The Crystal Structure of Glutamyl Endopeptidase from *Bacillus intermedius* Reveals a Structural Link between Zymogen Activation and Charge Compensation†

Rob Meijers,‡,§ Elena V. Blagova,|| Vladimir M. Levdikov,|| Galina N. Rudenskaya,⊥ Galina G. Chestukhina,@ Tatiana V. Akimkina,# Sergei V. Kostrov,# Victor S. Lamzin,‡ and Inna P. Kuranova*,||

European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, Institute of Crystallography, Russian Academy of Sciences, Leninski pr. 59, 119333 Moscow, Russia, Chemistry Department, Lomonosov Moscow State University, 119992 Moscow, Russia, Institute of Industrial microorganisms, Dorojnii pr. 1, 117545 Moscow, Russia, and Institute of Molecular Genetics, Kurchatov square, 123182 Moscow, Russia

Received July 30, 2003; Revised Manuscript Received December 5, 2003

ABSTRACT: Extracellular glutamyl endopeptidase from *Bacillus intermedius* (BIEP) is a chymotrypsin-like serine protease which cleaves the peptide bond on the carboxyl side of glutamic acid. Its three-dimensional structure was determined for $C222_1$ and $C2$ crystal forms of BIEP to 1.5 and 1.75 Å resolution, respectively. The topology of BIEP diverges from the most common chymotrypsin architecture, because one of the domains consists of a β-sandwich consisting of two antiparallel β-sheets and two helices. In the $C2$ crystals, a 2-methyl-2,4-pentanediol (MPD) molecule was found in the substrate binding site, mimicking a glutamic acid. This enabled the identification of the residues involved in the substrate recognition. The presence of the MPD molecule causes a change in the active site; the interaction between two catalytic residues (His47 and Ser171) is disrupted. The N-terminal end of the enzyme is involved in the formation of the substrate binding pocket. This indicates a direct relation between zymogen activation and substrate charge compensation.

Glutamyl specific endopeptidases (Glu-endopeptidases) cleave the peptide bond on the carboxyl end of a glutamic acid. Enzymes with such specificity were found in a number of microorganisms, including *Staphylococcus*, *Actinomyces*, *Streptomyces*, and *Bacillus*, as well as in some RNA viruses (*1*−*6*). Glu-endopeptidases from pathogenic microorganisms are responsible for several disease states. For instance, epidermolytic toxin A from *Staphylococcus aureus* (ETA) induces staphylococcal scaled skin syndrome in newborns (*7*). In spore-forming bacteria, these enzymes are activated prior to sporulation and probably participate in the regulation of this process (*8*). Viral enzymes cleave a polypeptide chain at the C-terminal side of glutamic acid (Glu) or glutamine (Gln), provided that the processing of viral polyproteins takes part in viral replication (*9*). These viral enzymes are specific to the nature of residues next to Glu; only small residues such as glycine, alanine, and serine can occupy the position following the scissile bond (*10*). Microbial enzymes are more specific and cleave only after a glutamic acid. This makes microbial Glu-endopeptidases widely applicable for protein sequencing, specific cleavage of fusion proteins, and enzymatic synthesis of biologically active peptides (*11*−*15*). Therefore, an understanding of the structural basis for the recognition of the glutamic acid side chain in the substrate binding groove is of major interest. The examination of three-dimensional structures of glutamyl endopeptidases from different sources will shed more light upon this recognition mechanism.

The majority of Glu/Gln-endopeptidases are serine proteases with a chymotrypsin-type fold (*1*). Among the viral enzymes, some are true serine proteases; others, like 3C proteases from hepatitis A and human rhino virus, are cysteine proteases, where a cysteine replaces serine in the active site. Despite a very low level of sequence similarity, the 3C cysteine proteases from rhino, polio, and hepatitis A viruses retain a chymotrypsin fold (*16*−*18*). These enzymes are denoted chymotrypsin-like (CHL)[1] proteases.

To date, the three-dimensional structures of several enzymes with glutamic acid specificity are available: epidermilytic A and B toxins (ETA and ETB, respectively) (*19*, *20*) as well as glutamyl endopeptidase from *Streptomyces griseus* (SGEP) (*21*) and nsp4 from arterivirus (*22*). A number of structures of glutamine specific enzymes have been solved: 3C proteinases from human rhino virus (*16*), hepatitis A virus (*17*), and polio virus (*18*) and 3C-like proteinases from human and porcine coronavirus (*23*). The 3C-like proteinases as well as nsp4 have an additional helical domain besides the chymotrypsin fold.

† This work was supported by the Russian Foundation for Basic Research (Grant 00-04-48280). I.P.K. thanks the European Molecular Biology Laboratory for supporting a visit to the Hamburg Outstation.

* Corresponding author. E-mail: inna@ns.crys.ras.ru.

‡ European Molecular Biology Laboratory.

§ Present address: Dana Farber Cancer Institute, 44 Binney St., Boston, MA 02115.

|| Institute of Crystallography of the Russian Academy of Sciences.

⊥ Lomonosov Moscow State University.

@ Institute of Industrial microorganisms.

Institute of Molecular Genetics.

[1] Abbreviations: BIEP, glutamylendopeptidase from *B. intermedius*; SGEP, glutamylendopeptidase from *S. griseus*; BLEP, glutamylendopeptidase from *B. licheniformis*; ETA, epidermolytic toxin A from *St. aureus*; CHL, chymotrypsin-like; MPD, 2-methyl-2,4-pentanediol.

10.1021/bi035354s CCC: $27.50
Published on Web 02/14/2004

Extracellular Glu-endopeptidase from *Bacillus intermedius* strain 3-19 (BIEP) is excreted at the late phase of bacterial growth. The biosynthesis of BIEP as well as other *Bacillus* glutamyl endopeptidases is enhanced before sporulation (*24*). The enzyme hydrolyzes the peptide bond on the carboxyl end of glutamic acid or after cysteic acid, when the β-chain of oxidized insulin is used as a substrate (*25*).

