Structure–function relationships in the cysteine proteinases actinidin, papain and papaya proteinase Ω

Three-dimensional structure of papaya proteinase Ω deduced by knowledge-based modelling and active-centre characteristics determined by two-hydronic-state reactivity probe kinetics and kinetics of catalysis

Christopher M. TOPHAM,* Erdjan SALIH,†§ Carlos FRAZAO,*|| Devanand KOWLESSUR,† John P. OVERINGTON,* Mark THOMAS,† Simon M. BROCKLEHURST,*¶ Manij PATEL,† Emrys W. THOMAS† and Keith BROCKLEHURST†

*Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, U.K., †Department of Biochemistry, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, U.K. and ‡Department of Biological Sciences, University of Salford, Salford M5 4WT, U.K.

1. A model of the three-dimensional structure of papaya proteinase Ω , the most basic cysteine proteinase component of the latex of papaya (*Carica papaya*), was built from its amino acid sequence and the two currently known high-resolution crystal structures of the homologous enzymes papain (EC 3.4.22.2) and actinidin (EC 3.4.22.14). The method used a knowledge-based approach incorporated in the COMPOSER suite of programs and refinement by using the interactive graphics program FRODO on an Evans and Sutherland PS 390 and by energy minimization using the GROMOS program library. 2. Functional similarities and differences between the three cysteine proteinases revealed by analysis of pH-dependent kinetics of the acylation process of the catalytic act and of the reactions of the enzyme catalytic sites with substrate-derived 2-pyridyl disulphides as two-hydronic-state reactivity probes are reported and discussed in terms of the knowledge-based model. 3. To facilitate analysis of complex pH-dependent kinetic data, a multitasking application program (SKETCHER) for parameter estimation by interactive manipulation of calculated curves and a simple method of writing down pH-dependent kinetic equations for reactions involving any number of reactive hydronic states by using information matrices were developed. 4. Papaya proteinase Ω differs from the other two enzymes in the ionization characteristics of the common (Cys)-SH/(His)-Im⁺H catalytic-site system and of the other acid/base groups that modulate thiol reactivity towards substrate-derived inhibitors and the acylation process of the catalytic act. The most marked difference in the Cys/His system is that the pK_a for the loss of the ion-pair state to form $-S^-/Im$ is 8.1–8.3 for papaya proteinase Ω , whereas it is 9.5 for both actinidin and papain. Papaya proteinase Ω is similar to actinidin in that it lacks the second catalytically influential group with pK_a approx. 4 present in papain and possesses a catalytically influential group with pK_a 5.5–6.0. 5. Papaya proteinase Ω occupies an intermediate position between actinidin and papain in the sensitivity with which hydrophobic interaction in the S_2 subsite is transmitted to produce changes in transition-state geometry in the catalytic site, a fact that may be linked with differences in specificity in P_2-S_2 interaction exhibited by the three enzymes. 6. The solvent-accessible areas of the binding sites in Ac-Ala-Ala-Phe-Ala-methylene-S-papain and in the free enzyme in the same conformation with water molecules removed in both cases were calculated. Residues losing more than 0.02 nm² (2 Å²) of solvent contact area on inhibitor binding are all conserved in the S_2 subsite of papaya proteinase Ω . The knowledge-based model of the structure of papaya proteinase Ω does not reveal any gross geometrical features in the S_1-S_2 intersubsite region or in the S_2 subsite that would readily account for the marked difference in its behaviour from that of papain. The most obvious structural differences involve the electrostatic fields to which the active centres are exposed. 7. The origins of the pK_a values additional to those of the catalytic site Cys–His interactive system that modulate in different ways the behaviour of actinidin, papain and papaya proteinase Ω remain to be discovered. The conservation in papaya proteinase Ω of the remote ion pair (Glu-52/His-81) that is present in papain but lacking in actinidin suggests that this feature is not responsible for the kinetically influential ionizations with pK_a values about 4 and 8 additional to those of the ion pair in papain (pK_a values about 4 and 9.5). The other feature remote from the catalytic-site regions (hydrogen-bonded Tyr-61/Tyr-67 in papain, non-hydrogen-bonded His-61/Tyr-67 in papaya proteinase Ω , and Arg-61/Tyr-67 in actinidin) might contribute to the different pH-dependent modulatory character of the three enzymes in alkaline media. Dehydration across pK_a 8.0–8.25 in papain decreases the catalytic competence and neighbouring group character of the $-S^-/Im^+H$ ion pair, dehydration with pK_a 9.5 increases the catalytic competence of the analogous ion pair in papaya proteinase Ω , and there appears to be no analogous modulatory group in actinidin. There are no obvious candidates for the other modulatory ionization with pK_a about 4 in papain, pK_a 5.5 in actinidin and pK_a 5.8 in papaya proteinase Ω . The new evidence discussed in the present paper now that a papaya proteinase Ω structure is available suggests that the previous assignment of the kinetically influential pK_a of 5.5 in actinidin to Asp-161, deduced from the structures of actinidin and papain, is probably incorrect and a different assignment needs to be sought.

Abbreviations used: L-Bz-Arg-Nan, N^ω-benzoyl-L-arginine *p*-nitroanilide; $-S^-/Im^+H$, thiolate/imidazolium; RMS, root mean square; Ac, acetyl.

§ Present address: Department of Chemistry, Brandeis University, Edison-Lecks 211, Waltham, MA 02254, U.S.A.

|| Present address: Centro de Tecnologia Química e Biológica, Ruada Quinta Grande 6, Apartado 127, 2780 Oeiras, Portugal.

¶ Present address: Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.

INTRODUCTION

The natural structural variation that exists in the cysteine proteinases ([1]; and see [2–5] for reviews on these enzymes) provides opportunities for investigation of the consequences of different electrostatic effects and of different binding interactions for active-centre chemistry and catalytic mechanism (see e.g. [6,7]). Initial reports on a number of relatively little-studied members of the cysteine-proteinase family (see [3] and e.g. [8,9]) suggest behaviour considerably different from that of papain (EC 3.4.22.2), the putative cysteine-proteinase archetype. The only high-resolution crystal structures of members of this enzyme family that exist currently are those of papain [0.165 nm (1.65 Å)] [10] and actinidin (EC 3.4.22.14) [0.17 nm (1.7 Å)] [11], which differ in active-centre behaviour (see e.g. [6,7]) despite the closely similar structures. As part of a programme of co-ordinated structural and structure-related kinetic studies on the cysteine proteinases, we are extending the availability of three-dimensional structures of members of this enzyme family (for stem bromelain [12]; for chymopapain B [13]; for chymopapain M (papaya proteinase IV) [14]) by using a knowledge-based approach incorporated in the COMPOSER suite of programs [15–18], and refinement by interactive graphics and energy minimization. In the present work, COMPOSER and refinement by using FRODO [19,20] and the GROMOS program library [21] were used to build the three-dimensional structure of papaya proteinase Ω , the most basic, and relatively abundant, cysteine-proteinase component of the latex of papaya (*Carica papaya*). The resulting structure is compared with those of papain, the least basic, and minor, component of papaya latex, and actinidin. Many of the active-centre features common to papain and actinidin are conserved in papaya proteinase Ω . The most obvious differences between the three enzymes involve the electrostatic fields to which the three active centres are exposed and these differences might account for the gradation in the effectiveness of binding-site catalytic signalling here reported, with papain > papaya proteinase Ω > actinidin possibly in terms of variation in permitted conformational flexibility. Functional similarities and differences between papaya proteinase Ω , papain and actinidin revealed by analysis of pH-dependent kinetics including those of the reactions of these enzymes with substrate-derived 2-pyridyl disulphides as two-hydronic-state reactivity probes are reported and discussed in terms of the knowledge-based model. The analysis of pH-dependent kinetic data is facilitated by the development during the present work of SKETCHER, a multitasking application program for parameter estimation by interactive manipulation of calculated curves, and of a simple method of writing down pH-dependent kinetic equations for reactions involving any number of reactive hydronic states by using information matrices. Preliminary accounts of parts of this work have been reported [12,22].

MATERIALS AND METHODS

Papaya proteinase Ω

The following procedure produces not only papaya proteinase Ω , the most basic cysteine proteinase of papaya latex which is contained in the last activity peak of the elution profile, but also individual forms of the chymopapains. Spray-dried papaya latex of Indian origin [Powell and Scholefield Ltd., Liverpool, U.K. (currently J. A. Scholefield and Co., Runcorn, Cheshire, U.K.)] (15 g) was ground with sand (5 g) and Celite (5 g) for 30 min in 100 ml of 0.1 M-sodium acetate buffer, pH 5.0, containing 0.1 M-NaCl/1 mM-EDTA.

Insoluble material was then removed by vacuum filtration (Whatman type 54 filter paper). The filtrate was brought to 45 %

