Crystal Structure of Thrombin−Ecotin Reveals Conformational Changes and Extended Interactions†

Stephanie X. Wang,‡ Charles T. Esmon,§ and Robert J. Fletterick*,⊥

Graduate Program in Chemistry and Chemical Biology, University of California, San Francisco, California 94143-0446, The Cardiovascular Biology Research Program, Oklahoma Medical Research Foundation, and the Departments of Pathology, Biochemistry and Molecular Biology, University of Oklahoma Health Science Center, and the Howard Hughes Medical Institute, Oklahoma City, Oklahoma 73104, and Departments of Biochemistry and Biophysics and Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0448

Received April 6, 2001; Revised Manuscript Received May 31, 2001

ABSTRACT: The protease inhibitor ecotin fails to inhibit thrombin despite its broad specificity against serine proteases. A point mutation (M84R) in ecotin results in a 1.5 nM affinity for thrombin, 10^4 times stronger than that of wild-type ecotin. The crystal structure of bovine thrombin is determined in complex with ecotin M84R mutant at 2.5 Å resolution. Surface loops surrounding the active site cleft of thrombin have undergone significant structural changes to permit inhibitor binding. Particularly, the insertion loops at residues 60 and 148 in thrombin, which likely mediate the interactions with macromolecules, are displaced when the complex forms. Thrombin and ecotin M84R interact in two distinct surfaces. The loop at residue 99 and the C-terminus of thrombin contact ecotin through mixed polar and nonpolar interactions. The active site of thrombin is filled with eight consecutive amino acids of ecotin and demonstrates thrombin's preference for specific features that are compatible with the thrombin cleavage site: negatively charged-Pro-Val-X-Pro-**Arg**-hydrophobic-positively charged (P1 Arg is in bold letters). The preference for a Val at P4 is clearly defined. The insertion at residue 60 may further affect substrate binding by moving its adjacent loops that are part of the substrate recognition sites.

The serine protease α-thrombin is a key enzyme in the processes of thrombosis and haemostasis (*1*). It exhibits both pro- and anticoagulant activities, through converting fibrinogen to clot-forming fibrin (*2*, *3*) or through the thrombomodulin-mediated activation of protein C (*4*, *5*). Physiologically, the activities of thrombin are regulated by endogenous inhibitors including antithrombin III (*6*), α2-macroglobulin (*7*), heparin cofactor II (*8*), protein C inhibitor (*9*), and protease nexins (*10*). Alternatively, thrombin levels can be regulated by natural and synthetic protease inhibitors such as hirudin, triabin, and rhodniin (*7*, *11*). Inhibitors, especially macromolecular inhibitors, have been essential in elucidating the mechanisms of actions of various proteases. The interactions between thrombin and different inhibitors have been studied extensively due to thrombin's biomedical significance.

Ecotin, a 142-amino acid protein discovered in the periplasmic space of *Escherichia coli* (*12*), has the unique property of inhibiting most serine proteases of the chymotrypsin fold, including trypsin, chymotrypsin, and elastase (*12*, *13*). The inhibitory profiles of ecotin against various serine proteases share three common features (*14*, *15*): (i) Ecotin dimerizes and inhibits cognate proteases at 1:1 stoichiometry. Crystal structures of ecotin−trypsin, ecotin−collagenase, and ecotin−chymotrypsin all show the formation of a tetramer containing two ecotin and two protease molecules. (ii) Ecotin inhibits proteases via binding at two different protease contact sites: the primary and secondary binding sites. Primary site loops of ecotin bind to the active site of target proteases in a substrate-like manner as observed in canonical protease inhibitors such as BPTI[1] (*16*). The P1[2] residue in ecotin (Met84) mimics the interactions of a canonical P1 substrate residue. Unique secondary site loops of ecotin bind to a relatively flat surface of the protease that is distant from the active site. (iii) The protein−protein interaction surfaces are extremely large in ecotin−protease complexes; the sum of the buried surface area is around 6000 Å^2 (*15*, *18*).

Despite its apparent pan-specificity against chymotrypsin family serine proteases, ecotin fails to inhibit thrombin (*13*). Thrombin contains nine extra amino acid residues inserted at residue 60 in comparison to the sequence of chymotrypsin (*2*). This insertion loop contains bulky residues such as Pro, Trp, and Tyr, and appears relatively rigid forming a lip above

† The structure has been deposited with a PDB ID of 1ID5.

* Corresponding author: Dr. Robert Fletterick, Department of Biochemistry and Biophysics, Box 0448, University of California at San Francisco, San Francisco, CA 94143-0448. Phone number: 415-476-5080. Fax number: 415-476-1902. E-mail: flett@msg.ucsf.edu.

‡ Graduate Program in Chemistry and Chemical Biology, University of California at San Francisco.

§ Oklahoma Medical Research Foundation and University of Oklahoma Health Science Center.

⊥ Departments of Biochemistry and Biophysics and Pharmaceutical Chemistry, University of California.

[1] Abbreviations: BPTI: bovine pancreatic trypsin inhibitor; PPACK: D-Phe−Pro−Arg-chloromethyl ketone; rmsd: root-mean-square deviation.

[2] Substrate residues are labeled NH3...-P3-P2-P1-P1′-P2′-P3′...-COO- (*17*). P1-P1′ is the cleaved peptide bond. Correspondingly, the pockets on enzyme are ...S3, S2, S1, S1′, S2′, S3′... so that P1 residue in substrate/inhibitor binds the S1 pocket of the protease, etc.

10.1021/bi010712h CCC: $20.00
Published on Web 07/31/2001

the active site (*19*). Modeling analysis shows that Trp^{60D} (thrombin residue numbers are labeled as superscripts to distinguish them from ecotin residues) clashes with residues Cys50 and Asn51 in the e50's binding loop of ecotin (residue numbers specifying the binding loops in ecotin follow the letter e to differ from thrombin). Previous studies showed that the insertion loop at thrombin residue 60 blocked inhibition by BPTI (*20*). In that study, a thrombin desPPW mutant, created by removing three bulky rigid residues Pro^{60B}-Pro^{60C}-Trp^{60D}, permitted BPTI inhibition with a nanomolar K_i (*20*). While a single mutation of Met84 to Arg only improved ecotin binding to trypsin by 3-fold (*21*), the same mutation increased ecotin's inhibition against thrombin by 2×10^4 times (*13*). The inhibition constant (K_i) changed from approximately 30 μM for wild-type ecotin to 1.5 nM for the M84R mutant (data not shown). Deleting the residues that cause steric hindrance in wild-type thrombin explains BPTI's inhibition against the desPPW thrombin mutant (*20*). Even though thrombin clearly prefers an Arg at the P1 site, this crucial residue is only effective when bound to the active site (*22*). Since ecotin access to the active site appears to be blocked by Trp^{60D}, it is difficult to explain why the Met84 to Arg mutation helps ecotin to inhibit thrombin.