The bacteria synthesize BIEP as an inactive zymogen that contains 303 amino acid residues and includes the secretory peptide and activation peptide which precedes the sequence of the mature enzyme (*26*). The proenzyme can be activated *in vitro* with trypsin or subtilisin upon cleavage of a lysine−valine bond and N-terminal propeptide degradation. The mature protein consists of 215 residues. Sequence alignment with other serine proteases showed that BIEP belongs to the chymotrypsin family, and contains the conserved catalytic triad consisting of Ser171 (195), His47 (57), and Asp120 (102) (residues are numbered from the N-terminus of the active enzyme; numbers in parentheses denote the numbering according to the chymotrypsin template). The highest degree of sequence identity (~50%) was found between BIEP and glutamyl endopeptidase from *Bacillus licheniformis* (BLEP, Figure 1). The degrees of sequence identity with Glu-endopeptidases for which the three-dimensional (3D) structures have been determined are 28% for SGEP and 30% for ETA.

Here we present the first structure of a Glu-endopeptidase from *Bacillus*. The enzyme was crystallized in two crystal forms ($C222_1$ and $C2$) as described in ref *27*. The 3D structure was determined for both these forms to 1.5 and 1.75 Å resolution, respectively, using molecular replacement. The polypeptide fold and the active site residues in BIEP were identified. The substrate specificity pocket that accommodates the glutamic acid residue from the peptide was located, and amino acid residues that interact directly with the ϵ-carboxyl of the P1 glutamyl residue were identified (substrate residues (P) and enzyme substrate binding sites (S) are named according to the convention proposed by Schechter and Berger (*28*), with the scissile bond situated between P1 and P1′). The enzyme in the $C2$ crystals grown in the presence of 2-methyl-2,4-pentanediol (MPD) contained an MPD molecule in the substrate binding site and was described as a BIEP−MPD complex. The comparison of the substrate binding sites in BIEP, the BIEP−MPD complex, SGEP, and ETA revealed some themes and variations that shed light on the specificity modes of these enzymes.

RESULTS AND DISCUSSION

General Description of the Structure. The structure was determined using the molecular replacement method implemented in AMoRe (*29*) and ETA as the search model. The refinement was carried out with REFMAC (*30*), coupled with automated building and updating of the solvent structure using ARP/wARP 5.1 (*31*). The refinement converged at an R-factor of 15.6% (R_{free} = 18.4%), and statistics are given in Table 1.

The electron density map of the refined structure showed well-defined density for all the residues except for the side chains of Arg55, Lys60, and Arg154. Double conformations were built for residues Arg15, Leu36, Val73, Ser86, Thr110, Tyr113, Ser115, Ile129, the main chain of Ala160, Val185, Ser191, Glu207, and Tyr211. The model has good stereochemistry with all residues lying within the allowed regions of the Ramachandran plot.

Polypeptide Fold and Topology of BIEP. Like other proteases of the chymotrypsin family, BIEP consists of two domains separated by an extended crevice so that catalytic and substrate binding sites are situated on the interface between the two domains, readily accessible to the solvent (Figure 2).

One domain of BIEP consists of the elongated N-terminal part of the polypeptide chain (residues 1−21 and 121−201, denoted the C-domain following the chymotrypsin nomenclature) and contains the substrate binding loop (residues 166−192). The second domain is comprised of residues 22−111 and the C-terminal α-helix (residues 202−215) and is termed the N-domain. The domains are connected by a loop of nine surface-exposed residues (112−120). The residues of the catalytic triad are shared between the domains so that Ser171 (195) belongs to the C-domain whereas Asp97 (102) and His47 (57) are located in the N-domain.

For the majority of chymotrypsin-like proteases, both domains are topologically similar and have a tertiary structure built around six-stranded antiparallel β-sheets folded into a β-barrel. Only the C-domain in BIEP has such a topology (Figure 3). One of the strands of the β-barrel (3) is connected by a single S−S bridge to a short helix (α2) that contains catalytic His47.

The C-domain topology is characterized by antiparallel β-sheets I and II, consisting of five and four antiparallel β-strands, respectively. A topological comparison of BIEP against all protein structures deposited in the Protein Data Bank using the TOPS server (*32*) revealed that the topology of the C-domain is somewhat similar to that of domain II of zymogen E (*33*), which looks like a hybrid between a β-barrel and a β-sheet (Figure 3). Zymogen E is a precursor of protease E, which cleaves specifically after the carbonyl bonds of alanine, valine, serine, and threonine. A unique four-stranded β-sheet was found recently in the N-terminal domain of 2A proteinase from the common cold virus (*34*). In the case of BIEP, the N-terminus of the active enzyme is an essential part of the substrate binding domain. The typical β-barrel found in chymotrypsin is extended to a β-sheet to accommodate a β-strand situated at the N-terminus. The orientation of the N-terminus turns out to be essential for the formation of the substrate binding pocket.

Comparison of BIEP with Other Protein Structures. The DALI server (*35*) was used to compare the three-dimensional structure of BIEP with all known protein structures stored in the Protein Data Bank. Structural similarity is reported via a Z score, and parts of the two compared structures that align reasonably well are superimposed along their C_α atoms. Several protein structures could be superimposed on BIEP with an average rmsd of <3 Å: ETA (rmsd of 1.9 Å for 196 residues) (*20*), β-trypsin (rmsd of 2.5 Å for 191 residues) (*36*), heparin-binding protein (rmsd of 2.4 Å for 187 residues) (*37*), and the chymotrypsin-like 3C protease from rhino virus (rmsd of 2.6 Å for 152 residues) (*16*).

