Crystal Structure of the Proenzyme Domain of Plasminogen[†,‡]

Ezra Peisach,[§] Jieyi Wang,[||,⊥] Teresa de los Santos,[⊥,#] E. Reich,[⊥] and Dagmar Ringe*

Program in Biophysics and Structural Biology, Brandeis University, Waltham, Massachusetts 02454-9110, Department of Pharmacological Sciences, SUNY at Stony Brook, Stony Brook, New York 11794-8651, and Departments of Biochemistry and Chemistry and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02454-9110

Received May 17, 1999; Revised Manuscript Received June 24, 1999

ABSTRACT: We have solved the X-ray crystal structure of the proenzyme form of the catalytic domain of plasminogen, with the nonessential mutations M585Q, V673M, and M788L, to 2.0 Å resolution. The structure presents an inactive protease characterized by Asp740 (chymotrypsinogen 194) hydrogen bonded to His586 (chymotrypsinogen 40), preventing proper formation of the oxyanion hole and S1 specificity pocket. In addition, the catalytic triad residues are misplaced relative to the active conformation adopted by serine proteases in the chymotrypsin family. Finally, a unique form of zymogen inactivation is observed, characterized by a "foot-in-mouth" mechanism in which Trp761 (chymotrypsinogen 215) is folded into the S1 specificity pocket preventing substrate binding.

Plasminogen (Plg),[1] the inactive precursor of the general, trypsin-like, serine protease plasmin (Plm), is widely distributed in the body, is present at high concentrations in plasma, and represents a large reservoir of potential proteolytic activity that can be recruited for pericellular proteolysis by local secretion of plasminogen activators (PAs). Plasminogen activation is associated with a broad spectrum of physiological and pathological processes, including invasion and dissemination of tumor cells and bacteria. Plasminogen activation can also be initiated pharmacologically by intravascular administration of PAs for therapeutic lysis of blood clots. Studies of knockout mice have shown that complete plasminogen deficiency causes premature death from wasting (*1*, *2*), which probably explains why mutations that inactivate plasminogen/plasmin have never been encountered in homozygous form in humans. Heterozygous mutations have been observed and are associated with a propensity for thrombosis (*3*).

Human plasminogen is a single-chain protein of 791 residues (Figure 1), containing an N-terminal domain (Glu1−Lys77), five kringle domains of approximately 80 residues each, and a C-terminal serine protease catalytic domain as an inactive proenzyme (544−791) (*4*). The proenzyme domain belongs to the chymotrypsin family of serine proteases and is similar to the serine protease zymogens bovine chymotrypsinogen (40% identity/47% similarity) and bovine trypsinogen (43% identity/51% similarity) (Figure 1). The specificity of plasmin for basic residues (Lys/Arg) is similar to that of trypsin and thrombin (*5*).

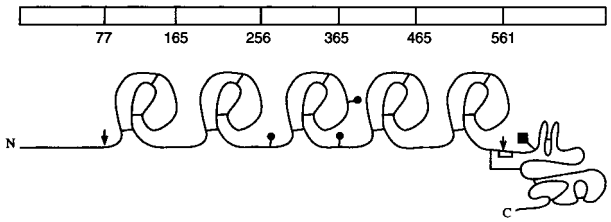

FIGURE 1: Schematic of plasminogen based on figure in Halkier (*44*). Arrows indicate cleavage sites in plasminogen. Balls indicate glycosylation sites and rectangle the phosphorylation site.

The full-length expressed plasminogen takes its name from the N-terminal residue, glu-plasminogen. The N-terminal domain is sometimes compared to a growth factor domain based on intron splicing and disulfide bonds, but its exact function is unknown (*6*). Full-length plasminogen is known to exist in at least three different conformations in solution, as determined by small-angle scattering, laser light scattering experiments and NMR, dependent on kringle domain binding of lysine analogues, such as 6-aminohexanoic acid (6-AHA) (*4*, *7*). At least three glycosylation sites are present upstream of the catalytic domain (*8*) and one phosphorylation site within it (*9*) (Figure 1). Three truncated forms of plasminogen have been described. Removal of the N-terminal domain by plasmin-catalyzed cleavage at residue 77 yields a shorter proenzyme (lys-plasminogen) that is more efficiently activated than the parent glu-plasminogen, and has altered affinities for small ligands (*10*). Cleavage by elastase removes the N-terminal domain and four kringles leaving miniplasminogen, which consists of a single kringle and the proenzyme domain. Removal of the last kringle leaves the

[†] This work was supported by grants from NIH (5 RO1 HL50320-05 to ER) and a Biophysics training grant to Brandeis (E.P.)

[‡] Coordinates have been submitted to the Protein Data Bank (http://www.rcsb.org/pdb/) with accession code 1QRZ.

* To whom correspondence should be addressed, Departments of Biochemistry and Chemistry and Rosenstiel Basic Medical Sciences Research Center, Brandeis University. Phone: (781) 736-4902. Fax: (781) 736-2405. E-mail: ringe@brandeis.edu.

[§] Program in Biophysics and Structural Biology, Brandeis University

[||] Present address: D48r, AP9-2124, Abbott Laboratories, 1 Abbott Park Road, Abbott Park, Illinois 60064.

[⊥] SUNY at Stony Brook.

[#] Present address: Department of Biochemistry, State University of New York at Stony Brook, 308 Life Sciences Building, Stony Brook, NY 11794-5251.

[1] Plg, plasminogen; Plm, plasmin; PAs, plasminogen activators; mPlg, microplasminogen; NPGB, nitrophenol-*p*-guanidinobenzoate; uPA, urokinase plasminogen activator; tPA, tissue plasminogen activator; PDB, protein data bank.

10.1021/bi991130r CCC: $18.00
Published on Web 08/04/1999

isolated proenzyme domain, microplasminogen (mPlg) (*10*). In their catalytic and substrate properties, all of the truncated forms are qualitatively identical to native glu-plasminogen.

A structure unique to plasminogen is the activation site within a nine residue sequence bounded by two cysteine residues (Cys558−Cys566) forming a disulfide-bonded loop. Maintaining this structure appears to be relevant to activation by the physiological activators tissue plasminogen activator (tPA) and urokinase (uPA) (*11*).