saturation with $(\text{NH}_4)_2\text{SO}_4$ (0.277 g/ml). The resulting precipitate was isolated by centrifugation at 4 °C for 30 min at 18 000 rev./min (40 000 g). The supernatant was dialysed against 3 × 5 litres of 0.1 M-sodium acetate buffer, pH 5.0, containing 1 mM-EDTA and 0.1 M-NaCl. The dialysed protein was then applied to a column (50 cm × 2.5 cm) of the cation-exchanger (SPC50) which had been equilibrated in the starting buffer. The column was washed with approximately 300 ml of the starting buffer until all of the unbound protein had been eluted. Multiple forms of chymopapain A were eluted from the column by using a linear gradient in which 1 litre of 0.1 M-sodium acetate buffer, pH 5.0, containing 1 mM-EDTA and 0.1 M-NaCl was mixed with 1 litre of 0.1 M-sodium acetate buffer containing 1 mM-EDTA and 0.25 M-NaCl; 15 ml fractions were collected. A second steeper gradient was then set up on the same column as follows: 1 litre of sodium acetate buffer, pH 5.0, containing 1 mM-EDTA and 0.25 M-NaCl was mixed with 1 litre of 0.1 M-sodium acetate buffer, pH 5.0, containing 1 mM-EDTA and 0.6 M-NaCl; 15 ml fractions were collected. A_{280} was monitored automatically during the elution and the fractions were subsequently analysed for catalytic activity towards N^α -benzoyl-L-arginine *p*-nitroanilide (L-Bz-Arg-Nan). The second gradient produced multiple forms of chymopapain B and finally papaya proteinase Ω . Fully active papaya proteinase Ω containing 1 mol of thiol per mol of protein was produced by using covalent chromatography by thiol-disulphide interchange [23,24] as described below. The pooled fractions containing papaya proteinase Ω were mixed with approximately 300 g of swollen agarose–glutathione–2-pyridyl disulphide gel which had been pre-equilibrated in 0.1 M-sodium acetate buffer, pH 4.0, containing 1 mM-EDTA and 0.3 M-NaCl. The gel was stirred gently for 1 h, isolated on a sintered glass funnel, and then washed with 4 litres of 0.1 M-sodium acetate buffer containing 1 mM-EDTA and 0.3 M-NaCl until values of both A_{280} and A_{343} fell to less than 0.03. The gel was then equilibrated with 4 litres of 0.1 M-Tris/HCl buffer, pH 8.0, containing 1 mM-EDTA and 0.3 M-NaCl. The gel was isolated by vacuum filtration and then mixed with 20 mM-cysteine in 0.1 M-Tris/HCl buffer, pH 8.0, containing 1 mM-EDTA and 0.3 M-NaCl. The gel was stirred gently for 2 h, collected by vacuum filtration and washed with a further 100 ml of 0.1 M-Tris/HCl buffer, pH 8.0, containing 1 mM-EDTA and 0.3 M-NaCl. The filtrate resulting from the cysteine treatment and subsequent wash was brought to 85 % saturation with respect to $(\text{NH}_4)_2\text{SO}_4$ (0.61 g/ml) at room temperature. The resulting precipitate was isolated by centrifugation at 40 200 g for 1 h at 4 °C and the pellet was redissolved in the minimum volume of phosphate buffer, pH 6.0, 1.0 mol/l containing 1 mM-EDTA and saturated with 2,2'-dipyridyl disulphide. The solution of reversibly inactivated enzyme (enzyme–S–S–2-Py) was stored in solution at 4 °C until required for kinetic studies. Then a small volume (usually 0.5 ml) was re-activated by addition of an equal volume of 20 mM-cysteine in Tris/HCl buffer, pH 8.0, 1.0 mol/l. After the solution had been allowed to stand at room temperature for 30 min, low- M_r material was removed by gel filtration on a column (2.5 cm × 40 cm) of Sephadex G-25 to produce fully active activator-free enzyme (approx. 20 ml of 20 μM). Papaya proteinase Ω protein was quantified by using A_{280} and $\epsilon_{280} = 5.6 \times 10^4 \text{ M}^{-1} \cdot \text{cm}^{-1}$ [25], total thiol content by titration at 343 nm ($\Delta\epsilon_{343} = 8.08 \times 10^3 \text{ M}^{-1} \cdot \text{cm}^{-1}$, [26]) with 2,2'-dipyridyl disulphide at pH 8.0 and intact catalytic-site content by an analogous titration at pH 4.0 (see [27] for a description of 2,2'-dipyridyl disulphide as a titrant of intact cysteine proteinase catalytic sites). Actinidin was isolated as described in [28].

Substrate and reactivity probes

The sources or syntheses of these compounds and most of the

procedures for using them in the study of cysteine proteinases have been described previously: L-Bz-Arg-Nan, methyl 2-pyridyl disulphide and 2,2'-dipyridyl disulphide (V) [29]; ethyl 2-pyridyl disulphide [30]; 2-(acetamido)ethyl 2'-pyridyl disulphide (II) and 2-(*N'*-acetyl-L-phenylalanyl)ethyl 2'-pyridyl disulphide (III) [31]; 2-(*N'*-acetyl-L-phenylalanyl)hydroxyethyl 2'-pyridyl disulphide (IV) [32].

Reactivity probe kinetics

Most of these reactions were studied at 25 °C and 1.0 mol/l in solutions containing 1 mM-EDTA with Dionex stopped-flow equipment connected to a Tektronix 5103N oscilloscope and interfaced with a Commodore 4032 microcomputer. Some kinetic runs in each set were recorded by photographing the oscilloscope screen, and conventional first-order analysis of these records always provided values of rate constants in good agreement with those produced by the on-line microcomputer. Release of pyridine-2-thione consequent on reaction of the enzyme thiol group with the 2-pyridyl disulphide probe reagents was monitored at 343 nm and quantified by using $\epsilon_{343} = 8.0 \times 10^3 / (1 + K/[H^+]) \text{ M}^{-1} \cdot \text{cm}^{-1}$, where $pK = 9.8$ [26]. The buffers, which were as described in [31], were prepared double strength such that dilution 1:2 during reaction in the mixing chamber gave 1.0 mol/l. All reactions were carried out under pseudo-first-order conditions with [enzyme] = 3–5 μM and [disulphide] = 50–150 μM in the mixing chamber of the stopped-flow machine. For all of the reactions, linear first-order logarithmic plots were obtained, and linear increase in the observed first-order rate constant with increase in [disulphide] at pH values around 4, 6 and 8 confirmed that, under the conditions of concentration used, the reactions obeyed second-order kinetics. A procedure similar in most respects to that described above was used to study the kinetics of the reaction of papaya proteinase Ω with methyl 2-pyridyl disulphide (I). The difference is that this reaction was studied by using an Applied Photophysics SF.17MV stopped-flow spectrophotometer, kinetic workstation and data-acquisition and analysis software.

Kinetics of catalysis

Catalytic activity was investigated by using L-Bz-Arg-Nan (Sigma Chemical Co.) and a continuous spectroscopic assay at 410 nm essentially as described for the ficin-catalysed hydrolysis of DL-Bz-Arg-Nan [33] using $\epsilon_{410} = 8.8 \times 10^3 \text{ M}^{-1} \cdot \text{cm}^{-1}$ [34] (which is essentially pH-independent at pH values above 3; the pK_a of the *p*-nitroanilinium cation is 1.0 [35]). When corrections are significant they are made by using

$$\epsilon_{410} = 8.8 \times 10^3 / (1 + 0.1[H^+]) \text{ M}^{-1} \cdot \text{cm}^{-1}$$

The stock solution of L-Bz-Arg-Nan was made up in water [containing 1% (v/v) dimethyl sulphoxide] rather than in buffer to facilitate pH-dependence studies in which buffer is added separately to the reaction mixtures in the kinetic runs. Buffers were prepared triple strength such that dilution 1:3 during reaction in a Cary 118C spectrophotometer gave 1.0 mol/l.

Evaluation of pH-dependent kinetic data

pH-dependent studies on reactions such as those investigated in the present work involving a multiplicity of reactive states require a rapid method of evaluating a variety of kinetic models differing in the number and characteristics of the reactive states. In the present work, a multitasking application program (SKETCHER) written in ANSI C [36] running under RISCOS on an Acorn Archimedes was developed to facilitate estimation of parameters in the generic set of equations for the various models by means of interactive manipulation of calculated curves. Values of the parameters thus obtained are generally close to the

final mean values obtained by weighted non-linear regression performed by using the AR computer program from the statistical software package BMDP [37] and a Compaq Deskpro 386/20e computer [22].

For a reaction with n reactive hydronic states, there are n pH-independent rate constants ($\tilde{k}$), and n or $(n+1)$ macroscopic pK_a values. For reactions of catalytic-site nucleophiles, there will normally be n pK_a values because the most hydron-deficient state will normally be nucleophilic. This contrasts with the situation for catalysis, where there will normally be $n+1$ pK_a values, because the analogous state will normally be inactive. Rate equations for reactions in any number of hydronic states may be written down or generated within the computer without the need for extensive algebraic manipulation. This is achieved by using a simple general expression, eqn. (1), and two information matrices (2) and (3).

$$k = \sum_{i=1}^n \frac{A_i}{\left(1 + \sum_{j=1}^n B_{i,j}\right)} \quad (1)$$

Eqn. (1) is a general expression that relates an experimentally determined pH-dependent rate constant k , and macroscopic acid dissociation constants K , where n = the number of reactive hydronic states; the numerator, $A_i (\equiv \tilde{k}_{x_{H_i}})$ is a generalized pH-independent rate constant where i (which may take values from 1 to n) specifies the particular hydronic state such that for the X -state, $i = 1$ and thus $\tilde{k}_{x_0} \equiv \tilde{k}_x$; the denominator:

$$\left(1 + \sum_{j=1}^n B_{i,j}\right)$$

provides for the pH-dependent variation in each of the contributions to k from reaction in a particular hydronic state, i.e. reaction associated with an individual pH-independent rate constant, $\tilde{k}$ [it provides the (extended) Michaelis pH function associated with reactivity in a given hydronic state]; each term generated by j in the denominator comprises a product of $[H^+]^m$ where m is an integer with $-n \leq m < +n$, and either an acid dissociation constant, K , or a product of two or more such constants.

The construction of specific kinetic equations for particular kinetic models from eqn. (1) involves determining expressions for $B_{i,j}$ by using two information matrices (2) and (3).

$$i = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \cdot \\ \cdot \\ \cdot \\ n \end{matrix} \quad j = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \dots n \\ \left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \dots n \\ -1 & 1 & 2 & 3 & 4 & 5 \dots n-1 \\ -2 & -1 & 1 & 2 & 3 & 4 \dots n-2 \\ -3 & -2 & -1 & 1 & 2 & 3 \dots n-3 \\ -4 & -3 & -2 & -1 & 1 & 2 \dots n-4 \\ -5 & -4 & -3 & -2 & -1 & 1 \dots n-5 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1-n & 2-n & 3-n & 4-n & 5-n & 6-n \dots 1 \end{array} \right) \end{matrix} \quad (2)$$

Matrix (2) provides values of m (i.e. powers of $[H^+]$) and matrix (3) the associated assemblies of macroscopic acid dissociation constants, each number p in a matrix element representing $K_{x_{H_p}}$, with a minus sign indicating a reciprocal K . For example, for $i = 3$ and $j = 1$, matrix (2) specifies a term in $1/[H^+]^2$ and matrix (3) a term containing $K_{x_{H_1}}K_{x_{H_2}}$. Thus this particular term in the denominator of the expression specified by j and associated with $\tilde{k}_{x_{H_3}}$ is $K_{x_{H_1}}K_{x_{H_2}}/[H^+]^2$.