The DALI algorithm indicated no clear global structural similarity between BIEP and SGEP or zymogen E. Superposition of the independent domains between BIEP and SGEP gives rmsds of 1.6 Å for the N-domain for the C_α

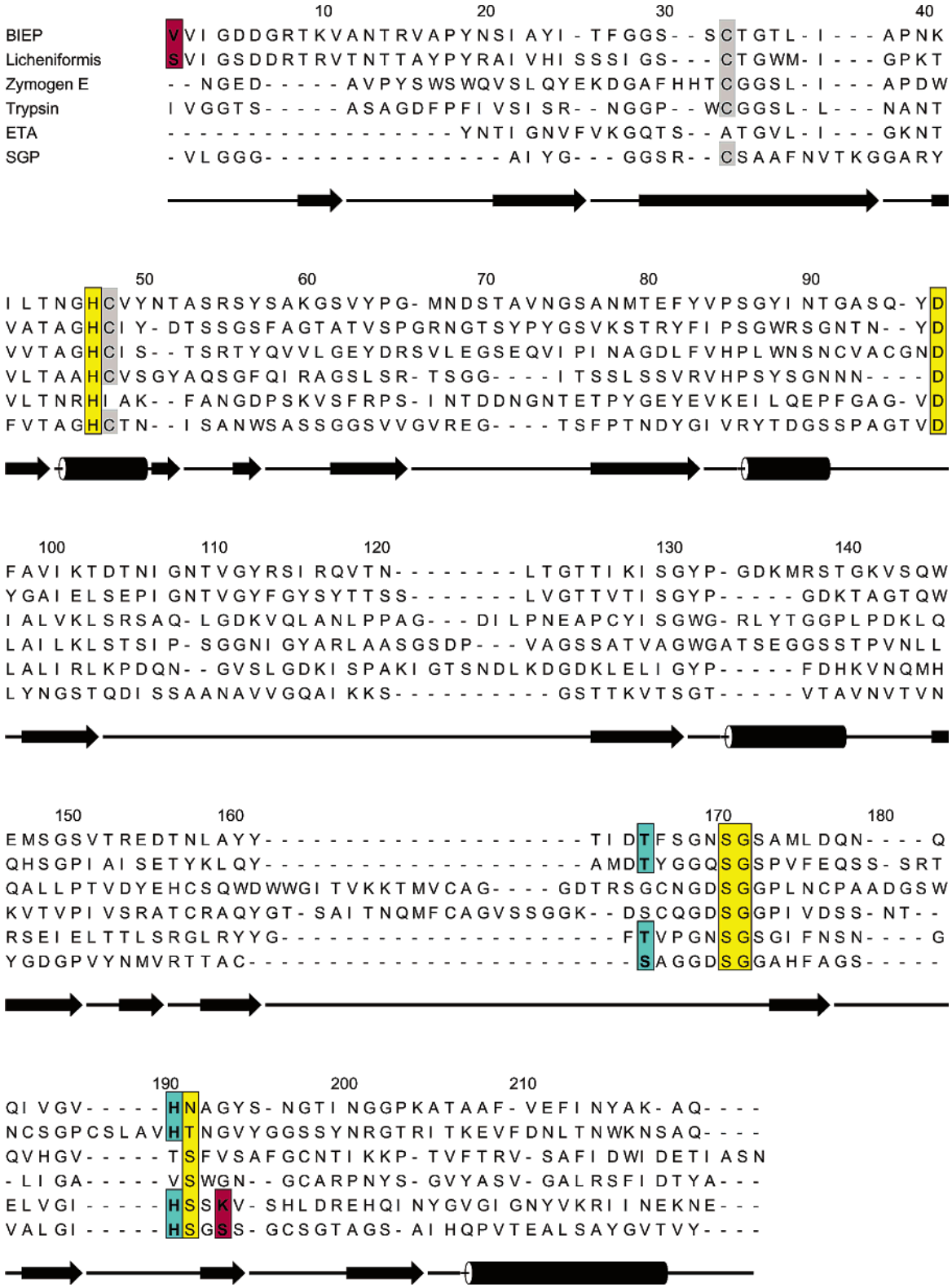

FIGURE 1: Sequence alignment of BIEP with related serine proteases. The sequence is numbered according to the BIEP sequence. The secondary structure elements of BIEP are displayed as arrows (β-strands) and barrels (α-helices). The one disulfide bond found in BIEP is shown in gray. Important catalytic residues are boxed and colored yellow. The residues that are part of the substrate binding site are colored as follows: cyan for residues that follow a general trend and crimson for residues that are specific for a certain enzyme.

atoms of 58 residues and 2.5 Å for the C-domain for the C_α atoms of 65 residues. For structural comparison, all atoms of the catalytic triad were used for superposition.

Disulfide Bonds in BIEP. The single disulfide bond of BIEP between Cys32 (42) and Cys48 (58) is approximately 4.5 Å from His47 (57) and maintains the conformation of the polypeptide chain near the active site. It seems that BIEP is an intermediate between SGEP and ETA in this respect. SGEP contains two disulfide bonds, one in a position similar to that of the disulfide in BIEP. ETA contains no disulfide bond, and is in this way reminiscent of viral CHL proteases (*10*).

Table 1: Statistics for Data Collection and Refinement for $C222_1$ and $C2$ Crystals of BIEP

	space group $C222_1$	space group $C2$
resolution range (Å)	20−1.5	20−1.72
overall completeness	97.5	97.3
overall $R(I)$ merge	6.9	10.8
no. of reflections	33091	18778
redundancy	3.7	3.2
I/σ	12.7	9.2
crystallographic R-factor (%)	15.6	16.6
R_{free} (%)	18.4	20.8
no. of protein atoms	1618	1618
no. of solvent atoms	155	58
no. of ligand atoms (MPD)	–	8
rmsd of bond lengths (Å)	0.013	0.013
rmsd of angles (deg)	0.031	0.037
average B-factor (Å^2) for all atoms, protein atoms, water molecules, MPD	18.6, 16.9, 37.0	23.4, 42.8, 37.3, 43.2

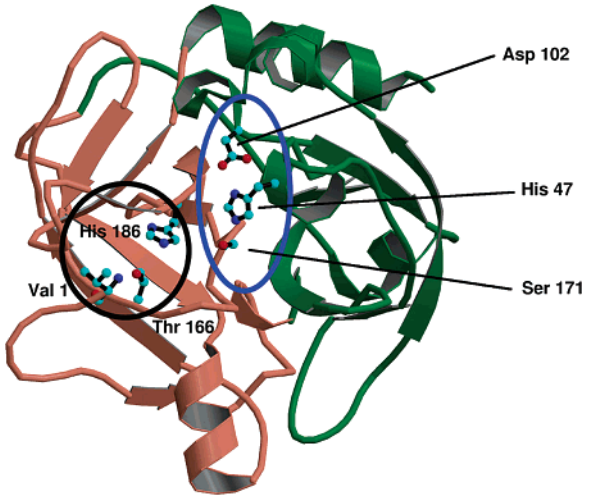

FIGURE 2: Ribbon diagram of Glu-endopeptidase from *B. intermedius*. The residues of catalytic and substrate binding sites are shown. This figure was generated with MOLSCRIPT (*48*) and rendered with Raster3D (*49*).

