THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 273, No. 32, Issue of August 7, pp. 20205–20212, 1998

Printed in U.S.A.

Crystal Structures of *Flavobacterium* Glycosylasparaginase

AN N-TERMINAL NUCLEOPHILE HYDROLASE ACTIVATED BY INTRAMOLECULAR PROTEOLYSIS*

(Received for publication, April 23, 1998, and in revised form, May 15, 1998)

Hwai-Chen Guo‡§, Qian Xu‡, Deirdre Buckley‡¶, and Chudi Guan‖

From the ‡Department of Biophysics, Boston University School of Medicine, Boston, Massachusetts 02118-2526 and the ‖New England Biolabs, Beverly, Massachusetts 01915-5599

Glycosylasparaginase (GA) is a member of a novel family of N-terminal nucleophile hydrolases that catalytically use an N-terminal residue as both a polarizing base and a nucleophile. These enzymes are activated from a single chain precursor by intramolecular autoproteolysis to yield the N-terminal nucleophile. A deficiency of GA results in the human genetic disorder known as aspartylglycosaminuria. In this study, we report the crystal structure of recombinant GA from *Flavobacterium meningosepticum*. Similar to the human structure, the bacterial GA forms an $\alpha\beta\beta\alpha$ sandwich. However, some significant differences are observed between the *Flavobacterium* and human structures. The active site of *Flavobacterium* glycosylasparaginase is in an open conformation when compared with the human structure. We also describe the structure of a mutant wherein the N-terminal nucleophile Thr^{152} is substituted by a cysteine. In the bacterial GA crystals, we observe a heterotetrameric structure similar to that found in the human structure, as well as that observed in solution for eukaryotic glycosylasparaginases. The results confirm the suitability of the bacterial enzyme as a model to study the consequences of mutations in aspartylglycosaminuria patients. They also suggest that further studies are necessary to understand the detail mechanism of this enzyme. The presence of the heterotetrameric structure in the crystals is significant because dimerization of precursors has been suggested in the human enzyme to be a prerequisite to trigger autoproteolysis.

Eukaryotic glycosylasparaginase (glycoasparaginase, N^4-(β-*N*-acetyl-D-glucosaminyl)-L-asparaginase, 1-aspartamido-β-*N*-acetylglucosamine amidohydrolase, aspartylglycosylaminase, aspartylglucosaminidase, EC 3.5.1.26) is a well known lysosomal enzyme that cleaves the amide bond of asparagine-linked glycoproteins (1). It is widely distributed in vertebrate tissues (2) and insect cells (3) and is also found in bacteria (4). Substrate preferences for this enzyme include free α-amino and α-carboxyl groups on the asparagine, and that position 6 of N-acetylglucosamine does not contain fucose. However, a recent study suggests that the α-amino and α-carboxyl groups on the asparagine part of the substrate may not strictly be required for hydrolysis (5).

A deficiency of glycosylasparaginase (GA)[1] results in accumulation of glycoasparagines in tissue lysosomes and leads to severe clinical symptoms, known as aspartylglycosaminuria (AGU). AGU is the most common disorder of glycoprotein degradation. It severely involves the central nervous system and causes skeletal abnormalities and connective tissue lesions. Among children in eastern Finland, AGU was found to be the leading genetic cause for mental retardation after trisomy 21 and fragile X syndrome (1).

Glycosylasparaginase has been biochemically characterized from different species and is composed of two nonidentical subunits of approximately 24 and 20 kDa, associated by noncovalent forces. These respective subunits are referred to as the a- and b-subunits (or heavy and light subunits). The enzyme is encoded by a single gene, and post-translational cleavage of the nascent polypeptide into a mature a/b heterodimer is required for activation. Neither the single chain precursor (6, 7) nor the isolated subunits (8) are enzymatically active by themselves. Expression of the a- and b-subunits of GA on separate DNA constructs showed that independently folded subunits lack enzyme activity, and even when co-expressed *in vitro* they fail to produce an active heterodimer (9). A common feature of GA from different species is a new N-terminal threonine of the C-terminal product (the b-subunit) resulting from the autoproteolytic activation (10). A study demonstrated that an irreversible inhibitor specifically reacts with the N-terminal threonine on the b-subunit of the human leukocyte enzyme via an α-ketone ether linkage with the hydroxyl side chain (8), indicating that this N-terminal threonine acts as a nucleophile during substrate hydrolysis. The crystal structure of human GA shows a topology similar to other N-terminal nucleophile hydrolases (11, 12) and reveals interactions between the N-terminal threonine and aspartate, one of the reaction products (13).

GA from *Flavobacterium meningosepticum* is the only prokaryotic homolog characterized so far. It differs from the human counterpart in several aspects: (i) sequence alignment of these two enzymes reveals only about 30% sequence identity and shows a difference in one gap/insertion of 31 residues (3, 4); (ii) part of the 31-residue insertion in the human enzyme is removed from the new C terminus of the a-subunit in the lysosome (6); no trimming occurs in the bacterial enzyme; (iii) the human enzyme contains *N*-linked glycans on both the a- and b-subunits (Asn^{15} and Asn^{285}) (14), whereas the bacterial enzyme is nonglycosylated (4); (iv) according to previous sequence alignments (3, 4), neither the position nor the pattern of disulfide bridges is conserved between these two enzymes. The

* This work was supported in part by Grant IRG-97 T from the American Cancer Society (to H.-C. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (codes 2GAW and 2GAC) have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY.

§ To whom correspondence should be addressed: Dept. of Biophysics, Boston University School of Medicine, 715 Albany St., Boston, MA 02118-2526. Tel.: 617-638-4023; Fax: 617-638-4041; E-mail: hguo@med-biophd.bu.edu.

¶ Present address: Dept. of Biochemistry, University College Cork, Lee Maltings, Cork, Ireland.

[1] The abbreviations used are: GA, glycosylasparaginase; AGU, aspartylglycosaminuria; MIR, multiple isomorphous replacement; r.m.s., root mean square.

