this study. We thank Dr E. Adman for the 7RXN coordinates of the structure of *Desulfovibrio vulgaris* rubredoxin refined at 1.5 Å resolution prior to their release from the Brookhaven Protein Data Bank.

References

ADMAN, E. T., SIEKER, L. C. & JENSEN, L. H. (1991). *J. Mol. Biol.* **217**, 337–352.
ADMAN, E. T., SIEKER, L. C., JENSEN, L. H., BRUSCHI, M. & LE GALL, J. (1977). *J. Mol. Biol.* **112**, 113–120.
AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
BAKER, E. N. & DODSON, E. J. (1980). *Acta Cryst.* A**36**, 559–572.
BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MAYER, E. F. JR, BRYCE, M. D., RODGERS, J. R., KENNARD, O., SIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.
CHAKRABARTI, P. (1989). *Biochemistry*, **28**, 6081–6085.
DAUTER, Z., TERRY, H., WITZEL, H. & WILSON, K. S. (1990). *Acta Cryst.* B**46**, 833–841.
FREY, M., SIEKER, L. C., PAYAN, F., HASER, R., BRUSCHI, M., PEPE, G. & LE GALL, J. (1987). *J. Mol. Biol.* **197**, 525–541.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.
KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.
LESLIE, A. G. W., BRICK, P. & WONACOTT, A. J. (1986). *CCP4 Newsl.* **18**, 33–39.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
PAULING, L. (1960). *The Nature of the Chemical Bond*, 3rd ed. New York: Cornell Univ. Press.
PIERROT, M., HASER, R., FREY, M., BRUSCHI, M., LE GALL, J., SIEKER, L. C. & JENSEN, L. H. (1976). *J. Mol. Biol.* **107**, 179–182.
RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.
READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
SERC Daresbury Laboratory (1979). *CCP4. A Suite of Programs for Protein Crystallography*. Daresbury Laboratory, Warrington, England.
STENKAMP, R. E., SIEKER, L. C. & JENSEN, L. H. (1990). *Proteins Struct. Funct. Genet.* **8**, 252–264.
VOORDOUW, G. (1988). *Gene*, **67**, 75–83.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1972). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 359–367.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.
WATENPAUGH, K. D., SIEKER, L. C. & JENSEN, L. H. (1979). *J. Mol. Biol.* **131**, 509–522.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
WILSON, K. S. (1989). *Synchrotron Radiation in Structural Biology*, edited by R. M. SWEET & A. D. WOODHEAD, pp. 47–54. New York, London: Plenum Press.

Acta Cryst. (1992). B**48**, 59–67

Structure of Monoclinic Papain at 1·60 Å Resolution

BY R. W. PICKERSGILL,* G. W. HARRIS AND E. GARMAN†

Protein Engineering Department, Protein Crystallography and Biomolecular Computing Section, AFRC Institute of Food Research, Reading Laboratory, Shinfield, Reading RG2 9AT, England

(*Received* 26 *July* 1990; *accepted* 3 *June* 1991)

Abstract

X-ray diffraction data to 1·60 Å resolution have been collected from monoclinic crystals of papain. The monoclinic model was derived from the orthorhombic one by molecular replacement, X-ray restrained molecular-dynamics simulation and least-squares refinement. Refinement against 1·60 Å data produced a model with reasonable stereochemistry and an *R* factor of 16·0%. The X-ray structures of orthorhombic and monoclinic papain are compared. The two structures are similar, the r.m.s. deviation between the two structures is 1·049 Å (mean difference 0·531 Å). The monoclinic model is shown to shift considerably (r.m.s. 3·083 Å) during refinement which indicates that bias due to the starting model may reasonably be expected to be low; in addition, solvent structure is independently determined. Tightly bound solvent occupies the same position in both structures and weakly bound solvent structures (high temperature factors) are different. Differences in protein structure are attributable to different crystal contacts, different covalent modification of the active-site cysteine, or different interpretation of weak density. The temperature factors for both structures show similar trends.

