Because the fluid holding the crystal becomes glass-hard at the low temperature, slippage of the crystal is completely eliminated. Redetermination of orientation matrices at the end of data collection revealed no observable change for the two test cases described above.

It is difficult to predict the effect of lower thermal motion on the process of structure solution. Normally a better data set leads to easier solution, but for some macromolecules it is conceivable that the greater visibility of the solvent in some way may interfere with the interpretation of electron density maps.

The observed variation in the cell dimensions of crambin point to a potential advantage in obtaining all data from one specimen. Crystals with slightly different cell dimensions will have slightly different structures, perhaps expressed as a difference in solvent arrangement. The averaging of data from different crystals can then result in smearing of electron densities in addition to that caused by thermal motion or disorder.

Dewan & Tilton (1987) have carried out low-temperature studies of ribonuclease, by use of a mounting technique first described by Hope & Power (1983). The crystals did not suffer phase separation, and virtual elimination of radiation damage was observed. Crystals of the same protein, mounted in coventional glass capillaries, were also successfully cooled; however, in this case the reduction in radiation damage was less pronounced. Although it would be of interest to determine the cause for this difference, it is not likely to be of immediate practical importance, since capillary mounting can be avoided. Available data do not appear to support the assumption of Dewan & Tilton that most protein crystals can be successfully cooled only after being transferred to a cryosolvent.

Further experiments in biological cryocrystallography are under way. Results will be described in forthcoming communications.

I thank Dr M. Teeter for supplying a sample of crambin and for instruction on the preparation of single crystals, and Dr W. Hendrickson for a supply of BPTI crystals.

References

DEWAN, J. C. & TILTON, R. F. (1987). *J. Appl. Cryst.* **20**, 130–132.
DREW, H. R., SAMSON, S. & DICKERSON, R. E. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4040–4044.
HARTMANN, H., PARAK, F., STEIGEMANN, W., PETSKO, G. A., RINGE, D., PONZI, D. R. & FRAUENFELDER, H. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4967–4971.
HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107–113.
HOPE, H. (1985). Proc. Am. Crystallogr. Assoc. Meet. Abstr. Vol. 13, p. 24.
HOPE, H. & NICHOLS, B. G. (1981). *Acta Cryst.* B**37**, 158–161.
HOPE, H. & POWER, P. P. (1983). *J. Am. Chem. Soc.* **105**, 5320–5324.
LOW, B. W., CHEN, C. C. H., BERGER, J. E., SINGMAN, L. & PLETCHER, J. F. (1966). *Prov. Natl Acad. Sci. USA*, **56**, 1746–1750.
PARAK, F., MÖSSBAUER, R. L., HOPPE, W., THOMANEK, U. F. & BADE, D. (1976). *J. Phys. (Paris) Colloq.* C6, **37**, 703–706.
PETSKO, G. A. (1975). *J. Mol. Biol.* **96**, 381–392.
TEETER, M. M. & HOPE, H. (1986). *Ann. N. Y. Acad. Sci.* **482**, 163–165.
WLODAWER, A., WALTER, J., HUBER, R. & SJÖLIN, L. (1984). *J. Mol. Biol.* **180**, 301–329.

Acta Cryst. (1988). B**44**, 26–38

Structure of Native Porcine Pancreatic Elastase at 1·65 Å Resolution*

BY EDGAR MEYER, GREG COLE AND R. RADHAKRISHNAN

Biographics Laboratory, Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843–2128, *USA*

AND OTTO EPP

*Strukturforschung, Max-Planck-Institut für Biochemie, D–*8033, *Martinsried, Federal Republic of Germany*

(*Received* 22 *September* 1986; *accepted* 24 *July* 1987)

Abstract

The structure of native porcine pancreatic elastase in 70% methanol has been refined using film data to 1·65 Å resolution, $R = 0{\cdot}169$. A total of 134 molecules of water (but no methanol) has been refined. This structure, because of its native state and modestly high resolution, serves as the basis for comparison with other elastase structures complexed with natural or synthetic ligands. Internal structured water occupies distinct regions. Two regions (*IW*1 and *IW*7) suggest a mechanism for equalizing 'hydrostatic pressure' related

* Presented at the IUCr Meeting 'Molecular Structure: Chemical Reactivity and Biological Activity', Beijing, China, September 1986.

0108-7681/88/010026-13$03.00

to ligand binding and release. A third region (*IW4*) forms part of a hydrogen-bonding network linking the catalytic Ser 195 Oγ with a remote (13·4 Å) surface of the enzyme. A comparison with the structures of all known serine proteases reveals that a linkage of Ser Oγ to remote surface is conserved in all cases, suggesting that the accepted catalytic mechanism of serine proteases needs to be re-evaluated. One possible mechanism for base catalysis of Ser OγH proton extraction is presented.

Introduction

The elastases (E.C. 3.4.21.11) are a subclass of serine proteases that possess the characteristic ability to degrade the desmosine-crosslinked protein, elastin, that forms the basis of highly flexible connective tissue, *e.g.*, in lung or arterial-wall tissues. The crystal structure of porcine pancreatic elastase (PPE) was reported to 2·5 Å resolution (Sawyer *et al.*, 1978) and is here reported to 1·65 Å resolution. In addition, the structure of human leucocyte (neutrophil, polymorphonuclear) elastase (HLE) has recently been reported to 1·7 Å resolution (Bode, Wei, Huber, Meyer, Travis & Neumann, 1986).

In the digestive process, the secreated, activated pancreatic proteases are controlled by proteolysis. Serum or leucocyte-released elastases are usually controlled by an abundant supply of α-1-antiprotease (Johnson & Travis, 1978). However, when control is absent, disease states result: pancreatitis, emphysema, acute adult respiratory distress syndrome, some forms of arthritis, and certain degenerative tissue disorders.

Because of the current interest in the design of drugs to inhibit elastase's degradation of human connective tissue, the insight afforded by a high-resolution view of the extended binding site of these enzymes, including conserved hydration, must be a crucial component in the design of inhibitors with both tight binding and high specificity. In order to obtain a clearer understanding of the relationship of enzyme structure and function, we initiated the study of the structure of PPE at the highest attainable resolution. The structure of the native enzyme provides a basis for comparison with inhibited PPE structures of comparable resolution (Meyer, Presta & Radhakrishnan, 1985; Meyer, Radhakrishnan, Cole & Presta, 1986; Meyer, Clore, Gronenborn & Hansen, 1987; Radhakrishnan, Presta, Meyer & Wildonger, 1987; Takahashi, Radhakrishnan, Rosenfield, Meyer & Trainor, 1987; Takahashi, Radhakrishnan, Rosenfield & Meyer, 1987) as well as with HLE (Bode, Wei *et al.*, 1986).

The serine proteases are a class of enzymes recognized to have conserved structural homology about the amino acids His 57, Asp 102 and Ser 195 [chymotrypsinogen (Freer, Kraut, Robertus, Wright & Xuong, 1970) numbering system]. We here propose that the equally conserved and structurally linked Ser 214 be included to comprise a catalytic *tetrad.* Support for this suggestion will be provided below. A charge-relay mechanism (Blow, Birktoft & Hartley, 1969) was proposed to facilitate H-atom extraction (general base catalysis) *via* oxyanion (Ser 195) attack on an amide or ester carbonyl C atom. Discrete steps in the micro-reversibility of catalysis have been discussed by Bizzozero & Dutler (1981). The proteases have recently been reviewed by Bode & Huber (1986). The most relevant new structure, for comparison purposes, is HLE complexed to turkey ovomucoid 3rd domain (Bode, Wei *et al.*, 1986).

There is only limited experimental evidence available to define the sequence of events in catalysis: His 57 can be deactivated by chloromethyl ketones (Robertus, Alden, Birktoft, Kraut, Powers & Wilcox, 1972), thus inactivating the enzyme. A neutron diffraction study of trypsin inhibited with a monoisopropylphosphoryl group (Kossiakoff & Spencer, 1981) shows the imidazole ring of His 57 to be doubly protonated, suggesting that the proton need go no further. Yet, when Asp 102 is converted by site-specific mutagenesis to Asn 102 (*a*) reactivity falls by 10^{-4} (Sprang *et al.*, 1987), (*b*) His and Ser are still chemically active, and (*c*) reactivity is closely correlated to pH, returning to 15% activity at pH 10·0. Both NMR (Steitz & Schulman, 1982) and crystallographic studies of the charge-relay system (Blow *et al.*, 1969) establish the catalytic role of Ser 195. The consensus may be reduced to agreeing that while further experimental evidence is needed, it is not easily available, due to the rapid turnover rates of the serine proteases under physiological conditions. This study discusses the homologous nature of the catalytic tetrad and the internal conserved water molecules, suggesting that additional chemical or physical evidence should be sought to probe the various structural states essential for catalysis.

The eucaryotic and procaryotic (trypsin-like) and procaryotic (subtilisin-like) serine proteases may be compared both structurally and functionally. They differ in their individual amino-acid sequences and functionally in their substrate specificity and catalytic rates. The structural homology seen by superimposing two or more Ser–His–Asp catalytic triads is striking. This homology is extended to a tetrad by including Ser 214 hydrogen bonded (2·76 Å) to Asp 102; this Ser 214 residue is also conserved in all mammalian and bacterial trypsin-like serine proteases (Marquart, Walter, Deisenhofer, Bode & Huber, 1983). In subtilisin, the Ser 221–His 64–Asp 32 triad is correspondingly hydrogen bonded (Bode, Papamokos, Musil, Seemueller & Fritz, 1986) to Thr 33 to form a similar tetrad. In all cases thus far studied, the tetrad is linked to surface contacts *via* hydrogen bonding to one (subtilisin-like),

two (trypsin-like), or three (elastase-like) water molecules.

