Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An analysis of subdomain orientation, conformational change and disorder in relation to crystal packing of aspartic proteinases

D. Bailey,[a]‡ E. P. Carpenter,[b]‡ A. Coker,[c] S. Coker,[c] J. Read,[d] A. T. Jones,[e] P. Erskine,[c] C. F. Aguilar,[f] M. Badasso,[g] L. Toldo,[h] F. Rippmann,[h] J. Sanz-Aparicio,[i] A. Albert,[i] T. L. Blundell,[j] N. B. Roberts,[k] S. P. Wood[c] and J. B. Cooper[c]*

[a]Incisive Media, 32–34 Broadwick Street, London W1A 2HG, England, [b]Structural Genomics Consortium, Old Road Campus Research Building, Old Road Campus, Roosevelt Drive, Headington, Oxford OX3 7DQ, England, [c]Laboratory for Protein Crystallography, Centre for Amyloidosis and Acute Phase Proteins, UCL Division of Medicine (Royal Free Campus), Rowland Hill Street, London NW3 2PF, England, [d]AstraZeneca, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, England, [e]Welsh School of Pharmacy, Cardiff University, Cardiff CF10 3NB, Wales, [f]Laboratorio de Biologia Molecular Estructural, Universidad Nacional de San Luis, Avenida Ejercito de los Andes 950, Bloque I, 5700 San Luis, Argentina, [g]Augsburg College, 2211 Riverside Avenue, Minneapolis, MN 55454, USA, [h]Merck KGaA, Merck Serono, Frankfurter Strasse 250, 64271 Darmstadt, Germany, [i]Department of Crystallography and Structural Biology, Instituto de Quimica-Fisica Rocasolano, CSIC, c/Serrano 119, E-28006 Madrid, Spain, [j]Department of Biochemistry, University of Cambridge, Old Addenbrooke's Site, 80 Tennis Court Road, Cambridge CB2 1GA, England, and [k]Clinical Chemistry Research Unit, School of Medicine, University of Liverpool, Liverpool L69 3BX, England

‡ To be considered as joint first authors.

Correspondence e-mail: jon.cooper@ucl.ac.uk

Received 22 November 2011
Accepted 4 February 2012

PDB References:
endothiapepsin–DB5, 3uri; endothiapepsin–DB4, 3urj; endothiapepsin–DB6, 3url; pepsin 3b, 3utl.

The analysis reported here describes detailed structural studies of endothiapepsin (the aspartic proteinase from *Endothia parasitica*), with and without bound inhibitors, and human pepsin 3b. Comparison of multiple crystal structures of members of the aspartic proteinase family has revealed small but significant differences in domain orientation in different crystal forms. In this paper, it is shown that these differences in domain orientation do not necessarily correlate with the presence or absence of bound inhibitors, but appear to stem at least partly from crystal contacts mediated by sulfate ions. However, since the same inherent flexibility of the structure is observed for other enzymes in this family such as human pepsin, the native structure of which is also reported here, the observed domain movements may well have implications for the mechanism of catalysis.

1. Introduction

Aspartic proteinases are ubiquitous in nature owing to their roles in numerous biological processes and have important commercial applications (Cooper, 2002; Dunn, 2002). They occur in retroviruses and fungi as well as in the entire plant and animal kingdoms. The digestive stomach enzymes of mammals include the aspartic proteinases pepsin and gastricsin as well as chymosin, which is found in many suckling mammals. The lysosomal enzymes cathepsin D and E also belong to this family. In addition, there is the plasma aspartic proteinase renin which is synthesized mainly by the kidneys and submaxillary glands and has a key role in regulating blood pressure. Fungal aspartic proteinases, such as endothiapepsin and mucorpepsin, are secreted and have an extracellular digestive function. In addition, enzymes of this family are found abundantly in the seeds, leaves and flowers of various plant species.

Aspartic proteinases are associated with numerous pathological processes including hypertension (renin), gastric ulcer disease (pepsin), muscular dystrophy and neoplastic diseases (cathepsins D and E), human immunodeficiency virus (HIV), tissue invasion and virulence of fungal pathogens (*e.g. Candida albicans*) and the avid digestion of haemoglobin by the malarial parasite (plasmepsins) (Cooper, 2010). Calf chymosin is specific for its substrate κ-casein, which is cleaved at the Phe105–Met106 bond and has low general proteolytic activity (Kumar *et al.*, 2010). In addition to chymosin, various microbial and plant aspartic proteinases have been used in milk coagulation by the cheese-production industry. The plant aspartic proteinases, such as those from barley and the flowers of *Cynara*, show the greatest sequence and structural similarity

to the lysosomal enzymes cathepsin D and saccharopepsin from the yeast vacuole (Frazão *et al.*, 1999).

Aspartic proteinases are a structurally homologous family of proteolytic enzymes that are generally most active under acidic conditions (Cooper, 2002; Dunn, 2002). They differ distinctly from the other major proteolytic enzymes, the serine proteases, cysteine proteases and metalloproteinases, by having two catalytically essential aspartic acid residues. The aspartates are in highly conserved Asp-Thr-Gly sequences and are found at positions 32 and 215 in the traditional family archetype porcine pepsin (Pearl & Blundell, 1984). Typically, enzymes in this class consist of a single polypeptide chain with a molecular weight of between 35 000 and 50 000 Da and an amino-acid sequence of between 320 and 440 residues in length. The retroviral proteinases have molecular weights of less than half of those of the eukaryotic enzymes and are catalytically active as dimers (Lapatto *et al.*, 1989; Wlodawer *et al.*, 1989).

Mammalian aspartic proteinases are synthesized as zymogens with an N-terminal pro-region and are subsequently activated to give functional mature enzymes. Structural and other studies of porcine pepsinogen have indicated the mechanism of activation (James & Sielecki, 1986). The precursor region inhibits the catalytic activity by virtue of a conserved lysine and two tyrosine side chains that form hydrogen bonds to the active-site aspartic acids. Upon secretion of the proenzyme into the acidic lumen of the stomach, protonation of the active-site carboxylates leads to a weakening of their interactions with the lysine of the pro-region. The resulting pH-induced conformational change is followed by limited proteolysis to yield the mature enzyme.

The active-site aspartates of enzymes in this family are located close to each other in the three-dimensional structure and are involved in hydrogen bonding to adjacent residues stabilizing the catalytic centre (Pearl & Blundell, 1984; Davies, 2000). The aspartic proteinase fold consists predominantly of β-sheet and has a bilobal structure with an extended binding cleft that accommodates 9–10 residues of the polypeptide substrate. The active-site cleft lies between the two domains of the molecule that are formed by the N- and C-terminal halves of the polypeptide. These two domains are very similar in topology and are related by a local twofold axis of pseudo-symmetry. The sequences related by this inter-domain dyad are most strongly conserved in the residue ranges 30–42 and 213–225, which include the two highly conserved Asp-Thr-Gly motifs. The active site is also covered by a β-hairpin 'flap' (residues 74–77) in the N-terminal lobe. The N-terminal domains are more conserved than the C-terminal domains, in which larger differences can occur in the loop regions of family members.

Endothiapepsin, which is produced by the fungus *Endothia parasitica*, was first isolated and crystallized by Sardinas (1968) and has been obtained in a number of different crystal forms with and without bound inhibitors (Tables 1 and 2). Moews & Bunn (1970) first reported three crystal forms (types I–III) grown at various pH values in the presence of various organic solvents with 2.2 *M* ammonium sulfate as the precipitant. Type I crystals grew at pH values between 4.5 and 6.3 when a small quantity of acetone (0.25–1.00%) was added. Crystals with the type I unit cell were used to solve the native endothiapepsin structure at 2.1 Å resolution (Blundell *et al.*, 1990). Whilst two further crystal forms, types II and III, were originally reported by Sardinas (1968), an additional form was reported more recently by Šali *et al.* (1989). Crystals of type IV appeared in the same crystallization conditions as type I and diffracted to appreciably higher resolution, although they usually took longer to grow, sometimes appearing in mother liquor which had previously yielded type I crystals. Type IV crystals were almost exclusively only obtained in the presence of active-site-directed inhibitors. The two crystal forms most suitable for X-ray analysis (types I and IV) both belong to space group $P2_1$ and have the following unit-cell parameters: type I, a = 54, b = 74, c = 46 Å, β = 110°; type IV, a = 43, b = 76, c = 43 Å, β = 97°.

High-resolution analyses of inhibitor complexes have shown that type I and type IV crystals of endothiapepsin have slightly different subdomain orientations. Comparison of the two forms showed that their conformational differences could be best described as a rigid-body rotation of residues in the range 190–302 with respect to the rest of the protein (Šali *et al.*, 1989, 1992).

Endothiapepsin has been used extensively to analyse the binding of numerous inhibitors that were designed to be specific for human renin as therapeutic antihypertensives (Bailey & Cooper, 1994). More recently, it has been subjected

Table 1
Unit-cell parameters for the different crystal forms of endothiapepsin.

Types I–III were originally reported by Moews & Bunn (1970), but only the unit-cell parameters of types I and II were determined rigorously.

