Structural Characterization of *Pseudomonas* 7A Glutaminase–Asparaginase[†,‡]

Jacek Lubkowski,[§] Alexander Wlodawer,[§] Herman L. Ammon,[∥] Terry D. Copeland,[⊥] and Amy L. Swain[*,§]

Macromolecular Structure Laboratory and Special Program in Protein Chemistry, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, Maryland 21702-1201, and Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Received April 7, 1994[⊛]

ABSTRACT: The amino acid sequence and a 2-Å-resolution crystallographic structure of *Pseudomonas* 7A glutaminase–asparaginase (PGA) have been determined. PGA, which belongs to the family of tetrameric bacterial amidohydrolases, deamidates glutamine and asparagine. The amino acid sequence of PGA has a high degree of similarity to the sequences of other members of the family. PGA has the same fold as other bacterial amidohydrolases, with the exception of the position of a 20-residue loop that forms part of the active site. In the PGA structure presented here, the active site loop is observed clearly in only one monomer, in an open position, with a conformation different from that observed for other amidohydrolases. In the other three monomers the loop is disordered and cannot be traced. This phenomenon is probably a direct consequence of a very low occupancy of product(s) of the enzymatic reaction bound in the active sites of PGA in these crystals. The active sites are composed of a rigid part and the flexible loop. The rigid part consists of the residues directly involved in the catalytic reaction as well as residues that assist in orienting the substrate. Two residues that are important for activity reside on the flexible loop. We suggest that the flexible loops actively participate in the transport of substrate and product molecules through the amidohydrolase active sites and participate in orienting the substrate molecules properly in relation to the catalytic residues.

Amidohydrolases catalyze the hydrolysis of asparagine and glutamine to their acidic forms. These metabolic enzymes are expressed in large amounts in bacteria and vary in specificity toward their substrates [for review, see Wriston and Yellin (1973)]. Bacterial amidohydrolases are active as tetramers of identical protein chains with molecular weights in the range of 34 000–36 000 per monomer. Some amidohydrolases, such as those from *Escherichia coli* (EcA)[1] and *Erwinia chrysanthemi* (ErA), have a relatively high specificity for asparagine and are referred to as asparaginases. These enzymes are used clinically in the therapy of certain lymphocytic leukemias (Roberts et al., 1976; Gallagher et al., 1989). Other amidohydrolases can catalyze the hydrolysis of both glutamine and asparagine with comparable efficiency and are, therefore, called glutaminase–asparaginases. These enzymes, examples of which are amidohydrolases from the bacteria *Pseudomonas* 7A (PGA) and *Acinetobacter glutaminasificans* (AGA), are not clinically useful against leukemias due to side effects caused by reduction in patient glutamine levels (Benezra et al., 1972; Howard & Carpenter, 1972). Recently, however, PGA has been demonstrated to inhibit retroviral replication (Roberts & McGregor, 1991), and this property may be exploited for potential therapeutic use in the treatment of retroviral diseases.

Crystallographic investigations of EcA, ErA, and AGA reached completion in the last two years (Swain et al., 1993; Miller et al., 1993; Lubkowski et al., 1994). The primary goal of these studies is to compare the structures of asparaginases and glutaminase–asparaginases and search for a basis for their different specificities. The structures of EcA and ErA, with ligands bound in their active sites, were solved at medium (2.3 Å) and high (1.8 Å) resolution, respectively, revealing the structural details of the active sites of asparaginases. The structure of AGA was solved at a lower resolution (2.9 Å), preventing a clear interpretation of the active site environment of glutaminase–asparaginases. Therefore, we pursued a crystallographic study of PGA that had already been crystallized and characterized with respect to molecular symmetry by Patterson techniques (Wlodawer et al., 1977; Ammon et al., 1983). PGA crystals (space group $P2_12_12_1$) contain a PGA tetramer with near-perfect 222 symmetry in the asymmetric unit. The published primary structure information for PGA is very limited, with only the sequence of the first 26 amino acid residues of this enzyme available (Holcenberg et al., 1978). Additionally, in the past, there was no sufficiently correct model that could be used to solve the structure by molecular replacement, and only a single heavy atom derivative had been identified (Murphy, 1983). Therefore, further progress of the study was limited.

We report here the crystal structure of PGA solved at 2.0-Å resolution by molecular replacement using the recently completed structure of EcA as the probe model. Simultaneously, the amino acid sequence was determined by conventional chemical means, supported by electron density interpretation. The apparent state of the enzyme in the PGA crystals is with ligand [i.e., the product(s) of the enzymatic reaction] bound in the active sites of a small population of the molecules.

[†] This research was sponsored in part by the National Cancer Institute, DHHS, under Contract NO1-CO-74101 with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply indorsement by the U.S. Government.

[‡] The coordinates have been deposited with the Brookhaven Protein Data Bank (Accession Number 3PGA).

* To whom correspondence should be addressed.

[§] Macromolecular Structure Laboratory, FCRDC.

[∥] University of Maryland.

[⊥] Special Program in Protein Chemistry, FCRDC.

[⊛] Abstract published in *Advance ACS Abstracts,* July 15, 1994.

[1] Abbreviations: EcA, *Escherichia coli* asparaginase; ErA, *Erwinia chrysanthemi* asparaginase; PGA, *Pseudomonas* 7A glutaminase–asparaginase; AGA, *Acinetobacter glutaminasificans* glutaminase–asparaginase; MR, molecular replacement.

