

# A Neutron Laue Diffraction Study of Endothiapepsin: Implications for the Aspartic Proteinase Mechanism<sup>†</sup>

L. Coates,<sup>‡</sup> P. T. Erskine,<sup>‡</sup> S. P. Wood,<sup>‡</sup> D. A. A. Myles,<sup>§</sup> and J. B. Cooper<sup>\*‡</sup>

*Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Southampton SO16 7PX, U.K., and EMBL Grenoble Outstation, Institut Laue-Langevin, 6 rue Jules Horowitz, 38042 Grenoble, France*

*Received March 28, 2001; Revised Manuscript Received July 10, 2001*

**ABSTRACT:** Current proposals for the catalytic mechanism of aspartic proteinases are largely based on X-ray structures of bound oligopeptide inhibitors possessing nonhydrolyzable analogues of the scissile peptide bond. However, the positions of protons on the catalytic aspartates and the ligand in these complexes have not been determined with certainty. Thus, our objective was to locate crucial protons at the active site of an inhibitor complex since this will have major implications for a detailed understanding of the mechanism of action. We have demonstrated that high-resolution neutron diffraction data can be collected from crystals of the fungal aspartic proteinase endothiapepsin bound to a transition state analogue (H261). The neutron structure of the complex has been refined at a resolution of 2.1 Å to an *R*-factor of 23.5% and an *R*<sub>free</sub> of 27.4%. This work represents the largest protein structure studied to date by neutron crystallography at high resolution. The neutron data demonstrate that 49% of the main chain nitrogens have exchanged their hydrogen atoms with D<sub>2</sub>O in the mother liquor. The majority of residues resisting exchange are buried within core β-sheet regions of the molecule. The neutron maps confirm that the protein has a number of buried ionized carboxylate groups which are likely to give the molecule a net negative charge even at very low pH, thereby accounting for its low pI. The functional groups at the catalytic center have clearly undergone H–D exchange despite being buried by the inhibitor occupying the active site cleft. Most importantly, the data provide convincing evidence that Asp 215 is protonated and that Asp 32 is the negatively charged residue in the transition state complex. This has an important bearing on mechanistic proposals for this class of proteinase.

The aspartic proteinases are a family of enzymes involved in a number of important biological processes (see ref 1 for a review). In animals, the enzyme renin has a hypertensive action through its role in the renin–angiotensin system. The retroviral aspartic proteinases, such as the HIV proteinase, are essential for maturation of the virus particle, and inhibitors have a proven therapeutic record in the treatment of AIDS. The lysosomal aspartic proteinase cathepsin D has been implicated in tumorigenesis, and the stomach enzyme pepsin plays a major physiological role in hydrolysis of acid-denatured proteins. All enzymes in this class are inhibited by the microbial peptide pepstatin which contains the unusual amino acid statine.

## Structure and Mechanism

Most eukaryotic aspartic proteinases are monomeric and consist of a single chain of ~330 amino acids which forms two similar domains with the active site located between them. In contrast, retroviral aspartic proteinases are dimeric, consisting of two identical subunits, each roughly equivalent

to one domain of a eukaryotic aspartic proteinase. Accordingly, the amino acid sequences of the eukaryotic enzymes have signs of an internal repeat relating the two halves of the molecule, their level of identity being greatest in the vicinity of the active site which involves two conserved Asp-Thr-Gly sequences (2). Hence, the eukaryotic aspartic proteinases are thought to have evolved divergently from a primitive dimeric enzyme (resembling the retroviral proteinase) by gene duplication and fusion. In all aspartic proteinases, the base of the active site cleft is made of β-strands which contain the catalytic aspartate residues (Asp 32 and 215 in porcine pepsin). The side chains of the aspartates are held in a coplanar arrangement and within hydrogen bonding distance by an intricate arrangement of H-bonds involving surrounding main chain and conserved side chain groups. A solvent molecule is found between both aspartate carboxyls in all native aspartic proteinase crystal structures and is presumed to be a water molecule. This water, which is hydrogen-bonded to both aspartate carboxyls, is thought to take part in the catalytic mechanism (3). In numerous chemical studies, the failure to trap covalently bound substrate indicates that the reaction involves an intermediate which binds noncovalently to the enzyme (4). NMR studies using an inhibitor with a ketone analogue of the scissile peptide bond (5) suggested that it binds to the enzyme in a hydrated *gem*-diol form [ $>C(OH)_2$ ]. Thus, current proposals for the catalytic mechanism (such as that shown in Figure

<sup>†</sup> We gratefully acknowledge the BBSRC (U.K.) for project grant support and the EPSRC for funding a research studentship (to L.C.).

<sup>\*</sup> To whom correspondence should be addressed. E-mail: J.B.Cooper@soton.ac.uk. Telephone: +44-(0)2380-595381. Fax: +44-(0)2380-594459.

<sup>‡</sup> University of Southampton.

<sup>§</sup> Institut Laue-Langevin.

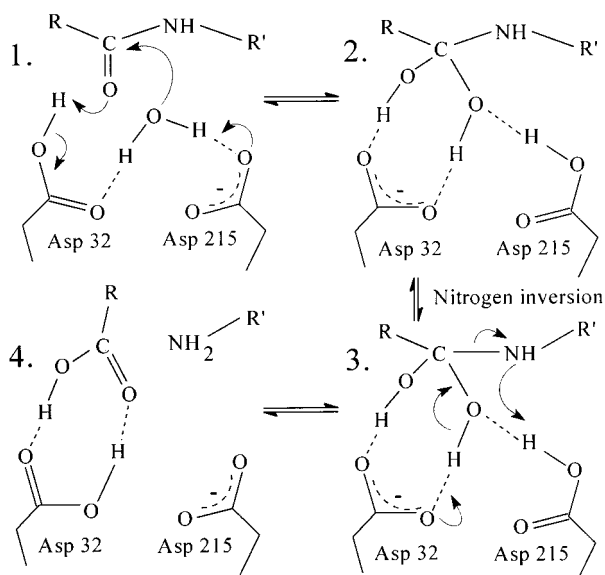


FIGURE 1: Catalytic mechanism proposed by Veerapandian et al. (6) which is based on the X-ray structure of a difluoroketone (*gem*-diol) inhibitor bound to endothiapepsin. A water molecule tightly bound to the aspartates in the native enzyme is proposed to nucleophilically attack the scissile bond carbonyl. The resulting tetrahedral intermediate (2) is stabilized by hydrogen bonds to the negatively charged carboxyl of Asp 32. Fission of the scissile C–N bond is accompanied by transfer of a proton to the leaving amino group either from Asp 215 (with nitrogen inversion) or from bulk solvent. Dashed lines represent hydrogen bonds.

1) invoke nucleophilic attack of the active site water molecule on the scissile bond carbonyl generating a tetrahedral *gem*-diol intermediate (6). The latter is stabilized by hydrogen bonds with the enzyme's active site aspartates. Studies of kinetic isotope effects using HIV proteinase have led to similar conclusions (7).