Unit cell	a (Å)	b (Å)	c (Å)	β (°)	Unit-cell volume (Å^3)	Space group
Native						
Type I	54	74	46	110	170000	$P2_1$
Type II	56	60	46	101	150000	$P2_1$
Type III†	75	85	50	—	—	—
Type IV	43	76	43	97	139000	$P2_1$
Complexed						
Type I	54	74	46	110	170000	$P2_1$
Type IV	43	76	43	97	139000	$P2_1$

† Unit-cell parameters were not rigorously determined.

Table 2
Crystallization conditions for different crystal forms of endothiapepsin.

Unit cell	pH	Buffer	Solvent	Reference
Native				
Type I	4.5–6.3	0.1 *M* NaOAc or NaH_2PO_4	0.25–1.00%(*w*/*w*) acetone	Moews & Bunn (1970)
Type II	4.5	0.1 *M* NaOAc or NaH_2PO_4	0.2% DMF	Moews & Bunn (1970)
Type III	4.5–6.3	0.1 *M* NaOAc or NaH_2PO_4	<0.1% acetone	Moews & Bunn (1970)
Type IV	4.5	0.1 *M* NaOAc	Acetone	This work
Complexed				
Type I	4.5	0.1 *M* NaOAc	Acetone	Foundling *et al.* (1987)
Type IV	4.5	0.1 *M* NaOAc	Acetone	Šali *et al.* (1989)

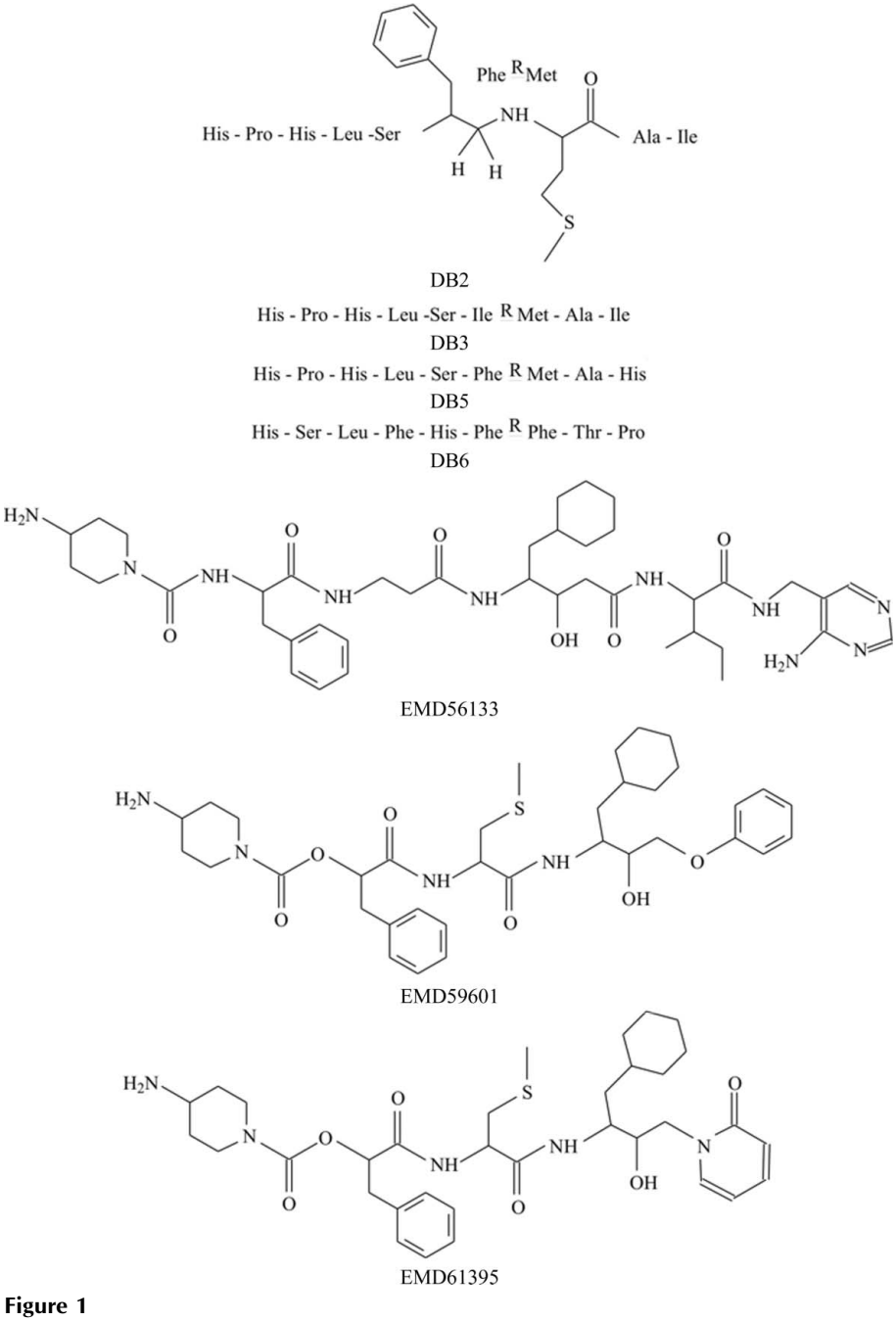

Figure 1
Chemical formulae of the inhibitors. All chiral centres are of *S* configuration.

to neutron diffraction and NMR studies, which have shed much light on the catalytic mechanism (reviewed in Coates *et al.*, 2006), and it continues to be used as a model system for fragment-based drug screening (Geschwindner *et al.*, 2007; Koster *et al.*, 2011). In this paper, we report a number of previously unpublished inhibitor complexes of endothiapepsin. The following inhibitors were successfully cocrystallized with the enzyme: DB2, DB5, DB6, EMD56133, EMD59601 and EMD61395 (formulae are shown in Fig. 1). We also report the structure of uncomplexed endothiapepsin in the type IV unit cell at high resolution. These crystals were formed in the presence of two inhibitors (DB3 and PD134685) which did not appear in the final electron-density map. No structural data for truly uncomplexed type IV crystals have been available until now. This work has allowed comparison of the crystal packing and rigid-body movements in the two crystal forms of endothiapepsin without a bound inhibitor. Type IV crystals are stabilized by the presence of three sulfate ions at points of intermolecular contact and the unit cell is significantly smaller in volume and thus in solvent content. We demonstrate that the presence of these three sulfate ions appears to be essential for the type IV crystals, as they appear not only in the newly reported native structures but also in all inhibitor complexes of this type, and never in type I crystals. We also present the X-ray crystal structure of human pepsin 3b in its native state and demonstrate that a similar rigid-body movement occurs in this enzyme.

2. Methods

2.1. Crystallization

A number of reduced-bond inhibitors prefixed DB (see Fig. 1) with sequences designed to match the known cleavage sites of chymosin and mucorpepsin in the insulin B-chain were synthesized using standard solid-phase Boc methods (details provided as Supplementary Material[1]). A second series of inhibitors prefixed EMD (Fig. 1) that were designed for activity against human renin were also studied. These have an N-terminal group similar to morpholine at P4, a cyclohexylstatine group spanning P1–P1′ and various C-terminal heteroaromatic groups. One of them, EMD56133, possesses a group at P2 which is similar to glycine but has an extra aliphatic backbone carbon, while the other two (EMD59601 and EMD61395) possess *S*-methylcysteine residues at this position.

For growth of the endothiapepsin inhibitor-complex crystals, a modification of the method of Moews & Bunn (1970) was used. Freeze-dried enzyme and inhibitor were mixed in 0.1 *M* acetate buffer pH 4.5 to give a tenfold molar excess of inhibitor over enzyme; the final enzyme concentration was

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: YT5038). Services for accessing this material are described at the back of the journal.

Table 3
Crystallographic data-collection and refinement statistics.

Note that the r.m.s.d.s for bond angles are actually the corresponding 1–3 atom distance deviations except where indicated with an asterisk. Dashes indicate parameters that are either not applicable or relate to legacy structures for which the reflection data or specific data-processing statistics are no longer available. Where possible, statistics for the outer resolution shell are given in parentheses and the Cruickshank diffraction precision index (DPI; Cruickshank, 1996) is given.

