNAc O6	Glu49 $O^{\varepsilon1}$	3.54	3.22	3.40	3.38
		Glu49 $O^{\varepsilon2}$	2.59	2.67	2.91	3.25
−5	GlcNAc N2	Wat O		2.64	3.00	3.04
	GlcNAc O3	Wat O			2.95	2.94

3.5. Comparison with saccharide binding in chitinase and HCGP-39

The SPG-40 structure presents a large contrast to chitinases, where chitin polymers move smoothly into the carbohydrate-binding groove without any hindrance (Fusetti *et al.*, 2002; Aronson *et al.*, 2003), although the overall scaffold of SPG-40 is essentially similar to that of chitinases, with the C^α traces of the two proteins showing an r.m.s. shift of 0.8 Å. The intermolecular hydrogen bonds (Table 4) and hydrophobic interactions between the oligosaccharide residues and protein atoms from various subsites indicate that subsite −2 contributes maximally to stability, followed by the other sites. This is

unlike chitinases and HCGP-39, where the most preferred sites are −1 and +1. These sites are close to the scissile bond. In the case of chitinase, the crystal structures of two complexes with $GlcNAc_7$ and $GlcNAc_8$ are known (Aronson *et al.*, 2003), while for human chitinase the crystal structure of the complex with a disaccharide has been determined (Fusetti *et al.*, 2002). It has been reported that the subsites from −6 to +2 are occupied, the most frequently occupied subsites being −2 and −1, and that the binding of oligosaccharides does not induce any change in the protein structure.

Another protein, HCGP-39, which is closely related to SPG-40 (r.m.s. shift for C^{α} atoms 0.7 Å) has been studied using two preparations, one from human chondrocytes (Houston *et al.*, 2003; HCGP-39A) and the other expressed in synovial cells (Fusetti *et al.*, 2003; HCGP-39B), in complex with oligosaccharides. As seen in Table 3, in both forms disaccharides and trisaccharides bind at distal position and occupy subsites −7 to −5, while longer oligosaccharides occupy subsites from −4 to +3, indicating a preferences for subsites −2 to +2. This occupancy behaviour of various oligosaccharides in SPG-40, chitinases and HCGP-39 clearly indicate easy diffusion in chitinases and moderate accessibility in HCGP-39, while diffusion appears to be considerably constrained in SPG-40. This is also reflected from the values of binding affinities of various oligosaccharides to SPG-40, HCGP-39 and chitinases (Table 2). For example, the binding affinity of hexasaccharide is nearly three times stronger for HCGP-39, while it is more than 250 times stronger in chitinases when compared with that of SPG-40.

4. Conclusions

The reason for the variations in subsite preference between SPG-40 and HCGP-39 can be attributed to differences in the intramolecular interactions in and around the carbohydrate-binding groove. Apart from the network of water molecules, the environment surrounding Trp78 in SPG-40 differs significantly from that found in HCGP-39 (Houston *et al.*, 2003; Fusetti *et al.*, 2003). Furthermore, the side chain of Asp186 in SPG-40 is aligned in such a way that it forms a link with the OH group of Tyr120 through a water molecule. This arrangement seems to constrain the carbohydrate-binding groove beyond subsite +1. On the other hand, Asp186 in HCGP-39 is oriented away from the groove because it is hydrogen bonded to Arg242, which is linked to Thr272. As a result, the carbohydrate-binding groove in HCGP-39 remains relatively unhindered. The observed hydrogen-bonding linkage in HCGP-39, Thr272···Arg242···Asp186, is absent in SPG-40 owing to the mutation of Thr272 to Ile272 in SPG-40. These observations indicate that the carbohydrate-binding grooves in SPG-40 and HCGP-39 differ appreciably.

Compared with the chitin-hydrolyzing enzyme chitinase, the mutation of the active-site residue from Glu to Leu and other specific structural variations in the sugar-binding groove observed in SPG-40 indicate a loss of chitinase activity as well as variation in the affinity of sugar binding. In addition to this, the presence of a flexible region exclusively observed in SPX-40 proteins indicates a potential protein-binding site. These three factors in SPG-40 suggest that it is a purely binding protein that seems to combine initial binding to glycan chains with subsequent formation of protein–protein interactions. This interesting signalling protein presents an excellent example of evolution in which primary sequence and structural homologies are retained but the function has altered dramatically. The results of these studies on this new class of signalling proteins are expected to enhance understanding of the precise role of SPG-40 in the processes of tissue remodelling during involution as well as in the progression of breast cancers.

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 the establishment of a proteomics facility is also acknowledged. The DST is also thanked for support under the 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

Aronson, N. N. Jr, Halloran, B. A., Alexyev, M. F., Amable, L., Madura, J. D., Pasupulati, L., Worth, C. & Roey, P. V. (2003). *Biochem. J.* **376**, 87–95.
Aslam, M. & Hurley, W. L. (1997). *J. Dairy Sci.* **81**, 748–755.
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.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
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.
Eftink, M. R. (1997). *Methods Enzymol.* **278**, 221–257.
Fang, W. & Sandholm, M. (1995). *J. Dairy Res.* **62**, 61–68.
Fukamizo, T., Saski, C., Schelp, E., Bortone, K. & Robertus, J. D. (2001). *Biochemistry*, **40**, 2448–2454.
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.* **27**, 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.
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.
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.
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. (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., Sharma, S., Das, U., Srinivasan, A. & Singh, T. P. (2006). *J. Struct. Biol.* **156**, 505–516.
Strange, R., Li, F., Saurer, S., Burkhardt, A. & Friis, R. R. (1992) *Development*, **115**, 49–58.
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.