Similarly for $i = 2$ and $j = 3$, matrix (2) specifies a term in $[H^+]^2$ and matrix (3) a term containing $1/K_{x_{H_1}}K_{x_{H_2}}$. For a reaction in four reactive hydronic states therefore the right hand side of eqn.

$$\begin{matrix}
 & j=1 & 2 & 3 & 4 & 5 & \dots n \\
 i=1 & -1 & -1.2 & -1.2.3 & -1.2.3.4 & -1.2.3.4.5 & \dots -1.2.3 \dots n \\
 2 & 1 & -2 & -2.3 & -2.3.4 & -2.3.4.5 & \dots -2.3.4 \dots n \\
 3 & 1.2 & 2 & -3 & -3.4 & -3.4.5 & \dots -3.4.5 \dots n \\
 4 & 1.2.3 & 2.3 & 3 & -4 & -4.5 & \dots -4.5.6 \dots n \\
 5 & 1.2.3.4 & 2.3.4 & 3.4 & 4 & -5 & \dots -5.6.7 \dots n \\
 6 & 1.2.3.4.5 & 2.3.4.5 & 3.4.5 & 4.5 & 5 & \dots -6.7.8 \dots n \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 n & 1.2.3 \dots n & 2.3.4 \dots n & 3.4.5 \dots n & 4.5.6 \dots n & 5.6.7 \dots n & \dots -n
 \end{matrix} \quad (3)$$

(1) is written as the sum of four terms each containing a numerator $\tilde{k}_{xH_3} - \tilde{k}_x$, and a denominator constructed by using matrices (2) and (3).

Knowledge-based model building of papaya proteinase Ω

The model of papaya proteinase Ω was built from its amino acid sequence [38] and the two currently known high-resolution crystal structures of the homologous enzymes, papain {0.165 nm (1.65 Å); [10]} and actinidin {0.17 nm (1.7 Å); [11]}, by using the COMPOSER suite of programs [15–18] and refined by using the interactive graphics program FRODO [19,20] on an Evans and Sutherland PS 390 and by energy minimization using the GROMOS program library [21]. Computations were performed on a Vax 11/750 or on a Micro Vax II computer.

The first stage of the procedure was the simultaneous rigid-body superposition of the three-dimensional structures of papain (9PAP) and actinidin (2ACT). The co-ordinates were obtained from the Brookhaven Data Bank [39], and an iteratively reweighted least-squares procedure was used to define a set of 199 topologically equivalent residues from an initial set of three residues in the catalytic-site region (Cys-25, His-159 and Asn-175; papain numbering) with a cut-off value of 0.35 nm (3.5 Å). This set was further reduced to the 174 residues shown in Fig. 1, which comprise the 'framework' or template for modelling. The reduction was achieved by peeling back the ends of the structurally conserved regions, where distances between topologically equivalent residues tend to the largest, and introducing breaks to avoid long contiguous stretches of framework. The papaya proteinase Ω sequence was then aligned to the template (see Fig. 1) in agreement with the alignment of Dubois *et al.* [38]. The framework was calculated as the weighted mean of each equivalent C_α position (Fig. 2a). The weights were a function of the sequence identity with papain (72.4%) and actinidin (52.3%) over the nine structurally conserved regions. Fragments of papain and actinidin that have the closest similarity (Table 1, column 1) to the corresponding section of the papaya proteinase Ω sequence were then least-squares fitted to the framework to give the model real protein geometry. The side chains were added using a set of environment-dependent rules [17] and the structurally conserved core region of the model was completed.

The next stage involved modelling the eight variable regions which connect the framework elements of secondary structure (see Table 1, columns 2 and 3, and Fig. 1). Protein fragments of the same length as the loops to be modelled and which are compatible with the geometry of a three-residue overlap with the core of the model and possess structural features compatible with the relevant papaya proteinase Ω loop sequence could all be found from the family of the two presently known structures. Loops of equivalent length in homologous proteins often have similar conformations [40]. The fragments were first identified by searching a local distance matrix database of all structures solved to a resolution of 0.2 nm (2 Å) or less in the Brookhaven Protein Data Bank (April 1990 release) including those of papain (9PAP) and actinidin (2ACT). The origins of the selected fragments are

detailed in Table 1. The five loops of three or more residues were annealed to the model core by using a novel ring closure algorithm (F. Eisenmenger, unpublished work) in which the main chain torsion angles of the loop are tweaked with minimal disturbance to the structurally conserved core, and the other three shorter loops (see Table 1) were annealed by using an earlier melding procedure [15,41]. The final choice of loop fragments was made on the basis of sequence similarity and geometric criteria such as initial differences in *N*- and *C*-terminal overlaps or changes in the torsion angles and root-mean-square (RMS) differences in C_α positions in the loop during ring closure.

The COMPOSER-built model had at this stage no bad C_α - C_α steric contacts [with a 0.35 nm (3.5 Å) cut-off], but was found to possess a number of short non-bonded steric contacts. This effect occurs because of the way in which the conformations of substituted side chains are selected since no account is taken of the environment of the side chain apart from the type of secondary structure it is in when the substitutions are made [17]. For example, in the model, a steric clash was found between the $O_{\epsilon 2}$ atom of Glu-118 and the side chain O atom (distance = 1.96 Å) of Tyr-207 (papain numbering, 203) as the side chain of Glu-118 had been modelled in the $g^+ t$ conformation from Glu-121 actinidin ($\chi_1 = -72.5^\circ$, $\chi_2 = -175.1^\circ$). Actinidin, however, has a residue with a much shorter side chain (Ala-209) at the position equivalent to the tyrosine (see Fig. 1). Papain, on the other hand, also has a tyrosine at position 203, and so the side chain torsion angles of Glu-118 were remodelled manually using interactive graphics on Gln-118 of papain in the tt position ($\chi_1 = -169.7^\circ$, $\chi_2 = 173.2^\circ$), thus permitting the formation of a side chain-side chain hydrogen bond between Glu-118 and Tyr-207. Other short non-bonded steric contacts and minor adjustments in side chain conformations were made manually in a similar way by following the set of rules described by Summers *et al.* [42]. The geometry of the whole molecule was then regularized using 10 or 20 cycles of the REFI option (based on the Hermans & McQueen method [43]) within FRODO on successive overlapping five-residue sections. Finally hydrogen atoms were added and the model was subjected to energy minimization using GROMOS (62 cycles of steepest descents and 200 cycles of conjugate gradients) until convergence was reached. The C_α positions of the final model are shown in Fig. 2(b).

RESULTS AND DISCUSSION

Papaya proteinase Ω : identification and the current status of structure-function relationships

The most basic of the cysteine proteinases present in the latex of *Carica papaya* was first isolated by Schack [44] who named it papaya peptidase A, and, with Kaarsholm [45], determined its pI as 11.7. Subsequently it has been called variously 'papaya peptidase II' [46], 'papaya proteinase A' [47] and 'papaya proteinase III' [48]. Because the number and nature of the cysteine proteinase components eluted from ion-exchange

columns between papain and the most basic enzyme vary with the source of the papaya latex and its treatment [49,50], we suggested that the most basic enzyme might usefully be called papaya proteinase Ω because it is most certainly the last cysteine proteinase to be eluted in ion-exchange fractionation of papaya latex [50,51]. This name, which is used in the present paper, has since been adopted by laboratories reporting results of structural investigations on the enzyme [38,52–55]. Buttle *et al.* [9], however, still use the name ‘papaya proteinase III’ for this enzyme, and consequently ‘papaya proteinase IV’ for the monothiol cysteine-proteinase component of the complex chymopapain elution band that we call ‘chymopapain M’ [14].

Relatively few substantive studies that bear on structure–function relationships for papaya proteinase Ω have been reported. The complete amino acid sequence is known [38], and the chain length (216 residues) is similar to those of papain (212) and actinidin (220). Like papain and actinidin, papaya proteinase Ω contains three disulphide bonds (Cys-22–Cys-63, Cys-56–Cys-95, and Cys-153–Cys-204) and Cys-25, which contributes the free thiol group that is essential for catalytic activity [56]. The minimal kinetic model for cysteine-proteinase-catalysed hydrolysis is the well-known three-step acyl-enzyme model (eqn. 4).

Both acylation (rate constant k_{+2}) and deacylation (rate constant k_{+3}) probably involve tetrahedral intermediates, and the possibility of kinetically significant conformational transitions cannot be ruled out [3]. By using thionoester substrates, the transient existence of the dithioacyl-enzymes, ES' , has been demonstrated for reactions catalysed by papaya proteinase Ω , and their formation and subsequent hydrolysis characterized at pH 6.5 by pre-steady-state and steady-state kinetics [57]. In view of the low values of $k_{\text{cat.}}/K_m$ for papaya proteinase Ω -catalysed hydrolysis of the thionoesters and of L-Bz-Arg-Nan, this parameter [which is equal to $k_{+1}k_{+2}/(k_{-1} + k_{+2})$] may be taken to be a close approximation to k_{+2}/K_s (i.e. $k_{+1}k_{+2}/k_{-1}$) [58]. The main mechanistic conclusion of the kinetic studies reported to date is that whereas the chemistry of the deacylation process is highly conserved among actinidin, papain and papaya peptidase Ω , there is considerable variation in the characteristics of the acylation process among the three enzymes. The conservation of the chemistry of the deacylation process deduced from kinetic study is supported by the close similarity of the resonance Raman spectra of the dithioacyl-enzymes. Comparison of the spectra of the derivatives of papaya proteinase Ω and actinidin [57] with those of the corresponding papain derivatives [59] permitted the conformation of the dithioacyl group in the catalytic site to be defined. Favourable or unfavourable contacts in the P_2 – S_2 region [54] are not reflected in a marked change in ligand geometry at P_1 , when this is assessed at the acyl-enzyme stage of catalysis, although P_2 – S_2 contacts probably have a marked effect during the acylation process.

In the present work, the pH-dependence of (a) $k_{\text{cat.}}/K_m$ ($= k_{+2}/K_s$) for the hydrolysis of L-Bz-Arg-Nan catalysed by papaya proteinase Ω and (b) the reactions of the catalytic-site thiol group of papaya proteinase Ω with a series of disulphides containing the 2-mercaptopyridine leaving group are reported and compared with analogous pH-dependences for the reactions of papain and actinidin. Differences in behaviour between papain and actinidin are probably due to a combination of differences in binding interactions and differences in electrostatic effects [6,7,29,32], and the experimental approaches indicated were considered

appropriate to reveal mechanistically relevant differences in the behaviour of papaya proteinase Ω which could be assessed in terms of the structure of the enzyme built by knowledge-based modelling. pH-dependent kinetic analysis makes use of H^+ as the least sterically demanding ligand and is particularly valuable in revealing electrostatic effects. Substrate-derived 2-pyridyl disulphides [31] are sensitive reactivity probes for study of the interrelationships of specificity and catalytic-site chemistry [6,7,30].