0006-2960/94/0433-10257$04.50/0

Table 1: Crystal Data

	Crystal I
space group[a]	$P2_12_12_1$
unit cell parameters (Å)[a]	a = 118.26, b = 130.75, c = 85.09
R_{merge} (%)	7.3
total no. of accepted observations ($F \geq 2\sigma_F$)[b]	
full	119294
partial	92859
all	212153
no. of independent reflections	81878

Intensity and Completeness Distribution (Cumulative Values)

resolution (Å)	$\langle F^2 \rangle$	$\langle \sigma_F \rangle$	obsd	theoretical	%
15.0	355.3	23.2	236	266	88.7
10.0	473.2	25.6	779	830	93.9
7.5	430.6	24.3	1 812	1 884	96.2
5.0	222.9	13.6	5 931	6 076	97.6
3.0	173.0	11.8	26 603	27 085	98.2
2.5	37.1	5.3	44 941	46 350	97.0
2.0	14.4	3.9	77 094	89 644	86.0
1.9	6.9	3.5	80 641	104 350	77.3
1.8	5.8	3.2	81 878	122 335	66.9

	Crystal II
R_{merge} (%)	10.4
total no. of accepted observations ($F \geq 2\sigma_F$)	
full	90992
partial	40187
all	131179
no. of independent reflections	60727

Intensity and Completeness Distribution (Cumulative Values)

resolution (Å)	$\langle F^2 \rangle$	$\langle \sigma_F \rangle$	obsd	theoretical	%
15.0	176.9	13.2	162	266	60.9
10.0	242.4	15.1	554	830	66.9
7.5	236.6	15.6	1 317	1 884	70.0
5.0	131.0	10.0	4 520	6 076	74.5
3.0	69.3	9.4	21 619	27 085	79.8
2.5	43.2	8.4	36 014	46 350	77.7
2.0	28.0	8.1	58 153	89 644	64.9
1.9	19.4	9.2	59 688	104 350	57.2
1.8	18.4	12.4	60 727	122 355	47.6

[a] For crystals I and II. [b] F is the observed structure factor and σ_F is the associated standard deviation.

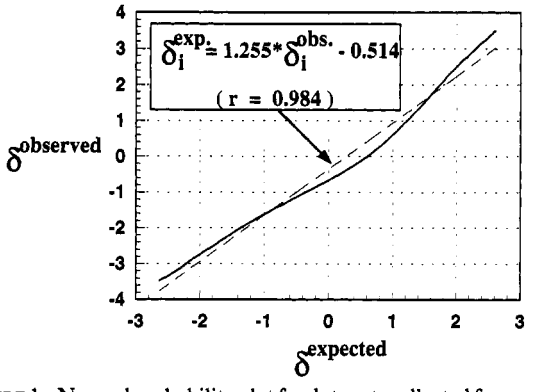

FIGURE 1: Normal probability plot for data sets collected from crystals I and II of PGA. For a description about the analysis with normal probability plots, see Abrahams and Keve (1971) and Stout and Jensen (1989).

MATERIALS AND METHODS

PGA was purified according to the procedure of Roberts (1976) and was stored as lyophilized powder. Before use in sequence determination, the sample was dissolved in water and dialyzed against 5% acetic acid. Aliquots for various procedures were removed and lyophilized.

Cyanogen Bromide Cleavage. A lyophilized aliquot of PGA was dissolved in 0.5 mL of 70% formic acid containing a crystal of CNBr and reacted in the dark for 48 h. After the solvent was removed with a stream of N_2 in a chemical hood, the sample was dissolved in 3 × 100 μL of water and blown dry; then the sample was dissolved in 2 × 2 mL of 50% acetonitrile and lyophilized. The sample was then dissolved in 200 μL of 8 M guanidine hydrochloride.

Enzymatic Digestion. Lyophilized aliquots were dissolved in 0.5 mL of 0.1 M ammonium bicarbonate and digested for 36 h with Lys-C endoproteinase or with Arg-C (Boehringer Mannheim) at a ratio of 1:50 (enzyme:substrate).

Reversed-Phase High-Performance Liquid Chromatography. The peptide samples were adjusted to pH 2 and applied to a Waters 3.9 × 300 mm Bondapak C_{18} column in a Waters 625 LC system. A gradient of 2–65% acetonitrile was developed over 120 min at a flow rate of 1 mL/min. Peptide peaks were collected manually.

Amino-Terminal Sequence Analysis. Intact protein and peptides were applied to derivatized disks and sequenced in an Applied Biosystems 477A sequencer equipped with an on-line 120A analyzer (Adachi et al., 1994).

Carboxyl-Terminal Sequence Analysis. Intact protein was digested with carboxypeptidase-P, and released amino acids were identified on a Beckman 6300 analyzer.

Crystallization, Data Collection, and Data Statistics. All procedures preceding data collection were performed at 4 °C. The PGA lyophilisate [~40% (w/w) protein] was dissolved in 10 mM phosphate buffer (pH 7.2) to 30–35 mg/mL. The solution was clarified by centrifugation, and the supernatant was dialyzed against the same phosphate buffer for 3 days with three buffer changes. The resulting sample was again centrifuged, and 2-methyl-2,4-pentanediol (MPD) was added to the supernatant to 30–33% (v/v). After 1–2 h, traces of precipitate that had formed were removed by centrifuging twice. Crystals were grown from the supernatant by the sitting drop method with a 5% higher concentration of MPD in the reservoir. Diamond-shaped crystals first appeared after 1 day and continued growing for 3–5 days. Since the crystals were unstable during manipulations that involved changes in temperature or environment, we stabilized them by cross-linking with glutaraldehyde (Quiocho & Richards, 1964). The crystals used for data collection (0.5 × 0.5 × 0.35 mm) were soaked in 0.03–0.04% (v/v) glutaraldehyde in mother liquor for 1.5–2.5 h. After being rinsed in mother liquor, the crystals were mounted in glass capillary tubes, and the temperature of the crystal assemblies was increased to ambient over 10–12 h.