Introduction

Papain is one of the cysteine proteases found in the latex of *Carica papaya*. The single polypeptide chain of 212 amino-acid residues is cross linked by three disulfide bridges. The monoclinic crystals reported here diffract to high angle and the structure of the orthorhombic form has been determined by multiple isomorphous replacement (Drenth, Jansonius, Koekoek, Swen & Wolthers, 1968). The orthorhombic

* To whom correspondence should be addressed.
† Laboratory of Molecular Biophysics, Rex Richards Building, University of Oxford, Oxford OX1 3QU, England.

0108-7681/92/010059-09$03.00

structure has subsequently been refined at 1·65 Å resolution (Kamphuis, Kalk, Swarte & Drenth, 1984). A model of the enzyme in monoclinic crystals will be useful in determining the accuracy to which the protein and solvent structure are determined and how the different packing arrangements affect the detailed conformation and dynamics of the protein. In addition, differences in the oxidation state of the active-site groups and the different pH values at which orthorhombic crystals (pH 9·3) and monoclinic crystals (pH 5·0) are grown may affect conformation.

Experimental

Monoclinic crystals of papain were grown from a 1·3% papain solution containing 25 m*M* NaCl and 25% (*v*/*v*) methanol buffered with 0·1 *M* sodium acetate at pH 5·0. This solution was equilibrated with a reservoir containing 67% (*v*/*v*) methanol. Crystals of up to 1 mm in length were grown. Precession photographs showed the crystals to be monoclinic with $a = 65{\cdot}7\ (1)$, $b = 50{\cdot}7\ (1)$, $c = 31{\cdot}5\ (1)$ Å and $\beta = 98{\cdot}4\ (1)^\circ$. Reflections were only present for $k = 2n$ along the b^* axis which is consistent with the space group $P2_1$. There is one papain molecule in the asymmetric unit.

Intensity data were collected using a Xentronics area detector and Cu $K\alpha$ radiation from a Rigaku Ru200H rotating-anode generator. The generator was operated at 50 kV and 50 mA and the focal size was 0·3 × 3·0 mm, Cu $K\alpha$ radiation was selected by a graphite monochromator. Data were collected from a single crystal to a resolution limit of 1·60 Å using a three-axis camera, crystal-to-detector distance of 8·5 cm and the detector at a 2θ angle of $-27{\cdot}0^\circ$. Data were recorded on a total of 1143 frames, each frame corresponding to an oscillation of 0·25° about ω. The data were collected in two parts; part 1 with $\varphi = 0{\cdot}0^\circ$ consisted of 943 frames and part 2 with $\varphi = 90{\cdot}0^\circ$ of 200 frames. All data were collected at 288 K. The data were scaled together in ω batches and the merging R factor (F) for the data to 1·60 Å was 5·07%; the data were 93% complete to 1·70 Å and 82% complete to 1·60 Å (Table 1).

Table 1. *Internal agreement of data and percentage recorded as a function of resolution*

21 719 reflections collected: 34% measured once, 49% measured twice, 15% measured three times. $R_{sym}(I) = \sum_n[\sum_{i=1}^{m}(I_i - \bar{I})/\sum_{i=1}^{m}\bar{I}]$, where there are m equivalents with average intensity $\bar{I}$ and n sets of equivalent reflections. $R_{sym}(F)$ is defined in a similar way with respect to F.

Lower-resolution shell limit (Å)	$R_{sym}(I)$	$R_{sym}(F)$	% complete
2·94	2·70	1·90	98
2·33	6·30	4·56	98
2·04	8·54	6·39	97
1·85	13·43	10·01	90
1·72	21·80	16·17	75
1·60	32·03	23·17	36
Total	5·14	5·07	82

Table 2. *Progress of the X-ray restrained molecular-dynamics refinement*

$R = \sum(|F_o| - F_c|)/\sum F_o$. The initial R factor for the molecular replacement solution was 40·2%. The temperature was 300 K throughout.

