Crystal Structure of Endo-β-*N*-acetylglucosaminidase F_1, an α/β-Barrel Enzyme Adapted for a Complex Substrate[†,‡]

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

Wadsworth Center, New York State Department of Health, Empire State Plaza, P.O. Box 509, Albany, New York 12201-0509, and Department of Biomedical Sciences, School of Public Health, and Department of Physics, University at Albany, Albany, New York 12222

Received June 23, 1994; Revised Manuscript Received August 22, 1994[⊗]

ABSTRACT: Endo-β-*N*-acetylglucosaminidase F_1 (Endo F_1) is an endoglycosidase, secreted by *Flavobacterium meningosepticum,* that cleaves asparagine-linked oligosaccharides after the first *N*-acetylglucosamine residue. The enzyme is selective for high-mannose oligosaccharide chains. The crystal structure of Endo F_1 has been determined at 2.0-Å resolution. The molecular fold consists of a highly irregular α/β-barrel, a commonly observed motif consisting of a cyclic 8-fold repeat of β-strand/loop/α-helix units with an eight-stranded parallel β-barrel at the center. Endo F_1 lacks two of the α-helices, those of units 5 and 6. Instead, the links after β-strands 5 and 6 consist of a short turn followed by a section in an extended conformation that replaces the helix and a long loop at the bottom of the molecule. The absence of any excursion on top of the molecule following β-strands 5 and 6 results in a pronounced depression in the rim of the barrel. This depression forms one end of a shallow cleft that runs across the surface of the molecule, over the core of the β-barrel to the area between the loops of units 1 and 2. The active site residues, Asp130 and Glu132, are located at the carboxyl end of β-strand 4 and extend into this cleft. These residues are surrounded by several tyrosine residues. The cleft area formed by loops 1 and 2 is lined with polar residues, mainly asparagines. The latter area is thought to be responsible for oligosaccharide binding and recognition while the protein moiety of the substrate would be located outside the molecule but adjacent to the area of loops 5 and 6. The absence of the α-helices in this area results in a narrower barrel rim with a more adaptable surface, which is thought to be important to facilitate interaction with many different glycoprotein substrates.

Endo-β-*N*-acetylglucosaminidase F_1 (Endo F_1)[1] is one of three endoglycosidases secreted by *Flavobacterium meningosepticum* (Plummer & Tarentino, 1991; Trimble & Tarentino, 1991). These enzymes hydrolyze the β-D-(1−4) bond between the two *N*-acetylglucosamine residues (GlcNAc) of the *N,N'*-diacetylchitobiose core of asparagine-linked oligosaccharides such that one GlcNAc remains with the protein and the other forms the reducing end of the released oligosaccharide. Endo F_1 is homologous to Endo H, a related enzyme secreted by *Streptomyces plicatus* that is extensively used as a tool in the study of glycoproteins (Tarentino et al., 1974, Trimble & Maley, 1984). Endo H and Endo F_1 have 32% sequence identity and have similar substrate specificity (Tarentino et al., 1992). Both enzymes are specific for high-mannose and hybrid oligosaccharide substrates. The minimum required substrate for activity is the core: Manα(1−3)Manα(1−6)Manβ(1−4)GlcNAcβ(1−4)-GlcNAc. In particular, the α(1−3)-linked mannose is absolutely required for enzymatic activity. The main difference in activity between Endo H and Endo F_1 is that a fucose, α(1−6) linked to the first GlcNAc residue, lowers the susceptibility of the substrate for processing by Endo F_1 but does not affect the activity of Endo H. The other endoglycosidases secreted by *F. meningosepticum*, Endo F_2 and Endo F_3, are less homologous to Endo H than Endo F_1 and differ in their substrate specificity (Tarentino et al., 1993). Endo F_2 prefers complex biantennary oligosaccharides, but high-mannose glycans are hydrolyzed, albeit at a much lower rate. Endo F_3 has evolved to accommodate both bi- and triantennary oligosaccharides and prefers these structures with a core-bound fucose substituent (Tarentino & Plummer, 1994).

Under nondenaturing conditions, Endo H and Endo F_1 process many but not all high-mannose asparagine-linked oligosaccharides. Inability to process some substrates appears to result from interference with binding by the protein moiety of the substrate because denaturation of the substrate renders all asparagine-linked high-mannose oligosaccharide chains susceptible to hydrolysis. This suggests that, although the endoglycosidases can process oligosaccharides that are not protein linked, nonspecific enzyme−protein interactions are likely to occur and affect activity.

The analysis of sequence data for these endoglycosidases (Tarentino et al., 1992) and for a class of bacterial chitinases that hydrolyze the β(1−4)-glycosidic link between GlcNAc residues (Kuranda & Robbins, 1991) has resulted in the

[†] Supported in part by Grants GM-50431 (P.V.R.) and GM-30471 (T.H.P. and A.L.T.) 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 2EBN.

* Author to whom correspondence should be addressed.

[§] Wadsworth Center.

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

[⊥] Department of Physics.

[⊗] Abstract published in *Advance ACS Abstracts,* October 15, 1994.

[1] Abbreviations: Endo, Endo-β-*N*-acetylglucosaminidase; fom, figure of merit; GlcNAc, *N*-acetylglucosamine; Man, mannose; MIR, multiple isomorphous replacement; rms, root mean square; TIM, triose-phosphate isomerase.

0006-2960/94/0433-13989$04.50/0 © 1994 American Chemical Society

recognition of several sections with high sequence homology and identification of several conserved residues. The most important homologous section includes residues 123−132, and contains five residues that are conserved among all endoglycosidases, Leu124, Asp125, Gly126, Asp130, and Glu132 (sequence numbering of Endo F_1 and Endo H). These homology studies and the mutagenesis studies of Watanabe et al. (1993), using the chitinase of *Bacillus circulans*, and of Schmidt et al. (1994), using Endo H, have resulted in the identification of residues Asp130 and Glu132 as the putative catalytic residues. The fact that the catalytic residues are nearly adjacent in the sequence is unusual for glycohydrolases. A similar configuration is observed only in bacterial (1,3-1,4)-β-D-glucan 4−glucanohydrolases (Juncosa et al., 1994) but not for the plant analogues (Chen et al., 1993; Varghese et al., 1994), lysozymes, and amylases. This suggests that the bacterial endoglycosidases, chitinases, and glycan hydrolases share a similar mechanism for activity and substrate recognition that is different from that of other related enzymes.

