The Crystallographic Structure of the Subtilisin Protease from *Penicillium cyclopium*[†,‡]

Stanley Koszelak, Joseph D. Ng, John Day, Tzu Ping Ko, Aaron Greenwood, and Alexander McPherson*

Department of Biochemistry, University of California, Riverside, Riverside, California 92521

Received December 30, 1996; Revised Manuscript Received March 14, 1997[⊗]

ABSTRACT: The major extracellular protease from the fungus *Pencillium cyclopium* was crystallized in the presence of *p*-phenylmethanesulfonyl fluoride (PMSF) and investigated by X-ray diffraction analysis. It was subsequently cloned and the amino acid sequence deduced from its cDNA. Although the sequence is only 49% identical to that of proteinase K of *Tritirachium album*, the three-dimensional structures of the two proteases are virtually identical. The model for *P. cyclopium* protease was refined by simulated annealing to an *R* of 18% at 1.7 Å resolution. The greatest variation from the proteinase K polypeptide is in loop 114−134 and is due to the absence of a disulfide bridge in the *P. cyclopium* protease that is present in proteinase K. A difference was also observed in the orientation of the histidine in the catalytic triad, though this could be due to the presence of PMSF at the active site. The coordination geometry of the strongly bound calcium in the *P. cyclopium* protease is octahedral and uses some different protein ligands than does proteinase K. In the protease from *P. cyclopium* there is no cysteine near the active site, nor is there a second calcium binding site as is found in proteinase K, suggesting that neither is important to catalytic activity.

The major extracellular protein of *Penicillium cyclopium* was purified and identified as a serine protease by this laboratory in 1986 (*1*). It has a molecular weight of 31 000, exhibits autoproteolysis, shows no evidence of Ca^{2+} activation, and demonstrates broad substrate specificity. Until now, its amino acid sequence was not known, nor was its relationship to other proteases defined. It is of some practical interest because *P. cyclopium*, and closely related fungal species, are major causes of soft fruit rot and produce significant post-harvest losses.

In the course of this analysis it became clear that *P. cyclopium* protease was of the subtilisin class and that it bore a striking resemblance to proteinase K from the fungus *Tritirachium album* (*2*, *3*). When, later, the amino acid sequence was obtained, this was further confirmed. Subtlisin proteases comprise one of the two largest classes of serine proteases, the other being those similar to trypsin (*4*). Both classes employ a catalytic triad of amino acids consisting of a serine, a histidine, and an aspartic acid disposed in a consistently similar manner (*5*). Polypeptide structural features of the two protease classes are otherwise very different. The subtilisin class now includes at least 80 known members from many organisms, though the precise three-dimensional structures of only a few, subtilisin (*5*), thermitase (*6*), and proteinase K (*3*), are available.

METHODS

Preparation and Crystallization. Protease from *P. cyclopium* was prepared from culture filtrates, purified, and crystallized by vapor diffusion from 15% PEG 3350 for X-ray diffraction as described previously (*1*).

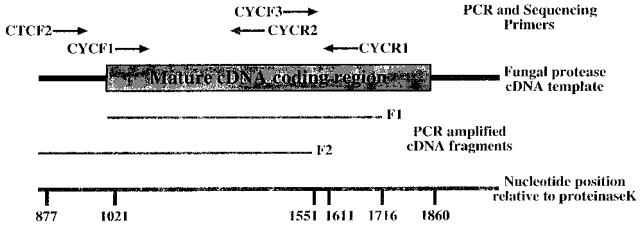

FIGURE 1: PCR cloning and sequencing strategy for the *P. cyclopium* protease. Oligonucleotide primers were designed to selectively amplify cDNA fragments containing the mature coding region of the protease. Primers CYCF1 and CYCR2 were used to amplify a 671 bp cDNA fragment (Fl) from which the direct sequence of its product was used to design the reverse primer CYCR2 and the forward primer CYCF3. CYCF2 is an oligonucleotide homologous to the proteinase K 5′ noncoding region and was used as a coupling primer to CYCR2 for the amplification of a 674 bp fragment (F2) containing the N-terminal sequence of the mature protease. Primer CYCF3 was used as a direct sequencing primer to obtain information in the C-terminal region. The *P. cyclopium* protease cDNA template is outlined and aligned to the relative nucleotide position of proteinase K.

Cloning and Sequence Determination. The cloning and sequencing strategy for *P. cyclopium* protease is illustrated in Figure 1. Partial cDNA fragments of the fungal protease were synthesized by the polymerase chain reaction (PCR) (*7*). Total RNA was isolated from fresh mycelium by methods described by Kirby (*8*) and Chirgwin *et al.* (*9*). ssDNA was generated from total RNA by reverse transcription (*10*), and single-stranded cDNA products were used as templates for PCR reactions. Reaction conditions included the standard PCR buffer (*7*), 2 units of Vent DNA polymerase purchased from New England BioLabs Inc., and 40 pmol of primer in 100 μL layered with an equal volume of mineral oil. PCR was carried out using an Ericomp DNA thermal cycler for 1 min 30 s at 96 °C initial denaturation,

[†] This work was supported by a grant from the National Aeronautics and Space Administration.

[‡] Coordinates for *P. cyclopium* protease have been deposited in the Brookhaven Data Bank (file name 4TST).

* Corresponding author.

[⊗] Abstract published in *Advance ACS Abstracts,* May 15, 1997.

S0006-2960(96)03189-3 CCC: $14.00

followed by 30 cycles for 30 s at 65 °C, 2 min at 72 °C, and 15 s at 96 °C. A final extension was carried out for 2 min at 55 °C and 2 min at 72 °C. The following oligonucleotide primers were used to selectively amplify cDNA segments coding only for the mature protease: 5′-GG(C/T)CTTTCTCG-CATCTCCAGC-3′ (CYCF1), 5′-CTTGACGAGAACATG-GTTCGGGTTCTCCGC-3′ (CYCF2), 5′-ACCGGCAACGT-G(A/G/T) GGAGTAGCCAT-3′ (CYCR1), 5′-TACATTC-GGAGCAGAAGCAGGGGAG-3′ (CYCR2), and 5′-CGT-TGTCGACCTCTA TGCTCCTGGGAAGGAC-3′ (CYCF3).