Like other regulatory proteases, the substrate specificity of thrombin extends beyond the P1 residue: other amino acids flanking the cleaved scissile bond are also important for the substrate recognition of thrombin (*23*, *24*). In comparison to digestive proteases, regulatory proteases often contain insertions of amino acids in the surface-located loops surrounding the active site. The characteristics of these loops are believed to be important for substrate recognition by the regulatory proteases (*25*). Details of the extended specificity of thrombin have been studied in vitro using synthetic substrates (*26*−*29*). These data define the cleavage site to be L/I/V/F-X-P-**R**-nonpolar/hydrophobic-hydrophobic/charged, from P4 to P2′ (X stands for any amino acid, P1 residue Arg is bold). Fibrinogen peptides are cleaved at GGV**R**-GP and FSA**R**-GH. Interactions at P4−P2′ sites can also be visualized by combining information from X-ray structures of many thrombin complexes with inhibitors or peptides (*19*, *30*−*32*). Among existing exogenous inhibitors, ecotin binds proteases over the largest interface, which may extend beyond the known interaction site for thrombin.

To elucidate how ecotin M84R and thrombin form a complex, we have determined the structure of the ecotin M84R−thrombin complex to 2.5 Å resolution.

MATERIAL AND METHODS

Purified bovine thrombin and ecotin M84R were prepared as described previously (*33*, *34*). Thrombin−ecotin M84R was formed by incubating 1:1 molar ratio of each component and subsequently purified on a superdex-200 size exclusion column (Pharmacia, Germany) in 50 mM potassium phosphate, pH 8.0, and 0.15 M NaCl.

The thrombin−ecotin complex isolated by gel filtration was used in the crystallization experiments. Orthorhombic crystals were grown at room temperature by hanging drops using the vapor diffusion technique, from 15% polyethylene glycol 6K, 0.1 M citric acid, pH 5.0, with 0.01 M spermine-tetrahydrochloride. Three diffraction data sets were collected at Advanced Light Source (ALS) beam line 5-2 with a CCD camera and at Stanford Synchrotron Radiation Laboratory (SSRL) beam line 9-1 with a MAR image plate system. The data sets were evaluated and integrated using SCALEPACK/DENZO (*35*). The crystal has a space group of $C222_1$. One asymmetric unit contains one thrombin and one ecotin. A heterodimer model of bovine thrombin (1UVT.pdb) bound to ecotin at the protease active site was generated using InsightII (Accelrys, San Diego, CA), based on the trypsin−ecotin structure (*14*). The 60's and 148's insertion loops were deleted from thrombin. The resulting structure was used as the search model to solve the structure of thrombin−ecotin M84R, using molecular replacement with rotational and translational functions from CNS 1.0 (*36*). Data to 3.5 Å were used in the rotational search. Translational search and rigid body fitting for the top solutions using data to 2.5 Å gave a final solution with an initial R value of 43.5%. Refinement was performed using programs from CNS 1.0 (*36*). Detailed data and refinement statistics are listed in Table 1 (PDB entry ID: 1ID5). Visualization and rebuilding of the model was done with Quanta98 (Molecular Simulations Inc., San Diego, CA).

Table 1: Data and Refinement Statistics for Ecotin M84R−Thrombin Complex

structure	M84R−thrombin
space group	$C222_1$
unit cell constants (Å)	
a	88.5
b	165.4
c	83.3
$\alpha = \beta = \gamma$	90°
highest resolution (Å)	2.5
total reflections	239 796
unique reflections	20 967
completeness	97.7%
highest resolution shell	88.7% (2.59 to 2.50 Å)
$\langle I/I_\sigma \rangle$	13.3
R_{merge}[a]	
highest resolution shell	29.8% (2.59 to 2.50 Å)
solvent content	55%
R[b] after molecular replacement	43.5%
refinement resolution range	6−2.5 (Å)
R	19.5%
free R[c]	25.3%
water molecules	158
rmsd in bond length (Å)	0.006
rmsd in bond angle (°)	1.3

[a] $R_{merge} = \sum|(I - \langle I \rangle)|/\sum(I)$. [b] $R = \sum_{h,k,l}(|F_{obs}(h,k,l)| - k|F_{calc}(h,k,l)|)^2/\sum_{h,k,l}|F_{obs}(h,k,l)|$. [c] Free R: cross-validation R calculated by omitting 10% of the reflections (*47*).

RESULTS

Analysis of Thrombin-Ecotin M84R Complex. Following incubation on ice, a mixture of bovine thrombin and ecotin M84R at 1:1 stoichiometry yields a complex that can be isolated by gel filtration. The observed molecular mass of the isolated complex is around 100 kDa, which corresponds to the sum of the molecular masses of two ecotin and two thrombin molecules. Denaturing gel electrophoresis confirms that the sample contains both thrombin and ecotin M84R. Native gel electrophoresis reveals a single band that is not found in either thrombin or ecotin M84R sample, indicating the formation of a stable complex between thrombin and ecotin M84R (data not shown). Under the same conditions,