Endo F_1, Endo F_2, and Endo F_3 provide a unique opportunity to correlate protein structure with well-defined oligosaccharide specificity. In addition, the complexity of the glycoprotein substrates also makes these enzymes of special interest for structural analysis. In this paper we report the crystal structure determination and refinement at 2.0-Å resolution of Endo F_1 and propose a model for substrate recognition and binding.

MATERIALS AND METHODS

Purification and Crystallization. Endo F_1 was purified from the culture filtrate of *F. meningosepticum* (Plummer & Tarentino, 1991) and crystallized by hanging drop vapor diffusion methods as previously described (Van Roey et al., 1994). The crystals belong to the hexagonal space group $P6_1$ and have unit cell dimensions of $a = b = 70.64$ Å, $c = 100.31$ Å, and $V = 433\,487$ Å^3. Assuming one molecule per asymmetric unit, this yields a $V_M = 2.41$ Å^3/D, corresponding to a solvent content of approximately 45% (Matthews, 1968). The crystals are hexagonal bipyramids that routinely grow to dimensions of about 0.4−0.7 mm. Diffraction to 1.6-Å resolution has been observed.

Data Collection. X-ray diffraction data were measured on a Rigaku R-axis IIc image plate area detector using an RU-200 rotating anode source with graphite monochromated Cu Kα radiation. The crystal used for the native data collection had approximate dimensions of 0.3 × 0.3 × 0.5 mm. Data to 2.0-Å resolution were measured using a single crystal. In total 141 175 data were measured, which were averaged to 19 158 unique data [$I/\sigma(I) > 1.0$] with an R_{mer} of 0.0296. The data are 93.6% complete. Derivative data were measured to 2.2- or 2.5-Å resolution depending on the quality of the crystals. The statistics for the derivative data are summarized in Table 1.

Structure Determination. The structure was determined by multiple isomorphous replacement (MIR) methods using five derivatives and the anomalous scattering data of two of them. The program package PHASES (Furey & Swaminathan, 1990) was used for the heavy atom position determination and refinement and for the phasing. All derivatives were prepared by soaking crystals in 5-mL solutions of precipitation buffer containing the heavy atom compounds at the concentrations listed in Table 1. The heavy atom derivatives included two mercury compounds, $HgCl_2$ and K_2HgI_4, and two platinum compounds, K_2PtCl_4 and $K_2Pt(NO_2)_4$. The fifth derivative was a double derivative, with K_2HgI_4 and K_2PtCl_4. The space group ambiguity ($P6_1$ or $P6_5$) was resolved by the inclusion of the anomalous data in the phasing. Table 2 lists the MIR phasing statisitics, and Table 3 lists the refined coordinates of the heavy atoms. All derivatives had multiple heavy atom binding sites, but the two Hg compounds had one site in common and the two Pt derivatives shared two sites. The double derivative had the highest Hg and the top two Pt sites. The thermal parameters of the heavy atoms were not refined because of their high correlation with the occupancy factors. The refined occupancy factors were maintained on an arbitrary scale. The structure was initially phased at 2.5-Å resolution, although only three of the derivatives contributed to the

Table 1: Statistics of the Heavy Atom Derivative Data Sets for Endo F_1

derivative	concn (mM)	time (days)	resolution (Å)	R_{sym}[a] (on F)	completeness (%)	$\langle F/\sigma(F)\rangle$ last shell	R_{mer}[a] (on F)
$HgCl_2$	1	8	2.2	0.081	98.2	4.16	0.140
K_2HgI_4	2	1	2.2	0.063	97.2	4.92	0.069
K_2PtCl_4	10	15	2.5	0.101	89.0	2.98	0.111
$K_2Pt(NO_2)_4$	10	20	2.2	0.056	95.2	4.13	0.140
K_2HgI_4 + K_2PtCl_4	2 10	2 10	2.2	0.085	91.0	3.57	0.115

[a] $R_{sym} = \Sigma|F - \langle F\rangle|/\Sigma\langle F\rangle$; $R_{mer} = \Sigma|F_{PH} - F_P|/\Sigma F_P$.

Table 2: MIR Phasing Statistics for Endo F_1

derivative	resolution (Å)	phasing power	R_{Cullis}	R_{Kraut}	reflections (fom > 0.25)
$HgCl_2$	2.5	2.16	0.470	0.191	9213
$HgCl_2$ anomalous	3.5	1.27		0.391	3354
K_2HgI_4	2.5	2.46	0.450	0.086	9303
K_2PtCl_4	3.0	1.12	0.611	0.178	5138
$K_2Pt(NO_2)_4$	2.5	1.70	0.532	0.213	9105
$K_2Pt(NO_2)_4$ anomalous	3.0	1.23		0.391	5277
$K_2HgI_4 + K_2PtCl_4$	3.0	1.63	0.596	0.157	5355