The knowledge-based model of the structure of papaya proteinase Ω

Development of the model. Fig. 1 shows the modelling framework, the alignment to it of the papaya proteinase Ω sequence, the structural alignment of papain and actinidin obtained by using the program COMPARE [60] and the consensus secondary structure. The structural alignment shows that residue physicochemical properties such as accessibility to solvent and involvement in hydrogen-bonding tend to be more conserved than residue identity. From the consensus secondary structure it is clear that the framework comprises mainly a series of helices and β strands with occasional connecting turns that are conserved. The framework calculated as the weighted mean of each equivalent C_α position is shown in Fig. 2(a), and the C_α positions of the final model, obtained by sequential application of the various procedures described in the Materials and methods section including energy minimization using GROMOS, are shown in Fig. 2(b). The fragments of papain and actinidin used to give the model real protein geometry are shown in Table 1 (column 1). The protein fragments used in modelling the eight variable regions that connect the framework elements of secondary structure are shown in Table 1 (columns 2 and 3).

Estimate of model error. It is important to have some indication of the likely errors in the final model (Fig. 2b) as a function of residue position in the structure. A good indicator of this appears to be the contact number [61] which is simply a count of the number of residues within a sphere of the atom under consideration. Using a sphere radius of 0.14 nm (14 Å), the reciprocal contact number is highly correlated with the RMS error of main chain C_α values for COMPOSER-built models of rattlesnake phospholipase A_2 and bovine trypsin when compared with the experimentally determined crystallographic structures [18]. Fig. 3 shows a plot of the reciprocal 0.14 nm (14 Å) contact number for each C_α position in the final model. Peaks in the plot occur where there are a few neighbouring residues and are likely to be associated with higher error than residue positions where there are many surrounding residues. Examination of Fig. 3 reveals that the major peaks coincide well with the variable loop regions of the papaya proteinase Ω structure. These occur at positions corresponding either to insertion or deletion sites in the structural alignment of actinidin and papain shown in Fig. 1 or to loop regions of identical length where mutations in the known structures have resulted in different stable local main chain conformations being taken up. The largest peak is at Gly-171 in the variable region 167–173 (GKSGGKG) corresponding to the segment 170–176 (GTEGGVD) in actinidin (see Table 1). The four-residue insertion at position 172 in actinidin (EGGV, see Fig. 1) is accommodated by an extension to an anti-parallel β -sheet by forming two additional hydrogen bonds, permitting a class 2:2 I' β turn [62]. The equivalent insertion in papaya proteinase Ω (SGGK) can, however, be expected to adopt a similar conformation since the pair of glycine residues are conserved and have been identified as structurally important or ‘key’ invariant residues in detailed studies of β -hairpin bends [62–64]. The next two highest peaks occur at Gly-198 and Gly-101 and the fragments containing these residues, modelled on

The sequences of papain and actinidin were aligned by using the three-dimensional structures and COMPARER [60]. The co-ordinates were obtained from the Brookhaven Protein Data Bank (PDB) [39] (PDB codes: 9PAP, papain; 2ACT, actinidin). The amino acid code is the standard one-letter code formatted by using the following convention [72]: UPPERCASE, solvent inaccessible; lowercase, solvent accessible; **Bold**, H-bond with amide proton; underline, H-bond to main chain carbonyl; tilde (~), H-bond with other side chain; breve (˘), *cis*-peptide bond; cedilla (ç), half-cystine. The consensus secondary structure is shown immediately below the alignment: α , α -helical positions; β , β -stranded positions; +, positions in a positive ϕ conformation. The framework of topologically equivalent positions in the two known structures used in the model building are indicated by = and the alignment of the amino acid sequence of papaya proteinase Ω to the framework (pp omega) is shown below the framework. Note that Gln-47, Gln-118 and Gln-135 recorded in 9PAP are glutamic acid residues according to the nucleotide sequence (see [3] for a summary of the differences between 9PAP and the nucleotide sequence reported by Cohen *et al.* [74]). Glu-135 is a component of the electrostatic field in the vicinity of the active-centre region of papain (see the Results and discussion section).

Fig. 2. C α positions of the model of papaya proteinase Ω
(a) The framework for modelling defined as the weighted average of the topologically equivalent C α positions of papain and actinidin indicated in Fig. 1, with a sequential numbering system for papaya proteinase Ω (see Table 1). (b) The C α positions of the final model of papaya proteinase Ω in the same orientation as (a) after regularization of geometry and energy minimization.

Table 1. Protein fragments used in the modelling of papaya proteinase Ω

A consecutive numbering scheme is used here for the papaya proteinase Ω structure. The structures for papain and actinidin are numbered according to the co-ordinate files deposited in the Brookhaven Data Bank [39] and identified by four-letter codes shown in parenthesis.

Structurally conserved region	Loop region	Loop sequence
Residues 1–13 Papain (9PAP) residues 1–13	Residues 14–15 Papain (9PAP) residues 14–15	TP TP
Residues 16–44 Papain (9PAP) residues 16–44	Residues 45–46 Actinidin (2ACT) residues 45–46	LV LI
Residues 47–55 Actinidin (2ACT) residues 47–55	Residues 56–62 Papain (9PAP) residues 56–62	CERRSHG CDRRSYG
Residues 63–75 Papain (9PAP) residues 63–75	Residues 76–79 Papain (9PAP) residues 76–79	AKNG AQYG
Residues 80–91 Papain (9PAP) residues 80–91	Residues 92–102 Papain (9PAP) residues 92–102	QGTCRAKQVGG QRYCRSREKGP
Residues 103–135 Papain (9PAP) residues 103–135	Residues 136–137 Papain (9PAP) residues 136–137	SK AA
Residues 138–166 Papain (9PAP) residues 138–166	Residues 167–173 Actinidin (2ACT) residues 170–176	GKSGGKG GTEGGVD
Residues 174–194 Papain (9PAP) residues 170–190	Residues 195–201 Papain (9PAP) residues 191–197	RAPGNSP RGTGNSY
Residues 202–216 Papain (9PAP) residues 198–212		

loop regions of the same length in papain (Table 1), are far removed from the active-centre region on the surface of the R and L domains respectively. Pickersgill *et al.* [55] have recently reported a preliminary X-ray analysis of the papaya proteinase Ω structure, and it will be of considerable interest to compare the COMPOSER-built model with the experimentally determined structure once the co-ordinates become available.

Fig. 3. Dependence of the reciprocal 1.4 nm (14 Å) contact number on residue position for the model-built structure of papaya proteinase Ω

The incidence of peaks is expected to be highly correlated with the RMS error of the model. The structurally variable regions (see Table 1) are indicated by horizontal brackets.

Fig. 4. Active-centre region of the papaya proteinase Ω model showing selected side chains and main chain N, C α , C' (carbonyl carbon) and O atom positions

Some mechanistically relevant structural features. The active-centre region of papaya proteinase Ω (Fig. 4) is similar in many respects to that of papain. Features that are conserved include the catalytic-site Cys-25/His-159 ion pair, Gln-19 and other residues surrounding the putative oxyanion hole, and Asn-179 which accepts a hydrogen bond from His-159. The topologies of the L-domain cleft walls are very similar, with the conserved Gly-65 and Gly-66 both modelled with a positive ϕ main chain torsion angle as observed in the X-ray structures of papain and actinidin (see Fig. 1), although the substitution of Asn-64 by Lys-64 would be expected to influence the electrostatic environment. Pro-68 is conserved at the *N*-terminus of the helix running from position -68 to position -77. Pro-68 is at the entrance of the S_2 subsite (a major specificity determinant in papain) which exhibits very few differences when compared with the analogous region in papain with Ser-209 (papain numbering, 205) and Val-133 and Val-157 conserved. Of the two candidates that have been tentatively suggested as alternatives for one of the kinetically influential pK_a values seen in reactions of papain [6,65], the remote ion pair Glu-52/His-81 (which is lacking in actinidin) is conserved in papaya proteinase Ω and the Tyr-61/Tyr-67 hydrogen-bonded pair is replaced by His-61/Tyr-67. The $N_{\epsilon 2}$ -atom of His-61 in papaya proteinase Ω is located too far [0.422 nm (4.22 Å)] in the model from the side chain oxygen atom of Tyr-67 for hydrogen-bond formation. In actinidin, Tyr-67 is conserved, but the residue analogous to Tyr-61 of papain is Arg-61. Differences in behaviour between papain and actinidin have been discussed in terms of differences in the electrostatic fields to which the two catalytic sites are exposed [7,66], and the analogous field in papaya proteinase Ω is thus of particular interest. In both papain and papaya proteinase Ω , Asp-158 is exposed to the same relatively long distance influence of Glu-135 which interacts also with Lys-156 in both enzymes. This contrasts with the situation in actinidin where Glu-135 is replaced by Asp-138 which can approach much closer to the carboxy group of Asp-161 and in which the cationic side chain of Lys-156 is replaced by the methyl group of Ala-159. It is of particular interest that in papaya proteinase Ω , the carboxy group of Asp-158 both engages in hydrogen-bond formation with the hydroxy group of Ser-136 and is exposed to the influence of the cationic side chain of Lys-137 [closest approach 158-O-137-N = 0.454 nm (4.54 Å)]. The electrostatic network around the catalytic site of papaya proteinase Ω is shown in Fig. 5. The residues in papain (136 and 137)

and in actinidin (139 and 140) that are analogous to Ser-136 and Lys-137 in papaya proteinase Ω are all alanine residues.