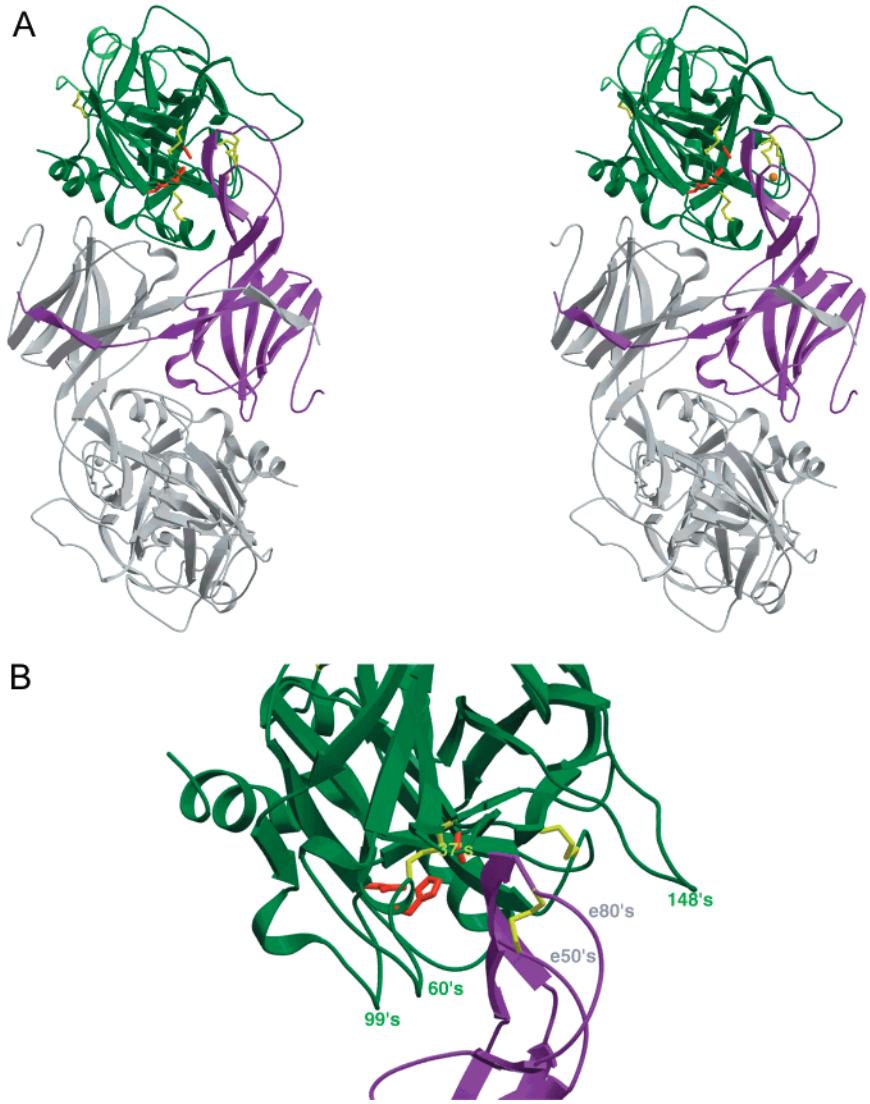

FIGURE 1: (A) Structure of bovine thrombin complexed to ecotin M84R in secondary structures. Light chain (L) and heavy chain (H) of thrombin are shown in green, with catalytic triad in red. Ecotin is shown in purple; disulfide bonds and calcium ion are highlighted in yellow and orange. The symmetry mate that helps to construct a full tetrameric complex is shown in gray. The figure was generated by Raster3D (*48*), with secondary structures rendered by Rasmol (*49*). (B) A close-up of the interface between thrombin and ecotin M84R at the active site of the protease denotes all surrounding surface loops in thrombin (the 37's, 60's, 99's, and 148's loops) and the two primary site loops in ecotin (the e50's and e80's loops).

no complex is observed from a mixture of wild-type ecotin and thrombin.

X-ray Structure of Thrombin-Ecotin M84R Complex. Ecotin M84R−thrombin complex is crystallized in the $C222_1$ space group. Two ecotin and two thrombin molecules form a tetramer as shown in Figure 1a. One asymmetric unit of the crystal contains only one thrombin and one ecotin; the other half of the tetramer is generated through crystallographic symmetry. The ecotin M84R−thrombin complex exhibits the same overall features as other ecotin−serine protease complexes, such as ecotin−trypsin and ecotin−collagenase (*14*, *15*). The ecotin M84R dimer binds to thrombin in the similar orientation as wild-type ecotin does in the ecotin−trypsin structure, contacting thrombin with two binding sites.

The primary binding site of ecotin M84R includes the e50's and e80's loops (*14*). The e80's loop binds to the active site of thrombin in a substrate-like manner with characteristic main chain interactions of canonical protease inhibitors (*37*). A total of eight residues from ecotin e80's loop, on both sides of the scissile bond, directly contact thrombin. The e50's loop functions by supporting and stabilizing the e80's loop through a disulfide bond between Cys50 and Cys87. The buried surface area at the primary contact area between thrombin and ecotin is 2290 Å^2, which is larger than any other known protease−ecotin interaction site (*14*, *15*). The secondary binding site, mainly consisting of the e60's and e110's loops, binds to the 99's loop and the C-terminal helix of thrombin that are about 27 Å away from the active site. The total buried surface area at the secondary binding site is 1250 Å^2, also larger than any other secondary site in known ecotin−protease complexes (*14*, *15*).

As compared with the structure of uncomplexed ecotin dimer (*38*), the two ecotin molecules adjust their relative position slightly upon binding to thrombin. However, they do not change their conformations on inhibiting thrombin as they do on inhibiting collagenase (*15*). The structure of ecotin M84R mutant exhibits few significant differences from the structure of wild-type ecotin when bound to trypsin (*14*) or in its unbound form (*38*), with rmsd of 0.39 and 0.48 Å

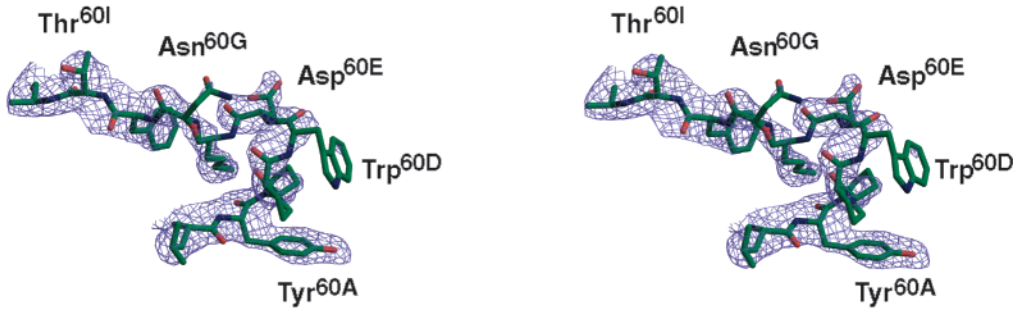

FIGURE 2: Electron density ($2F_{obs}-F_{calc}$) of the 60's insertion loop from the bovine thrombin-ecotin M84R complex contoured at 1σ. Density for the side chains of Trp^{60D} and Asn^{60G} and the peptide bond between residues Asp^{60E} and Lys^{60F} is missing. Both are confirmed by a composite omit map, which possibly results from the high B-factors of most residues in this insertion loop. The figure was produced using Raster3D (*48*).

for superimposing the conserved core structure that includes residues 12−43, 59−63, 70−76, and 103−131 (*15*).