Recently, it has been shown that the yeast aspartic proteinase saccharopepsin has a macromolecular inhibitor which forms an  $\alpha$ -helix occupying the enzyme's active site cleft (8). In contrast, short oligopeptide inhibitors of aspartic proteinases bind in the active site in extended conformations with up to 10 amino acids occupying the cleft. The best synthetic inhibitors are those in which one or both of the hydroxyl groups of the putative transition state are mimicked (intermediate 2 in Figure 1). One hydroxyl binds by hydrogen bonds to both of the catalytic aspartates in the same position as the solvent molecule in the native enzyme, and most of the transition state analogues (e.g., statine) mimic this group alone. In contrast, fluoroketone analogues ( $-\text{CO}-\text{CF}_2-$ ) mimic both hydroxyls of the putative intermediate since they readily hydrate to the *gem*-diol form  $[-\text{C}(\text{OH})_2-\text{CF}_2-]$ .

In catalysis, the active site water molecule may become partly displaced upon substrate binding and polarized by one of the aspartate carboxyls (9). The water may then nucleophilically attack the scissile bond carbonyl group to form the tetrahedral intermediate. The latest proposals for the mechanism based on crystal structures (6, 9–11) are consistent in general terms, but they lack direct experimental evidence for the protonation states of the active site groups. Since the active site hydrogen atoms cannot be located by X-ray analysis of proteins even at high resolution, their putative positions have to be inferred from the local geometry of surrounding polar atoms. One of the key features of the

mechanism proposed by Veerapandian et al. (6) (shown in Figure 1) is the stabilization of the transition state by a negative charge localized on Asp 32. The assignment of a negative charge to this residue was made on the basis that its hydrogen bonding capacity is satisfied to a greater extent than that of Asp 215 in complex with the *gem*-diol inhibitor. This implies that a negative charge on Asp 32 would be more stable than if it were on Asp 215 in the transition state complex. However, the protonation states of the catalytic aspartates in the complexes had not been determined with certainty at that stage.

#### Background to Neutron Studies

Information about the location of hydrogen atoms in proteins can be obtained from X-ray cryocrystallographic studies if atomic resolution data can be collected ( $d < 1.2$  Å). Hydrogen atom positions can be determined for particularly well ordered regions of the main chain and side chains, but disorder and high temperature factors can still render many more labile hydrogen atoms invisible. In addition, hydrogen atoms involved in low-barrier hydrogen bonds (found at enzyme active sites) will be distributed between two positions  $\sim 0.5$  Å apart (12) and may therefore have high apparent temperature factors. There is evidence for these effects in neutron studies of carboxylic acid dimers (13). Recent work has shown that a number of functional groups in proteins, including carboxylates, are particularly vulnerable to radiolytic damage at the radiation doses that might be required to collect such atomic resolution data from protein crystals (14, 15). Aspartate and glutamate residues, including those involved in buried salt bridge interactions, are found to become disordered rapidly and appear to suffer from decarboxylation of the side chain even at 100 K. These effects could make the process of locating hydrogen atoms on active site carboxylate groups with X-ray data difficult.

However, atomic resolution X-ray crystallography provides an alternative means of defining the protonation states of carboxylate groups. Neutral carboxyl groups have a significant difference between the C–OH and C=O bond lengths (typically, 1.20 Å for the C=O bond and 1.30 Å for the C–OH bond), whereas ionized carboxylates are expected to have identical C–O bond lengths (typically, 1.25 Å) due to resonance. Thus, unrestrained refinement using atomic resolution data is a very powerful tool for determination of the protonation state of each residue and avoids the difficulties of locating weak electron density due to the hydrogens themselves (J. Helliwell, personal communication). However, it must be remembered that the apparent bond lengths can be influenced by the temperature factors of the atoms involved. Unrestrained refinement of native endothiapepsin at 0.9 Å (P. Erskine, unpublished observations) has shown that one cannot assign a charge to either aspartate carboxyl at the catalytic center since all four C–O bond lengths refine to almost identical values. It is highly unlikely that both aspartates are ionized at the same time in view of their proximity. Therefore, this effect must be due to the presence of two canonical forms at the catalytic center; i.e., in half of the enzyme molecules Asp 32 is ionized, and in the remaining half Asp 215 is ionized. However, an asymmetric charge distribution may arise in the presence of substrate (see Figure 1).

Neutron protein crystallography offers a powerful adjunct to X-ray analysis at the resolutions typical of most X-ray protein structure determinations ( $d > 1.5 \text{ \AA}$ ) by enabling key details of hydrogen atom positions and solvent structure to be revealed (16). Hydrogen and/or deuterium atoms may be more readily located in a corresponding neutron analysis because the scattering lengths of hydrogen ( $-3.7 \text{ fm}$ ) and deuterium ( $6.7 \text{ fm}$ ) for neutrons are closely similar to those of other biological atoms:  $6.6 \text{ fm}$  for carbon,  $9.4 \text{ fm}$  for nitrogen,  $5.8 \text{ fm}$  for oxygen, and  $2.8 \text{ fm}$  for sulfur. The difference in amplitude and phase means hydrogen atom positions appear as characteristic negative density features in neutron maps, while deuterium-labeled positions have density indistinguishable from that of carbon atoms. Hydrogen- and deuterium-labeled or -exchanged positions can therefore be distinguished in the crystal even at resolutions of  $\sim 2.0 \text{ \AA}$ . Nitrogen and oxygen atoms can also be distinguished, allowing the carbonyl and amides of asparagine and glutamine residues to be properly assigned. While cancellation of neighboring positive and negative scattering density can reduce the visibility of some hydrogen-rich groups in medium-resolution analyses [e.g.,  $\text{H}_2\text{O}$  ( $-1.7 \text{ fm}$ ) and  $\text{CH}_3$  ( $-4.6 \text{ fm}$ )], the opposite is true for deuterium-labeled groups where strong reinforcement from neighboring positively scattering atoms can greatly enhance their visibility [e.g.,  $\text{D}_2\text{O}$  ( $19.1 \text{ fm}$ ) and  $\text{CD}_3$  ( $26.7 \text{ fm}$ )]. This is particularly important in  $\text{D}_2\text{O}$  solvent structure analysis, where scattering from both the deuterium and oxygen atoms can allow the orientation and geometry of well-ordered water molecules to be determined and can render disordered or mobile water molecules more visible than in a corresponding X-ray analysis.

Prior to exposure, crystals are usually subjected to  $\text{H}_2\text{O}$ – $\text{D}_2\text{O}$  exchange to reduce the otherwise large hydrogen incoherent scattering background. This soaking procedure enables all solvent accessible OH and NH groups on the protein to undergo H–D exchange. Since neutron maps readily discriminate between H and D density in the structure, analysis of the pattern and extent of this H–D exchange can provide an elegant probe of group accessibility, mobility, and the dynamics of the protein.

