Crystal Structure of Peptide-N^4-(*N*-acetyl-β-D-glucosaminyl)asparagine Amidase F at 2.2-Å Resolution[†,‡]

Peter Kuhn,[§,∥] Anthony L. Tarentino,[§] Thomas H. Plummer, Jr.,[§] and Patrick Van Roey*[,§,⊥]

Wadsworth Center, New York State Department of Health, Albany, New York 12201-0509, Physics Department, University at Albany, Albany, New York 12222, and Department of Biomedical Sciences, School of Public Health, University at Albany, Albany, New York 12222

Received June 7, 1994; Revised Manuscript Received August 1, 1994[⊛]

ABSTRACT: Peptide-N^4-(*N*-acetyl-β-D-glucosaminyl)asparagine amidase F (PNGase F) is an amidase that cleaves the β-aspartylglucosylamine bond of asparagine-linked glycans. The 34.8-kDa (314 amino acids) enzyme has a very broad substrate specificity and is extensively used for studies of the structure and function of glycoproteins. Enzymatic activity of PNGase F requires recognition of both the peptide and the carbohydrate components of the substrate. Only limited information regarding the mechanism of action of the enzyme is available. The three-dimensional structure of PNGase F has been determined by X-ray crystallography at 2.2-Å resolution. The protein folds into two domains comprising residues 1–137 and 143–314, respectively. Both domains have eight-stranded antiparallel β-sandwich motifs that are very similar in geometry. Both sandwiches have parallel principal axes and lie side by side. The covalent link between the domains is located at the top end of the molecule. Extensive hydrogen-bonding contacts occur along the full length of the interface between the two domains. Three different areas, all at the interface between the two domains, have been identified as possible locations for the active site of the enzyme. These include a hydrophobic bowl of about 20 Å in diameter on one surface of the molecule, a long polar cleft on the opposite side, and a cleft at the bottom, which is lined with large aromatic residues including eight tryptophans.

Peptide-N^4-(*N*-acetyl-β-D-glucosaminyl)asparagine amidase F (PNGase F)[1] from *Flavobacterium meningosepticum* (Tarentino et al., 1985) is an amidase/amidohydrolase that cleaves the β-aspartylglucosylamine bond of asparagine-linked glycans, converting the asparagine residue to an aspartic acid (Plummer et al., 1984). The released 1-aminooligosaccharide spontaneously degrades to ammonia and the intact oligosaccharide (Risley & Van Etten, 1985; Rasmussen et al., 1992). The enzyme has a very broad substrate specificity (Plummer et al., 1984) and is extensively used as a biochemical tool in the study of glycoproteins, including the analysis of the oligosaccharide structure and of the role of oligosaccharides in the function and activity of glycoproteins (Tarentino & Plummer, 1993a). PNGase F hydrolyzes all types of asparagine-linked oligosaccharides, provided that both the amino and carboxyl groups of the asparagine are in polypeptide linkage. The enzyme recognizes mainly the inner di-*N*-acetylchitobiose moiety of the carbohydrate chain. In fact, the broad substrate specificity of PNGase F is due to the lack of recognition of the peripheral carbohydrate moiety of the various asparagine-linked glycans. However, the binding of the enzyme to the di-*N*-acetylchitobiose core is a highly discriminating process: fucose linked α1,3 to the asparagine-proximal *N*-acetylglucosamine residue completely blocks PNGase F activity (Tretter et al., 1991), but asparagine-linked oligosaccharides with the more common α1,6-fucose substituent are excellent substrates for the enzyme. PNGase F appears to be related in activity to glycosylasparaginases which occur in both mammalian lysosomes (Tollersrud & Aronson, 1992) and bacterial cells (Tarentino & Plummer, 1993b).

Very few details regarding the mechanism of action and residues involved in activity are known from previous biochemical studies. Much of the preliminary data is based on analogy with enzymes that have related substances: glycosylasparaginase (Tollersrud & Aronson, 1992) and L-asparaginase (Swain et al., 1993). Glycosylasparaginases are heterodimeric α,β proteins that are similar in size to PNGase F and that deglycosylate asparagine-linked glycans by a similar mechanism but without the requirement for polypeptide linkage. The amino-terminal residue of the β-chain is a threonine that is conserved in all glycosylasparaginases and is absolutely required for activity (Fisher et al., 1993). L-Asparaginase catalyzes the hydrolysis of L-asparagine to L-aspartate and ammonia. The crystal structure of *Escherichia coli* L-asparaginase (Swain et al., 1993) has been determined and shows an active site that includes a triad consisting of an aspartic acid, a threonine, and a lysine. A histidine residue may also be involved in either the activity or the substrate recognition of PNGase F because the enzyme can be inactivated by incubation with low-concentration diethyl pyrocarbonate (Tarentino and Plummer, unpublished results), which alkylates histidine (Li et al., 1993).

The crystal structures of several oligosaccharide-releasing enzymes are under investigation in this laboratory with the goal of understanding their substrate specificities and mechanisms of action (Van Roey et al., 1994a,b). Here we report the crystal structure of PNGase F at 2.2-Å resolution and

[†] Supported in part by Grants GM-50431 (P.V.R.) and GM-30471 (A.L.T. and T.H.P.) from the National Institutes of Health, USPHS.

[‡] The crystallographic coordinates have been deposited with the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973, under the file name 1PNG.

* To whom correspondence should be addressed at Wadsworth Center, New York State Department of Health.

[§] Wadsworth Center.

[∥] Physics Department, University at Albany.

[⊥] Department of Biomedical Sciences, School of Public Health.

[⊛] Abstract published in *Advance ACS Abstracts,* September 1, 1994.

[1] Abbreviations: PNGase F, peptide-N^4-(*N*-acetyl-β-D-glucosaminyl)-asparagine amidase F; rmsd, root-mean-square deviation.