	PD134685	DB2	DB3	DB5	DB6	EMD56133	EMD59601	EMD61395	Pepsin 3b
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Type	IV	I	IV	IV	IV	I	I	I	—
Unit-cell parameters									
a (Å)	43.2	53.8	43.1	43.0	43.1	54.2	53.9	53.5	50.9
b (Å)	75.7	74.5	75.7	75.6	76.0	74.7	74.1	73.8	75.3
c (Å)	42.9	46.0	42.9	42.8	42.9	46.2	45.8	45.3	87.1
β (°)	97.1	109.8	97.0	97.0	97.0	109.5	110.1	110.1	90
Data collection									
Source	Rotating anode	Rotating anode	Rotating anode	Sealed tube	Rotating anode	Rotating anode	Rotating anode	Rotating anode	SRS 9.5
Detector	FAST	MAR IP	FAST	CAD-4	MAR IP	MAR IP	MAR IP	MAR IP	MAR IP
Processing	*MADNES*	*MOSFLM*	*MADNES*	*CADRAL*	*MOSFLM*	*MOSFLM*	*MOSFLM*	*MOSFLM*	*DENZO*
Resolution (Å)	2.0	2.05	1.9	2.1	2.0	2.05	1.9	2.05	2.6
Unique reflections	18487	21510	19248	10301	17271	21735	26434	20712	10426
Completeness (%)	92.4	98.9 (94.3)	90.7	60.0†	92.8 (61.7)	93.1 (90.5)	97.3 (85.5)	94.8 (89.2)	96.7 (82.0)
Multiplicity	3.0	3.8 (3.7)	4.3	—	3.3 (2.7)	4.0 (4.0)	3.7 (3.5)	3.8 (3.7)	4.2 (2.4)
$R_{\rm merge}$ (%)	6.6	12.3 (46.1)	4.6	15.0	12.7 (41.9)	10.2 (32.9)	9.2 (26.3)	12.5 (37.9)	3.3 (12.4)
Refinement									
R factor (%)	18.9	18.2	14.9	14.9	14.7	18.6	17.6	19.1	13.0
$R_{\rm free}$ (%)	—	23.7	21.3	25.3	23.7	22.1	20.0	23.4	20.0
R.m.s.d. bonds (Å)	0.025	0.003	0.006	0.008	0.006	0.003	0.003	0.003	0.015
R.m.s.d. angles (Å)	0.049	0.009	0.021	0.022	0.051	0.010	0.010	0.010	1.8*
DPI (Å)	—	0.18	0.18	—	0.20	0.20	0.13	0.21	0.26
PDB code	‡	1e5o	3uri	3urj	3url	1e80	1e82	1e81	3utl

† The low completeness of the DB5 data set is a consequence of radiation damage during the long data collection using a single-counter diffractometer, which required the use of two crystals. Accordingly, the DPI could not be determined for this structure. However, the electron density was of satisfactory quality, perhaps because the data to medium resolution (∼3 Å) are substantially complete. ‡ This structure is equivalent to that obtained with DB3 (*i.e.* type IV native endothiapepsin) and therefore has not been deposited as a separate entry in the PDB.

2 mg ml^{-1}. In cases where the inhibitor was poorly soluble, it was stirred together with the enzyme overnight. Finely powdered ammonium sulfate was then added to give a 2.2 *M* (55% saturated) solution, which was then Millipore-filtered. Any remaining turbidity was removed by the addition of a few drops of acetone. Each crystallization batch consisted of approximately 2 ml of the above solution that was sealed in a glass bijou bottle. Crystals of the complexes were obtained within a few weeks or months and were stable almost indefinitely in the mother liquor.

Crystals of human pepsin 3b purified from gastric juice (Jones *et al.*, 1993) were grown by vapour diffusion from a 16–30 mg ml^{-1} protein solution in 30% saturated ammonium sulfate buffered with 200 m*M* formate pH 5.2.

2.2. Structure analysis of type IV endothiapepsin crystals

X-ray data from the crystals of endothiapepsin grown in the presence of various inhibitors were collected using in-house rotating-anode sources and were processed using the software described in Table 3, in which the corresponding detectors that were used are also indicated. The coordinates of the enzyme moiety in the type IV crystal form were used with the above data for the calculation of $\sigma_{\rm A}$-weighted difference Fourier maps (Read, 1986). Following interpretation of the resulting electron density, the model of the enzyme, inhibitor, solvent structure and bound sulfates was refined by stereochemically restrained least squares using *RESTRAIN* (Haneef *et al.*, 1985) or *SHELX* (Sheldrick, 2008).

2.3. Structure analysis of human pepsin

Crystals of human pepsin 3b were mounted in glass capillaries and synchroton data were then collected on Daresbury SRS beamline 9.5 using a MAR Research image plate with the crystal cooled to a temperature of 277 K. The data were processed using *DENZO* (Otwinowski & Minor, 1997) to a resolution of 2.6 Å, which showed that the crystal belonged to space group $P2_12_12_1$, with unit-cell parameters a = 50.9, b = 75.3, c = 87.0 Å. The structure was determined by molecular replacement using porcine pepsin as the search model with the programs *AMoRe* (Navaza, 1994) and *TFFC* (Driessen *et al.*, 1991). The structure was then refined using *X-PLOR* (Brünger *et al.*, 1998), *RESTRAIN* (Haneef *et al.*, 1985) and, more recently, *REFMAC* (Murshudov *et al.*, 2011).

2.4. Inhibition kinetics

Kinetic studies of the DB inhibitor series were carried out using the chromogenic substrate KPLEFF$^{\rm NO_2}$RL, which has $\lambda_{\rm max}$ = 279 nm and ε_{279} = 10 000 M^{-1} cm^{-1}, thereby allowing the initial substrate concentration to be determined at this wavelength. Peptide-hydrolysis rates were monitored on a Lambda 2 Perkin Elmer double-beam UV–Vis spectrophotometer. A bathochromic shift occurs on hydrolysis of the above substrate, allowing the reaction to be followed at a

wavelength of 300 nm. The enzyme and substrate concentrations were 100 p*M* and 40 µ*M*, respectively, and the temperature was maintained at 310 ± 0.2 K. The EMD inhibitor series was studied with the same substrate at the same concentration using an Applied Photophysics Bio Sequential SX.17MV stopped-flow reaction analyser and an enzyme concentration of 10 n*M*. The pH of each assay was maintained at 4.6 using 0.1 *M* sodium acetate buffer with 0.1 *M* NaCl and all solutions were passed through a filter of 0.22 µm pore size.

2.5. Structural validation and comparison

The quality of each protein structure has been assessed using the program *PROCHECK* (Laskowski *et al.*, 1993). Since we report for the first time a number of legacy structures which were solved around 20 years prior to publication, reflection data are available for most, but not all, coordinate sets. However, all contemporary refinement statistics are satisfactory which, together with the fact that these structures all belong to a closely related and well studied family, suggests that there is no cause for concern as to their validity. Wherever possible, the coordinates of structures reported in this paper and reflection data have been submitted to the Protein Data Bank (http://www.wwpdb.org), and the corresponding accession codes are shown in Table 3.

Pairwise superposition of protein molecules and rigid-body domains was performed using *DynDom* (Poornam *et al.*, 2009). Intermolecular contacts were identified by the program *CONTACTS* (Winn *et al.*, 2011) using a cutoff distance of 4.0 Å.

The B_{iso} isotropic displacement parameters in all type I and type IV structures were scaled to those of the 1.9 Å resolution native type IV structure to provide a convenient means of comparison. The scaling was based on the average B_{iso} value of each structure and was intended to correct empirically for systematic differences and for the fact that in some of the structures U_{iso} values were refined instead of B_{iso} values (where $B_{iso} = 8\pi^2 U_{iso}$). Overall, these scaled and residue-averaged B_{iso} values ($B_{ave,j}$) enabled qualitative comparison of the disorder in regions with and without lattice contacts.

3. Results and discussion

3.1. Quality of structures

The stereochemical quality analysis program *PROCHECK* (Laskowski *et al.*, 1993) showed that all structures reported in this work score satisfactorily for their respective resolutions. The χ_1 side-chain angles show the familiar grouping into the three energetically favourable conformations, with ideal angles of ±60° or 180° (Janin *et al.*, 1978). Interestingly, endothiapepsin has a large number of serine (49/330) and threonine (47/330) residues and this leads to the least favourable χ_1 conformation (g^-) being the second most populated of the three. Indeed, atomic resolution analyses of native and inhibitor-bound endothiapepsin established that many of these side chains exhibit dual and in some cases triple conformations (Coates *et al.*, 2002, 2006; Erskine *et al.*, 2003).

Essentially full-length electron density was visible for the three EMD compounds analysed as well as for DB5 and DB6. All of the bound inhibitors were observed to adopt extended β-strand conformations in the active-site cleft and to form a common set of hydrogen bonds which have been described in detail elsewhere (see, for example, Bailey & Cooper, 1994). However, the maps for endothiapepsin cocrystallized with PD134685 and DB3 did not reveal any electron density for the inhibitor moiety. Intriguingly, both of these inhibitors gave type IV crystals, which generally only form when an inhibitor is bound. The absence of bound inhibitor with PD134685 was probably a result of the low solubility of the inhibitor (5–10 µg ml^{-1}) in the crystallization liquor. However, the other inhibitor, DB3, was very soluble, suggesting that other factors precluded its analysis.