Table 3: Refined Heavy Atom Sites for Endo F_1

	atom	x	y	z	B	occupancy
$HgCl_2$	Hg	0.554 94	0.113 07	−0.002 85	25.0	1.241 41
	Hg	0.465 87	0.158 06	0.412 77	25.0	0.728 57
	Hg	0.759 40	0.233 16	0.898 13	25.0	0.807 75
$HgCl_2$ anomalous	Hg	0.553 57	0.113 65	−0.001 85	25.0	0.823 12
	Hg	0.466 90	0.159 54	0.411 71	25.0	0.621 22
	Hg	0.758 92	0.233 63	0.897 21	25.0	0.559 30
K_2HgI_4	Hg	0.552 81	0.112 52	1.001 19	25.0	1.529 35
	I	0.558 96	0.088 87	0.987 21	25.0	1.236 40
K_2PtCl_4	Pt	0.566 89	0.161 18	0.027 45	25.0	1.021 24
	Pt	0.366 96	0.369 87	0.012 33	25.0	1.039 24
	Pt	0.389 04	0.011 51	0.123 21	25.0	0.718 02
$K_2Pt(NO_2)_4$	Pt	0.570 14	0.155 77	0.029 67	25.0	1.354 67
	Pt	0.367 40	0.364 45	0.014 50	25.0	0.945 16
	Pt	0.550 08	0.182 73	0.024 69	25.0	0.920 76
	Pt	0.337 81	0.022 87	0.249 86	25.0	0.867 23
	Pt	0.292 43	0.112 09	0.581 03	25.0	0.645 39
$K_2Pt(NO_2)_4$ anomalous	Pt	0.572 24	0.155 52	0.030 92	25.0	1.178 67
	Pt	0.366 65	0.363 61	0.013 60	25.0	0.934 31
	Pt	0.549 20	0.185 19	0.025 04	25.0	0.793 48
	Pt	0.336 53	0.022 16	0.248 24	25.0	0.763 12
	Pt	0.295 57	0.114 54	0.584 02	25.0	0.553 10
K_2HgI_4 + K_2PtCl_4	Hg	0.556 39	0.114 64	0.000 10	25.0	1.081 17
	Pt	0.575 32	0.155 25	0.027 25	25.0	0.822 23
	Pt	0.367 07	0.373 81	0.007 80	25.0	0.703 54

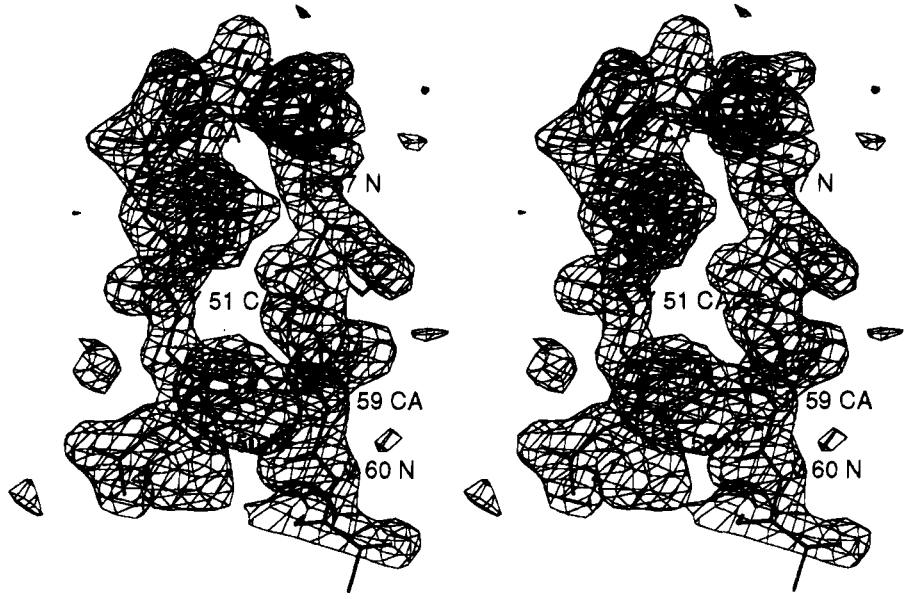

FIGURE 1: Stereo diagram of the $2F_o - F_c$ map, based on the present model, in the area of the loop containing residues 50−60. The map is contoured at the 1σ level. This loop forms the turn of the hairpin loop of the excursion following β-strand 2.

phasing at that resolution. A derivative was used in the phasing only up to the resolution where its phasing power dropped below 1.00. The average figure of merit (fom) after the heavy atom refinement was 0.709 for 9374 data. Density modification, using the procedure of Wang (1985) as implemented in the PHASES package, was used to improve the phases at 2.5-Å resolution. A 45% solvent content estimate was used for calculation of the solvent mask. The phasing statistics after solvent flattening are $R = 0.203$, correlation coefficient = 0.972, and average fom = 0.878.

The resulting electron density map was displayed with the program CHAIN (Sack, 1988) and revealed continuous density for 284 of the 289 residues. Almost all side chains could be readily identified so that no difficulties in making the correct residue and connectivity assignments were encountered. Only the first five residues were missing from this electron density map. The 284-residue model was built in the map and then refined using the simulated annealing procedures of the program package X-PLOR (Brünger et al., 1987; Brünger, 1988).

The simulated annealing refinement brought the R-value for all data with $F > 2\sigma(F)$ at 2.5-Å resolution from 0.403 before refinement to 0.286 after 40 cycles of positional refinement and to 0.217 after heating to 2000 K for 1 ps and slow cooling. At this stage the model was corrected and further refined, using the positional refinement routines of X-PLOR at 2.2-Å resolution. The model was again checked using $2F_o - F_c$ and $F_o - F_c$ maps. Omit maps were also used to examine the areas of weak electron density or unusual molecular geometry. The first set of solvent atoms were included in the model at this stage. Water molecules were included only if peaks in the $2F_o - F_c$ and $F_o - F_c$ maps were observed within hydrogen-bonding distance of oxygen or nitrogen atoms. Solvent atoms were considered noise and removed if their thermal parameters exceeded 65.0 Å^2 at the end of a round of refinement. Further refinement, including individual isotropic thermal parameters, was performed by restrained least-squares methods, using the program PROFFT (Hendrickson & Konnert, 1980; Finzel, 1987). The present model, refined to an R-value of 0.156 for all 18 856 data with $F > 2\sigma(F)$ in the range of 10.0- and 2.0-Å resolution, includes 2210 protein atoms, from all except the first four residues, one Zn^{2+} ion, and 242 water molecules. The root-mean-square deviations from ideality are bond lengths, 0.016 Å; angle distances, 0.033 Å; dihedral distances, 0.045 Å; and planarity, 0.014 Å.

