Sugar-mediated lattice contacts in crystals of a plant glycoprotein

Tanis Hogg,[a] Ivana Kuta Smatanova,[b] Karel Bezouska,[cd] Norbert Ulbrich[e] and Rolf Hilgenfeld[a]*

[a]*Department of Structural Biology and Crystallography, Institute of Molecular Biotechnology, Beutenbergstrasse 11, D-07745 Jena, Germany,* [b]*Institute of Physical Biology, University of South Bohemia at Ceske Budejovice, Novy zamek 136, CZ-37333 Nove Hrady, Czech Republic,* [c]*Department of Biochemistry, Faculty of Science, Charles University Prague, Hlavova 8, CZ-12840 Praha 2, Czech Republic,* [d]*Laboratory of Protein Architecture, Department of Immunology and Gnotobiology, Institute of Microbiology, Czech Acadamy of Science, CZ-14220 Praha 4, Czech Republic, and* [e]*BioCol GmbH, Jageralle 15, D-14469, Potsdam, Germany. E-mail: hilgenfd@imb-jena.de*

Pokeweed antiviral protein, PAP-S_{aci}, isolated from seeds of the Chinese pokeweed plant, *Phytolacca acinosa*, belongs to the family of type-1 ribosome-inactivating proteins (RIPs). Type-1 RIPs are ~30-kDa N-glycosidases that inactivate eukaryotic and prokaryotic ribosomes *via* a site-specific depurination of ribosomal RNA (rRNA). Here we describe the preliminary X-ray structure determination at 1.7 Å resolution of one PAP isoenzyme from seeds, PAP-$S1_{aci}$, after crystallisation from a heterogeneous mixture of two isoenzymes. PAP-$S1_{aci}$ possesses a rare type of glycosylation, specifically, N-linked *N*-acetyl-D-glucosamine monosaccharide (GlcNAc) substitutions at canonical Asn-Xaa-Ser/Thr sequons. One GlcNAc residue was found to play a critical role in crystal lattice formation, forming a packing interface across a crystallographic two-fold with the identical sequon of an adjacent monomer. This observation suggests that deglycosylation protocols for the crystallisation of glycoproteins should be designed to allow for exploitation of the crystal packing potential of the innermost core sugar residue (N-linked GlcNAc or O-linked GalNAc).

Keywords: pokeweed antiviral protein; PAP; ribosome-inactivating protein; RIP; glycosylation; deglycosylation; glycoprotein; N-acetyl glucosamine; X-ray crystal structure; crystallogenesis; crystal contacts

1. Introduction

A wide variety of plant species produce enzymes that are potent inactivators of the ribosomal machinery and are accordingly termed ribosome-inactivating proteins (RIPs). Interest in RIPs has been on a continual upswing due to their potentially useful applications in medical science as immunotoxins against cancer and AIDS (Irvin & Uckun, 1992; Sandvig & van Deurs, 2000) and in crop science, including applications in the construction of transgenic plants exhibiting broad-spectrum resistance to viral and fungal infection (Zoubenko *et al.*, 1997; Wang *et al.*, 1998). Intensive investigations are being carried out by several groups on pokeweed antiviral proteins (PAPs), type-1 RIPs isolated from various members of the *Phytolacca* genus (pokeweed). Classical type-1 RIPs, including PAPs as well as other members such as trichosanthin, momorchardin, saporin and agrostin, are monomeric proteins of approximately 30 kDa that exhibit RNA N-glycosidase (Endo *et al.*, 1987, 1991) and polynucleotide:adenosine glycosidase (Barbieri *et al.*, 1997) activities. In contrast, type-2 RIPs, such as ricin and gelonin, are heterodimers consisting of an A chain which is structurally and functionally homologous to the type-1 RIPs, and a sugar-binding B chain which attaches the RIP to the surface of target cells. RIPs inhibit protein biosynthesis by recognizing the highly conserved α-sarcin/ricin loop of the large ribosomal RNA (rRNA) species and catalyzing a site-specific deadenylation reaction (targeting A_{4324} in rat liver ribosome 28S rRNA, for example) (Endo *et al.*, 1987). Ribosomes thus modified lack the ability to interact with elongation factors 1a and 2 (EF-Tu and EF-G in bacteria, respectively) (Montanaro *et al.*, 1975; Gessner & Irvin, 1980; Hausner *et al.*, 1987; Macbeth & Wool, 1999; Wool *et al.*, 2000), resulting in the irreversible inhibition of protein biosynthesis. Although type-2 RIPs are more cytotoxic due to their carbohydrate-binding B-chain, their activity is more or less restricted to animal ribosomes (Barbieri *et al.*, 1993). In contrast, type-1 RIPs such as PAP have a broad specificity for eukaryotic and prokaryotic ribosomes. The broad spectral activity of PAP has been suggested to stem from specific interactions between PAP and ribosomal protein L3, the latter of which is highly conserved throughout the different taxa (Hudak *et al.*, 1999).