X-ray diffraction data were collected from two crystals by using an R-axis II image plate detector mounted on a Rigaku RU-200 rotating-anode generator, operated at 50 kV and 100 mA. The crystals were stable for only 10 h of continuous X-irradiation, after which time the resolution of the data decreased substantially. After a 7–9-h break from X-ray exposure, the diffraction power of the crystals was restored almost completely. Data collection was performed over 4

Table 2: PGA Refinement Statistics

R-factor	0.165
resolution (Å)	10.0–2.0
no. of reflections	78178
no. of atoms with occupancy = 1.0	9366
no. of water molecules	407
rms deviations from ideality	
distance restraints (Å)	
bond distance	0.019
angle distance	0.059
chiral center restraints (Å^3)	0.290

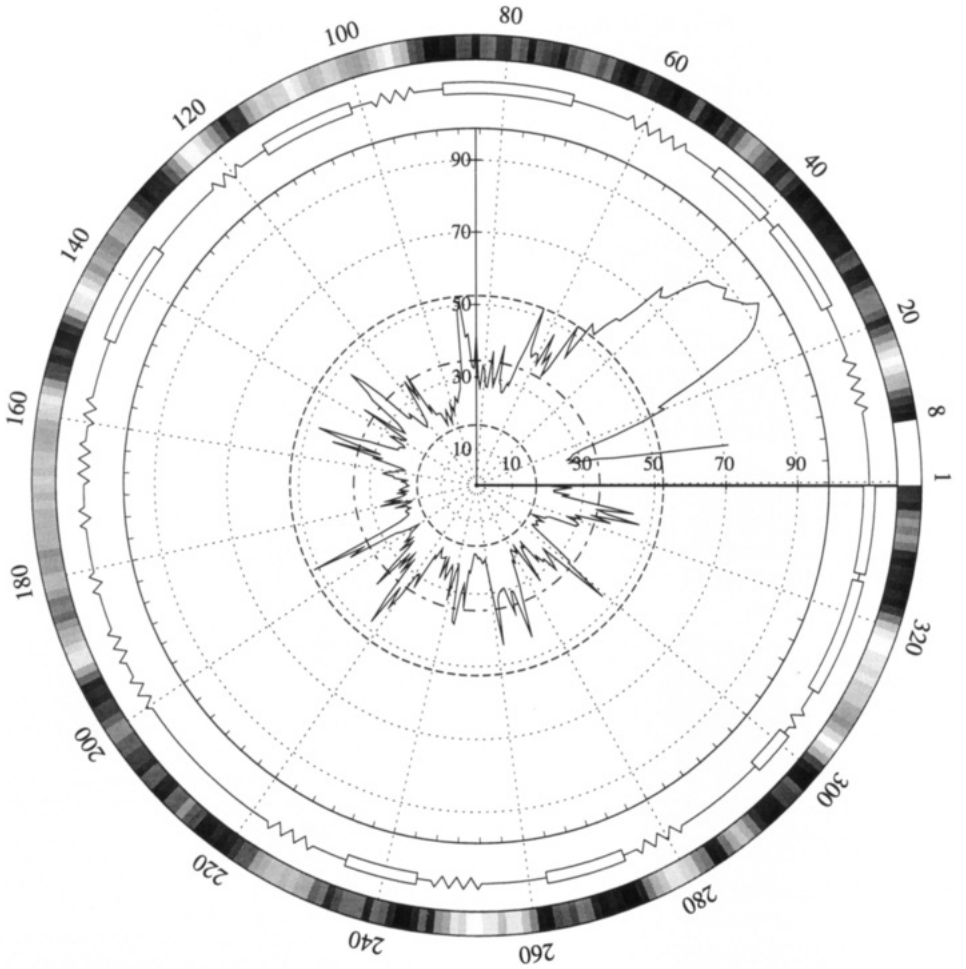

FIGURE 2: Circular representation of three characteristics as a function of residue number for PGA. The temperature factor (B) distribution is shown in the internal area in the form of a "spider" plot. The value of B used for each residue (except those in the loop from residues 20 to 40) was obtained by double averaging, first over all atoms of the residue and then over the four monomers of PGA. The scale is shown on the horizontal and vertical radii, and concentric dotted circles are provided to aid in the identification of B for each residue. The three thick-dashed circles have radii equal to $B_{av} - \sigma_B$, B_{av}, and $B_{av} + \sigma_B$, where B_{av} is the B averaged over all residues in the monomer and σ_B is the standard deviation of B. The spider plot is surrounded by a diagram of secondary structure elements, with β-strands shown as zigzag lines and α-helices as rectangles. The most external ring is a gray-scale representation of the modified solvent-accessible area (A) of each residue, with black equal to 1 and white equal to 0. The A values for every atom in the tetramer were calculated with program MOLSV from the QCPE collection (Smith, 1990). The following values were used for atomic radii (Connolly, 1983): main chain, $r(C_\alpha) = 1.70$ Å, $r(O) = 1.52$ Å, $r(C) = 1.80$ Å, and $r(N) = 1.55$ Å; side chain, r(all atoms) = 1.80 Å; probe radius (solvent molecule) = 1.4 Å. Initial values of A_i (value A for residue i) were obtained by summation of A for all atoms of residue i and then averaging over the four monomers. The value of every A_i was then modified (to give smoother distribution) according to the formula $A_i^{mod} = 0.075A_{i-2} + 0.175A_{i-1} + 0.5A_i + 0.175A_{i+1} + 0.075A_{i+2}$, and the modified values were finally scaled to the range $\langle 0, 1 \rangle$, where the minimum and maximum values of A^{mod} were normalized to values of 0.0 and 1.0. For terminal residues (i = 8, 9, 336, and 337) the A_i^{mod} values were calculated similarly. The modification of A values removes possible artifacts that can make it difficult to analyze the distributions. For example, when a large residue (e.g., Gln) located on the surface of the protein is linked to small residues (such as Gly or Thr), the size of the large residue will mask the true exposure of the small residues to solvent molecules. Values of B and A for the first seven residues are not shown because these residues were excluded from the model. Distributions of both B and A values were very similar for all four monomers, justifying the averaging of these quantities. For the loop region, residues 20–40, B and A values are those for monomer A. The averaged values could not be calculated for the loop region, however, because it was fit to density only in monomer A.

days (per crystal) with 7–9-h breaks after each 10-h X-ray exposure to restore diffraction. The statistics for the two data sets are shown in Table 1.