0006-2960/94/0433-11699$04.50/0

Table 1: Statistics on the Native and the Heavy Atom Derivative Data

compound	native	$Pt(en)Cl_2$[a]	$K_2UO_2F_5$	K_2OsCl_6	$(NH_4)_3IrCl_6$
soaking concn (mM)		1	14	10	10
soaking time (days)		14	7	10	16
data collection statistics					
resolution (Å)	2.20	3.00	3.30	3.20	2.60
complete (%)	98.1	95.2	84.5	95.8	92.6
observed	86222	52219	24628	41176	32550
unique	21034	8637	5808	7183	8371
R_{sym}[b]	0.056	0.093	0.097	0.110	0.090
phasing statistics					
resolution used (Å)		3.1	3.3	3.7	3.1
data used		7040	5200	4404	6816
R_{iso}[c]		0.104	0.123	0.115	0.112
R_{Cullis}[d]		0.586	0.552	0.615	0.572
R_{Kraut}[e]		0.075	0.083	0.078	0.065
total mean figure of merit		0.64			
phasing power[f]		1.51	1.70	1.37	1.70
binding site		Met255	Asp139 Lys302	Glu206	Glu206

[a] en = ethylenediamine. [b] $R_{sym} = \Sigma|I_i - \langle I\rangle|/\Sigma\langle I\rangle$, where I = observed intensity and $\langle I\rangle$ = average intensity obtained from multiple observations of symmetry-related reflections. [c] $R_{iso} = \Sigma|I_{PH} - I_P|\Sigma(I_{PH} + I_P)$. [d] $R_{Cullis} = \Sigma|(F_{PH} \pm F_P) - F_H|/\Sigma|F_{PH} - F_P|$, where F_{PH} = heavy atom derivative structure factor amplitude and F_P = protein structure factor amplitude. [e] $R_{Kraut} = \Sigma|F_{PH} - F_{PH(calc)}|/\Sigma|F_{PH}|$. [f] Phasing power = root-mean-square ($|F_H|/E$), where F_H = heavy atom structure factor amplitude and E = residual lack of closure.

present preliminary information about possible locations of the active site.

MATERIALS AND METHODS

Crystallization and X-ray Data Collection. PNGase F was purified from the culture filtrate of *F. meningosepticum* (Tarentino et al., 1985) and crystallized (Kuhn et al., 1994) as previously reported. The crystals belong to space group $C222_1$ with cell dimensions of a = 87.16 Å, b = 125.10 Å, c = 79.33 Å, and V = 864 991 Å^3. There are eight molecules in the unit cell, one per asymmetric unit. With a M_r of 34 779, this yields a V_m = 3.13 Å^3/Da, which corresponds to a solvent content of about 60% (Matthews, 1968).

Diffraction data were measured on a Rigaku R-Axis IIc image-plate detector with Cu Kα radiation generated by a Rigaku RU-200 rotating-anode source. The native data set was measured from one crystal. A total of 95 693 observations were merged to yield 22 074 unique reflections ($F > 0$) with an R_{sym} of 0.056, where $R_{sym} = \Sigma|I_i - \langle I\rangle|/\Sigma\langle I\rangle$. The data are 98.1% complete for all 21 034 data to 2.2-Å resolution and 93.0% for the 19 017 reflections with $F^2/\sigma(F^2) > 1$.

Heavy atom derivatives were prepared by soaking crystals in 1-mL solutions of stabilizing buffer containing the heavy atom compounds at concentrations of 1–20 mM. Soaking experiments in the original crystallization buffer of 0.3 M $(NH_4)_2SO_4$ and 30% PEG 3350 did not yield any results. This problem was resolved by changing the concentrations for the stabilizing buffer to 0.4 M $(NH_4)_2SO_4$ and 25% PEG 3350, increasing the solubility of the heavy atom compounds. Soaking times and concentrations of the heavy atom preparations and the X-ray data collection statistics are summarized in Table 1.

Structure Determination. The structure was determined by multiple isomorphous replacement methods using four derivatives. All calculations were carried out with the program package PHASES (Furey et al., 1990). The Patterson maps of all four derivatives showed single heavy atom sites. Cross-Fourier calculations confirmed the heavy atom positions and the lack of significant secondary sites. Phasing statistics are summarized in Table 1. An initial electron density map was calculated at a resolution of 3.2 Å and displayed with the program CHAIN (Sack, 1988). This map allowed clear identification of protein and solvent regions and showed continuous density for much of the protein. Phases were improved by solvent flattening (Wang, 1985) and phase extension to 3.1 Å as implemented in PHASES. The mask separating the protein and solvent regions was constructed on the basis of a solvent fraction of 0.60. The resulting statistics for 8093 phased reflections were as follows: map inversion R-factor = 0.261 ($R = \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively), F_o/F_c correlation coefficient = 0.951, and mean figure of merit = 0.806. The solvent-flattened map was of good quality and revealed connectivity for 310 of the 314 residues. The density of most of the side chains was clear in the map. After the tracing of the backbone, the primary sequence of PNGase F (Tarentino et al., 1990) could be assigned, and the three disulfide bridges in the protein (Cys51–Cys56, Cys204–Cys208, Cys231–Cys252) were readily identified.