Pokeweed produces multiple PAP isoenzymes, each appearing in different tissues and/or at different stages in the plant's life-cycle. The best-characterized pokeweed species, *Phytolacca americana*, is known to produce the following isoenzymes: a) PAP-I, from spring leaves; b) PAP-II, from early summer leaves; c) PAP-III, from late summer leaves; d) PAP-R, from roots; and e) PAP-S, from seeds (Irvin & Uckun, 1992; Barbieri *et al.*, 1993). In addition, a clone encoding a PAP protein with C- and N-terminal extensions, termed α-PAP, was isolated from the genomic DNA of *P. americana*, and shown to exhibit RNA N-glycosidase activity (Kataoka *et al.*, 1992). PAP-S from *P. americana* has been shown to posses a relatively rare type of glycosylation, specifically, N-linked *N*-acetyl-D-glucosamine (GlcNAc) monosaccharide substitutions at three canonical Asn-Xaa-Ser/Thr sequons (Islam *et al.*, 1991), although the exact role of these sugars is yet to be resolved. More recent work has shown that seeds from *Phytolacca americana* actually contain two different PAP isoforms, PAP-S1 and PAP-S2, with 83 % sequence identity and different inhibitory activities on ribosomes (Honjo *et al.*, 2002). An earlier crystallographic paper reported on the crystallisation of PAP-S from *P. americana* (Li *et al.*, 1998), although the authors indicated that only one isoform was detected by SDS-PAGE analysis on lyophilized samples. Monzingo *et al.* (1993) first described the 2.5 Å X-ray crystal structure of PAP-I from *P. americana* at room temperature, both uncomplexed and together with the substrate analog formycin 5'-monophosphate. This work was complemented years later by crystal structure determinations of PAP-I at cryo temperatures to 2.0-2.1 Å resolution, both as the free enzyme and in complexes with the active-site RIP inhibitor pteoric acid, as well as with adenine and guanine, both products of N-glycosidase reactions (Kurinov *et al.*, 1999a; 1999b). In addition, the X-ray structure of α-PAP (Ago et al., 1994) and the crystallisation of PAP-II (Kurinov et al., 2000) have been published. Although the structural and functional properties of PAPs from *P. americana* have been examined in quite some detail, less attention has been paid to these enzymes from other species. In order to obtain a better understanding of the structural elements responsible for governing the pronounced differences in specificity and enzymatic activity between the different species and isoforms, structural studies of other PAP proteins are necessary. Messenger RNA encoding a leaf isoform of PAP has been sequenced from the Chinese pokeweed plant, *Phytolacca acinosa* (formerly called *Phytolacca esculenta*) (TrEMBL entry Q941G8), and we have isolated two major PAP

isoenzymes from *P. acinosa* seeds, which we term PAP-S1$_{aci}$ and PAP-S2$_{aci}$. This heterogeneous PAP-S$_{aci}$ seed extract exhibits significantly higher activity against plant virus infections (by tobacco mosaic virus as well as tomato or potato mosaic viruses) than PAP from *Phytolacca americana* (unpublished results). Here we report the preliminary X-ray structure of one isoenzyme, PAP-S1$_{aci}$, to 1.7 Å, after crystallisation from a heterogeneous mixture of PAP-S1$_{aci}$ and PAP-S2$_{aci}$, each of which may in turn contain several isoforms, as indicated by preliminary mass-spectrometric analysis (KB, unpublished). Although the sequence of PAP-S1$_{aci}$ is not yet known, the excellent quality of the electron density maps, combined with the known sequences of the two PAP-S isoforms from the related *P. americana* pokeweed plant (Honjo et al., 2002), has allowed us to undertake an 'X-ray sequencing' approach with a high degree of confidence. We present structural evidence that a GlcNAc-asparagine structure similar to that discovered biochemically for PAP-S from *P. americana* (Islam et al., 1991) is also present in PAP-S1$_{aci}$. Furthermore, we show that one GlcNAc residue is critically involved in crystal lattice formation. Implications for the general pre-treatment of glycoproteins prior to screening for crystallisation conditions are discussed on the basis of these findings.

2. Materials and methods

2.1. Protein preparation and crystallisation

PAP-S$_{aci}$ was isolated from mature seeds of *P. acinosa* utilizing an initial purification procedure similar to that established by Barbieri *et al.* (1982) for the related protein from *P. americana*. Subsequent to the initial purification step employing a carboxymethyl cellulose (CMC 52) column, PAP-S$_{aci}$ fractions were pooled and dialysed against 15-20% PEG 20K in 5 mM sodium phosphate, pH 6.7, and concentrated to ~10 mg/ml at 4ºC. This probe was further processed by gel filtration through Sephadex G75 (2 cm x 140 cm) in 5 mM sodium phosphate, pH 6.7, at 6 ml per hour. 3-ml fractions were collected and analyzed by standard SDS-PAGE under reducing conditions. Fractions displaying PAP-S$_{aci}$ were collected, dialysed in the cold against demineralised water, and further purified by FPLC chromatography on Highload S. Bound protein was eluted with a salt gradient consisting of 0–200 mM NaCl in 5 mM sodium phosphate, pH 6.8, at 0.6 ml per min and a final wash with 1 M NaCl, and lyophilised for storage.