Methods

Pure PPE was obtained from Serva (20909), Heidelberg, and crystallized without further purification by vapor diffusion using the Zeppezauer method [*cf.* Meyer, Radhakrishnan *et al.* (1986) for a more detailed description of crystal growth]. Crystals were then converted to a cryosolvent, 70% methanol, 30% acetate, 0·1*M*, pH 5·0, by vapor diffusion or by dialysis. The cryobuffer environment with a high organic component was chosen in order to facilitate subsequent studies involving relatively insoluble inhibitors. The resulting crystals were approximately isomorphous with those used in previous studies (Sawyer *et al.*, 1978; Meyer, Radhakrishnan *et al.*, 1986): $a = 52{\cdot}1$, $b = 58{\cdot}1$, $c = 75{\cdot}2$ Å, $P2_12_12_1$. The a axis of crystals in 70% methanol buffer was approximately 3% longer than that reported by Sawyer *et al.* (1978) for elastase crystals in aqueous buffer (50·8 Å).

Crystals were mounted in thin-walled glass capillaries containing excess cryobuffer. Using crystals of approximate size 0·3 × 0·3 × 0·7 mm, intensity data to 1·65 Å resolution were collected on a modified Nonius precession camera (Delft, Holland), using the rotation mode with a film-to-crystal distance of 47 mm and Cu $K\alpha$ radiation (pyrolytic graphite monochromator) from a Rigaku Denki rotating-anode generator operating at 3·6 kW and a 0·5 mm diameter collimator. Osray T4 X-ray film was used to record all reflections. Data were taken at 293 K.

Rotation photographs from four crystals, covering 3° wedges about the c axis, completed the asymmetric unit corresponding to a total rotation of 90°, with appropriate overlaps. The last crystal was then mounted about the b axis and a supplementary 20° rotation range was measured. The films were digitized using an Optronics rotating densitometer and evaluated using program *FILME* (Schwager, Bartels & Jones, 1975). Intensities for about 60 000 reflections were found to be greater than *FILME*'s 1·0σ significance level (partial reflections were not included). Structure factors for 19 232 unique reflections were obtained by scaling and merging these data using the *PROTEIN* program system (Steigemann, 1974). The R_{merge} value for all merged data, defined as $\{\sum[(I_{hi} - \langle I\rangle_h)^2]/\sum(I_{hi})^2\}^{1/2}$, was 0·093 (ranging from 0·063 to 0·126), I_{hi} being the intensity value of an individual measurement and $\langle I\rangle_h$ the corresponding mean value; the summation is over all measurements common to two or more films. The completeness ratio, which is the ratio of the number of measured reflections (above 1σ) to the number of possible reflections, was 76% to 1·78 Å and 66% to 1·65 Å resolution. No corrections for absorption or decay were included. An independent analysis of effective resolution (Swanson, 1988) gives a value of 1·88 Å.

Refinement of atomic coordinates and temperature factors

For refinement, the *PROTEIN* system (Steigemann, 1974) and Deisenhofer's program *EREF* (Levitt, 1974; Jack & Levitt, 1978; Deisenhofer, Remington & Steigemann, 1985) were used to calculate the difference Fourier map, extract the derivative matrix and calculate weighted shifts to positional or thermal parameters. The program *EREF* minimizes the function $F = E + kX$, where E is the potential energy [using parameters from Levitt (1974)] and X is the crystallographic residual $\sum(|F_{\text{obs}}| - |F_{\text{calc}}|)^2$; k is the weight given to the crystallographic term relative to the energy term. By increasing the weighting term k from 1×10^{-4} in steps of 5×10^{-5} it was possible first to stress and then gradually to relax the geometrical restraints in favor of the crystallographic term. The refinement methods used here are comparable to other analyses (Deisenhofer *et al.*, 1985; Wlodawer, Deisenhofer & Huber, 1987).

Using the previously published coordinates of PPE (Sawyer *et al.*, 1978) and these data in the range 5·0–2·5 Å resolution, an initial R factor of 0·323 was obtained; this value was reduced to 0·272 by one cycle of refinement of positional parameters. The dramatic decrease in the R factor is not surprising when one remembers that the coordinate shifts during this cycle were compensating for the 3% increase in the length of the a axis. Seven cycles of *EREF* refinement, using data from 5·0 to 2·5 Å resolution, interspersed by two manual fittings of the model to the density map using the computer-graphics program *FIT* (Morimoto & Meyer, 1976; Meyer, Cole, Presta, Rosenfield & Swanson, 1982), together with an addition of 52 solvent molecules, reduced the R factor to 0·22.

Careful evaluation of the $2F_o - F_c$ map, using the computer-graphics program *FRODO* (Jones, 1978), revealed 22 additional solvent molecules. Data from 7·0 to 2·0 Å resolution were used for further refinement. When the R factor dropped to 0·20 we felt enough confidence in our data to include the rest of the high-resolution data to 1·65 Å resolution. Reflections were omitted from the refinement if the rejection ratio $2(|F_{\text{obs}}| - |F_{\text{calc}}|)/(|F_{\text{obs}}| + |F_{\text{calc}}|)$ exceeded 1·2. Further refinement caused the R factor to improve to 0·190. With three additional cycles of refinement of the coordinates and one cycle of temperature-factor refinement the R factor decreased to 0·188 and 37 more solvent molecules were located. During the initial stages of refinement, the temperature factors of the side-chain and backbone atoms were averaged and applied to each residue. The minimum and maximum temperature factors were set to 6·0 and 40·0 Å^2. During the final stages of refinement the temperature factors of the

Table 1. *Final parameters for PPE*

Final R factor	0·169
No. of reflections (7·0 to 1·65 Å)	17993
Overall temperature factor (Å^2)	14·1
Standard deviation for bond length (Å)	0·018
Standard deviation for bond angle (°)	2·5
Min., max. cutoff for temperature factors (Å^2)	4·0, 45·0
Min., max. electron density in the final ΔF map (e Å^{-3})	−0·40, 0·38
Estimated (Luzatti, 1952) positional error (Å)	0·2
Effective resolution (Swanson, 1988) (Å)	1·88

individual atoms were allowed to vary: the minimum and maximum were set to 4·0 and 45·0 Å^2. Experimental parameters and results are summarized in Table 1.*

Results

Comparison with PPE at 2·5 Å resolution

Sawyer *et al.* (1978) described in considerable detail the structural environment of PPE to 2·5 Å resolution (their Table 7). We especially wish to note the following changes in definition of residues that previously were not clearly defined by electron density, but which now, thanks to the benefits of refinement with higher-resolution data, help us define specific residues. We contoured the final $2F_o - F_c$ map at $0{\cdot}66\sigma$ (0·38 e Å^{-3}). The maximum density is 3·3 e Å^{-3}.

For Ser 36C, Ser 37, Ser 170A and Ser 170B, density was present at the above level for all but the Oγ atoms. Density was missing for Arg 107 after Cγ and likewise for Arg 125 beyond Cγ. The guanidinium group is without density for Arg 61, which is near the active site but entirely 'out in solution'. Likewise, Arg 223 on the terminal α-helix is missing density for the NεH1 and NεH2 atoms. For Gln 150 and 153, there is a break in density at the above level between Cγ and Cδ atoms. We find clearly defined density for Asn 72, Thr 96 and Gln 110, all of which have refined properly, although the side chains of the latter residue had high thermal parameters.

After refinement at higher resolution, specific regions of the molecule were defined more precisely. More molecules of hydration were located and refined with confidence. Likewise, several residues formerly located (Sawyer *et al.*, 1978) in ill-defined density were placed into crisply defined density. By studying the enzyme in its native state, a view of the relaxed receptor geometry and especially the loci of the catalytic tetrad was obtained, undistorted by the presence of ligand binding.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 3EST, R3ESTSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centers at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37020 (as microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Relation of temperature factors to structure

The variation of temperature factors was plotted with respect to the sequential order of the residues (Fig. 1). At each position along the abscissa a line is drawn to represent the value of the average temperature factor of the side-chain atoms. The average temperature factor of the backbone atoms of each residue is represented by the thicker continuous line. At the lower end of each line, the one letter code of each residue is also given and may be compared with the historical residue number. Residues comprising α-helices and β-sheets are marked,