TABLE I
Data collection and phasing statistics

	Wild type	T152C	Derivatives			
			$Hg(OAc)_2$	CH_3HgCl	K_2PtCl_4	$KAuCN_2$
Resolution (Å)	2.2	2.1	2.5	2.5	4.5	2.5
Reflections (total/unique)	42,281/23,999	59,059/30,530	50,741/18,514	40,157/18,532	5,298/3,078	55,117/18,816
I/σ	6.9	8.3	9.5	6.2	7.0	8.9
Completeness (%)	86.7	96.2	97.7	96.0	91.1	98.7
R_{symm}[a]	0.11	0.09	0.09	0.11	0.09	0.08
R_{iso}[b]			0.24	0.34	0.26	0.14
Number of sites			4	6	6	2
MIR phases (12–2.5 Å) (MLPHARE)						
Phasing power[c]			1.43	1.56	1.49	1.03
Cullis R-factor[d]			0.78	0.78	0.75	0.86
Overall figure of merit				0.550		
Density modification (DM)						
Overall figure of merit				0.785		

[a] $R_{symm} = \Sigma|I_{obs} - I_{avg}|/\Sigma I_{obs}$.
[b] $R_{iso} = \Sigma|I_{der} - I_{nat}|/\Sigma I_{nat}$, where I_{nat} is the observed native (T152C) intensity.
[c] Phasing power $= \Sigma|F_H|/\Sigma\|F_{PH,obs}| - |F_{PH,calc}\|$.
[d] Cullis R-factor $= \Sigma\|F_{PH,obs}| - |F_{PH,calc}\|/\Sigma|F_{PH,obs}| - |F_{P,obs}\|$.

disulfide bonds have been shown to be essential for initial protein folding and activation of human GA (15). Together, these differences suggest the possibility that the structure of *Flavobacterium* GA may be significantly different from the human enzyme. Here we report the crystal structures of recombinant GA from *F. meningosepticum*. Structures of the wild type GA and a single amino acid mutant in their mature forms have been determined at 2.2 and 2.1 Å, respectively.

MATERIALS AND METHODS

Crystallization and Crystal Preparation—Protein expression and purification will be described elsewhere.[2] Crystals were grown in hanging drops equilibrated by vapor diffusion against well solutions of 15% PEG 3300, 100 mM HEPES, pH 7.5, 0.1% sodium azide. Microseeding was routinely used to improve crystal quality. Crystals were harvested into a modified well solution containing 20% PEG 3300. Shortly before data collection, crystals were placed in dialysis buttons and dialyzed stepwise against harvest buffer supplemented with a progressively higher concentrations of glycerol. The final glycerol concentration was 20%, and the transfer steps were between 5 and 10% glycerol and varied from 3 h to a couple of days for each step. No correlation was observed between transfer procedure and the quality of the diffraction data. Heavy atom derivatives were obtained by soaking the native crystals in the harvest buffer supplemented with 10% glycerol and 0.1–10 mM of heavy atom compound, from 15 h to a few days, before step up to the final glycerol concentration.

Data Collection and Processing—Oscillation data were collected from crystals frozen at 100 K, mounted in a thin film of harvest buffer plus 20% glycerol, and supported by a loop made of dental floss. Diffraction data were collected using an R-AXIS IIC image plate detector mounted on a Rigaku RU300 rotating anode generator. All intensity data were processed and scaled using the programs DENZO and SCALEPACK (16) and converted to structure factors using TRUNCATE from the CCP4 software package (17). The space group was determined by examination of the differences in intensities of potential pairs of reflections across putative mirror planes within the data and confirmed by examining the electron density maps.

Experimental Phases—Heavy atom positions were initially obtained from isomorphous Patterson maps calculated in XTALVIEW (18). The heavy atom parameters were then refined by MLPHARE from the CCP4 package (17) and were confirmed by the cross Fourier. At this stage, the figure of merit was 0.55 (0.71 calculated by XTALVIEW). The MIR phases were then extended to the resolution range 12–2.5 Å and were improved by noncrystallographic symmetry averaging, solvent flatterning, and histogram matching using the program DM (17) with a figure of merit of 0.785.

Model Building and Refinement—The first map was calculated at 2.5 Å resolution (see Fig. 3) and skeletonized to build a Cα trace in the program O (19). The first 244 of 275 residues were built based on the DM-modified MIR map. Automated refinement included rigid body, overall temperature factor, positional, and restrained atomic temperature factor refinement, as well as simulated annealing using a slow cooling protocol in X-PLOR (20). After the first round of manual rebuilding, without the N-terminal nucleophile (amino acid 152), the structure, after rigid body fit by AMORE (17), was used for refinement of both the wild type and the T152C mutant. SIGMAA (17) was used in the early cycles of refinement and manual rebuilding to combine model phases with experimental phases. Initially, strict noncrystallographic symmetry constraints were applied, and in later stages of refinement, tight noncrystallographic symmetry restraints were applied, exclusive of residues that were involved in crystal contacts. After a few rounds of model rebuilding, stepwise resolution extension, and automated refinement, clear electron density could be seen for all residues in the final model. Refinement protocols were aimed at decreasing the R_{free} (21) rather than the conventional R_{cryst} to avoid errors introduced by overfitting of the data. When the R_{free} appeared to have reached a minimum at the final resolution, water molecules were added and subjected to another round of automated refinement and manual rebuilding. The statistics of the final structures are shown in Table II, with a root mean square (r.m.s.) deviation of 0.20–0.23 Å for main chain atoms between crystallographically independent molecules.

Structural Comparisons—All superimpositions of different structures were performed using LSQKAB (17). For Fig. 4, all atoms of residues contacting the reaction product, aspartate, in the human structure are superimposed (atoms equivalent to those in bacterial Thr^{152}, Thr^{170}, Arg^{180}, Asp^{183}, Thr^{203}, and Gly^{204}).

RESULTS

Description of the Structure—The enzymes crystallized in space group $P2_1$ with unit cell constants $a = 46.2$Å, $b = 97.3$Å, $c = 61.8$Å, and $\beta = 90.3°$. The initial phases were obtained by MIR method with four heavy atom derivatives (Table I). The wild type structure has been determined at 2.2 Å and refined to an R_{free} (21) of 29.70% and an R_{cryst} of 24.65% with all reflections (Table II). There are two a/b heterodimers per asymmetric unit. In the final model, each heterodimer comprises 136 residues (3–138) of the a-subunit and 139 residues (152–290) of the b-subunit. No electron density is observed for the 13 residues spanning the segment (139–151) that connects the a- and b-subunits in the precursor protein. In the crystal, this linker segment appears to face into the solvent channels. 93% of the nonglycine residues fall in the most favored regions of Ramachandran plot, as defined in PROCHECK (22), and no residues are in the disallowed regions.

Overall, the topology of *Flavobacterium* GA is very similar to its human counterpart (13). Both the a- and b-subunits together form a four-layer α-β-β-α structure (Fig. 1), with two β-sheets packed against each other to form a core that is "sandwiched" by two layers of α-helices. Eight β-strands from both the a- and b-subunits form the first β-sheet, with topology aS4, aS3, aS2, bS2, bS1, aS1, bS7, and bS8. All these β-strands

[2] T. Cui, T. J. O'Loughlin, C. Guan, and H.-C. Guo, manuscript in preparation.