Lyophilised PAP-S$_{aci}$ was dissolved in redistilled water to a concentration of 20 mg/ml. Hanging-drop crystallisation trials were carried out using VDX plates (Hampton Research, Laguna Niguel, CA) at 289 K. 4 µl droplets of protein solution were mixed with 4 µl of PEG 4K (42-44%) and 2 µl 1M sodium citrate, all in 100 mM phosphate buffer, pH 7.2. Drops were equilibrated over reservoirs containing 42-44% PEG 4K and 100 mM phosphate buffer, pH 7.2. The compositions of the purified protein sample and of resolubilized PAP-S$_{aci}$ crystals were analyzed by SDS-PAGE with a precast 15% Tris-glycine gel (Bio-Rad).

2.2. Data collection and molecular replacement

A low-temperature data set was collected by transferring a loop-mounted crystal directly from the mother liquor to a stream of nitrogen gas at 100 K. Diffraction data to a resolution of 1.7 Å were collected using synchrotron radiation (λ = 0.8033 Å) at the IMB Jena – University of Hamburg – EMBL Beamline X13, DESY (Hamburg, Germany), on a MAR CCD camera (X-ray Research, Hamburg) at a crystal-to-detector distance of 150.0 mm. Indexing of the initial diffraction images indicated that the crystal lattice arrangement was body-centered orthorhombic (either I222 or $I2_12_12_1$, which possess identical extinction rules), with unit cell constants a = 78.63, b = 84.19 and c = 90.88 Å. A total of 600 frames were collected at rotation steps of 0.5 degrees. All data were processed and scaled using the packages *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997), resulting in a dataset of 33,596 unique reflections derived from 417,357 observations. Data collection statistics are summarized in Table 1.

Table 1 X-ray data collection and refinement statistics.

Values in parentheses are for the outer shell, with a resolution of 1.74-1.70 Å.

Data-collection statistics	
Temperature (K)	100
Number of crystals	1
Resolution range (Å)	30.0-1.70
Space group	I222
Unit-cell parameters (Å)	
a	78.63
b	84.19
c	90.88
Reflections	
Measured	417,357
Unique	33,596
Completeness (%)	99.9 (100.0)
Redundancy	12.4 (12.4)
R_{merge}[a] (%)	4.3 (38.9)
$R_{r.i.m.}$[b] (%)	4.5 (40.6)
$R_{p.i.m}$[c] (%)	1.3 (11.4)
Mean $\langle I/\sigma(I)\Sigma$	55.7 (7.1)
Refinement statistics	
Resolution range (Å)	30.0-1.70
Reflections used [>0σ(*F*)]	32,752
R factor[d (%)]	18.1
Free *R* factor (%)	21.9
Total number of non-H atoms	2396
Number of water molecules	350
R.m.s. deviations from ideals	
Bond lengths (Å)	0.012
Bond angles (º)	1.60
Dihedral angles (º)	23.0
Improper angles (º)	0.96
Average B factors (Å^2)	
All atoms	25.3
Protein	23.3
Water	36.9
N-acetyl glucosamine	31.8
Ramachandran plot	
Residues in most favorable regions (%)	91.7
Residues in additional allowed regions (%)	8.3
Residues in disallowed regions (%)	0

[a] $R_{merge} = \sum_{hkl}\sum_{i=1}^{N}\langle I^{hkl}\rangle - I_i^{hkl}| / \sum_{hkl}\sum_{i=1}^{N} I_i^{hkl}$

[b] $R_{r.i.m.} = \sum_{hkl}(N/N\text{-}1)^{\frac{1}{2}}\sum_{i=1}^{N}|\langle I^{hkl}\rangle - I_i^{hkl}| / \sum_{hkl}\sum_{i=1}^{N} I_i^{hkl}$ where N is the multiplicity of the observed reflection (Weiss & Hilgenfeld, 1997).

[c] $R_{p.i.m} = \sum_{hkl}(1/N\text{-}1)^{\frac{1}{2}}\sum_{i=1}^{N}|\langle I^{hkl}\rangle - I_i^{hkl}| / \sum_{hkl}\sum_{i=1}^{N} I_i^{hkl}$ where N is the multiplicity of the observed reflection (Weiss & Hilgenfeld, 1997).

[d] R factor$=\sum_{hkl\subset W}||F_{obs}| - k|F_{calc}|| / \sum_{hkl\subset W}|F_{obs}|$, where hkl⊂W indicates all reflections belonging to the working set W of unique reflections. The free R factor was calculated with 5% of the data excluded from the refinement.

The structure was solved by molecular replacement, using the atomic coordinates of *P. americana* PAP-I (Monzingo *et al.*, 1993; PDB code 1paf) as a search probe. Calculation of the Matthews coefficient, V_m (Matthews, 1968), excluded the possibility of more than one PAP-S$_{aci}$ molecule in the asymmetric unit (V_m = 2.5 Å^3/Da with an estimated M_r of 30 kDa). The rotation function and translation searches were carried out with *CNS* 1.0 (Brünger *et al.*, 1998), using data between 15.0-4.0 Å resolution. For the rotation

function, a fast direct protocol was selected. The asymmetric unit of the rotation function (Rao *et al.*, 1980) was initially sampled over a coarse angular grid of 13.6°, with the top 20 peak clusters subjected to a finer grid sampling of 2.7° intervals. The best 1000 solutions were filtered by the Patterson correlation refinement procedure producing one prominent peak in the function, with angles θ_1 = 221.8, θ_2 = 100.8, θ_3 = 98.2°. This peak, which represented the correct solution, was 6.8 σ above the average of the rotation function, with a difference of 4.2 σ to the next peak. Translation searches were performed in both space groups I222 and $I2_12_12_1$, and unambiguously revealed I222 as the true space group. The highest peak was well separated from other solutions with T = 0.56 (the next largest peaks being T = 0.18 and T = 0.12). The corresponding top peak in the $I2_12_12_1$ space group had T = 0.34, with the next peaks having T = 0.19 and T = 0.14. The rotational and translational parameters of the monomer were further refined using the rigid-body protocol of *CNS*, resulting in an R factor of 39.5 % for 30.0-3.5 Å resolution data.