So far, the structure of the entire Plg molecule has not been determined. However, by a divide and conquer approach, the structures of a number of domains of plasminogen have been established. For instance, the structures of kringle 1 (*12*, *13*), kringle 4 (*14*, *15*), and kringle 5 (*16*) with various ligands bound have been determined by NMR and X-ray crystallography. The zymogen form of the proenzyme domain is being reported here. Recent reports describe the structures of streptokinase (*17*) and staphylokinase (*18*) complexed with the catalytic domain of plasmin.

Plasminogen activators are of two kinds, direct and indirect. Direct activators, produced by vertebrate cells, are highly specific limited serine proteases that cleave a single Arg561−Val562 bond at the activation site to yield plasmin. Indirect activators, streptokinase and staphylokinase, are produced by bacteria and have no proteolytic function. These indirect activators form complexes with plasminogen and plasmin, and the complexes act as PAs. In so doing, an active site, resembling that of plasmin, is formed without covalent modification within the zymogen domain in the complex. It is therefore of interest to define the structure of the active site in the zymogen both before and after complex formation with a bacterial activator. We have crystallized human microplasminogen and report here its structure at 2.0 Å resolution.

MATERIALS AND METHODS

1.1. Microplasminogen. Microplasminogen was expressed in insect cell culture (High Five cells) during infection with recombinant baculovirus. Construct 53, consisting of the N-terminal sequence of Plg EPLDDYVNTQG to which was attached the proenzyme domain beginning with residue P544[2], contained three mutated residues, M585Q, V673M, and M788L; these mutations produced no detectable changes either in activation of mPlg or in catalytic properties of the resulting mPlm (*19*). This mutant is the only form which has crystallized to date.

The culture supernatants were centrifuged at 18000*g* at 4° for 30 min, and solid ammonium sulfate was added to a final concentration of 1 M. The protein was adsorbed onto a column of phenyl-Toyopearl equilibrated with 1 M ammonium sulfate, 10^{-5} M NPGB, 0.02 M phosphate buffer, pH 6.2. The column was washed with 2 column volumes of the same buffer, and the adsorbed protein eluted in 0.5 M urea and NPGB 0.02 M phosphate, pH 6.2. The eluate was concentrated by dialysis against aquacide, dialyzed against 20 mM Tris, pH 7.8−8.0, NPBG, and adsorbed onto a column of SP-agarose equilibrated with the same buffer. After washing the column, the protein was eluted with a linear gradient of NaCl in the same buffer, with 0 and 0.3 M NaCl as limiting concentrations in a total of 10 column volumes. The peak fractions were dialyzed against 20 mM cacodylate buffer, pH 6.5, NPGB, adsorbed onto a column of heparin-Toyopearl equilibrated with the same buffer, and eluted with a linear gradient of NaCl (0 to 0.5 M) also in the same buffer (total of 7 column volumes). The peak fractions were pooled, benzamidine was added to 50 mM to inhibit any plasmin formed accidentally, and the protein precipitated by addition of solid ammonium sulfate to 75% saturation. The precipitate was concentrated by centrifugation, dissolved in the minimal volume of cacodylate buffer (100 mM, pH 6.2, NPGB) dialyzed against the same buffer containing 50 mM benzamidine, and lyophilized.

Table 1: Statistics for Plasminogen Data Collection and Refinement

dataset	$P2_1$	$P2_12_12_1$
Data Collection		
temp	cryo	4 °C
unit cell		
a (Å)	81.66	48.88
b (Å)	73.56	68.08
c (Å)	81.36	75.67
β (deg)	108.73	
resolution (Å)	2.0	2.8
total no. of reflections	199 470	26 903
no. of unique reflections $[I/\sigma(I) > 0]$	63 217	5822
completeness of data Å (%)	99.3 [79.4 $I > \sigma(I)$]	87.6
resolution range (Å)	35−2.0	10−2.8
R_{merge}[a] (%)	16.5 [9.2 $I > \sigma(I)$]	20.6
Refinement		
resolution range (Å)	35−2.0	8−3
reflections used(working/free)	55 335/2869	4170/439
R_{factor}[b,c]/R_{free} (%)	23.5/29.4	
R_{factor}[b,d]/R_{free} (%)	22.5/29.8	
no. of protein atoms	7568	
no. of waters	518	
temperature factor model	restrained individual	fixed overall
B values (Å^2)		
main and side chain	17.5	
side chains	21.8	
waters	20.6	
deviations observed		
rms dev (bond lengths, Å)	0.014	
rms dev (bond angles, deg)	1.9	
rms dev (planar groups, Å)	1.2	

[a] $R_{merge} = \sum|I_{obs} - I_{avg}|/\sum I_{avg}$, the sums being taken over all symmetry related reflections. [b] $R_{factor} = \sum|F_o - F_c|/\sum F_o$, the sums being taken over all reflections. [c] Maximum likelihood target function. [d] Standard crystallographic refinement.

1.2. Crystallization of Microplasminogen. Lyophilized protein was reconstituted in 50 mM benzamidine, 100 mM sodium cacodylate buffer, pH 6.2, to a concentration of 15 mg/mL. Crystals were grown by the hanging drop method using a well solution of 3% 2-propanol/3% poly(ethylene glycol) (PEG) 4000 and 20 mM citrate buffer, pH 5.6. An equal volume of protein solution and well solution were mixed and allowed to equilibrate over the well solution. Diffraction quality crystals grew in under a week and were approximately rectangular with dimensions 0.5 mm × 0.25 mm × 0.1 mm.

1.3. Data Collection. A number of datasets were collected (Table 1). One dataset was collected at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratories, beam-line X12B using an ADSC Quantum 4 detector. One crystal with dimensions approximately 0.3 mm

[2] The residue numbering refers to the position in full-length glu-plasminogen.

× 0.2 mm × 0.1 mm was frozen in liquid nitrogen after transferring into cryoprotecting oil (available from Håkon Hope, Department of Chemistry, University of California, Davis) and removing surface mother liquor. In total, 169 frames of 1 min exposure and 1.0° oscillation were collected under cryogenic temperatures, at a crystal to detector distance of 150 mm, at a wavelength of 0.9924 Å. Data were processed with the program DENZO (*20*) and scaled with the program SCALEPACK (*20*). The crystal belongs to space group $P2_1$ with unit cell dimensions $a = 81.66$ Å, $b = 73.56$ Å, $c = 81.36$ Å, $\beta = 108.73°$ and contains four molecules per asymmetric unit.