Based on structural similarities apparent in electron density maps, it was reasonable to assume that the *P. cyclopium* protease sequence would be highly homologous to that of proteinase K (*11*). Therefore, primers were designed to contain sequences that matched regions in the proteinase K gene which were highly conserved when compared with closely related proteases. An N-terminal portion of the mature *P. cyclopium* protease was determined by direct sequencing of a cyanogen bromide fragment. Subsequently, degenerate oligonucleotides were synthesized using the most commonly used codon for each amino acid found in the N-terminal region of the protease. Primers CYCF1 and CYCR1 are coupled primers used in the amplification of a 671 bp cDNA fragment (F1) containing the majority of the coding region. Similarly, primer CYCF2 is coupled to CYCR2 in the PCR amplification of a 674 bp fragment (F2) containing the sequence of the N-terminus of the mature protease. The sequence of CYCR2 was determined from direct sequencing of fragment F1. All oligonucleotide primers were synthesized using phosphoramidite chemistry at the Biotechnology Instrumentation Facility at UCR or purchased directly from Cruachem Inc., Sterling, FL.

The amino acid sequence of the mature protease was deduced from partial length cDNAs that contained the corresponding coding region. Purified PCR product either was directly sequenced or was subcloned into a sequencing vector for analysis. Direct sequencing was performed using Sanger's dideoxy method (*12*) with a Sequenase kit, version 2.0, from United States Biochemical Co. according to directions supplied by the manufacturer. Nucleotide and protein sequence analysis were performed using the Wisconsin Package (*13*) version 8.0 run on a DEC VAX computer.

Structure Determination by Isomorphous Replacement. Multiple isomorphous replacement (MIR) structure determination was carried out using X-ray diffraction data collected on a CAD-4 automated diffractometer, and the data were processed and scaled in a conventional manner using procedures previously described (*14*). For refinement of the structure, native data of high redundancy were collected for the protease crystals to 1.7 Å resolution using a San Diego Multiwire Systems two-detector system (*15*, *16*). The X-ray generator was a Rigaku RU-200 fitted with a Supper monochromator.

Formation of heavy atom derivatives of the protease crystals was problematic. The most common experiences were loss of diffraction pattern, non-isomorphism, or no substitution. Eventually, however, heavy atom derivatives were prepared and refined (*17*). Statistics describing their phasing value are found in Table 1. Even with these derivatives, non-isomorphism continued to be a serious problem and limited the use of MIR phasing to about 3.2 Å resolution.

Table 1: Isomorphous Replacement Statistics

data set	resol limit (Å)	no. of reflections	completeness (%)	R_{sym} (%)	R(der-nat) (%)	R_c (%)	rms (F)/ rms (ϵ)	X	Y	Z	B ($Å^2$)	relative occupancy	nearest aa [distance (A)]
two crystals													
PMSF inhibited (native)	1.6	35803	100	6.5				0.356	0.057	0.754	15	1.16	Thr 262 (6.9)−Asn 273 (7.4)−Asn 108 (6.9)
Na_2WO_4	3.4	3936	100		34	70	1.23	0.301	0.047	0.715	12	0.93	*a*
								0.298	0.080	0.683	1	0.87	*a*
								0.378	0.092	0.700	12	1.03	*a*
								0.314	0.118	0.739	17	1.00	*a*
								0.026	0.185	0.421	15	0.25	Val 152 (2.1)−Met 133 (3.3)−Val 157 (2.9)−Val 119 (3.1)
								0.124	0.473	0.229	2	0.17	Asp 64 (1.2)−Asn 61 (1.9)
$PtBr_2(NH_3)_2$ (set 1)	3.9	2134	83	4.2	26	51	1.27	0.074	0.052	0.389	22	1.55	Met 135 (1.2)−Met 115 (25)
								0.736	0.118	0.561	68	1.62	Arg 55 (1.0)
								0.000	0.320	0.002	22	0.66	His 73 (2.1)
								0.367	0.017	0.834	29	0.72	Asn 108 (1.2)−Met 145 (2.4)
Pt-dibenzyl, diammonium chloride	5.0	985	79	4.3	29	54	1.85	0.069	0.054	0.391	16	1.78	Met 135 (1.3)−Met 115 (2.4)
								0.370	0.019	0.831	23	1.09	Asn 108 (1.1)−Met 145 (2.4)
K_2IrCl_6 (set 1)	3.9	2122	83	4.2	17	64	1.22	0.313	0.425	0.072	24	0.98	Asp 46 (2.3)
								0.416	0.146	0.970	28	0.38	Asn 165 (2Å)−Ser 228 (4.0)
								0.398	0.321	0.124	28	0.22	Asn 2 (2.9)
								0.287	0.445	0.733	15	0.66	Asp 67 (3.7)−Gln (5.9)−Lys 6 (4.8)
$PtBr_2(NH_3)_2$ (set 2)	3.4	3650	95	4.3	14	62	1.00	0.496	0.174	0.985	37	0.58	His 73 (2.1)
								0.736	0.113	0.561	52	0.53	Arg 55 (1.1)
K_2IrCl_6 (set 2)	3.4	3907	100	4.6	12	68	0.70	0.312	0.425	0.071	17	0.53	Asp 46 (2.4)
								0.286	0.443	0.734	23	0.31	Asp 67 (3.7)−Gln 5 (6.0)−Lys 6 (4.7)

[a] All WO4 sites were in a space between molecules and their sites in the neighborhood of that of the first site. No cell dimensions of any crystal used in Mir analysis varied by more than 0.5% from those of the native. Centric figure of merit for 771 reflections was 0.72. Noncentric figure of merit for 3249 reflections was 0.63. Overall figure of merit for 4020 reflections was 0.65.

A figure of merit weighted electron density map based on the "best" phase angles was computed (*18*). While much of the map was clear, we could not produce an unambiguous tracing for the polypeptide chain. Application of the density modification procedure of Wang (*19*) significantly improved the map, not only increasing the contrast with solvent at the margins of the molecule but substantially improving the connectivity and clarity of the internal structure as well. The resemblance of the *P. cyclopium* protease to that of proteinase K from *T. album*, whose structure had been solved in the laboratory of Saenger in Berlin (*2*), at this stage became apparent.