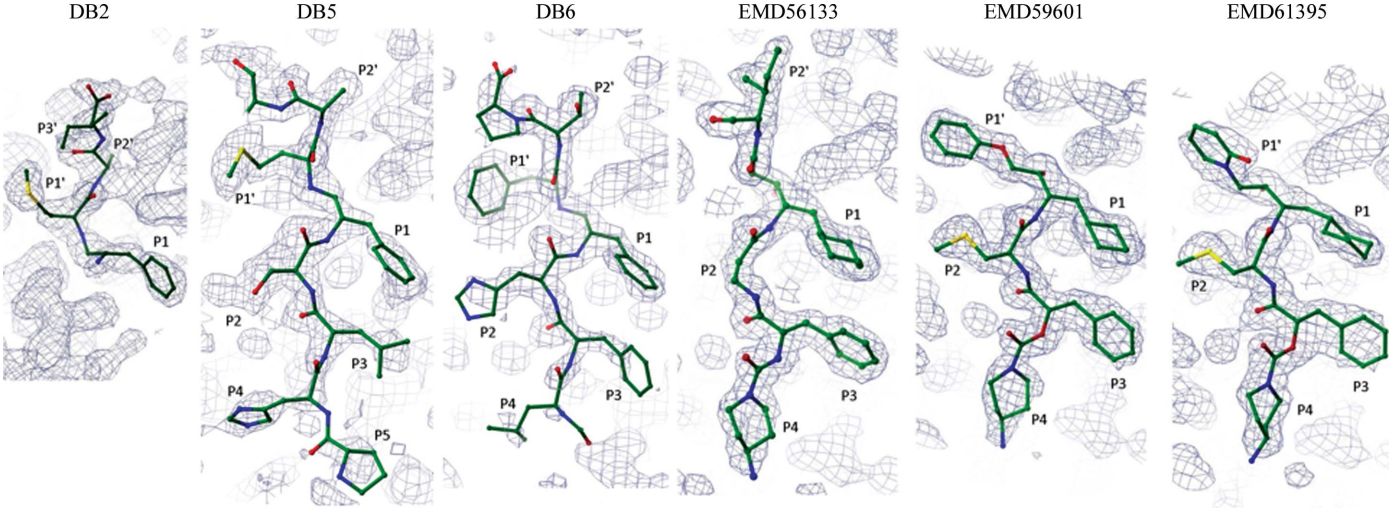

Figure 2
The refined electron density for the bound inhibitors drawn with a contour level of 1.0$\sigma(\rho)$.

3.2. Inhibitor conformation, stability and kinetics

The intriguing absence of electron density for DB3 in the structure might arise from either weak binding to the enzyme or degradation of the inhibitor during cocrystallization. To test these possibilities, HPLC was used to analyse samples taken from the mother liquor and these were compared with samples of the inhibitor DB3 alone, of the enzyme alone and of fresh enzyme–inhibitor mixtures that were incubated for different times over 24 h. This was performed using a Varian 5000 liquid chromatograph with a Brownlee Aquapore (300 Å pore size) silica C18 column running an acetonitrile gradient. Distinct peaks attributable to the enzyme and inhibitor were observed in the samples from a fresh enzyme–inhibitor mixture. The inhibitor peak was seen to decrease with increasing incubation time, and additional peaks, most likely arising from degradation products, were seen to increase (Bailey, 1994). With the sample taken from the mother liquor, which was approximately ten months old at the time of analysis, very little intact DB3 was present in spite of it being added in a large molar excess.

Thus, X-ray and HPLC analysis of the endothiapepsin–DB3 complex showed that this inhibitor was actually cleaved by the enzyme. Similarly, our analysis of the complex of endothiapepsin with DB2 suggested that cleavage of the inhibitor between the P2 and P1 residues had occurred since electron density was only visible for the P1–P3′ residues in this structure (Fig. 2). However, the high-resolution structures of endothiapepsin complexed with DB5 and DB6 clearly showed full-length electron density for these inhibitors which was essentially contiguous from P4 to P3′ (Fig. 2), thereby confirming the correct synthesis of these compounds.

Williams *et al.* (1972) delineated the specificity of endothiapepsin by its action on the oxidized β-chain of insulin. The enzyme has a preference for hydrophobic regions, as reflected in the relative rates of hydrolysis:

$$\text{Phe24-Phe25} > \text{Tyr16-Leu17} > \text{Gln4-His5} >>> \text{Leu11-Val12} > \text{Asn3-Gln4}.$$

Intriguingly, Drohse & Foltmann (1989) found that the milk-clotting activity of endothiapepsin does not arise from cleavage of κ-casein at the expected Phe-Met dipeptide which is cleaved by chymosin and other commercially used milk-clotting enzymes. Instead, endothiapepsin cleaves at the preceding dipeptide, which has the sequence Ser-Phe, as indicated in Fig. 3.

Unfortunately, there are no Ser-Phe dipeptides in the β-chain of insulin to allow measurement of the relative rate of cleavage at this sequence. However, it seems plausible that endothiapepsin may have cleaved DB3 at the Ser-Phe peptide bond and the remaining peptide fragments would then simply dissociate owing to weak binding. With this inhibitor, the presence of the β-branched side chain of Ile at P1, which is unfavourable for binding to the pepsins in general (see, for example, Fruton, 1976), may have precluded the normal binding mode of the inhibitor in which the transition-state analogue resides at the catalytic centre. Accordingly, modelling an Ile at P1 shows that the branching of the side chain at C^{β} causes unfavourable van der Waals contacts with the enzyme.

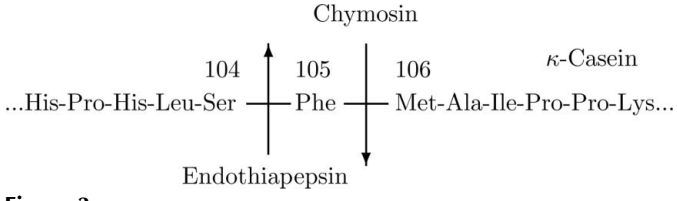

Figure 3
Cleavage of κ-casein by endothiapepsin and chymosin.

Table 4
IC_{50} values determined for endothiapepsin.

The values in parentheses indicate the corresponding K_i (n*M*). The dash indicates where a measurement was not made.

Inhibitor	IC_{50} (n*M*)
DB2	20 (2)
DB3	2000 (162)
DB4	10 (1)
DB5	10 (1)
DB6	—
EMD56133	400 (32)
EMD59601	400 (32)
EMD61395	6000 (486)

Kinetic measurements of the DB inhibitor series with endothiapepsin (shown in Table 4) establish that DB3 has an affinity for endothiapepsin that is two orders of magnitude lower than those of DB2 and DB5. This difference in IC_{50} suggests why DB2, DB5 and DB6 were observed in the crystal structures (Fig. 2) but the more weakly binding and proteolytically vulnerable inhibitor DB3 was not. The IC_{50} values for the EMD compounds were determined with endothiapepsin and found to be in the nanomolar range. For human renin, the IC_{50} values were found to be 56 n*M* for EMD59601 and 41 n*M* for EMD61395, thus confirming their tight-binding nature.

Of the EMD series of compounds, the inhibitor EMD56133 reveals an intriguing effect owing to the unusual residue at the P2 position which essentially mimics glycine with an additional main-chain CH_2 group. The extra main-chain carbon might be expected to introduce a frameshift in the binding mode of the inhibitor, potentially disrupting some of the main-chain hydrogen-bonding interactions. In addition, the lack of a P2 side chain might also be expected to weaken binding affinity. However, it is very clear that all of the conserved main-chain hydrogen bonds are preserved in this inhibitor complex. This is achieved partly by virtue of the P2 main chain sitting appreciably deeper in the otherwise empty S2 pocket than it does in the other complexes. Whilst there are slight repercussions for the P1–P2′ residues, resulting in them being appreciably displaced by about 0.5 Å with respect to the other inhibitors in this series, the essential hydrogen-bond interactions are conserved, as are the side-chain conformations. The inhibitors EMD59601 and EMD61395 possess large non-peptidic aromatic groups at the P1′ position and in spite of their dissimilarity from the physiological substrate they bind in conformations in the S1′ subsite that are broadly similar to

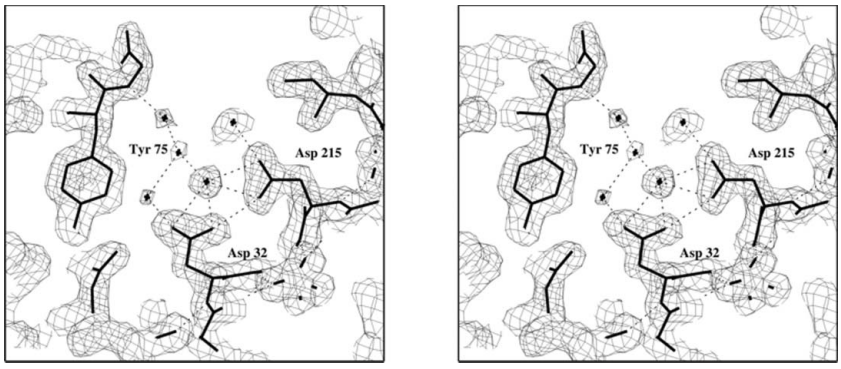

Figure 4
A stereoview of the electron density in the active-site cleft for the inhibitor DB3 which did not bind to the enzyme. The catalytic aspartates (32 and 215), intervening water molecule(s) and Tyr75 of the active-site flap are clearly visible.