Active Site Area. The active site area in serine proteases can be divided into three functionally different parts: the residues of the catalytic triad (Ser-His-Asp), the oxyanion hole, and the substrate binding region, which comprises the S1 substrate pocket (*38*). The latter accommodates the characteristic P1 peptide residue of the substrate and, consequently, is responsible for the enzyme specificity.

Catalytic Triad. The position and geometry in BIEP of the catalytic triad of Ser171 (195), His47 (57), and Asp102 (97) coincide with those of other serine proteases. The mechanism of catalysis by these enzymes and the role of the individual amino acid residues have been reviewed in detail (*38*, *39*). It is assumed that enzymatic cleavage of the polypeptide bond proceeds via nucleophilic attack of the serine hydroxyl on the carbonyl carbon atom of the scissile peptide bond. The tetrahedral intermediate is transferred into an acyl enzyme, which is subsequently hydrolyzed at the final step of the reaction. The Asp^-/His^+ pair, which arises as a result of acceptance of a proton from the catalytic serine, provides electrostatic stabilization of the transition state. Hence, the lengths of the hydrogen bonds between the residues of the catalytic triad are important characteristics of the enzyme. Like in other related proteases, the hydrogen bond between

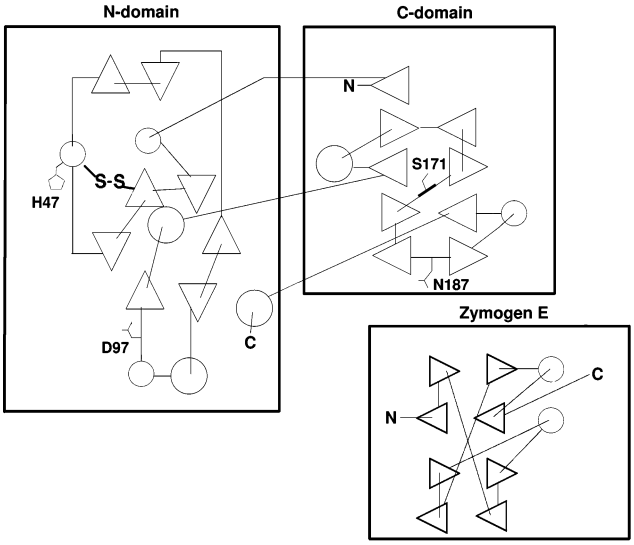

FIGURE 3: Topology diagrams generated with TOPS (*32*) for BIEP and the substrate binding domain of zymogen E (*33*). Large circles denote α-helices, small circles denote 3^{10} helices. The position of the catalytic triad and the single disulfide bond in BIEP are marked. Both zymogen E and BIEP have a β-sheet topology in the C-domain.

Table 2: Lengths of Some Essential Hydrogen Bonds in Several Serine Proteases

	hydrogen bond length (Å)	
protease	Ser195[a] Oγ−His57[a] Nϵ2	His57[a] Nδ1−Asp102[a] Oδ2
BIEP	2.83	2.69
ETA	2.97	2.61
trypsin	3.03	2.71
SGEP	3.39	2.91
rhino virus[b]	3.98	2.08

[a] According to chymotrypsin numbering. [b] In 3C protease from rhino virus, Ser195 is replaced with a cysteine and Asp102 with a glutamic acid.

the catalytic serine and histidine in BIEP is longer than that between histidine and aspartate (Table 2). The latter (2.69 Å) has been described as a low-barrier hydrogen bond (*40*).

Catalytic residue Asp97 (102) forms a hydrogen bond with His186 (213) and, consequently, with Asn45 and Asn187, which is homologous to Ser214 in chymotrypsin. Asn187 belongs to the substrate binding region, while His186 is a crucial residue in the S1 specificity pocket (see below). Thus, the cluster made up of the catalytic aspartate and two asparagine residues connects the catalytic and substrate binding region in BIEP. Because of the hydrogen bond with Asp97 (102), Asn187 is able to relegate the proton extraction from the catalytic serine. In addition, Asn187 (214) contacts the solvent through a water molecule and thus can mediate between the catalytic triad and the surface of the enzyme. According to Meyer *et al*. (*41*), a similar role was attributed to Ser214, which is invariant in the majority of mammalian and bacterial serine proteases. For this reason, Ser214 (Asn187 in BIEP) was considered as an additional member of the catalytic site.

Oxyanion Hole. The oxyanion hole is made up by the backbone of Ser171 and two preceding Gly residues (Gly169 and Gly170) and stabilizes electrostatically the negatively charged tetrahedral intermediate that arises during catalysis.

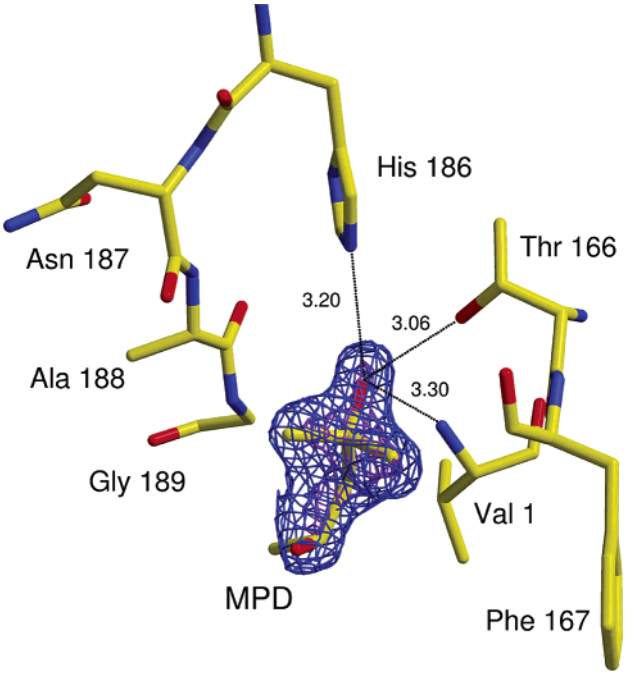

FIGURE 4: View of the substrate binding pocket with the electron density displayed around the MPD molecule, with contour levels of 1σ and 2σ from a $2mF_o - DF_c$ map (*30*), where $1\sigma = 0.45$ e/Å^3. The distances (angstroms) between the O2 atom of the MPD molecule and its protein ligands are given.