The two data sets were merged using the XSCALE option from the XDS data processing package (Kabsch, 1988). The completeness (97.2%; 87 144 independent reflections at 2.0-Å resolution) and redundancy (total number of observations = 343 332) of the merged data set was higher than those for each of the two separately collected sets. This merged data set was used in the early model building and refinement processes. However, since statistics indicated a significantly lower quality of both the merged data and the data set for crystal **II**, in comparison to the data set for crystal **I**, only the data collected from crystal **I** were used in the later stages of refinement.

Solution and Refinement of the Structure. The structure was solved by molecular replacement (MR) using the crystal structure of EcA as the probe model. Rigid body, conventional, and simulated annealing refinements were performed using X-PLOR (Brünger et al., 1990) followed by several cycles of conventional refinement with PROFFT (Finzel, 1987). These procedures decreased the R-factor from an initial value of 0.34 to 0.29. The first set of electron density maps was calculated on the basis of this model. Since the asymmetric unit of PGA contains four identical protein chains, an averaging procedure was used to improve the maps. The first step of rebuilding involved converting the amino acid identities of the MR model, which corresponded to the sequence of EcA, to those that matched the electron density maps. The PGA sequence was not available at that time. Additionally,

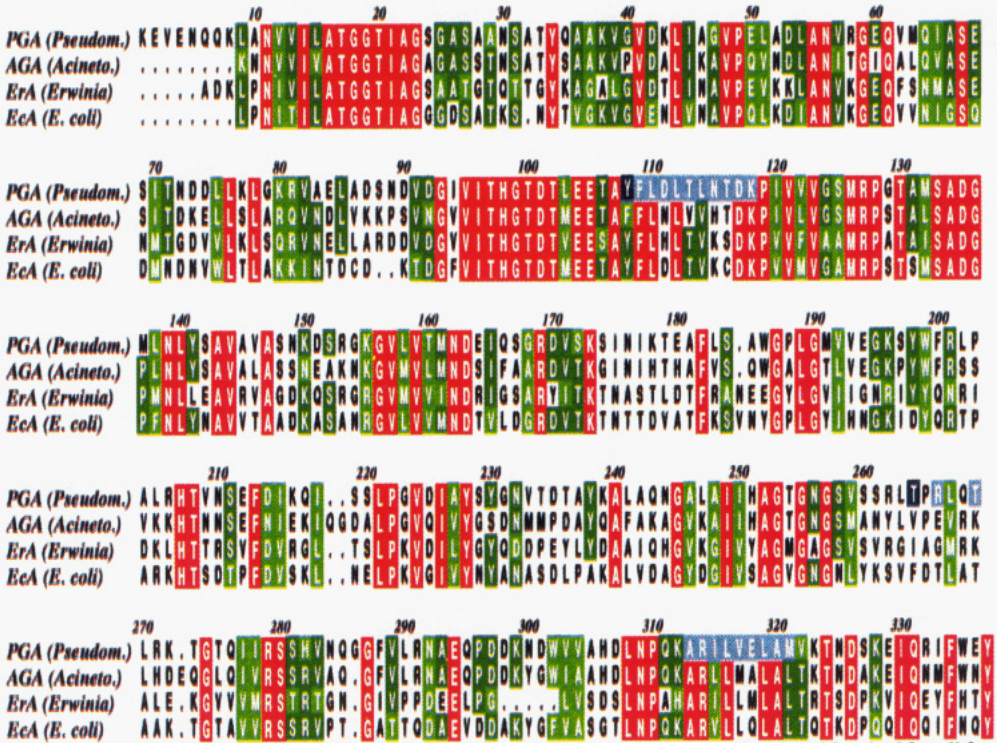

FIGURE 3: Amino acid sequence alignment of PGA, EcA, ErA, and AGA, with PGA numbering. Residues identical for all four amidohydrolases are shown in red. Residues identical for three amidohydrolases are shown in dark green, and conserved residues are shown in light green. The residues that were not identified by chemical sequencing are shown in light blue, and those for which the identities based on the electron density did not agree with the tentative chemical sequencing assignment are shown in dark blue.

the geometry was improved for the model of the entire tetramer. The model rebuilding process was followed by refinement with X-PLOR and PROFFT, and over many cycles of rebuilding and refinement, the value of the *R*-factor decreased to 0.21.

The PGA amino acid sequence that was being concomitantly determined was used to replace the estimated sequence. Ultimately, 93% of the PGA amino acid sequence was identified by conventional chemical methods. Approximately 70% of the sequence estimated from the electron density matched the conventionally determined sequence. The sequence information, together with the model geometry improvement, was reflected by a decrease in the *R*-factor value to 0.18. Although the geometry and the fit to the electron density were satisfactory for most of the model at this stage, the positions of ~20 residues per monomer near the active sites were still uncertain. Further improvement of the model was limited, so we attempted to improve the quality of the data. In order to estimate the quality of the merged data set, we prepared a normal probability plot (Abrahams & Keve, 1971) for the initial data sets, which had 46 958 common reflections. This plot, together with its linear least-squares approximation, is shown in Figure 1. The significant deviation of the intercept value from zero for the linear approximation indicates that the two compared data sets are statistically different and cannot be scaled to produce a qualitatively superior data set. Additionally, both the completeness and R_{merge} favor the data set collected from crystal **I** over that for crystal **II** (Table 1). Therefore, we exchanged the merged data set for that collected from crystal **I** only.

More rigorous geometry restraints were introduced, and subsequently the model was completed by inclusion of water molecules. The statistics for final refinement are presented in Table 2. A quantitative characterization of PGA is shown in Figure 2. The coordinates resulting from this refinement have been deposited with the Brookhaven Protein Data Bank.

RESULTS

Primary Structure. The amino acid sequence of PGA and its alignment with the sequences of EcA, ErA and AGA are given in Figure 3. Each monomer of PGA consists of 337 amino acids and is longer than that of the other amidohydrolases by several residues located at the amino-terminal end. By the conventional methods described above, 314 amino acids were identified, and most of them were confirmed by the degeneracy of the peptides produced in the sequencing procedures. The seventh PGA residue, previously identified as His (Holcenberg et al., 1978), has now been determined to be Gln. The sequence alignment and electron density shapes were used as additional identification criteria. This procedure led to the correct identification of residues such as Tyr108 and Thr264, which were tentatively assigned by the chemical sequencing methods as Phe and Gly, respectively. For 21 residues that were not characterized by conventional methods, the electron density and the sequence alignment were the only sources used for defining the sequence. Although the identities of these 21 residues are not as certain as the rest of the sequence, the electron density maps in these regions are clearly defined.