Experimental evaluation of some of the structural features revealed by the knowledge-based model of papaya proteinase Ω

Detection, characteristics and perturbation of the catalytic-site Cys-25/His-159 ion pair and its modulation by other ionizations. The characterization of the catalytic-site ion pairs of papain and actinidin and of their perturbation by binding and electrostatic interactions has been carried out previously by using a range of time-dependent inhibitors such as (I)–(V) as reactivity probes [6,32]:

In each case the thiol-specific probe reagent contains a chromophoric 2-mercaptopyridine leaving group whose activation, either by hydronation (pK_a 2.45–2.90) or by hydrogen-bonding with the catalytic site (His)-Im ^+H , permits the detection and characterization of the -Im ^+H /S $^-$ ion pairs and of other interactions that affect the reactivity characteristics of the ion pairs and their components.

In reactions of actinidin and of papain with a probe containing a simple alkyl side chain as in (I), or another 2-mercaptopyridine group (V), activation is by formal hydronation of the pyridyl N atom and the pH-*k* profile contains a striking bell-shaped component at low pH, a minimum at pH 6–7 and a plateau at high pH. That the same general form of pH-*k* profile is obtained for analogous reactions of papaya proteinase Ω is shown in Figs. 6(a) and 6(b). Analysis of these profiles by the procedures given in the Materials and methods section provides values for the molecular pK_a values characteristic of the formation of the ionization state containing the -S $^-$ /Im ^+H ion pair and its conversion into -S $^-$ /Im as 3.4 and 8.1–8.3 respectively. The

Fig. 5. Stereoview of the electrostatic influences on the catalytic site -S $^-$ /Im ^+H ion pair of the model-built structure of papaya proteinase Ω

Fig. 6. pH-dependence of the reaction of the catalytic site (Cys-25) thiol group of papaya proteinase Ω with (a) 2,2'-dipyridyl disulphide (2-Py-S-S-2-Py) and (b) methyl 2-pyridyl disulphide ($\text{CH}_3\text{-S-S-2-Py}$) at 25 °C and at 10.1

The points were determined by stopped-flow kinetic analysis at 343 nm and in each case the line is theoretical for a reaction in three reactive hydronic states with for (a) $\text{p}K_1 = 2.45$, $\text{p}K_{II} = 3.4$, $\text{p}K_{III} = 8.1$, $\bar{k}_1 = 2.8 \times 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$, $\bar{k}_2 = 650 \text{ M}^{-1} \cdot \text{s}^{-1}$, $\bar{k}_3 = 2.1 \times 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$, and for (b) $\text{p}K_1 = 2.8$, $\text{p}K_{II} = 3.4$, $\text{p}K_{III} = 8.3$, $\bar{k}_1 = 1.5 \times 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$, $\bar{k}_2 = 2.5 \times 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$, $\bar{k}_3 = 1.8 \times 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$, determined by using SKETCHER (see the Materials and methods section). The values of $\text{p}K_1$ are identical with the $\text{p}K_a$ values of 2-Py-S-S-2-Py⁺H and $\text{CH}_3\text{-S-S-2-Py}^+\text{H}$ determined spectrophotometrically; the values of $\text{p}K_{II}$ and $\text{p}K_{III}$ are identical with or closely similar to kinetically influential $\text{p}K_a$ values observed in the pH-dependence of $k_{\text{cat.}}/K_m$ for the hydrolysis of L-Bz-Arg-Nan (Fig. 7). Preliminary data for the reaction of 2-Py-S-S-2-Py did not permit a clear identification of $\text{p}K_1$ with the reagent $\text{p}K_a$ [56].

corresponding $\text{p}K_a$ values for actinidin are 3.1–3.2 and 9.5, and for papain are 3.8 and 9.5.

Another difference between the pH- k profiles for the reactions of 2-Py-S-S-2-Py and $\text{CH}_3\text{-S-S-2-Py}$ with papaya proteinase Ω and papain is the lack in the profiles for the former enzyme of the additional enzyme $\text{p}K_a$ close to 4 that is characteristic of the kinetic behaviour of papain in these and other reactions. The absence of this additional $\text{p}K_a$ value close to 4 has been reported for the pH-dependence of $k_{\text{cat.}}/K_m$ in acidic media for the catalysis of the hydrolysis of a number of substrates by papaya proteinase Ω [67].

To explore the possibility that the minimal catalytic device of papaya proteinase Ω (the Cys-25/His-159 ion pair) might be modulated by electrostatic and binding interactions and to compare any modulatory characteristics with those of papain and actinidin, we determined the characterizing parameters of the pH- $(k_{\text{cat.}}/K_m)$ profile of the catalysed hydrolysis of L-Bz-Arg-Nan by using SKETCHER and collected kinetic data for the reactions of papaya proteinase Ω with probe reagents (I)–(IV) in

Fig. 7. (a) pH-dependence of $k_{\text{cat.}}/K_m$ for the hydrolysis of L-Bz-Arg-Nan catalysed by papaya proteinase Ω at 25 °C and 10.1 and (b) comparison with analogous pH-dependences for papain- and actinidin-catalysed hydrolysis

(a) The data points in the range 2.6–10.5 were determined by measuring v_i with $[\text{E}]_T = 2\text{--}3 \mu\text{M}$ and $[\text{S}]_0 = 0.5 \text{ mM}$, i.e. $[\text{S}]_0 \ll K_m$ (K_m values determined at pH values 4.0, 7.4 and 9.2 were 18 mM, > 18 mM and 16 mM respectively). Values of $k_{\text{cat.}}/K_m$ calculated from values of $k_{\text{cat.}}$ and K_m determined separately at pH values around 4 and 9 lie within the cluster of points shown. The line corresponds to the following values of the parameters for a reaction in three reactive hydronic states determined by using SKETCHER (see the Materials and methods section): $\text{p}K_1 = 3.4$, $\text{p}K_{II} = 5.8$, $\text{p}K_{III} = 8.0$, $\text{p}K_{IV} = 9.4$; $(\bar{k}_{\text{cat.}}/\bar{K}_m)_1 = 8.4 \text{ M}^{-1} \cdot \text{s}^{-1}$; $(\bar{k}_{\text{cat.}}/\bar{K}_m)_2 = 22.0 \text{ M}^{-1} \cdot \text{s}^{-1}$; $(\bar{k}_{\text{cat.}}/\bar{K}_m)_3 = 18.0 \text{ M}^{-1} \cdot \text{s}^{-1}$. (b) The lines are theoretical for the appropriate multihydronic state rate equations with the following values of the parameters. (i) Papain: ordinate scaling factor, $n = 1$, $\text{p}K_1 = 3.7$, $\text{p}K_{II} = 4.2$, $\text{p}K_{III} = 8.0$, $\text{p}K_{IV} = 9.6$, $(\bar{k}_{\text{cat.}}/\bar{K}_m)_1 = 50 \text{ M}^{-1} \cdot \text{s}^{-1}$, $(\bar{k}_{\text{cat.}}/\bar{K}_m)_2 = 220 \text{ M}^{-1} \cdot \text{s}^{-1}$, $(\bar{k}_{\text{cat.}}/\bar{K}_m)_3 = 68 \text{ M}^{-1} \cdot \text{s}^{-1}$ (C. M. Topham, E. Salih & K. Brocklehurst, unpublished work); (ii) actinidin: $n = 25$, $\text{p}K_1 = 3.1$, $\text{p}K_{II} = 5.5$, $\text{p}K_{III} = 9.7$, $25 \times (\bar{k}_{\text{cat.}}/\bar{K}_m)_1 = 68.75 \text{ M}^{-1} \cdot \text{s}^{-1}$, $25 \times (\bar{k}_{\text{cat.}}/\bar{K}_m)_2 = 212.5 \text{ M}^{-1} \cdot \text{s}^{-1}$ [29]; (iii) papaya proteinase Ω : $n = 10$ and the values of the parameters are those given in (a). The arrows indicate the extent of the data.

pH regions that would reveal binding-site-catalytic-signalling effects of the type observed with papain but not with actinidin [6,32]. The results of the studies on the substrate are shown in Fig. 7 and those on the substrate-derived reactivity probes in Table 2.

Dubois *et al.* [52] reported, without showing any experimental data, that pH-dependences of $k_{\text{cat.}}/K_m$ for the hydrolysis of L-Bz-Arg-Nan catalysed by both papaya proteinase Ω and papain are bell-shaped with $\text{p}K_1$ 4.2–4.3 and $\text{p}K_{II}$ 8.5–8.6. (Actually DL-Bz-Arg-Nan was used but presumably the results relate to the L-isomer; see [68].) Fig. 7 shows that the pH-dependences of $k_{\text{cat.}}/K_m$ for the hydrolysis of papaya proteinase Ω , papain and actinidin are all more complex than bell-shaped curves and that the shape of the pH- $(k_{\text{cat.}}/K_m)$ profile for the papaya proteinase

Table 2. Demonstration of the intermediate position occupied by papaya proteinase Ω between actinidin and papain in the transmission of binding effects to catalytic-site chemistry: effect of structural variation in the probe reagent (R-S-S-2-Py) on the pH-dependence of the kinetics of reaction with the three cysteine proteinase variants