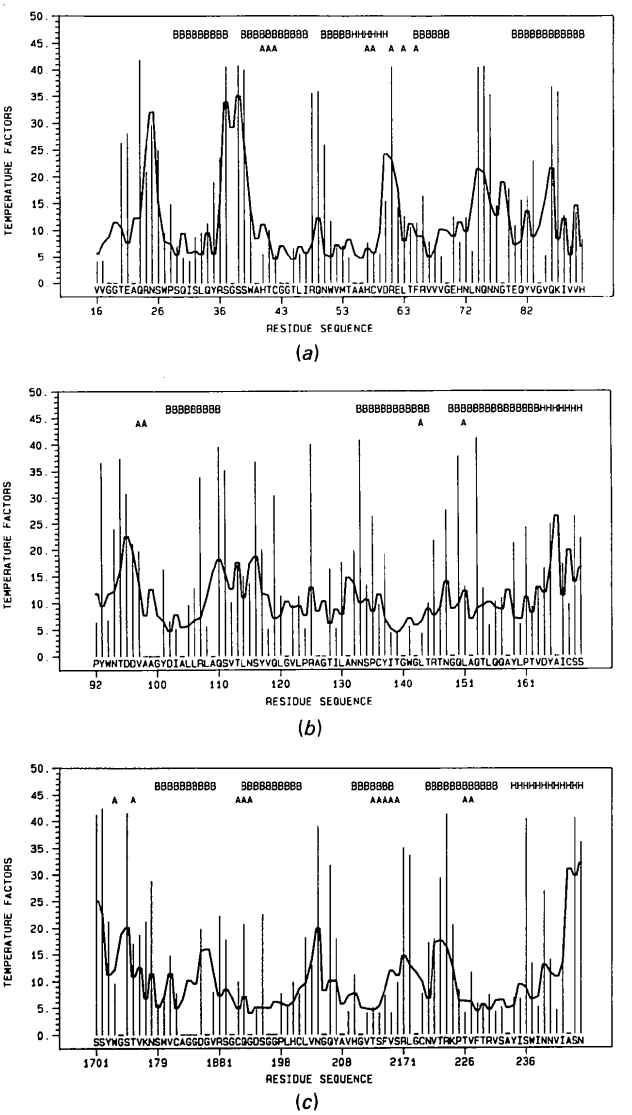

Fig. 1. Plot of isotropically refined temperature factors [B(Å^2)], side-chain (bar) and backbone (continuous plot) atoms as a function of sequence number (according to chymotrypsinogen) with amino acids indicated by the appropriate letter. Above the plot, secondary structure is indicated: H for helix, B for β structure. Amino acids involved with substrate binding ($S4$ to $S2'$ subsites, Table 2) are marked with an A.

Table 2. *List of residues in PPE comprising subsites of the extended binding site (Presta, 1985)*

Subsite notation	Enzyme residues
S5	Trp 172, Thr 175, Ser 217, Arg 217A
S4	Val 99, Ala 99A, Trp 172, Thr 175, Phe 215, Arg 217A
S3	Gln 192, Val 216, Ser 217
S2	His 57, Val 99, Gln 192, Phe 215
S1	Cys 191, Phe 192, Gly 193, Ser 195, Thr 213, Ser 214, Phe 215, Val 216, Thr 226, Val 227
S1'	Thr 41, Cys 42–S–S–Cys 58, Arg 61, Leu 63, Phe 65
S2'	His 40, Phe 41, Leu 143, Leu 151, Gln 192, Gly 193
S3'	Tyr 35, Ser 36A, Ala 39, His 40, Thr 41, Leu 63

respectively, with *H* or *B*. Active site and subsite [*S4* to *S2'*: notation of Schechter & Berger (1967)] residues are marked with an '*A*'; these subsites are summarized in Table 2 (Presta, 1985). Fig. 1 shows that one residue, Arg 61, has an exceptionally high temperature factor. Owing to the flexibility afforded by five torsion angles, it can provide a hydrophilic or charged anchor to the *S1'* subsite (leaving-group region), thus offering potential interactions thus far not utilized in drug-design strategies.

General structural features

The secondary and tertiary structure of mammalian serine proteases including elastase may be summarized as follows: two antiparallel β-barrel cylindrical domains 'left' and 'right' (displaced by *ca* 60° about an axis perpendicular to the barrel vectors) form the primary rib cage of the molecule (Fig. 2). The extended binding site and catalytic active site are found in the crevice between the two domains. The two domains are held together by the crossings of the termini of the polypeptide chain. There is no disulfide bridge connecting these two domains. The amino terminus (Val 16) crosses over to the 'left' domain to form a salt bridge with the atom Oδ2 of Asp 194. Sawyer *et al.* (1978) give an excellent discussion of the secondary and tertiary structure of PPE at lower resolution. A detailed discussion of the secondary structure of PPE is provided by Cole (1986).

Intramolecular hydrogen bonding

Hydrogen bonds were assigned on the basis of the distance between a hydrogen-bond donor–acceptor pair being $< 3{\cdot}2$ Å and visualization on the graphics display revealing a geometry appropriate to hydrogen bonding. A number of hydrogen bonds are reported to external water molecules which were not observed in the earlier structure determination. In some cases, these additional molecules of hydration contribute to the network of internal water channels described below.

Because there are no large, exposed hydrophobic patches on the contact surface of elastase, intermolecular hydrogen bonds and salt bridges are principally responsible for stability of the elastase crystal. The arrangement of polar and charged residues is summarized in Table 3.

Sulfate ions

There are two sulfate ions bound to the elastase molecule. The relative insolubility of the sulfate ion in bulk 70% methanol buffer may be a contributing factor in facilitating binding at these loci. One ion is on the exterior of the protein, hydrogen bonded to the guanidinium groups of Arg 230 (2·70 Å) and Arg 145 (3·10 Å) (from a symmetry-related molecule), and to three other water molecules (Sol 534, Sol 579, Sol 582). Sawyer *et al.* (1978) also report this sulfate; however, they do not report the surrounding water molecules. Although one of the O atoms of the sulfate group is poorly resolved, the tetrahedral shape, the volume and the density precluded that this be treated as acetate or any other chemically plausible ion. This sulfate ion is directly on the α-helix axis situated at the carboxyl terminus of the protein.

The other sulfate ion is in the active-site region making hydrogen bonds with Ser Oγ (2·8 Å), Gly 193 N (2·6 Å) and to a water molecule (Sol 571, 2·8 Å). One of the O atoms of the sulfate is at a distance of 3·0 Å from His 57 Nε, suggesting a weak hydrogen bond. As the charge on the sulfate ion is negative, we

Fig. 2. Stereoview of α-carbon backbone of PPE, looking towards the catalytic tetrad, with the *S* and *S'* regions of the extended binding site to the left and right, respectively. Vectors are drawn to represent the centers of the β-barrel domains. Seven water domains are drawn and labelled (*WT1–WT7*).

can presume that the His 57 is protonated, which is supported by the fact that this work was done at pH 5·0, reasonably below the pK_a of histidine. The fourth O atom makes van der Waals contacts (3·3 Å) with the Gln 192 side chain. These contacts partially characterize this sulfate ion as a 'receptor-bound ligand'.

Internal water molecules

Sawyer *et al.* (1978) reported 25 water molecules inside the elastase structure (labelled *IW*1–25) and another 23 water molecules external to the enzyme (labelled *EW*1–23). In most cases, our observations concur. We have numbered water molecules corresponding to internal water molecules reported by Sawyer *et al.* (1978) from Sol 301 (=*IW*1) to Sol 332. Correspondingly, external water molecules are numbered beginning with number Sol 401 (=*EW*1). Additional internal water molecules are numbered Sol 501–590.

The common understanding of soluble, globular proteins is that they have a hydrophobic core with a hydrophilic surface. We find that significant channels, wide enough to accept an ion or water molecule but little else, are found penetrating from the surface into the β-barrel domains. Two such water 'channels' connect the surface directly with the buried, primary specificity pocket.

*Cluster WT*1. The primary specificity (*S*1) pocket extends from the active site at the intersection of the two β-barrel regions diagonally across and down into the 'left' β-barrel region (Fig. 2). The bottom of the *S*1 pocket is fixed by Ser 189 [Asp in trypsin (Huber *et al.*, 1974) and kallikrein (Chen & Bode, 1983)]. At the back of the *S*1 site adjacent to Ser 189 and in the 'left' β-barrel, an included water domain (*WT*1) is found to be gated by Val 216, an estuary chain: Sol 301 to Sol 302 to Sol 303 and Sol 304 connects to bulk external water *via* Sol 328 and Sol 326 (Fig. 3). This cluster is partially described by Sawyer *et al.* (1978) as cluster 5. HLE (Bode, Wei *et al.*, 1986) also has a Val at 216. A closely analogous channel contains Sol 568 at the entrance, near Val 216. HLE Sol 703, nearby, sets up two branches in opposite directions: (1) Sol 567–629–560–696–624 is closely analogous to cluster *WT*1; (2) Sol 572–571 is screened from bulk solvent by flexible Arg 217 and Phe 215 (behind the anti-parallel '*S*' binding region).