Secondary and Tertiary Structure. The topology of each PGA monomer is identical to that of the other known asparaginases with the exception of a disordered flexible loop (residues 20–40) near the active site. Sufficient electron density to trace this loop was observed for only one monomer (Figure 4A), designated monomer A. The conformation of the active site loop in PGA is significantly different with respect to that observed in other asparaginase structures. In the PGA structure reported here, the loop contains a helical element than can be considered as one bent helix or as two separate, short helices. This topological feature is described in detail in the next sections. The distribution of secondary structure elements as a function of residue numbers is shown in Figure 2.

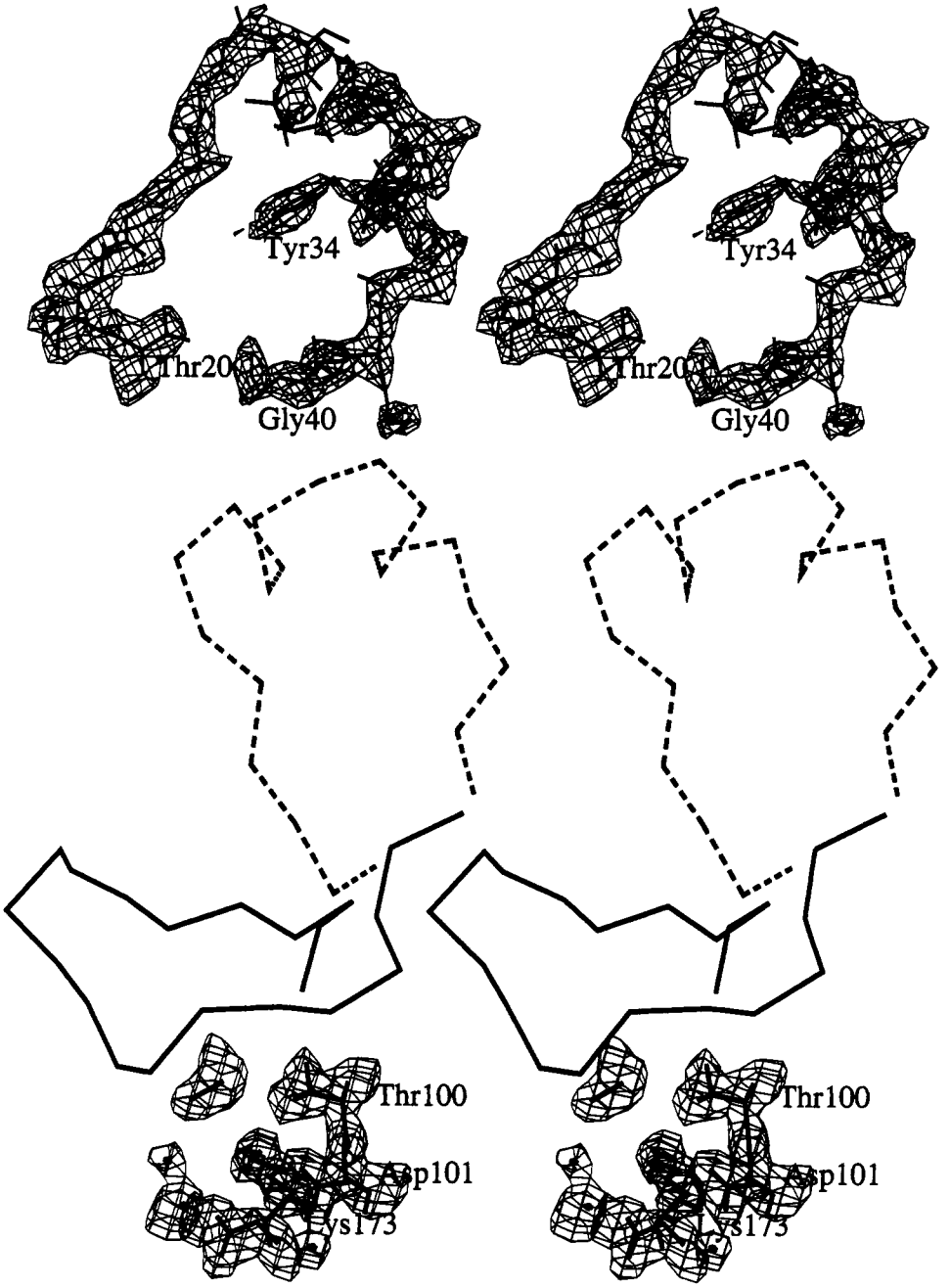

FIGURE 4: Active site loop of the PGA monomer. (A, top) Residues 20–40 are shown with $2F_o - F_c$ electron density contoured at the 0.9σ level. (B, bottom) C_α representations of residues 20–40 are shown in the open conformation (dashed line), as observed from the electron density of monomer A, and in the closed conformation (solid line), as modeled according to traces of electron density and the positions of the active site loops in EcA and ErA. The catalytic residues are shown with $2F_o - F_c$ electron density contoured at the 1.0σ level.

Each PGA monomer is divided into two domains, linked by a single strand of ~20 residues. The amino-terminal domain has 10 β-strands and four α-helices. An eight-stranded β-sheet has two helices on each side. Five β-strands of the sheet are parallel and four are antiparallel. As in all asparaginase structures determined thus far, the connection between the fourth and fifth parallel β-strands is left-handed (Miller et al., 1993). A β-hairpin extends away from the sheet between the sixth and ninth strands. The smaller carboxyl-terminal domain has a four-stranded parallel β-sheet with four flanking α-helices.