2.3. Refinement

The model was subjected to an initial refinement protocol consisting of simulated-annealing refinement (Brünger *et al.*, 1987) using a slow-cool scheme (Brünger *et al.*, 1990) from 2000 to 100 K and 30.0-1.70 Å data, followed by 150 cycles of positional and 20 cycles of atomic B-factor refinement. A bulk-solvent correction (Jiang & Brünger, 1994) was utilized throughout the refinement. The resulting model exhibited an R factor of 31.4% (R_{free} 34.9%) for all data between 30.0 and 1.70 Å. At this point, electron-density map interpretation and model building was carried out, using the molecular graphics program O (Jones *et al.*, 1991). The excellent map quality allowed for an 'X-ray sequencing' approach, as several amino acid exchanges with respect to the sequence of PAP-I from *P. americana* were clearly evident. In parallel, and since the sequence of PAP-S1$_{aci}$ was not yet known, the model phases were used as input into the *ARP/wARP* 5.0 package (Perrakis *et al.*, 1997) for phase improvement and creation of a dummy atomic model, which was then used as an atomic template to assist in manual rebuilding. This procedure, along with the exceptional quality of the electron density map, allowed us to assign the identity of almost all amino-acid side chains with a high degree of certainty (with the exception of the Asn/Asp and Gln/Glu ambiguities). During rebuilding and refinement, a well-defined patch of $2F_o$-F_c and F_o-F_c density extending from the side chain of Asn-255 (numbering is according to the homologous *P. americana* PAP-S1 sequence) clearly indicated the presence of a single covalently attached GlcNAc residue, prompting inclusion of the latter into the model. Prior to any addition of water molecules, the R/R_{free} was 25.4/27.8% for all data between 30.0 and 1.70 Å resolution. Water molecules were then modelled into positions where spherical densities were above 1.0 σ in the $2F_o$-F_c map and 3.0 σ in the F_o-F_c map, provided that stereochemically reasonable hydrogen bonds were possible. These criteria ensured that water molecules were not erroneously placed into the uninterpreted density of any misidentified amino-acid side chains. A structural appraisal of the present model of PAP-S1$_{aci}$ using *PROCHECK* (Laskowski *et al.*, 1993) indicated that 91.7% of the residues are in the most favorable regions of the Ramachandran plot (Ramachandran *et al.*, 1974), and none in the generously allowed or disallowed regions. The refined structure includes residues 2 - 261, one GlcNAc moiety and 350 water molecules, yielding an R factor of 18.1% and a free R factor of 21.9%. The r.m.s deviations in bond lengths and angles are 0.012 Å and 1.6°, respectively. A summary of the refinement statistics is given in Table 1. Structure superposition was carried out with the program DALI (Holm & Sander, 1999). Figures 2 and 3 were prepared using Molscript (Kraulis, 1991), Bobscript (Esnouf, 1997) and Raster3D (Merritt & Bacon, 1997).

3. Results and discussion

SDS-PAGE analysis of resolubilized PAP-S$_{aci}$ indicated that the sample consisted of at least two proteins, with apparent molecular masses of approximately 29 kDa and 30 kDa, respectively (Fig. 1).

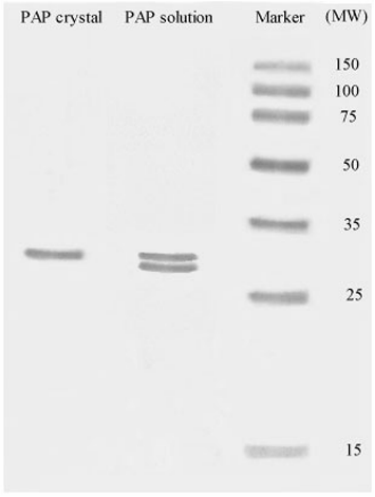

Figure 1

Coomassie Blue-stained SDS-PAGE showing the purified PAP-S$_{aci}$ preparation. The protein solution used for crystallisation contained two major isoenzymes, shown in lane 2. The crystals consisted only of the high-M_r isoenzyme (lane 1).