Table 3. *Arrangement of charged and polar residues*

Val 16	BSB	Asp 194
Glu 21	open –	near His 71
Gln 30	buried	
Gln 34	+ : +	Arg A65
Arg 36	PSB	Asp 98
Arg 48	open +	
Gln 49	open	
His 57	BSB	Asp 102
His 57	H bond	Ser 195
Arg 61	open +	
Glu 70	SB	Asn 77
Glu 70	SB	calcium 325
His 71	open +	near Glu 21
Asn 72	open	
Asn 74	to contact	Gln 153
Gln 75	open	
Asn 77		calcium 325
Glu 80	SB to	calcium 325
Gln 86	open	
Lys 87	+ to	Asn 245
Asn 95	open	
Asp 97	PHB	Gln 110 *via* HOH 504
Asp 102	H bond to	Ser 214
Arg 125	PSB	Thr 147
Asn 133	open	
Asn 148	open	
Gln 150	packing to	Asn 178
Asp 164	open –	
Asp 186	BSB	Arg A188
Asn 204	open	
Gln 206	packing to	Thr 93
Arg A217	packing + to	Val 59: Arg 61 β bend
Arg 223	open +	
Arg 230	BSB to	sulfate 419
Asn 239	open	
Asn 245	BSB to	Arg 107
Asn 245	to +	Lys 87

Notes: SB = salt bridge, BSB = buried salt bridge, PSB = packing salt bridge to symmetrically related neighbor, PHB = packing H bond to symmetrically related neighbor, open +, – = unbalanced charged residues.

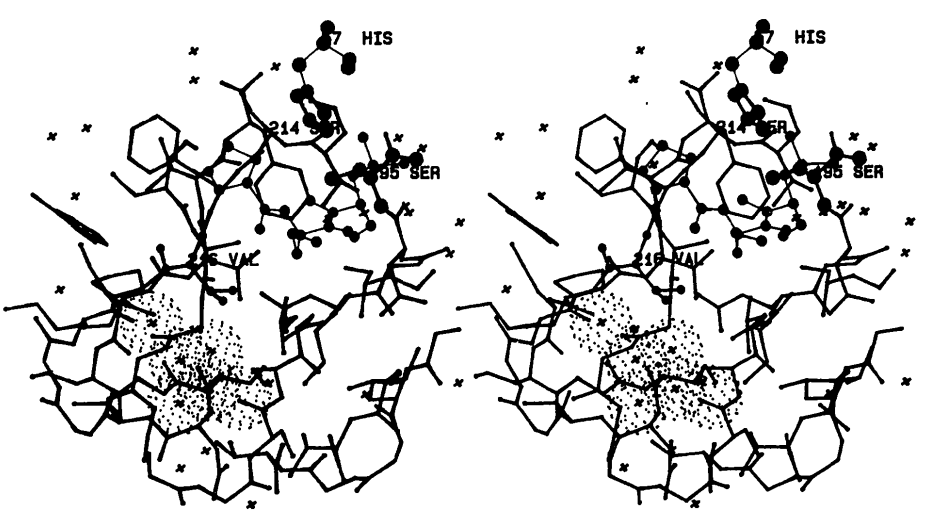

Fig. 3. Stereoview of the *WT*1 internal bound-water region of PPE in approximately the same orientation as Fig. 2. A portion of the hexapeptide (Pro-nor-valine-*N*-methylleucine; Meyer *et al.*, 1987) is shown in binding sites *S*2–*S*1–*S*1′, drawn with small balls for atoms and thin lines for bonds. Ser 195 and His 57 of the catalytic triad are drawn as larger balls and labelled. Van der Waals' surfaces are used to illustrate the size and shape of this water tunnel that extends into the 'left' β-barrel at the back of the *S*1 primary specificity site.

Cluster WT2. Buried water molecules Sol 307–305–306 branch to 308–317–318 in another shallow pocket at the top (Fig. 2) of the 'left' β-barrel, actually extending across the surface of the β-barrel domain.

Buried water WT3. Sol 316 is all by itself at the center of PPE at the interface of the two β-barrel regions, hydrogen bonding to Gly 196 O (2·8 Å) and Thr 45 Oγ (2·8 Å).

Cluster WT4. A precisely defined water molecule tunnel may be established involving Sol 319–320–321 and linking Ser 214 to the surface of the enzyme. This channel is virtually superimposable (within 1 Å) with the comparable channel in HLE: Ser 214 Oγ (2·82 Å) Sol 564 (2·73 Å) Sol 552 (2·86 Å) Sol 549 at the surface. Fig. 4 illustrates water cluster *WT4*; the topological proximity of important channels is illustrated in Fig. 5.

Cluster WT5. The 'right' cylinder (Fig. 2) extends the length of the protein with a 'split' at both top and bottom. Included water cluster *WT5* is only one water molecule deep, but with the aid of Gln 34 it makes a 'trough' to Sol 329–Sol 327 to Ca1 325 to Sol 331–Sol 330–Sol 532.

Calcium binding site

The calcium ion (325) binds at the calcium binding site, identified for PPE by Dimicoli & Beith (1977) and consisting mainly of the loop formed by residues 69–80. Sawyer *et al.* (1978) treated this as a water molecule (*IW25*). They indeed expected to find Ca here but could not distinguish it on the basis of residual density at the lower resolution of their study. This calcium ion has approximate octahedral coordination with six O atoms (Glu 70 Oε2 2·5; Asn 72 O 2·4; Glu 75 O 2·4; Asn 77 Oδ1 2·7; Glu 80 Oε2 2·4; Sol 327 2·5 Å). The values are comparable with the ones given by Bode & Huber (1978) for trypsinogen and Bode & Schwager (1975) for trypsin. It should be noted that the guanidinium group of Asn 77 can be fitted to density and refined according to two assignments of atomic positions; either the O or the N atom can be placed nearer to calcium. We have preferred to keep it so that the O atom will be nearer to calcium to complete the octahedral coordination (Fig. 6). The Nε2 atom of this Asn 77 is now at a distance of 2·8 Å from the NεH1 of Arg 24, which is rather short. Shotton & Hartley (1970) remarked that this and three other similar asparagines were particularly sensitive to deamination. Given the environment, it is plausible that deamination may have occurred yielding an aspartate. By this transformation not only the unfavorably short contact would be removed but also a favorable hydrogen bond would be introduced in the structure. We still report this residue 77 as asparagine for the sake of consistency with previous work. This assignment of Asn is supported by the gene sequence (Barrett & McDonald, 1980). Relevant water molecules in this region include Sol 329–331, 532, 554.

Cluster WT6 [Sawyer *et al.* (1978) called this cluster 1]. At the back, 'split' end of the 'right' cylinder, a chain of water molecules leads into the β-barrel from Sol 311–310–312, spreads to Sol 332–309 and penetrates into the enzyme *via* Sol 324 to Sol 323 and 322. A backbone bridge (69–70–71) joins the pair, Sol 322 and 323, which reaches across to the opposite side of the barrel structure.

Cluster WT7. At the interface of the two barrel domains, at the heart of the enzyme, a cluster of buried water molecules [described by Sawyer *et al.* (1978) as cluster 2] is immediately adjacent to the buried primary specificity pocket, *S*1 (Fig. 7). The *S*1 site is partitioned by the bulky side chain of Thr 213 (*cf.* trypsin : Val; kallikrein : Thr; HLE : Ala). The Oγ atom of Thr 213 on the *S*1 surface begins a hydrogen-bond chain that extends through internally bound water molecules (313–314–315), the far side of which is bound to and bounded by the imidazole ring of His 40 (HLE, trypsin, and chymotrypsin : His, kallikrein : Phe), which is situated on the surface of the enzyme molecule in the crevice between the two barrels and helps define the inner surface of receptor subsites *S*2′ and *S*3′. (A similar channel in HLE consists of Sol 537–Ser 197 Oγ–Sol 550.) Thus, a 'canal' of internally bound water

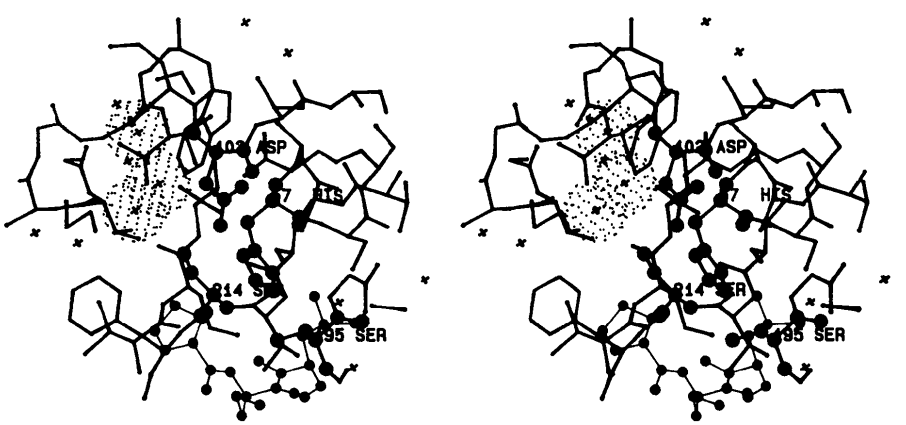

Fig. 4. Buried water channel, *WT4*, connecting the catalytic tetrad with bulk solvent, is illustrated in this stereoview. As in Fig. 3, the tripeptide is shown bound to the receptor sites *S*2–*S*1–*S*1′. The catalytic tetrad is drawn with larger balls representing individual atoms. The water molecules Sol 319–320–321 are drawn with van der Waals' radii, illustrating the size and shape of the tunnel.

molecules is terminated by two 'locks', Thr 213 and His 40, which can pivot about their Cα–Cβ bonds to expose the canal to exterior regions, thus effectively linking the primary specificity site with the hydrated surface of the protein. The static crystallographic model shows these 'canal' regions to be closely packed at van der Waals' contacts. The Thr 139 O atom forms a link in this hydrogen-bonding chain. Glu 30 is the only buried positively charged residue within the protein and is 3·3 Å from Sol 314.