No electron density was observed for the seven amino-terminal residues in any of the four protein chains. There is no evidence that the amino-terminal part of the chain folds into any type of secondary structure or adheres to the rest of the protein globule. These residues are most likely disordered and have been excluded from the model.

Active-Site Architecture. There are four active sites in the PGA tetramer, situated between adjacent monomers and related by noncrystallographic 222 symmetry. Each of the four active sites is formed by a topological switchpoint (Brändén & Tooze, 1991) between the first and third parallel β-strands in the amino-terminal domain and residues extending from loops of the carboxyl-terminal domain of another monomer. Although the active sites in PGA are exposed to solvent, several residues known to be involved in catalysis are particularly rigid with side chains clearly defined in the electron density (Figure 4B).

There were several peaks of $F_o - F_c$ electron density within the active sites of PGA. Electron density corresponding to aspartate molecules was observed within the active sites both of EcA and ErA (Swain et al., 1993; Miller et al., 1993).

Therefore, aspartate and glutamate molecules were alternately modeled to the $F_o - F_c$ density peaks in the PGA active sites. After refinement, large peaks of negative $F_o - F_c$ electron density were observed in the positions where the ligand molecules had been modeled. As an alternative, water molecules were placed into the $F_o - F_c$ electron density peaks. Following refinement, the water molecules were appropriately centered in $2F_o - F_c$ electron density peaks. Most of the active site water molecules have relatively high *B*-values (>35 Å^2), which is the result of the overall disorder of the active sites.

Relatively continuous and interpretable electron density for the active site loop (residues 20–40), which contains residues known to be important for activity, was observed for only one PGA monomer, A (Figure 4A). For the same monomer we also observed traces of $2F_o - F_c$ and $F_o - F_c$ electron density indicating an alternate path of the protein chain for a closed conformation of the active site loop, as observed for other asparaginases (Swain et al., 1993; Miller et al., 1993; Lubkowski et al., 1994). Although the electron density for the loop in the closed conformation was too sparse to fully describe its geometry, this density indicated the presence of an alternate conformation for the active site loop. For the other three monomers (B, C, and D), we observed peaks of electron densities along the paths of both the open and closed conformations of the loop. Therefore, we performed the refinement of the PGA structure with two alternate conformations of the loop for each monomer, varying the relative occupancies. For monomer A, the open conformation of the loop was fit to the electron density maps. In the cases of monomers B, C, and D the loops were modeled (without tracing electron density) in the same open conformation as for monomer A. The loops for all four monomers were also modeled (without tracing electron density) in the closed conformations as they were observed for other asparaginases (Swain et al., 1993; Miller et al., 1993; Lubkowski et al., 1994). Although the refinements of PGA with alternate loop conformations slightly decreased the value of the *R*-factor, they did not improve the clarity of the electron density maps. Therefore, in the final refinement, we included both conformations of the loop for all monomers but, with the exception of the loop in the open conformation for monomer A, we assigned zero occupancies. Although atoms with occupancies equal to 0 do not affect refinement results, we included the atoms for completeness of the PGA model. Our intention for including these atoms is to reflect the evidence we observe for alternate geometries of the flexible loop in all monomers.

Quaternary Structure. The tetramer is roughly spherical in shape and is the only active form of the enzyme (Shifrin et al., 1971). This compact construction, as well as enzymatic activity, is preserved by intermonomer contacts. The contact is defined here as the presence of two atoms in different monomers, separated by a distance less than or equal to 3.5 Å. There are 100 contacts between monomers A and C and 91 contacts between monomers B and D, forming an arrangement of two dimers within the tetramer. The number of contacts between any of the other two monomers is approximately one-third of the "intradimer" contacts. These contacts must be equally important, however, because only the tetrameric form of the enzyme exhibits activity (Cammack et al., 1972).

A ribbon diagram of the PGA tetramer is shown in Figure 5. The flexible loop that was observed in monomer A was superimposed on the other monomers. The loop, as observed in monomer A, makes contacts with a symmetry-related tetramer that presumably fix the loop in place in a significant

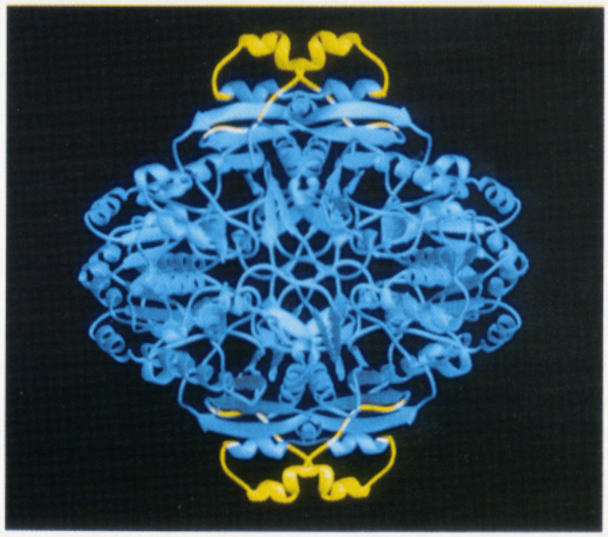

FIGURE 5: PGA tetramer. The four protein chains are shown in blue, except for the mobile active site loops in the open conformation, which are represented in gold.

population of the molecules in the crystals. Since the crystallographic asymmetric unit contains an entire tetramer of PGA, each loop forms different contacts with symmetry-related molecules. The overall effect of these interactions may be energetically favorable only in the case of monomer A, providing a plausible explanation for why the distribution of positions of the loop is different for each monomer.

DISCUSSION

Sequence/Activity. The four amidohydrolase sequences aligned in Figure 3 have 100 identical residues and 78 that are conserved. The numbers of identical residues for PGA and AGA, PGA and EcA, and PGA and ErA are 203, 159, and 153, respectively. The PGA amino acid sequence is most similar to that of AGA. This observation is consistent with the fact that these enzymes have the most similar activity profiles with respect to glutaminase vs asparaginase activity. Both enzymes are most active in the pH 6–9 range. PGA has a glutaminase:asparaginase activity ratio of ~2.0 (Roberts, 1976); for AGA this value is ~1.3 (Roberts et al., 1972).