Preliminary N-terminal sequencing indicated that the two major components are both PAPs, exhibiting ~80% sequence identity over the first 40 residues (KB, unpublished). In contrast, the two recombinant *P. americana* PAP-S isoforms were shown to exhibit identical mobility, corresponding to a molecular mass of 30 kDa (Honjo *et al.*, 2002). An earlier paper describing the crystallisation of *P. americana* PAP-S, which was purified from its natural source and presumably contained the two isoforms, also showed a single band upon SDS-PAGE analysis (Li *et al*, 1998). Despite the obvious heterogeneity displayed by our sample, crystallisation experiments were initiated before investing in further purification efforts. Surprisingly, well-diffracting crystals with dimensions of 0.5 × 0.2 × 0.2 mm grew within one month against a background of amorphous precipitate. The crystallisation conditions identified were similar to those described for α-PAP from *P. americana* (Ago et al., 1994). Crystals were shown by SDS-PAGE (after dissolution) to consist only of the high M_r protein (Fig. 1). The body-centered orthorhombic crystal form which we obtained (Table 1) contains a single monomer in the asymmetric unit and appears to have the same unit cell as that reported earlier for PAP-S from *P. americana* (Li *et al.*, 1998). From genomic clones, Honjo *et al.* (2002) have deduced the amino acid sequences of PAP-S1 and PAP-S2 from *P. americana*, which were found to exhibit 83% sequence identity among themselves. Judging from our high-quality electron density maps and the low R factors of our current model (Table 1), we can conclude that we have crystallised an isoenzyme which contains a very similar sequence to the PAP-S1 isoform of *P. americana*. Hence, we have designated the name PAP-S1$_{aci}$ to our crystallised isoenzyme in order to indicate its origin.

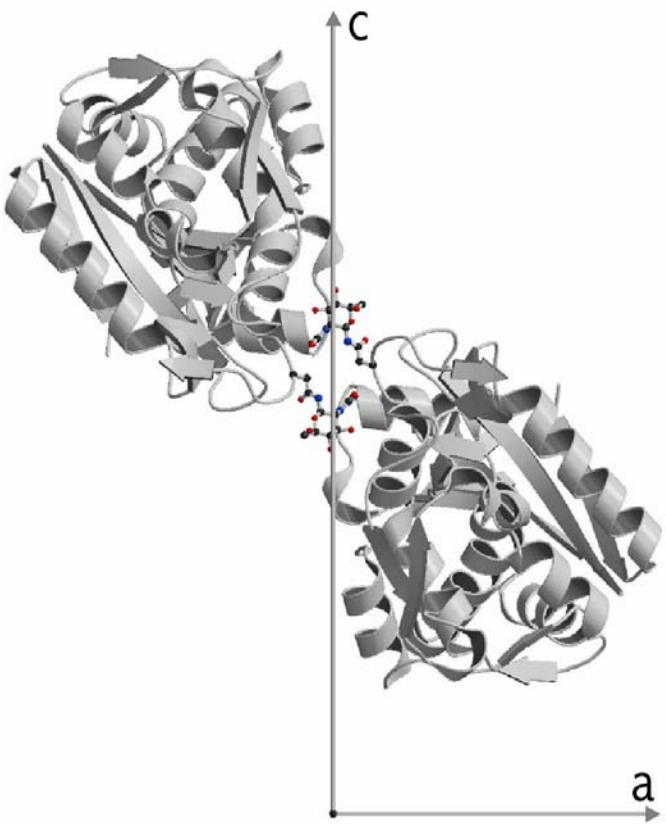

Figure 2

Packing diagram showing a PAP-S1$_{aci}$ monomer and a symmetry mate, forming a twofold parallel to the crystallographic *b* axis. The GlcNAc residues involved in the packing interaction are rendered in ball-and-stick. Vectors running along the unit-cell edges *a* and *c* are also depicted.

The tertiary structure of PAP-S1$_{aci}$ is similar to that of PAP-I from *P. americana* (Monzingo *et al.*, 1993; Kurinov *et al.*, 1999a; 1999b), with several notable exceptions. However, since the amino-acid sequence of PAP-S1$_{aci}$ has not yet been fully determined, a detailed comparison of PAP-S1$_{aci}$ with the known PAPs from *P. americana* would be premature. Despite our confidence in the electron density-based amino acid assignment, it is clear that the distinction between Asn and Asp, or Gln and Glu, or, sometimes, Thr and disordered Ser, is difficult. Therefore, a thorough description of the tertiary structure of PAP-S1$_{aci}$ will be published elsewhere when the primary structure of PAP-S1$_{aci}$ is established by Edman degradation and mass spectrometry. Here we restrict ourselves to the description of an interesting crystal contact that critically depends on the presence of a single sugar residue.

During model building and refinement, it became quite evident from the electron density that an Asn residue located near the C-terminus (corresponding to Asn-255 in PAP-S1 from *P. americana*) was glycosylated with a single GlcNAc residue. This is in agreement with a previous biochemical report which had identified this modification at the homologous position in PAP-S (presumably a mixture of isoforms) from *P. americana* (Islam et al., 1991). In our crystal structure, the tripeptide forming the N-glycosylation site (Asn-255, Gly-256, and Thr-257) protrudes from the surface of the protein by adopting a bent conformation (but not a β-turn), with the central glycine displaying dihedral angles easily accessible only to glycine residues ($\varphi = 94^{o}$, $\psi = 165^{o}$). The χ_1 and χ_2 torsion angles of the Asn side chain (-164° and 57°, respectively) are situated within a potential energy minimum that is commonly occupied by glycosylated asparagines (Imberty & Pérez, 1995).