TABLE II
Refinement statistics

	Wild type	T152C
Resolution range (Å)	6–2.2	6–2.1
Unique reflections		
All F	22,932	29,161
$F > 3\sigma$	18,089	25,661
Non-hydrogen atoms		
Protein[a]	4,188	4,186
Water	134	246
R_{cryst} (%)[b]		
All F	24.65	23.32
$F > 3\sigma$	22.43	21.93
R_{free} (%)[c]		
All F	29.70	28.06
$F > 3\sigma$	26.95	26.69
r.m.s. deviation		
Bond length (Å)	0.011	0.011
Bond angle (°)	2.7	2.8
B-value (Å^2)[d]	4.3	5.3
Two a/b heterodimers (Å)[e]	0.20	0.23
Average B-value (Å^2)		
Main chain	9.3	14.3
Side chain	10.3	15.6
Water	23.1	23.8
Residues in Ramachandran plot (%)		
Most favored	93.3	92.9
Additional	6.2	6.7
Disallowed	0.0	0.0

[a] Model contains residues 3–138 and 152–290 of both a/b heterodimers 1 and 2 in the asymmetric unit.
[b] $R_{cryst} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$.
[c] R_{free} was calculated with 5 and 10% of data for wild type and T152C structures, respectively.
[d] Atomic B-values are for all protein atoms.
[e] Value is for all main chain atoms between a/b heterodimers 1 and 2 in the asymmetric unit.

are antiparallel except aS4, which is parallel to aS3. The other β-sheet is comprised of four antiparallel β-strands from the b-subunit, with the topology bS3, bS4, bS5, and bS6. The α-helix layer packed against the outside of the eight-strand sheet is formed by five α-helices from the a-subunit: aH1, aH2, aH3, aH4, and aH5. The other layer of α-helix packed outside of the four-strand sheet is formed by three α-helices (bH1, bH2, and bH3) plus a 3_{10} helix (bH4) from the b-subunit. All inter-layer loops cluster at one side of the structure to provide functional groups for the active site (top center of the structure in Fig. 1, toward the viewer), whereas the intra-layer loops are located on the other side. The structure-based alignment of amino acid sequences between *Flavobacterium* and human enzymes is shown in Fig. 2. They have similar secondary structural elements, with the exception that the 3_{10} helix (bH4) is unique to the bacterial structure. The bacterial GA differs from the human form as follows. There is a 1-residue insertion at the N-terminal end of the aH1 helix, a 1-residue deletion between strands aS2 and aS3, a 31-residue deletion at the C-terminal end of the a-subunit, a 7-residue insertion between helices bH3 and bH4, a 4-residue deletion between strands bS6 and bS7, a 2-residue deletion between strands bS7 and bS8, and 2 extra residues at the C-terminal end of the b-subunit.

A deficiency of GA results in the human genetic disorder known as AGU (1). In Fig. 2, we have mapped five known AGU mutations onto the structure-based sequence alignment. Only the double mutant (human Arg^{138} to Gln and Cys^{140} to Ser) maps outside of the shared secondary structural elements. This double mutant results in the loss of a disulfide bond (Cys^{140}-Cys^{156}) that stabilizes the loop formed by a portion of the 31-residue insertion at the C-terminal end of the human a-subunit (Fig. 1*c*). The remaining four mutations may disturb secondary structural elements, such as aH1 (with a 2-amino acid insertion), aH2, aS3, bS5, and thus may disturb the correct folding of the enzyme.

Structural Differences between Flavobacterium and Human Structures—An r.m.s. deviation of 1.4 Å is obtained by superimposing the common 1,068 main chain atoms (excluding insertions/deletions) of the *Flavobacterium* and human GA structures. This is significantly larger than the r.m.s. deviation of 0.26–0.45 Å found between the two human structures (13). Moreover, the r.m.s. deviation between our two bacterial structures is 0.22 Å (see below), similar to that observed between two heterodimers in the asymmetric unit (Table II). A number of peptide fragments within the structure deviate by more than 2 Å (Fig. 1*c*); most of them are in loops connecting elements of secondary structure. The largest difference of 8.5 Å is near the 7-residue insertion in the bacterial structure. Deviations greater than 2 Å are also observed in the common secondary structural elements (see below). These data are consistent with the observation that molecular replacement using the human structure proved difficult with the data of *Flavobacterium* GA.[3]

The human enzyme contains four disulfide bonds (Fig. 2) that are important for protein folding, autoproteolysis, and enzyme activity (15). These four disulfide bonds are conserved among mammalian enzymes. The insect enzyme retains all but one (Cys^{263}-Cys^{283}) of the disulfide bonds (3). However, no conserved disulfide bond is found between the *Flavobacterium* and eukaryotic GA. Indeed, there are no disulfide bonds among the five cysteines in the bacterial a/b heterodimer. One cysteine pair in the bacterial structure (Cys^{68}-Cys^{168}) has side chains in close proximity that may potentially form a disulfide bond, but this was ruled out based on several observations: (i) Cys^{68} and Cys^{168} bind to heavy atoms $Hg(OAc)_2$ and CH_3HgCl, respectively; (ii) the initial MIR map indicates that the side chain of Cys^{168} point away from Cys^{68}; (iii) the simulated annealed omit maps also show these two side chains to be in nonbridged conformations; (iv) a Cys to Ser mutation at either of these two cysteines does not significantly affect either protein stability or enzymatic activity[4]; and (v) the a- and b-subunits can be separated on a nonreducing SDS protein gel (data not shown).

Although the overall protein folds are similar in the bacterial and human structures, the location or length of some secondary structural elements differ. For example, the bacterial enzyme has a unique 3_{10} helix (bH4), whereas the human enzyme carries a C-terminal additional loop on its a-subunit (Fig. 1*c*). Furthermore, in the bacterial structure, the insertion of Gly^{14} extends helix aH1 at its N-terminal end by two residues. At the C-terminal end of helix aH1, Ser^{26} is designated as part of the helix in the human structure, but the equivalent Lys^{27} in the bacterial structure is not assigned as part of the helix by PROCHECK (22). This is apparently because of a significant deviation of the main chain traces between these two structures (Fig. 3). When these two structures are superimposed by their common secondary structural elements, Cα of Lys^{27} deviates by 3.5 Å from its equivalent atom in the human structure. The C-terminal end of aH2 helix also deviates by more than 2 Å. No crystal contact either in the human or bacterial structure can account for these deviations. In the bacterial structure, the 7-residue insertion in the b-subunit also extends the bH3 helix by 4 residues at its C-terminal end.

Active Site and Mechanism—The loops connecting different layers of α-helices and β-sheets form a deep funnel-shaped active site centered at the N-terminal Thr^{152} of the b-subunit (Fig. 1). The funnel in the bacterial enzyme is wider than that of the human enzyme, mainly because of deviation of the loop between helix aH2 and strand aS2 as well as lack of the

[3] H.-C. Guo and Q. Xu, unpublished observation.
[4] T. Cui, unpublished results.