RESULTS

Molecular Structure. The structure of Endo F_1 was determined by multiple isomorphous replacement methods and refined to an R-value of 0.156 at 2.0-Å resolution using restrained least-squares methods. Details of the structure determination and refinement are given in the Materials and Methods section. The present model consists of 2210 protein atoms, from 285 of the 289 residues, one Zn^{2+} ion, and 242 water molecules. Figure 1 shows the electron density map of the loop containing residues 50−60. The four amino-terminal residues are not present in the current model. The amino terminus is known to be ragged in the naturally produced enzyme and extends into the solvent area. Partial occupancy and disorder make it difficult to locate these residues, even at high resolution. Figure 2 shows the Ramachandran plot for the current model. Of the residues other than proline or glycine, 86.9% are in the most favorable region, and only one non-glycine residue, Ser174, is in the generously allowed region. There are two *cis*-peptide bonds, for Pro139 and for Ser46. Figure 3 shows the main-chain thermal parameters. Large thermal parameters, indicative of thermal motion or disorder, are observed in the amino and carboxyl termini and for residues 142−144, which are part of a loop that extends into the solvent area.

The structure of Endo F_1 is shown in Figure 4 as a RIBBON diagram (Priestle, 1988). The basic folding motif of the molecule is that of an irregular TIM-type α/β-barrel. The TIM-barrel motif is named for triose-phosphate isomerase, the first protein which was observed to have this fold (Banner et al., 1975). The TIM barrel consists of a cyclic 8-fold repeat of a β-strand/loop/α-helix unit. The β-strands form a central eight-stranded parallel β-barrel that is surrounded

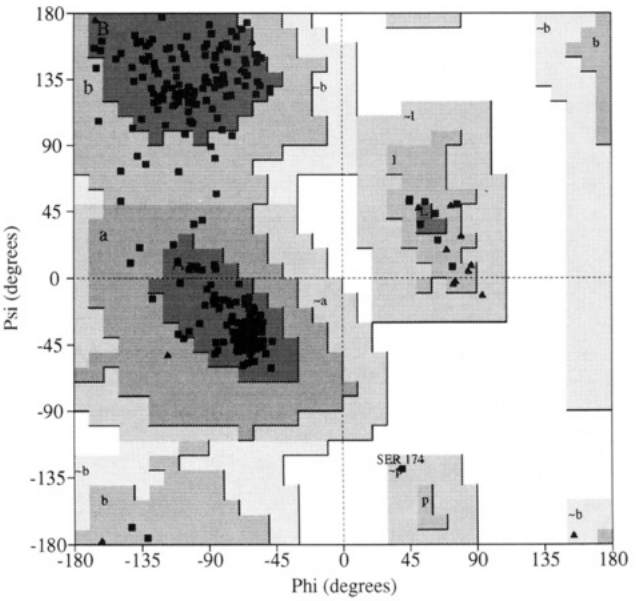

FIGURE 2: Ramachandran plot for the current model of Endo F_1. The figure is produced using the program PROCHECK (Morris et al., 1992). Ser174 is the only non-glycine residue in a generously allowed area.

by the α-helices. The helices run parallel with the β-strands and connect the carboxyl-terminal end of one β-strand to the amino-terminal end of the next one. The loops connecting the β-strands to the α-helices tend to be long and frequently contain additional structural elements, while the loops connecting the α-helices to the β-strand tend to be very short.

Figure 5 shows the α-carbon tracing of the Endo F_1 α/β-barrel with every 15th residue labeled. Consistent with the definitions used in the description of other α/β-barrels, the area of the carboxyl-terminal residues of the β-strands is referred to as the top of the barrel. The amino-terminal residues, 5−8, form an extended section along the bottom of the barrel rim. β-Strand 1 starts at residue 9. α-Helix 1 is short and is followed by a three-residue β-strand (residues 28−30) before reaching the bottom of the molecule. Relatively large excursions from the simple α/β-barrel are observed in the loops that connect β-strands 2, 3, and 4 to the following α-helices. β-Strand 2, residues 41−52, is almost twice as long as the others and folds back to form a small antiparallel β-hairpin with residues 58−61 on top of the molecule before turning into helix 2. This helix is kinked at residue Arg73. The β-hairpin of loop 2 is located above α-helix 3, which is long and almost perpendicular to the main axis of the β-barrel. The loops following β-strands 3 and 4 are long and have random coil geometry. The conformations of the next two β/loop/α units are unusual. The β-strands are short and the α-helices are essentially missing. The link between β-strands 5 and 6 has only a small loop at the top of the barrel. This loop consists primarily of a single-turn 3_{10} helix, residues 174−177, before turning into a section that runs to the bottom of the molecule. This section is in an extended conformation and replaces the α-helix. At the bottom of the molecule, there is a rather long loop that extends below helix 4 and contains a single-turn α-helix, residues 187−190, close to the amino-terminal end of β-strand 6. The loop following β-strand 6 makes a sharp turn into an extended section that runs from the top to the bottom of the molecule. Only two residues have α-helical conformation. These are located at the bottom of the barrel, adjacent to the amino terminus of β-strand 7. β/loop/α units 7 and 8 are similar to units 1 and 4, in that the β-strand is followed by a loop that forms the upper surface of the barrel before turning into the α-helices which run essentially parallel to the β-strands. α-Helix 8 ends at residue 274, leading into a type I reverse turn. Residues 275−289 run up from the bottom to the top of the barrel adjacent to helix 1. Most of this section is in an extended conformation. Residues 279−281 have a β-strand conformation and form a small two-stranded antiparallel β-sheet with the three-residue β-strand (30−28) that follows helix 1.

When viewed down the main axis of the β-barrel, the molecule is elliptical in shape with a short axis of about 32 Å, located near β-strand 2, and a long axis of 47 Å, where axis lengths correspond to the distances between α-carbons on the exterior surfaces of the molecule. The area of α-helices 2, 3, and 4 extends somewhat further out from the center of the β-barrel than the area of α-helices 1, 7, and 8. The central cavity in the β-barrel also is elliptical in shape. It has a short axis of about 11 Å and a long axis of about 16 Å. The long axis of the central cavity ellipsoid makes an angle of about 35° with the short axis of the molecule and is located near β-strand 3. The interior of the β-barrel consists of three layers of large hydrophobic residues, including several phenylalanine and methionine residues. These three layers are oriented perpendicular to the barrel axis in a typical arrangement for the packing of residues in the core of α/β-barrels (Lesk et al., 1989).