Application of Molecular Replacement. A model based on the proteinase K atomic coordinates from the Brookhaven Data Bank was superimposed upon the partial model for the protease from *P. cyclopium*. The α-helices and the strands of the parallel β-sheet served as reliable fiducial marks and served to align the two models with a high degree of precision. The model for proteinase K, which had been well refined by Betzel *et al.* (*3*) to a resolution of 1.5 Å and a final residual of 0.18, was by this means accurately placed upon the electron density of the *P. cyclopium* protease unit cell.

The proteinase K model was subjected to rigid body refinement in the *P. cyclopium* unit cell with the program CORELS (*20*) using data between 12 and 3.5 Å resolution. Convergence was achieved after only a few cycles with a resulting residual of $R = 0.32$. When the correct amino acid sequence for the *P. cyclopium* protease became known from cloning and was substituted for the proteinase K sequence, some significant rebuilding of the model was necessitated, and adjustments were made throughout.

Refinement of Structure. Refinement by simulated annealing using the program X-PLOR (*21*−*23*) was initiated by employing a temperature gradient of 3000 → 300 K in 25 deg increments and a time step of 0.0005 ps. $2F_o - F_c$ omit maps were computed in 10 amino acid increments and adjustments made to side chains and, where necessary, main chain amino acid groups. Refinement by simulated annealing and rebuilding were alternated until no further improvement in structure or statistics was apparent. Water molecules were added to the model and temperature factors refined only in the last cycles of refinement. Model building was carried out on an Evans and Sutherland PS 390 using the program FRODO (*24*). Additional analyses and final images were prepared for presentation using SETOR (*25*) and O (*26*) on an SGI 340 VGX.

Before refinement was begun, the data set was partitioned into a working set (90%) and a test set (10%) as described by Brunger (*27*). The test set was used to calculate the statistical quantity R_{free} which was used to monitor the course of refinement. The *R*-factor was 0.18 and the R_{free} was 0.22 for the final model which contains 280 amino acids, 110 water molecules, a Ca^{2+} ion, and one molecule of PMSF. Other relevant statistics (*27*−*30*) regarding the refined model are found in Table 2.

Table 2: Data and Refinement Statistics

space group	$P2_12_12_1$
unit cell	$a = 59.22$ Å, $b = 61.64$ Å, $c = 70.85$ Å
data completeness	27 779 unique reflections ($F > 3\sigma$) from 1.7 to 8.0 Å

resolution range (Å)	% completeness	% accumulative ($F > 3\sigma$)
3.33−8.00	75.58	75.58
2.67−3.33	99.97	87.58
2.34−2.67	99.97	91.65
2.14−2.34	99.97	93.68
1.98−2.14	99.75	94.88
1.87−1.98	98.77	95.52
1.78−1.87	98.17	95.89
1.70−1.78	97.44	96.08

refinement statistics	
resolution limits	1.7−8.0 Å
R-value[a]	18.0%
no. of unique reflections in working set ($F > 3\sigma$)	25016
free *R*-value	21.6%
no. of unique reflections in test set ($F > 3\sigma$)	2763
no. of molecules in asymmetric unit	1
no. of total non-hydrogen atoms in final model	2060
no. of solvent molecules	110
no. of calcium ions	1
average protein *B*-value	51 Å^2
rmsd in protein bond lengths	0.006 Å
rmsd in protein bond angles	1.2°

[a] $R\text{-value} = \sum_{hkl}(||F_{o_{hkl}}| - |F_{c_{hkl}}||)/\sum_{hkl}|F_{o_{hkl}}|$, where $|F_{o_{hkl}}|$ and $|F_{c_{hkl}}|$ are the observed and calculated structure factor amplitudes.

RESULTS

Sequence Comparison with Proteinase K. The amino acid sequence of *P. cyclopium* protease when compared, in Figure 2, with that of proteinase K shows 49% identity, with a high degree of homology among the remaining amino acids. The greatest variation occurs in the final carboxy-terminal amino acids where a substantial variation was detected even within clones of the *P. cyclopium* enzyme.

The *P. cyclopium* protease is one amino acid longer than proteinase K (280 vs 279), with two additional residues at the amino terminus but two fewer at the carboxy terminus. There are two insertions of amino acids in the former with respect to the latter, Ser 11 and Ala 63, and one deletion, at Gly 259 of proteinase K. Insertions and deletions occur in surface loops and produce little conformational perturbation. Residues of the catalytic triad of *P. cyclopium* were immediately recognized by homology to be Ser 228, His 73, and Asp 42. Other residues highly conserved among the subtilisin proteases (*31*, *32*), underlined in Figure 2, are present as well. Among the differences with proteinase K are that Cys 34 and Cys 123 in proteinase K are Val 37 and Ala 127 in the *P. cyclopium* protease, while Cys 178 and Cys 249 in the former are conserved in the latter (Cys 182, Cys 254). Thus, while proteinase K has two disulfide bonds, *P. cyclopium* protease has but one.

Proteinase K has a free cysteine at position 73 lying very near the active site and which has been suggested to play a role in catalysis (*3*). *P. cyclopium* protease has no corresponding cysteine, the homologous residue being threonine. Proteinase K also has a secondary, weak Ca^{2+} binding site using ligands provided by Thr 16 and Asp 260. In *P. cyclopium* protease there are no suitable ligands at the corresponding locations, and this Ca^{2+} binding site is absent.