The Thrombin Structure. The overall thrombin structure in the complex maintains the classic morphology of serine proteases of the chymotrypsin fold (*19*). The globular structure of thrombin consists mainly of β strands, with only a few helical segments and various surface-located loops. The catalytic residues: His^{57}, Asp^{102}, and Ser^{195} are located in the junction cleft formed by the two beta barrel domains of thrombin. The conformations of the catalytic triad and surrounding regions are well preserved (*19*). Our thrombin core structure, excluding the light chain and surface loops (Figure 1b) that include the 60's (Leu^{60} to Thr^{60I}), the 148's (Trp^{148} to Glu^{149E}), the 37's (Arg^{35} to Glu^{39}), and the 99's insertion loops (Arg^{93} to Leu^{99}), compares with published coordinates with rmsd smaller than 0.4 Å for over 160 equivalent C_α atoms. The four surface loops, however, have undergone significant movements to permit the binding to ecotin M84R.

The 60's insertion loop contains a fragment of nine amino acids with the sequence of Y^{60A}PPWDKNFT^{60I} (according to chymotrypsin nomenclature) (*7*). This loop is unique to thrombin with its amino acid sequence highly conserved across different species (*39*), indicating its significance in thrombin substrate recognition. Previous structural analyses have shown that the 60's insertion forms a rigid loop above the active site, and it exhibits little conformational change upon inhibitor binding (*19*, *30*). This loop's clashing into the e50's loop from ecotin has been hypothesized to account for wild-type ecotin's failure to bind thrombin. In our structure, the 60's loop moves significantly away from the substrate-binding cleft. The side chains of Trp^{60D} and Asn^{60G} become disordered upon ecotin binding (Figure 2). Moreover, the C_α of Trp^{60D} moves nearly 7 Å as compared to the position of the same atom when thrombin is bound with small peptide inhibitors (Figure 3a). The apparent rigidity of the 60's loop suggests that there may be a significant energy cost to promote movements of this loop.

The 148's loop also moves significantly upon ecotin binding (Figure 3b), as compared to the conformations when bound to small inhibitors. This loop contains a shorter insertion of T^{149A}SVAE^{149E} (*7*) and is located on the opposite side of the active site from the 60's loop. The sequence of the 148's insertion is not as conserved as that of the 60's loop (*39*), neither is its conformation (*19*, *30*, *40*). In our structure, the 148's loop adopts an extended conformation (Figure 3b) and is slightly disordered. No direct contacts are formed between the 148's loop and ecotin. This loop seems intrinsically flexible, suggesting a small energy cost for its conformational change.

The 37's and the 99's loops, two surface loops adjacent to the 60's insertion loop, also show minor conformational change upon ecotin binding (Figure 3c,d), when compared to the conformations of these two loops when bound to other inhibitors (*19*, *24*, *25*, *30*, *41*). In particular, Trp^{96}, a residue important for the catalytic activity of thrombin (*42*), forms a hydrogen bond with Tyr100 from ecotin. Moreover, the side chain of this residue clearly differs from its conventional position in other structures (Figure 3c). Previous studies showed that the 37's and the 99's loops are involved in substrate binding at P2 and P1′ sites (*26*, *43*). The small changes in the positions of these loops may result in size changes of the S2 and S1 substrate binding pockets, therefore allowing different amino acid recognition.

Interactions between Ecotin Primary Binding Loop and Thrombin Active Site. The primary site e80's loop of ecotin M84R interacts extensively with the active site of thrombin. Although not optimized for recognition (except for the Met to Arg mutation), residues Ser79 (P6) to Ala86 (P2′) from ecotin span across the active site cleft of thrombin, providing a platform for predicting amino acid preference at each substrate recognition site of thrombin.

Arg84, the only mutated residue in ecotin, fits exceptionally well into the S1 binding pocket of thrombin. Arg84 is the ecotin residue that forms most interactions with thrombin: it makes five direct hydrogen bonds and a salt bridge with five residues in thrombin in addition to two water-mediated hydrogen bonds and various hydrophobic interactions (Figure 4a). Gly^{193}, Asp^{194}, and Ser^{195} coordinate with the main chain -N and the carboxyl -O of Arg84. Asp^{189}, which lies at the bottom of the S1 pocket, contributes to Arg side chain recognition by forming a salt-bridge. Gly^{219} makes an additional hydrogen bond with the terminal -$N_{\eta 2}$ of Arg84. Gly^{216} and Phe^{227} form two water-mediated hydrogen bonds with Arg84 side chain. Furthermore, Ala^{190} and Val^{213} coordinate the aliphatic side chain of Arg84 through hydrophobic interactions. Most interactions established between the side chain of Arg84 and thrombin could not be formed if this residue were a Met. Additional energy from these interactions may be essential for the change in inhibition profile.

It is not obvious whether Thr83 is the optimized residue at the P2 position for a thrombin inhibitor. The S2 pocket appears to have a predominately hydrophobic character. Leu^{99}