Active Site and Oligosaccharide Binding Cleft. The contrast between the short turns of the loops that follow β-strands 5 and 6 and the longer excursions of the other loops results in a depression in the rim of the barrel above these two strands. This depression is part of a shallow but well-defined cleft that runs from this area over the core of the β-barrel to make a turn to the left to exit between loop 1 and the β-hairpin in loop 2. Figure 6 shows a space-filling model of the top of the molecule, illustrating the cleft and the residues that form its surface. The cleft is about 40 Å long and is lined with a large number of aromatic and polar residues, most of which are contributed by the turns and excursions following β-strands 2, 3, and 4 on the one side and 7, 8, and 1 on the other side. Figure 5 shows this cleft with the residues that have side chains extending into it. The catalytic residues, Asp130 and Glu 132, are located in this cleft adjacent to the core of the β-barrel. Asp130 is buried deep in the core of the β-barrel but extends up so that its carboxylate group forms part of the base of the cleft. Glu132 is the last residue of β-strand 4, and its side chain extends into the cleft from the side. The side chains of the two residues are in hydrogen-bonding contact with each other and with several water molecules but not with any other residues. Figure 7 shows catalytic residues and the surrounding residues that form part of the cleft in the area of the β-barrel. Table 4 lists all residues that have side chains extending into the cleft. No other acidic residues are close enough to Asp130 or Glu132 to form a typical hydrolase active site. The only other acidic residues in the vicinity of the β-barrel are Glu17, Asp198, and Glu223. The shortest distances of their carboxylate oxygens to the carboxylate oxygens of Glu132 are 10.3, 7.0, and 10.4 Å, respectively. Aromatic residues, primarily tyrosines, are prevalent at the front end of the cleft, above β-strands 5 and 6 and around

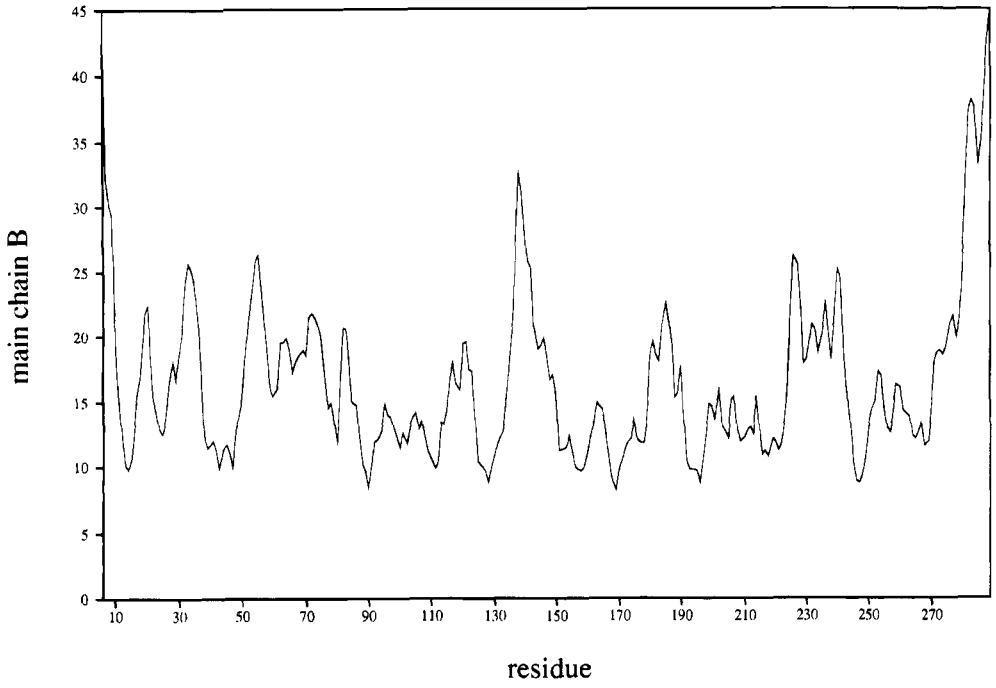

FIGURE 3: Main-chain thermal parameters (Å^2) for the current model of Endo F_1.

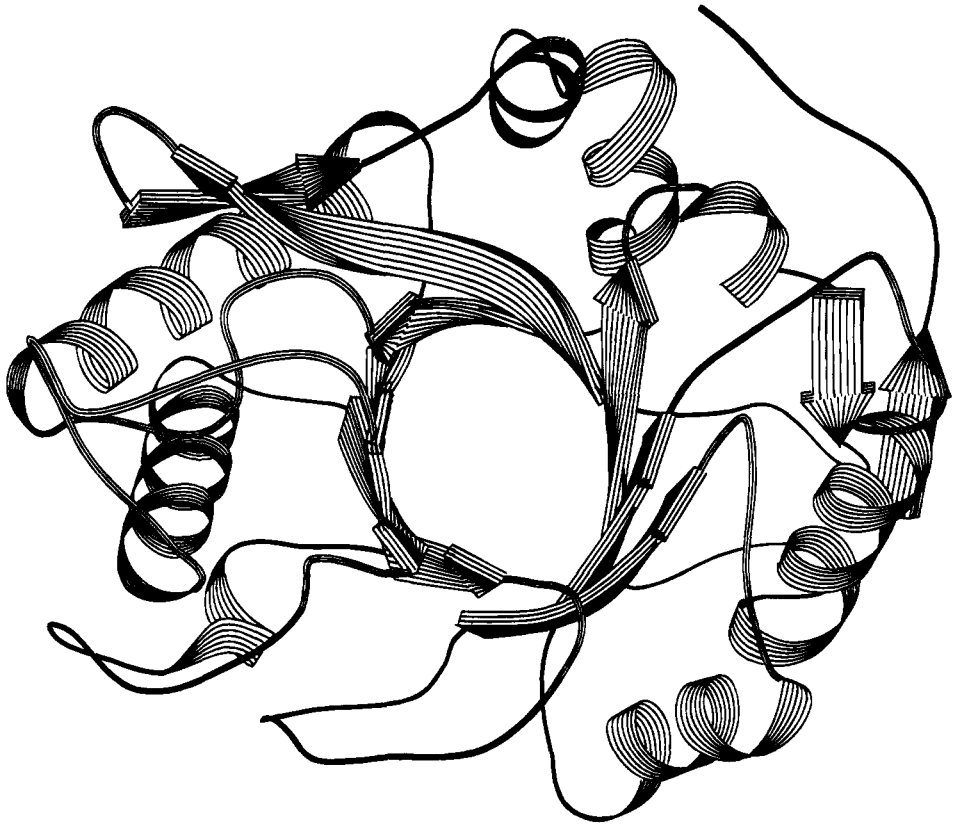

FIGURE 4: Idealized representation of the main-chain tracing of Endo F_1 produced with the program RIBBON (Priestle, 1988). The α-helices are represented as spiral ribbons and the β-strands as solid arrows. The most important features of the folding include the β-hairpin following β-strand 2 (top left) and the absence of α-helices in units 5 and 6 (bottom left).