Time (ps)	Resolution (Å)	R (%)
2·0	10·0 5·0	33·9
0·5	10·0 4·0	27·6
0·5	10·0 3·0	35·0
0·5	10·0–2·0	34·5

Molecular replacement

The start coordinates were those of the orthorhombic structure (Kamphuis *et al.*, 1984) and the program *MERLOT* was used for all molecular replacement calculations (Fitzgerald, 1988). The monoclinic data and orthorhombic coordinates rotated through $\beta = 90^\circ$ were used in a Crowther fast rotation function (Crowther, 1972). The origin was removed, a cut-off radius of 24 Å, and 10 Å to 4 Å data were used in the Patterson expansion. The search was over 180° in α (2·5° steps) and β (5·0° steps) and 360° in γ (5·0° steps). The solution $\alpha = 75$, $\beta = 95$, $\gamma = 265^\circ$ was 15·97 r.m.s. above background. This solution was refined using the Lattman rotation function (Lattman & Love, 1972) to $\alpha = 75$, $\beta = 96$ and $\gamma = 266^\circ$. This solution was used in a Crowther–Blow translation function (Crowther & Blow, 1967). The step size was 0·02 of the cell edge and the solution $fa = 0{\cdot}27$, $fc = 0{\cdot}15$ was at 7·37 r.m.s. above background.

Refinement

The molecular-replacement solution was used in X-ray restrained molecular-dynamics simulation (Van Gunsteren, Berendsen & Fujinaga, 1988). This was with the aim of reducing bias due to the starting structure. The temperature was maintained at 300 K and the quantity of data increased as the simulation proceeded. The R factor was reduced from 40·2% for 10 to 5 Å data to 33·9% for these data after 2·0 ps of simulation. A further 1·5 ps of simulation reduced the R factor to 34·5% for 10 to 2 Å data (Table 2).

At this stage a $2F_o - F_c$ map was calculated using data to 2·0 Å and displayed on an Evans and Sutherland PS390 graphics system using *FRODO* (Jones, 1985). The map revealed 107 water molecules and several side chains occupying solvent density. In addition, seven methanols were tentatively assigned, and one atom, initially assumed to be an

Table 3. *Progress of the restrained least-squares refinement of monoclinic papain*

Rebuilds 1, 2, 3 and 4 have 107, 150, 210 and 226 water molecules and 7, 7, 7 and 1 methanol molecules, respectively.

No. of cycles	Resolution (Å) (No. of reflections)	R (%)	Comment
2	2·0–10·0 (13 217)	27·02	After rebuild 1, O atom added to Sγ of Cys 25. Refine overall MSDA (1652 protein atoms, 115 O atoms)
3	2·0–10·0 (13 217)	21·80	Refine isotropic MSDAs
5	1·6–10·0 (20 172)	19·72	Extend resolution
10	1·6–10·0 (20 172)	17·80	After rebuild 2, refine second site occupancies for Ile 148, Asn 155 (1659 protein atoms, 166 O atoms)
34	1·6–10·0 (20 172)	16·59	After rebuild 3, refine second-site occupancies Ile 148, Asn 155, Glu 112 and Asn 169* (1667 protein atoms, 225 O atoms)
25	1·6–10·0 (20 172)	15·96	After rebuild 4, residue changes: extra atoms added for Lys 139, Gln 47, 118, 135⇒Glu 47, 118, 135. Only second site for Asn 169† (1659 protein atoms, 250 O atoms)

* Refined occupancies for Asn 169: site 1, 0·616; site 2, 0·384.
† Refined occupancies for Asn 169: site 1, 0·474; site 2, 0·526.

oxygen, added to the Sγ of Cys 25. The course of the restrained least-squares refinement using *RESTRAIN* (Driessen, Haneef, Harris, Howlin, Khan & Moss, 1989) is summarized in Table 3. *RESTRAIN* is a least-squares refinement program with pseudo-energy restraints. *RESTRAIN* uses an approximation to a normal matrix where contributions to off-diagonal terms are included for the energy restraints and 3×3 blocks are used for the contribution from the positional parameters of the atoms. All other off-diagonal terms are taken as zero.