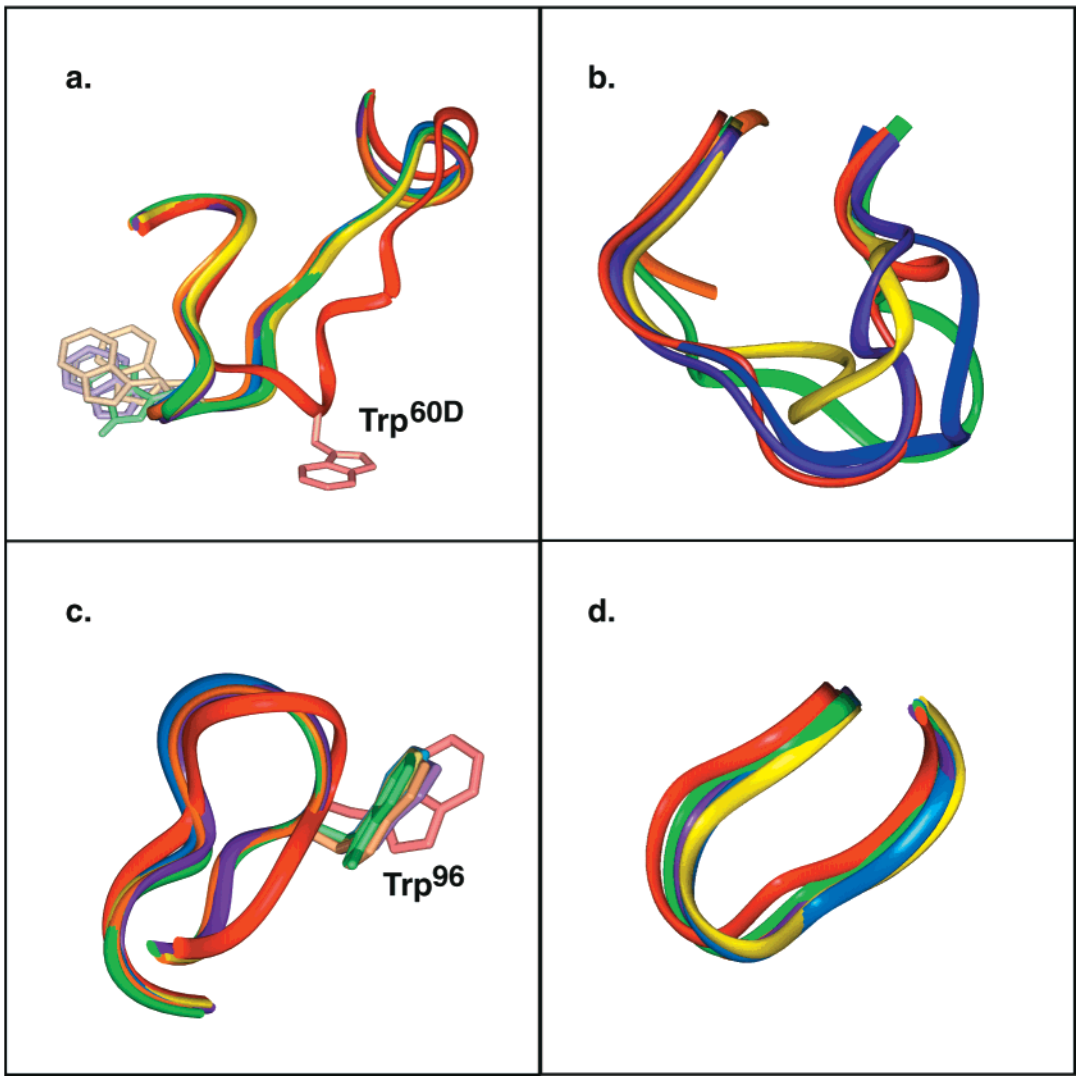

FIGURE 3: Comparisons of conformations of (a) the 60's, (b) the 148's, and (c) the 99's (d) the 37's loops bound with different inhibitors. Bovine thrombin structures, 1ETR.pdb (green, bound to small inhibitor 2MQPA), 1BBR.pdb (blue, bound to fibrinopeptide α2 7−16), 1TBR.pdb (yellow, bound to rhodniin), 1TOC.pdb (purple, bound to ornithodorin), 1UVT.pdb (gold, bound to small inhibitor MB14.1248) and the ecotin bound thrombin molecule (red), are superimposed based on core residues. Trp^{60D} in all structures in (a) are shown in full stick model. The conformations of all loops are shown in ribbon diagram. All figures were generated with InsightII (Accelrys, San Diego, CA).

and Tyr^{94} set the boundary in one direction; Trp^{215} and Tyr^{60A} further define this volume in the other direction. Trp^{96} forms a lid to close the binding pocket. As shown in Figure 4a,b, the S2 pocket seems small enough to exclude large residues such as Phe, Trp, Arg, and Lys. In vitro studies with positional scanning libraries of fluorogenic peptides have shown that the S2 site of thrombin exhibits a strong preference for Pro (*29*). The preference for a Pro at the P2 position is well supported by the structure of human α-thrombin with PPACK (*19*). In this structure, the 60's insertion loop is right above the active site and creates a small rigid pocket that seems to be selective for a Pro. Furthermore, Pro is found as the P2 residue in many natural substrates or inhibitors of thrombin (Figure 4c). Although the thrombin−ecotin M84R structure indicates that the restriction by the 60's loop no longer exists upon the binding of certain macromolecules, Pro may still be favored as the P2 residue for having the additional role of providing rigidity to the substrate backbone to enhance extended binding.

Ser82 occupies the P3 position in ecotin. Two hydrogen bonds with Gly^{216} well coordinate the main chain -N and carboxyl -O in Ser82. Two water-mediated hydrogen bonds are formed between its side chain and main chain -N of Glu^{192} and carboxyl -O of Gly^{219}. There is no obvious preference for a Ser residue. As the S3 pocket is partially solvent exposed, a variety of residues can be docked in the S3 pocket such as Asp, Lys, and Arg. Conversely, the P4 position is much more selective with Val81 well recognized by thrombin. A local hydrophobic environment, which is limited in size, explains the preference for a Val at this position. As depicted in Figure 4b, Trp^{215} paves the base of the pocket and Ile^{174} makes significant hydrophobic contact with Val83. $Met^{180}$and Leu^{99}, not shown in the figure, also fortify a well-formed hydrophobic binding pocket suitable for a Val, Ile, or Leu. Our observations at S3 and S4 substrate binding sites are consistent with in vitro studies with combinatorial substrate libraries (*29*) and with the sequences of some physiological substrates of thrombin (Figure 4c).

The role of Pro80 at the P5 position appears to be to create a rigid turn so that Ser79 at the P6 position can form a favorable hydrogen bond interaction with the side chain of Arg^{221A} from thrombin (Figure 4a,b). Lys^{224} is also in the