One of the most interesting features displayed by this modification is the role of the sugar in crystal lattice formation, as it makes no less than two direct and two water-mediated hydrogen bonds to a crystallographically related monomer (Figs. 2 & 3). Most remarkably, the sugar establishes crystal contacts with the identical modification site on the adjacent monomer, resulting in the GlcNAc–Asn moieties circumscribing a twofold rotation axis parallel to the crystallographic *b* axis (Figs. 2 & 3). The carbonyl group of the acetyl function (O7') of the carbohydrate moiety accepts a hydrogen bond from the main-chain amide of GlcNAc–Asn-255 in the symmetry-related monomer. Additional intermolecular contacts are provided by the exocyclic 3' and 4' hydroxyl groups of GlcNAc. O3' forms a hydrogen bond with the side-chain hydroxyl of Thr-212$_{sym}$ (all residue number assignments are based on the sequence of PAP-S1 from *P. americana* (Honjo *et al.*, 2002)). Also, O3' makes an additional hydrogen bond through a water molecule to the backbone carbonyl of Tyr-253$_{sym}$. A second water-mediated contact is established between the O4' hydroxyl and the side-chain Oγ of Thr-212$_{sym}$. It should also be mentioned that an additional stabilizing interaction is provided by the parent monomer to GlcNAc *via* a water-mediated hydrogen bond between the main-chain amide of Thr-257 of the glycosylation motif and the exocyclic O6' moiety (Fig. 3). Based on these observations, we would predict that the present crystal form of PAP-S1$_{aci}$ could not have been obtained in the absence of the carbohydrate moiety attached to Asn-255. The presence of this strong sugar-mediated crystal contact led us to ask whether this interaction might have distorted the stereochemistry of the GlcNAc–Asn moiety itself, or of the surrounding loop region. In accord with most other glycoprotein structures solved to date, the pyranose ring of GlcNAc forms a typical chair conformation and extends away from the surface of PAP-S1$_{aci}$. The conformations of the exocyclic moeities are also quite typical (Imberty & Pérez, 1995), with the hydroxymethyl group at C5' adopting a favorable *gauche-gauche* orientation and the *N*-acetyl group at C2' exhibiting an ω_2 torsion angle (C1'–C2'–N2'–C7') of 115°. The latter is nearly identical to the ω_2 angle of 112.5° found in the crystal structure of isolated GlcNAc–Asn (Delbaere, 1974). The conformation of the glycopeptide linkage is also nearly ideal, with observed Φ_N (O5'–C1'–N1'–Cγ) and Ψ_N (C1'–N1'–Cγ–Cβ) values of -97° and 177°, respectively, which place this linkage in a low steric energy minimum (Imberty & Pérez, 1995). We would therefore conclude that the crystal contact does not impose any unfavorable distortions in the stereochemistry of the GlcNAc–Asn structure. In order to assess any possible alterations in the structure of surrounding loop region which may have resulted from the presence of a glycosidic linkage and/or the crystal contact, we compared the regional fold of our current PAP-S1$_{aci}$ model to that of α-PAP (Ago *et al.*, 1994). α-PAP possesses a homologous N-glycosylation motif (Asn-260–Gly-261–Thr-262), which is not utilised in the protein studied by X-ray crystallography because the gene was expressed in the glycosylation-incompetent *Escherichia coli* (Kataoka *et al.* 1992). Interestingly, α-PAP was crystallised in a different space group and a crystal packing analysis shows that the homologous Asn-260 forms a weak (>3.3 Å) crystal contact through its side chain to the main chain of an adjacent monomer. A superposition study between our current PAP-S1$_{aci}$ model and that of α-PAP demonstrated that neither the glycosylation nor the crystal contact had any detectable influence on the conformation of the surrounding loop region: the average Cα root mean square deviation of the pentapeptide region centered on Asn-255$_{aci}$/Asn-260$_{alpha}$ was found to be 0.5 Å, which is lower than the overall average Cα r.m.s.d. between the two proteins (0.9 Å). Based