The Levenberg–Marquardt method is used to remove ill-conditioning and the normal equations are solved using the Gauss–Seidel method (Haneef, Moss, Stanford & Borkakoti, 1985). The rebuilt model was first regularized over 15 cycles followed by five cycles of structure-factor restrained least-squares refinement with a high-resolution cut-off of 2·0 Å. The high-resolution cut-off was then extended to 1·60 Å over a further five cycles. For the first two cycles, an overall MSDA (mean-square displacement amplitude or U value) was refined; for all subsequent cycles, isotropic MSDAs were refined for all atoms. Equal weights ($w = 0{\cdot}02$) were applied to all reflections throughout. A second $2F_o - F_c$ map was calculated which revealed both incorrectly positioned side chains and solvent molecules, and additional water and methanol molecules to be included. The map showed evidence of second sites for Ile 148 and Asn 155. Following this rebuild, eight cycles of geometric

Table 4. *Details of the least-squares refinement of monoclinic papain*

(*a*) Least-squares refinement

No. of reflections	20 172
High-resolution cut off (Å)	1·60
Low-resolution cut off (Å)	10·0
No. of protein atoms	1659
No. of bound atoms (O)	1
No. of water molecules currently included	226
No. of methanol molecules currently included	1
Residual[i]	0·1596
Weighted residual[ii]	0·1678
Correlation coefficient[iii]	0·9621
No. of positional parameters	5727
No. of thermal parameters	1909
No. of overall parameters	2
No. of occupancies refined	1
Total No. of parameters refined	7639

(*b*) Crystallographic R factor as a function of resolution for the final structure

Resolution (Å)	R factor	No. of structure amplitudes
10·00–2·89	0·1236	4443
2·89 2·30	0·1454	4351
2·30 2·01	0·1648	4187
2·01–1·83	0·2062	3618
1·83 1·70	0·2892	2724
1·70–1·60	0·4156	849
Total	0·1596	20 172

Notes: (i) $R = \sum||F_o| - G|F_c||/\sum|F_o|$; (ii) $wR = [\sum W(|F_o| - G F_c|)^2/\sum W F_o^2]^{1/2}$ where $W = 0{\cdot}02$; (iii) $C = n\sum(|F_o|G F_c|) - \sum F_o \sum G F_c|/\{[n\sum F_o|^2 - (\sum F_o)^2][n\sum(G|F_c|)^2 - (\sum G|F_c|)^2]\}^{1/2}$, where n is the number of amplitudes used.

Table 5. *R.m.s. deviation of papain structures from the final structure*

Structure	R.m.s. deviation (Å)
Molecular-replacement solution	3·083
X-ray restrained molecular-dynamics refined	2·137
Least-squares refinement (before last cycle)	0·005
Orthorhombic	1·049

regularization followed by ten cycles of structure-factor least-squares refinement with a high-resolution cut-off of 1·60 Å reduced the R factor to 17·80%. Second sites were included for Ile 148 and Asn 155 and the occupancies refined. A $3F_o - 2F_c$ map was then calculated and examined on the graphics system at a contour level of 0·52 e Å^{-3}. At this stage, most of the residues fitted the density reasonably well, apart from mobile side chains such as Lys 10, Arg 98, Glu 99, for which there was little density. Side chains were adjusted where necessary and more water molecules added. The map showed evidence for second sites for Gln 112 and Asn 169, which were included in the next stage of the refinement. Five cycles of geometric regularization followed by a further 34 cycles of least-squares structure-factor refinement reduced the R factor to 16·59%. A further $3F_o - 2F_c$ map was calculated and examined on the

Table 6. *Statistical information on the geometry of the final structure for monoclinic papain*

The values quoted by Kamphius *et al.* (1984) for orthorhombic papain are very similar

	Number	R.m.s. deviation (Å)
Distances d (Å)		
$d < 2{\cdot}12$	1648	0·022
$2{\cdot}12 < d < 2{\cdot}62$	2210	0·049
$d > 2{\cdot}62$	25	0·069
Planes		
Type 1 (peptide)	212	0·020
Type 2 (non-peptide)	82	0·010
Chiral centres*	120	0·028

* Chiral restraints are applied as distance restraints along the edges of chiral tetrahedra with target distance ($d_t \leq 2{\cdot}12$ Å).