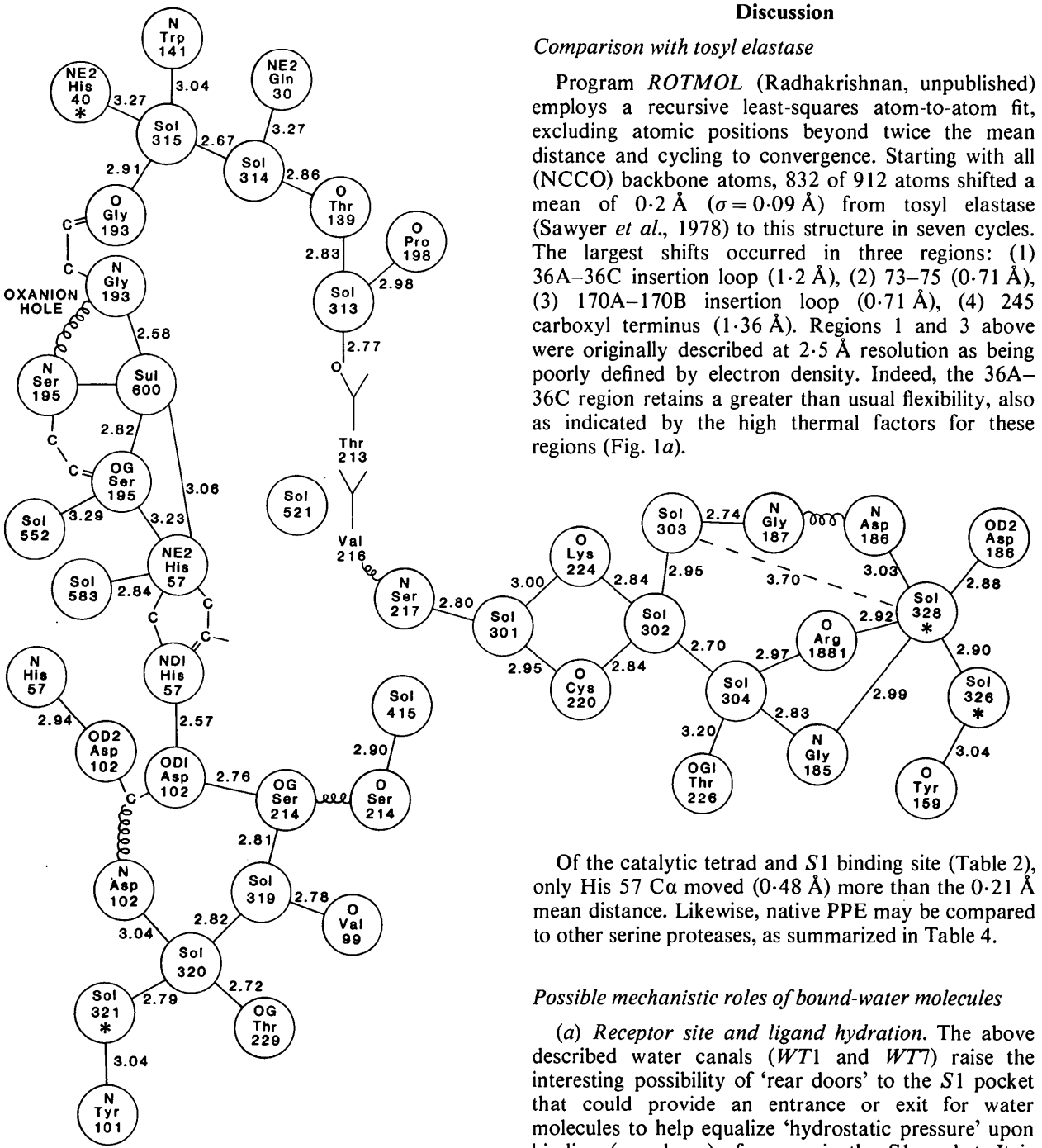

Fig. 5. A topological view of active-site residues and water channels *WT*1, *WT*4 and *WT*7 is given, together with bond lengths of hydrogen bonds. The sulfate (600) ion is also shown. Surface water molecules are marked with an asterisk.

Discussion

Comparison with tosyl elastase

Program *ROTMOL* (Radhakrishnan, unpublished) employs a recursive least-squares atom-to-atom fit, excluding atomic positions beyond twice the mean distance and cycling to convergence. Starting with all (NCCO) backbone atoms, 832 of 912 atoms shifted a mean of 0·2 Å ($\sigma = 0{\cdot}09$ Å) from tosyl elastase (Sawyer *et al.*, 1978) to this structure in seven cycles. The largest shifts occurred in three regions: (1) 36A–36C insertion loop (1·2 Å), (2) 73–75 (0·71 Å), (3) 170A–170B insertion loop (0·71 Å), (4) 245 carboxyl terminus (1·36 Å). Regions 1 and 3 above were originally described at 2·5 Å resolution as being poorly defined by electron density. Indeed, the 36A–36C region retains a greater than usual flexibility, also as indicated by the high thermal factors for these regions (Fig. 1*a*).

Of the catalytic tetrad and *S*1 binding site (Table 2), only His 57 Cα moved (0·48 Å) more than the 0·21 Å mean distance. Likewise, native PPE may be compared to other serine proteases, as summarized in Table 4.

Possible mechanistic roles of bound-water molecules

(*a*) *Receptor site and ligand hydration.* The above described water canals (*WT*1 and *WT*7) raise the interesting possibility of 'rear doors' to the *S*1 pocket that could provide an entrance or exit for water molecules to help equalize 'hydrostatic pressure' upon binding (or release) of groups in the *S*1 pocket. It is known that a water molecule helps compensate for side-chain length differences between Arg and Lys in complexes in trypsin and kallikrein (Bode, Schwager &

Huber, 1976; Chen & Bode, 1983), so it is reasonable to assume that the *S*1 subsite is hydrated prior to, and as required, during substrate binding.

The mechanistic role of solvent reorganization has been recognized (Stein 1985*a*, and especially 1985*b*) in kinetic studies. Active-site water molecules must play a primary role (*e.g.*, be removed) to permit substrate binding. Indeed, the displacement of solvent molecules attached to both the *S*1 specificity pocket as well as the *P*1 residue must surely play a structural role both in the formation of the Michaelis complex as well as in the final dissociation steps in product release. This high-resolution crystallographic study establishes the existence of three canals of internally bound water molecules and suggests that this may be a functionally important substructural characteristic of 'globular' proteins. It further suggests that the structural homology characteristic of the serine proteases may be extended considerably farther than the catalytic site.

The static constraints of the X-ray experiment only illuminate local minima that help define the possibility of a 'rear door' for water molecules. While one can model pivoting side chains to provide access to the water canals, these internal motions are really part of a dynamic model; thus dynamical breathing of the enzyme with a portal action of the pivotal Thr, His and Val residues is a plausible hypothesis which must be addressed by other physical methods. It also presents an interesting case that should be considered in dynamical force-field calculations in substrate-binding studies. It must be left to other physical methods to probe the dynamic and mechanistic roles of these canals. It does prompt one to view critically the hypotheses and calculations that have neglected the role of bound water in substrate binding and catalysis or inhibition.

(*b*) *Hydrogen-bonding to the catalytic tetrad.* Much effort has been expended in attempts to elucidate the mechanism by which serine proteases hydrolyze peptide bonds. The mechanism of the 'proton shuttle' (Kraut, 1977) reviewed by Bode & Huber (1986) linking Asp 102–His 57–Ser 195 has received much discussion. The structural evidence for, or against, the proton-relay system is taken from hydrogen-bonding schemes within

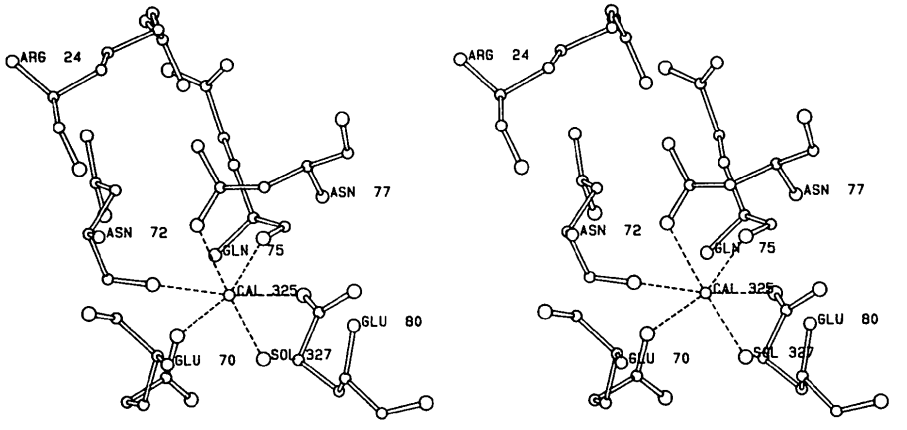

Fig. 6. Stereoview of the calcium binding site in PPE.