Nine positions in the sequence, which are identical for the other three amidohydrolases, are different in the case of PGA. Seven of these substitutions are relatively conservative: Ala29 from Thr, Gly47 from Ala, Leu114 from Val, Ser172 from Thr, His282 from Arg, Met321 from Leu, and Val322 from Thr. The other two substitutions are not conservative: Ala83 from Asn and Met137 from Pro. However, no conformational rearrangements are observed as a direct result of any of these substitutions.

Active Site. Discussion of the active site loop is based on the geometry and topology determined for that in monomer A where it is observed in the open conformation. Two of several residues known to be crucial for activity of amidohydrolases (Bagert & Röhm, 1989; Harms et al., 1991), $Thr20_{12}$ and $Tyr34_{25}$, are located on this loop (EcA numbering is given in subscripts). The PGA monomer is shown superimposed on an EcA monomer in Figure 6A. In the case of the EcA model, with the loop in the closed conformation, the Thr12 and Tyr25 side chains are oriented directly into the active site region, and interactions with the surrounding solvent are excluded. Although these two residues in the PGA model (Thr20 and Tyr34) are still oriented toward the protein core, both are located significantly further from the active site and are

FIGURE 6: Superposition of PGA on EcA. (A, top) Superposition of PGA monomer A (blue) on an EcA monomer (red). (B, bottom) Closeup of the superimposed active site loops from PGA (blue) and EcA (red).

exposed to solvent. This shift in the position is largest for the Tyr residue, which is located almost at the tip of the loop (Figure 6B). An additional difference in the geometry of the two loop conformations is the presence of a helical fragment in the open conformation. The open loop participates in crystal contacts that may stabilize the observed conformation. Specifically, the side-chain and main-chain atoms of Thr33 interact through hydrogen bonds with atoms of Asn292 and Asp298 from a symmetry-related tetramer (Figure 7).

Although neither substrate nor product was included in the crystallization medium, no special effort was made to remove them from the PGA sample. If substrate or product were present, it would likely be bound in the active sites. This was indeed the case for the structure of EcA (Swain et al., 1993). Although continuous electron density for the open conformation of the loop was observed only for monomer A, significant traces of electron density were observed for a closed conformation. We postulate that the open form of the loop corresponds to an unoccupied active site and that the closed

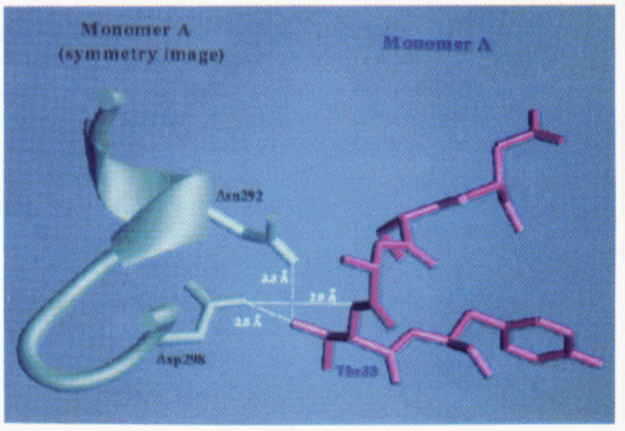

FIGURE 7: Crystal contacts involving the active site loop observed in monomer A.

form corresponds to the state of the active site with enzymatic reaction product(s) bound. Since we do not observe continuous

electron density for the closed position of the active site loop, and since we do not observe clearly defined density within the active sites, as we do in the cases of EcA and ErA, we conclude that the active sites of PGA have a low occupancy of ligand.

Our results showing that the loop (residues 20–40) is very flexible are consistent with NMR studies. Bagert and Röhm (1989) used ^{1}H NMR to examine the effect of aspartate binding on active site residues of EcA. They found that a Tyr residue is a constituent of the active site. Their observation was confirmed by crystallographic analyses of EcA [Tyr25 (Swain et al., 1993)], ErA [Tyr29 (Miller et al., 1993)], and AGA [Tyr26 (Lubkowski et al., 1994)]. The analogous residue in PGA, Tyr34, is located on the tip of the loop. Additionally, Bagert and Röhm (1989) observed broadening of resonances contributed by aromatic residues upon ligand binding, indicating a local conformational change. This may be a reflection of the active site loop motion resulting from ligand binding that we observe statically in the crystalline state. If there are ligands bound to PGA in the crystals, the distribution of occupancies is probably random, based on the fact that there is no cooperativity between asparaginase active sites (Citri et al., 1972). Diffraction data collected from PGA crystals with ligand present at high occupancy will likely reflect more order.

Quantitative evidence for the flexibility of the active site loop is given in Figure 2. The values calculated for two different parameters, averaged *B*-factor (*B*) and solvent accessibility (*A*), provide consistent evidence for the inherent mobility of this loop. Residues with higher *B* values are located on the protein surface. Residues with low *A* values indicate their participation mainly in hydrophobic or protein–protein hydrogen bond interactions such as those in helices. Since asparaginases exist in solution as tetrameric aggregates, the characteristics represented in Figure 2 may be more general than for the crystalline state. The most notable feature is the peak in the *B*-value plot in the range of residues 20–40, which corresponds to the flexible, disordered active site loop of PGA. The high values of *B* also reflect some uncertainty of the model in this area. For all other residues, the values of *B* are distributed around the mean (35 Å^2), within one standard deviation.

The rigid part of the active site of PGA is clearly discernible from the electron density (Figure 4B). Residues Thr100$_{89}$, Asp101$_{90}$, and Lys173$_{162}$, which are postulated to participate directly in the enzymatic reaction (Swain et al., 1993), are clearly defined in all four PGA active sites and are located in almost identical positions and orientations compared to those residues in EcA, ErA, and AGA. This absolute conservation of identity and conformation in the midst of surrounding disorder strongly suggests a common enzymatic mechanism for deamidation.