Refinement and Validation of the Model. The model was refined at a resolution of 2.2 Å with the program suite XPLOR (Brünger et al., 1987). The first two rounds of refinement were carried out with simulated annealing using the slow-cooling protocol (Brünger, 1988) at resolutions of 2.8 and 2.5 Å, respectively. Starting at an initial temperature of 3000 K and performing 50 steps of molecular dynamics (time step of 0.5 fs), we then reduced the temperature by 25 K and repeated the procedure until the temperature reached 300 K. The initial R-factor for 12 632 data between 5.0- and 2.5-Å resolution with $F > 2\sigma(F)$ dropped from 0.454 to 0.242. This procedure was followed by alternating cycles of standard positional and individual thermal parameter refinement including data to a resolution of 2.2 Å. New $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps were calculated, and the model was interactively corrected with the program CHAIN after each round of XPLOR refinement. After the first round of refinement at a resolution of 2.2 Å (R = 0.213), water molecules were included if the $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps showed density within hydrogen-bonding distance to protein atoms. Water molecules whose B-factors refined to values greater than 65 Å^2 were considered noise and rejected. The statistics for the current model after four rounds of refinement are as follows: R = 0.183 for data with $F > 2\sigma(F)$, rmsd bonds = 0.013 Å, and rmsd angles = 1.97°. Figure 1 shows the $2|F_o| - |F_c|$ map for the region of sheet II in domain B. The present model includes all but the four amino-terminal residues and 222 water molecules. The amino

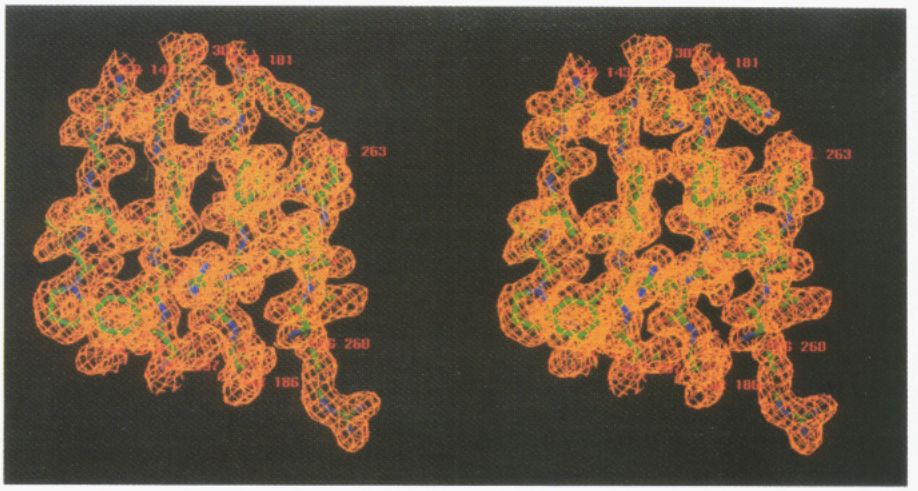

FIGURE 1: $2|F_o| - |F_c|$ map of sheet II in domain B contoured at 1.0σ. The molecular figures have been prepared with the program SETOR (Evans, 1993).

Table 2: Statistics on the Current Refined Model[a]

resolution	8.0–2.2 Å	B_{av} (main chain)	25.2 Å^2
R-factor	0.183 ($F > 2\sigma(F)$)	B_{av} (side chain)	27.9 Å^2
	0.190 ($F > 0$)	bond length msd	0.013 Å
no. of reflections	19001 ($F > 2\sigma(F)$)	bond angle msd	1.97°
used	20065 ($F > 0$)	dihedral msd	27.5°
protein atoms	2989 (2433 non-hydrogen)	impropers rmsd	1.79°
water molecules	222		

[a] $R = \sum_{hkl}||F_o| - |F_c||/\sum_{hkl}|F_o|$, where F_o and F_c are the observed and calculated structure factor. rmsd = root-mean-square deviation from ideality.

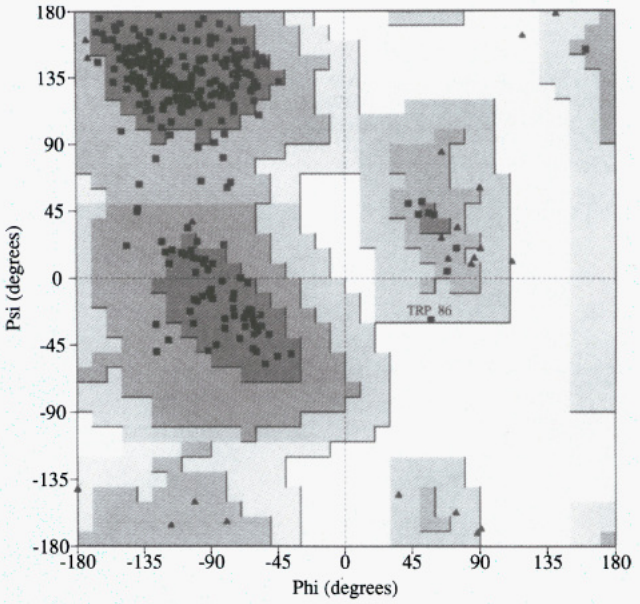

FIGURE 2: Ramachandran plot of PNGase F including 310 of the 314 residues. The plot was produced with the program PROCHECK (Laskowski et al., 1993). The different gray shadings represent the population density of the φ,ψ angles as calculated on the basis of 462 protein structures (Morris et al., 1992). Decreasing probability levels are indicated by a decreasing gray scale: core (dark gray), allowed (gray), generous (pale gray), and disallowed (white). Glycine residues are shown as triangles. Trp86 is the only nonglycine residue in the generously allowed region.

terminus is known to be ragged, and the residues extend into the solvent area where they are disordered.

RESULTS

Details about the structure determination are given in the Materials and Methods section. The structure was determined by multiple isomorphous replacement methods and refined to an R-factor of 0.183 at a resolution of 2.2 Å using the program XPLOR (Brünger et al., 1987). The current model contains 310 amino acids (of a total of 314) with 2433 non-hydrogen protein atoms and 222 water molecules. The stereochemical parameters of the model were checked with the geometry analysis features of XPLOR and the program PROCHECK (Laskowski et al., 1993). The refinement and stereochemical statistics for the current model are listed in Table 2. All other parameters listed by PROCHECK are well within the acceptable limits. Figure 2 shows the Ramachandran plot (Ramachandran et al., 1968) of the refined model. A total of 88.8% of the residues are in the most favored regions. Only residue Trp86, which is tightly bound in the interior of the molecule, has a generously allowed main-chain conformation. Residues Pro196, Ala205, and Pro241 have *cis*-peptide conformations.