The ability to visualize and discriminate individual hydrogen atoms, and more especially the deuterium isotope, at resolutions of  $> 1.5 \text{ \AA}$  and at room temperature is of much potential value in the study of enzymatic processes. The use of neutron diffraction in understanding enzyme mechanisms at these resolutions was pioneered by Kossiakoff and Spencer (17), whose analysis of trypsin contributed significantly to the current understanding of serine proteinase catalysis. Similar work on lysozyme (18) and ribonuclease (19) has established the protonation states of the catalytic groups with important implications for their mechanisms of action.

Despite this potential wealth of information, the problems associated with the relatively low flux of available neutron beams have until now restricted neutron crystallography to only a few proteins which produce very large crystals. While neutron analysis imparts no observable radiation damage upon biological samples, the long time scales required to collect data by conventional means have made most studies prohibitive. To address these issues, a neutron Laue diffractometer (LADI) for protein crystallography was developed by the EMBL Grenoble Outstation and the Institut Laue-

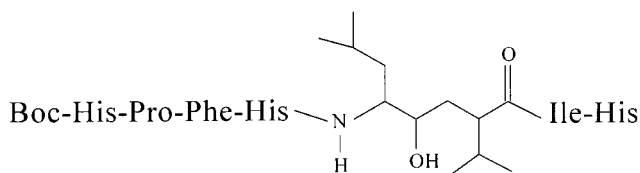


FIGURE 2: Chemical structure of the inhibitor H261. The central region of the inhibitor mimics a Leu-Val dipeptide. Here the inhibitor possesses a hydroxyethylene transition state analogue.

Langevin (ILL) (20). One of the benefits of the LADI diffractometer is the speed with which high-resolution neutron data can be collected. The combination of a broad band-pass quasi-Laue geometry (e.g.,  $\lambda = 3.5 \text{ \AA}$ ,  $d\lambda/\lambda = 25\%$ ) with a novel cylindrical, neutron sensitive image plate detector that completely surrounds the sample provides 10–100-fold gains in efficiency compared with those of conventional monochromatic neutron diffractometers (21). This has made feasible studies of larger biological complexes and smaller crystals than was previously possible and has enabled the location of important hydrogen and/or deuterium atom positions in a number of systems at the resolutions typical of the majority of protein structure analyses (22–25). This approach therefore offers great promise in the study and location of key hydrogen atom positions involved in enzymatic mechanisms, proton pumping, and shuttling processes and in protein–ligand bonding interactions.

We have collected neutron Laue data on an inhibitor complex of the fungal aspartic proteinase endothiapepsin, allowing the structure to be refined at a resolution of  $2.1 \text{ \AA}$ . The bound inhibitor H261 possesses a hydroxyethylene transition state analogue (see Figure 2). The neutron structure of the endothiapepsin complex reveals the locations of crucial protons at the active site with important implications for the mechanism of action. The structure confirms earlier suggestions that the low pI values of many aspartic proteinases can be attributed to buried negatively charged carboxyl groups. This work also represents the largest macromolecular neutron structure to be analyzed to date.

## EXPERIMENTAL PROCEDURES

**Crystallization and Deuteration.** The cocrystals of endothiapepsin complexed with H261 were obtained by an adaptation of the native enzyme crystallization (26). This involved adding a 10-fold molar excess of inhibitor to enzyme at a concentration of  $2.0 \text{ mg/mL}$  in  $100 \text{ mM}$  sodium acetate buffer at  $\text{pH } 4.5$ . Finely ground ammonium sulfate was added until slight turbidity became apparent ( $\sim 55\%$  saturation), at which point the solution was Millipore filtered and, if necessary, a few drops of acetone was added to clear any remaining turbidity. Crystals generally grew after several weeks. Those used for neutron diffraction had been stored in mother liquor for a total of 11 years. To reduce the contribution of the large incoherent neutron scattering cross section of hydrogen to the experimental background, crystals of the endothiapepsin–H261 complex were subjected to hydrogen–deuterium exchange by vapor diffusion for several months prior to data collection.

The H–D exchange involved mounting selected crystals in large diameter capillaries and equilibrating against successive changes of a 90% deuterated mother liquor solution and, finally, against several changes of a 95% deuterated

solution. This method has been shown to be effective in other studies (24). The decision to use vapor diffusion rather than soaking directly in D<sub>2</sub>O was made to avoid shocking the crystals since only a few were large enough for neutron data collection and they were 11 years old at the time of the experiment.

Preliminary neutron analysis identified several crystals that were suitable for neutron Laue data collection. Data collection was started with the intention of obtaining a complete data set with a single crystal (1.8 mm × 1.4 mm × 1.4 mm). A partial data set to 2.2 Å was collected in a preliminary experiment (27); however, the crystal later dried out, and a second crystal of similar volume (3 mm<sup>3</sup>) from the same batch was used to complete the data set. Both crystals had been deuterated for the same length of time, were monoclinic (*P*2<sub>1</sub>), and had the following unit cell dimensions: *a* = 43.1 Å, *b* = 75.7 Å, *c* = 42.9 Å, and β = 97.0°.

**Data Collection and Processing.** Quasi-neutron Laue data were collected on the LADI instrument installed on end-station LADI/T17 of cold neutron guide H142 at ILL (21). The detector consists of a large Gd<sub>2</sub>O<sub>3</sub> doped neutron sensitive image plate (400 mm × 800 mm) mounted on a cylindrical camera (318 mm diameter) that completely encircles the sample (20). A Ni/Ti multilayer wavelength selector was used to select a narrow band-pass (dλ/λ = 25%) centered at 3.20 Å. This serves to reduce both the reflection overlap at the detector and the incoherent neutron scattering background that comes largely from unexchanged hydrogens in the sample. A further benefit of using a narrow quasi-Laue band-pass is that, since the proportion of energy-overlapped (harmonic) reflections is dependent on the ratio dλ/λ and is independent of the size of unit cell, almost all reflections (>99%) recorded on the LADI detector are singlets and thus amenable to standard crystallographic analysis. The spectral range was determined by time-of-flight techniques.

Data for the endothiapepsin–H261 complex were collected at room temperature. A total of 61 frames were obtained with a φ separation of typically 8° and exposure times between 22 and 32 h per frame. The observed diffraction peaks were indexed and matched to a wavelength range of 2.7–3.6 Å and to a *d*<sub>min</sub> of 1.95 Å using the program LAUEGEN (28, 29). Integrated intensities were extracted using the *I*/σ(*I*) method (30, 31) as implemented in the program INTEGRATE. Data from groups of images were scaled together and wavelength-normalized (to account for the spectral distribution) using the program LAUESCALE. Data recorded at wavelengths of <2.8 Å were in poor agreement with the remainder of the data, most likely due to the falloff in the spectral intensity distribution, and were excluded from wavelength normalization calculations. Final scaling and merging of the data frames were performed with SCALA of the CCP4 suite (32).