Attempts were made at finding heavy-atom derivatives in order to solve the structure by multiple isomorphic replacement (MIR), but a suitable derivative was never found, due either to nonisomorphism or lack of heavy-atom binding. In the process, it was found that soaking a plasminogen crystal in 1 mM ethyl-mercury-phosphate (EMP), 3% 2-propanol/3% PEG 4K, 10 mM Hepes, pH 7.5, for 1 h reproducibly caused a spacegroup transition to $P2_12_12_1$ with unit cell dimension $a = 48.88$ Å, $b = 68.08$ Å, $c = 75.67$ Å. In this crystal form, there is only one molecule per asymmetric unit. Crystallographic data were collected to 2.8 Å resolution on a Siemens multiwire area detector mounted on an Eliot GX6 generator running at 29 kV, 20 mA, and crystal to detector distance of 110 mm. Two sweeps totaling 1090 0.2° 10 min oscillations were collected at 4 °C. The program XDS (*21*) was used to process and analyze the data, and the program XSCALE (*21*) was used to merge the data from the two sweeps.

1.4. Molecular Replacement. Because of the difficulty of finding suitable heavy-atom derivatives, we attempted to solve the structure by molecular replacement using AMORE (*22*) in the CCP4 package (*23*). We chose to use as a starting model a monomer of the 1.8 Å resolution crystal structure of bovine γ-chymotrypsinogen [pdb accession code 2cga (*24*−*26*)]. This model was chosen because of a conserved disulfide bridge, Cys 548−Cys 666 (Cys1−Cys122 in chymotrypsinogen), which is conserved in plasminogen and chymotrypsinogen but not in other zymogens. The 245 amino acid model, 40% identical/47% similar to Plg Pro544−Asn791, was used without any trimming or compensating for differences in sequence. Temperature factors were left as in the original structure.

Using the $P2_12_12_1$ dataset (1 molecule per asymmetric unit), the search model was first translated to the origin and rotated to align the molecule along an axis. The rotation function was solved with data from 20−4 Å resolution using a sphere radius of 23 Å and box size of 80 Å × 75 Å × 70 Å. A clear solution resulted with a correlation of 19.6% (12.7% for next highest peak) with $\alpha = 84.17°$, $\beta = 86.25°$, $\gamma = 275.54°$.

The translation function gave a clear solution at (0.1159, 0.3742, 0.2285), with a correlation of 30.3% and R_{factor} of 46.5%. Rigid body fitting in AMORE resulted in an improved correlation of 46.2% and R_{factor} of 46.5%. Initial examination of a generated unit cell with this solution did not show any packing problems.

A number of other models were also used in an attempt to solve the structure by molecular replacement. It is interesting to note that both the zymogen and active forms of other serine proteases would also have been successful. For instance, molecular replacement using as a model γ-chymotrypsin [1.8 Å; pdb accession code 1afq (*25*−*27*); 39.6% identity] or ϵ-trypsin [1.8 Å; pdb accession code 1ept (*28*); 46.7% identity] were equally successful.

1.5. Refinement and Rebuilding in $P2_12_12_1$ Space Group. Phases calculated using XPLOR (*29*) from the chymotrypsinogen molecular replacement solution gave an R_{factor} of 46.1% and R_{free} of 43.8% (10−5 Å resolution data). Rigid body refinement of the entire model reduced the R_{factor} to 45.7% and R_{free} to 41.7%. This was followed by positional refinement using data from 8 to 3 Å resolution resulting in an R_{factor} of 34.8% and R_{free} of 46.7%. Initial electron density maps with coefficients $2mF_o - DF_c$ and sigma weighting (*30*) were calculated. Disulfide bridges were clearly visible as well as density for large residues, which existed as smaller ones in the search model. Numerous rounds of model building in O (*31*) and refinement in XPLOR were carried out, removing entire residues that did not fit into density, trimming residues to alanine if the side chain did not fit, and building in the proper sequence whenever possible. In general, the most progress was made by building out from "anchor" points, areas of high identity which are conserved among the chymotrypsin family serine proteases.

With the relatively low sequence identity between plasminogen and chymotrypsinogen in loop regions (Figure 2), ensuring the proper registration of sequence was particularly difficult at 2.8 Å resolution. For instance, the sequence from 603 to 646 contains a deletion of six residues, stretches of residues with small side chains, and stretches of residues with long side chains. Consequently, it was hard to confirm the registration of the sequence as it was being built. Another such stretch of residues with similar difficulties occurs from 687 to 695. Structurally conserved residues that could serve as anchor points from which to build the proper sequence outward were found by creating a three-dimensional sequence alignment combining both sequence and secondary structural elements. Structures in the Protein Data Bank (PDB) were found with a FASTA (*32*) search using the sequence of the catalytic domain of plasminogen. A sequence alignment alone does not correctly align residues that appear in analogous positions structurally. For example, a sequence alignment of plasminogen and chymotrypsinogen pairs Glu627 with Glu78, respectively. Structurally, Glu627 aligns with Ser77 (Figure 2).

1.6. Molecular Replacement in $P2_1$ Space Group. The $P2_1$ dataset provides higher resolution but a more difficult molecular replacement problem with four molecules per asymmetric unit. Using the above $P2_12_12_1$ model with 229 of the 259 residues placed, molecular replacement in the $P2_1$ space group was performed. Of these 229 residues, 209 were of the correct sequence and 20 were represented by alanines or serines. These 20 residues either flanked eight chain breaks or were substituted for larger amino acids for which the electron density maps failed to show a clear position for the side chain.

A self-rotation map of the $P2_1$ dataset calculated using the program GLRF (*33*) with data from 10 to 3.5 Å clearly showed two sets of 2-fold axes at ($\phi = 0°$, $\psi = 36°$, $\kappa = 180°$) and ($\phi = 180°$, $\psi = 54°$, $\kappa = 180°$), which are equivalent, at 99% the height of the symmetry related peak. We interpreted the presence of a single 2-fold axis in terms of multiple pairs of molecules being related by a pure

FIGURE 2: Sequence alignment of plasminogen, trypsinogen, factor D, and chymotrypsinogen. Gap insertions and alignment based on the three-dimensional structures. Residue numbering on top is based on plasminogen, on bottom, chymotrypsinogen. Secondary structural elements as determined by DSSP (*45*) refer to the structure presented here. Boxed sequences represent conserved residues and gray shaded residues represent identical residues. Figure produced by ALSCRIPT (*46*).

translation. This was supported by a native Patterson map showing a peak at (1/2, 0, 1/2) (pseudo B-centering).