Reactivity probe	Nature of side chain and of potential binding interactions	Enzyme	pH	Second-order rate constant (k) ($M^{-1} \cdot s^{-1}$)	Comments	Reference
(I)	CH ₃ -CH ₂ -; no opportunities for hydrogen-bonding with the enzymes or for occupancy of the S ₂ subsites	Actinidin	3.5	2.05×10^3	k is a minimum in approximately neutral media, which demonstrates that probe activation is by formal hydration of the -S-2-Py leaving group at low pH in the reactions of all three enzymes with this 'featureless' reactivity probe	Present work
			4.1	1.0×10^3		
			5.1	4.2×10^2		
			7.0	1.5×10^2		
			9.2	2.9×10^2		
			10.0	4.0×10^2		
(I)		Papaya proteinase Ω	3.5	1.95×10^4		Present work
			4.0	9.8×10^3		
			5.0	2.0×10^3		
			6.5	1.0×10^3		
			8.0	2.4×10^3		
			9.0	3.1×10^3		
(I)		Papain	4.1	3.25×10^4		[30]
			5.0	1.35×10^4		
			7.0	4.9×10^3		
			10.4	6.0×10^3		
(II)	$\text{CH}_3-\text{C}-\text{NH}-\text{CH}_2-\text{CH}_2-;$ $\quad \quad \quad \parallel$ $\quad \quad \quad \text{O}$ <p>this probe contains a structural feature that approximates to the P₁-P₂ amide bond of substrates, and model building suggests that it could bind across the active-centre clefts of these enzymes by hydrogen-bonding to the backbone N-H of Gly-66 in papain and papaya proteinase Ω and Gly-68 in actinidin and to the backbone C=O of Asp-158 in the first two enzymes and Asp-161 in actinidin</p>	Actinidin	3.50	2.5×10^3	k is a maximum in approximately neutral media only for the papain reaction; a rate maximum in this pH region provides evidence for a substantial contribution from an -Im ⁺ H-assisted transition state (see the text). For the other two reactions, k is a minimum; for actinidin, $k_{\text{pH } 6.5}/k_{\text{pH } 10.1} = 0.16$; for papaya proteinase Ω , $k_{\text{pH } 6.2}/k_{\text{pH } 10.5} = 0.36$; for papain, $k_{\text{pH } 6.6}/k_{\text{pH } 10.0} = 1.5$	[6]
			4.25	1.5×10^3		
			6.50	4.9×10^2		
			7.50	5.0×10^2		
			9.00	1.35×10^3		
			10.10	3.0×10^3		
(II)		Papaya proteinase Ω	3.50	1.8×10^4		Present work
			4.00	1.25×10^4		
			6.20	6.1×10^3		
			7.50	8.5×10^3		
			9.20	1.5×10^4		
			10.50	1.7×10^4		
(II)		Papain	3.20	8.0×10^3		[6]
			3.70	2.0×10^4		
			5.00	2.9×10^4		
			6.60	2.9×10^4		
			8.40	2.2×10^4		
			10.00	1.95×10^4		
(III)	$\text{CH}_3-\text{C}-\text{NH}-\text{CH}-\text{C}-\text{NH}-\text{CH}_2-\text{CH}_2-;$ $\quad \quad \quad \parallel \quad \quad \quad \parallel$ $\quad \quad \quad \text{O} \quad \quad \quad \text{O}$ $\quad \quad \quad \text{CH}_2$ $\quad \quad \quad \text{C}_6\text{H}_5$ <p>this probe contains the P₁-P₂ amide bond available also in probe (II) but in addition it contains a phenylalanine side chain as a potential occupant for the S₂ subsites of the enzymes</p>	Actinidin	4.0	3.0×10^5	k is a maximum in approximately neutral media for the reaction of papain and papaya proteinase Ω but not for the reaction of actinidin; the rate maximum is much more highly developed in the papain reaction ($k_{\text{pH } 6.0}/k_{\text{pH } 10.0} = 6.7$) than in the papaya proteinase Ω reaction ($k_{\text{pH } 6.2}/k_{\text{pH } 10.0} = 1.4$)	[7]
			5.2	2.2×10^5		
			6.0	1.2×10^5		
			7.3	6.0×10^4		
			8.5	5.0×10^4		
			10.1	5.0×10^4		
(III)		Papaya proteinase Ω	2.6	9.0×10^4		Present work
			3.3	1.5×10^5		
			4.0	1.75×10^5		
			4.9	1.9×10^5		
			6.2	1.9×10^5		
			7.5	1.6×10^5		
(III)		Papain	9.9	1.4×10^5		[6]
			3.3	8.0×10^5		
			4.0	3.0×10^6		
			6.0	6.0×10^6		
			7.5	5.0×10^6		
			8.6	3.0×10^6		
			10.0	9.0×10^5		

Table 2. (cont.)

Reactivity probe	Nature of side chain and of potential binding interactions	Enzyme	pH	Second-order rate constant (k) ($\text{M}^{-1} \cdot \text{s}^{-1}$)	Comments	Reference
(IV)	$\text{CH}_3\text{-C(=O)-NH-CH(CH}_2\text{-C}_6\text{H}_5\text{)-C(=O)-O-CH}_2\text{-CH}_2\text{-}^-$ <p>this probe retains the potential to bind in the S_2-subsites and to accept a hydrogen bond from Gly-66 in papain and papaya proteinase Ω and Gly-68 in actinidin but lacks the potential to donate a hydrogen bond to the backbone C=O of Asp-158 (or Asp-161 in actinidin)</p>	Actinidin	3.6	1.4×10^5	The general shapes of the pH- k profiles are as for the reactions of the three enzymes with probe (III)	Present work
			5.0	9.0×10^4		
			6.2	3.8×10^4		
			8.3	1.4×10^4		
			10.0	1.4×10^4		
(IV)		Papaya proteinase Ω	2.2	1.8×10^4		Present work
			4.4	6.1×10^4		
			6.0	6.8×10^4		
			8.5	6.0×10^4		
			10.0	5.3×10^4		
(IV)		Papain	3.3	2.4×10^5		[7]
			4.0	7.1×10^5		
			5.3	1.15×10^6		
			6.5	1.2×10^6		
			9.0	5.2×10^5		
			10.0	2.0×10^5		

Ω catalysis resembles that of actinidin much more than that of papain. The profile for the papaya proteinase Ω catalysis exhibits the pK_a values (3.4 and 8.0) characteristic of the formation and dehydronation of the ion pair (3.4 and 8.1–8.3) determined by reactivity-probe kinetics (Fig. 6). In addition, however, it is clear from Fig. 7 that the reactivity of the $\text{-S}^-/\text{-ImH}^+$ ion pair in papaya proteinase Ω is markedly modulated by an ionization with pK_a 5.8, a situation similar to that observed with actinidin. The alkaline limb of the papaya proteinase Ω profile contains an additional sigmoid wave with pK_a 9.4. It is possible that, with more limited data and in the absence of the evidence from reactivity probe kinetics, analysis of the alkaline limb might appear to provide a single pK_a of approx. 8.5 as reported by Dubois *et al.* [52] (see [69] for an account of overlapping kinetically influential ionizations).

The most obvious interpretation of the data in Fig. 7 for the papaya proteinase Ω -L-Bz-Arg-Nan reaction is that formation of the acylenzyme (thioester) intermediate requires the (Cys-25)- $\text{S}^-/(\text{His-159})\text{-ImH}^+$ ion pair produced as the pH is raised across pK_a 3.4 and that the effectiveness of the acylation process is enhanced by dehydronation of a group in the enzyme associated with pK_a 5.8 and further enhanced by dehydronation of another group in the enzyme associated with pK_a 9.4. This is postulated to offset the decrease in $k_{\text{cat.}}/K_m$ resulting from the loss of an ion pair across pK_a 8.0. For the analogous actinidin-catalysed hydrolysis [29], the effectiveness of the ion pair (produced across pK_a 3.1–3.2 [6]) is increased by dehydronation across pK_a 5.5, and $k_{\text{cat.}}/K_m$ decreases as the pH is raised across pK_a approx. 9.7, which is close to the value (9.5) for the dehydronation of the ion pair to produce $\text{-S}^-/\text{-Im}$ deduced from reactivity-probe kinetics. Thus there is no evidence in the case of actinidin for the second modulatory ionizing group seen with papaya proteinase Ω .

The behaviour of papain in both L-Bz-Arg-Nan hydrolysis and its reactions with reactivity probes is quite different in a number of respects. Its catalysis of the hydrolysis of L-Bz-Arg-Nan is not modulated by a pK_a around 5.5–6.0 but, uniquely, the production of an ion pair across pK_a approx. 3.8 is coupled with a co-operative enzyme ionization of closely similar pK_a . The pK_a associated with the dehydronation of the ion pair to form $\text{-S}^-/\text{-Im}$ deduced from reactivity probe kinetics is 9.5. Decrease of

$k_{\text{cat.}}/K_m$ in alkaline media, however, is associated mainly with dehydronation across pK_a 8.0, with only residual activity lost across pK_a 9.5 (C. M. Topham, E. Salih & K. Brocklehurst, unpublished work). This phenomenon occurs also in the reaction of papain with the substrate-derived probe (compound III), whereby the high reactivity at pH values around 6, which is attributed to binding-site-catalytic-site signalling, is lost in alkaline media mainly across pK_a 8.25, with only residual reactivity lost across pK_a 9.5 [6]. Thus dehydronation in alkaline media appears to decrease the effectiveness of the catalytic site ion pair in papain whereas it appears to increase it in the case of papaya proteinase Ω .

The combination of structural and kinetic results reported in the present work casts doubt on the assignment of the kinetically influential pK_a of approx. 5.5 in actinidin-catalysed hydrolysis as the carboxy group of Asp-161 [7,57]. These results comprise: (a) the fact that the electrostatic environment of Glu-135 in papaya proteinase Ω is analogous to that of Glu-135 in papain [i.e. in ion-pair formation with Lys-156 (Fig. 5)] rather than to that of Asp-138 in actinidin (where the residue analogous to Lys-156 is Ala-159), and (b) the fact that the forms of the acid limbs of the $k_{\text{cat.}}/K_m$ -pH profiles for the papaya proteinase Ω - and actinidin-catalysed hydrolyses of L-Bz-Arg-Nan are similar in that each contains a kinetically influential pK_a in the range 5.5–6.0, which is lacking in the analogous profile for papain (Fig. 7). Thus the environment of Asp-158 in papaya proteinase Ω would not be expected to provide it with a pK_a value of 5.5–6.0 because of the lack of proximity of the negative charge of Glu-135 and because of the presence of the additional hydrogen bond and cationic site donors, Ser-136 and Lys-137 (Fig. 5). This contrasts markedly with the situation in actinidin, where Asp-138 exerts an electrostatic influence on Asp-161. Thus if the kinetically influential pK_a in the range 5.5–6.0 observed for both actinidin and papaya proteinase Ω has a common cause, it is unlikely to be attributable to Asp-161 in actinidin and Asp-158 in papaya proteinase Ω . The recent demonstration that the Asp-158-Asn mutant of papain produced by site-directed mutagenesis still displays a second pK_a around 4 in the pH-dependence of $k_{\text{cat.}}/K_m$ [70] supports the present suggestion that the additional ionizations observed in the pH-dependent behaviour of the ion-pair systems of actinidin,

papaya proteinase Ω and papain are not attributable to the carboxy group of the residue adjacent to the catalytic-site histidine residue.