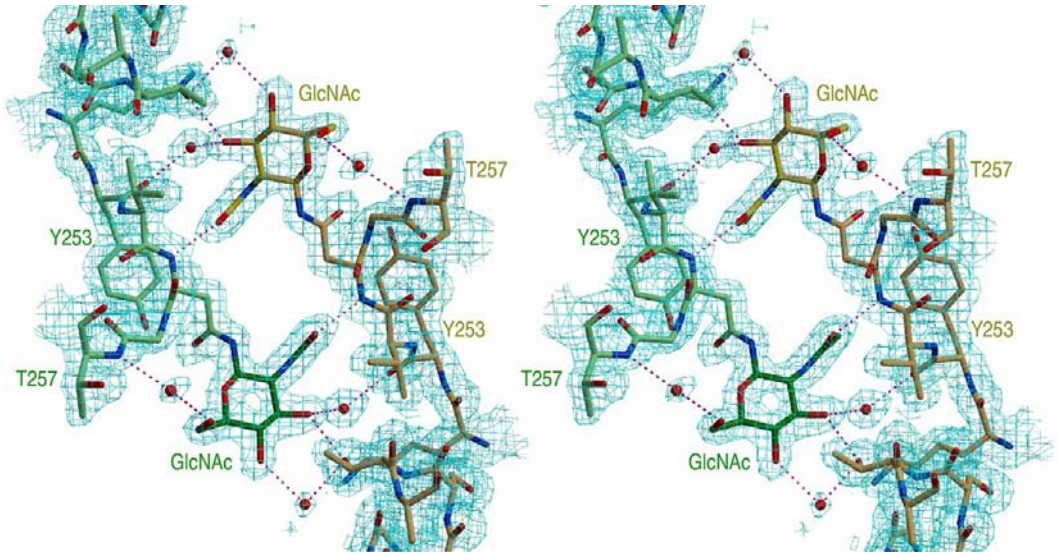

Figure 3

A detailed view of the sugar-mediated crystal contact depicted in Fig. 2. The current $2|F_o| - |F_c|$ electron density is shown, contoured at 1 σ. Hydrogen bonds are depicted as dashed lines.

on this, we can deduce that neither the glycosylation at Asn-255 nor the crystal contact formation had an impact on the conformation of the surrounding loop region to any significant extent.

4. Concluding remarks

Obtaining high-quality crystals suitable for X-ray diffraction studies remains a major challenge facing macromolecular crystallographers. This problem can be particularly frustrating when attempting to crystallise eukaryotic proteins, which are often glycosylated. It is well known that the presence of attached oligosaccharides can negatively influence the crystallisability of proteins. This phenomenon is presumably due to *i)* excess surface conformational entropy exhibited by the sugar chains, and *ii)* microheterogeneity of glycan chain length and composition. To circumvent this, it has been suggested that thorough deglycosylation be carried out as a routine method for the crystallisation of glycoproteins (Baker *et al.*, 1994; Grueninger-Leitch *et al.*, 1996). Indeed, several recent reports have emphasized the fundamental role of controlled deglycosylation in obtaining suitably diffracting crystals (Hoover *et al.*, 1999; Dale *et al.*, 2000; Davis *et al.*, 2001).

Based on our observations with the crystal structure of PAP-S1$_{aci}$ described here, we would suggest that when implementing a deglycosylation protocol for crystallisation of glycoproteins, partial deglycosylation should first be attempted by using selective endoglycosidases (such as endo F_3 or endo H) which leave the first core GlcNAc linked to the glycosylated Asn. Endoglycosidase F_3 is to be favored for removing complex oligosaccharide chains, whereas endoglycosidase H is most efficient with high mannose or hybrid oligosaccharide chains. Endoglycosidase A (N-glycosidase A from almond) apparently works for both types of carbohydrate chains (Tai *et al.*, 1997; Foddy *et al.*, 1986; Trimble *et al.*, 1986). In cases where the inner-core GlcNAc bears an additional fucosyl moiety, a combined strategy using selective endoglycosidases and α-fucosidases should produce similar results. As we have discovered with our current work on PAP-S1$_{aci}$, the advantages of a partial N-linked core go beyond the already well-established improvements in protein solubility and stability (Wyss & Wagner, 1996). The restricted mobility (van Zuylen *et al.* 1998), bulky structure and rich hydrogen-bonding potential of a single attached GlcNAc residue can provide an attractive landscape for crystal contact formation.

Acknowledgements

This work was supported in part by BMBF (through DESY-HS), grant # 05SH8BJA, and by Deutsche Forschungsgemeinschaft (grant # Hi 611/1-3). Beamline X13 is supported by BMBF (grant # 05KSIGKA/5). Access of IKS to the Centre for Design and Structure in Biology (CDSB), Jena, was supported by grant # HPRI-CT-1999-0097 from the European Commission. IKS and KB acknowledge support from the Ministry of Education of the Czech Republic (grants # LN00A141 to Dalibor Stys, and MSM 113100001, respectively). Research stays of RH at the Czech Academy of Science were supported by the KONTAKT program of the same ministry, by BMBF (WTZ program), and by the Howard Hughes Medical Institute (grant # 75195-540701).. RH also thanks the Fonds der Chemischen Industrie.

References

Ago, H., Kataoka, J., Tsuge, H., Habuka, N., Inagaki, E., Noma, M. & Miyano, M. (1994). *Eur. J. Biochem.* **225**, 369-374.
Baker, H. M., Day, C. L., Norris, G. E. & Baker, E. N. (1994). *Acta Cryst.* D**50**, 380-384.
Barbieri, L., Aron, G. M., Irvin, J. D. & Stirpe, F. (1982). *Biochem. J.* **203**, 55-59.
Barbieri, L., Battelli, M. G. & Stirpe, F. (1993). *Biochim. Biophys. Acta*, **1154**, 237-282.
Barbieri, L., Valbonesi, P., Bonora, E., Gorini, P., Bolognesi, A. & Stirpe, F. (1997). *Nucleic Acids Res.* **25**, 518-522.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science* **235**, 458-460.
Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Acta Cryst.* A**46**, 585-593.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905-921.
Dale, G. E., D'Arcy, B., Yuvaniyama, C., Wipf, B., Oefner, C. & D'Arcy, A. (2000). *Acta Cryst.* D**56**, 894-897.
Davis, S. J., Ikemizu, S., Collins, A. V., Fennelly, J. A., Harlos, K., Yvonne Jones, E. & Stuart, D. I. (2001). *Acta Cryst.* D**57**, 605-608.
Delbaere, L. T. J. (1974). *Biochem. J.* **143**, 197-205.
Endo, Y., Mitsui, K., Motizuki, M. & Tsurugi, K. (1987). *J. Biol. Chem.* **262**, 5908-5912.
Endo, Y., Gluck, A. & Wool, I. G. (1991). *J. Mol. Biol.* **254**, 848-855.

Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132-134.
Foddy, L., Feeney, J. & Hughes, R. C. (1986). *Biochem. J.* **233,** 697-706.
Gessner, S .L. & Irvin, J. D. (1980). *J. Biol. Chem.* **255**, 3251-3253.
Grueninger-Leitch, F., D'Arcy, A., D'Arcy, B. & Chène, C. (1996). *Protein Sci.* **5**, 2617-2622.
Hausner, T. P., Atmadja, J. & Nierhaus, K. H. (1987). *Biochimie*, **69**, 911-923.
Holm, L. & Sander, S. (1999). *Nucleic Acids Res.* **27**, 244-247.
Honjo, E., Dong, D., Motoshima, H. & Watanabe, K. (2002). *J. Biochem.* **131**, 225-231.
Hoover, D. M., Schalk-Hihi, C., Chou, C.-C., Menon, S., Wlodawer, A. & Zdanov, A. (1999). *Eur. J. Biochem.* **262**, 134-141.
Hudak, K. A., Dinman, J. D. & Tumer, N. E. (1999). *J. Biol. Chem.* **274**, 3859-3864.
Imberty, A. & Pérez, S. (1995). *Protein Eng.* **8**, 699-709.
Irvin, J. D. & Uckun, F. D. (1992). *Pharmacol. Ther.* **55**, 279-302.
Islam, M. R., Kung, S.-S., Kimura, Y. & Funatsu, G. (1991). *Agric. Biol. Chem.* **55**, 1375-1381.
Jiang, J. S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100-115.
Jones, T. A., Cowan, S., Zou, J.-Y. and Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110-119.
Kataoka, J., Habuka, N., Masuta, C., Miyano, M. & Koiwai, A. (1992). *Plant Mol. Biol.* **20**, 879-886.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946-950.
Kurinov, I. V., Myers, D. E., Irvin, J. D. & Uckun, F. M. (1999a). *Protein Sci.* **8**, 1765-1772.
Kurinov, I. V., Rajamohan, F., Venkatachalam, T. K & Uckun, F. M. (1999b). *Protein Sci.* **8**, 2399-2405.
Kurinov, I. V., Mao, C., Irvin, J. D. & Uckun, F. M. (2000). *Biochem. Biophys. Res. Commun.* **275**, 549-552.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283-291.
Li, H.-M., Zeng, Z.-H., Hu, Z. & Wang, D.-C. (1998). *Acta Cryst.* D**54**, 137-139.
Macbeth, M. & Wool, I. G. (1999). *J. Mol. Biol.* **285**, 567-580.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491-497.
Merritt, E. A. & Bacon, D. J. (1997). *Meth. Enzymol.* **277**, 505-524.
Montanaro, L., Sperti, S., Mattioli, A., Testoni, G. & Stirpe, F. (1975). *Biochem. J.* **146**, 127-131.
Monzingo, A. F., Collins, E. J., Ernst, S. R., Irvin, J. D. & Robertus, J. D. (1993). *J. Mol. Biol.* **233**, 705-715.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307-326.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448-455.
Ramachandran, G. N., Kolaskar, A. S., Ramakrishnan, C. & Sasisekharan V. (1974). *Biochim. Biophys. Acta*, **359**, 298-302.
Rao, S.N., Jih, J.-H. & Hartsuck, J.A. (1980). *Acta Cryst.* A**36**, 878-884.
Sandvig, K. & van Deurs, B. (2000). *EMBO J.* **19**, 5943-5950.
Tai, T., Yamashita, K. & Kobata, A. (1997). *Biochem. Biophys. Res. Commun.* **78**, 434-441.
Trimble, R. B., Atkinson, P. H., Tarentino, A. L., Plummer T. H. Jr, Maley, F. & Tomer, K. B. (1986). *J. Biol. Chem.* **261**, 12000-12005.
van Zuylen, C. W., de Beer, T., Leeflang, B. R., Boelens, R., Kaptein, R., Kammerling, J. P. & Vliegenthart, J. F. (1998). *Biochemistry*, **37**, 1933-1940.
Wang, P., Zoubenko, O. & Turner, N. E. (1998). *Plant Mol. Biol.* **38**, 957-964.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203-205.
Wool , I. G., Correll, C. C. & Chan, Y.-L. (2000). *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions*, edited by R.A. Garrett, S. R. Douthwaite, A. Liljas, A. T. Matheson, P. B. Moore & H. F. Noller, pp. 461-473. Washington, DC: ASM Press.
Wyss, D. F. & Wagner, G. (1996). *Curr. Opin. Biotechnol.* **7**, 409-416.
Zoubenko, O., Uckun, F., Hur, Y., Chet, I. & Turner, N. (1997). *Nature Biotechnol.* **15**, 992-996.

 Received 30 April 2002 • Accepted 8 August 2002