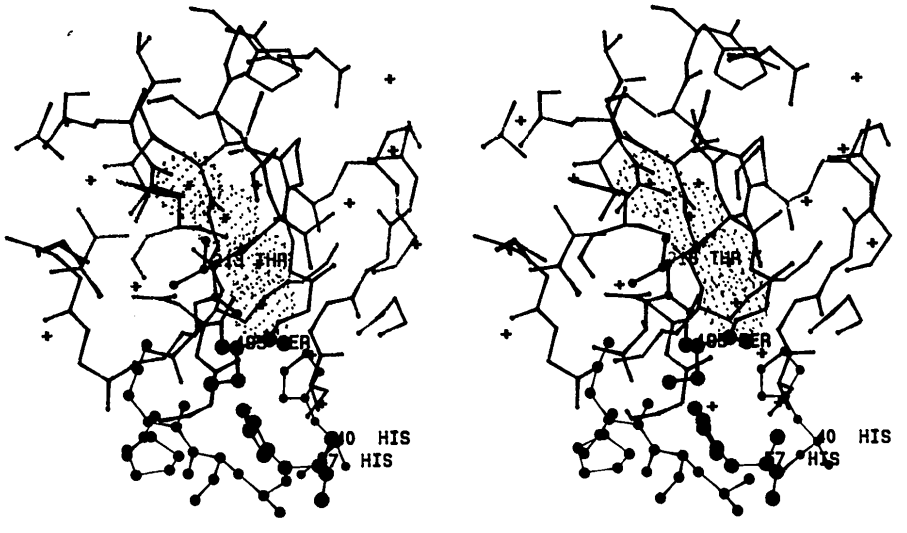

Fig. 7. As with Figs. 3 and 4, a tripeptide is drawn (small balls) into the active site; the catalytic tetrad is drawn as larger balls. The buried water channel is bounded by interior Thr 213 and His 40 on the surface. The norvaline side chain (Meyer *et al.*, 1987) is inserted into the *S*1 pocket; the Thr 213 side chain may act as a pivot, permitting water molecules to be expelled from (or to enter) the *S*1 pocket upon ligand binding.

Table 4. *Root-mean-square fit of serine proteases to PPE*

Backbone (NCCO) atoms are fitted recursively; the initial and final number of atoms used in the fit are given, together with the root-mean-square (r.m.s.) agreement.

Name	Resolution (Å)	Atoms	⟨r.m.s. fit⟩ (Å)
ELAS	2·5	912:832	0·207
HLE	1·7	188:111	0·356
4CHA	1·68	846:646	0·65
TTNA	1·9	848:491	0·416
TLLA	1·5	848:536	0·472
KPTI	2·5	860:541	0·440
RAT2	--	832:752	0·432

References: ELAS, tosyl elastase (Sawyer *et al.*, 1978); HLE, human leucocyte elastase (Bode, Wei *et al.*, 1986); 4CHA, 1·68 Å α-chymotrypsin (Tsukada & Blow, 1985); TTNA, 1·9 Å trypsin + BPTI (Huber *et al.*, 1974); TLLA, 1·5 Å trypsin + leupeptin (Walter, Radhakrishnan, Meyer & Bode, 1988); KPTI, 2·5 Å kallikrein + BPTI (Chen & Bode, 1983); RAT2, predicted rat elastase II (Carlson *et al.*, 1986).

the catalytic triad. Even as the active-site geometry is modified upon zymogen activation and N-terminus binding (Huber & Bode, 1978), here it appears that substrate binding (or lack thereof) influences the geometry of the catalytic triad. While the Asp 102–His 57 O···N distance is 2·6 Å in this native structure, it is 2·7 Å for the PPE : APA5 complex (Meyer, Radhakrishnan *et al.*, 1986). The distance between Ser 195 Oγ and His 57 Nε is 3·2 Å and the angle between Ser 195 Cβ–Ser 195 Oγ–His 57 Nε is 85·1° indicating that this interaction is weak. Thus, even though the PPE:APA complex involves unproductive (backwards) binding, the resulting geometry more strongly supports both the extended proton-relay system as well as the subtle 'induced fit' of the active site as a result of ligand binding. The active site hydrogen-bonding geometry is comparable with the description of Marquart *et al.* (1983). In fact they remark that the interaction between these two atoms is weaker in liganded compounds. Our study supports their finding. The distances between His 57 Nδ, Oδ1 and Oδ2 of Asp 102 are 3·0 and 2·6 Å, suggesting that there is a strong hydrogen bond between these two groups. It is interesting to note that Ser 214 is conserved not only sequentially but also structurally. The distance between Ser 214 Oγ and Asp 102 Oδ1 is 2·8 Å which may be compared with values given by Marquart *et al.* (1983) (2·6–2·9 Å).

Water 'tunnel' *WT*4 occupies a special place in the enzyme that suggests a plausible catalytic role. The well-recognized hydrogen-bonding network in the active site, the Ser 105–His 57–Asp 102 linkage, is actually a conserved tetrad, with Asp 102 firmly linked to Ser 214 (here 2·8 Å, *cf.* Figs. 4, 5). Ser 214 is then immediately linked to the chain of water molecules, Sol 219–220–221, forming a 'tunnel' whose entrance is flanked by hydrophilic (Tyr 101 and 234, Ser 179) or basic residues (Lys 177, Arg 230). For the native enzyme (this structure) both the active site Ser 195 and His 57 residues are hydrated (Fig. 5). However, upon binding of a substrate which has atoms occupying the *S*1′ and *S*2′ receptor subsites, bulk solvent (water and ions) must be excluded [*cf.* trypsin + BPTI (Huber *et al.*, 1974), PPE + APA5 (Meyer, Radhakrishnan *et al.*, 1986)]. Water is still abundant near the His residue of such complexes but only perpendicular to the plane of the imidazole ring, forced into a proximity very unfavorable for a hydrogen bond (*i.e.*, acute donor–H-acceptor angle) so long as His occupies the productive 'in' orientation. His 57 has also been reported in the 'out' location in several inhibitor complexes (Meyer *et al.*, 1985; Radhakrishnan *et al.*, 1987). While torsional rotation of the imadazole ring about the Cα–Cβ bond costs little energy, breaking the two (Ser, Asp) hydrogen bonds does; His 57 rotation has thus not been given serious consideration as a crucial, microreversible catalytic state.

Upon study of the family of serine proteases, the Ser 214–Sol 319–320–321 'tunnel' (Figs. 4, 5) *WT*4 has two pertinent features:

(1) The serine proteases may be divided into two classes: (*a*) trypsin-like (Ser 214 is invariably conserved in all published sequences) and (*b*) subtilisin-like (Thr or Ser 33 provides the –Oγ H group). Thus, the Oγ H functional group is uniformly conserved at the fourth amino acid of the tetrad (Ser or Thr).

(2) This water tunnel (*WT*4) is structurally conserved in many of the trypsin-like proteases (and hydrogen-bond linked to Ser 214 in those cases where water coordinates are available). For some trypsin-like proteases, *e.g.*, α-lytic protease (Fuginaga, Belbaere, Brayer & James, 1985), a major set of sequence deletions has Ser 214 linked to bulk water *via* Tyr 171 OH, obviating the need for this tunnel.

Possible functional roles attributable to this buried water tunnel include: (1) enhanced polarization of the catalytic tetrad, as suggested by Sawyer *et al.* (1978) or (2) a clearly defined, concerted hydrogen-bonding link from the catalytic Ser Oγ to bulk water, some 12 Å removed. Indeed, other functional possibilities may exist. The functional role cannot be determined from the crystallographic experiment here reported, which must limit itself to reporting the existence of this water 'tunnel', the establishment of the structural homology of the 'tunnel' and a few inferences as to structure–function relationships.

A crucial limitation with the catalytic triad (tetrad)/proton-relay mechanism involves the unanswered question: what base is strong enough to extract a proton from Ser 195 Oγ to initiate nucleophilic attack by the Oγ oxyanion?

Concerted, multiple hydrogen-bond systems have been studied in small molecules by neutron diffraction (Saenger, 1979) and with *PCILO* calculations (Lesyng & Saenger, 1981). One possibility is that proton extraction from Sol 321 by some basic ion on the surface may be transmitted in a concerted fashion all

the way to Ser 195. The mechanistic role of multiple proton transfer as a part of the rate-limiting step of catalysis has been measured with D_2O/H_2O concentration dependent proton inventory analysis and was shown to be an essential component of carbonic anhydrase catalysis (Venkatasubban & Silverman, 1980; Stein, 1983). The size and functionality of the substrate has been shown to play a role in the proton inventory studies of the serine proteases. Trypsin was shown (Elrod, Hogg, Quinn, Venkatasubban & Schowen, 1980) to have at least two protonic sites catalytically active in the transition state. In seeking to discern mechanistic properties from various experimental reports, it is important to separate those studies with larger substrates which must exclude active-site water molecules from those studies involving small or absent 'leaving' groups, which must include hydration in any mechanistic process. Relevant proton inventory studies have been reviewed by Schowen (1987).

(*c*) *Bound water in homologous structures.* We here report a structural homology that thus far has been missed in the structural literature, even though individual studies have reported some, or all, of the extant water molecules. This may be due, in part, to the conservative methods used to locate and refine water molecules. Here, it is noteworthy that the temperature values of Sol 319–320–321 are all quite low, *ca* 6 Å^2, indicating a high degree of order of the time-averaged X-ray structure.