Secondary Structure Comparison with Proteinase K. As described for proteinase K (*2*, *3*), the secondary structure of *P. cyclopium* protease, seen in Figure 3, consists of a central, seven-stranded parallel β-sheet framed above and below by long α-helices and containing, as well, several antiparallel β-ribbons near the extremities of the molecule. The Ca^{2+}

```
                  10         20         30         40         50         60
Fungal protease   ANVVQLASPSWGLSRISSKKSGATDYVYDSTAGEGIVIYGVDTGIDIGHADFGGRAETGT
                    ::| ::| |||:||||:::|:::| ||::||:|  :| :||||: :|::|:|||:  :
Proteinase K        AAQTNAP-WGLARISSTSPGTSTYYYDESAGQGSCVYVIDTGIEASHPEFEGRAQMVK
                         10         20         30         40         50

                        70         80         90        100        110        120
Fungal protease   NTADNDDTDGNGHGTHTASTAAGSKFGVAKKASVVAVKVLGADGSGTNSQVIAGMDWAVK
                  :    :: |||||||| |:|::::::|||||: : :||||:::|||  | :|||||::::
Proteinase K      TY-YYSSRDGNGHGTHCAGTVGSRTYGVAKKTQLFGVKVLDDNGSGQYSTIIAGMDFVAS
                    60         70         80         90        100        110

                       130        140        150        160        170        180
Fungal protease   DSKSRGATGKSVMNMSLGGAYSRAMNDAAANVVRSGVFLSVAAGNEAQDASNSSPASAPN
                  |:::|: :   | ::||||:||:::|:|||:: :|||:::|||||:::||:| ||||:|:
Proteinase K      DKNNRNCPKGVVASLSLGGGYSSSVNSAAARLQSSGVMVAVAAGNNNADARNYSPASEPS
                   120        130        140        150        160        170

                       190        200        210        220        230        240
Fungal protease   VCTIAASTNSDGSASFTNFGSVVDLYAPGKDITAAYPGGGSKTLSGTSMATPHVAGLASY
                  |||::|| : | ::||:|:|||:|:::||::| ::: ||:::::|||||||||||||:|
Proteinase K      VCTVGASDRYDRRSSFSNYGSVLDIFGPGTSILSTWIGGSTRSISGTSMATPHVAGLAAY
                   180        190        200        210        220        230

                       250        260        270        280
Fungal protease   LMALEKTTASAACNRIVQDATT-KIAGAPKGTTNLLAYNGI
                  ||:|:|||||::||: |:: |:: ::: | ||:||||||:
Proteinase K      LMTLGKTTAASACRYIADTANKGDLSNIPFGTVNLLAYNNYQA
                   240        250        260        270
```

FIGURE 2: Sequence comparison of the *P. cyclopium* protease with proteinase K from *T. album* Limber. The two proteins are 49% identical to each other as indicated by vertical bars in the 274 amino acid overlap. When conservative amino acid substitutions are considered (double dots), the sequences show about 88% similarity. The protein sequence analysis was performed using the FASTA Pearson and Lipman search for homologous sequences in the Wisconsin Package version 8.0 (*13*) run on a VAX alpha computer.

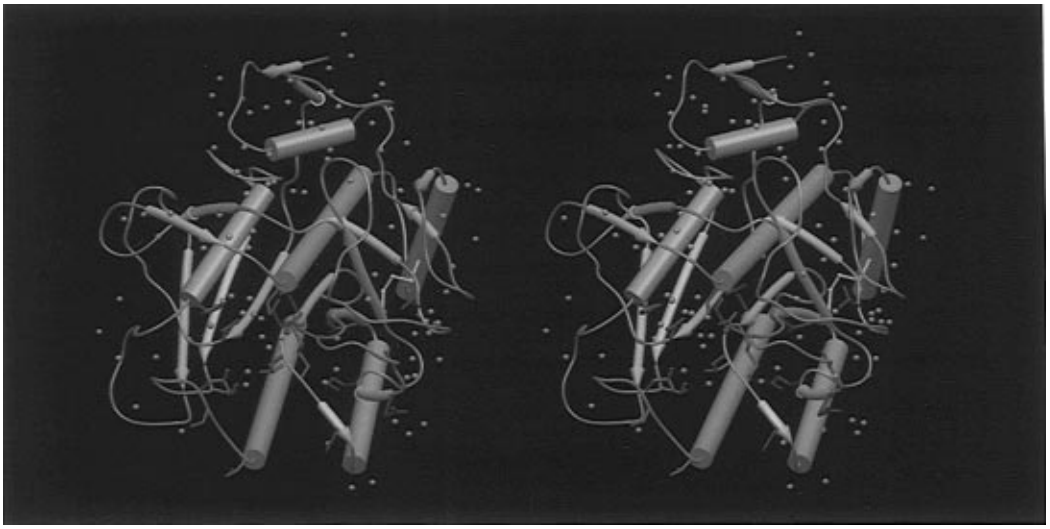

FIGURE 3: *P. cyclopium* protease structure along with the 110 water molecules, shown as white spheres, included in the model. Proteinase K has 178 waters associated with it in its crystal structure. Helices are in green and sheet strands in yellow. The catalytic triad is shown in red along with Asn 165, which binds the water molecule, the red sphere, present in the oxyanion hole. Analysis of over 90 subtilisin-like proteases reveals that 12 amino acid residues are almost perfectly conserved and these, which also include the catalytic residues in red, are shown in purple. The conserved residues are not randomly distributed over the protein but are restricted to roughly a quarter of the molecule.

binding site and the active site of the enzyme lie on opposite sides of the protein. The catalytic triad is solvent accessible but is found at the bottom of one end of a rather long cleft across a face of the molecule.

As seen in Figure 4, major secondary structural elements of the two proteins are virtually superimposable, particularly the parallel β-sheet and helices comprising the core of the molecule. The rms difference in C_α coordinates between homologous residues is 0.85 Å. The only significant deviations are in the connecting loops, particularly that comprised of residues 114−134. Other notable variations occur in loops containing amino acids 179, 246, and 265.

Proteinase K has two unusual helices in that one, α4, contains a proline (228), and another, α2, contains a small extra loop (amino acids 79−83) before completing its final turn. *P. cyclopium* protease contains a proline (232) embedded in helix α4 as well. *P. cyclopium* protease also contains the small extra loop, though different in sequence, near the end of helix α2. Again, this allows conservation of the structure observed in proteinase K. Proline 171 in proteinase