The negative charge is in part neutralized by the interaction with hydrogens of the imido groups of conserved Gly193 and Ser195.

The geometry of the oxyanion hole has a dramatic influence on the enzymatic activity. A distorted geometry of the oxyanion loop is found in ETA, where the carbonyl oxygen of Pro192 fills the oxyanion hole, thus blocking substrate binding (*19*, *20*). It has been suggested that this proline residue may move away to open the oxyanion hole when ETA binds another unknown activating molecule. In BIEP, the oxyanion hole is occupied by two water molecules, which should be displaced upon substrate binding. The geometry of the oxyanion hole is compatible with that of the majority of serine proteases. The distance between the main chain nitrogen atoms of Gly169 (193) and Ser171 (195) in BIEP is nearly the same as in trypsin (*42*), 4.2 and 4.3 Å, respectively. On the other hand, in zymogen E, the inactive form of protease E, Gly193 is turned away from the catalytic residues and the corresponding distance is 6.3 Å (*33*).

Substrate Binding Region. The structure of the BIEP−MPD complex with an MPD molecule in the substrate binding pocket has been studied. Crystals grown in the presence of 3% MPD had a lower symmetry, and the structure was determined using the free enzyme as a search model by molecular replacement. The BIEP−MPD complex was refined to an *R*-factor of 16.3% ($R_{free} = 20.9\%$) based on the $C222_1$ model. Superposition of the two crystal structures, $C222_1$ and *C*2, gives a root-mean-square deviation (rmsd) of 0.19 Å for all C_α positions.

The main difference from the free enzyme is the presence of residual electron density inside the substrate (S1) binding pocket. The electron density was interpreted as an MPD molecule (Figure 4). The oxygen at the C2 atom of the MPD molecule makes hydrogen bonds with two residues identified

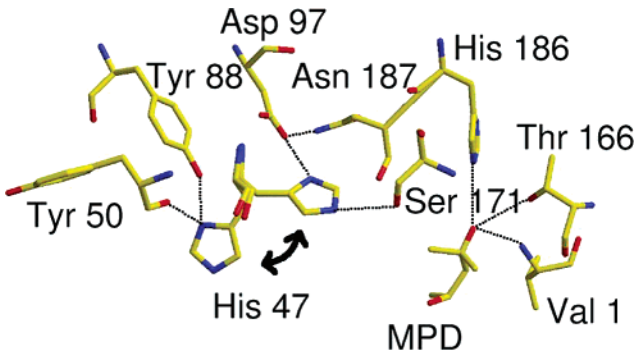

FIGURE 5: View of the active site area and the substrate binding pocket, with the MPD molecule. His47 (57) adopts two alternate conformations in the *C*2 crystal form: the classical conformation with the imidazole ring pointing at Ser171 (195) and a conformation in which the imidazole points away from the active site forming hydrogen bonds with Tyr88 and the main chain carbonyl oxygen of Tyr50.

previously in ETA and SGEP to bind the Glu substrate, Thr166 and His186. A third hydrogen bond is formed with the α-NH_2 group of the valine that forms the very N-terminus of the active enzyme. There is no significant rearrangement of the residues in the S1 pocket after accommodation of MPD. However, the water molecule that bridges Thr166 and His186 in the free enzyme is absent in the enzyme−MPD complex, and the hydrogen bond between those residues is broken.

Another difference between the two crystal forms is a change in the active site, which illustrates the sensitivity of the hydrogen bonding network connecting the substrate binding pocket to the active site. The association of a small polar compound like MPD with the specificity pocket initiates a conformational change in His47 (57) so that in the BIEP−MPD complex it has a double conformation. One conformation is similar to the one in the free enzyme, where His47 (57) points toward the catalytic Ser171 (195). In a second, unproductive conformation, the hydrogen bond between the Nϵ2 atom of His47 and the Oγ atom of Ser171 is broken. The imidazole ring of His47 is flipped, forming a new hydrogen bond with the hydroxyl group of Tyr88 (2.77 Å) and with the main chain carbonyl oxygen of Tyr50 (3.14 Å, Figure 5). A water molecule bound to Ser171 (195) in the free enzyme is shifted closer to the Oγ atom (2.6 Å instead of 3.2 Å). The productive conformation of the His47 side chain is prevalent with an average atomic displacement parameter (ADP) of 10 Å^2 ; the unproductive conformation has an average ADP of 19 Å^2 given equal occupancy. MPD is a precipitating agent that is routinely used to crystallize proteins. It is a small amphiphilic molecule that can bind in grooves and cavities on the protein surface, sometimes in a substrate binding site (*43*). The enzymatic activity of BIEP in the presence of 15% MPD was tested, and no inhibitory effect was observed. The presence of MPD may have decreased the dielectric constant locally, which affected both crystal packing and the hydrogen bond between His47 and Ser171.