Table 7. *Comparison of monoclinic and orthorhombic papain structures*

The matrix to superpose the orthorhombic structure on the monoclinic structure (orthogonal coordinates in Å) is:

Rotation =	−0·39403	0·90996	0·12929
	0·08157	−0·10549	0·99107
	0·91547	0·40106	−0·03266
Translation =	−36·18653	−8·17597	−51·38811

(*a*) Superposition of domains

	R.m.s. (mean) (Å)	
	Main chain only	All protein atoms
Whole molecule	0·287 (0·246)	1·049 (0·531)
First domain	0·248 (0·213)	1·245 (0·588)
Second domain	0·212 (0·176)	0·738 (0·397)

	No. of atoms	
	Main chain only	All protein atoms
Whole molecule	848	1652*
L domain	428	844
R domain	420	808

(*b*) Superposition of secondary structural elements, main-chain atoms only

	R.m.s. (mean) (Å)	No. of atoms
All atoms	0·287 (0·246)	848
α and β only	0·220 (0·198)	392
α only	0·220 (0·194)	224
β only	0·146 (0·130)	168

* Excluding Lys 139, Cδ, Cε, Nζ.

graphics system. Of the residues with second-site positions, only Asn 169 had convincing electron density for two distinct sites. For Ile 148, the 'second site' was the preferred site (occupancy of 0·616) and was retained.

At this stage all atoms of Lys 139 were included in the refinement and in accordance with DNA sequence data (Cohen, Coghlan & Dihel, 1986) Gln 47, 118 and 135 changed to Glu residues. A further 25 water molecules were added. Further refinement reduced the R factor to 15·96%. The refined occupancy for Asn 169 was site 1: 0·474; site 2: 0·526. Details of the least-squares refinement and final crystallographic R factor as a function of resolution are given in Tables 3, 4(*a*) and 4(*b*). A final $3F_o - 2F_c$ map was calculated to check the quality of the refinement. The protein structure is well defined in this final map except for a few side chains and one main-chain atom. The solvent O atoms were critically reviewed on the graphics and those with poor electron density and high MSDA were deleted from the model (17 water molecules and one methanol molecule were deleted, a further two methanols were replaced by water molecules). The progress of the structure determination is presented in terms of the deviation from the final structure in Table 5. Coordinates of this refined structure have been deposited with the Protein Data Bank.*

Table 8. *Average MSDAs for monoclinic and orthorhombic papain*

	Monoclinic		Orthorhombic	
	No. of atoms	Average MSDA (Å^2)	No. of atoms	Average MSDA (Å^2)
Main-chain atoms	848	0·15	848	0·16
Side-chain atoms	807	0·24	804*	0·25
All protein atoms	1655	0·19	1652*	0·20
Water O atoms	226	0·48	195	0·52

* Lys 139, Cδ, Cε, Nζ not present in structure.

Table 9. *Number of main-chain hydrogen bonds for monoclinic and orthorhombic structures*

	Monoclinic papain	Orthorhombic papain
Total No. of H bonds	113	117
No. of α-helix H bonds	39	38
No. of β-sheet H bonds	40	43
No. of β-turn H bonds	34	36

Structure of papain in monoclinic crystals

Quality of the structure. The accuracy and quality of the monoclinic and orthorhombic structures are comparable as judged from the resolution limits, R factor and statistics of the geometry of the final structures (Table 6). The R factor and resolution limit are 16·0% at 1·60 Å for the monoclinic structure. The average error in both structures as estimated from a Luzatti plot is in the range 0·15–0·30 Å (at high resolution the error is overestimated due to errors in intensity measurement). The error will be lowest for well ordered and relatively rigid atoms and highest for poorly ordered or flexible groups (high MSDAs). The overall structure of the enzyme determined from the monoclinic crystal is similar to

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory, and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37048 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 10. *Main-chain hydrogen-bond geometry for monoclinic papain*

Mean values (angles in °, bond lengths in Å).

	C—O—H	O—H—N	C—O—H—N	O—H—N—C*A*	C*A*—C—O—H	O—H
All H bonds*	135·36	149·03	−8·03	−5·29	−50·20	2·08
α-Helix H bonds	145·92	157·13	−11·08	−34·38	115·75	2·09
β-Sheet H bonds	148·29	157·48	−9·83	25·92	−6·74	1·98

* β-turns excluded.