the catalytic residues. Especially, tyrosine residues 133, 171, and 173 surround Asp130 and Glu132 but without making hydrogen-bonding contacts. The cleft area beyond the β-barrel contains many polar residues, primarily asparagines. Tyrosines and asparagines have been implicated in the binding and recognition of oligosaccharides in lectins, antibodies, and binding proteins. Therefore, this cleft is thought to be the oligosaccharide binding and recognition site. The positions of Asp130 and Glu132 above each other appear to allow approach of the substrate from the top and from the bottom of the molecule rather than from opposite sides of the cleft.

The cleft is occupied by several water molecules and at least one Zn^{2+} ion. The Zn^{2+} ion is coordinated by His95, by Glu272 of a symmetry-related molecule, and by at least one water molecule. No Zn^{2+} ions or other cations are

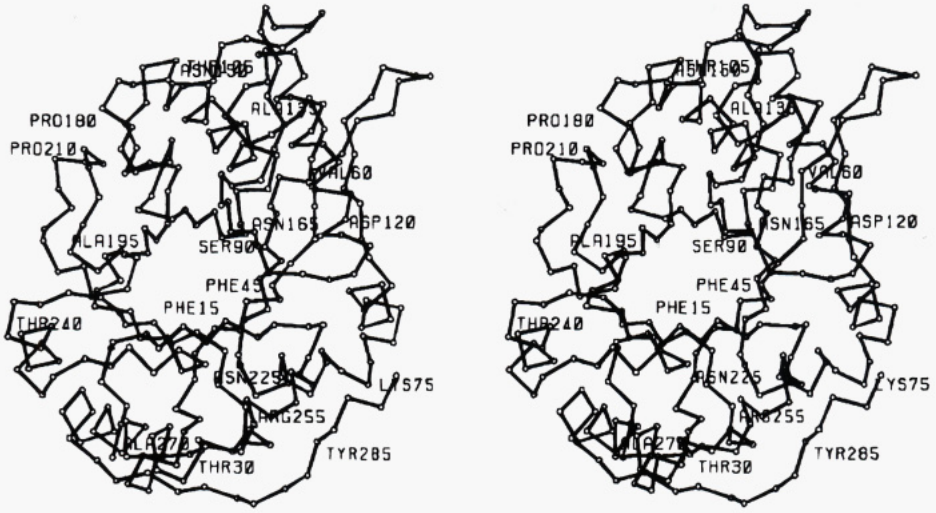

FIGURE 5: Stereo diagram showing the α-carbon tracing of Endo F_1 with every 15th residue labeled.

FIGURE 6: Stereo diagram showing the space-filling model of the structure of Endo F_1 viewed down the barrel axis. Residues are colored according to the following scheme: red, Asp, Glu; magenta, Tyr, Phe, Trp; dark blue, Arg, Lys; light blue, Asn, Gln; gray-blue, His; brown, Pro, Ala, Val, Leu, Ile; green, Ser, Thr; yellow, Cys, Met; gray, Gly. The molecule is shown in the same orientation as in Figure 4. The oligosaccharide binding cleft runs from the bottom of the figure (loops 5 and 6) to the top but makes a left turn past the central area of the barrel. The catalytic residues, Asp130 and Glu132, are the red residues in the left center of the figure and are flanked by the aromatic residues (magenta), Tyr171 and Tyr173 to the lower end of the figure and Tyr133 at the other side.

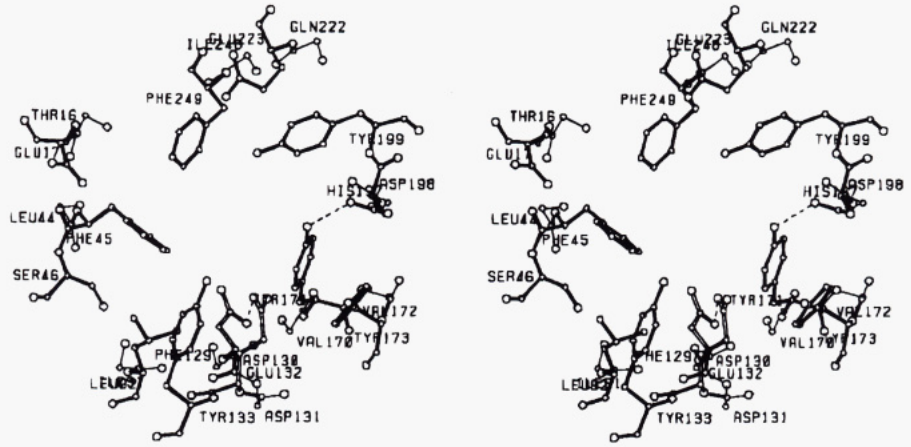

FIGURE 7: Stereo diagram showing the residues that form the active site and the upper area of the β-barrel. The catalytic residues, Asp130 and Glu 132, have been drawn with open bonds and can be seen in the middle at the bottom of the figure. All other residues that form part of the binding cleft are shown with heavy solid bonds. Main-chain atoms for the residues preceding and following the residue of interest are included in fine line to indicate the orientation of the barrel.