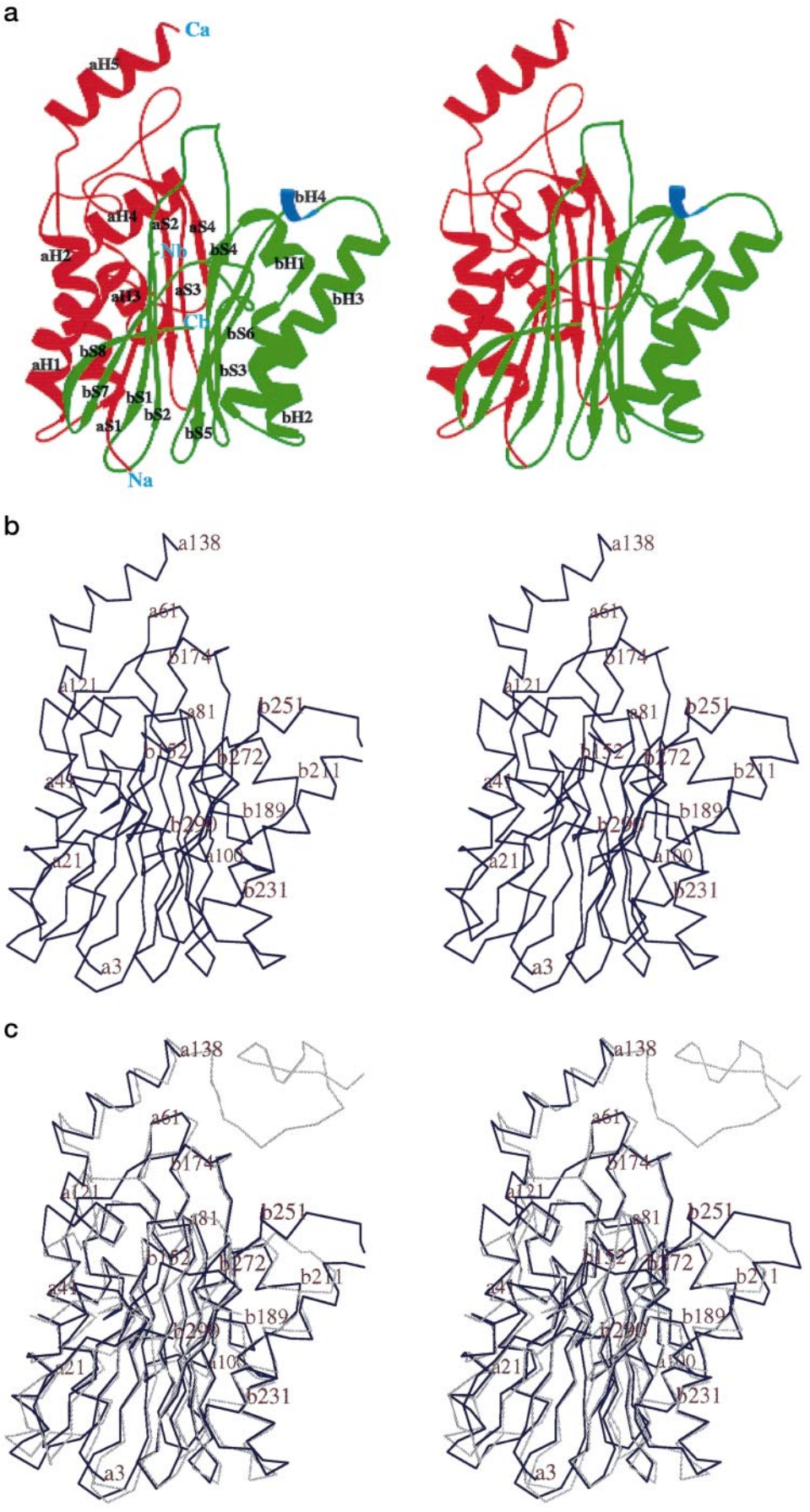

FIG. 1. **The structure of glycosylasparaginase from *F. meningosepticum.*** *a*, stereo ribbon representation of the *Flavobacterium* GA structure. One heterodimer is shown with a- (*red*) and b-subunits (*green*). The active site is at *top center* of the structure toward the viewer and around the N-terminal end of the b-subunit (*green*) (labeled *Nb* in *light blue*). *b*, stereo diagram of Cα traces of *Flavobacterium* GA. *c*, stereo diagram

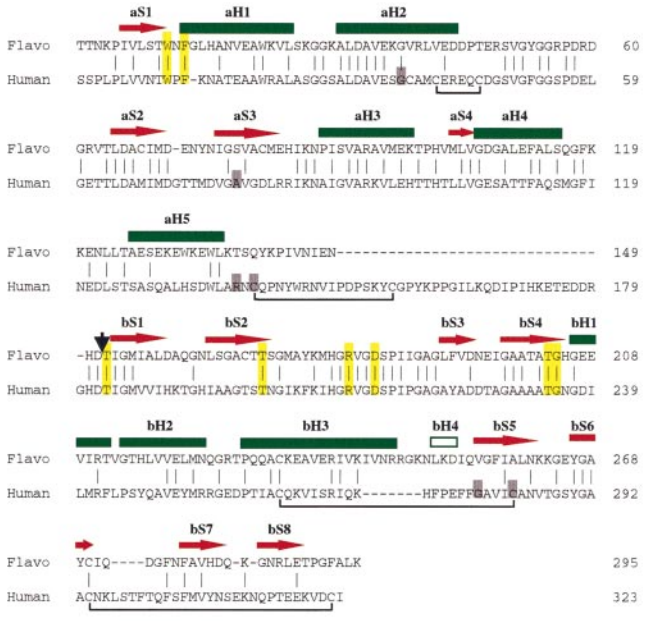

FIG. 2. **Structure-based sequence alignment between the *Flavobacterium* (*Flavo*) and human glycosylasparaginase (*Human*).** The *dashed lines* represent incorporated gaps that bring the sequences into alignment. The *vertical lines* represent identical matches. The *bold black arrow* represents the autoproteolytic site. *Yellow shading* shows conserved residues in the active site (see Fig. 4 and text). *Purple shades* highlight the mutations identified in human AGU disease (1); not shown is an Asp-Ala insertion between Ala^{19} and Ala^{20} in the human sequence. The secondary structural elements are based on this study: *red arrows* for β-strands, *green rectangles* for α-helices, and *open rectangle* for 3_{10} helix. Cysteine residues forming disulfide bonds in the human structure are linked by *lines* below the human sequences: Cys^{41}-Cys^{46}, Cys^{140}-Cys^{156}, Cys^{263}-Cys^{283}, and Cys^{294}-Cys^{322}.