that from orthorhombic crystals as described by Kamphius *et al.* (1984). The r.m.s. deviation of main-chain atoms is 0·287 Å after superposition using *EULER* (Moss, Howlin & Haneef, 1985). Details of the superposition are given in Table 7 and a comparison of the MSDA values in Table 8. After superposition the mean-square deviation is (0·287 × 0·287) 0·0824 Å^2. If we assume the main-chain structures are identical and the error is equally distributed between the two structures then the estimated error is ($\sqrt{}$0·0824/2) 0·144 Å. This is an upper limit to the error in each structure as some differences will be due to real conformational differences. When the first (1 to 111) and second (112 to 212) domains of the orthorhombic and monoclinic structures are compared separately the second domains are most similar (Table 7*a*). This is primarily due to differences in the region 43 to 106 which contains little secondary structure. Also, the second domain comprises predominantly β-sheet, while the first domain has more α-helical regions. Comparing secondary structural elements for the orthorhombic and monoclinic papain structures, the β-sheet regions are more similar than the α-helical region (Table 7*b*) as the β-strands are less free to move relative to one another than are helices. Details of the overall number or hydrogen bonds in the structure and their geometry is given in Tables 9 and 10. H atoms bonded to the peptide nitrogens were placed on the bisector of the angle C—N—Cα, at a distance of 1·02 Å from the N atom. Possible hydrogen bonds were identified using a cut-off distance of O···H less than 2·4 Å [for details of calculation, see Harris, Borkakoti, Moss, Palmer & Howlin (1987)].

The cDNA sequence of papain (Cohen *et al.*, 1986) has glutamic acids at positions 47, 118 and 135. The residues are isosteric with the previous sequence containing glutamine residues. The changes may have some importance in our understanding of the stability and action of the enzyme. Glu 47 is now part of the inter-domain salt bridges in papain and a possible role of Glu 135 has been discussed previously (Pickersgill, Sumner, Collins & Goodenough, 1989).

Flexibility and disorder. The main-chain conformation is unambiguously defined throughout the final $3F_o - 2F_c$ map except for the carbonyl oxygen of Ser 21 which is not seen. This oxygen is probably highly mobile. At 0·43 e Å^{-3} the density for Arg 98, Arg 111, Lys 139, Arg 145 and Ser 196 is discontinuous or absent, and the ends of Asn 84, Gln 142, Lys 156 and Asn 212 are not seen.

The atoms which are not seen have MSDAs greater than 0·60 Å^2. At the higher contour level examined, 0·65 e Å^{-3}, the ring density was poor for Tyr 61, 78, 94, 123 and 197 and for the hydroxyl of Tyr 67. The atoms of the rings with poor density have MSDAs greater than 0·57 Å^2. It is remarkable that the disordered side chains in the monoclinic structure correspond almost exactly to those disordered in the orthorhombic structure. The carbonyl of Ser 21 was poorly defined in the orthorhombic structure, whereas the carbonyl oxygen was not seen in our map. However, Glu 9 is well defined in the monoclinic structure but disordered in the orthorhombic one. The MSDAs are higher for the orthorhombic structure than the monoclinic structure despite the similar resolution of the data used in refinement. In this refinement there was no coupling between neighbouring atoms as employed in refinement of the orthorhombic structure. Hence no absolute comparison of the observed temperature factors is possible. However, apart from minor differences the variations in MSDAs for main-chain atoms are remarkably similar for the two structures despite the different crystal packing (Fig. 1), and

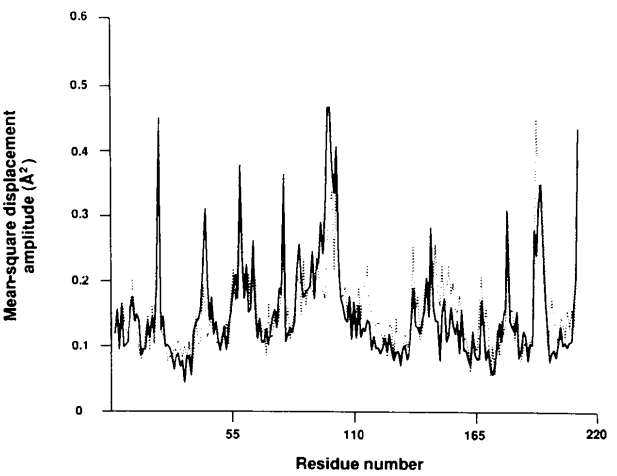

Fig. 1. Plot of the mean-square displacement amplitude (MSDA) of main-chain atoms for each residue. The MSDA is averaged over the main-chain atoms in each residue. MSDAs for the monoclinic structure are in full line and for the orthorhombic structure in broken line.

consequently represent true internal molecular motion rather than an artifact of the crystal lattice or experimental error.