**Refinement.** The coordinates of the endothiapepsin–H261 complex were obtained from the earlier X-ray analysis of this complex at 1.6 Å resolution (33). Least-squares refinement with the neutron data set was performed initially using XPLOR (34) assuming the molecule to be fully deuterated. Inspection of the neutron density maps (2*F*<sub>o</sub> – *F*<sub>c</sub> and *F*<sub>o</sub> – *F*<sub>c</sub>) using the program TURBO-FRODO (Bio-Graphics, Marseille, France) allowed deuterium sites to be identified and modeled into the structure. Refinement was then switched to SHELX-97 (35) which allowed incorporation

of stereochemical restraints for the deuterated and modified amino acids. *R*<sub>free</sub> was calculated from 5% of the reflection data which were omitted from the least-squares calculations. These rounds of refinement were interspersed with inspection of the maps to locate extra sites of deuteration and solvent molecules, including D<sub>2</sub>O sites. During graphical rebuilding, the neutron structure of the complex was checked to ensure that significant deviations from the room-temperature X-ray structure of the same complex (33) had not occurred. During refinement, the coordinates and isotropic *B*-factors for all atoms, including the hydrogen and deuterium atoms, were refined with stereochemical restraints. Deuterium sites at the catalytic center were located by inspection of difference Fourier maps weighted by the sigma-A method (36). In the refinement, a working high-resolution cutoff of 2.1 Å was applied since inclusion of data beyond this was found to raise the refinement *R*-factor by several percent. This may stem from the fact that the second of the two crystals did not diffract quite as well as the first at high resolution. To test the locations of the active site deuteriums, the structure was refined with two different active site models involving the D atoms in two canonical forms. The resulting difference Fourier maps were inspected to assess the likelihood of each model being correct. Finally, the occupancies of the catalytic D atoms in both models were refined. The coordinates of the preferred model and structure factors have been deposited with the Protein Data Bank (code 1gkt).

## RESULTS AND DISCUSSION

The final neutron data set for the H261 complex comprised 102 312 reflections where *I* > 2.0σ(*I*). These were reduced to 13 738 unique reflections to 2.1 Å resolution with a merging *R*-factor of 7.5% and a completeness of 84.5%. Full statistics for the data processing and refinement are given in Table 1. The structure contains 4674 protein atoms (including hydrogens and deuteriums), 159 inhibitor atoms, and 256 solvent sites; 41 of these solvent sites have well-defined density for D<sub>2</sub>O molecules and were therefore modeled as D<sub>2</sub>O. While it might be expected that more D<sub>2</sub>O sites would have been apparent at this resolution, the majority of solvent density peaks were spherical and could be refined satisfactorily as O atoms as is usual in X-ray studies. The neutron visibility of the deuterium atoms in solvent molecules would be reduced if they have high temperature factors. H<sub>2</sub>O molecules scatter neutrons very weakly due to cancellation of the H and O scattering lengths which have opposite signs. Thus, the solvent sites which appear to only have density for the central O atom are likely to be orientationally disordered D<sub>2</sub>O molecules. The majority of the well-defined D<sub>2</sub>O molecules are close to the surface of the protein, and some are partially buried within the protein.

The final neutron refinement *R*-factor is 23.5% (27.4% for *R*<sub>free</sub>). These *R* values are slightly high in comparison with the values expected for a refined X-ray structure at comparable resolution. However, they are consistent with *R* values obtained in other analyses which have used neutron Laue data (22–25). The neutron structure superimposes very well with the previously determined X-ray structure of this complex (33); the rms deviation between the two is only 0.2 Å for all Cα atoms. While deuterium sites in the protein generally are very well defined by the map, the neutron density for hydrogen atoms, which appear as negative

Table 1: Neutron Data and Refinement Statistics for the Endothiapepsin–H261 Complex

unit cell	
<i>a</i> (Å)	43.1
<i>b</i> (Å)	75.7
<i>c</i> (Å)	42.9
$\beta$ (deg)	97.0
space group	$P2_1$
no. of crystals	2
entire data set	
resolution range (Å)	100.0–2.1
$R_{\text{merge}}$ (%)	7.5
completeness (%)	84.5
multiplicity	3.4
mean $I/\sigma(I)$	5.4
outer shell	
resolution range (Å)	2.2–2.1
$R_{\text{merge}}$ (%)	11.9
completeness (%)	72.6
multiplicity	2.4
mean $I/\sigma(I)$	1.7
refinement (preferred model)	
resolution range (Å)	20.0–2.1
no. of reflections	13548
overall $R$ -factor (%)	23.46
overall $R_{\text{free}}$ (%)	27.42
outer shell $R$ -factor (%)	33.10
rmsd for bond lengths (Å)	0.009
rmsd for bond angle distances (Å)	0.014
rmsd for bumps (Å)	0.051
rmsd for chiral volumes (Å <sup>3</sup> )	0.027
rmsd for planes (Å)	0.023
no. of atoms in asymmetric unit	
C	1569
N	379
O	772
S	2
H	2055
D	393
total	5170

features in the  $2F_o - F_c$  map, was found to be weaker and harder to interpret. However, this is to be expected since the scattering amplitude of hydrogen is approximately half that of deuterium in addition to being of opposite sign. For this reason, CH<sub>2</sub> and CH<sub>3</sub> groups commonly have weak density since the scattering lengths of the atoms involved tend to cancel each other out.

**Extent of Deuteration.** While deuteration of the crystals was undertaken primarily to reduce incoherent scatter from bulk solvent, the neutron data show that the endothiapepsin molecule itself has become extensively deuterated by the vapor diffusion protocol. The low pH of the mother liquor (pH 4.5) would not be favorable for H–D exchange of the groups on the protein. Indeed, in NMR experiments, lowering the pH is used to deliberately quench H–D exchange, for example, in time-resolved studies of protein folding. However, the length of time involved in the H–D exchange of the endothiapepsin crystals (2–3 months) appears to have allowed it to proceed to a reasonable extent. The extent of main chain deuteration is shown in Figure 3. The majority of amino acids have become deuterated either in the side chain or in the main chain with the exposed secondary structure elements being the most affected. In total, 215 of 330 residues (65%) have undergone side chain and/or main chain deuteration. In the backbone, a total of 161 main chain amides (49%) have exchanged. The parts of the molecule most protected from exchange are the buried  $\beta$ -strand regions. In contrast, the strands which have become most