C-terminal loop in the a-subunit (Fig. 1*c*). Several conserved residues surround the nucleophilic center Thr^{152} of the bacterial active site, including Thr^{170}, Arg^{180}, Asp^{183}, Thr^{203}, and Gly^{204}, which are highlighted in *yellow* in Fig. 2. These residues had been described to interact with aspartate, one of the two reaction products (13). As depicted in Fig. 4*a*, the human equivalent to Arg^{180} forms hydrogen bonds with the α-carboxyl group of aspartate. Both human equivalent residues of Asp^{183} and Gly^{204} make hydrogen bonds with the α-amino group of aspartate. Human residues equivalent to Thr^{152}, Thr^{203}, and Gly^{204} also form hydrogen bonds with the Oδ1 of aspartate. In addition, human residue equivalent to Thr^{170} makes a hydrogen bond with the Oγ of Thr^{152}.

Additional conserved residues not described previously also might participate in either ligand binding or catalysis. Residue Trp^{11} has a putative role in carbohydrate binding (13). It is also near the N-terminal nucleophile and may participate in catalysis, possibly through a bridging water molecule. In line with this suggestion, mutation of Trp^{11} to Ser (W11S) affects enzyme specificity for substrates with or without carbohydrate moiety (23). Furthermore, the k_{cat} of the W11S mutant was reduced by more than 400-fold, suggesting an additional role of Trp^{11} in regulating enzyme catalysis. Contrary to a previous suggestion (13), human Phe^{278} might not contribute to carbohydrate binding, because sequence alignment shows that this residue is not conserved either in the bacterial (Gln^{254} in Fig. 2) or insect (Met^{278} in Ref. 3) enzymes. Nonetheless, aromatic side chains have been suggested to be involved in protein-carbohydrate interactions (24). Here we propose that two conserved aromatic residues, Phe^{13} and Trp^{11} (Fig. 4), form part of the carbohydrate binding site.

Based on structural and biochemical studies, the reaction mechanism of GA is similar to serine proteases and hence utilizes a cycle of enzymatic acylation and deacylation. However, the free α-amino group on the N-terminal threonine acts as the base, probably through a bridging water molecule, to enhance the nucleophilicity of its own side chain hydroxyl group. This intra-residue base on the threonine replaces the well characterized histidine base in the hydrogen-bonded triad that is present in the active site of many serine proteases (25). The activated Oγ of Thr^{152} attacks the amide carbon of a substrate to form a tetrahedral transition state structure that is stabilized by an oxyanion hole. The structure then collapses to form a covalent enzyme-acyl (β-aspartyl) intermediate with release of the carbohydrate product. Deacylation is accomplished by a nucleophilic attack by an entering water molecule on the same carbon to release aspartate, the second product.

The identity of the oxyanion hole that stabilizes the negatively charged carbonyl oxygen on the tetrahedral transition state is still unclear in the current GA structures. Oinonen *et al.* (13) proposed that the side chain of human residue equivalent to bacterial Thr^{203} and the main chain equivalent to Gly^{204} act as the oxyanion hole, based on the structure of human enzyme-product complex. However, when the active sites of the bacterial and human structures are superimposed (Fig. 4*a*), there are conformational differences with respect to the nucleophilic Oγ of the N-terminal Thr^{152} (human Thr^{183}) and the proposed oxyanion hole Oγ of Thr^{203} (human Thr^{234}). We suggest that the current structure of the bacterial enzyme appears to be in an open conformation, whereas the human enzyme adopts a closed conformation that grasps the reaction product, aspartate (Fig. 4*a*). In the bacterial structure, the Oγ of Thr^{203} is displaced by 1.9 Å and the Oγ of Thr^{152} is shifted in the opposite direction by 0.7 Å (the r.m.s. deviation of all other residues hydrogen-bonded to Asp is 0.64 Å between the bacterial and human enzymes, and 0.25 Å between bacterial wild type and the T152C mutant). As a result, the relative distance between these two atoms has changed by 2.3 Å. In the case of isocitrate dehydrogenase (26), small changes in distance (<1.55 Å) and orientation of reacting groups results in a large reduction (10^{-3} to 10^{-5}) in the reaction rate. The differences in the GA case could result from the binding of ligand (aspartate) in the human complex. However, the structure of the unliganded human enzyme (13) also has a similar closed conformation. This raises the possibility that the differences observed in the position of Thr^{203} in the bacterial structure may represent differences in mechanism relative to the human enzyme. Mutagenesis studies also indicate that the side chain of bacterial Thr^{203} may not be as important in stabilizing the negative oxyanion intermediate as previously suggested for the human enzyme (13), because replacement by Ala (T203A mutant) in the bacterial enzyme decreases k_{cat} only about 10-fold (23). Further studies are necessary to determine whether Gly^{204} together with a main chain component of Thr^{203} (or other residues) actually form the oxyanion hole.

Structure of the T152C Mutant—Thr^{152} plays a key role in catalysis (4, 7, 8). Substitution of the N-terminal nucleophile Thr^{152} by a thiol group (T152C mutant) reduces k_{cat} by 5 orders of magnitude (23). Autoproteolysis in this mutant is also very slow but can be accelerated by hydroxylamine (10). In this study, we have also determined the three-dimensional struc-

of Cα traces of *Flavobacterium* (*dark blue*) and human (*gray*) GA. Superimposition is based on all common main chain atoms (excluding insertions/deletions). The residues labeled are in *Flavobacterium* sequence number.

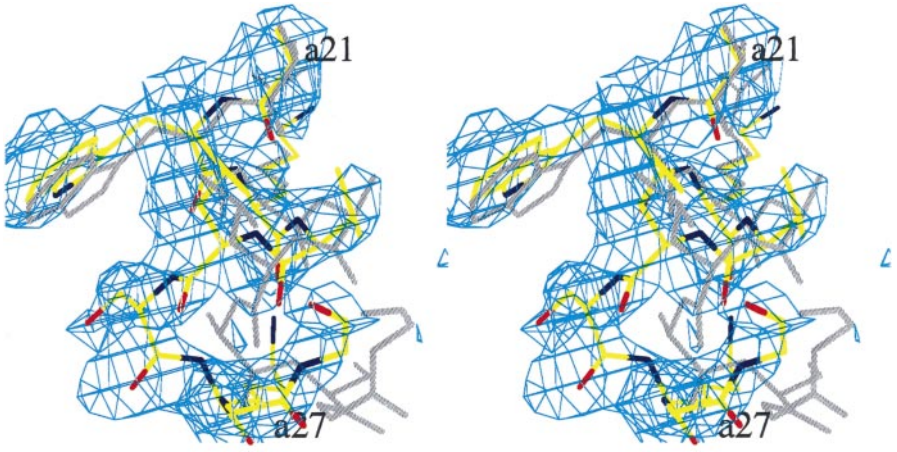

FIG. 3. **Quality of experimental map at 2.5 Å obtained with MIR phases modified using the program DM (17).** The map is contoured at 1.1 σ. The final model of the C-terminal end of α-helix aH1 is shown by atom type: *yellow* for carbons, *blue* for nitrogens, and *red* for oxygens. The best fit structure of human GA based on the common secondary structural elements is also superimposed (shown in *gray*) to visualize the differences.