The substrate binding region and the specificity pocket in BIEP were further verified by superposition of the active site residues of the BIEP and SGEP molecules. The SGEP structure was determined in a complex with the inhibitor tetrapeptide Boc-Ala-Ala-Pro-Glu. The examination of the

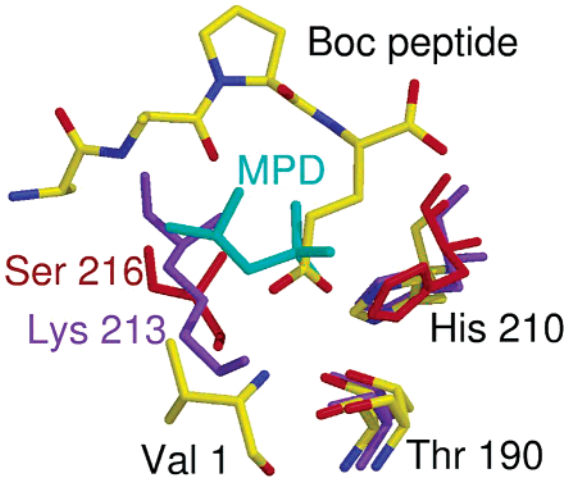

FIGURE 6: Superimposition of the substrate binding pocket for three Glu specific endopeptidases, BIEP, ETA (violet), and SGEP (crimson). The Boc tetrapeptide of SGEP is displayed, as well as the MPD molecule (cyan) found in the *C*2 crystal form of BIEP. The two conserved residues, Thr190 and His210, are numbered according to the SGEP nomenclature. The third residue involved in binding to the carboxy group of glutamic acid is different between enzymes, but occupies a similar position.

environment of the tetrapeptide fitted into the BIEP structure confirms the positions of amino acid residues involved in the binding of the substrate (Figure 6). The position of the MPD molecule in complex with BIEP is comparable to the glutamate of the peptide in the SGEP structure. In comparison to SGEP, the substrate binding region in BIEP is more open and accessible to solvent. The same region in SGEP is in part covered by the loop spanning residues 170−176 (SGEP numbering) and contains the bulky side chain of Tyr171. This loop is absent in BIEP.

The main chain of the modeled tetrapeptide is stretched along the crevice between the domains, aligning along the main chain segment of residues 187−189 of the N-terminal domain. Several amino acid residues of the protein can interact directly with the tetrapeptide. The main chain of residues P1 (Glu) and P2 (Ala) of the tetrapeptide form hydrogen bonds with main chain Asn187 (Ser 214) and Gly189 (Ser216) of the protein, respectively. It seems that Asn187, which is structurally equivalent to Ser214, participates in the binding of the substrate. Since only main chain atoms of residue 214 are directly involved in substrate binding through hydrogen bond formation, the chemical nature of the side chain of residue 214 may not be critical. However, in a viral serine protease from Sindbis core virus, where Ser214 is replaced with a leucine, the enzymatic activity is quite low (*44*). Possibly, the large hydrophobic residue in position 214 hinders substrate binding. On the other hand, the replacement of Ser214 with Ala in trypsin increases the enzymatic activity (*44*). This confirms the influence of residue 214 on the enzyme activity. It is unclear whether mutations in this position affect substrate binding, or catalytic activity, since this residue in most proteases also forms a hydrogen bond with catalytic residue Asp97.

S1 Site in BIEP. The Glu side chain of the tetrapeptide that was modeled by superimposition of the SGEP structure is located close to the polypeptide chain residues 183−189 (210−216), 196−199 (226−229), and 163−169 (184−193). The side chains of Thr166 (192), His186 (213), and the N-terminal Val (Val1) are close to the carboxylate group of the side chain of the Glu residue. All these residues (Val1, Thr166, and His186) are connected by hydrogen bonds. The α-carboxyl group of the P1 Glu residue is situated close to the enzyme's catalytic site. One oxygen of the α-COOH group is directed toward the oxyanion hole and is close to Oγ of catalytic Ser171 (195); another oxygen is within hydrogen bonding distance of the Nϵ2 atom of His47 (57). There is also a hydrogen bond connecting the substrate binding pocket and the active site, between His186 and Asp97.

While His186 and Thr/Ser166 are invariant in the specificity loop of all Glu-endopeptidases, the BIEP structure reveals a strikingly elegant mechanism for zymogen activation. Since the amino group of Val1 is protonated under physiological conditions, it is able to contribute to the compensation of the negative substrate charge. Moreover, it is likely that the substrate binding pocket can only be formed upon cleavage of the Lys−Val bond in the proenzyme. Compared to SGEP, BIEP has an extended N-terminal segment of 17 residues which allows Val1 to reach the S1 binding pocket. A similar elongation with a high degree of sequence similarity is observed in BLEP (*6*) (Figure 1), and it is likely that this enzyme is able to include the N-terminal group in the S1 site for substrate charge compensation.

Comparison of the S1 Site in Three Microbial Glu-Endopeptidases. The mechanism of substrate recognition and substrate charge compensation responsible for the highly restricted specificity of Glu-endopeptidases is a point of particular interest. According to sequence alignment, the specificity loop is the most conserved part in the structure of Glu-endopeptidases (*19*). The enzymes from different sources (*Bacillus*, *Streptomyces*, *Staphylococcus*, ChTL serine, and cysteine proteases from viruses) have two conserved residues (His and Thr/Ser) in the specificity loop. It was shown by site-directed mutagenesis in SGEP, that His213 (186 in BIEP) is required for enzyme activity, while Thr166 can sometimes be replaced with serine (*45*). The structural conservation in this area indicates that the enzymes of this family bind the incoming polypeptide in a similar manner.

We compared the structure of the S1 site, which accommodates the P1 glutamyl moiety, in BIEP with two other microbial Glu-endopeptidases, SGEP (*21*) and ETA (*19*, *20*). The S1 sites in all enzymes are characterized by a cluster composed of three amino acid residues that interacts directly with the carboxylate group of the side chain of the glutamic acid of the substrate. The conserved threonine (or in some cases serine) and histidine residues are oriented in a similar way in the substrate binding pocket in all three proteases (Figure 6).

The enzymes are distinguished by the nature and the position of the third residue, responsible for specific binding. This is Ser216 in SGEP, Val1 in BIEP, and Lys213 in ETA. Ser216 in SGEP and Lys193 in ETA are situated in a similar position of the substrate binding loop.