associated with the purified enzyme, nor are cations required for activity. Therefore, the Zn^{2+} ion is incorporated into the crystal lattice from the zinc acetate in the crystallization buffer, and its role is limited to that of an effector of intermolecular interaction in the crystal packing. Zn^{2+}–ligand distances are 2.32 Å for His95, 2.21 and 2.58 Å for the carboxylate oxygens of Glu272, and 2.84 Å for the closest water molecule. This water molecule also refines with a

Table 4: Residues That Line the Oligosaccharide Binding Cleft of Endo F_1[a]

loop no.	residue	loop no.	residue
2	Phe59 (Tyr)		
	Ser61 (His)		
	Asn63 (Asn)		
	Asn65 (Asn)		
2	Asn50 (Asn)	1	Asn19 (Asn)
	Asn48 (Asn)		Glu17 (Glu)
	Ser46 (Ala)		
	Phe45 (Phe)		
3	His95 (His)	8	Phe249 (Tyr)
	Asn94 (Asn)		
	Leu92 (Leu)		
4	Asp130 (Asp)	7	Asn225 (Gly)
	Glu132 (Glu)		Gln226 (Arg)
	Tyr133 (Tyr)		Glu223 (Glu)
5	Tyr173 (Tyr)	6	Asp198 (Pro)
	Tyr171 (Tyr)		Tyr199 (Tyr)

[a] The residues have been organized according to their respective positions on the basis of the orientation of Figure 6 so that the area of loops 5 and 6 is at the bottom of the table. The corresponding residues of Endo H have been included in parentheses.

very low thermal parameter, less than 2 Å^2, and is surrounded by several other peaks at distances less than 2.0 Å. It is possible that this density corresponds to that of a cacodylate anion. This possibility will be further investigated during refinement at higher resolution.

DISCUSSION

Comparison with Other α/β-Barrels. The TIM-type α/β-barrel has been observed in at least 25 different proteins that have little or no sequence homology or related functions. The TIM-barrel motif has been analyzed extensively because this represents an unusually large percentage of all known structures (Brändén, 1991; Farber, 1993). All but one of the other proteins that have the TIM-type α/β-barrel fold are enzymes. All enzymes have their active site at the top of the β-barrel, where residues of the loops that connect the β-strands with the α-helices contribute to their functionality. In all but one case, the α-subunit of tryptophan synthetase (Hyde et al., 1988), the catalytic residues are contributed by different β-strands. Many of the TIM-barrel proteins have deviations from the standard TIM-barrel fold. These can range from minor excursions to large external domains. All excursions or additional domains either precede the first β-strand, replace the last α-helix, or are inserted in the loops on top of the barrel. In contrast to the loops at the top of the molecule, the turns at the bottom tend to be short and often have one of two conserved conformations, one with one and one with three residues between the α-helix and the β-strand (Scheerlinck et al., 1992).

The Endo F_1 barrel follows the pattern observed for all others except for the following characteristics: α-helices 5 and 6 are missing, helix 2 runs perpendicular to the β-strands, and there are significant excursions at the bottom of the barrel in the loops that connect β-strands 5 to 6 and 6 to 7. The structure of Endo F_1 shows the first TIM barrel that lacks two α-helices in consecutive units. These helices are replaced by sections in extended conformation followed by excursions at the bottom of the molecule. However, these loops do not appear to be important for the activity or structural integrity of the molecule. It would therefore seem logical that it is the absence of the helices that must be an important feature of the structure.

The geometry of TIM barrels has been reviewed and discussed extensively in relation to convergent or divergent evolution (Farber, 1993). Farber and Petsko (1990) classified the barrels into four different structural classes on the basis of criteria such as the domain structure, the orientation of the principal axis of the β-barrel, and the requirement for cofactors. Endo F_1 does not fit either one of these classes, nor do many of the other new TIM-barrel structures discovered since the Farber and Petsko analysis. The location of the major axis of Endo F_1 places it in the same group as α-amylase (Matsuura et al., 1984; Buisson et al., 1987) and cyclodextrin glycosyltransferase (Klein & Schultz, 1991), which are enzymes that share some functional similarity with Endo F_1 in that they also hydrolyze polysaccharides. However, Endo F_1 lacks the additional domains and the requirements for cofactors that these two enzymes share. Although the active site of the amylases is located at the top of the β-barrel, the additional domain plays a major role in the functionality of the enzyme. The additional domain is the binding site for the essential Ca^{2+} ion, and the polysaccharide binding cleft is located between the two domains (Boel et al., 1990; Larsen et al., 1994; Qian et al., 1994).

Recently, the structures of two enzymes that are more closely related to Endo F_1 have been published (Varghese et al., 1994). These are the plant β-glycan endohydrolases, (1−3)-β-glucan glucanohydrolase and (1,3-1,4)-β-glucan glucanohydrolase. They hydrolyze internal (1,3)-β-glycosidic links in β-glucans. These enzymes fold as regular circular-shaped TIM barrels with a 40 Å long cleft across the top of the barrel. This cleft is similar in size and composition to the Endo F_1 cleft but is straight. It is lined primarily with tyrosine and asparagine residues. The active site residues are contributed by β-strand 7 and the loop between β-strand and α-helix 8. Although different loops are involved, the catalytic site resembles that of Endo F_1 because one of the catalytic residues, Glu231, is buried in the base of the cleft while the other, Glu288, comes in from the side. Both acidic residues are surrounded by several tyrosine residues, which are thought to interact with the faces of the glucosyl residues.

Interestingly, the plant glycan endohydrolases have a fold and a spatial arrangement of the active site residues that are closely related to those of Endo F_1, but the bacterial glycan endohydrolases, which have a similar sequence relationship to the catalytic residues of Endo F_1, are structurally unrelated (Keitel et al., 1993). This suggests that there are at least two different combinations of primary and tertiary structure that can lead to a functional glycohydrolase active site.