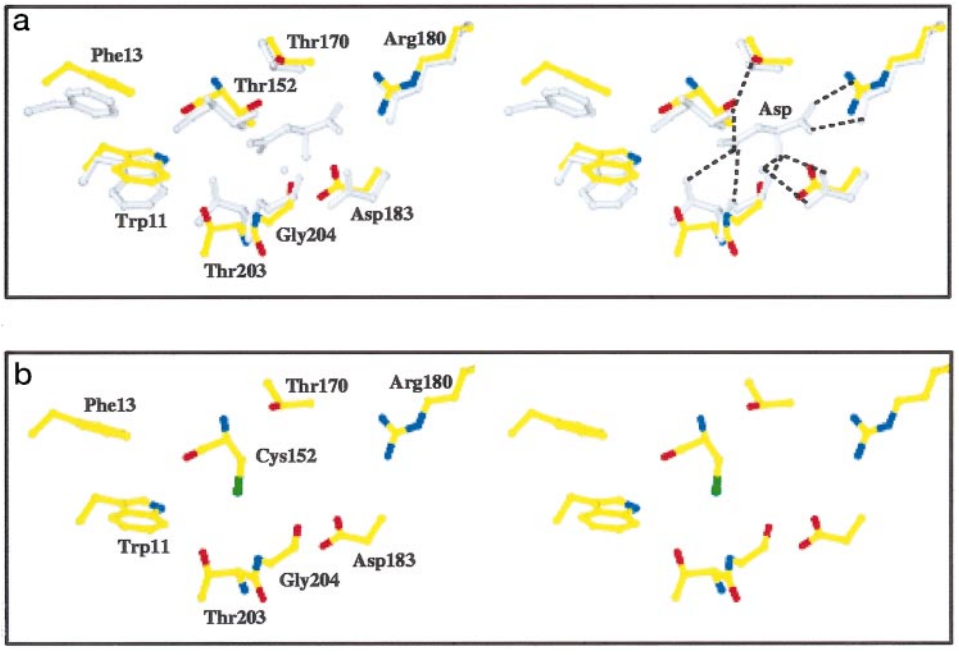

FIG. 4. **Stereo view of the active site of glycosylasparaginase.** *a*, stereo view of superimposition of the active sites between *Flavobacterium* (shown according to atom type: *yellow* for carbons, *blue* for nitrogens, and *red* for oxygens) and human (shown in *gray*) GA. Also shown is aspartate in the human enzyme/product structure (13). *Dashed lines* correspond to the hydrogen bonds described in the human structure. *b*, the same stereo view of active site in the T152C mutant. The color scheme is the same as the wild type in (*a*), except that sulfur of the thiol group is shown in *green*.

ture of the T152C mutant in its mature form at 2.1 Å and refined to an R_{free} of 28.06% and an R_{cryst} of 23.32% with all reflections (Table II). The structure of this mutant is essentially identical to that of the wild type enzyme with an r.m.s. deviation of 0.22 Å for all the main chain atoms and 0.25 Å for the active site atoms (Fig. 4*b*). This indicates that the reduction of reaction rate of this mutant is because of the change of chemical groups at the side chain of residue 152.

The active site of the T152C mutant also has the open conformation as described above. Like the wild type structure, the distance between the Cβ atoms of Cys^{152} and Thr^{203} is 2.0 Å further apart than in the human structure. Furthermore, the thiol group of Cys^{152} points in the opposite direction and is 2.9 Å removed from the wild type nucleophile Oγ of Thr^{152}. This appears to be because of a favorable packing of the thiol group into a small pocket formed between side chains of Cys^{168} and Thr^{203} and main chain atoms of the β-sheet bS1. Such an inactive conformation has also been observed in the glutaminase domain of glucosamine 6-phosphate synthase, where Cys^{1} is the wild type N-terminal nucleophile (27). In the native GA enzyme, packing of the γ-methyl group of Thr^{152} into this pocket, as well as a hydrogen bond formation between Oγ of Thr^{152} and Oγ of Thr170 (Fig. 4*a*), positions the nucleophile in an active conformation. We propose that in the presence of substrate, the thiol group switches to the active conformation by a rotation of 120 ° around the Cα-Cβ bond and a small angular adjustment around the Cα-C bond. Further studies are needed to determine whether the Cys^{152} adopts our proposed active conformation in the presence of substrate.

Quaternary Structure of GA—Bacterial GA forms a dimer of a/b heterodimers in the crystals (Fig. 5*a*). A similar quaternary structure is also observed in the crystals of human GA in different crystal packings (13). The surface interactions between pairs of heterodimers are extensive and mainly involve hydrogen bonds and hydrophobic contacts (Fig. 5, *b* and *c*). Basically, both heterodimers use the same hydrophobic surface