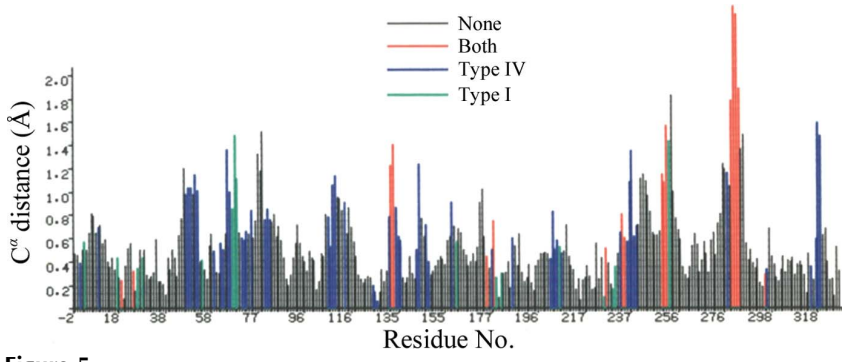

Figure 5
Local differences following superposition of the rigid bodies. The distances in Å between the C^{α} atoms of type I and type IV endothiapepsin following superposition of the two rigid-body subdomains are shown *versus* residue number. Residues involved in lattice contacts in either the type I or type IV crystal forms or both are colour-coded green, blue and red, respectively.

those of other inhibitors (Fig. 2). Since these two inhibitors only differ appreciably at the P1′ position, the significantly weaker binding of EMD61395 must stem from the less favourable binding of its bulky P1′ group within the tight contraints of the S1′ pocket. The C-terminal residues in inhibitor EMD56133 (the largest of the three EMD compounds) follow the usual path, although interestingly the substituted ring at P3′ occupies what is typically a main-chain rather than a side-chain position. Of the DB inhibitor series, DB5 and DB6 were observed to fully occupy the enzyme active-site cleft from P5 to P3′ and form the expected set of hydrogen-bond interactions (Bailey & Cooper, 1994). In contrast, as mentioned above, DB2 appears to have undergone cleavage of the P2–P1 peptide bond such that electron density is only visible from P1 to P3′.

3.3. The type IV native endothiapepsin room-temperature structure

Both of the high-resolution type IV native structures (obtained with DB3 and PD134685) show well defined water molecules in the active site but no electron-density features for the inhibitor. The electron-density maps for these structures show evidence of the water molecule that is tightly bound between the two catalytic aspartates, as shown for DB3 in Fig. 4. Other solvent molecules bound to Asp32 and Asp215 have no equivalents in the type I crystals of the native enzyme. A network of solvent molecules extends from the catalytic water towards the active-site flap.

The C^{α} r.m.s. difference between the two native type IV structures after superposition was 0.15 Å (330 equivalent C^{α} atoms), suggesting that they are identical within the errors of the analysis. The r.m.s. difference between the room-temperature native enzyme structure (Blundell *et al.*, 1990) and the inhibitor complexes forming type I crystals is usually in the range 0.1–0.2 Å (330 equivalent C^{α} atoms). When the type I native structure was superposed on the highest resolution type IV native structure (DB3), the C^{α} r.m.s. difference was significantly greater at 0.68 Å (330 equivalent C^{α} atoms). The difference between the enzyme in the type I and type IV crystals can be attributed to a rigid-body shift in the molecule which affects a large part of the C-terminal domain (residues 190–302; Šali *et al.*, 1989). The two rigid bodies in the type I and type IV native structures were then fitted onto each other and a plot of the distance between each C^{α} atom against residue number was drawn (Fig. 5). This indicates regions where further local differences exist between the type I and type IV structures and shows, as is to be expected, that they differ most in the regions that are involved in intermolecular contacts in one or both of the crystal forms.

3.4. Rigid-body shifts

The rigid-body shifts between the type I and type IV endothiapepsin structures were originally analysed using the method of Šali *et al.* (1989) and, more recently, with *DynDom* (Poornam *et al.*, 2009). Specifically, the structures were initially superposed with the type I native structure using rigid body 1 alone. The rotations and translations of rigid body 2 (residues 190–302) were then determined relative to this region in the type I native structure (Fig. 6). This indicated that the rigid-body movements found in the type IV native structures were similar to those found in the type IV inhibitor-complexed structures. The rigid-body shift would therefore appear to be mainly associated with the change in crystal form rather than the presence of an inhibitor in the active site.

When inhibitors are bound, there are small but significant differences in the rigid-group rotations between different complexes in both the type I and the type IV crystal forms (see Fig. 6). The fact that the rigid-group rotations and translations within the molecule are very similar for the ‘complexes’ in which the ligand could not be ‘seen’, as shown in Fig. 6,

provides some evidence that the inhibitors were not bound when the enzyme crystallized.

3.5. Sulfate ions

Inspection of the electron-density maps for type IV native structures showed three sulfate ions (Fig. 7) in the same positions as those identified in the type IV endothiapepsin–inhibitor complexes. All three are involved in intermolecular contacts at the protein surface (see Supplementary Table 5). Sulfate 1 is found near the N-terminus of helix h_{N2} (nomenclature of Blundell *et al.*, 1990; residues 108–114), *i.e.* it is in a region which is expected to have a partial positive charge owing to the helix dipole. All three sulfates are hydrogen-bonded to symmetry-related molecules, although it can be seen in Fig. 7(*c*) (and Supplementary Table 5) that sulfates 1 and 3 are involved in more intermolecular contacts than sulfate 2. Most contacts (∼70%) are polar and all residues involved except one (Pro133) are polar.

All type IV endothiapepsin structures have sulfate ions in the same place, but none of the type I structures appear to have any. Sulfate binding therefore appears to be characteristic of type IV crystal formation.

3.6. Intermolecular contacts

The smaller unit-cell volume of type IV crystals (139 000 Å^3) compared with type I crystals (170 000 Å^3) gives them an approximately 20% lower solvent content and many more intermolecular contacts. As expected, the type I and type IV crystals have completely different intermolecular contacts; those which are conserved in each crystal form (*i.e.* are present in more than 70% of type I and IV crystals) are listed in Supplementary Tables 6 and 7.

In type IV crystals each molecule has six symmetry-related neighbours making 43 conserved protein contacts closer than 4 Å, 22 of which are polar. This compares with type I crystals in which each molecule has three symmetry-related neighbours making 23 conserved protein–protein intermolecular contacts including six conserved polar contacts. Thus, there are many more intermolecular contacts in type IV crystals than in type I crystals.

In type I crystals, the conserved set of contacts made with the molecule at $(x, y, 1 + z)$ occurs between the loops 175–181 on one molecule and 279–281 on the other. The contacts with the molecule at $(x + 1, y, 1 + z)$ involve the loops 250–252 and the two loops 66–68 and 133–134. The contact with the molecule at $(1 - x, y + 1/2, -z)$ involves the loop between residues 17 and 26 interacting with the h_C helix (residues 225–233). None of the protein–protein contacts involve residues from one molecule interacting with more than one rigid body of another molecule.

In the type IV crystals, a sulfate ion forms a link between the molecule at (x, y, z) and residues 318 and 319 of the adjacent molecule at $(x, y, 1 + z)$. This sulfate is important for the formation of type IV structures, since it is also involved in interactions with the molecule at $(-x, y + 1/2, 1 - z)$, thus further stabilizing the structure. This intermolecular contact involves a large interface including the sheet strands a'_N (residues 70–74) and b'_N (residues 80–81) and the helix h_{N2} (residues 106–108) on one molecule. These interact with strand a'_C (residues 245–249) and the preceding loop of the second molecule. There are two sulfates involved in this contact, one of which can form hydrogen bonds with both residues 132 and 133 of the symmetry-related molecule, while the other forms a connection with the first type IV contact region, as described above.

Of particular interest are the interactions between the molecules at (x, y, z) and $(x, y, 1 + z)$ since they involve residues of both the N-terminal (47–52, 109 and 113) and C-terminal (278–280) rigid bodies of one molecule interacting with residues of both the N-terminal (144–149 and 317–319) and C-terminal (177 and 178) rigid bodies of the adjacent molecule, as shown in Fig. 8. This interaction therefore spans

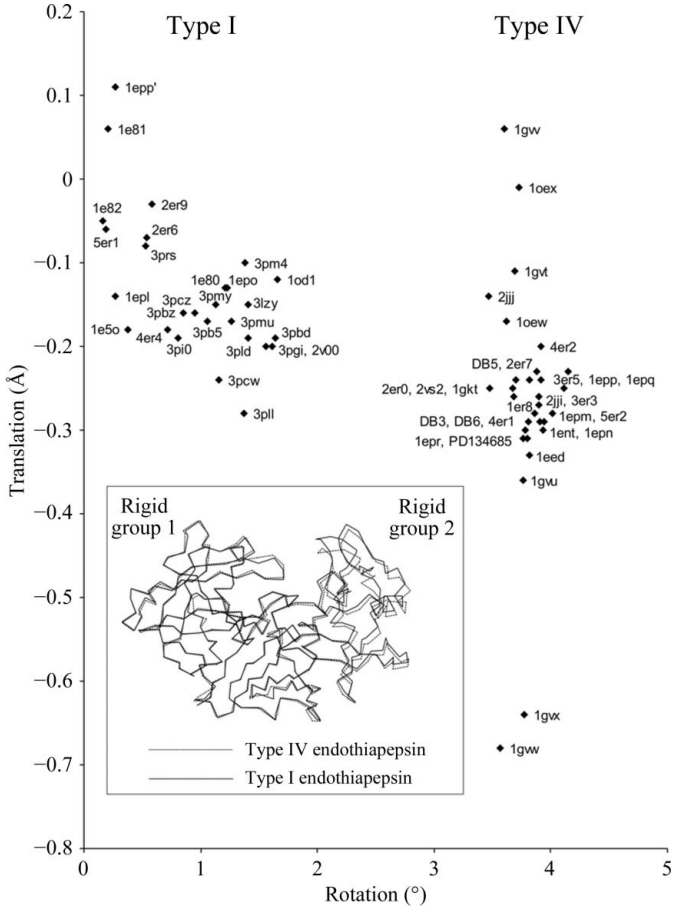

Figure 6
Rigid-group domain movements. The rotations and translations of rigid body 2 (residues 190–302) relative to the type I native endothiapepsin crystal structure (PDB entry 4ape; not shown) are drawn on the graph. The rigid-group parameters for the inhibitor complexes obtained in the type I and type IV crystal forms cluster into two distinct groups (shown as types I and IV). The inhibitor complexes are identified by their PDB codes, with the exception of DB3, DB5, DB6 and PD134685. Structures 2vs2 and 1gkt were solved by neutron diffraction. Note that the structure of the PD130693 complex has been solved in the type IV crystal form (PDB entry 1epp) and also in the type I form (shown as 1epp′). The inset figure indicates the two rigid groups of endothiapepsin and the small but appreciable difference in domain orientation that occurs between the two crystal forms of the native enzyme.