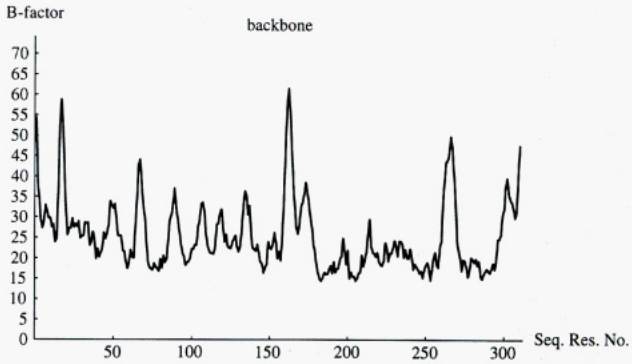

FIGURE 3: Temperature factor plot showing the average B-factor (Å^2) per residue for main-chain atoms.

The average temperature factor (B-factor) is 25.3 Å^2 for the main-chain atoms and 27.9 Å^2 for the side-chain atoms. The water molecules have B-factors in the range of 15–65 Å^2 with an average of 42 Å^2. The B-factors for the main-chain atoms are somewhat higher than observed for most structures. This can probably be attributed to rather loose crystal packing. The crystals tend to be very fragile and there are a few intermolecular contacts. Only 25% of the total surface area of the molecule is inaccessible to a probe with a radius of 1.6 Å. Figure 3 shows a plot of the average B-factor per residue for main-chain atoms. The three loops formed by residues Phe19–Gly22, Gly162–Gly168, and Gly270–Phe273 have the highest B-factors. All three regions are highly exposed to the solvent. The overall main-chain B-factor distribution shows

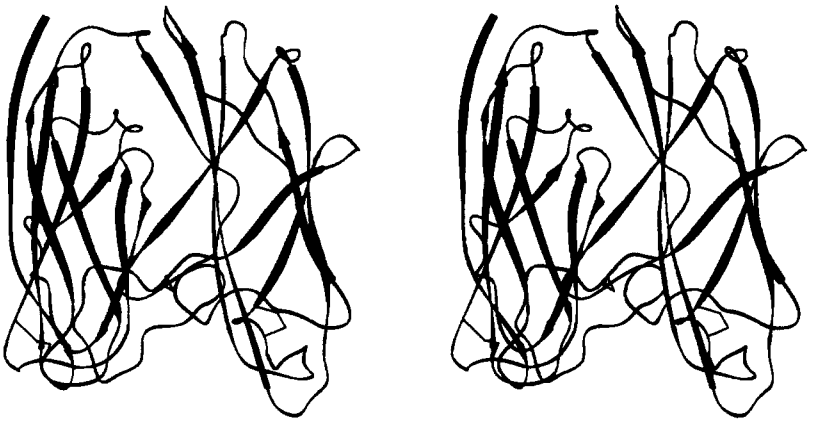

FIGURE 4: Stereo ribbon diagram of PNGase F showing the β-strands as arrows. Domain A is on the left with sheet I (strands 1, 8, 3, and 6) in front and sheet II (strands 2, 7, 4, and 5) in the back. Domain B is on the right with sheet I on the left and sheet II on the right. The covalent link between the domains is on top of the molecule. The three disulfide bridges at the bottom (left, right, and middle of the molecule) are shown as solid lines. The description of the structure is based on the orientation shown in this figure.

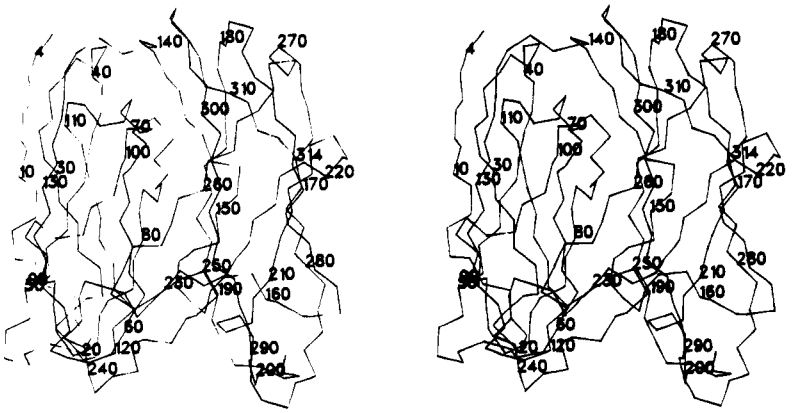

FIGURE 5: Backbone structure of PNGase F showing all α-carbon atoms with every 10th α-carbon and N-terminal and C-terminal residues labeled.

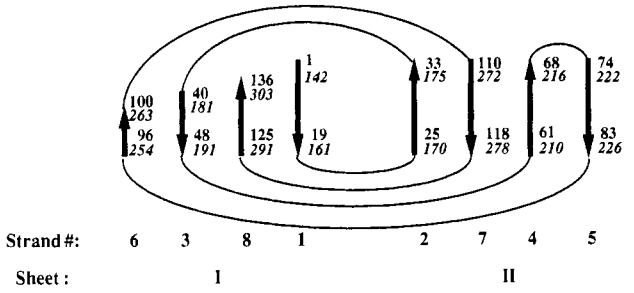

FIGURE 6: Schematic diagram showing the jelly-roll topologies of domains A and B. All β-strands are numbered, while residue numbers at the beginning and end of each strand are indicated. Residue numbers for domain B are printed in italics. The strands in sheets II remain constant in size while those of sheets I decrease in size in their spatial order 1-8-3-6. Only the last six residues of the β-strand formed by residues 254–263 are part of domain B (strand 6, sheet I). Residues 254–257 are in hydrogen-bonding contact with strand II-5 of domain A. This β-strand forms a crossover between the two domains.

an almost continuous decrease from the top of the molecule to the bottom.