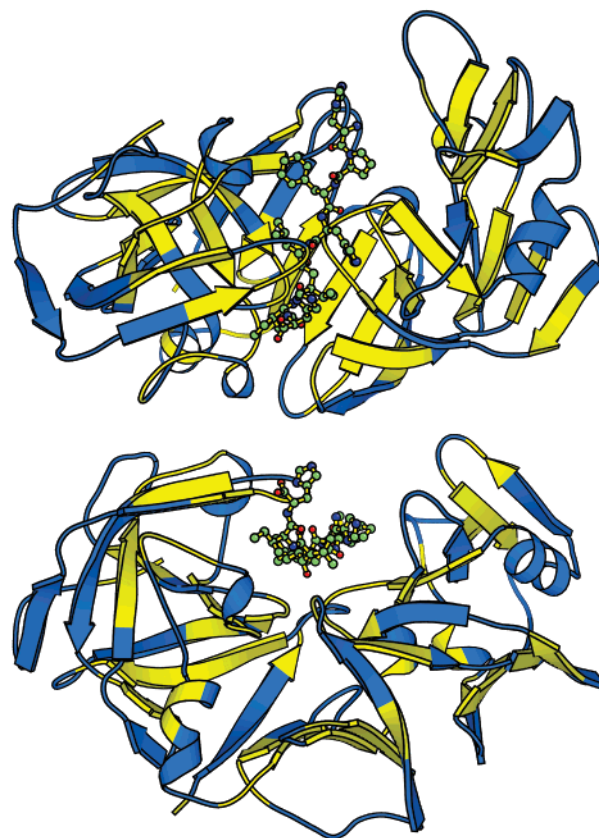


FIGURE 3: Extent of main chain deuteration of the endothiapepsin molecule in two orthogonal views. Yellow shows the regions which have not exchanged, whereas those regions shown in blue are segments where the amino acids have become deuterated in the main chain. In each view, the inhibitor (H261) can be seen occupying the active site cleft.

deuterated are generally those at the exposed edges of the sheets. The majority of polar side chains have exchanged, and in general, the loops and helical regions have exchanged to a greater extent than the  $\beta$ -sheet regions. Although the active site cleft is occupied by a tight-binding inhibitor, many residues in the vicinity of the catalytic center and specificity pockets of the enzyme have undergone H–D exchange. It is perhaps surprising that the protein has not become deuterated to a greater extent. However, it should be remembered that the crystals have a low solvent content (39%), and the efficacy of capillary vapor diffusion for H<sub>2</sub>O–D<sub>2</sub>O exchange is not well characterized as yet.

**Buried Carboxylates and the *pI* of the Enzyme.** A number of aspartic proteinases have *pI* values which are unusually low; e.g., pepsin has a *pI* value of  $\sim 2.0$  (*1*). It has been suggested that these low *pI* values are due to buried carboxylate groups which remain deprotonated even at very low pH. Evidence for this effect is provided by the neutron data since there are a number of buried carboxylates which appear to be deprotonated. In general, the carboxylate groups of these residues interact with buried main chain  $>NH$  groups or polar side chain atoms. While a number of these make salt bridge interactions (i.e., they will not contribute to the net charge of the molecule), some do not and instead interact with neutral polar groups which may function to stabilize the negative charge on the carboxylates. A definitive example of a buried, negatively charged carboxylate is Asp 87 which is an almost completely conserved residue. The

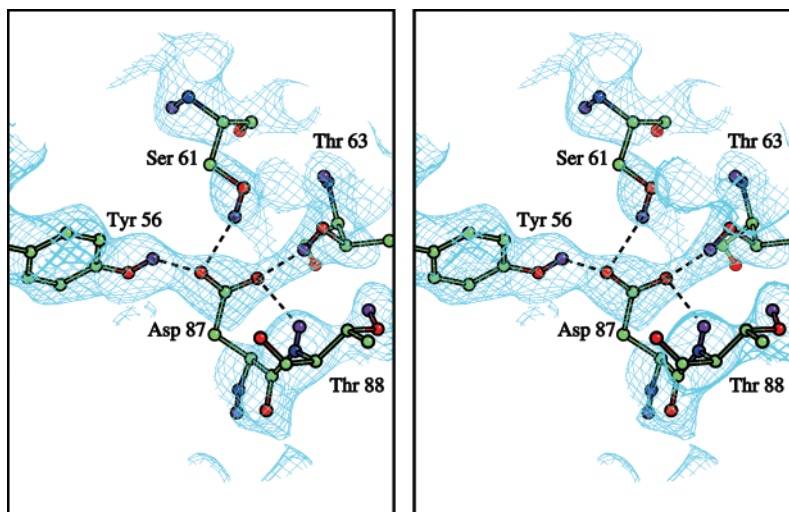


FIGURE 4: A  $2F_o - F_c$  density map (cyan) for residues in the vicinity of Asp 87 contoured at 1.0 rms. The map shows that the groups hydrogen bonding with the carboxyl of this residue are all deuterated, confirming that the aspartate is deprotonated. Deuterium atoms are shown in purple, and dashed lines represent hydrogen bonds.

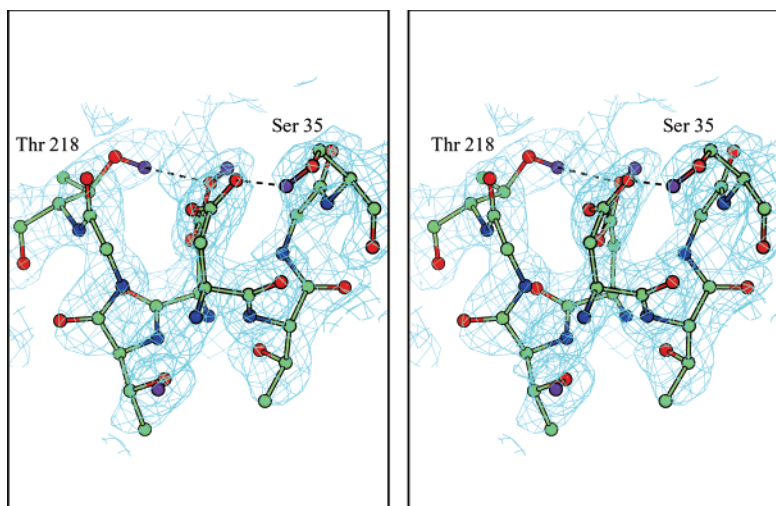


FIGURE 5: A  $2F_o - F_c$  density map (cyan) for residues in the vicinity of the catalytic aspartates contoured at 1.2 rms. The active site aspartate residues are shown in the middle with Asp 32 in the foreground and Asp 215 behind. The two residues which form hydrogen bonds with the outer oxygens of the carboxyl diad (Ser 35 and Thr 218) are also shown with the hydrogen bonds represented as dashed lines. The deuterium atoms are colored purple.

side chain of this residue accepts hydrogen bonds from the side chains of Tyr 56, Ser 61, and Thr 63 as well as the main chain >NH group of Thr 88. The neutron density clearly shows that these four groups are deuterated (Figure 4) which strongly indicates that the aspartate is negatively charged. Of the residues making these interactions, the Ser and Thr side chains are strongly conserved; the tyrosine residue is also conserved but to a lesser extent.