In the BIEP sequence, the backbone of Gly189 is in a position similar to the backbone of Ser216 (SGEP) and Lys213 (ETA). Like Ser216 in SGEP, Gly189 takes part in substrate binding through the main chain oxygen that forms a hydrogen bond with the backbone nitrogen of the P3 residue, but it does not interact with the P1 Glu residue. The N-terminal valine in BIEP that binds to the carboxylate group

of the side chain of the substrate is separated from the specificity loop and is not homologous to Lys213 and Ser216 in sequence. Val1 in BIEP and Lys213 in ETA both carry a positive charge on the α- and ϵ-amino group, respectively. In this respect, both enzymes resemble trypsin-like proteases whose specificity is based on the presence of a counter-charged residue in the binding pocket.

Unlike those of BIEP and ETA, all three residues involved in P1 binding in SGEP are neutral. Even His186 (His213 according to SGEP numbering) acts at physiological pH in a nonprotonated form, and the binding of the charged substrate must be accomplished by means of hydrogen bonds only (*45*).

MATERIALS AND METHODS

Enzyme Purification and Crystallization. The enzyme was isolated from cultural fluid and purified by ion exchange chromatography as described in refs *25* and *26*. Crystals of BIEP were grown from a protein solution in 0.01 M Tris-HCl buffer (pH 7.0) containing 2 mM $CaCl_2$ and 1.2 M potassium phosphate without and with 3% MPD (*27*). The enzyme was crystallized in two different forms, depending on the crystallization conditions. Crystals grown with potassium phosphate as a precipitant belonged to space group $C222_1$ and diffracted to 1.5 Å. Their unit cell parameters are as follows: $a = 59.49$ Å, $b = 85.44$ Å, and $c = 82.15$ Å. The crystals grown using a potassium phosphate solution containing 3% MPD belonged to space group $C2$ and diffracted to 1.74 Å. The unit cell parameters are as follows: $a = 61.94$ Å, $b = 55.76$ Å, $c = 60.19$ Å, and $\beta = 118.17°$.

X-ray Data Collection. X-ray data for two crystal forms of BIEP were collected on the X11 beam line at the synchrotron source at EMBL Outstation (Hamburg, Germany) using a MAR imaging plate detector. The crystals were mounted in a capillary, and data were collected at room temperature. The data collection statistics are given in Table 1. All X-ray data were integrated with DENZO (*46*) and scaled with SCALEPACK. There is one molecule in the asymmetric unit with a solvent content of 43% and a Matthews coefficient of 2.18 $Å^3$/Da.

Structure Determination and Refinement. The structure of BIEP was determined by molecular replacement using AmoRe (*29*). The crystal structure of epidermolytic toxin A (ETA) was used as a search model (*19*). The full stretch of homologous residues in the sequence alignment did not give any significant solution. Systematic deletion of secondary structure elements from the search model was performed. Only after the deletion of two loops (residues 105−120 and 138−151 following the numbering of ETA) was a marginal solution obtained with a correlation coefficient (CC) of 25.1% and an *R*-factor of 50.8% after rigid body refinement. The second, incorrect, solution had a CC of 18.0% and an *R*-factor of 52.5%.

The structure refinement was performed with REFMAC (*30*), applying a bulk solvent correction and an overall anisotropic temperature factor. Since the diffraction extended to 1.5 Å, we attempted to trace the model automatically with ARP/wARP version 5.1 (*31*). However, only 14 residues (7% of the total) were traced. This clearly indicated that the molecular replacement phases were too biased toward the search model, and there was a substantial structural difference between ETA and BIEP. SIGMAA was used to obtain figures of merits (FOMs) for the calculated structure factors of the molecular replacement model. To improve the FOMs, density modification was performed on the starting model using DM in the multiresolution mode over the whole resolution range. On the basis of the FOMs that DM produced after histogram matching and solvent flattening, a FoFOM map was calculated. The resulting FoFOM map was used to build a starting model for ARP/wARP. The map was automatically parametrized by free atoms that were placed according to the density criteria of ARP/wARP until the number of atoms was roughly equivalent to the expected number of atoms for the final structure. A random shuffle was applied to the free atom model to create four separate models. Unrestrained refinement with REFMAC and optimization of the free atom model with ARP/wARP were performed in three overall cycles. Free atoms within 1.1 and 1.8 Å of existing atoms were added if density criteria were fulfilled. After each cycle, the coordinates of the free atoms were randomly shaken and spurious atoms removed.

The calculated structure factors from the optimized free atom models were combined with the initial MR model and the hybrid model containing the stretch of 14 residues. The combined structure factors were weighted so that the ones with more similar contributions were given a higher figure of merit. An electron density map calculated with the set of weighted structure factors was used as input for automated tracing in ARP/wARP, and 210 of 215 residues were traced. The side chains were docked using the program option side-dock. XtalView (*47*) was used for inspection of the model and manual adjustments. The refinement converged at an *R*-factor of 15.6% ($R_{free} = 18.4\%$), and statistics are given in Table 1. The coordinates and structure factors have been deposited in the Protein Data Bank as entries 1P3C and 1P3E.

REFERENCES

1. Barbosa, J. A., Saldanha, J. W., and Garratt, R. C. (1996) *Protein Eng. 9*, 591−601.
2. Breddam, K., and Meldal, M. (1992) *Eur. J. Biochem. 206*, 103−107.
3. Drapeau, G. R. (1978) *Can. J. Biochem. 56*, 534−544.
4. Yoshida, N., Tsuruyama, S., Nagata, K., Hirayama, K., Noda, K., and Makisumi, S. (1988) *J. Biochem. 104*, 451−456.
5. Wellink, J., and van Kammen, A. (1988) *Arch. Virol. 98*, 1−26.
6. Svendsen, I., and Breddam, K. (1992) *Eur. J. Biochem. 204*, 165−171.
7. Marrack, P., and Kappler, J. (1990) *Science 248*, 1066.
8. Klier, A., Msadek, T., and Rapoport, G. (1992) *Annu. Rev. Microbiol. 46*, 429−459.
9. Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M., and Koonin, E. V. (1989) *FEBS Lett. 243*, 103−114.
10. Parks, G. D., and Palmenberg, A. C. (1987) *J. Virol. 61*, 3680−3687.
11. Kakudo, S., Kikuchi, N., Kitadokoro, K., Fujiwara, T., Nakamura, E., Okamoto, H., Shin, M., Tamaki, M., Teraoka, H., Tsuzuki, H., and Yoshida, N. (1992) *J. Biol. Chem. 267*, 23782−23788.
12. Ishizaki, J., Tamaki, M., Shin, M., Tsuzuki, H., Yoshikawa, K., Teraoka, H., and Yoshida, N. (1992) *Appl. Microbiol. Biotechnol. 36*, 483−486.
13. De Filippis, V., and Fontana, A. (1990) *Int. J. Pept. Protein Res. 35*, 219−227.
14. Seetharam, R., and Acharya, A. S. (1986) *J. Cell. Biochem. 30*, 87−99.
15. Dancer, S. J., Garratt, R., Saldanha, J., Jhoti, H., and Evans, R. (1990) *FEBS Lett. 268*, 129−132.
16. Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A.,