Using data from 30 to 4 Å resolution and a cell size of 90 Å^3, four solutions to the rotation function were found in two sets related by 180° rotations by AMORE. The solutions found (α°, β°, γ°, correlation) are (11.78, 127.04, 352.19, 27.4%), (168.22, 52.96, 172.19, 27.4%), (191.38, 56.83, 338.83, 27.0%), and (348.62, 123.17, 158.83, 27.0%), which agree with the self-rotation map. The next highest rotation function solution has a correlation of 9.3%. Each translation solution was refined in AMORE before incorporating it into the phased-translation function search for the next monomer. The correlation coefficients for the $P2_1$ model as each monomer was added were 16.5, 61.2, 62.1, and 76.8%. The final AMORE rigid body refined correlation is 79.7% with an R_{factor} of 35.9%. The translation function solutions agree with the pseudo-B centering observed for the data.

1.7. Refinement of P2$_1$ Dataset. The initial R_{factor} for the model with phases calculated from the four molecules found by molecular replacement was 36.6% and R_{free} 38.3% to 4 Å resolution. The four molecules were first refined in rigid groups with CNS (*34*) using a maximum likelihood target function to an R_{factor} of 40.7% and R_{free} of 42.0%. Positional refinement using strict noncrystallographic symmetry and data to 3.0 Å resolution was carried out to a resulting R_{factor} of 39.5% and R_{free} of 42.9%. Electron density maps were calculated both in CNS and in CCP4 using phases modified with DM (*35*) (solvent flattening, histogram matching, averaging). Again, an iterative method of rebuilding and refining was followed to connect most of the backbone gaps and to place side chains.

Rebuilding of a representative model using strict noncrystallographic symmetry (NCS) progressed until 93% of the model was built resulting in an R_{factor} of 38.1% and R_{free} of 40.3% to 2.6 Å resolution. Attempts to impose any form of noncrystallographic restraints at higher resolution did not appear to improve the electron density maps. NCS restraints were removed from the refinement and positional refinement to 2.4 Å resolution carried out. This resulted in an immediate lowering of both the R_{factor} and R_{free} to 31.4 and 39.5%, respectively.

Refinement of group temperature factors was incorporated at this point. As resolution was gradually increased, water molecules were added provided they looked spherical in electron density maps calculated with coefficients $2mF_o - DF_c$ and $F_o - F_c$ contoured at 1.5σ and 2.5σ, respectively, and were within reasonable hydrogen-bonding distance of a donor or acceptor atom. Restrained individual temperature factor refinement was not used until all the data to 2.0 Å resolution were included. Simulated annealing omit maps were calculated for the whole structure to confirm the correctness of the final model. The final R_{factor} was calculated in two ways. The first from refinement using the maximum likelihood target function as described by Adams et al. (*36*) resulting in an R_{factor}/R_{free} of 23.5%/29.4%. The standard residual R_{factor} was calculated by refining the final model against the standard crystallographic target function resulting in an R_{factor}/R_{free} of 22.5%/29.8%.

The final model contains four monomers of Plg each representing residues 546−791 and 518 waters (Table 1). The N-terminal 11 residues and residues Pro544/Ser545, representing approximately 5% of the scattering mass, are disordered and not included. The only residue which does not fall into the preferred main-chain angle combination ϕ/ψ is Lys750. This residue is at the apex of a tight three residue turn connecting two β-strands that is stabilized by internal hydrogen bonds.

RESULTS AND DISCUSSION

2.1. Structure. We have solved the structure of the proenzyme domain of plasminogen to 2.0 Å resolution. The overall architecture of this domain is very similar to other chymotrypsin-like serine proteases (Figure 3), made up of two β-sheet barrels connected by loops (*37*). Comparison

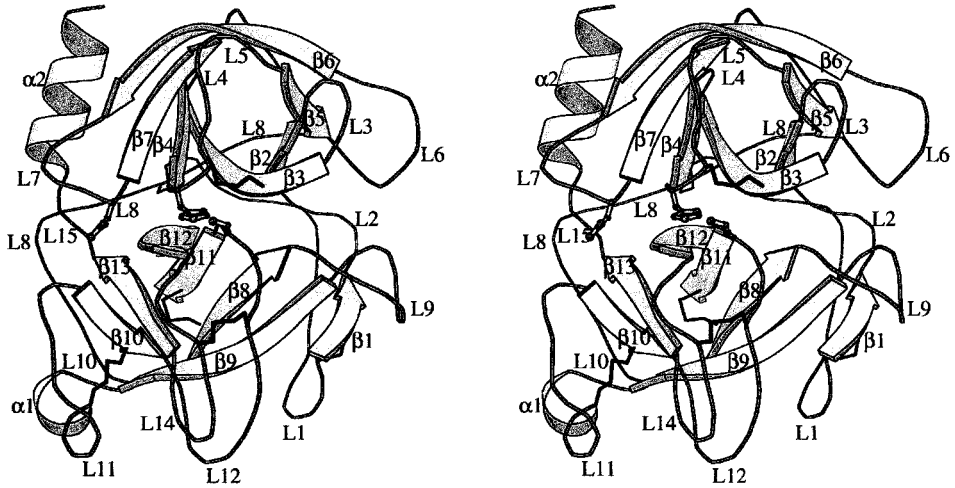

FIGURE 3: Stereo ribbon diagram of microplasminogen. The catalytic triad consisting of Asp646, His603, Ser741 is indicated. The sequence alignment of plasminogen with other proteases is presented in Figure 2 with the secondary structural elements labeled. Figure produced with MOLSCRIPT (*47*).

Table 2: RMS Deviation between of Plasminogen and Other Serine Proteases

protein	resolution (Å)	no. of Cα atoms	RMSD (Å)	ref[a]
chymotrypsinogen	1.8	202	1.19	2cga (*24*)
γ-chymotrypsin	1.6	200	1.18	1ab9 (*40*)
trypsin	1.8	171	1.06	1btx (*41*)
trypsinogen	1.65	172	1.10	1tgn (*42*)
factor D	2.0	173	1.06	1dsu (*43*)

[a] The references column provides the primary citation and accession code in the PDB (*25*).

of this structure with those of other serine proteases in this family shows that this fold is structurally conserved. For instance, using a 3.8 Å distance cutoff for α-carbons, the RMS deviation between Plg and a number of other zymogens and active forms of serine proteases is approximately 1.2 Å (Table 2).