A graphical survey of other high-resolution structures (Bernstein *et al.*, 1977) shows that a few missing links may have caused investigators to overlook these chains of water molecules in homologous structures. In chymotrypsin (Tsukada & Blow, 1985), the Val 213–His 40 canal has two of the three water molecules (H 503–504) just behind Ser 195. In trypsin (Bode & Schwager, 1975), two of the three water molecules in cluster *WT*7 were located, corresponding to Sol 313 and 315. In the 2·5 Å kallikrein + BPTI complex (Chen & Bode, 1983) HOH 110 alone is found in the channel, near Thr 713 Oγ (213). In both trypsin and chymotrypsin, cluster *WT*1 is openly in contact with the *S*1 site because of the lack of a portal side chain on Gly 216. Conversely, the exterior region has folds of β-sheets across the β-barrel that, in a rigid sense, block the canal from open water. Thus, in the cases of both high-resolution structures of chymotrypsin and trypsin, a missing component may have kept the authors from observing the connection between surface waters and the deeply buried *S*1, primary specificity site. In the 2·5 Å resolution kallikrein + BPTI structure (Chen & Bode, 1983), HOH 104 is near Asp 689 (*i.e.*, 189) and Lys 915 (P1Lys): the channel is terminated by Ser 721 (221) and His 717 (217). While no water molecules are reported, these hydrophilic residues are free to rotate 'out' into solution, effectively opening the channel to the flow of water. Elastase, which may need these channels less, has them clearly delineated in the native and complexed structures (Meyer *et al.*, 1985; Meyer, Radhakrishnan *et al.*, 1986). A striking confirmation is to be found in the recently determined structure of human leucocyte elastase (Bode, Wei *et al.*, 1986) where our water clusters *WT*1, *WT*4 and *WT*7 are conserved, even though the primary sequence and α-carbon skeletons differ markedly.

(*d*) *Structural generality of a functionally active hydrogen-bonding chain.* A major challenge to this hypothesis about the functional (catalytic) role of the tetrad + buried water comes from the comparison with the subtilisin-like subfamily of serine proteases. Inspection of the 1·3 Å resolution structure of subtilisin Carlsberg + Eglin (Bode, Papamokos *et al.*, 1986) reveals the usual catalytic triad: Ser 221 Oγ (2·7 Å) His 64 Nε–His 64 Nε (2·68 Å) Asp 32 Oδ2. This hydrogen-bonding triad is geometrically established and universally recognized. What has been overlooked until now is the extended hydrogen-bonding chain linking the triad with the surface of the enzyme (13·2 Å removed) and hence the bulk solvent. Here only one buried water molecule is needed. The hydrogen-bonding chain continues with Asp 32 Oδ2 (2·8 Å) Thr 33 N (3·0 Å) Thr 33 Oγ1 (2·7 Å) Asp 60 Oδ1 (3·0 Å) Sol 523: to surface bulk water.* Thus for all serine protein structures examined, a hydrogen-bonding chain exists from Ser 195 through to another surface locus of the enzyme. This does not prove any particular mechanism. However, it strongly suggests a vital structure–function relationship must hold for all serine proteases thus far examined.

(*e*) *Predictions based upon structural homology and force-field calculations.* There has been some interest in predicting the structures of homologous proteins. The internal, structured water was recognized to be an essential and established part of this homology and therefore was included in a recent predictive model for rat elastase II (Carlson, MacDonald & Meyer, 1986) that employed modelling and force-field refinement of bound-water molecules.

External water molecules

Careful inspection of difference Fourier maps failed to indicate a single bound-methanol molecule even though the study employed 70% methanol–acetate buffer. Density peaks are found in positions corresponding to external waters *EW*1–*EW*4, *EW*8–*EW*12, *EW*15, *EW*17, *EW*18 and *EW*21. However, by virtue of higher-resolution data, we do find a number of

* *Note added in proof*: For a group to participate in concerted 'flip-flop' hydrogen bonds, it is essential that it be both donor and acceptor. This would exclude the amide >N–H group (Thr 33 N). However, a small, 20° rotation of the Thr 33 side chain places the Oγ atom directly into the chain, thus permitting the possibility of 'flip-flop' hydrogen bonding.

additional bound-water molecules exterior to PPE. In positions near each of those which Sawyer *et al.* (1978) reported for *EW*8, *EW*9 and *EW*21, two (rather than one) distinct water density peaks are found. An additional 32 water molecules making multiple hydrogen bonds with atoms of the enzyme were located. Furthermore, 70 solvent molecules making a single hydrogen bond were located. Some of these external water molecules are part of water inlets into the interior of the enzyme. The remainder of the ordered solvent molecules tend to cluster on polar indentations on the enzyme's surface rather than forming a continuous monolayer.

Sequence-dependent tunnels

The terminal solvent molecule in cluster *WT*4 (Sol 319) makes contact with the side chain of Val 99; a constituent of the *S*2 and *S*4 binding subsites. Were it not for the Val 99 side chain, cluster *WT*4 would actually constitute a tunnel throughout the enzymes connecting the *S*2 and *S*4 binding subsites with a point on the surface of the enzyme some 10 Å distant from the substrate binding site. Since the side chain of Val 99 is also responsible for dividing a single groove on the enzyme's surface into two (*S*2 and *S*4) binding subsites, those researchers contemplating site-directed mutagenesis in the elastase gene might find that a valine to glycine mutation (CTT to CCT) at residue 99 might produce a variety of interesting effects. A similar situation exists in inlet cluster *WT*1 which is a 12 Å channel entering the peptide loop consisting of residues 220–226 and terminates against the side chains of Val 126 and Thr 226. These side chains form one wall of the *S*1 subsite. Were it not for these side chains, the inlet would be a tunnel connecting *S*1 with a distant point of the enzyme's surface.

Specificity of elastase

Until recently, all PPE complexes of peptide-like substrates demonstrated 'reverse' binding modes. 'Forward' binding is exemplified by the location of the bovine pancreatic trypsin inhibitor (BPTI) bound to trypsin (Huber *et al.*, 1974) and most recently in the complex of human leucocyte elastase with turkey ovomucoid inhibitor 3rd domain (Bode, Wei *et al.*, 1986). Evidence for 'reversed' binding includes the crystal structures of complexes Ac-Ala-Ala-Ala-OH and Ac-Pro-Ala-Pro-Ala-OH (Shotton, White & Watson, 1971) and a trifluoroacetyl-Lys-Ala derivative (Hughes, Sieker, Beith & Dimicoli, 1982), but was first described as such in the PPE : APA complex (Meyer, Radhakrishnan *et al.*, 1986). Reverse binding is primarily the result of the small number of backbone–backbone hydrogen bonds that characterize receptor + ligand binding.

We (Meyer & Presta, 1986; Presta & Meyer, 1987) have used model-building techniques to explore the possible binding modes of elastase substrates. While 'reversed' binding modes have been observed in PPE crystallographically, these studies suggest that it is highly unlikely that the binding modes observed in these complexes mimic enzyme-substrate complexes where the substrate is actually hydrolyzed. Only in the normal binding mode can the scissile amide or ester linkage be brought into correct juxtaposition with the oxyanion hole of the active site. Taken together, these results suggest that while PPE has a propensity to bind peptide-like substrates and inhibitors in a backwards manner, complexes exhibiting this 'reversed' binding are not good models of productive Michaelis complexes and thus represent unproductive binding, which will still contribute to kinetics and binding values (Clore, Gronenborn, Carlson & Meyer, 1986). In terms of drug-design methods, the anomalous results exhibited by PPE alert the investigator to caution. With no ionic interactions and a small number of (sometimes, backwards!) hydrogen-bonding sites, binding specificity is dominated by the composite van der Waals' interactions. This makes elastase an ideal model compound for studying the role of weak interactions in the absence of more dominant factors. This lack of specificity also raises the possibility of using PPE to catalyze 'peptide- or ester-like' bonds in non-peptide substrates.

Elastase subsites *S*1, *S*2, *S*4 and *S*1′ (Table 2) are small, weakly polar indentations on the surface of the enzyme capable of binding small amino-acid side chains. The *S*3 subsite is a larger groove of more polar character capable of contact with unbranched side chains. A number of non-productive binding modes may compete with productive ones because stark features distinguishing *S*1, *S*2, *S*4 and *S*1′ do not exist to direct the substrate into a productive binding mode. Because *S*3 is rather unique in being able to accommodate a large side chain and being more polar than the other subsites, we may expect that choice of the proper residue at *P*3 (Asp, for example) may direct the substrate towards a productive binding mode just as effectively as proline has been shown to do at *P*2.

As inhibitor binding studies proceed, with addition of the structure of human leucocyte elastase (Bode, Wei *et al.*, 1986), the 1·65 Å resolution structure analysis of native PPE provides a basis for comparison with a variety of inhibitor classes. It also suggests that important mechanistic and catalytic functions of the serine proteases have yet to be discovered.

Diffraction data were measured in the laboratory of Professor Robert Huber, to whom we are especially grateful. Dr W. Bode provided many useful suggestions as well as access to the HLE and subtilisin + eglin atomic coordinates. Drs Richard E. Rosenfield Jr,

Leonard Presta and Lori Takahashi contributed critical discussions. We thank Mrs Sharyll Pressley for typing assistance. Dr Stanley Swanson materially assisted with many computational aspects of this study. We acknowledge the Robert A. Welch Foundation, NATO, the Tobacco Research Fund, USA, ICI Americas, Office of Naval Research and the Texas Agricultural Experiment Station for financial support.