**Deuteration of the Active Site Residues.** The residues forming the catalytic center of aspartic proteinases are two strongly conserved Asp-Thr-Gly-Ser/Thr sequences. These are provided by the two domains of the enzyme where they associate to form the active site cleft. The two aspartate carboxyls are involved in numerous hydrogen bonds which keep them approximately coplanar. The outer oxygens of the aspartate diad accept hydrogen bonds from the Ser/Thr side chains in the above consensus sequence. It is clear from inspection of the neutron maps that these Ser and Thr residues (Ser 35 and Thr 218) have deuterated side chains despite having very low solvent accessibility due to the

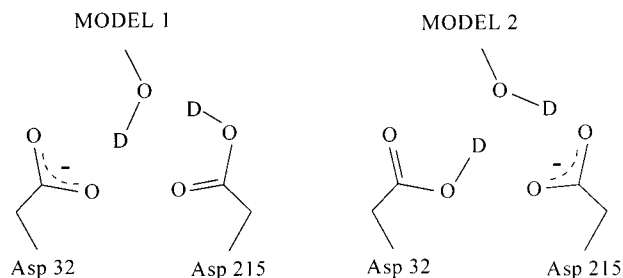


FIGURE 6: Possible proton (or deuterium) positions at the catalytic center. These models were tested by refinement with the neutron data. Model 1 assumes that Asp 32 is negatively charged, whereas model 2 assumes that Asp 215 is the negatively charged residue.

presence of the inhibitor in the active site. The refined neutron structure and maps for the active site are shown in Figure 5. Since the active site flap and the inhibitor help to shield the catalytic residues from the surrounding solvent, the fact that buried active site residues have become deuterated may indicate that the network of hydrogen bonds

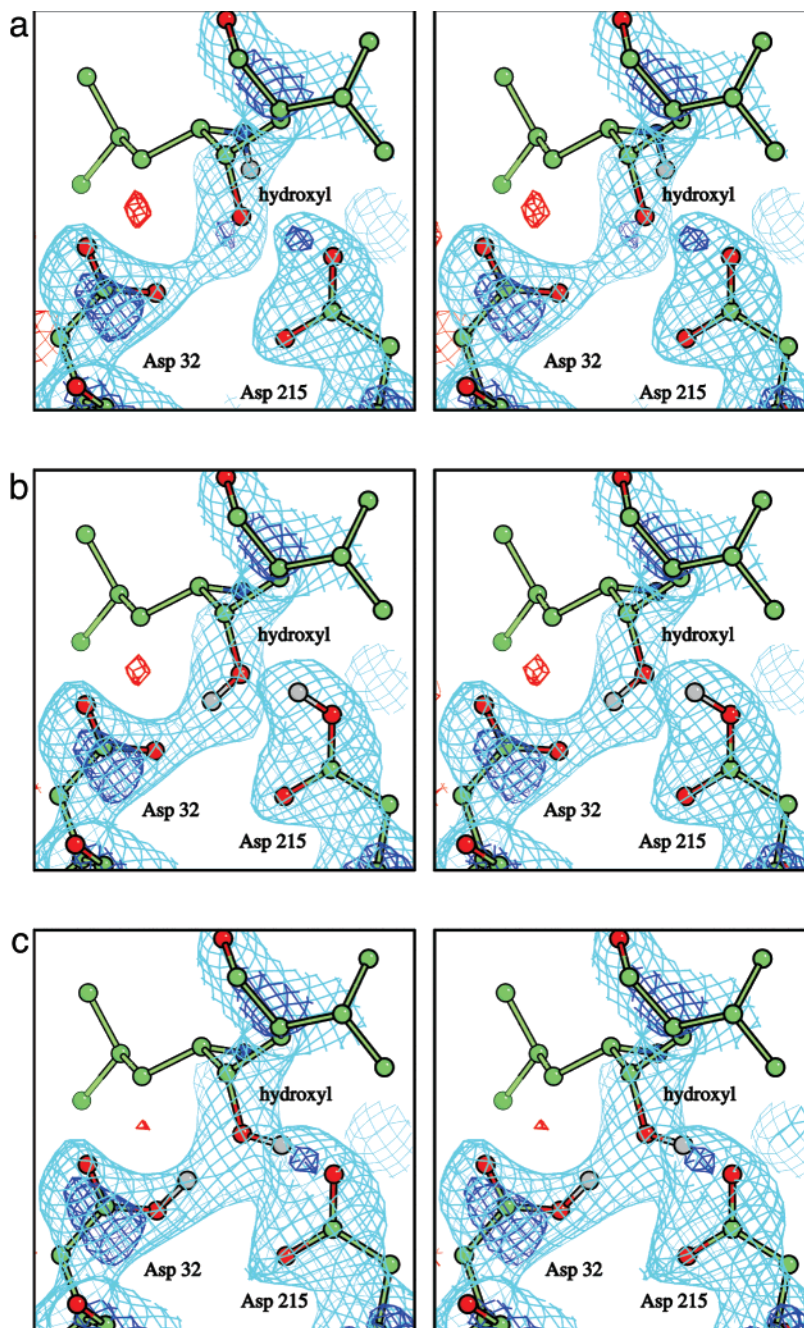


FIGURE 7: Neutron density at the catalytic center. Asp 32 and 215 are shown along with the transition state isostere. (a) Original difference density ( $F_o - F_c$ ) in dark blue and red lines which correspond to 2.5 and  $-2.5$  rms, respectively. The  $2F_o - F_c$  density is shown in cyan contoured at 1.2 rms. The regions of positive difference density between the aspartates and ligand strongly suggest that the inhibitor hydroxyl has a D substituent oriented toward Asp 32 and that Asp 215 is deuterated on its outer carboxyl oxygen. The region of negative difference density close to Asp 32 cannot be interpreted easily since a proton located here would not make hydrogen bonds with other groups. The refined neutron density maps for active site models 1 and 2 are shown in panels b and c, respectively, with the same contour levels as shown for panel a. Asp 32 and 215 are shown as is the hydroxyl of the transition state isostere. The deuterium atoms at the catalytic center are shown as gray spheres. Some parts of the inhibitor appear to be without density, but this is only because a narrow slab of each map is displayed for clarity. The best fit is provided by model 1, shown in panel b.

in the active site cleft provides a means of exchanging protons with the bulk solvent.