In the following analysis, we compare Plg to chymotrypsinogen [2cga (*24*)] as they share a common disulfide linkage Cys548(1)[3]−Cys666(122) not found in trypsinogen. Most of the differences occur in the loops between secondary structural elements and take two forms: differences in length of the loop due to insertions or deletions, or completely different conformations. The significantly different loops in Plg compared to chymotrypsinogen occur at residues 546−565 (L1), 581−585 (L3), 602−614 (L5), 641−646 (L7), 653−678 (L8), 685−697 (L9), 712−723 (L11), 728−743 (L12), and 760−771 (L14) (Figure 3). Of these, loops L3, L5, L8, and L11 are areas in which chymotrypsin, trypsin, and factor *D* all differ due to insertions and deletions in the sequences.

Different structural features distinguish the remaining of the above loops from their homologues. The conformation of the N-terminal 20 residues of Plg (546−565) visible in this structure, up to the first structural element, is very similar to that of chymotrypsinogen. This is because of a conserved disulfide linkage at Cys548(1)−Cys666(122). In chymotrypsinogen, this region is characterized by a β-strand-turn-β-strand motif, but in plasminogen, the first strand is deformed by the unique Cys558−Cys566 bond. This sequence contains the conserved activation cleavage site in this family of serine proteases. Loop L7 (641−646) ending at the catalytic Asp646(102) is shortened compared to other proteases by a six residue deletion in the plasminogen sequence. The effect of this shortening may allow for a more open active site, permitting entry of larger molecules, such as cross-linked fibrin clots. Loops L9, L12, and L14 have the highest backbone temperature factors in this structure, and their conformations are significantly different from those of the corresponding loops in chymotrypsin, trypsin, factor *D*, and their zymogens. Loop L9 is in an area in which rearrangement takes place in the conversion of chymotrypsinogen to chymotrypsin (143−156) and trypsinogen to trypsin (143−156).

Loop L12 is part of the S1 specificity pocket of chymotrypsin and trypsin. The conformation of loop L12 (728−743) closely follows those in chymotrypsinogen and trypsinogen residues 181−194 along the right-hand side of the S1 specificity pocket (Figure 4). In all serine proteases in this family, this loop includes the conserved disulfide bridge Cys191−Cys220 and Asp194 of the putative oxyanion hole and is significantly rearranged during zymogen activation. In chymotrypsinogen, Asp194 is hydrogen bonded to His40. In a similar paradigm, plasminogen Asp740 Oδ1 is hydrogen bonded to His586 Nε2 at 3.1 Å and the backbone NH of 686 at 2.7 Å. The other oxygen of Asp740, Oδ2, forms hydrogen bonds with Trp685 NH (2.8 Å) and a water molecule (2.8 Å).

Loop L14 (760−771), the equivalent of residues 214−221 in chymotrypsinogen, forms the left side of the S1 specificity pocket. This loop in Plg has a significantly different conformation from the equivalent loop in any other serine protease zymogen for which coordinates are available. In addition, this loop has to rearrange in chymotrypsinogen when forming the active protease (Figure 4, bottom). In Plg, the conformation is perturbed by the close proximity of loop L12 (728−743) and held close to this loop by the disulfide bridge Cys737(191)−Cys765(220) (Figure 3). A key residue in this loop is Trp761(215), which always faces outward from the S1 specificity pocket in all other zymogens and proteases

[3] Numbers in parentheses represent important equivalent chymotrypsinogen residues.