The folding of PNGase F is shown as a ribbon diagram in Figure 4 and an α-carbon tracing in Figure 5. The view shown in these figures corresponds to the orientation used for the discussion below. The molecular fold consists of two eight-stranded antiparallel β-sandwich domains, A and B, including residues 1–137 and 143–314, respectively. The two domains lie side by side with parallel principal axes so that the twist axes (Murzin et al., 1994) of all four β-sheets are in one plane. The twist axes of the sheets in domain A make an angle of about 50° with those of domain B. Residues 138–142 form the only covalent link between the domains and have an extended conformation. This link is located at the top end of the molecule.

The two domains have identical jelly-roll topologies (Richardson, 1981) as shown in Figure 6. The spatial order of the eight β-strands in each sandwich is I-1 I-8 I-3 I-6 and II-2 II-7 II-4 II-5, where I and II refer to the two sheets in one sandwich. Figure 7 shows the two domains superimposed onto each other, illustrating their similar geometries. The C_α atoms of the 51 residues in the core of the two β-sheets in domain A were least-squares fitted to the corresponding C_α atoms in domain B. The average displacement between these C_α atoms is 1.24 Å. The similarity in structure includes all β-sheets and the turns on top of the molecule but breaks down at the bottom. The loops at the bottom of domain B are larger than in domain A, extending the length of the domain by about 10 Å.

There is a difference in the size of the β-strands in the two sheets of one domain that is consistent between the two domains. The number of residues per β-strand in their spatial order decreases in sheets I but remains constant in sheets II. The β-strands of sheet I contain 5–15 residues in domain A and 10–20 residues in domain B. The β-strands of sheets II have 9 residues in domain A and 7 residues in domain B. The distance between the two sheets in each sandwich is about 10 Å. The surface of domain A is highly polar and contains many charged residues. Its interior contains only hydrophobic residues, especially phenylalanines. Sandwich B is more hydrophobic on the surface but less hydrophobic in the interior.

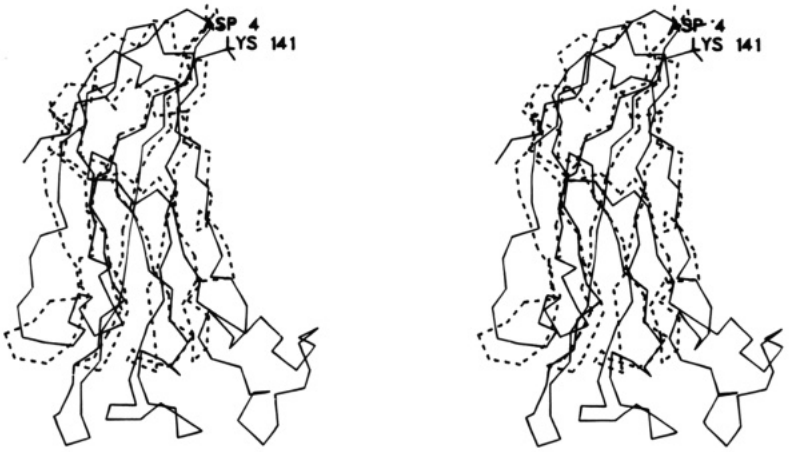

FIGURE 7: Stereo diagram showing the superposition of the α-carbon tracing of domains A (dashed line) and B (solid line) generated by least-squares fitting of 51 pairs of α-carbon atoms of the β-strands. The links between the β-strands on top of the molecule maintain similar conformations in both domains. The links at the bottom are much longer in domain B than in domain A.

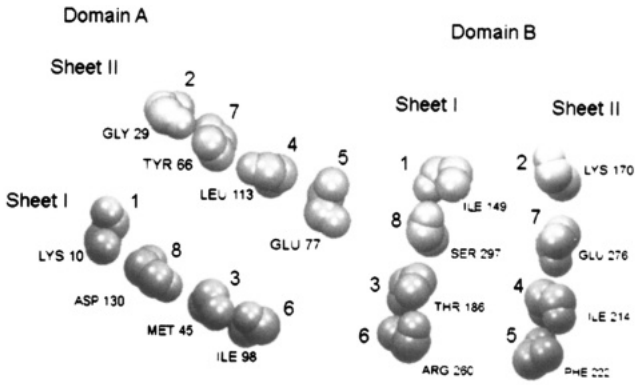

FIGURE 8: Cross section through the middle of the molecule, approximately at the height of the twist axes of the β-sheets, showing the relative locations and orientations of the two sandwiches.

In the center core, residues Ser297 and Thr186 of sheet I form hydrogen contacts with Tyr279 of sheet II.

The geometries of the links between the β-strands on top of the molecule are very similar in both domains. They are all shorter than ten residues. A nine-residue loop with no intraloop main-chain hydrogen bonds connects strand II-2 to strand I-3. The turn connecting strand II-4 to strand II-5 is a three-residue tight turn as defined by Richardson (1985). The connection between strands I-6 and II-7 is formed by two interlocked tight turns with 3_{10}-helical conformation. Following β-strand I-8, an additional loop crosses from sheet I to sheet II. The carbonyl terminus, residues 312–314, has a β-strand conformation and forms a small extension of sheet II, adjacent to strand II-2.