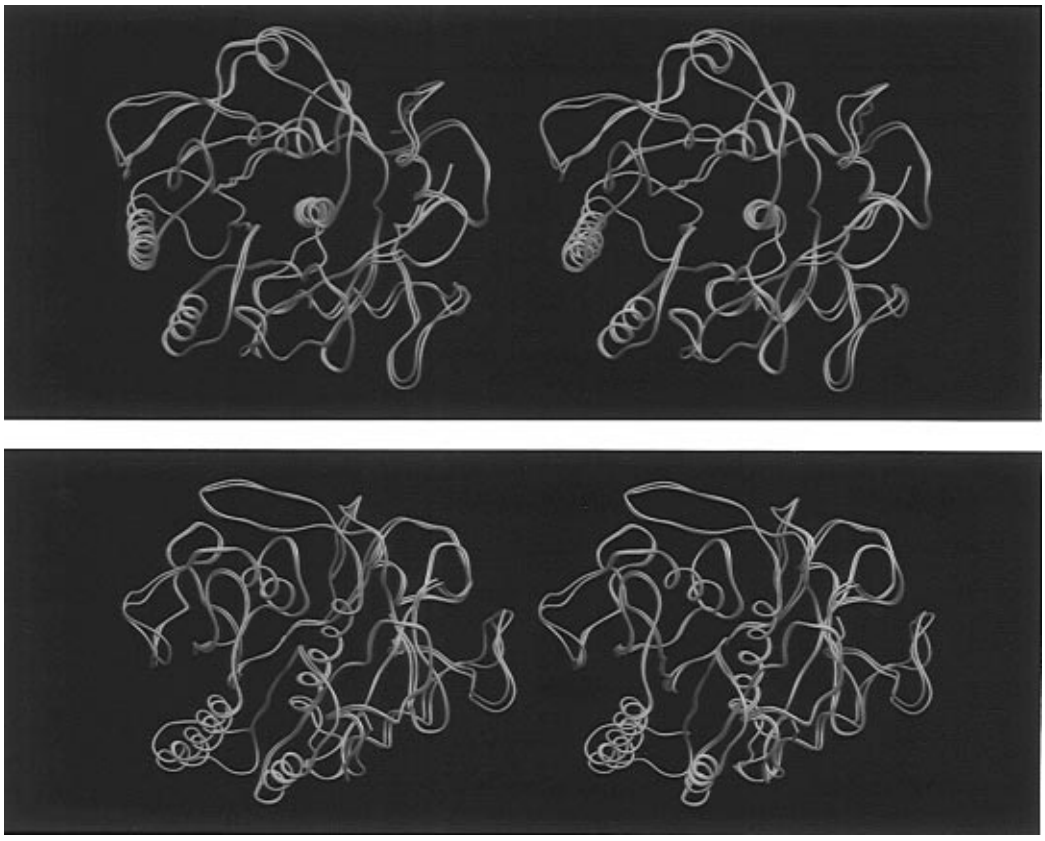

FIGURE 4: C_α backbone tracing of *P. cyclopium* protease shown in yellow in two different orientations superimposed on the backbone of proteinase K. Although there is only a 49% amino acid identity between the two proteins, the degree of congruence is exceptional.

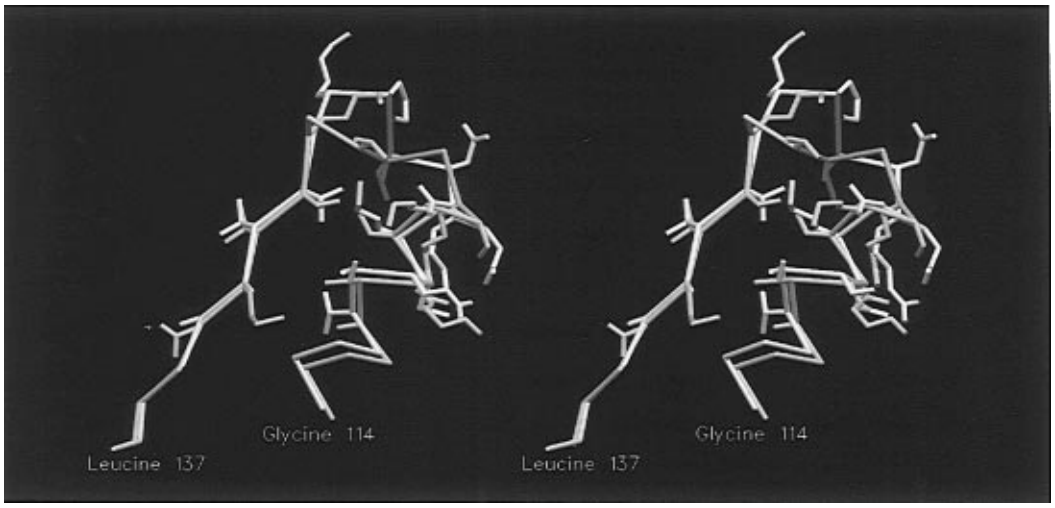

FIGURE 5: Disulfide bridge between Cys 34 and Cys 123 (shown in red) in proteinase K replaced in *P. cyclopium* protease by Ala 127 (shown in green). A consequence of removing this constraint is the largest variation in relative dispositions of polypeptide chains between proteinase K and *P. cyclopium* protease. This involves the loop between Gly 114 and Leu 137 which assumes a substantially different conformation in the two molecules.

K was observed to have a cis conformation as, indeed, all proteases of this class so far analyzed by X-ray diffraction (*3*). *P. cyclopium* protease also has a proline (175) at the corresponding position, and it is also in the cis conformation.

A significant difference in the structures of proteinase K and *P. cyclopium* protease arises from a comparison of disulfide bridges. Both enzymes have one homologous linkage, Cys 178−Cys 249 in proteinase K and Cys 182−Cys 254 in *P. cyclopium* protease. The second disulfide in proteinase K, between Cys 34 and Cys 123, would have corresponding residues Val 37 and Ala 127 in *P. cyclopium* protease; hence no second disulfide is possible. The Cys 34−Cys 123 disulfide in proteinase K was described as an anchor to link helix α3 and strand βII4 (*3*). A consequence of its removal is shown in Figure 5. The loop in *P. cyclopium* protease containing residues 114−134 and, corresponding to strand βII4 in proteinase K, has assumed a substantially different conformation and has swung out away from the end of helix α3. This divergence of loop conformations represents the greatest difference in backbone conformations seen anywhere in the molecule.

The Active Site. Native crystals for this analysis were grown from protein exposed to an excess of PMSF during preparation and crystallization to prevent autoproteolysis.