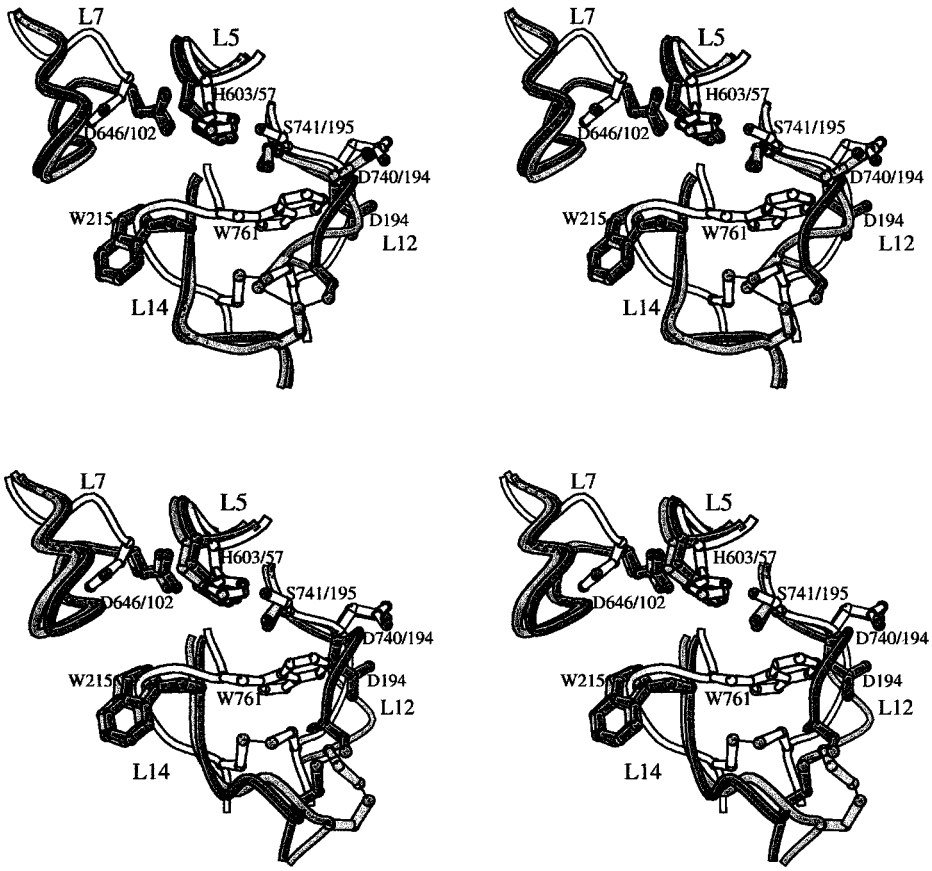

FIGURE 4: (Top) Comparison of the active site of plasminogen (white) to that of bovine trypsinogen [light gray, 1tgb (*48*)] and trypsin [dark gray, 1btx (*41*)]. (Bottom) Comparison of the active site of plasminogen (white) to that of bovine chymotrypsinogen [light gray, 2cga (*24*)] and chymotrypsin [dark gray, 1ab9 (*40*)]. Portions of loops L5 (601−604) and L7 (643−646), from which the active site histidine and aspartate derive, loops L12 (736−742) and L14 (760−766), from which the catalytic serine and the S1 pocket derive, have been labeled. Because the active-site residues and loops surrounding the S1 pocket have such different configurations from the expected active form of plasmin, major structural reorganization in the region must take place. In chymotrypsinogen and trypsinogen the catalytic triad residues Asp102, His57, Ser195 are already in the active configuration. These residues in plasminogen, Asp646, His603, Ser741, must rearrange to form an active enzyme. In trypsinogen, the lower region of the S1 pocket is in the correct configuration while plasminogen and chymotrypsinogen must undergo further rearrangements here. In all three zymogens, the oxyanion hole has not yet formed. Finally, plasminogen has a unique form of inactivation: Trp761 blocks the S1 pocket preventing substrate binding. Figure produced with MOLSCRIPT (*47*).

of this family. In Plg, however, Trp761(215) is facing the "wrong" direction, into the putative S1 pocket, blocking entrance into this pocket. This conformation makes it impossible for this zymogen to be catalytically active (Figure 4).

Loops L6 (620−630), L9, and L11 have been reported to be involved in interactions with streptokinase in the plasmin-streptokinase complex (*17*). In addition, loops L7, L11, and L14 have been reported to interact with staphylokinase in the plasmin-staphylokinase-plasmin complex (*18*). As stated previously, some of these loops have conformations that are the same as the equivalent loop in chymotrypsinogen and trypsinogen, however, some are clearly different in these regions.

Relative to chymotrypsin and trypsin, the catalytic triad of Plg, made up of Asp646(102), His603(57), and Ser741-(195), is incorrectly positioned for catalysis (Figures 4 and 5). The side chain of His603(57) is clearly facing in the wrong direction. Hydrogen bonds are observed from Nδ1 to a water molecule (2.5 Å) which in turn hydrogen bonds to the Oγ of Ser741(195) (2.5 Å). On the other side, His603-(57) Nϵ2 hydrogen bonds to the side chain of Glu 606 (3.1 Å) and another water molecule (2.8 Å). Asp646(102) is facing away from the active site with its main-chain carbonyl sterically blocking its own side-chain from properly rotating and interacting with His603(57). Ser760(214) is facing the correct direction to hydrogen bond with Asp646(102) if the latter were positioned correctly.

Unique to plasminogen is the presence of a seven residue loop, closed by a disulfide bridge (Cys558−Cys566), that contains the activation cleavage site Arg561−Val562. The function of this disulfide bridge is unknown. Removal of the disulfide bridge (*11*) and changes in length and content of this turn (*19*) have profound effects on the rate of activation and activity of plasmin. For instance, shortening the loop by one residue and cleaving the appropriate peptide bond prevents the resulting plasmin mutant from showing any activity (*19*). Mutations of Val562 and Val563 can result in marked decrease of catalytic activity of microplasmin. In the native activated plasmin, Val562 is expected to fold into a hydrophobic pocket and form an ionic interaction with the carboxylate of Asp740. Substitution of Val562 or Val563 with larger residues precludes access to this position. Mutation of either of these Val's to Gly may have the effect of not steering the newly formed terminal end of plasmin into the pocket.

Structurally, this Cys588−Cys566 disulfide bridge creates an inwardly kinked turn with Val562 facing toward the

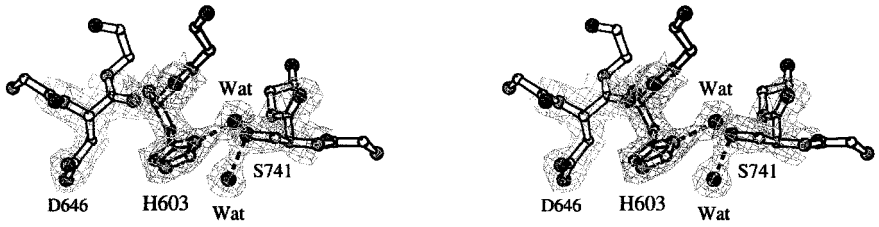

FIGURE 5: Sigma weighted electron density map with coefficients $2F_o - F_c$ of the catalytic triad of plasminogen contoured at the 1σ level. Electron density is shown for Asp646, His603, Ser741 and two water molecules. Backbone atoms for residues 645, 647, 604, 740, and 742 are shown for context. The orientation of the figure is approximatly that of Figure 3. Figure produced with MOLSCRIPT (*47*).

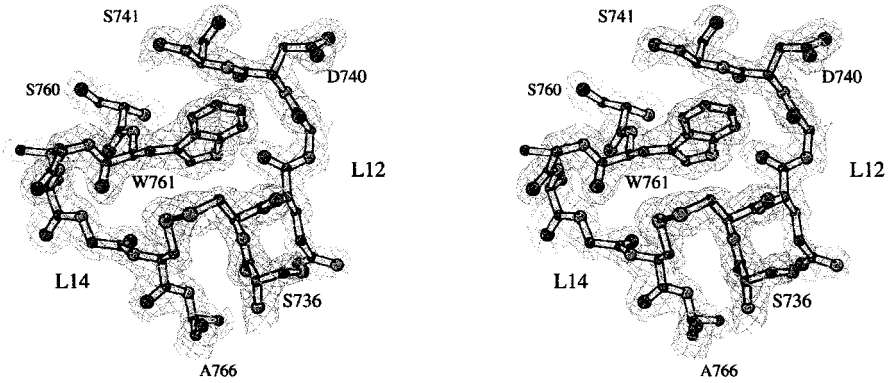

FIGURE 6: Sigma weighted electron density map with coefficients $2F_o - F_c$ of the S1 subsite of plasminogen contoured at the 1σ level. Portions of loops L12 (736−741) and L14 (760−766) are shown. Figure produced with MOLSCRIPT (*47*).

disulfide bridge. It is interesting to note that the conformation of the loop from Pro559−Arg561 mirrors the bent conformation of the Glu-Gly-Arg chloromethyl ketone in the structure of the uPA catalytic domain (*38*). Possibly, this kinked conformation of the Cys558−Cys566 loop might effect the recognition of the Arg561−Val562 site by its physiological activators.