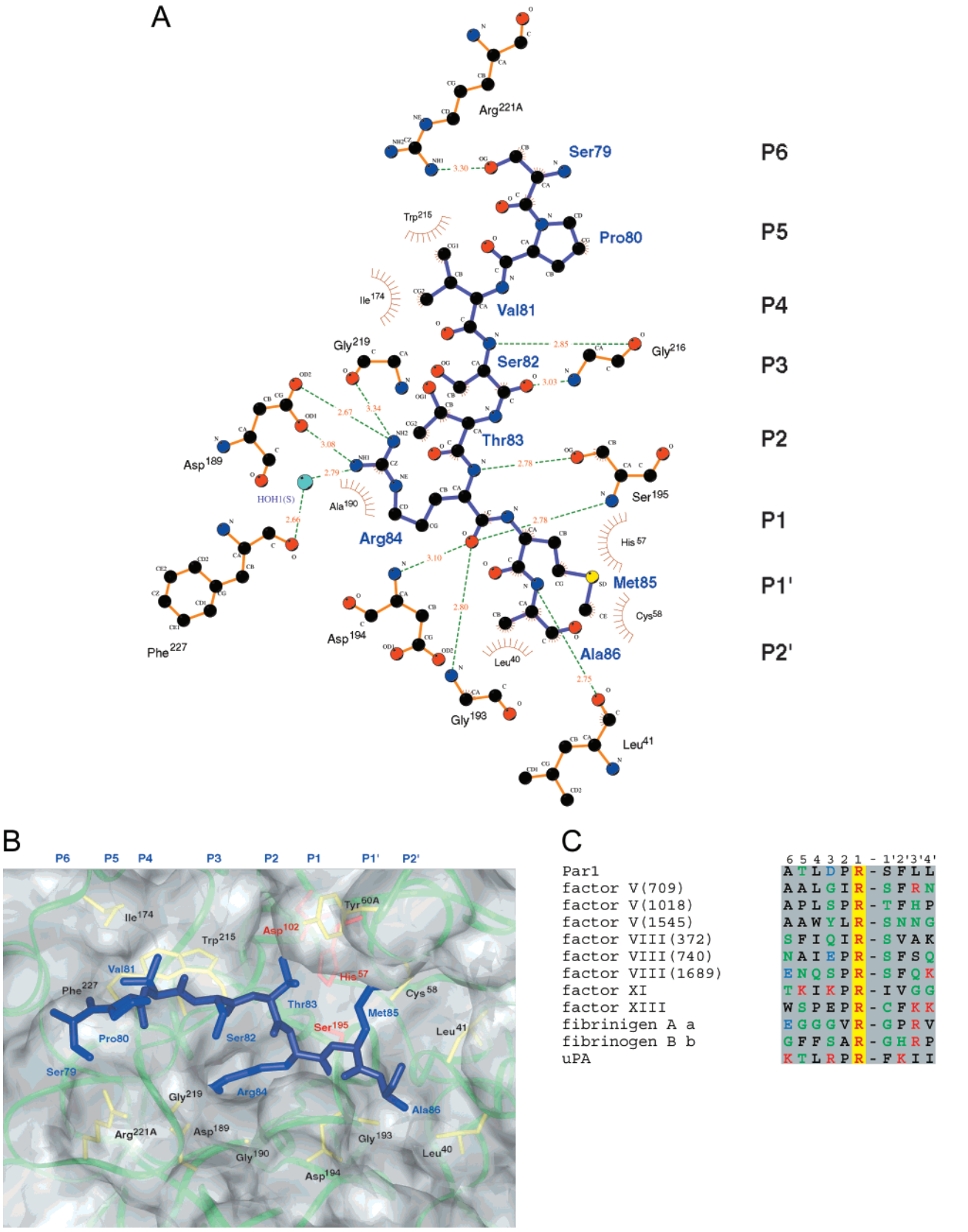

FIGURE 4: Details of the primary binding site interaction between bovine thrombin and ecotin M84R mutant are shown in (A) a two-dimensional plot and (B) a three-dimensional surface representation of thrombin surface. In (A), residues 79 through 86 in ecotin are laid out vertically as the P6 to P2′ site. Residues from thrombin that make hydrogen bonds (linked by green dash line with distance labeled in red) or hydrophobic contacts (in brown with residues labeled in black) are drawn in each side of ecotin ligand. In (B), the same fragment from ecotin is shown in stick model horizontally from P6 to P2′, with main chain atoms in navy and side chain atoms in blue. Thrombin is shown as a transparent surface with its C_α trace shown as green ribbon. Side chains of the residues directly contributing in ecotin binding are shown in stick mode in yellow. The catalytic triad is in red. Panels A and B are generated using Ligplot (*50*) and InsightII (Accelrys, San Diego, CA), respectively. Panel C lists the cleavage sites of some physiological substrates of thrombin.

vicinity of Ser79. Ser79 is the furthest residue N-terminal to the scissile bond that directly contacts thrombin; its side chain -OH is solvent exposed.

Our structure also reveals ecotin and protein interactions C-terminal to the P1 residue because ecotin stops the catalysis of serine protease prior to the cleavage of the scissile bond (*14*). Figure 4b shows that the P1′ residue Met85 takes an extended conformation in binding the substrate binding pocket, underneath the insertion loop at residue 60 of thrombin. The side chain of Met85 is sandwiched between two disulfide bonds: Cys^{42}−Cys^{58} from thrombin and Cys50−Cys87 from ecotin. Tyr^{60A} defines the bottom of the S1′ pocket. Pro^{60B}, Phe^{60H}, and the aliphatic side chain of the buried Lys^{60F} are all within 6 Å to Met85. Leu^{41} and His^{57} further enhance the hydrophobicity of the S1′ pocket. The P2′ site is occupied by Ala86, the last ecotin residue in the e80's loop that directly contacts thrombin. Its main chain -N forms a hydrogen bond with the carboxyl -O in Leu^{41}. Its methyl side chain interacts extensively with the aliphatic side chain of Leu^{40}. A Val, Leu, or Ile may also be well accommodated because of the local hydrophobicity at the S2′ site. However, the S2′ pocket is only partially hydrophobic. In the direction toward the exposing surface of the protease, Arg^{73}, Asn^{143}, Gln^{151}, and Glu^{192} are within 8 Å of Ala86. Therefore, a charged residue with long side chain, such as an Arg or a Lys can possibly go through the local hydrophobic area and establish interactions with these polar residues.

The geometric vicinity of the S1′ and S2′ sites creates a combined specificity, i.e., residue bound in S1′ affects the choice of residues at S2′ site. When the S1′ site is bound with a small noncharged residue, the S2′ site can accommodate either a large hydrophobic residue such as a Phe to fill the large hydrophobic space, or a small hydrophobic or noncharged residue followed by a similar residue at the S3′ site. In the presence of a large hydrophobic P1′ residue such as a Phe, the S2′ pocket will be limited to a large charged residue like a Lys, as found in uPA (Figure 4c). An Arg may also fit; other alternatives include small noncharged residues such as Cys, Ser, Ala, or Gly.

DISCUSSION

A simple docking and fitting model does not explain the differences in interactions of thrombin with wild type or ecotin M84R mutant. The conformations of thrombin likely shift between a "closed" state and an "open" state; only in the "open" state, the active-site-restricting surface loops move away to permit ligand binding to thrombin. It is possible that a transient complex between thrombin and ecotin M84R exists prior to the stable final complex we observe in the static X-ray structure. The major energetic contribution for the formation of the final ecotin M84R−thrombin complex comes from the numerous interactions the Arg side chain establishes with thrombin (Figure 4b). A similar transient complex also likely exists between thrombin and wild-type ecotin, but it lacks the interaction affinity for the transition to a more stable complex that can then be isolated by gel filtration. Conformational changes, especially those in the surface loops surrounding thrombin active site, are required for the formation of the transient complexes. Such conformational changes often require high energetic costs, which