McElroy, H. E., Gribskov, C. L., and Worland, S. (1994) *Cell 77*, 761−771.
17. Bergmann, E. M., Mosimann, S. C., Chernaia, M. M., Malcolm, B. A., and James, M. N. (1997) *J. Virol. 71*, 2436−2448.
18. Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S., and James, M. N. (1997) *J. Mol. Biol. 273*, 1032−1047.
19. Cavarelli, J., Prevost, G., Bourguet, W., Moulinier, L., Chevrier, B., Delagoutte, B., Bilwes, A., Mourey, L., Rifai, S., Piemont, Y., and Moras, D. (1997) *Structure 5*, 813−824.
20. Vath, G. M., Earhart, C. A., Rago, J. V., Kim, M. H., Bohach, G. A., Schlievert, P. M., and Ohlendorf, D. H. (1997) *Biochemistry 36*, 1559−1566.
21. Nienaber, V. L., Breddam, K., and Birktoft, J. J. (1993) *Biochemistry 32*, 11469−11475.
22. Barrette-Ng, I. H., Ng, K. K., Mark, B. L., Van Aken, D., Cherney, M. M., Garen, C., Kolodenko, Y., Gorbalenya, A. E., Snijder, E. J., and James, M. N. (2002) *J. Biol. Chem. 277*, 39960−39966.
23. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., and Hilgenfeld, R. (2002) *EMBO J. 21*, 3213−3224.
24. Gabdrakhmanova, L. A., Shakirov, E. V., Balaban, N. P., Sharipova, M. R., Rudenskaya, G. N., and Leshchinskaya, I. B. (1999) *Microbios 100*, 97−108.
25. Leshchinskaya, I. B., Shakirov, E. V., Itskovitch, E. L., Balaban, N. P., Mardanova, A. M., Sharipova, M. R., Blagova, E. V., Levdikov, V. M., Kuranova, I. P., Rudenskaya, G. N., and Stepanov, V. M. (1997) *Biochemistry (Moscow) 62*, 903−908.
26. Leshchinskaya, I. B., Shakirov, E. V., Itskovitch, E. L., Balaban, N. P., Mardanova, A. M., Sharipova, M. R., Viryasov, M. B., Rudenskaya, G. N., and Stepanov, V. M. (1997) *FEBS Lett. 404*, 241−244.
27. Kuranova, I. P., Blagova, E. V., Levdikov, V. M., Rudenskaya, G. N., Balaban, N. P., and Shakirov, E. V. (1999) *J. Cryst. Growth 196*, 313−318.
28. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.
29. Navaza, J. (2001) *Acta Crystallogr. D57*, 1367−1372.
30. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D53*, 240−255.
31. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol. 6*, 458−463.
32. Gilbert, D., Westhead, D., Viksna, J., and Thornton, J. (2001) *Comput. Chem. 26*, 23−30.
33. Pignol, D., Gaboriaud, C., Michon, T., Kerfelec, B., Chapus, C., and Fontecilla-Camps, J. C. (1994) *EMBO J. 13*, 1763−1771.
34. Petersen, J. F., Cherney, M. M., Liebig, H. D., Skern, T., Kuechler, E., and James, M. N. (1999) *EMBO J. 18*, 5463−5475.
35. Holm, L., and Sander, C. (1998) *Proteins 33*, 88−96.
36. Johnson, A., Gautham, N., and Pattabhi, V. (1999) *Biochim. Biophys. Acta 1435*, 7−21.
37. Iversen, L. F., Kastrup, J. S., Bjorn, S. E., Rasmussen, P. B., Wiberg, F. C., Flodgaard, H. J., and Larsen, I. K. (1997) *Nat. Struct. Biol. 4*, 265−268.
38. Bode, W., and Huber, R. (1986) in *Molecular and Cellular Basis of Digestion* (Desnuelle, P., Sjostrom, H., and Noren, O., Eds.) Elsevier Science, Amsterdam.
39. Warshel, A., Naray-Szabo, G., Sussman, F., and Hwang, J. K. (1989) *Biochemistry 28*, 3629−3637.
40. Frey, P. A., Whitt, S. A., and Tobin, J. B. (1994) *Science 264*, 1927−1930.
41. Meyer, E., Cole, G., Radhakrishnan, R., and Epp, O. (1988) *Acta Crystallogr. B44* (Part 1), 26−38.
42. Rypniewski, W. R., Ostergaard, P. R., Norregaard-Madsen, M., Dauter, M., and Wilson, K. S. (2001) *Acta Crystallogr. D57*, 8−19.
43. Anand, K., Pal, D., and Hilgenfeld, R. (2002) *Acta Crystallogr. D58*, 1722−1728.
44. Tong, L., Wengler, G., and Rossmann, M. G. (1993) *J. Mol. Biol. 230*, 228−247.
45. Stennicke, H. R., Birktoft, J. J., and Breddam, K. (1996) *Protein Sci. 5*, 2266−2275.
46. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W., and Sweet, R. M., Eds.) Academic Press, San Diego.
47. McRee, D. E. (1999) *J. Struct. Biol. 125*, 156−165.
48. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
49. Meritt, E. A., and Bacon, D. J. (1997) in *Methods in Enzymology* (Carter, C. W., and Sweet, R. M., Eds.) pp 505−524, Academic Press, San Diego.

BI035354S