2.2. Inactivity of the Zymogen. Chymotrypsin family serine proteases have evolved at least three prominent conformational mechanisms to ensure inactivity of the zymogen form. The most significant conformational change is in the loop from Ser189 to Asp194 (chymotrypsinogen numbering) (L12 in Plg). In both chymotrypsin and trypsin, this segment forms the oxyanion hole and one side of the S1 recognition pocket (Figure 4). In the zymogen form, Asp194 is hydrogen bonded to His40 (not shown). Upon activation, Asp194 reorients such that the carboxylate group forms a salt bridge with the newly formed N-terminal Ile16. This reorientation of Asp194 induces loop rearrangement in this region. This inactivation mechanism is shared by trypsinogen and chymotrypsinogen and presumably shared by plasminogen because Asp740 (194) is hydrogen bonded to His586(40) and the conformation of the loop is thus similar to that in chymotrypsinogen and trypsinogen.

A second zymogen inactivation mechanism which is observed in chymotrypsinogen but not trypsinogen involves the loop from Ser214 to Cys220 (part of loop L14, Ser760 to Cys765 in Plg). These residues make up the opposite side of the S1 subsite (Figure 4, bottom) and, in chymotrypsinogen and factor D, result in a narrowed active-site pocket. During activation, this loop rearranges, possibly pulled by the movement of the Cys191−Cys220 disulfide bridge. In trypsinogen, this loop is already preformed in the correct orientation (Figure 4, top). As reported in Parry et al. (*18*), this loop has reoriented in plasmin to the same configuration as observed in trypsin.

In chymotrypsinogen and trypsinogen, the catalytic residues are already aligned in a catalytically active conformation (Figure 4). However, in complement factor *D* (*43*), these residues are misorientated and the same is true in mPlg. In factor *D*, the histidine and aspartate residues are stabilized in an inactive conformation. It is believed that upon substrate binding, an induced conformational change takes place rearranging these residues. In plasminogen His603(57), Asp646(102), and Ser741(195) are in an incorrect conformation and must reorient, whether as part of the activation process or through an induced fit conformational change upon binding of substrate. As reported in Parry et al. (*18*), this rearrangement of the catalytic triad has occurred in the structure of plasmin inactivated by a chloromethyl ketone. These catalytic residues now have the same conformations as found in the active forms of serine proteases. An unusal form of inactivation is also observed in plasminogen in that the side chain of Asp646 of the catalytic triad is sterically blocked by its own carbonyl oxygen from rotating into the proper orientation for interaction with His603. Another example of steric blockage is present in the relatively common genetic mutation Ala601-Thr, known as plasminogen Tochigi, which occurs in heterozygous form in 2−3% of the Japanese population (*39*). This mutation causes steric hindrance of the catalytic His603, preventing the proper functioning of plasmin.

Finally, plasminogen presents a novel and unique form of inactivation characterized by a "foot-in-mouth" mechanism, blocking the entrance to the S1 subsite. Tryptophan 761 (215) folds into the S1 subsite, preventing substrate binding (Figures 4 and 6). In addition, the backbone from Trp761 to Cys765 bulges outward from the subsite, pulling the Cys765−Cys737 disulfide bond upward toward the mouth of the pocket. For activation to occur, it will be necessary for Trp761 to leave the pocket and for the disulfide

bridge to be pulled downward as the walls of the S1 subsite form.

2.3. Zymogen Activation. While we do not have the coordinates of the active form of the catalytic domain of plasmin, we can speculate as to how Plg is activated to Plm. The proteases chymotrypsin, trypsin, elastase, and thrombin are structurally very similar in the active-site region. Therefore, we expect that activation of Plg should lead to a similar conformation. Activation of these zymogens occurs through multiple rearrangements to go from inactive to active forms. For example, in chymotrypsinogen, activation is initiated by the cleavage of the peptide bond Arg15−Ile16. The resulting new N-terminal Ile16 folds into a pocket and forms an ion bridge with the carboxylate of the Asp194 side chain. The effect of this interaction is rearrangement of the loop containing Asp194 (Gly193−Ser195) forming the oxyanion hole and one side of the S1 pocket. The other side of the S1 subsite also rearranges, probably pulled by the Cys191−Cys220 disulfide bond. In trypsin, this site is already formed and the latter rearrangement is unnecessary.

If Plg follows the same activation pathway as chymotrypsinogen and trypsinogen, then the following events need to take place. Activation of plasminogen is initiated by the cleavage of the Arg561−Val562 bond by tPA or uPA. The new free amino group of Val562(16) needs to interact with Asp740(194). This can occur by rearrangement of loop L12 (728−743), reorienting Asp740(194) in the correct direction, and forming the oxyanion hole in the process. Possibly, loop L9 (685−697) could also rearrange such that the nitrogen of Val562 will also interact with the carbonyls of Gly686 and Glu687 stabilizing its position. This would then parallel what is observed in trypsin.

In the rearrangement of loop L12 (728−743), one side of the S1 subsite is formed. Loop L14 (760−771) would then be pulled by L12 via the Cys737−Cys765 disulfide bridge forming the other side of the S1 pocket. Trp761(215) must also leave the S1 subsite for catalytic activity to develop; this might be accomplished by the movement of L14, pulling Trp761 out of the S1 pocket and leaving the pocket open for substrate binding.

Somewhere along this pathway, loop L7 should shift slightly allowing Asp646(102) free rotation to interact properly with His603(57) and Ser760(214). Presumably, when the protein has been converted to plasmin by this process, the catalytic residues either will be in the proper orientation for catalysis or will be able to reorient upon substrate binding. In the staphylokinase-plasmin structure described by Parry et al. (*18*), the catalytic residues are disordered in one of the plasmin molecules. It is possible that these residues position themselves correctly when substrate is bound. It is also possible that their structure represents an intermediate which is sufficient for the activation of another molecule of plasminogen.

2.4. Summary. We have determined the structure of the proenzyme form of the catalytic domain of microplasminogen to 2.0 Å resolution. If, as we expect, the active conformation of plasminogen (plasmin) is similar to those of chymotrypsin and trypsin, then the zymogen must undergo rearrangements in both traditional and new ways. "Traditional" activation occurs through rearrangement of the S1 subsite induced by the interaction between Asp194 and residue 16. In plasminogen, the active site S1 subsite is blocked by Trp761, which must also move out the way for substrate recognition. The catalytic triad residues Asp646-(102), His603(57), and Ser741(195) are all displaced relative to the active form and must presumably rearrange to form an active protease. Finally, the activation loop is clearly observed in this structure and seems to have a conformation that mimics the conformation required for recognition by the activating protease uPA.