The conformations of the links at the bottom of the molecule are different in the two domains. The loops in domain B are longer than in domain A. Both links between strands I-3 and II-4 contain disulfide bridges in identical positions but with very different loop lengths: three residues in domain A and 20 residues in domain B. The longest link in the structure occurs between strands II-5 and I-6 of domain B. It contains 27 residues from Gly227 to Pro253. The most remarkable structural feature is an Ω-loop (Leszczynski, 1986), which is formed by residues Cys231–Pro246 with a distance of 3.6 Å between the termini. The side chain of Trp244 fills the interior of this loop.

Figure 8 shows a cross section through the molecule at the height of the twist axes, illustrating the relative orientations and positions of the two sandwiches. The interface is formed mainly by strand II-5 of domain A making contacts with strands I-1, I-8, and I-3 of domain B and by strand I-6 of domain A making contacts with strand I-6 of domain B. The closest contact, between strand II-5 of domain A and strand I-8 of domain B, is about 8 Å. There are no main-chain to main-chain hydrogen bonds between the β-strands of the two sandwiches. However, residues 254–257, which are the beginning of strand I-6 of domain B, are hydrogen-bonded to residues 85–80 of strand II-5 in domain A in what appears to be a distorted extension to sheet II of this domain. The interface contains mainly polar residues, which form extensive contacts along the full length of the interface. The interdomain hydrogen-bonding contacts are listed in Table 3.

Table 3: Hydrogen Bonding between Domain I and Domain II

residue no.	acid	atom	residue no.	acid	atom	distance (Å)	domain-sheet-strand	
62	Tyr	OH	152	Asn	O	2.67	A-II-4	B-I-1
64	Asn	ND2	152	Asn	O	3.12	A-II-4	B-I-1
66	Tyr	OH	152	Asn	O	3.38	A-II-4	B-I-1
75	Trp	O	150	Gln	NE2	3.01	A-II-5	B-I-1
77	Glu	N	150	Gln	OE1	3.11	A-II-5	B-I-1
77	Glu	OE2	152	Asn	ND2	3.00	A-II-5	B-I-1
77	Glu	OE2	296	Ser	OG	2.80	A-II-5	B-I-8
78	Ile	O	185	Arg	NH2	2.97	A-II-5	B-I-3
80	Arg	NH2	187	Thr	O	2.78	A-II-5	B-I-3
80	Arg	NH1	189	Ser	OG	2.97	A-II-5	B-I-3
80	Arg	NH1	294	Ala	O	2.76	A-II-5	B-I-8
80	Arg	O	257	Val	N	2.86	A-II-5	B-I-6
82	Ile	O	254	Gly	N	2.78	A-II-5	B-I-6
82	Ile	N	255	Met	O	3.08	A-II-5	B-I-6
82	Ile	O	255	Met	N	3.29	A-II-5	B-I-6
83	Thr	O	254	Gly	N	3.34	A-II-5	B-I-6
85	Tyr	N	239	Gln	OE1	2.77	A-II-5	B-Ω-loop
99	Asp	O	185	Arg	NH1	3.36	A-I-6	B-I-3
102	Asp	OD1	183	Tyr	OH	2.57	A-I-6	B-I-3
102	Asp	OD2	185	Arg	NH1	2.85	A-I-6	B-I-3

The alignment of the domains and the interaction between them result in a bowl on the front surface of the molecule that is formed by strands I-6 and II-5 of domain A and strands I-3, I-6, and II-5 of domain B. The lower end of this bowl is formed by the long link between strands II-5 and I-6 of domain B, which contains the Ω-loop and the contact area, residues 254–257, with domain A. Additional features at the interface are a long cleft along the full length of the molecule on the back surface, formed by sheet II of domain A and sheet I of domain B, and a wide U-shaped cleft at the bottom of the molecule.

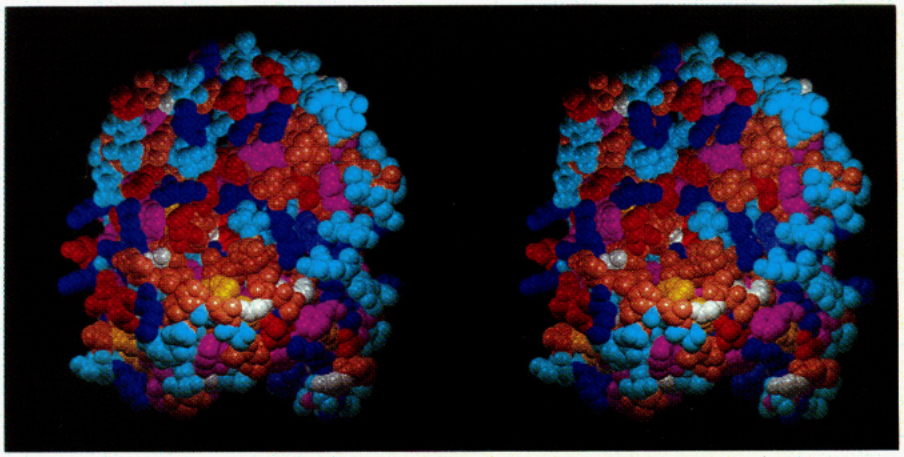

FIGURE 9: Stereo van der Waals figure, in the same orientation as in Figure 4, showing the bowl in the front face of the molecule. The color code for the residues is as follows: violet, Phe, Trp, and Tyr; red, Asp and Glu; cyan, Ser, Thr, Asn, and Gln; blue, Arg, Lys, and His; brown, Pro, Val, Ala, Ile, and Leu; yellow, Cys and Met; gray, Gly. The lower half of the bowl is mainly hydrophobic, and the upper part is polar.