the active-site cleft: a situation which does not occur in type I crystals, where the most similar interaction involves only N-terminal rigid body to C-terminal rigid body contacts, namely the 175–181 loop in one molecule and the 279–281 loop in the other.

3.7. Isotropic displacement parameters

The unscaled B_{iso} values of the type IV crystals of native endothiapepsin are appreciably lower than those of crystals of type I. This may in part be a consequence of the tighter crystal packing and the more extensive interactions as well as the decreased volume of disordered solvent and differences in the data-collection and refinement strategy. Therefore, to facilitate comparison of the type I and type IV crystals the B_{iso} values of the type I structures were scaled empirically to those of the highest resolution type IV uncomplexed structure (DB3). A plot of the differences between residue-averaged temperature factors in the type IV and type I native structures ($B_{ave_{IV},j} - B_{ave,j}$) against residue number j (Fig. 9*a*) shows that the largest decreases in B_{iso} occur where there are crystal contacts, indicating a substantial decrease in thermal or static displacement of these residues.

The h_{N2} α-helix (residues 108–114) shows a large difference in B_{iso} values between type I and type IV native structures (Fig. 9*a*). The occurrence of significantly lower B_{iso} values in this region of inhibitor-complexed endothiapepsin has previously been attributed to inhibitor binding in the P3 pocket (Bailey & Cooper, 1994). However, the existence of low B_{iso} values in type IV crystals without inhibitors bound indicates that this effect may also stem from crystal contacts involving the nearby sulfate anion.

The differences in mean B_{iso} values for residues of the active-site flap in two representative inhibitor complexes that

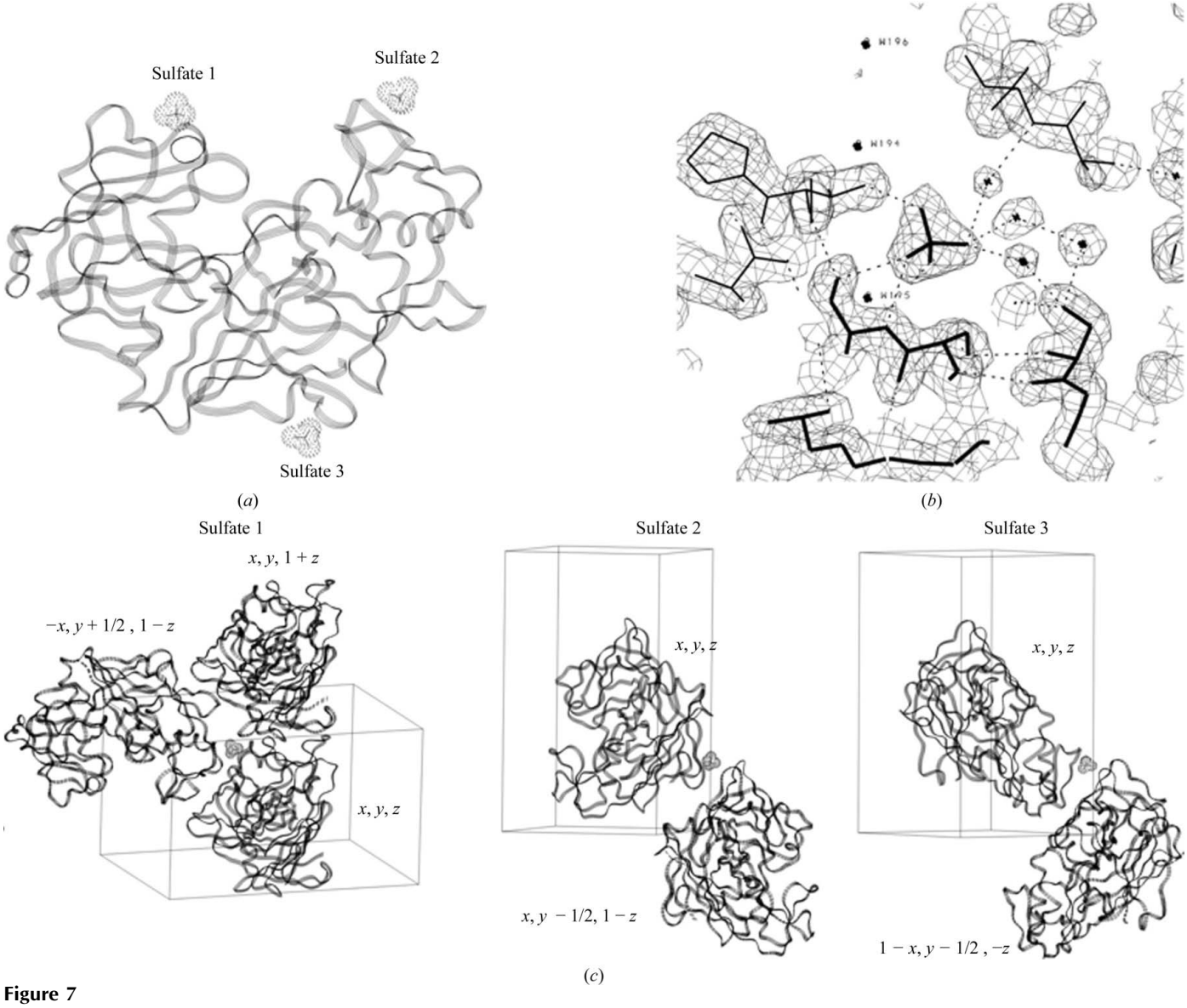

Figure 7
The sulfate groups in type IV endothiapepsin. (*a*) The positions of the three sulfate groups; (*b*) the electron density for the best defined sulfate (sulfate 1) in type IV native endothiapepsin. The contacts that each makes with symmetry-related protein molecules are indicated in (*c*).

form type IV crystals (H-189 and pepstatin A) and the highest resolution type I native structure are shown in Figs. 9(*b*) and 9(*c*), respectively. The B_{iso} values of residues in the active-site flap (74–77) are clearly greater in type IV crystals of native endothiapepsin than in the inhibitor-complexed forms. A similar effect has been noted for the type I crystal form (Bailey & Cooper, 1994). However, reference to Fig. 9(*a*) shows that intermolecular contacts also contribute to the lower disorder in the flap of type IV structures when compared with type I structures, in which the flap is not involved in intermolecular contacts.

3.8. Human pepsin 3b structure and domain movements

In the pre-genomic era, extensive biochemical studies of porcine pepsin, owing to its abundance and ease of preparation, led to it being regarded as the archetypal aspartic proteinase (Fruton, 1976, 2002). Indeed, the catalytic aspartates and other residues of other enzymes in this family are often numbered according to the porcine pepsin scheme. A number of crystal structures are available for pepsin from this organism (Abad-Zapatero *et al.*, 1990; Cooper *et al.*, 1990; Sielecki *et al.*, 1990) as well as of the human enzyme (Fujinaga *et al.*, 1995). Indeed, porcine pepsin was one of the first enzymes to be crystallized and was the first to be analysed by X-ray diffraction (Bernal & Crowfoot, 1934). The crystal form of native human pepsin 3b that we report here is distinct from that obtained with inhibitor complexes and has a lower solvent content (49 *versus* 62%), although it is similar to that reported for uropepsin (Canduri *et al.*, 2001). Human pepsin A has three chromatographically distinct isoforms 1, 3a, 3b and 3c, with pepsin 3b being the major variant. They are encoded by the same gene and therefore have identical amino-acid sequences, but mass-spectrometric analysis suggests that pepsin 3a is phosphorylated (Jones *et al.*, 1995).