**Catalytic Protons.** The inhibitor H261 possesses a hydroxyethylene analogue ( $-\text{CHOH}-\text{CH}_2-$ ) in place of the scissile peptide bond (see Figure 2). This analogue mimics one hydroxyl of the putative tetrahedral intermediate (see Figure 1). X-ray structural studies of the complex of H261 with endothiapepsin (33) as well as many similar inhibitor complexes at high resolution (37) showed that this hydroxyl replaces the water molecule found at the catalytic center of

the native enzyme. The hydroxyl appears to form short hydrogen bonds with the inner carboxyl oxygen of Asp 32 and the outer carboxyl oxygen of Asp 215, indicating that protons or deuterons must reside between these atoms. This would allow them to form hydrogen bonds in one of two possible tautomeric forms shown in Figure 6. The other two possible hydrogen bonds to the hydroxyl group involve the outer carboxyl oxygen of Asp 32 and the inner oxygen of Asp 215, but both have unfavorable geometry (e.g., donor-acceptor distances that are too long). These findings were

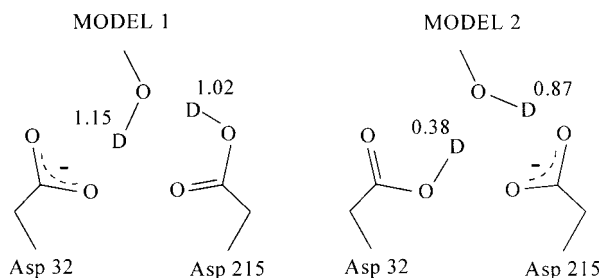


FIGURE 8: Two models for the proton positions at the active site showing the refined occupancies for the D atoms. The occupancies indicate that model 1 is most likely to be the correct one.

confirmed by recent atomic resolution X-ray studies of the same complex at 1.1 Å resolution (P. Erskine, unpublished observations).

In the deuterated H261 complex, there is significant evidence that residues in the immediate vicinity of the catalytic center have undergone H–D exchange (see above). A difference Fourier map for the structure without hydrogens or deuteriums modeled at the catalytic center (Figure 7a) shows positive difference density for deuterium atoms on the inhibitor hydroxyl and the outer carboxyl of Asp 215.

Since the neutron difference density for the deuterium sites at the catalytic center seemed to bridge the gaps between the inhibitor and the aspartates, it was considered necessary to investigate different possible locations for the deuteriums. Therefore, the structure was refined using the two separate models of the deuterium positions shown in Figure 6, and the resulting maps were inspected. In model 1, Asp 32 is treated as being negatively charged, whereas in model 2, Asp 215 is treated as the negatively charged aspartate. From inspection of the initial neutron difference Fourier maps, model 1 was considered to be the most likely scenario. In the neutron refinement of these two models, the occupancies of all atoms, including deuteriums, were fixed at unity. The refined neutron maps ( $2F_o - F_c$  and  $F_o - F_c$ ) for models 1 and 2 are shown in panels b and c of Figure 7, respectively. From inspection of these maps, it is clear that model 1 provides a qualitatively better fit to the data. In contrast, the deuterium atoms in model 2 are at nodes in the  $2F_o - F_c$  density, and a feature of positive difference density remains close to the outer oxygen of Asp 215, indicating that this residue is deuterated. These results suggest that Asp 215 is protonated in the transition state complex which is in accord with the earlier proposal for the catalytic mechanism based on X-ray analysis at 2.0 Å resolution of a fluoroketone inhibitor (6). In this mechanism, Asp 32 was designated as the negatively charged aspartate in the transition state complex since its hydrogen bonding capacity is better satisfied than that of Asp 215.

Finally, refinement of the occupancies of just the active site deuteriums was undertaken for the two models, and the results are shown in Figure 8. The refined occupancies indicate that model 1 is more consistent with the data than model 2 which is in accord with our initial interpretation of the original difference Fourier. This is also corroborated by atomic resolution X-ray analysis of the H261 complex at 100 K in which the pattern of aspartate C–O bond lengths obtained by unrestrained refinement is in accord with model 1 (P. Erskine, unpublished observations), although the protein

was not deuterated for this study. The ability of the neutron Laue data to discriminate between the models indicates that the data are of high quality. In accord with the recent findings of other groups (38), our results suggest that this technique will be of much value in other mechanistic studies.

The results of this study indicate that model 1 is the most likely scenario since there are strong indications that Asp 215 is deuterated. However, it may be significant that occupancy refinement with the neutron data does not discriminate between models 1 and 2 as well as might be hoped. This may reflect on properties of the catalytic diad itself. The donor–acceptor distances for the hydrogen bonds made between the inhibitor hydroxyl and the catalytic aspartates are very short ( $\sim 2.6$  Å). This was suggested by many previous X-ray studies at  $\sim 2.0$  Å resolution (37) and has recently been corroborated by the atomic resolution data and NMR (L. Coates, P. Erskine, and M. P. Crump, manuscript in preparation). The hydrogen bond involving Asp 32 is consistently the shortest ( $< 2.6$  Å). Hydrogen bonds as short at 2.5–2.6 Å are termed low-barrier hydrogen bonds since the proximity of the donor and acceptor atoms reduces the energy barrier which normally prevents transfer of the hydrogen atom from the donor to the acceptor group (12). Thus, rapid exchange of the proton between the donor and acceptor atoms can occur, and this has been proposed as an important effect in the catalytic mechanisms of a number of enzymes, including citrate synthase and the serine proteinases. Thus, we propose that some of the difficulty in determining the exact positions of the active site D atoms is due to formation of a low-barrier hydrogen bond between the hydroxyl of the inhibitor and the carboxyl of Asp 32. While our neutron results indicate strongly that Asp 215 is deuterated in the transition state analogue complex, the formation of a low-barrier hydrogen bond between the inhibitor hydroxyl and Asp 32 would tend to delocalize the neutron density between these two groups. Atomic resolution X-ray studies and NMR indicate that the short hydrogen bonds are absent in the native structure where a water molecule is bound to both carboxyls. In accord with studies of the serine proteinases, this indicates that low-barrier hydrogen bond formation is due to steric compression upon inhibitor binding.

## ACKNOWLEDGMENT

The Institut Laue-Langevin is thanked for the provision of neutron beam time and for support during data collection. We are also grateful to Professor T. L. Blundell, the late Dr. V. Dhanaraj, and Professor J. R. Helliwell for many useful discussions and Professor M. Szelke for provision of the inhibitor (H261).

## REFERENCES

1. Davies, D. R. (1990) The structure and function of the aspartic proteinases, *Annu. Rev. Biophys. Biophys. Chem.* 19, 189–215.
2. Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A., and Blundell, T. L. (1978) Structural evidence for gene duplication in the evolution of acid proteases, *Nature* 271, 618–621.
3. Pearl, L. H., and Blundell, T. L. (1984) The active site of aspartic proteinases, *FEBS Lett.* 174, 96–101.
4. Hofmann, T., Dunn, B. M., and Fink, A. L. (1984) Cryoenzymology of penicillopepsin. appendix: mechanism of action of aspartyl proteinases, *Biochemistry* 23, 5247–5256.