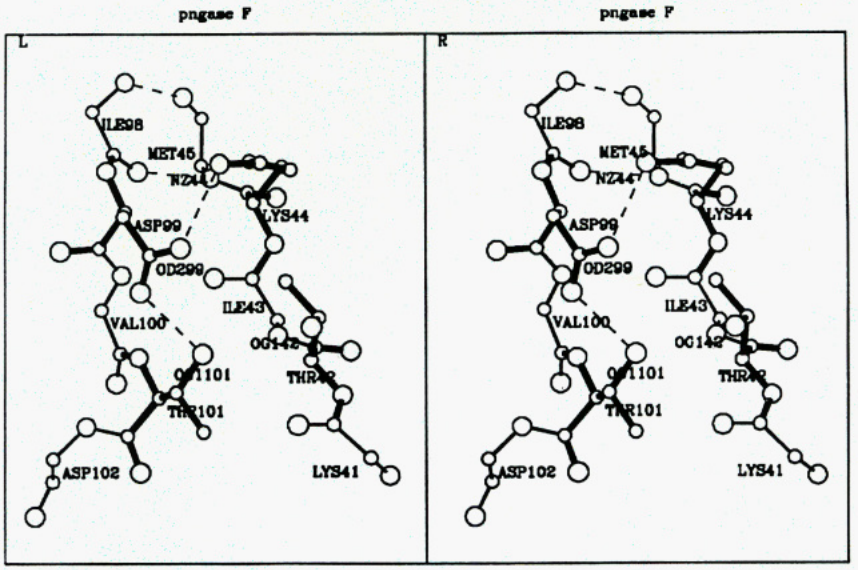

FIGURE 10: Ball-and-stick stereo figure of the possible active-site residues Thr42, Thr101, Asp99, and Lys44. The arrangement of these residues, consisting of a hydrogen-bonded triad of Thr101, Asp99, and Lys44 and an adjacent threonine (Thr42), resembles the geometry of the active site of L-asparaginase. This figure has been prepared with the program PLUTO as implemented in the CCP4 program suite (CCP4, 1979).

DISCUSSION

Examination of the molecular structure of PNGase F does not uniquely identify a single active-site area but suggests three areas that merit further investigation: the bowl on one surface and the clefts on the opposite surface and at the bottom of the molecule. The limited biochemical data available, the possible relationship to the L-asparaginase active site, and the possible involvement of a threonine and histidine have been taken into account in the following analysis. In addition, the unusual large size and complexity of the substrate put major constraints on the active-site geometry. Major conformational changes upon interaction with the substrate cannot be excluded at this time because of the hydrophilic nature of the contacts in the interface between the two domains and the lack of information regarding the mechanism of action of the enzyme. However, there is some evidence that the interaction between the two domains is very stable. The crystal structure of a different crystal form of PNGase F has been determined independently by Norris et al. (personal communication). This form was obtained under very different crystallization conditions: at pH 8.0 instead pH 4.3 (Norris et al., 1994). The two structures of PNGase F are identical in folding (the rmsd of the main-chain atoms is 0.74 Å), and the geometry of the interface between the two domains is identical. Therefore, major movement of the two domains is not likely and was not taken into consideration in the following analysis.

The first potential active- and binding-site area is the bowl in the front face of the molecule, formed by the solvent-exposed parts of strands II-6 and I-5 of domain A, strands I-3, I-6, and II-5 of domain B, and the upper surface of the loop between strands II-5 and I-6 of domain B. Figure 9 shows a van der Waals model of the bowl, which has a diameter of about 20 Å. The lower half is lined mainly with small hydrophobic residues, and the upper half contains polar and charged residues. This area contains a cluster of three residues, Thr101, Asp99, and Lys44, that are hydrogen-bonded to each other. Figure 10 shows this triad and Thr42, which is also within close contact distance. The arrangement of these four residues is somewhat reminiscent of the L-asparaginase active-site geometry (Swain et al., 1993). However, because of its location in a depression in the middle of the molecule, it is difficult to

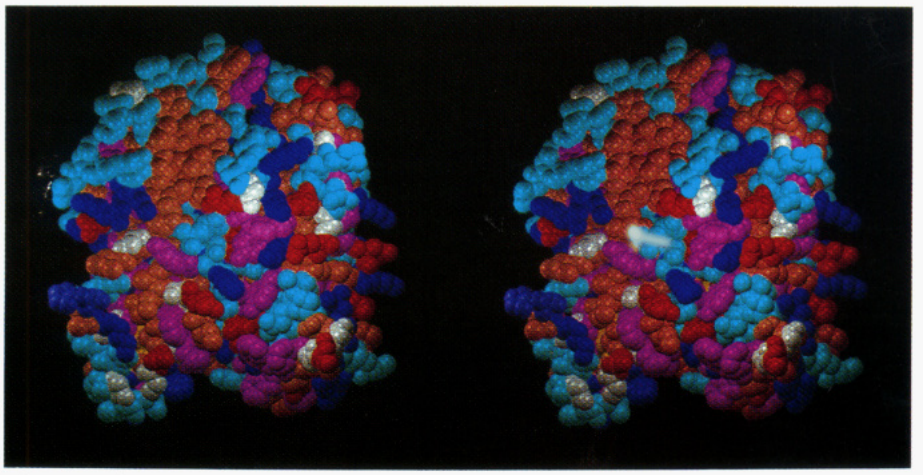

FIGURE 11: Stereo van der Waals figure of the back of the molecule showing the S-shaped cleft. The color code is equivalent to that in Figure 9. The cleft is formed by sheet II of domain A and sheet I of domain B and contains many polar residues.