References

BARRETT, A. J. & MCDONALD, J. K. (1980). *Mammalian Proteases*, Vol. 1, *Endopeptidases*. New York: Academic Press.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., ROBERS, J. R., KENNARD, O., SCHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BIZZOZERO, S. A. & DUTLER, H. (1981). *Bioorg. Chem.* **10**, 46–62.

BLOW, D. M., BIRKTOFT, J. J. & HARTLEY, B. S. (1969). *Nature (London)*, **221**, 337–340.

BODE, W. & HUBER, R. (1978). *FEBS Lett.* **90**, 265–269.

BODE, W. & HUBER, R. (1986). *Crystal Structures of Pancreatic Serine Endopeptidases*. In *Molecular and Cellular Basis of Digestion*, edited by P. DESNUELLE, H. SJOESTROEM & O. NOREN. Amsterdam: Elsevier.

BODE, W., PAPAMOKOS, E., MUSIL, D., SEEMUELLER, U. & FRITZ, H. (1986). *EMBO J.* **5**, 813–818.

BODE, W. & SCHWAGER, P. (1975). *J. Mol. Biol.* **98**, 693–717.

BODE, W., SCHWAGER, P. & HUBER, R. (1976). In *Proteolysis and Physiological Regulation, Miami Winter Symposia* 11, p. 43-076. New York: Academic Press.

BODE, W., WEI, A.-Z., HUBER, R., MEYER, E., TRAVIS, J. & NEUMANN, S. (1986). *EMBO J.* **5**, 2453–2458.

CARLSON, G. M., MACDONALD, R. J. & MEYER, E. F. (1986). *J. Theor. Biol.* **119**, 107–124.

CHEN, Z. & BODE, W. (1983). *J. Mol. Biol.* **164**, 283–311.

CLORE, M. G., GRONENBORN, A., CARLSON, G. M. & MEYER, E. F. (1986). *J. Mol. Biol.* **190**, 259–267.

COLE, G. M. (1986). PhD Dissertation, Texas A&M Univ., USA.

DEISENHOFER, J., REMINGTON, S. J. & STEIGEMANN, W. (1985). *Methods in Enzymology*, Vol. 115B, edited by H. W. WYCKOFF, C. H. W. HIRS & S. N. TIMASHEFF, pp. 303–323. New York: Academic Press.

DIMICOLI, J. L. & BEITH, J. (1977). *Biochemistry*, **16**, 5532–5537.

ELROD, J. P., HOGG, J. L., QUINN, D. M., VENKATASUBBAN, K. S. & SCHOWEN, R. L. (1980). *J. Am. Chem. Soc.* **102**, 3917–3922.

FREER, S. T., KRAUT, J., ROBERTUS, J. D., WRIGHT, H. T. & XUONG, N. H. (1970). *Biochemistry*, **9**, 1977.

FUGINAGA, M., DELBAERE, L. T. J., BRAYER, G. D. & JAMES, M. N. G. (1985). *J. Mol. Biol.* **184**, 479.

HUBER, R. & BODE, W. (1978). *Acc. Chem. Res.* **11**, 114–122.

HUBER, R., KUKLA, D., BODE, W., SCHWAGER, P., BARTELS, K., DEISENHOFER, J. & STEIGEMANN, W. (1974). *J. Mol. Biol.* **89**, 73–101.

HUGHES, D. L., SIEKER, L. C., BEITH, J. & DIMICOLI, J. L. (1982). *J. Mol. Biol.* **162**, 645–658.

JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931–935.

JOHNSON, D. & TRAVIS, J. (1978). *J. Biol. Chem.* **253**, 7142–7144.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KOSSIAKOFF, A. A. & SPENCER, S. A. (1981). *Biochemistry*, **20**, 6462–6474.

KRAUT, J. (1977). *Ann. Rev. Biochem.* **46**, 331–358.

LESYNG, B. & SAENGER, W. (1981). *Biochim. Biophys. Acta*, **678**, 408–413.

LEVITT, M. (1974). *J. Mol. Biol.* **82**, 393–420.

LUZATTI, V. (1952). *Acta Cryst.* **5**, 802–810.

MARQUART, M., WALTER, J., DEISENHOFER, J., BODE, W. & HUBER, R. (1983). *Acta Cryst.* B**39**, 480–490.

MEYER, E. F., CLORE, G. M., GRONENBORN, A. M. & HANSEN, H. A. S. (1987). *Biochemistry*. In the press.

MEYER, E. F., COLE, G. M., PRESTA, L. G., ROSENFIELD, R. E. & SWANSON, S. M. (1982). *Modelling the Binding of Small Molecules to Proteins*. In *Structure of Complexes Between Biopolymers and Low-Molecular-Weight Molecules*, edited by W. BARTMANN & G. SNATZKE, pp. 77–85. London: Wiley–Heyden.

MEYER, E. F. & PRESTA, L. G. (1986). *The Modelling and Verification of Complexes of Elastase and Other Serine Proteases*. In *Molecular Graphics and Drug Design*, edited by A. V. S. BURGEN, G. C. K. ROBERTS & M. D. TUTE, pp. 307–321. Amsterdam: Elsevier.

MEYER, E. F., PRESTA, L. G. & RADHAKRISHNAN, R. (1985). *J. Am. Chem. Soc.* **107**, 4091–4093.

MEYER, E. F., RADHAKRISHNAN, R., COLE, G. M. & PRESTA, L. G. (1986). *J. Mol. Biol.* **189**, 553–559.

MORIMOTO, C. N. & MEYER, E. F. (1976). *Information Retrieval, Computer Graphics, and Remote Computing*. In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 488–496. Copenhagen: Munksgaard.

PRESTA, L. G. (1985). PhD Dissertation, Texas A&M Univ., USA.

PRESTA, L. G. & MEYER, E. F. JR (1987). *Biopolymers*, **26**(8), 1207.

RADHAKRISHNAN, R., PRESTA, L. G., MEYER, E. F. & WILDONGER, R. (1987). *J. Mol Biol.* In the press.

ROBERTUS, J. D., ALDEN, R. A., BIRKTOFT, J. J., KRAUT, J., POWERS, J. C. & WILCOX, P. E. (1972). *Biochemistry*, **11**, 2439–2449.

SAENGER, W. (1979). *Nature (London)*, **279**, 343–344.

SAWYER, L., SHOTTON, D. M., CAMPBELL, J. W., WENDELL, P. L., MUIRHEAD, H., WATSON, H. C., DIAMOND, R. & LADNER, R. C. (1978). *J. Mol. Biol.* **118**, 137–208.

SCHECHTER, I. & BERGER, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.

SCHOWEN, R. L. (1987). In *Principles of Enzyme Activity*, Vol. 9, *Molecular Structure and Energetics*, edited by J. F. LIEBMAN & A. GREENBERG. Deerfield Beach, FL: VCH Publishers.

SCHWAGER, P., BARTELS, K. & JONES, T. A. (1975). *J. Appl. Cryst.* **8**, 275–280, with modifications by W. S. BENNET, unpublished.

SHOTTON, D. M. & HARTLEY, B. S. (1970). *Nature (London)*, **225**, 802–806.

SHOTTON, D. M., WHITE, N. J. & WATSON, H. C. (1971). *Cold Spring Harbor Symp.* No. XXXVI, pp. 91–105.

SPRANG, S., STANDING, T., FLETTERICK, R., FINGER-MOORE, J., STROUD, R., XUONG, N.-H., HAMLIN, R., RUTTER, W. & CRAIK, C. S. (1987). *Science*, **237**, 905–909.

STEIGEMANN, W. (1974). PhD Thesis, Technische Univ. Munchen, Federal Republic of Germany.

STEIN, R. L. (1983). *J. Am. Chem. Soc.* **105**, 5111–5116.

STEIN, R. L. (1985*a*). *J. Am. Chem. Soc.* **107**, 6039–6042.

STEIN, R. L. (1985*b*). *J. Am. Chem. Soc.* **107**, 7768–7769.

STEITZ, T. A. & SCHULMAN, R. G. (1982). *Annu. Rev. Biophys. Bioeng.* **11**, 419.

SWANSON, S. M. (1988). *Acta Cryst.* A**44**. Submitted.

TAKAHASHI, L. H., RADHAKRISHNAN, R., ROSENFIELD, R. E. JR & MEYER, E. F. (1987). *J. Mol. Biol.* Submitted.

TAKAHASHI, L. H., RADHAKRISHNAN, R., ROSENFIELD, R. E. JR, MEYER, E. F. & TRAINOR, D. A. (1987). *J. Am. Chem Soc.* Submitted.

TSUKADA, H. & BLOW, D. M. (1985). *J. Mol. Biol.* **184**, 703–711.

VENKATASUBBAN, K. S. & SILVERMAN, D. N. (1980). *Biochemistry*, **19**, 4948–4989.

WALTER, J., RADHAKRISHNAN, R., MEYER, E. F. & BODE, W. (1988). In preparation.

WLODAWER, A., DEISENHOFER, J. & HUBER, R. (1987). *J. Mol. Biol.* **193**, 145–156.