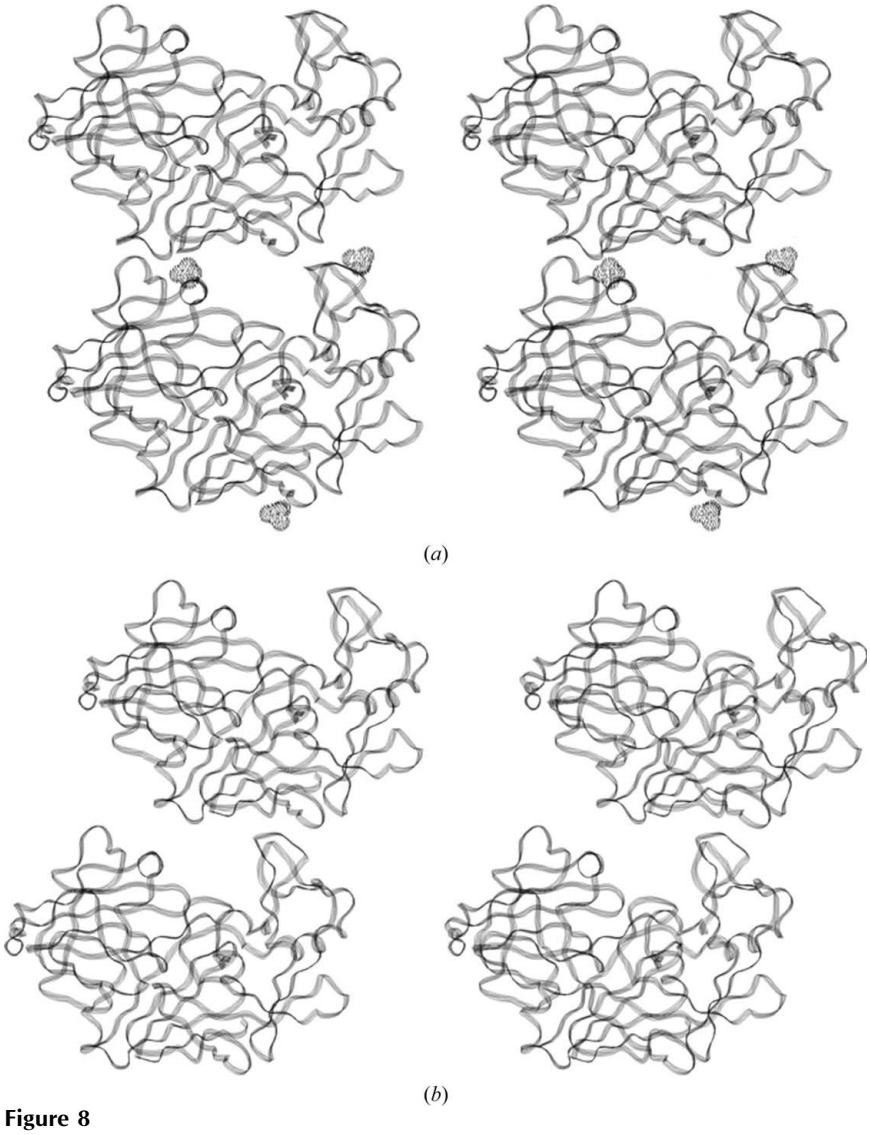

Figure 8
Crystal packing in endothiapepsin crystals. (*a*) Stereoview of the extensive interaction between molecules related by a unit-cell translation along the crystallographic *z* axis in the type IV crystal form. Both rigid bodies of both symmetry-related molecules are involved in these contacts. This contrasts with the more limited nature of contacts in the type I crystal form, as shown in stereo in (*b*).

Superposition of our native human pepsin 3b structure with those of human pepsin 3a complexes (PDB entries 1pso, 1psn and 1qrp) determined by Fujinaga *et al.* (1995) established that essentially the same rigid-body movements that were observed for endothiapepsin appear to occur upon inhibitor binding to the human enzyme. The mean rigid-body rotation for the three complexes is 3.3° (±0.2°) and the mean translation is 0.10 Å (±0.05 Å). There are very few amino acids which change conformation upon binding of the inhibitors, although one notable exception is Glu13, which is very well defined in the electron density for human native pepsin. The side chain of this residue points into the hydrophobic S3 pocket in the native enzyme, whereas binding of an inhibitor causes it to adopt a conformation pointing towards the bulk solvent. Glu13 is highly conserved in pepsins and has been implicated in the pH-dependent activation of the zymogen form of the enzyme (James & Sielecki, 1986; Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992).

4. Conclusions

Protein domain movements are critical for innumerable biological functions such as enzyme catalysis, molecular transport and signal transduction. Such movements represent a major challenge in terms of crystallographic analysis since they occur on a timescale that is much smaller than the duration of a typical data collection. Although there have been remarkable

advances in the speed of data collection using third-generation synchrotron sources, meaningful crystallographic analysis of any dynamic process requires that the movements within the crystal do not disrupt the lattice and are concerted or synchronized during the experiment, both of which are difficult to achieve. The timescale for aspartic proteinase turnover is of the order of 50 ms ($k_{cat} \simeq 20\ s^{-1}$; Dunn *et al.*, 1986), which is within the timescale that can be analysed using special facilities for time-resolved diffraction (Bourgeois *et al.*, 2007). However, synchronizing turnover events within the crystal represents a major hurdle in applying these techniques. The use of synthetic chemistry to prepare analogues which judiciously mimic the transition states of an enzyme-catalysed reaction provides a very convenient tool for harnessing high-energy intermediate states which are, by definition, hard or impossible to isolate otherwise, as well as defining any local or global conformational changes that may take place in the reaction cycle.

It has long been suggested that domain movement plays a significant role in substrate binding and release in the aspartic proteinase family (Šali *et al.*, 1989, 1992). This stems from the observation of appreciable domain movements in the crystal structures upon the binding of inhibitors, although these effects are often correlated with a change in crystal form. The occurrence of either type I or type IV crystals of endothiapepsin does not have any clear experimental determining factors. The initial conditions of crystallization are the same for both types; indeed, the native enzyme as well as some inhibitor complexes were found to crystallize in both forms. Sometimes type I crystals were seen to grow first and to deteriorate over time; type IV crystals then grew from the same mother liquor. The lower solvent content of the type IV crystals suggests that this crystal form is the more stable of the two as the mother liquor becomes more dehydrated. Type IV crystals are characterized by a greater number of lattice contacts and the presence of sulfate anions mediating intermolecular interactions. These observations suggest that the main determinant of the rigid-body shift, rather than being the binding of an inhibitor, may instead be a physical factor such as the ionic strength of the medium and/or the respective crystal lattice contacts. The involvement of the sulfate ions in the lattice may be entirely fortuitous and may depend on local sulfate concentration and other factors during nucleation. However, the interactions that they make appear to be pivotal in the lattice and since they involve both N- and C-terminal rigid bodies of two adjacent molecules they may be responsible for the change in their relative orientation.

All of the inhibitor-complexed forms of endothiapepsin were obtained by cocrystallization rather than soaking since diffusion of inhibitors into native type I crystals was observed to cause deterioration. The first crystallographic evidence that appreciable domain movements occur on inhibitor binding to the enzyme was reported by Šali *et al.* (1989, 1992). In this work, two different crystal forms of endothiapepsin inhibitor complexes were compared with the original type I native structure. It was found that for structures in the type IV form the C-terminal domain was rotated and translated along a screw axis, giving a mean r.m.s. C^{α} deviation of 0.67 Å (±0.02 Å). For structures in the type I form (which has a much larger unit cell) smaller shifts occur, with a mean r.m.s. C^{α} deviation of 0.26 Å (±0.06 Å). In the current work, we have shown that two type IV native endothiapepsin structures have C-terminal domain shifts that are comparable to those that occur when inhibitors were bound in the type IV crystal form. The key to the domain shift may be the presence of three sulfate anions bound to the surface of endothiapepsin in the type IV form. There is further evidence to suggest that crystal-packing forces are at least partly responsible for this domain movement. One inhibitor complex has been solved in both crystal forms (PD130693) and the domain shifts are clearly different in each form (Fig. 6), with only the type IV form having sulfate ions bound. Of note are the two inhibitor complexes (PDB entries 2vs2 and 1gkt) that were solved by neutron diffraction (Coates *et al.*, 2001, 2008); these are observed to lie well within the type IV distribution. Intriguingly, the largest outliers in the type IV distribution (PDB entries 1gvv, 1oex, 1gvx and 1gvw in Fig. 6) are structures that were solved at atomic resolution using data collected from