Substrate Binding. The unusual features of the Endo F_1 α/β-barrel are interpreted to be important for the binding and deglycosylation of glycoproteins, which are very complex and bulky substrates when compared with the substrates of other TIM barrels. The oligosaccharide component of the glycoprotein substrate binds in the cleft across the top of the barrel. The length of the cleft allows for binding and recognition of five to six oligosaccharide units, as is required in view of the substrate specificity. The bend in the cleft may be an adaptation to allow binding of a branched substrate because the cleft of the glycan endohydrolases, which process linear substrates, is straight. The putative catalytic residues are located above the β-barrel but toward the end of the cleft formed by the depression in the rim at loops 5 and 6. This

asymmetry of the cleft about the active site suggests that the substrate could bind with the reducing end of the oligosaccharide above loops 5 and 6 and the nonreducing end in the area of loops 1 and 2. This would place the first GlcNAc residues above the turns of loops 5 and 6 and the protein moiety of the substrate glycoprotein in front of the face where the α-helices are absent. This interpretation is supported by the observation that the absence of α-helices results in a narrower and conformationally more flexible barrel rim in this area, facilitating interaction with the protein moieties of a variety of the glycoprotein substrates.

Most of the residues that line the cleft are conserved between Endo H and Endo F_1 as is shown in Table 4. This includes almost all residues contributed by loops 1, 2, 3, 4, and 5 but not many of those of loops 6, 7, and 8. This is consistent with the substrate alignment described above because the area encompassing loops 7 and 8 is the contact region for the core α(1−6)-fucose residue, which is the only basis for differences in activity. In this alignment, the tyrosines surrounding the catalytic residues are responsible for the binding and positioning of the GlcNAc residues while the asparagines at the other end of the cleft are responsible for binding the Man residues and are important for determination of the substrate specificity. The analyses of the crystal structures of Endo H (Van Roey et al., 1994) and Endo F_3 are in progress and will allow evaluation of this proposed binding mechanism.

REFERENCES

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., & Waley, S. G. (1975) *Nature 255*, 609−614.

Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L., & Woldike, H. F.(1990) *Biochemistry 29*, 6244−6249.

Brändén, C.-I. (1991) *Curr. Opin. Struct. Biol. 1*, 978−983.

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

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

Buisson, G., Duée, E., Haser, R., & Payan, F. (1987) *EMBO J. 6*, 3909−3916.

Chen, L., Fincher, G. B., & Høj, P. B. (1993) *J. Biol. Chem. 268*, 13318−13326.

Evans, S. V. (1993) *J. Mol. Graphics 11*, 134−138.

Farber, G. K. (1993) *Curr. Opin. Struct. Biol. 3*, 409−412.

Farber, G. K., & Petsko, G. A. (1990) *Trends Biochem. Sci. 15*, 228−234.

Finzel, B. C. (1987) *J. Appl. Crystallogr. 20*, 53−55.

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

Hendrickson, W. A., & Konnert, J. H. (1980) in *Biomolecular structure, function, conformation and evolution* (Srinivasan, R., Ed.) pp 43−57, Pergamon Press, Oxford.

Hyde, C. G., Ahmed, S. A., Padlan, E. A., Miles, E. W., & Davies, D. R. (1988) *J. Biol. Chem. 263*, 17857−17871.

Juncosa, M., Pons, J., Dot, T., Querol, E., & Planas, A. (1994) *J. Biol. Chem. 269*, 14530−14535.

Keitel, Y., Simon, O., Borriss, R., & Heinemann, U. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5287−5291.

Klein, C., & Schulz, G. E. (1991) *J. Mol. Biol. 217*, 737−750.

Kuranda, M. J., & Robbins, P. W. (1991) *J. Biol. Chem. 266*, 19758−19767.

Larson, S. B., Greenwood, A., Cascio, D., Day, J., & McPherson, A. (1994) *J. Mol. Biol. 235*, 1560−1584.

Lesk, A. M., Brändén, C.-I., & Chothia, C. (1989) *Proteins 5*, 139−148.

Matsuura, Y., Kusunoki, M., Harada, W., & Kakudo, M. (1984) *J. Biochem. (Tokyo) 95*, 697−702.

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

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992) *Proteins 12*, 345−364.

Plummer, T. H., Jr., & Tarentino, A. L. (1991) *Glycobiology 1*, 257−263.

Priestle, J. P. (1988) *J. Appl. Cryst. 21*, 572−576.

Qian, M., Haser, R., Buisson, G., Duée, E., & Payan, F. (1994) *Biochemistry 33*, 6284−6294.

Sack, J. S. (1988) *J. Mol. Graphics 6*, 244−245.

Scheerlinck, J.-P. Y., Lasters, I., Claessens, M., De Mayer, M., Pio, F., Delhaise, P., & Wodak, S. J. (1992) *Proteins 12*, 299−313.

Schmidt, B. F., Ashizawa, E., Jarnagin, A. S., Lynn, S., Noto, G., Woodhouse, L., Estell, D. A., & Lad, P. (1994) *Arch. Biochem. Biophys. 311*, 350−353.

Tarentino, A. L., & Plummer, T. H., Jr. (1994) *Methods Enzymol. 230*, 44−57.

Tarentino, A. L., Plummer, T. H., Jr., & Maley, F. (1974) *J. Biol. Chem. 249*, 818−824.

Tarentino, A. L., Quinones, G., Schrader, W. P., Changchien, L.-M., & Plummer, T. H., Jr. (1992) *J. Biol. Chem. 267*, 3868−3872.

Tarentino, A. L., Quinones, G., Changchien, L.-M., & Plummer, T. H., Jr. (1993) *J. Biol. Chem. 268*, 9702−9708.

Trimble, R. B., & Maley, F. (1984) *Anal. Biochem. 141*, 515−522.

Trimble, R. B., & Tarentino, A. L. (1991) *J. Biol. Chem. 266*, 1646−1651.

Van Roey, P., Silva, G. H., Rao, V., Plummer, T. H. Jr., Tarentino, A. L., & Guan, C. (1994) *J. Mol. Biol. 237*, 157−159.

Varghese, J. N., Garrett, T. P. J., Colman, P. M., Chen, L., Høu, P. B., & Fincher, G. B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2785−2789.

Wang, B.-C. (1985) *Methods Enzymol. 115*, 90−112.

Watanabe, T., Kobori, K., Miyashita, K., Fujii, T., Sakai, H., Uchida, M., & Tanaka, H. (1993) *J. Biol. Chem. 268*, 18567−18572.