Modulation of catalytic-site ion-pair reactivity by binding interactions in the S_2 subsite and S_1 - S_2 intersubsite regions

The results of the investigation of binding-site-catalytic-site signalling in papaya proteinase Ω and their comparison with data for the reactions of papain and actinidin are presented in Table 2. The approach, involving the pH-dependent kinetic analysis of the reactions of the catalytic-site thiol group of the enzyme with reactivity probes such as (I)–(IV), has been discussed previously [6,32] in connection with studies on papain and actinidin. In summary, a rate minimum at pH 6–7 implies a reaction pathway in which nucleophilic attack of the (Cys-25)-S[−] component of the ion pair is not assisted to a significant extent by association of (His-159)-Im⁺H with the leaving group, whereas a rate maximum in this pH region indicates substantial assistance by the catalytic-site imidazolium cation, as is envisaged in the productive collapse of a tetrahedral intermediate in the catalysed hydrolysis of substrates [69].

When the side chain of the probe contains no binding opportunities as in (I) (where R = unsubstituted alkyl, e.g. C₂H₅ or CH₃; see Fig. 6b), a rate minimum at pH 6–7 is observed in the reactions of all three enzymes. Introduction of a P₁-P₂ amide bond [probe (II), where R = CH₃-CONH-CH₂-CH₂-] produces a shallow rate maximum for the reaction with papain (with $k_{pH6.6}/k_{pH10.0} = 1.5$) but not for those of the other two enzymes (for papaya proteinase Ω , the ratio $k_{pH6.2}/k_{pH10.5} = 0.36$, being somewhat larger than that for actinidin, $k_{pH6.5}/k_{pH10.1} = 0.16$). When the side chain contains both a P₁-P₂ amide bond and a phenylalanine side chain as a potential occupant for the S_2 subsites of the enzymes [probe (III), where R = CH₃CONH-CH(CH₂Ph)CONH-CH₂-CH₂-] a rate minimum is still observed for the actinidin reaction, a shallow rate maximum for the papaya proteinase Ω reaction ($k_{pH6.2}/k_{pH10.0} = 1.4$) and a striking well-developed rate maximum for the papain reaction ($k_{pH6.0}/k_{pH10.7} = 6.7$).

pH-*k* profiles of similar shape to those obtained in reactions with probe (III) are obtained also for all three enzymes when the probe lacks the NH- component of the P₁-P₂ amide bond [probe (IV), the P₁-P₂ ester analogue of probe (III)].

The main conclusions drawn from the data reported in Table 2 are (i) that provision of the opportunity for the reactivity probe to bind across the active-centre clefts by hydrogen-bonding involving its P₁-P₂ amide bond provides for a sufficiently effective -Im⁺-H-assisted pathway to produce a (shallow) rate maximum at pH 6–7 in the case of the papain reaction but not in the reactions of the other two enzymes and (ii) that provision, additionally, of the potential for binding of a phenylalanine side chain in the S_2 subsites substantially enhances the effectiveness of the -Im⁺-H-assisted reaction in the papain reaction, provides for a sufficiently effective assistance in the papaya proteinase Ω reaction to produce a shallow rate maximum at pH 6–7, and still fails to produce a rate maximum for the actinidin reaction. The observation that for papain, the -NH component of the P₁-P₂ amide bond is essential for significant -Im⁺-H- assistance in the absence of a S_2 -P₂ interaction, but not in its presence, has been discussed previously [32].

It is clear from Table 2 that papaya proteinase Ω represents an intermediate situation between actinidin and papain in the sensitivity with which specific binding interactions provide for sufficient -Im⁺-H- assistance to leaving-group expulsion to produce a rate maximum at pH 6–7. The knowledge-based model of papaya proteinase Ω (Figs. 2 and 4) suggests that the differences in the electrostatic fields to which the active-centre regions of the

three enzymes are exposed (see Fig. 5 and associated text) might explain this gradation in properties, possibly in terms of the degree of conformational flexibility permitted in the enzyme-ligand complexes.

The lower sensitivity of response to binding of an L-phenylalanine side chain in the S_2 subsite of papaya proteinase Ω than to analogous binding to papain, detected by reactivity-probe kinetics, appears to correlate to some extent with the differences in proteolytic cleavage patterns produced by the two enzymes [54]. The S_2 subsite of papaya proteinase Ω appears to exert a preference for leucine, valine and, to a lesser extent, for threonine, whereas papain prefers the aromatic residues phenylalanine or tyrosine. The S_2 binding pocket of papain is lined with the hydrophobic side chains of Tyr-67, Pro-68, Trp-69, Phe-207, Val-133, Val-157 and Ala-160 [4]. Most of these residues are conserved in papaya proteinase Ω with the exception of Trp-69, which is replaced by Pro-69, and Phe-207, which becomes Tyr-211 (see Fig. 4).

Drenth *et al.* [71] investigated the binding mode in papain of three different chloromethyl ketone inhibitors, each possessing an L-phenylalanine residue at the P₂ position, by the difference-Fourier technique in order to explain the substrate specificity of the enzyme. By using the co-ordinates of the complex formed by the reaction of Ac-Ala-Ala-Phe-Ala-CH₂Cl with the sulphur atom of Cys-25 (Protein Data Bank code: 1PAD), we have calculated the solvent-accessible area of each residue both in the complex (with the water molecules removed from the co-ordinate set) and in the free enzyme (in the same conformation as in the inhibitor complex and with the water molecules removed). The results for the residues in the S_2 subsite are recorded in Table 3. It is interesting that residues losing more than 0.02 nm² (2 Å²) of solvent contact area on binding of the inhibitors are all conserved in papaya proteinase Ω , and so there

Table 3. Residue solvent contact areas in the S_2 subsite of papain and its Ac-Ala-Ala-Phe-Ala-methylene derivative

The residues in the S_2 binding pocket of papain are those defined by Baker & Drenth [4]. The Protein Data Bank [39] co-ordinate set 1PAD [71] was used to calculate solvent contact areas with and without the bound inhibitor by applying the method of Richmond & Richards [73]. A solvent probe radius of 0.14 nm (1.4 Å) was used. Both main chain and side chain atoms were included in the calculations.

S_2 subsite residues in papain with analogous residues in papaya proteinase Ω in parenthesis	Solvent accessibility in free enzyme [nm ² (Å ²)]	Solvent accessibility in complex [nm ² (Å ²)]	Δ accessibility (complex – free enzyme) [nm ² (Å ²)]
Tyr-67 (Tyr-67)	0.221 (22.07)	0.068 (6.81)	−0.153 (−15.26)
Pro-68 (Pro-68)	0.031 (3.09)	0.007 (0.66)	−0.024 (−2.4)
Trp-69 (Pro-69)	0.317 (31.69)	0.300 (30.02)	−0.017 (−1.69)
Phe-207 (Tyr-211)	0.066 (6.62)	0.066 (6.62)	0 (0)
Val-133 (Val-133)	0.056 (5.63)	0.002 (0.23)	−0.054 (−5.40)
Val-157 (Val-157)	0.155 (15.55)	0.092 (9.24)	−0.063 (−6.31)
Ala-160 (Ala-160)	0.013 (1.30)	0 (0)	−0.013 (−1.30)

would not appear to be any obvious explanation in purely geometrical terms for the reported differences in specificity with respect to the S_2 subsite in the two enzymes.

Concluding remarks

The addition of detailed structural and structure-related kinetic data for papaya proteinase Ω to those existing for actinidin and papain has contributed to the developing picture of structure-function relationships in the cysteine proteinases in a number of respects. Papaya proteinase Ω differs from the other two enzymes in the ionization characteristics of the common (Cys)-SH/(His)-Im⁺H catalytic-site system and of the other acid/base groups that modulate thiol reactivity towards substrate-derived inhibitors and the acylation process of the catalytic act. Also, it occupies an intermediate position between actinidin and papain, in the sensitivity with which hydrophobic interaction in the S_2 subsite is transmitted to produce changes in transition-state geometry in the catalytic site, a fact that may be linked with differences in specificity in P_2 - S_2 interaction during proteolysis exhibited by the three enzymes. The most obvious structural differences involve the electrostatic fields to which the active centres are exposed.

It is striking that dehydronation with pK_a 8.0–8.25 in papain decreases the catalytic competence and neighbouring-group character of the -S⁻/Im⁺H ion pair, whereas dehydronation with pK_a 9.5 increases the catalytic competence of the analogous ion pair in papaya proteinase Ω . There appears to be no analogous modulatory group in actinidin.

Although these cysteine proteinases make use of a common -S⁻/Im⁺H ion pair in the rate-determining acylation process of the catalytic act, it is clear that the modulation of the chemical behaviour of the ion pair by binding interactions and by electrostatic effects varies markedly from enzyme to enzyme. A complete description of the mechanism of cysteine-proteinase-catalysed hydrolysis needs the structural basis of these modulatory effects to be identified. The addition of three-dimensional structures of other members of the cysteine-proteinase family to the knowledge base should contribute to this. When only the structures of actinidin and papain were available, cogent argument suggested the assignment of the kinetically influential pK_a of 5.5 in actinidin to Asp-161. Addition of the structure of papaya proteinase Ω to the knowledge base now permits Asp-161 to be eliminated as a candidate for this controlling electrostatic effect and the remote ion pair (Glu-52/His-81) of papain to be eliminated as a potential candidate for the modulatory feature with pK_a values of about 4 and 8 that is unique to this enzyme. Our working hypothesis for the observed variation in the effectiveness with which binding interactions are linked with changes in transition-state geometry in reactions of the three enzymes with time-dependent inhibitors is that the different electrostatic fields to which the three active centres are exposed provide differential control of necessary conformational adjustment in the transition state. Expansion of the structural knowledge base will increase the effectiveness with which structural change by site-directed mutagenesis may be designed for kinetic assessment.

We thank the Science and Engineering Research Council, the European Community and Pfizer for financial support for Postdoctoral Research Assistantships and for Earmarked Studentships, and Mrs. Liz Beech and Miss Joy Smith for preparation of the manuscript.