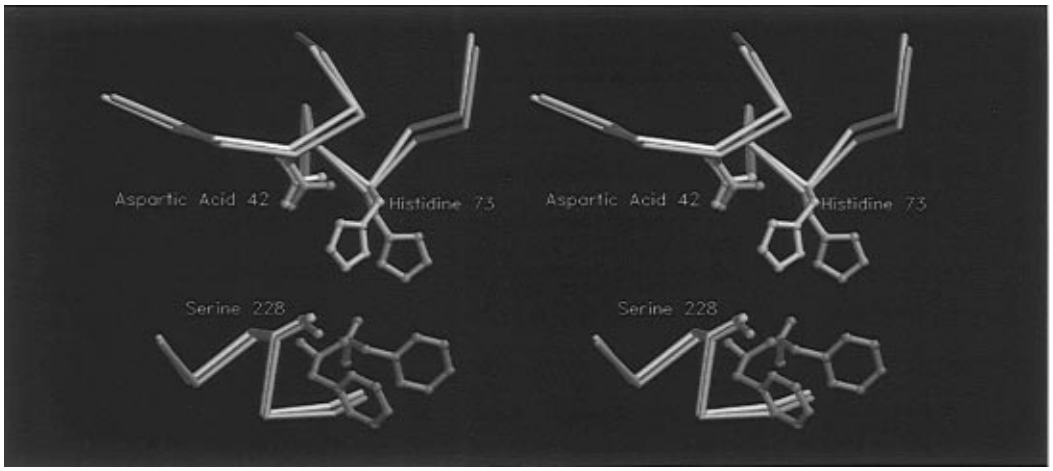

FIGURE 6: Alternate dispositions of the PMSF molecule bound at the active site of *P. cyclopium* protease. For that in blue the benzene ring is arbitrarily placed, as it was not visible in the electron density map. The orientation of the PMSF in green occupies a portion of the substrate binding site. The C_α backbones bearing the catalytic side chains of Ser 228, His 73, and Asp 42 for *P. cyclopium* protease (yellow) and the corresponding residues for proteinase K (white) are shown superimposed. The superposition of Ser 228 and Asp 42 side chains is, within experimental error, exact. The imidazole ring of His 73, however, is rotated dramatically in *P. cyclopium* protease compared to that observed in proteinase K.

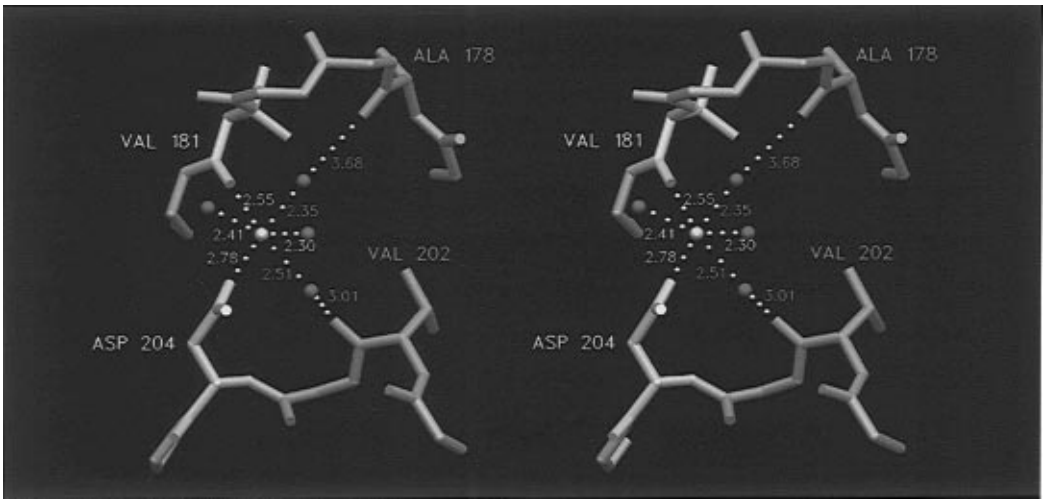

FIGURE 7: Octahedrally coordinated Ca^{2+} ion shown here as a yellow sphere with the two segments of the polypeptide chain which provide ligands. Three water molecules, seen as red spheres, and the carbonyl oxygen of Val 181 form the equatorial plane. The carboxyl of Asp 204 is at one apical position. The sixth position is occupied by a water molecule, also seen as a red sphere, hydrogen bonded to the carbonyl oxygen of Ala 178.

Indeed, no cleavage of the polypeptide chain was evident. We expected to observe a single PMSF molecule at the active site but found a somewhat more complicated situation. The hydroxyl oxygen of active site Ser 228 lies immediately adjacent to a dense peak which we assume to be the sulfur of a bound PMSF molecule. The benzene ring of this PMSF is, however, apparently disordered as we see no density to indicate its location. A second, bilobal peak of electron density having the proper characteristics of a PMSF molecule is also close by, but its sulfur, again a dense peak, lies about 4 Å from the refined position of the serine hydroxyl. Rotation about the $C_\alpha-C_\beta$ bond, however, brings the serine hydroxyl oxygen within covalent bonding distance of this second sulfur as well. While the second PMSF could be reversibly bound by secondary interactions at the active site, we interpret the two sulfur positions to likely indicate alternate positions of a single covalently bound PMSF. The electron density is most consistent with two orientations of a PMSF, marked by the sulfur peaks, in one case ordered by the substrate binding site and in the second case disordered in the active site cleft. These alternate positions are shown in Figure 6.

Even though the hydroxyl of Ser 228 is bonded to a PMSF sulfur, there appears to be no distortion of the main chain and almost none for the serine side chain from the conformation observed in proteinase K. The ordered PMSF lies in a groove formed by segments Leu 137-Gly 138-Gly 139 and Gly 164-Asn 165-Glu 166 and has beneath it Ser 173 and Asn 198.

The catalytic triads of both proteinase K (*3*, *33*) and *P. cyclopium* protease are shown superimposed in Figure 6. There is no significant difference in the local conformations of the respective polypeptide chains bearing the catalytic residues. In addition, the active site Ser 228 and Asp 42 side chains of *P. cyclopium* protease are congruent with the corresponding side chains in proteinase K (Ser 224 and Asp 39). The major difference is in the orientation of the side chain of His 73, which has assumed a position corresponding to approximately a 90° rotation about the $C_\alpha-C_\beta$ bond. Whether the change in orientation is a consequence of the

PMSF bound to Ser 228 we cannot be certain. We observe that *P. cyclopium* protease, as for proteinase K, and indeed most other serine proteases, also contains a water molecule in the "oxyanion hole", in turn hydrogen bonded to Asn 165.