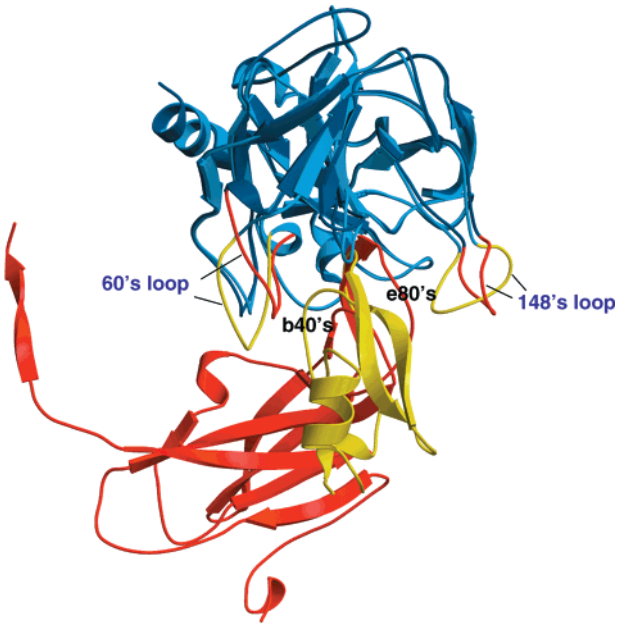

FIGURE 5: Thrombin from E192Q−BPTI complex (BPTI and the 60's and 148's loop shown in yellow) is superimposed with the ecotin M84R bound thrombin (the 60's, 148's loops and bound ecotin M84R in red). The conserved thrombin structures are in light gray. The 40's loop from BPTI and the e80's loop from ecotin are labeled. Figure was generated with Raster3D (*48*).

suggests the metastable nature of the transient complexes. A survey of existing structures of thrombin−inhibitor complexes indicates that the main energy cost probably comes from the conformational change of the 60's loop because of its apparent rigidity in most thrombin structures. The 148's insertion loop seems to have a more flexible nature and likely a small energy cost for conformational changes.

Two approaches can be taken to permit stable thrombin−inhibitor complex formation, to reduce the cost in conformational change by deleting or mutating residues in the 60's loop such as the desPPW mutant (*20*), or to increase the energy of interactions as found in ecotin M84R−thrombin. Ecotin M84R−thrombin complex is not the only example where positive binding energy inputs overcomes the energy cost due to conformational changes, particularly in the 60's loop. The interactions between thrombin and ecotin M84R showed similar characteristics to those observed in the BPTI−thrombin E192Q complex (*25*). BPTI weakly inhibits wild-type thrombin with a μM level K_i. A point mutation in thrombin: E192Q increased the inhibition of BPTI against thrombin by 500-fold (*44*). The E192Q mutation in thrombin eliminates a negative repulsion between the terminal -O in Glu^{192} side chain and BPTI Cys14 and adds a hydrogen bond between Gln^{192}-$N_{\epsilon 2}$ and Cys14-O of BPTI. The "extra" interactions from both point mutations are critical for the final stabilization of the thrombin−M84R or E192Q−BPTI complex.

The structures of thrombin−ecotin M84R and thrombin E192Q−BPTI are the only two existing examples where "rigid" insertion loop at thrombin residue 60 moves significantly upon inhibitor binding. Figure 5 shows that the 60's and the 148's loops in thrombin adopt different conformation when bound to different inhibitors since BPTI (58 amino acids) and ecotin (142 amino acids) differ much in their sizes and three-dimensional structures. Except for the conformation

of the P1 residue, the reactive site loop of BPTI differs considerably from that of ecotin. When the C_α atoms of seven residues from each reactive site loop are superimposed, the rmsd value is greater than 1.5 Å. The 40's loop in BPTI creates a bulkier surface than the corresponding surface in ecotin M84R. As a result, the 60's loop moves further away from the active site in the BPTI bound state than in the ecotin M84R bound state. On the opposite side, the residues at the end of the e80's loop in ecotin M84R protrude more than the corresponding side of BPTI. Consequently, the insertion loop at thrombin residue 148 in ecotin bound state moves further away from the active site than its BPTI bound state. It appears that the extent of the loop movement is determined by the conformations of the incoming inhibitors.

Further analyses suggest that the 60's insertion loop in thrombin may be involved in functions more than accommodating an incoming ligand. Analyses of the packing of this loop by the Naccess program (*45*) on the current structure and over 10 published thrombin coordinates show that little change has occurred in the side-chain solvent accessibility for almost all residues in the insertion loop at thrombin residue 60. The same analysis reveals variations in side-chain solvent accessibility for residues in the insertion loop at thrombin residue 148. This observation and the highly conserved sequence suggest that the 60's insertion loop may also participate in substrate recognition. However, most residues in the 60's insertion loop of thrombin are too far above the active site cleft to make contacts with bound substrates. Anchoring residues for the loop, Tyr^{60A} and Phe^{60H}, are exceptions that contribute to the hydrophobicity of the S2 and S1 sites. Distance difference matrix plots (data not shown) show that the movement of the 60's insertion loop is always accompanied by movements of the 37's and the 99's loops which are indeed involved in substrate recognition (*19*, *43*). Because the 60's insertion loop is sandwiched between the 37's and the 99's loops, it may influence substrate binding indirectly: by affecting the position of the adjacent 37's and 99's loops. On the other hand, ligand binding at the 37's or 99's loop may also cause conformational changes in the 60's insertion loop, although this possibility has not been seen beyond the observation reported.

Designing selective inhibitors has been challenging since thrombin and other blood coagulation factors share considerable similarities in primary sequence and three-dimensional structures. Extended substrate specificity revealed here may be employed in the quest for more specific inhibitors. For example, placing a sharp turn followed by a negatively charged residue at P5, P6 and an extended hydrophobic group at P1′ may produce a more potent inhibitor of thrombin. A Val at P4 position is also a good choice. Finally, we showed that a single point mutation, M84R in a surface interaction of 6000 $Å^2$, enables strong binding. Our work shows this occurs because the Arg84 appears to be a "hot spot" for the interactions between ecotin and thrombin (*46*). It forms seven hydrogen bonds, one salt bridge as well as several hydrophobic interactions through its aliphatic side chain. Energy from these interactions is critical for stabilizing conformational changes required for binding and for the 10^4 times increase in affinity.

ACKNOWLEDGMENT

The authors thank Dr. Sarah Gillmor for her work on human thrombin and ecotin and Christopher Eggers for his help in K_i measurements. This work was funded by the UC Biotechnology STAR Project and NIH Grant DK39304.