REFERENCES

1. Bugge, T. H., Flick, M. J., Daugherty, C. C., and Degen, J. L. (1995) *Genes Dev. 9*, 794−807.
2. Ploplis, V. A., Carmeliet, P., Vazirzadeh, S., Van Vlaenderen, I., Moons, L., Plow, E. F., and Collen, D. (1995) *Circulation 92*, 2585−2593.
3. Tsutsumi, S., Saito, T., Sakata, T., Mlyata, T., and Ichinose, A. (1996) *Thromb. Haemost. 76*, 135−138.
4. Ponting, C. P., Marshall, J. M., and Cederholm-Williams, S. A. (1992) *Blood Coagul. Fibrinol. 3*, 605−614.
5. Clavin, S. A., Bobbitt, J. L., Shuman, R. T., and Smithwick, E. L., Jr. (1977) *Anal. Biochem. 80*, 355−365.
6. Gherardi, E., Manzano, R. G., Cottage, A., and Hawker, K. (1997) in *Plasminogen-Related Growth Factors*, pp 24−35, John Wiley Sons Ltd., West Sussex, England.
7. Marshall, J. M., Brown, A. J., and Ponting, C. P. (1994) *Biochemistry 33*, 3599−3606.
8. Pirie-Shepherd, S. R., Stevens, R. D., Andon, N. L., Enghild, J. J., and Pizzo, S. V. (1997) *J. Biol. Chem. 272*, 7408−7411.
9. Wang, H., Prorok, M., Bretthauer, R. K., and Castellino, F. J. (1997) *Biochemistry 36*, 8100−8106.
10. Shi, G. Y., and Wu, H. L. (1988) *J. Biol. Chem. 263*, 17071−17075.
11. Linde, V., Nielsen, L. S., Foster, D. C., and Petersen, L. C. (1998) *Eur. J. Biochem. 251*, 472−479.
12. Mathews, I. I., Vanderhoff-Hanaver, P., Castellino, F. J., and Tulinsky, A. (1996) *Biochemistry 35*, 2567−2576.
13. Rejante, M. R., and Llinas, M. (1994) *Eur. J. Biochem. 221*, 939−949.
14. Atkinson, R. A., and Williams, R. J. (1990) *J. Mol. Biol. 212*, 541−552.
15. Mulichak, A. M., Tulinsky, A., and Ravichandran, K. G. (1991) *Biochemistry 30*, 10576−10588.
16. Chang, Y., Mochalkin, I., McCance, S. G., Cheng, B., Tulinsky, A., and Castellino, F. J. (1998) *Biochemistry 37*, 3258−3271.
17. Wang, X., Lin, X., Loy, J. A., Tang, J., and Zhang, X. C. (1998) *Science 281*, 1662−1665.
18. Parry, M. A., Fernandez-Catalan, C., Bergner, A., Huber, R., Hopfner, K. P., Schlott, B., Guhrs, K. H., and Bode, W. (1998) *Nat. Struct. Biol. 5*, 917−923.
19. Wang, J., Brdar, B., and Reich, E. (1995) *Protein Sci. 4*, 1758−1767.
20. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.) Vol. 276, pp 307−326, Academic Press, San Diego.
21. Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 67−71.
22. Navaza, J. (1994) *Acta Crystallogr., Sect. A 50*, 157−163.
23. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
24. Wang, D., Bode, W., and Huber, R. (1985) *J. Mol. Biol.* 185, 595−624.
25. Abola, E. E., Sussman, J. L., Prilusky, J., and Manning, N. O. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.), Vol. 277, pp 556−571, Academic Press, San Diego.
26. Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., an O. Ritter, J. P., and Abola, E. E. (1998) *Acta Crystallogr., Sect. D 54*, 1071−1077.
27. Kashima, A., Inoue, Y., Sugio, S., Maeda, I., Nose, T., and Shimohigashi, Y. (1998) *Eur. J. Biochem. 255*, 12−23.
28. Huang, Q., Wang, Z., Li, Y., Liu, S., and Tang, Y. (1994) *Biochim. Biophys. Acta 1209*, 77−82.

29. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science 235*, 458−460.
30. Read, R. J. (1986) *Acta Crystallogr., Sect. A 42*, 140−149.
31. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
32. Pearson, W. R. (1990) *Methods Enzymol. 183*, 63−98.
33. Tong, L., and Rossmann, M. G. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.) Vol. 276, pp 594−611, Academic Press.
34. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
35. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, number 31, pp 34−38.
36. Adams, P. D., Pannu, N. S., Read, R. J., and Brünger, A. T. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5018−5023.
37. Shotton, D. M., and Watson, H. C. (1970) *Nature 225*, 811−816.
38. Spraggon, G., Phillips, C., Nowak, U. K., Ponting, C. P., Saunders, D., Dobson, C. M., Stuart, D. I., and Jones, E. Y. (1995) *Structure 3*, 681−691.
39. Higuchi, Y., Furihata, K., Ueno, I., Ishikawa, S., Okumura, N., Tozuka, M., and Sakurai, N. (1998) *Br. J. Haematol. 103*, 867−870.
40. Yennawar, N. H., Yennawar, H. P., and Farber, G. K. (1994) *Biochemistry 33*, 7326−7336.
41. Katz, B. A., Finer-Moore, J., Mortezaei, R., Rich, D. H., and Stroud, R. M. (1995) *Biochemistry 34*, 8264−8280.
42. Kossiakoff, A. A., Chambers, J. L., Kay, L. M., and Stroud, R. M. (1977) *Biochemistry 16*, 654−664.
43. Narayana, S. V., Carson, M., el Kabbani, O., Kilpatrick, J. M., Moore, D., Chen, X., Bugg, C. E., Volanakis, J. E., and DeLucas, L. J. (1994) *J. Mol. Biol. 235*, 695−708.
44. Halkier, T. (1991) *Mechanisms in Blood Coagulation, Fibrinolysis and the Complement System*, Cambridge University Press, Cambridge.
45. Kabsch, W., and Sander, C. (1983) *Biopolymers 22*, 2577−2637.
46. Barton, G. J. (1993) *Protein Eng. 6*, 37−40.
47. Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
48. Fehlhammer, H., Bode, W., and Huber, R. (1977) *J. Mol. Biol. 111*, 415−438.

BI991130R