The Calcium Binding Site. Coordination of the calcium ion in *P. cyclopium* protease is not the same as described for proteinase K although several ligands are the same. As illustrated in Figure 7, three water molecules and the carbonyl oxygen of Val 181 form a square planar, equatorial array around the Ca^{2+}. In proteinase K the homologous Val 177 carbonyl was also an equatorial ligand. The carbonyl oxygen of Val 202 is situated in such a way that it can hydrogen bond to water molecules in the equatorial plane. Apical ligands are the carboxyl group of Asp 204, identical in function to the carboxyl of Asp 200 in proteinase K, and a fourth water molecule. In proteinase K, this apical ligand was provided directly by the carbonyl oxygen of Pro 175.

The equatorial plane of the proteinase K Ca^{2+} binding site contained five ligands, and the coordination geometry was a pentagonal bipyramid (*3*, *34*). In *P. cyclopium* protease the Ca^{2+} is octahedrally coordinated with several, perhaps all, of the water ligands hydrogen bonding to protein oxygen atoms. The waters thus bridge between the Ca^{2+} and the protein. The second apical ligand is not the carbonyl oxygen of Pro 177 as seen in proteinase K but a water bridging the Ca^{2+} to the carbonyl oxygen of Ala 178. We examined our electron density maps for alternatives that might allow us to utilize the carbonyl oxygen of Pro 177, but the density was clear. The proline was rotated in such a manner that this was simply not possible for *P. cyclopium* protease.

CONCLUSIONS

The protease from *P. cyclopium* and proteinase K have 49% sequence identity and otherwise a high degree of homology, in spite of the substantial phylogenetic difference of the two fungi which produce them (*35*). The polypeptide backbones are virtually superimposable with an rms C_α difference overall of 0.85 Å. The only significant variation appears in the loop between residues 114 and 134, and this is a consequence of the absence in *P. cyclopium* protease of a disulfide bridge that is present in proteinase K.

Over 90 proteases of the subtilisin class have been characterized in terms of their amino acid sequences. Examination of these sequences, extending from bacteria and fungi through higher plants and animals (*31*, *32*), reveals that there are 12 amino acid residues that are completely conserved. Among them are those comprising the catalytic triad. In addition, however, others seen in Figure 3 suggest crucial structural or physiological functions.

Val 99, Leu 100, Thr 107, Asn 134, Ser 136, Gly 138, and Ala 140 are residues that are located at or adjacent to the active site. Val 157 is situated opposite the active site between the end of a β-sheet and the beginning of an external helix. Ala l70, Asn 172, and Pro 175 are external residues located across from another external conserved residue, Val 203. The amino acid segment Gly 70−Ala 82, which corresponds to a helix adjacent to a core helix, also has very high similarity when compared to other known proteases.

The structures of proteinase K and *P. cyclopium* protease are most similar in the active site region. Proteinase K has a substrate recognition site that includes residues Gly 100−Ser 101 and Ser 132−Gly 134 (*3*, *33*). *P. cyclopium* has identically those same residues at the corresponding sites and in the same orientations. Both proteases have very low substrate specificities, and it seems reasonable to assume that they are essentially the same.

The polypeptide atoms of the catalytic triad are superimposable for proteinase K and *P. cyclopium* protease, as are the side chains of the serine and aspartic acid. This is in spite of the covalent attachment of a PMSF molecule to Ser 228 of *P. cyclopium* protease, a modification which appears to produce no substantial perturbation. The side chain of His 73 has adopted a significantly different orientation, however, rotating about the C_α-C_β bond by about 90°. Similar alternate conformations of this histidine have been observed in the catalytic triads of both trypsin-like and subtilisin-like proteases complexed with inhibitors (*36*, *37*, and R. Bott, personal communication). The oxyanion hole present in all serine proteases so far analyzed is occupied by a water molecule hydrogen bonded to Asn 165. Unlike proteinase K, and several other subtilisin like proteases, there is no cysteine residue adjacent to the histidine at the active site, being threonine in *P. cyclopium* protease. Thus we conclude that no sulfhydryl group plays any significant role in the catalytic mechanism.

A significant variation we do observe in the structures of proteinase K and *P. cyclopium* protease that could suggest long-range mechanistic differences is the alternate coordination geometry of the calcium. The difference in the binding of the Ca^{2+} ion may reflect the flexibility that ion shows in the arrangement of its ligands, a flexibility seen across a range of other protein structures containing calcium ions, and within proteinase K itself, where its weakly bound calcium ion is octahedrally coordinated.

Bajorath *et al.* (*34*) noted that the strong Ca^{2+} site of proteinase K was linked through a network of hydrogen bonds to the active site 17 Å away. They further found that removal of the strongly bound Ca^{2+}, though promoting no large local structural changes, nonetheless produced subtle alterations at the substrate binding site. These in turn led to a significant reduction in enzymatic activity over a period of hours. The protease from *P. cyclopium* appears to be otherwise, perhaps as a consequence of the different Ca^{2+} coordination geometry and subsequent modifications to the associated network of hydrogen bonds. *P. cyclopium* protease does not lose activity in the presence of high concentrations of EDTA, and it crystallizes isomorphously both in its absence and presence. Thus the Ca^{2+} requirement for full activity appears to be relaxed, or perhaps it has not yet evolved.

ACKNOWLEDGMENT

The authors thank Mr. Derrick Wang and Ms. Negeen Zareh for their technical assistance and Robert Lucas for preparing the illustrations.

REFERENCES

1. Day, J., Koszelak, S., Cascio, D., and McPherson, A. (1986) Isolation, characterization and preliminary X-ray diffraction data for a serine protease from *Penicillum cyclopium*, *J. Biol. Chem. 261*, 1957−1961.
2. Pähler, A., Dattagupta, J. K., Fujiwara, T., Lindner, K., Pal, G. P., Suck, D., Weber, G., and Saenger, W. (1984) Three-dimensional structure of fungal proteinase K reveals similarity to bacterial subtilisin, *EMBO J. 3*, 1311−1314.