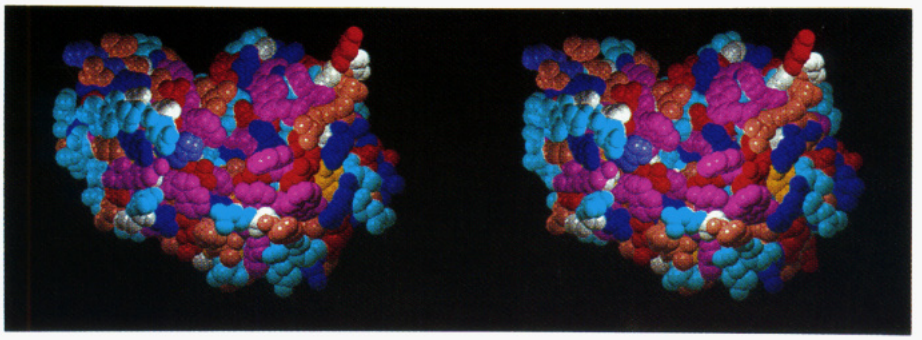

FIGURE 12: Stereo van der Waals figure of the bottom of the molecule showing the wide U-shaped cleft lined almost exclusively by aromatic residues, especially eight tryptophans.

predict how this active site could be accessible for a large substrate such as a glycoprotein.

The second possible site is a shallow S-shaped cleft that runs along the full length of the interface on the back side of the molecule. Figure 11 shows a van der Waals model of the cleft. It is formed by sheet II of domain A and sheet I of domain B. The S-shape of the cleft results from the protrusions formed by the tight β-turn of the link between strands II-4 and II-5 of domain B and the β-bulge formed by residues Tyr151–Pro160 of strand I-1 of domain B. The cleft contains many charged and polar residues, which would be suitable for substrate recognition. However, no obvious active-site area has been recognized at this time.

The third site is a cleft of about 25-Å width and 20-Å depth at the bottom of the molecule. Figure 12 shows a van der Waals model of this cleft. It is formed by the bottom residues of sheet II of domain A and sheet I of domain B and covered in front by the Ω-loop. The most remarkable feature of this cleft is that it contains almost exclusively aromatic residues including eight tryptophan residues (of a total of nine in the molecule), three tyrosine residues, one histidine residue, and three arginine residues. In addition, a hole of about 5 Å in diameter runs into the center of the molecule from the middle of the cleft. Several features of this cleft make it a likely substrate recognition site. These include the aromatic residues, which are known to be involved in oligosaccharide binding, the presence of the histidine residue, and the overall shape, which is suitable for interaction with a large substrate. Although this cleft would be the most logical candidate for the binding site, no obvious active site has been recognized at this time.

These three sites are the main targets for further studies of the activity of PNGase F, including cocrystallization experiments with different substrates or inhibitors and site-directed mutagenesis.

REFERENCES

Brünger, A. T. (1988) *J. Mol. Biol. 203*, 803–816.

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

CCP4 (1979) *The SERC (U.K.) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography*, distributed from Daresbury Laboratory, Warrington WA4 4AD, U.K.

Evans, S. E. (1993) *J. Mol. Graphics 11*, 134–137.

Fisher, K. J., Klein, M., Park, H., & Aronson, N. N. (1993) *FEBS Lett. 323*, 271–275.

Furey, W., & Swaminathan, S. (1990) American Crystallographic Meeting, New Orleans, LA, Abstract PA33.

Kuhn, P., Tarentino, A. L., & Van Roey, P. (1994) *J. Mol. Biol. 241*, 622–623.

Laskowski, R. A., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283–291.

Leszczynski, J. F., & Rose, G. D. (1986) *Science 234*, 849–855.

Li, C., Moore, D. S., & Rosenberg, R. C. (1993) *J. Biol. Chem. 268*, 11090–11096.

Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491–497.

Morris, A. L., Macarthur, M. W., & Thornton, J. M. (1992) *Proteins 12*, 345–364.

Murzin, A. G., Lesk, A. M., & Chothia, C. (1994) *J. Mol. Biol. 236*, 1369–1381.

Norris, E. N., Flaus, A. J., & Baker, E. N. (1994) *J. Mol. Biol.* (in press).

Plummer, T. H., Jr., Elder, J. H., & Tarentino, A. L. (1984) *J. Biol. Chem. 259*, 10700–10704.

Ramachandran, G. N., & Sasisekharan, V. (1968) *Adv. Protein Chem. 28*, 283–437.
Rasmussen, J. R., Davis, J., & Van Etten, R. (1992) *J. Am. Chem. Soc. 114*, 1124–1126.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–330.
Risley, J. M., & Van Etten, R. (1985) *J. Biol. Chem. 260*, 15488–15494.
Sack, J. S. (1988) *J. Mol. Graphics 6*, 244–245.
Swain, A. L., Jaskolski, M., & Wlodawer, A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1474–1478.
Tarentino, A. L., & Plummer, T. H., Jr. (1993a) *Trends Glycosci. Glycotechnol. 5*, 163–170.
Tarentino, A. L., & Plummer, T. H., Jr. (1993b) *Biochem. Biophys. Res. Commun. 197*, 179–186.
Tarentino, A. L., Gomez, C. M., & Plummer, T. H., Jr. (1985) *Biochemistry 24*, 4665–4671.
Tarentino, A. L., Quinones, G., & Plummer, T. H. (1990) *J. Biol. Chem. 265*, 6961–6966.
Tollersrud, O. K., & Aronson, N. N., Jr. (1992) *Biochem. J. 282*, 891–897.
Tretter, V., Altman, F., & März, L. (1991) *Eur. J. Biochem. 199*, 647–651.
Van Roey, P., Silva, G. H., & Guan, C. (1994a) *J. Mol. Biol. 294*, 157–159.
Van Roey, P., Rao, V., & Tarentino, A. L. (1994b) American Crystallographic Meeting, Atlanta, GA, Abstract PGL02.
Wang, B. C. (1985) *Methods Enzymol. 115*, 90–112.