REFERENCES

1. Fenton, J. W. d. (1988) *Semin. Thromb. Hemostasis 14*, 234−40.
2. Blomback, B., Blomback, M., Hessel, B., and Iwanaga, S. (1967) *Nature 215*, 1445−8.
3. Hogg, D. H., and Blomback, B. (1978) *Thromb. Res. 12*, 953−64.
4. Kisiel, W., Canfield, W. M., Ericsson, L. H., and Davie, E. W. (1977) *Biochemistry 16*, 5824−31.
5. Esmon, N. L., Owen, W. G., and Esmon, C. T. (1982) *J. Biol. Chem. 257*, 859−64.
6. Rosenberg, R. D., and Damus, P. S. (1973) *J. Biol. Chem. 248*, 6490−505.
7. Stubbs, M. T., and Bode, W. (1993) *Thromb. Res. 69*, 1−58.
8. Sheehan, J. P., Tollefsen, D. M., and Sadler, J. E. (1994) *J. Biol. Chem. 269*, 32747−51.
9. Rezaie, A. R., Cooper, S. T., Church, F. C., and Esmon, C. T. (1995) *J. Biol. Chem. 270*, 25336−9.
10. Gronke, R. S., Bergman, B. L., and Baker, J. B. (1987) *J. Biol. Chem. 262*, 3030−6.
11. Lombardi, A., De Simone, G., Galdiero, S., Staiano, N., Nastri, F., and Pavone, V. (1999) *Biopolymers 51*, 19−39.
12. Chung, C. H., Ives, H. E., Almeda, S., and Goldberg, A. L. (1983) *J. Biol. Chem. 258*, 11032−8.
13. Seymour, J. L., Lindquist, R. N., Dennis, M. S., Moffat, B., Yansura, D., Reilly, D., Wessinger, M. E., and Lazarus, R. A. (1994) *Biochemistry 33*, 3949−58.
14. McGrath, M. E., Erpel, T., Bystroff, C., and Fletterick, R. J. (1994) *EMBO J. 13*, 1502−7.
15. Perona, J. J., Tsu, C. A., Craik, C. S., and Fletterick, R. J. (1997) *Biochemistry 36*, 5381−92.
16. Perona, J. J., Tsu, C. A., Craik, C. S., and Fletterick, R. J. (1993) *J. Mol. Biol. 230*, 919−33.
17. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−62.
18. Gillmor, S. A., Takeuchi, T., Yang, S. Q., Craik, C. S., and Fletterick, R. J. (2000) *J. Mol. Biol. 299*, 993−1003.
19. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R., and Hofsteenge, J. (1989) *EMBO J. 8*, 3467−75.
20. Le Bonniec, B. F., Guinto, E. R., MacGillivray, R. T., Stone, S. R., and Esmon, C. T. (1993) *J. Biol. Chem. 268*, 19055−61.
21. Yang, S. Q., Wang, C. I., Gillmor, S. A., Fletterick, R. J., and Craik, C. S. (1998) *J. Mol. Biol. 279*, 945−57.
22. Laskowski, M., and Qasim, M. A. (2000) *Biochim. Biophys. Acta 1477*, 324−37.
23. Stubbs, M. T., Oschkinat, H., Mayr, I., Huber, R., Angliker, H., Stone, S. R., and Bode, W. (1992) *Eur. J. Biochem. 206*, 187−95.
24. Martin, P. D., Malkowski, M. G., DiMaio, J., Konishi, Y., Ni, F., and Edwards, B. F. (1996) *Biochemistry 35*, 13030−9.
25. van de Locht, A., Bode, W., Huber, R., Le Bonniec, B. F., Stone, S. R., Esmon, C. T., and Stubbs, M. T. (1997) *EMBO J. 16*, 2977−84.
26. Le Bonniec, B. F., Myles, T., Johnson, T., Knight, C. G., Tapparelli, C., and Stone, S. R. (1996) *Biochemistry 35*, 7114−22.
27. DiBella, E. E., and Scheraga, H. A. (1996) *Biochemistry 35*, 4427−33.
28. Rezaie, A. R., and Olson, S. T. (1997) *Biochemistry 36*, 1026−33.
29. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. A. (2000) *Nat. Biotechnol. 18*, 187−93.
30. Martin, P. D., Robertson, W., Turk, D., Huber, R., Bode, W., and Edwards, B. F. (1992) *J. Biol. Chem. 267*, 7911−20.

31. van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., Hoffken, W., and Bode, W. (1995) *EMBO J. 14*, 5149−57.
32. van de Locht, A., Stubbs, M. T., Bode, W., Friedrich, T., Bollschweiler, C., Hoffken, W., and Huber, R. (1996) *EMBO J. 15*, 6011−7.
33. Owen, W. G., Esmon, C. T., and Jackson, C. M. (1974) *J. Biol. Chem. 249*, 594−605.
34. McGrath, M. E., Erpel, T., Browner, M. F., and Fletterick, R. J. (1991) *J. Mol. Biol. 222*, 139−42.
35. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W. J., and Sweet, R. M., Eds.) pp 307−26, Academic Press, New York.
36. Brunger, 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. D Biol. Crystallogr. 54*, 905−21.
37. Bode, W., and Huber, R. (1992) *Eur. J. Biochem. 204*, 433−51.
38. Shin, D. H., Song, H. K., Seong, I. S., Lee, C. S., Chung, C. H., and Suh, S. W. (1996) *Protein Sci. 5*, 2236−47.
39. Banfield, D. K., and MacGillivray, R. T. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2779−83.
40. Grutter, M. G., Priestle, J. P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J., and Stone, S. R. (1990) *EMBO J. 9*, 2361−5.
41. Slon-Usakiewicz, J. J., Sivaraman, J., Li, Y., Cygler, M., and Konishi, Y. (2000) *Biochemistry 39*, 2384−91.
42. DiBella, E. E., and Scheraga, H. A. (1998) *J. Protein Chem. 17*, 197−208.
43. He, X., Ye, J., Esmon, C. T., and Rezaie, A. R. (1997) *Biochemistry 36*, 8969−76.
44. Guinto, E. R., Ye, J., Le Bonniec, B. F., and Esmon, C. T. (1994) *J. Biol. Chem. 269*, 18395−400.
45. Hubbard, S. J., and Thornton, J. M. (1993) *NACCESS*, Computer Program, Dept. of Biochemistry and Molecular Biology, University College London.
46. Clackson, T., and Wells, J. A. (1995) *Science 267*, 383−6.
47. Kleywegt, G. J., and Brunger, A. T. (1996) *Structure 4*, 897−904.
48. Merritt, E. A., and Bacon, D. J. (1997) in *Methods in Enzymology*; pp 505−524, Academic Press, New York.
49. Sayle, R. A., and Milner-White, E. J. (1995) *Trends Biochem. Sci. 20*, 374.
50. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng. 8*, 127−34.

BI010712H