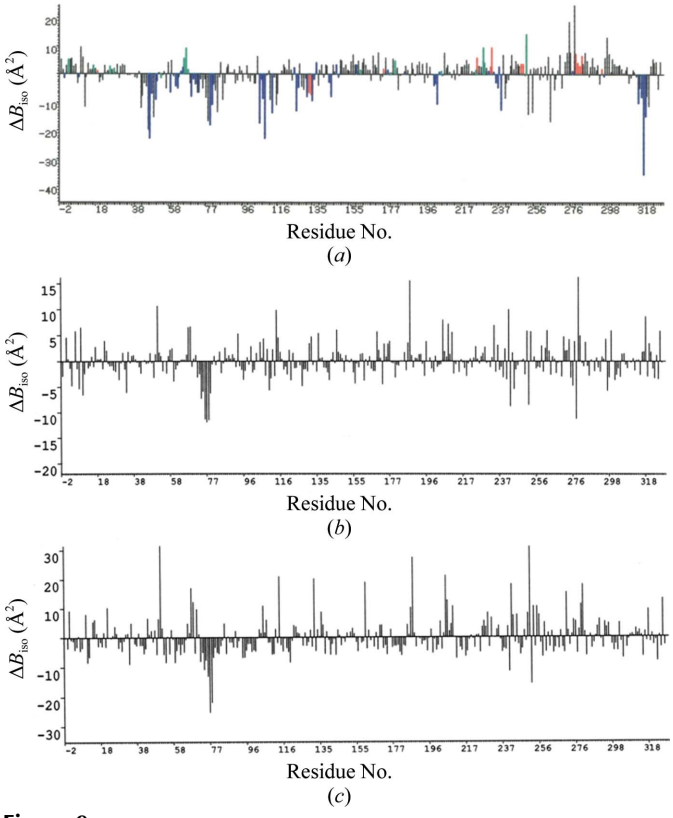

Figure 9
Analysis of mean displacement parameters. (*a*) The difference between the mean B_{iso} of residues in the type IV native structure and those in the type I native crystal form. The vertical bar for each residue is colour-coded according to whether that residue forms lattice contacts in type I crystals only (green), in type IV crystals only (blue), in both type I and type IV crystals (red) or in neither crystal form (black). (*b*) and (*c*) show the differences in mean B_{iso} between the two typical type IV inhibitor complexes (H-189 and pepstatin, respectively) and the type IV native structure (DB3) reported here. Both of these inhibitor complexes and the type IV native structure were solved at the same resolution (1.9 Å).

crystals that were cryocooled to 100 K. Since the domain shift occurring in these structures appears to be significantly affected by crystal freezing, it is likely to be caused primarily by crystal-packing effects rather than the presence or otherwise of an active-site ligand.

Needless to say, the change in domain orientation apparently caused by crystal packing must reflect on a propensity of the fold to flex in the same manner, which may well have catalytic importance. Šali *et al.* (1992) have comprehensively and persuasively reviewed the structural and other evidence that domain movement in aspartic proteinases has a catalytic role. Intriguingly, the assertion that the domain movement in endothiapepsin upon inhibitor binding could arise from lattice contacts alone would appear to be countered by our observation that essentially the same domain movement occurs upon inhibitor binding to human pepsin in two entirely different crystal forms.

We would like to thank Peter Strop and Milan Soucek of the Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, Prague 6, Czech Republic for provision of facilities for and expertise in peptide synthesis. We thank Dr J. K. Cockcroft, UCL Department of Chemistry for much assistance with data retrieval.

References

Abad-Zapatero, C., Rydel, T. J. & Erickson, J. (1990). *Proteins*, **8**, 62–81.

Bailey, D. (1994). PhD thesis, University of London.

Bailey, D. & Cooper, J. B. (1994). *Protein Sci.* **3**, 2129–2143.

Bernal, J. D. & Crowfoot, D. (1934). *Nature (London)*, **133**, 794–795.

Blundell, T. L., Jenkins, J. A., Sewell, B. T., Pearl, L. H., Cooper, J. B., Tickle, I. J., Veerapandian, B. & Wood, S. P. (1990). *J. Mol. Biol.* **211**, 919–941.

Bourgeois, D., Schotte, F., Brunori, M. & Vallone, B. (2007). *Photochem. Photobiol. Sci.* **6**, 1047–1056.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Canduri, F., Teodoro, L. G. V. L., Fadel, V., Lorenzi, C. C. B., Hial, V., Gomes, R. A. S., Neto, J. R. & de Azevedo, W. F. (2001). *Acta Cryst.* D**57**, 1560–1570.

Cruickshank, D. W. J. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 11–22. Warrington: Daresbury Laboratory.

Coates, L., Erskine, P. T., Crump, M. P., Wood, S. P. & Cooper, J. B. (2002). *J. Mol. Biol.* **318**, 1405–1415.

Coates, L., Erskine, P. T., Mall, S., Gill, R., Wood, S. P., Myles, D. A. A. & Cooper, J. B. (2006). *Eur. Biophys. J.* **35**, 559–566.

Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. A. & Cooper, J. B. (2001). *Biochemistry*, **40**, 13149–13157.

Coates, L., Tuan, H.-F., Tomanicek, S., Kovalevsky, A., Mustyakimov, M., Erskine, P. T. & Cooper, J. B. (2008). *J. Am. Chem. Soc.* **130**, 7235–7237.

Cooper, J. B. (2002). *Curr. Drug Targets*, **3**, 155–174.

Cooper, J. B. (2010). *Methods Princ. Med. Chem.* **45**, 71–105.

Cooper, J. B., Khan, K., Taylor, G., Tickle, I. J. & Blundell, T. L. (1990). *J. Mol. Biol.* **214**, 199–222.

Davies, D. (2000). *Annu. Rev. Biophys. Biophys. Chem.* **19**, 189–215.

Driessen, H. P. C., Bax, B., Slingsby, C., Lindley, P. F., Mahadevan, D., Moss, D. S. & Tickle, I. J. (1991). *Acta Cryst.* B**47**, 987–997.

Drohse, H. B. & Foltmann, B. (1989). *Biochim. Biophys. Acta*, **995**, 221–224.

Dunn, B. M. (2002). *Chem. Rev.* **102**, 4431–4458.

Dunn, B. M., Jimenez, M., Parten, B. F., Valler, M. J., Rolph, C. E. & Kay, J. (1986). *Biochem. J.* **237**, 899–906.

Erskine, P. T., Coates, L., Mall, S., Gill, R. S., Wood, S. P., Myles, D. A. A. & Cooper, J. B. (2003). *Protein Sci.* **12**, 1741–1749.

Foundling, S. I. *et al.* (1987). *Nature (London)*, **327**, 349–352.

Frazão, C., Bento, I., Costa, J., Soares, C. M., Vérissimo, P., Faro, C., Pires, E., Cooper, J. B. & Carrondo, M. A. (1999). *J. Biol. Chem.* **274**, 27694–27701.

Fruton, J. S. (1976). *Adv. Enzymol.* **44**, 1–36.

Fruton, J. S. (2002). *Q. Rev. Biol.* **77**, 127–147.

Fujinaga, M., Chernaia, M., Tarasova, N., Mosimann, S. & James, M. N. G. (1995). *Protein Sci.* **4**, 960–972.

Geschwindner, S., Olsson, L.-L., Albert, J. S., Deinum, J., Edwards, P. D., deBeer, T. & Folmer, R. H. A. (2007). *J. Med. Chem.* **50**, 5903–5911.

Haneef, I., Moss, D. S., Stanford, M. J. & Borkakoti, N. (1985). *Acta Cryst.* A**41**, 426–433.

Hartsuck, J. A., Koelsch, G. & Remington, S. J. (1992). *Proteins*, **13**, 1–15.

James, M. N. G. & Sielecki, A. R. (1986). *Nature (London)*, **319**, 33–38.

Janin, J., Wodak, S., Levitt, M. & Maigret, B. (1978). *J. Mol. Biol.* **125**, 357–386.

Jones, A. T., Green, B. N., Wood, S. P. & Roberts, N. B. (1995). *Adv. Exp. Med. Biol.* **362**, 83–89.

Jones, A. T., Keen, J. N. & Roberts, N. B. (1993). *J. Chromatogr.* **646**, 207–212.

Koster, H., Craan, T., Brass, S., Herhaus, C., Zentgraf, M., Neumann, L., Heine, A. & Klebe, G. (2011). *J. Med. Chem.* **54**, 7784–7796.

Kumar, A., Grover, S., Sharma, J. & Batish, V. K. (2010). *Crit. Rev. Biotechnol.* **30**, 243–258.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lapatto, R., Blundell, T. L., Hemmings, A., Overington, J., Wilderspin, A., Wood, S. P., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Hawrylik, S. J., Lee, S. E., Scheld, K. G. & Hobart, P. M. (1989). *Nature (London)*, **342**, 299–302.

Moews, P. C. & Bunn, C. W. (1970). *J. Mol. Biol.* **54**, 395–397.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pearl, L. H. & Blundell, T. L. (1984). *FEBS Lett.* **174**, 96–101.

Poornam, G., Matsumoto, A., Ishida, H. & Hayward, S. (2009). *Proteins*, **76**, 201–221.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Šali, A., Veerapandian, B., Cooper, J. B., Foundling, S. I., Hoover, D. J. & Blundell, T. L. (1989). *EMBO J.* **8**, 2179–2188.

Šali, A., Veerapandian, B., Cooper, J. B., Moss, D. S., Hofmann, T. & Blundell, T. L. (1992). *Proteins*, **12**, 158–170.

Sardinas, J. L. (1968). *Appl. Microbiol.* **16**, 248–255.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S. & James, M. N. G. (1990). *J. Mol. Biol.* **214**, 143–170.

Sielecki, A. R., Fujinaga, M., Read, R. J. & James, M. N. G. (1991). *J. Mol. Biol.* **219**, 671–692.

Williams, D. C., Whitaker, J. R. & Caldwell, P. V. (1972). *Arch. Biochem. Biophys.* **149**, 52–61.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. (1989). *Science*, **245**, 616–621.