5. Rich, D. H., Bernatowicz, M. S., and Schmidt, P. G. (1982) Direct  $^{13}\text{C}$  NMR evidence for a tetrahedral intermediate in the binding of a pepstatin analogue to porcine pepsin, *J. Am. Chem. Soc.* *104*, 3535–3536.
6. Veerapandian, B., Cooper, J. B., Šali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., and Hoover, D. J. (1992) Direct observation by X-ray analysis of the tetrahedral “intermediate” of aspartic proteinases, *Protein Sci.* *1*, 322–328.
7. Rodrigues, E. J., Angeles, T. S., and Meek, T. D. (1993) Use of nitrogen-15 kinetic isotope effects to elucidate details of the chemical mechanism of human immunodeficiency virus protease, *Biochemistry* *32*, 12380–12385.
8. Li, M., Phylip, L., Lees, W., Winther, J., Dunn, B., Wlodawer, A., Kay, J., and Gustchina, A. (2000) The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix, *Nat. Struct. Biol.* *7*, 113–117.
9. Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D., and Davies, D. R. (1987) Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: implications for a mechanism of action, *Proc. Natl. Acad. Sci. U.S.A.* *84*, 7009–7013.
10. James, M. N. G., Sielecki, A. R., Hayakawa, K., and Gelb, M. H. (1992) Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides, *Biochemistry* *31*, 3872–3886.
11. Silva, A. M., Cachau, R. E., Sham, H. L., and Erickson, J. W. (1996) Inhibition and catalytic mechanism of HIV-1 aspartic protease, *J. Mol. Biol.* *255*, 321–340.
12. Cleland, W. W., Frey, P. A., and Gerlt, J. A. (1998) The low-barrier hydrogen bond in enzymatic catalysis, *J. Biol. Chem.* *273*, 25529–25532.
13. Wilson, C. C. (2001) Hydrogen atoms in acetylsalicylic acid (Aspirin): the librating methyl group and probing the potential well in the hydrogen bonded dimer, *Chem. Phys. Lett.* *335*, 57–63.
14. Burmeister, W. P. (2000) Structural changes in a cryo-cooled protein crystal owing to radiation damage, *Acta Crystallogr. D56*, 328–341.
15. Ravelli, R. B. G., and McSweeney, S. M. (2000) The fingerprint that X-rays leave on structures, *Structure* *8*, 315–328.
16. Schoenborn, B. (1994) *Neutrons in Biology*, Basic Life Sciences, Vol. 27, Plenum, New York.
17. Kossiakoff, A. A., and Spencer, S. A. (1980) Neutron diffraction identifies His 57 as the catalytic base in trypsin, *Nature* *288*, 414–416.
18. Mason, S., Bentley, G. A., and McIntyre, G. (1984) Deuterium exchange in lysozyme at 1.4 Å resolution, in *Neutrons in Biology* (Schoenborn, B., Ed.) pp 323–334, Plenum, New York.
19. Wlodawer, A., Miller, M., and Sjolín, L. (1983) Active site of RNase: neutron diffraction study of a complex with uridine vanadate, a transition state analogue, *Proc. Natl. Acad. Sci. U.S.A.* *80*, 3628–3631.
20. Cipriani, F., Castagna, J. C., Wilkinson, C., Oleinek, P., and Lehmann, M. S. (1996) Cold neutron protein crystallography using a large position-sensitive detector based on image-plate technology, *J. Neutron Res.* *4*, 79–85.
21. Myles, D. A. A., Bon, C., Langan, P., Cipriani, F., Castagna, J. C., Lehmann, M. S., and Wilkinson, C. (1998) Neutron Laue diffraction in macromolecular crystallography, *Physica B* *241*, 1122–1130.
22. Niimura, N., Minezaki, Y., Nonaka, T., Castagna, J. C., Cipriani, F., Hoghoj, P., Lehmann, M. S., and Wilkinson, C. (1997) Neutron Laue diffractometry with an imaging plate provides an effective data collection regime for neutron protein crystallography, *Nat. Struct. Biol.* *4*, 909–914.
23. Habash, J., Raftery, J., Weisgerber, S., Cassetta, A., Lehmann, M. S., Hoghoj, P., Wilkinson, C., Campbell, J. W., and Helliwell, J. R. (1997) Neutron Laue diffraction study of concanavalin A: the proton of Asp 28, *J. Chem. Soc., Faraday Trans. 93*, 4313–4317.
24. Bon, C., Lehmann, M. S., and Wilkinson, C. (1999) Quasi-Laue neutron diffraction study of the water arrangement in crystals of triclinic lysozyme from hen egg-white, *Acta Crystallogr. D55*, 978–987.
25. Habash, J., Raftery, J., Nuttall, R., Price, H. J., Wilkinson, C., Kalb (Gilboa), A. J., and Helliwell, J. R. (2000) Direct determination of the positions of the deuterium atoms of the bound water in concanavalin A by neutron Laue crystallography, *Acta Crystallogr. D56*, 541–550.
26. Moews, P., and Bunn, C. W. (1970) An X-ray crystallographic study of the rennin-like enzyme of *Endothia parasitica*, *J. Mol. Biol.* *54*, 395–397.
27. Myles, D. A. A., and Cooper, J. B. (2000) A preliminary neutron Laue diffraction study of the aspartic proteinase endothiapepsin, *Acta Crystallogr. D56*, 246–248.
28. Campbell, J. W. (1995) LAUEGEN: an X-windows based program for the processing of Laue X-ray diffraction data, *J. Appl. Crystallogr.* *28*, 228–236.
29. Campbell, J. W., Hao, Q., Harding, M. M., Nguti, N. D., and Wilkinson, C. (1998) LAUEGEN version 6.0 and INTLDM, *J. Appl. Crystallogr.* *31*, 496–502.
30. Wilkinson, C., Khamis, H. W., Stansfield, R. F. D., and McIntyre, G. J. (1988) Integration of single-crystal reflections using area multidetectors, *J. Appl. Crystallogr.* *21*, 471–478.
31. Prince, E., Wilkinson, C., and McIntyre, G. J. (1997) Comparison of the  $\sigma(I)/I$  and least-squares methods for integration of Bragg reflections, *J. Appl. Crystallogr.* *30*, 133–137.
32. Collaborative Computational Project No. 4 (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr. D50*, 760–763.
33. Veerapandian, B., Cooper, J., Šali, A., and Blundell, T. L. (1990) Three-dimensional structure of endothiapepsin complexed with a transition-state isostere inhibitor of renin at 1.6 Å resolution, *J. Mol. Biol.* *216*, 1017–1029.
34. Brunger, A. T. (1987) *X-PLOR: A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.
35. Sheldrick, G. M. (1998) High-resolution refinement, in *Crystallographic Computing* (Watenpaugh, K., and Bourne, P. E., Eds.) Vol. 7, Oxford University Press, Oxford, U.K.
36. Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors, *Acta Crystallogr. A42*, 140–149.
37. Bailey, D., and Cooper, J. B. (1994) A structural comparison of 21 inhibitor complexes of the aspartic proteinase from *Endothia parasitica*, *Protein Sci.* *3*, 2129–2143.
38. Helliwell, J. R. (1997) Neutron Laue diffraction does it faster, *Nat. Struct. Biol.* *4*, 874–876.