Enzymatic Mechanism of Asparaginase. The preliminary mechanism of the enzymatic reaction was proposed as a result of the structural studies of other asparaginases (Swain et al., 1993; Miller et al., 1993). The mechanism is reminiscent of that of serine proteases (Blow et al., 1969). However, Thr, rather than Ser, is postulated to play the role of nucleophile in asparaginases (Swain et al., 1993; Miller et al., 1993). A neighboring Lys may serve as a base to enhance the nucleophilicity of the Thr. An Asp juxtaposed to the Lys could stabilize the protonation state of Lys during catalysis. In PGA, the identities of the catalytic triad are thus Thr100, Lys173, and Asp101 (Figure 4B).

Each asparaginase active site is formed by parts of the protein with significantly different motion potential. The core part, which consists of Ser67$_{58}$, Glu68$_{Gln59}$, Met126$_{115}$, besides the three catalytic residues, and Ser258$_{Asn248}$ and Glu294$_{283}$ from the adjacent monomer, is quite rigid. These side chains have well-determined conformations. However, the loop, which includes residues Thr20 and Tyr34, is exposed to the solvent and is very mobile. The complete active site is intact when the flexible loop assumes the closed conformation, presumably stabilized by bound ligand.

We suggest the following update to the mechanism proposed previously (Swain et al., 1993): The flexible loop near the active site plays a dual role. In the open conformation, Tyr34 is exposed to the solvent environment, possibly playing a role in recognition by the substrate. The flexibility of this part of the active site significantly increases the exposure of the catalytic residues to substrate in the solvent medium. In the later stage of the catalytic process, the loop undergoes a series of complex conformational changes, and several residues assist in determining the proper orientation of the substrate molecule with respect to the rigid part of the active site. During the reaction, the closed conformation of the loop traps the substrate molecule, maintaining it in a fixed position. After catalysis, interactions between the product and the carbonyl oxygen of Ser125 create a charge repulsion (Miller et al., 1993). Subsequently, the loop refolds from its closed position, and the products leave the reaction center.

ACKNOWLEDGMENT

We gratefully acknowledge the invaluable efforts of Kenan C. Murphy in the purification of PGA and thank Drs. Mariusz Jaskólski, Maria Miller, and J. K. Mohana Rao for their comments and Anne Arthur for editorial assistance.

REFERENCES

Abrahams, S. C., & Keve, E. T. (1971) *Acta Crystallogr. A27*, 157–165.

Adachi, Y., Pavlakis, G. N., & Copeland, T. D. (1994) *J. Biol. Chem. 269*, 2258–2262.

Ammon, H. L., Murphy, K. C., Sjölin, L., Wlodawer, A., Holcenberg, J. S., & Roberts, J. (1983) *Acta Crystallogr. B39*, 250–257.

Bagert, U., & Röhm, K. H. (1989) *Biochim. Biophys. Acta 999*, 36–41.

Benezra, D., Pitaro, R., Birkenfeld, V., & Hochman, A. (1972) *Nature, New Biol. 236*, 80–82.

Blow, D. M., Birktoft, J. J., & Hartley, B. S. (1969) *Nature 221*, 337–340.

Brändén, C., & Tooze, J. (1991) in *Introduction to Protein Structure*, p 51, Garland Publishing, Inc., New York and London.

Brünger, A. T., Krukowski, A., & Erickson, J. (1990) *Acta Crystallogr. A46*, 585–593.

Cammack, K. A., Marlborough, D. I., & Miller, D. S. (1972) *Biochem. J. 126*, 361–379.

Citri, N., Kitron, N., & Zyk, N. (1972) *Biochemistry 11*, 2110–2116.

Connolly, M. L. (1983) *Science 221*, 709–713.

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

Gallagher, M. P., Marshall, R. D., & Wilson, R. (1989) *Essays Biochem. 24*, 1–40.

Harms, E., Wehner, A., Aung, H.-P., & Röhm, K. H. (1991) *FEBS Lett. 285*, 55–58.

Holcenberg, J. S., Ericsson, L., & Roberts, J. (1978) *Biochemistry 17*, 411–417.

Howard, J. B., & Carpenter, F. H. (1972) *J. Biol. Chem. 247*, 1020–1030.

Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 916–924.

Lubkowski, J., Wlodawer, A., Housset, D., Weber, I. T., Ammon, H. L., Murphy, K. C., & Swain, A. L. (1994) *Acta Crystallogr.* (in press).

Miller, M., Rao, J. K. M., Wlodawer, A., & Gribskov, M. R. (1993) *FEBS Lett. 328,* 275–279.

Murphy, K. C. (1983) Ph.D. Dissertation, University of Maryland, College Park, MD.

Quiocho, F. A., & Richards, F. M. (1964) *Proc. Natl. Acad. Sci. U.S.A. 52,* 833–839.

Roberts, J. (1976) *J. Biol. Chem. 251,* 2119–2123.

Roberts, J., & McGregor, W. G. (1991) *J. Gen. Virol. 72,* 299–305.

Roberts, J., Holcenberg, J. S., & Dolowy, W. C. (1972) *J. Biol. Chem. 247,* 84–90.

Roberts, J., Schmid, F. A., Old, L. J., & Stockert, E. (1976) *Cancer Biochem. Biophys. 1,* 175–178.

Shifrin, S., Luborsky, S. W., & Grochowski, J. (1971) *J. Biol. Chem. 246,* 7708–7714.

Smith, G. M. (1990) *QCPE509 22,* 38.

Stout, G. H., & Jensen, L. H. (1989) in *X-ray Structure Determination—A Practical Guide,* pp 382–384, John Wiley & Sons, Inc., New York.

Swain, A. L., Jaskólski, M., Housset, D., Rao, J. K. M., & Wlodawer, A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90,* 1474–1478.

Wlodawer, A., Roberts, J., & Holcenberg, J. S. (1977) *J. Mol. Biol. 112,* 515–519.

Wriston, J. C., Jr., & Yellin, T. O. (1973) *Adv. Enzymol. 39,* 185–248.