Crystal packing and conformation. A list of intermolecular contacts for the monoclinic structure is given in Table 11. Contacts between residues 1 and 150 and between 140 and 191 are found in both structures. The different packing arrangements lead to a small number of differences in side-chain conformation. Differences in χ_1 and χ_2 values are plotted for each residue in Fig. 2. The differences due to different packing include Tyr 94, Arg 96 (Fig. 3) and Asn 64. The region around Tyr 94 is shown, together with symmetry-related molecules, in Fig. 4(*a*) and 4(*b*) for the monoclinic and orthorhombic structures respectively. The α-carbon of residue 94 moves by about 1·0 Å in response to the change in packing. Other main-chain atoms that differ by more than 1·0 Å are the carbonyl oxygens of residues 9, 21 and 180. The carbonyl oxygen 9 is involved in a crystal contact (at 3·28 Å) to N*E*2 of symmetry-related Gln 114. This crystal contact is not present in the orthorhombic papain and arises from the difference in

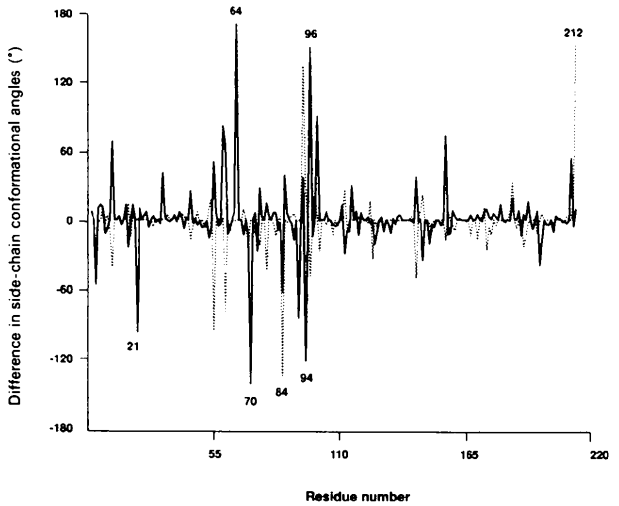

Fig. 2. Difference in the side-chain conformational angles χ_1 and χ_2 for each residue. The difference is that between the angle in the monoclinic and that in the orthorhombic structure. χ_1 difference is in full line, χ_2 in the broken line. Residues with differences greater than 90° are numbered.

Table 11. *Intermolecular contacts in monoclinic crystals*

All contacts between protein atoms in adjacent molecules separated by less than 3·2 Å are listed. The contacts in orthorhombic papain can be found in Table 7 of Kamphius *et al.* (1984).

Atom 1		Atom 2		*d* (Å)	Sym*	*U*1 (Å²)	*U*2 (Å²)
1	N Ile	150	O Val	2·68	1	0·087	0·189
42	O Thr	93	NH2 Arg	3·14	2	0·223	0·385
44	C*G* Asn	93	NH2 Arg	2·84	2	0·480	0·385
44	O*D*1 Asn	93	NH2 Arg	2·42	2	0·700	0·385
44	N*D*2 Asn	93	N*E* Arg	2·89	2	0·434	0·782
44	N*D*2 Asn	93	C*Z* Arg	2·53	2	0·434	0·474
44	N*D*2 Asn	93	NH1 Arg	3·00	2	0·434	0·268
44	N*D*2 Asn	93	NH2 Arg	2·71	2	0·434	0·385
47	O*E*1 Glu	59	N*E* Arg	3·02	3	0·345	0·388
47	O*E*1 Glu	59	NH2 Arg	3·16	3	0·345	0·293
47	O*E*2 Glu	59	NH2 Arg	3·14	3	0·345	0·293
58	NH1 Arg	108	O*D*1 Asp	3·16	4	0·477	0·193
64	O*D*1 Asn	78	O Tyr	3·07	4	0·756	0·330
77	O*E*1 Gln	92	O Gln	3·00	5	0·269	0·222
77	N*E*2 Gln	92	N Gln	2·88	5	0·162	0·136
92	O*E*1 Gln	106	N*Z* Lys	2·88	4	0·210	0·151
140	O*D*1 Asp	191	NH1 Arg	2·91	6	0·191	0·098
140	O*D*2 Asp	191	NH2 Arg	2·74	6	0·201	0·140
156	C*E* Lys	184	N*D*2 Asn	3·09	7	0·734	0·212