REFERENCES

1. Brocklehurst, K. (1986) in *Cysteine Proteinases and their Inhibitors* (Turk, V., ed.), pp. 307–325, Walter de Gruyter, Berlin
2. Brocklehurst, K. (1987) in *Enzyme Mechanisms* (Page, M. I. & Williams, A., eds.), pp. 140–158, Royal Society of Chemistry, London
3. Brocklehurst, K., Willenbrock, F. & Salih, E. (1987) in *Hydrolytic Enzymes* (Neuberger, A. & Brocklehurst, K., eds.), pp. 39–158, Elsevier, Amsterdam
4. Baker, E. N. & Dreath, J. (1987) in *Biological Macromolecules and Assemblies* (Jurnak, F. & McPherson, A., eds.), vol. 3, pp. 313–368, John Wiley and Sons, New York
5. Polgár, L. (1989) in *Mechanism of Protease Action* (Polgár, L., ed.), pp. 123–155, CRC Press, Florida
6. Brocklehurst, K., Brocklehurst, S. M., O'Driscoll, M., Patel, G., Salih, E., Templeton, W., Thomas, E., Topham, C. M. & Willenbrock, F. (1988) *Biochem. J.* **256**, 543–555
7. Kowlessur, D., O'Driscoll, M., Topham, C. M., Templeton, W., Thomas, E. W. & Brocklehurst, K. (1989) *Biochem. J.* **259**, 443–452
8. Laycock, M. V., Hirma, T., Hasnain, S., Watson, D. & Storer, A. C. (1989) *Biochem. J.* **263**, 439–444
9. Buttle, D. J., Kumbhavi, A. A., Sharp, S. L., Shute, R. E., Rich, D. H. & Barrett, A. J. (1989) *Biochem. J.* **261**, 469–476
10. Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A. & Dreath, J. (1985) *J. Mol. Biol.* **179**, 233–256
11. Baker, E. N. (1980) *J. Mol. Biol.* **141**, 441–484
12. Topham, C. M., Overington, J., Kowlessur, D., Thomas, M., Thomas, E. W. & Brocklehurst, K. (1990) *Biochem. Soc. Trans.* **18**, 579–580
13. Topham, C. M., Overington, J., O'Driscoll, M., Salih, E., Thomas, M., Thomas, E. W. & Brocklehurst, K. (1990) *Biochem. Soc. Trans.* **18**, 933–934
14. Topham, C. M., Overington, J., Thomas, M., Kowlessur, D., Thomas, E. W. & Brocklehurst, K. (1990) *Biochem. Soc. Trans.* **18**, 934–935
15. Blundell, T., Carney, D., Gardiner, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J., Singh, D. A., Sibanda, L. & Sutcliffe, M. (1988) *Eur. J. Biochem.* **172**, 513–520
16. Sutcliffe, M. J., Haneef, I., Carney, D. & Blundell, T. L. (1987) *Protein Eng.* **1**, 377–384
17. Sutcliffe, M. J., Hayes, F. R. F. & Blundell, T. L. (1987) *Protein Eng.* **1**, 385–392
18. Topham, C. M., Thomas, P., Overington, J. P., Johnson, M. S., Eisenmenger, F. & Blundell, T. L. (1991) *Biochem. Soc. Symp.* **57**, 1–9
19. Jones, T. A. (1978) *J. Appl. Crystallogr.* **11**, 268–272
20. Jones, T. A. & Thirup, S. (1986) *EMBO J.* **5**, 819–822
21. van Gunsteren, W. F. & Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Nijenborgh vol. 16, Groningen, The Netherlands
22. Brocklehurst, S. M., Topham, C. M. & Brocklehurst, K. (1990) *Biochem. Soc. Trans.* **18**, 598–599
23. Brocklehurst, K., Carlsson, J., Kierstan, M. J. P. & Crook, E. M. (1973) *J. Biol. Chem.* **248**, 573–584
24. Brocklehurst, K., Carlsson, J. & Kierstan, M. P. J. (1985) *Top. Enzyme Ferment. Biotechnol.* **10**, 146–188
25. Robinson, G. W. (1975) *Biochemistry* **14**, 3695–3700
26. Stuchbury, T., Shipton, M., Norris, R., Malthouse, J. P. G., Brocklehurst, K., Herbert, J. A. L. & Suschitzky, H. (1975) *Biochem. J.* **151**, 417–432
27. Brocklehurst, K. (1982) *Methods Enzymol.* **87C**, 427–469
28. Brocklehurst, K., Baines, B. S. & Malthouse, J. P. G. (1981) *Biochem. J.* **197**, 739–746
29. Salih, E., Malthouse, J. P. G., Kowlessur, D., Jarvis, M., O'Driscoll, M. & Brocklehurst, K. (1987) *Biochem. J.* **247**, 181–193
30. Brocklehurst, K., Kowlessur, D., Patel, G., Templeton, W., Quigley, K., Thomas, E. W., Wharton, C. W., Willenbrock, F. & Szawelski, R. J. (1988) *Biochem. J.* **250**, 761–772
31. Brocklehurst, K., Kowlessur, D., O'Driscoll, M., Patel, G., Quenby, S., Salih, E., Templeton, W., Thomas, E. W. & Willenbrock, F. (1987) *Biochem. J.* **244**, 173–181
32. Kowlessur, D., Topham, C. M., Thomas, E. W., O'Driscoll, M. & Brocklehurst, K. (1989) *Biochem. J.* **258**, 755–764
33. Malthouse, J. P. G. & Brocklehurst, K. (1976) *Biochem. J.* **159**, 221–234
34. Erlanger, B. F., Kokowsky, N. & Cohen, W. (1961) *Arch. Biochem. Biophys.* **95**, 271–278
35. Bascombe, K. N. & Bell, R. P. (1959) *J. Chem. Soc.*, 1096–1184
36. Kernighan, B. W. & Ritchie, D. M. (1988) *The C Programming Language*, 2nd edn., Prentice Hall, Englewood Cliffs, NJ

37. Dixon, W. J. (chief ed.), Brown, M. B., Engelman, L., Frane, J. W., Hill, M. A. & Jennrich, R. I. (1988) *BMDP Statistical Software Manual*, vol. 1, pp. 389–417, University of California Press, Berkeley, CA
38. Dubois, T., Kleinschmidt, T., Schnek, A. G., Looze, Y. & Braunitzer, G. (1988) *Biol. Chem. Hoppe-Seyler* **369**, 741–754
39. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542
40. Greer, J. (1981) *J. Mol. Biol.* **153**, 1027–1042
41. Sutcliffe, M. J. (1988) PhD. Thesis, University of London
42. Summers, N. L., Carlson, W. D. & Karplus, M. (1987) *J. Mol. Biol.* **196**, 175–198
43. Hermans, J., Jr. & McQueen, J. E., Jr. (1974) *Acta Crystallogr.* **A30**, 730–739
44. Schack, P. (1967) *C.R. Trav. Lab. Carlsberg* **36**, 67–83
45. Kaarsholm, N. C. & Schack, P. (1983) *Acta Chem. Scand. Ser. B.* **37**, 607–611
46. Brocklehurst, K. & Salih, E. (1983) *Biochem. J.* **213**, 559–560
47. Polgár, L. (1984) *Biochem. J.* **221**, 555–556
48. Buttle, D. J. & Barrett, A. J. (1984) *Biochem. J.* **223**, 81–88
49. Salih, E. (1984) PhD. Thesis, University of London
50. Brocklehurst, K., McKee, R., Salih, E. & Smith, H. (1985) *J. Protein Chem.* **4**, 103–127
51. Brocklehurst, K., Salih, E., McKee, R. & Smith, H. (1985) *Biochem. J.* **228**, 525–527
52. Dubois, T., Jacquet, A., Schnek, A. G. & Looze, Y. (1988) *Biol. Chem. Hoppe-Seyler* **369**, 733–740
53. Jacquet, A., Kleinschmidt, T., Schnek, A. G., Looze, Y. & Braunitzer, G. (1989) *Biol. Chem. Hoppe-Seyler* **370**, 425–434
54. Jacquet, A., Kleinschmidt, T., Dubois, T., Schnek, A. G., Looze, Y. & Braunitzer, G. (1989) *Biol. Chem. Hoppe-Seyler* **370**, 819–829
55. Pickersgill, R. W., Sumner, I. G. & Goodenough, P. W. (1990) *Eur. J. Biochem.* **190**, 443–444
56. Baines, B. S. & Brocklehurst, K. (1982) *Biochem. J.* **205**, 205–211
57. Brocklehurst, K., Carey, P. R., Lee, H., Salih, E. & Storer, A. C. (1984) *Biochem. J.* **223**, 649–657
58. Brocklehurst, K. & Topham, C. M. (1990) *Biochem. J.* **270**, 561–562
59. Storer, A. C., Lee, H. & Carey, P. R. (1983) *Biochemistry* **22**, 4789–4796
60. Šali, A. & Blundell, T. L. (1990) *J. Mol. Biol.* **212**, 403–428
61. Oobatake, M. & Ooi, T. (1977) *J. Theor. Biol.* **67**, 567–584
62. Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989) *J. Mol. Biol.* **206**, 759–777
63. Sibanda, B. L. & Thornton, J. M. (1985) *Nature (London)* **316**, 170–174
64. Wilmot, C. M. & Thornton, J. M. (1988) *J. Mol. Biol.* **203**, 221–232
65. Templeton, W., Kowlessur, D., Thomas, E. W., Topham, C. M. & Brocklehurst, K. (1990) *Biochem. J.* **266**, 645–651
66. Pickersgill, R. W., Sumner, I. G., Collins, M. E. & Goodenough, P. W. (1989) *Biochem. J.* **257**, 310–312
67. Schack, P. & Kaarsholm, N. C. (1984) *Biochemistry* **23**, 631–635
68. Kowlessur, D., Thomas, E. W., Topham, C. M., Templeton, W. & Brocklehurst, K. (1990) *Biochem. J.* **266**, 653–660
69. Brocklehurst, K., Willenbrock, S. J. F. & Salih, E. (1983) *Biochem. J.* **211**, 701–708
70. Ménard, R., Khouri, H. E., Plouffe, C., Dupras, R., Ripoll, D., Vernet, T., Tessier, D. C., Laliberté, F., Thomas, D. Y. & Storer, A. C. (1990) *Biochemistry* **29**, 6706–6713
71. Drenth, J., Kalk, K. H. & Swen, H. M. (1976) *Biochemistry* **15**, 3731–3738
72. Overington, J., Johnson, M. S., Šali, A. & Blundell, T. L. (1990) *Proc. R. Soc. London Ser. B* **241**, 132–145
73. Richmond, T. J. & Richards, F. M. (1978) *J. Mol. Biol.* **119**, 537–555
74. Cohen, L. W., Coghlan, V. M. & Ditel, L. C. (1986) *Gene* **48**, 219–227

Received 13 March 1991/20 May 1991; accepted 28 May 1991