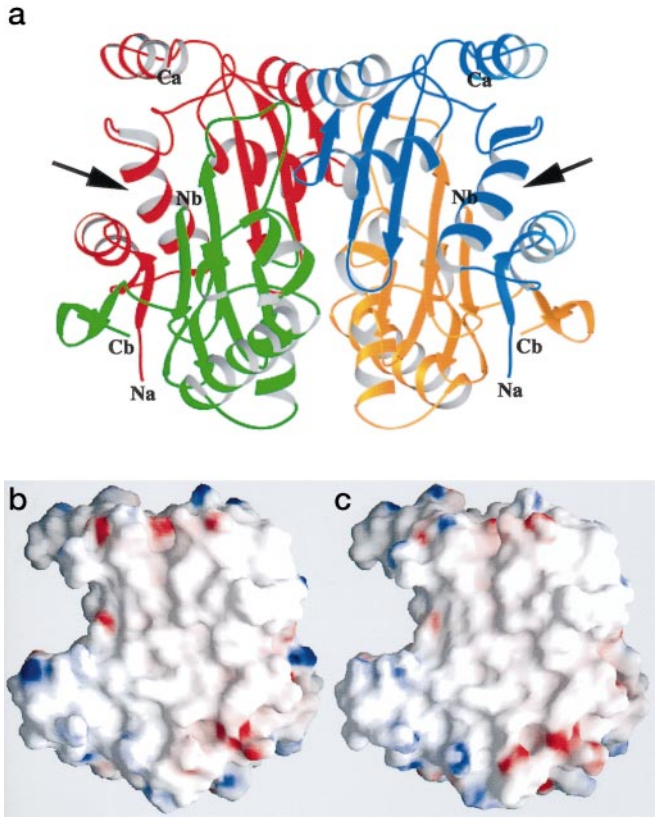

FIG. 5. **Dimer structure of GA a/b heterodimers from *F. meningosepticum.*** *a*, ribbon drawing of the quaternary structure of $(a/b)_2$ heterotetramer. One heterodimer is in *red* (a-subunit) and *green* (b-subunit), and the other is in *blue* (a-subunit) and *orange* (b-subunit). The site of enzyme activity (and putative site of autoproteolysis) in each heterodimer is pointed out by an *arrow*. The view orientation is similar to that of Fig. 2 in Ref. 13. *b*, surface potential representation of the dimerization interface of the *left* heterodimer in *a*, produced using Grasp (39). The orientation has been rotated 90 ° to the left along the vertical axis in *a*. *c*, surface potential representation of the dimerization interface of the *right* heterodimer in *a*, produced using Grasp (39). The orientation has been rotated 90 ° to the right along the vertical axis in *a*.

for the $(a/b)_2$ tetramer formation, reminiscent of hand shaking. The main interface interactions come from the strand aS4, helix bH2, and the loops between aH3 and aS4, aS4 and aH4, bS2 and bS3, and bH1 and bH2 in both heterodimers. In addition, the bacterial enzyme has unique interactions between the 7-residue insertion and the loop connecting β-strands aS2 and aS3. The human structure has substantially more interactions from the unique C-terminal loop of the a-subunit. Dimerization of a/b heterodimers sequesters a solvent-accessible surface area of 1882 Å^2 from each a/b heterodimer of the bacterial GA, compared with 2485 Å^2 for the human enzyme. The smaller interface and thus weaker dimer interaction between two heterodimers of bacterial GA may explain why no $(a/b)_2$ tetramers of the bacterial GA are observed on sizing columns.[5] Nonetheless, the existence of bacterial $(a/b)_2$ tetramers in the crystals suggest that the heterotetramers of bacterial GA may also exist in solution. This heterotetrameric structure of $(a/b)_2$ has been observed in solution for the human (8), chicken (2), and insect GA (3).

DISCUSSION

Aspartylglycosaminuria—The physiological importance of the glycosylasparaginase is revealed by the occurrence of a human genetic disorder, known as AGU, because of a deficiency of this lysosomal hydrolase (1). Many mutations in the GA gene that cause AGU have been reported, and more are likely to be found. However, a major obstacle to studying the consequences of these mutations is the difficulty to obtain recombinant human enzyme in sufficient quantities (28). In this study, four known AGU single mutations have been mapped onto the shared secondary structural elements between the bacterial and human enzymes (Fig. 2). A double mutant (human Arg^{138} to Gln and Cys^{140} to Ser) maps outside of the secondary structural elements and appears to result from the loss of a disulfide bond (Cys^{140}-Cys^{156}) that stabilizes a unique loop in the human enzyme. Thus, our work confirms the suitability of the bacterial enzyme as a model to analyze the consequences of mutations in AGU patients at the molecular level.

Structural Comparisons—Glycosylasparaginase belongs to a newly classified family of enzymes that have a novel N-terminal threonine, serine, or cysteine that provides the nucleophile in their reaction mechanism (11). Previously reported structures of this family include glutamine 5-phosphoribosyl-1-pyrophosphate amidotransferase from *Bacillus subtilis* (29), *Escherichia coli* penicillin amidohydrolase (30), the 20 S proteasome from the archaebacterium *Thermoplasma acidophilum* (31) and yeast (32), human glycosylasparaginase (13), and the glutaminase domain of *E. coli* glucosamine 6-phosphate synthase (27). All of these enzymes have a similar protein fold comprised of a sandwich of antiparallel β sheets surrounded on either side by layers of α helices. Many of these enzymes are activated by cleavage of the peptide bond to free the α-amino group to form the N-terminal nucleophile (10). A different protein fold has recently been described for the autoprocessing domain of *Drosophila* Hedgehog protein (33). It is an all β structure that is distinct from the GA structure but is related to the intein domain of PI-*Sce*I endonuclease (34, 35).

Enzyme Mechanism—Crystal structures described in this study, on the other hand, also raise questions about the detail mechanism of GA. Structure of the wild type GA from *F. meningosepticum* in its mature form has been determined at 2.2 Å resolution. Although the topology of the bacterial enzyme is very similar to that of the human structure, several significant differences have been observed. The active site of *Flavobacterium* GA is in an open conformation, whereas the human enzyme adopts a closed conformation that grasps the reaction product, aspartate (Fig. 4*a*). Moreover, the side chain of Thr^{203} may not be as important in stabilizing the negative oxyanion intermediate as previously suggested (13). This is consistent with a mutagenesis study in which replacement of Thr^{203} by Ala (T203A mutant) in the bacterial enzyme does not dramatically decrease the reaction rate (23). A three-dimensional structure of the enzyme-substrate complex is necessary to clarify the role of Thr^{203} side chain in the enzymatic mechanism.

In addition, we also report the structure at 2.1 Å resolution of a T152C mutant wherein the N-terminal nucleophile Thr^{152} of the b-subunit is replaced with Cys. The T152C mutant has a dramatically reduced rate of autoproteolysis or enzyme catalysis (23) and thus is a good candidate for future crystallographic studies of the precursor structure and enzyme-substrate complex. Similar to the glutaminase domain of glucosamine 6-phosphate synthase, Cys^{152} in the T152C mutant appears to be in an inactive conformation (27). We propose that binding of substrate would switch the thiol group into an active conformation.

Autoprocessing for Enzyme Activation—Cis-autoproteolysis involves the intramolecular catalytic cleavage of a peptide bond and is required to activate many enzymes (12). In addition to GA, these include penicillin acylase (30), proteasomes (31, 36, 37), as well as the *hedgehog* family of eukaryotic developmental

[5] C. Guan, unpublished results.

regulatory proteins (38). Autoproteolytic cleavage also serves as a mechanistic component for protein splicing (35). In contrast to the activation of zymogens, such as chymotrypsinogen and trypsinogen through proteolysis by another trypsin molecule, the autoproteolysis of GA is an intramolecular reaction (10). Human GA is also believed to undergo autoproteolysis to form the active enzyme but with some differences. First, the disulfide bridges in the human enzyme are essential for early folding and for autoproteolytic processing (15). In contrast, there are no disulfide bridges in the bacterial enzyme. Furthermore, part of the 31-residue insertion in the human enzyme is removed from the C terminus of the autoproteolyzed a-subunit in the lysosome by a second cleavage (6). No such trimming occurs for the bacterial enzyme.