3. Betzel, C., Pal, G. P., and Saenger, W. (1988) Three-dimensional structure of proteinase K at 0.15-nm resolution, *Biochemistry 27*, 155−171.
4. Warshel, A. Naray-Szabo, G., Sussman, F., and Hwang, J.-K. (1989) How do serine proteases really work, *Biochemistry 28*, 3629−3637.
5. Wright, C. S., Alden, R. A., and Kraut, J. (1969) Structure of subtilisin BPN′ at 2.5 Å resolution, *Nature 221*, 235−242.
6. Teplyakov, A V., Kuranova, I. P., Harutyunyan, E. H., and Vainshtein, B. K. (1990) Crystal structure of thermitase at 1.4 Å resolution, *J. Mol. Biol. 214*, 261−279.
7. Mullis, K. B., Fallona, F. A., Charf, S. J., Saikli, R. K., Horn, G. T., and Erlich, H. A. (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction, *Cold Spring Harbor Symp. Quant. Biol. 51*, 263−273.
8. Kirby K. S. (1968) Isolation of nucleic acids with phenolic solvents, *Methods Enzymol. 12B*, 87−89.
9. Chirgwin, J. M., Przbyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease, *Biochemistry 18*, 5294−5299.
10. Verma, I. M. (1981) Reverse transcriptase, in *The Enzymes* (Boyer, P. D., Ed.) 3rd ed., Vol. 14, p 87, Academic Press, New York.
11. Gunkel, F. A., and Gassen, H. G. (l989) Proteinase K from *Tritirachium album* Limber. Characterization of the chromosomal gene and expression of the cDNA in *Escherichia coli*, *Eur. J. Biochem. 179*, 85−94.
12. Sanger, F., Nicklen, S., and Coulson, A. R. (l977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci U.S.A.* 74, 5463−5467.
13. Pearson, W. R., and Lipman, D. J. (1988) Improved tools for biological sequence comparison, *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444−2448.
14. McPherson, A., Brayer, G. D., and Morrison, R. D. (1986) Crystal structure of RNAse A complexed with $d(pA)_4$, *J. Mol. Biol. 189*, 305−327.
15. Xuong, N.-H., Nielson, C., Hamlin, R., and Anderson, D. (1985) Strategy for data collection from protein crystals using a multiwire counter detector diffractometer, *J. Appl. Crystallogr. 18*, 342−360.
16. Hamlin, R., Cork, C., Howard, A., Nielson, C., Vernon, W., Matthews, D., and Xuong, N.-H. (1981) Characteristics of a flat multiwire area detector for protein crystallography, *J. Appl. Crystallogr. 14*, 85−89.
17. Terwilliger, T., and Eisenberg, D. (1983) Unbiased three dimensional refinement of heavy atoms parameters by correlation of origin removed Patterson functions, *Acta Crystallogr., Sect. A 39*, 813−817.
18. Ten Eyck, L., Weaver, L. A., and Matthews, B. W. (1976) A method of obtaining a stereochemically acceptable protein model which fits a set of atomic coordinates, *Acta Crystallogr., Sect. A 32*, 349−355.
19. Wang, B. C. (1985) Resolution of phase ambiguity in macromolecular crystallography, *Methods Eng. 115*, 90−112.
20. Sussman, J. L. (1983) Protein model building by the use of a constrained-restrained least squares procedure, *J. Appl. Crystallogr. 16*, 144−150.
21. Brunger, A. T., Kuriyan, J., and Karplus, M. (1987) Crystallographic R factor refinement by molecular dynamics, *Science 235*, 458−460.
22. Brunger, A. T. (1988) Crystallographic refinement by simulated annealing; application to a 2.8 Å resolution structure of aspartate aminotransferase, *J. Mol. Biol. 203*, 803−816.
23. Brunger, A. T. (1991) Simulated annealing in crystallography, *Annu. Rev. Phys. Chem. 42*, 197−223.
24. Jones, T. A. (1985) Interactive computer graphics: FRODO, in *Methods in Enzymology* (Wyckoff, H. W., Hirs, C. H. W., and Timasheff, S. N., Eds.) Vol. 115, Chapter 12, pp 157−171, Academic Press, New York.
25. Evans, S. V. (1993) SETOR: Hardware lighted three dimensional solid model representations of macromolecules, *J. Mol. Graphics 11*, 134−138.
26. Jones, T. A., and Kjeldgaard, M. (1994) *O—The Manual*, Uppsala University Press, Uppsala, Sweden.
27. Brunger, A. T. (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472−474.
28. Ramachandran, G. N., and Sassiekharan, V. (1968) Conformation of polypeptides and proteins, *Adv. Protein Chem. 23*, 283−437.
29. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates, *Proteins 12*, 345−364.
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1992) *PROCHECK v.2. Programs to check the stereochemical quality of protein structures*, Oxford Molecular Ltd., Oxford, England.
31. Siezen, R. J. (1992) Subtilases: Subtilisin-like serine proteases—Update 1992, in *Proceedings of the International Symposium on Subtilisin Enzymes* (Betzel, C., Ed.) Sept, Hamberg, Germany.
32. Siezen, R. J., de Vos, W. M., Leunissen, J. A. M., and Dijkstra, B. W. (1991) Homology modeling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases, *Protein Eng. 4*, 719−737.
33. Betzel, C., Pal, G. P., Struck, M., Jany, K.-D., and Saenger, W. (1986) Active-site geometry of proteinase K. Crystallographic study of its complex with a dipeptide chloromethyl ketone inhibitor, *FEBS Lett. 197*, 105−110.
34. Bajorath, J., Raghunathan, S., Hinrichs, W., and Saenger, W. (1989) Long-range structural changes in proteinase K triggered by calcium ion removal, *Nature 337*, 481−484.
35. Ainsworth, G. C., Sparrow, F. K., and Sussman, A. S. (1973) *The Fungi, An Advanced Threatise. Volume IVA—a Taxonomic Review w/kigs: Ascomycetes & Fungi Omperfecti*, Academic Press, New York and London.
36. Kossiakoff, A. A. (1987) *Catalytic properties of Trypsin, from Biological Macromolecules and Assemblies* (Jurnak, F. A., and McPherson, A., Eds.) Vol. 3, Chapter 7, p 399, John Wiley, New York.
37. Sprang, S. R., Fletterick, R. J., Gráf, L., Rutter, W. J., and Craik, C. S. (1988) Studies of specificity and catalysis in trypsin by structural analysis of site-directed mutants, *Crit. Rev. Biotechnol. 8*, 225−236.

BI963189T