This work reveals an $(a/b)_2$ quaternary structure that has been observed in solution or crystals of the eukaryotic GA. Furthermore, an amino acid substitution (equivalent to bacterial Ile^{186}) at this interface in the human enzyme disrupts the dimer formation of the precursor protein and also prevents proteolytic activation of the enzyme (15). Therefore, it appears that in the human enzyme dimerization of precursors is a prerequisite to trigger autoproteolysis. In contrast, only the a/b heterodimer is observed on sizing gels and columns for the bacterial GA.[5] Nonetheless, a dimer of a/b heterodimers exists in the crystals of bacterial GA (Fig. 5*a*) that is similar to the quaternary structure observed in the crystals of human GA (13). This raises the possibility that dimerization of bacterial GA, although it has not been observed yet, might also occur in solution.

Further studies are necessary to determine whether dimerization of the single chain precursor proteins occurs and, if so, to determine the significance of this dimerization in autoproteolysis. Unless there is a large conformational change as a result of autoproteolysis, the location of the key cleaved Thr^{152} in the enzyme active site suggests that the autoproteolytic site is near or overlaps with the active site. In line with intramolecular autoproteolysis, the two active sites in the dimer of GA are facing apart with the autoproteolyzed N-terminal threonines 32 Å away (arrows in Fig. 5*a*). The size and shape of the active site funnel also appear to be difficult for any proteolytic enzyme to approach Thr^{152} for peptide bond cleavage. However, it remains unclear how this dimerization triggers autoproteolytic activation of these enzymes. It is still possible that dimerization of precursor proteins results in a conformational change to trigger autoproteolysis. In our group, crystallographic studies on the precursor proteins are underway.

Acknowledgments—We thank Drs. G. G. Shipley, C. W. Akey, and C. J. McKnight for helpful discussions and comments on the manuscript, T. Cui for sharing unpublished results of Cys to Ser mutations in *Flavobacterium* GA, and members of the lab for helpful suggestions.

REFERENCES

1. Mononen, I., Fisher, K. J., Kaartinen, V., and Aronson, N. N., Jr. (1993) *FASEB J.* **7,** 1247–1256
2. Tollersrud, O. K., and Aronson, N. N., Jr. (1992) *Biochem. J.* **282,** 891–897
3. Liu, Y., Dunn, G. S., and Aronson, N. N., Jr. (1996) *Glycobiology* **6,** 527–536
4. Tarentino, A. L., Quinones, G., Hauer, C. R., Changchien, L.-M., and Plummer, T. H., Jr. (1995) *Arch. Biochem. Biophys.* **316,** 399–406
5. Risley, J. M., and Xia, Y. (1996) *FASEB J.* **10,** A1104 (abstr.)
6. Ikonen, E., Julkunen, I., Tollersrud, O. K., Kalkkinen, N., and Peltonen, L. (1993) *EMBO J.* **12,** 295–302
7. Fisher, K., Klein, M., Park, H., Vettese, M. B., and Aronson, N. N., Jr. (1993) *FEBS Lett.* **323,** 271–275
8. Kaartinen, V., Williams, J. C., Tomich, J., Yates, J. R., III, Hood, L. E., and Mononen, I. (1991) *J. Biol. Chem.* **266,** 5860–5869
9. Riikonen, A., Tikkanen, R., Jalanko, A., and Peltonen, L. (1995) *J. Biol. Chem.* **270,** 4903–4907
10. Guan, C., Cui, T., Rao, V., Liao, W., Benner, J., Lin, C.-L., and Comb, D. (1996) *J. Biol. Chem.* **271,** 1732–1737
11. Brannigan, J. A., Dodson, G., Duggleby, H. J., Moody, P. C. E., Smith, J. L., Tomchick, D. R., and Murzin, A. G. (1995) *Nature* **378,** 416–419
12. Aronson, N. N., Jr. (1996) *Glycobiology* **6,** 669–675
13. Oinonen, C., Tikkanen, R., Rouvinen, J., and Peltonen, L. (1995) *Nat. Struct. Biol.* **2,** 1102–1108
14. Halila, R., Baumann, M., Ikonen, E., Enomaa, N., and Peltonen, L. (1991) *Biochem. J.* **276,** 251–256
15. Riikonen, A., Rouvinen, J., Tikkanen, R., Julkunen, I., Peltonen, L., and Jalanko, A. (1996) *J. Biol. Chem.* **271,** 21340–21344
16. Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S., eds) pp. 56–62, SERC Daresbury Laboratory, Warrington, UK
17. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr.* **D50,** 760–763
18. McRee, D. E. (1992) *J. Mol. Graphics* **10,** 44–46
19. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr.* **A47,** 110–119
20. Brünger, A. T. (1992) *X-PLOR Manual: A System for X-ray Crystallography and NMR*, Version 3.1, Yale University Press, New Haven, CT
21. Brünger, A. T. (1992) *Nature* **355,** 472–475
22. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
23. Liu, Y., Guan, C., and Aronson, N. N., Jr. (1998) *J. Biol. Chem.* **273,** 9688–9694
24. Toone, E. J. (1994) *Curr. Opin. Struct. Biol.* **4,** 719–728
25. Barrett, A. J., and Rawlings, N. D. (1995) *Arch. Biochem. Biophys.* **318,** 247–250
26. Mesecar, A. D., Stoddard, B. L., and Koshland, D. E., Jr. (1997) *Science* **277,** 202–206
27. Isupov, M. N., Obmolova, G., Butterworth, S., Badet-Denisot, M.-A., Badet, B., Polikarpov, I., Littlechild, J. A., and Teplyakov, A. (1996) *Structure* **4,** 801–810
28. Mononen, I., Heisterkamp, N., Dunder, U., Romppanen, E.-L., Noronkoski, T., Kuronen, I., and Groffen, J. (1995) *FASEB J.* **9,** 428–433
29. Smith, J. L., Zaluzec, E. J., Wery, J.-P., Niu, L., Switzer, R. L., Zalkin, H., and Satow, Y. (1994) *Science* **264,** 1427–1433
30. Duggleby, H. J., Tolley, S. P., Hill, C. P., Dodson, E. J., Dodson, G., and Moody, P. C. E. (1995) *Nature* **373,** 264–268
31. Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995) *Science* **268,** 533–539
32. Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D., and Huber, R. (1997) *Nature* **386,** 463–471
33. Hall, T. M. T., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) *Cell* **91,** 85–97
34. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) *Cell* **89,** 555–564
35. Perler, F. B. (1998) *Cell* **92,** 1–4
36. Seemüller, E., Lupas, A., and Baumeister, W. (1996) *Nature* **382,** 468–470
37. Chen, P., and Hochstrasser, M. (1995) *EMBO J.* **14,** 2620–2630
38. Johnson, R. L., and Tabin, C. (1995) *Cell* **81,** 313–316
39. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11,** 281–293