* Symmetry operations: (1) second atom generated by $-x$, $y-\frac{1}{2}$, $-z$; (2) second atom generated by $-x+1$, $y-\frac{1}{2}$, $-z+1$; (3) second atom generated by x, y, $z+1$; (4) second atom generated by $-x+1$, $y+\frac{1}{2}$, $-z$; (5) second atom generated by $-x+1$, $y-\frac{1}{2}$, $-z$; (6) second atom generated by $-x$, $y+\frac{1}{2}$, $-z$; (7) second atom generated by $-x$, $y+\frac{1}{2}$, $-z-1$.

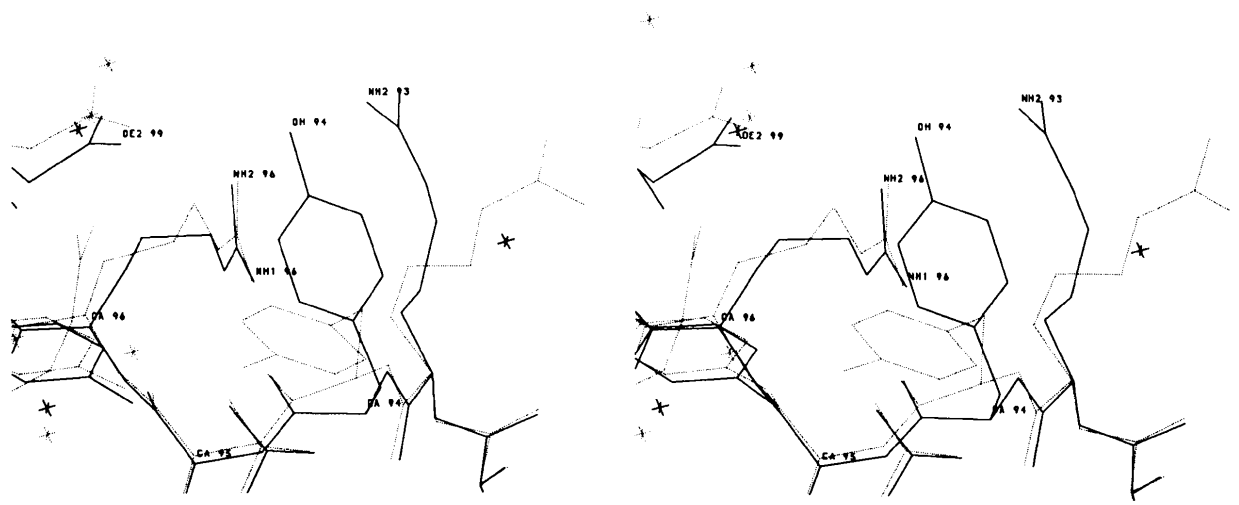

Fig. 3. Comparison of the monoclinic (full line) and orthorhombic (broken line) structures around residues 94 to 96 for which there is rearrangement due to crystal-packing differences.

crystal packing. The carbonyl oxygens of 21 and 180 are mobile atoms with that of 180 close to 94. Ser 21, Asn 84 and Asn 212 have high MSDAs, whilst Ser 70 could make an alternative hydrogen bond. Other differences are relatively minor.

Solvent structure. The final number of solvent molecules in the monoclinic structure is 226 water molecules and one methanol. The structures differ in the number of methanols observed as the orthorhombic structure has 29 methanols (Kamphius *et al.*, 1984) compared to the one in the monoclinic structure.

The solvent structures in orthorhombic and monoclinic crystals have been compared. The number of water molecules at similar positions (within 1·5 Å) in the two structures is 125 and 50 water molecules are within 0·5 Å. Detailed comparison is given in Table 12. The matched waters have a mean MSDA of 0·38

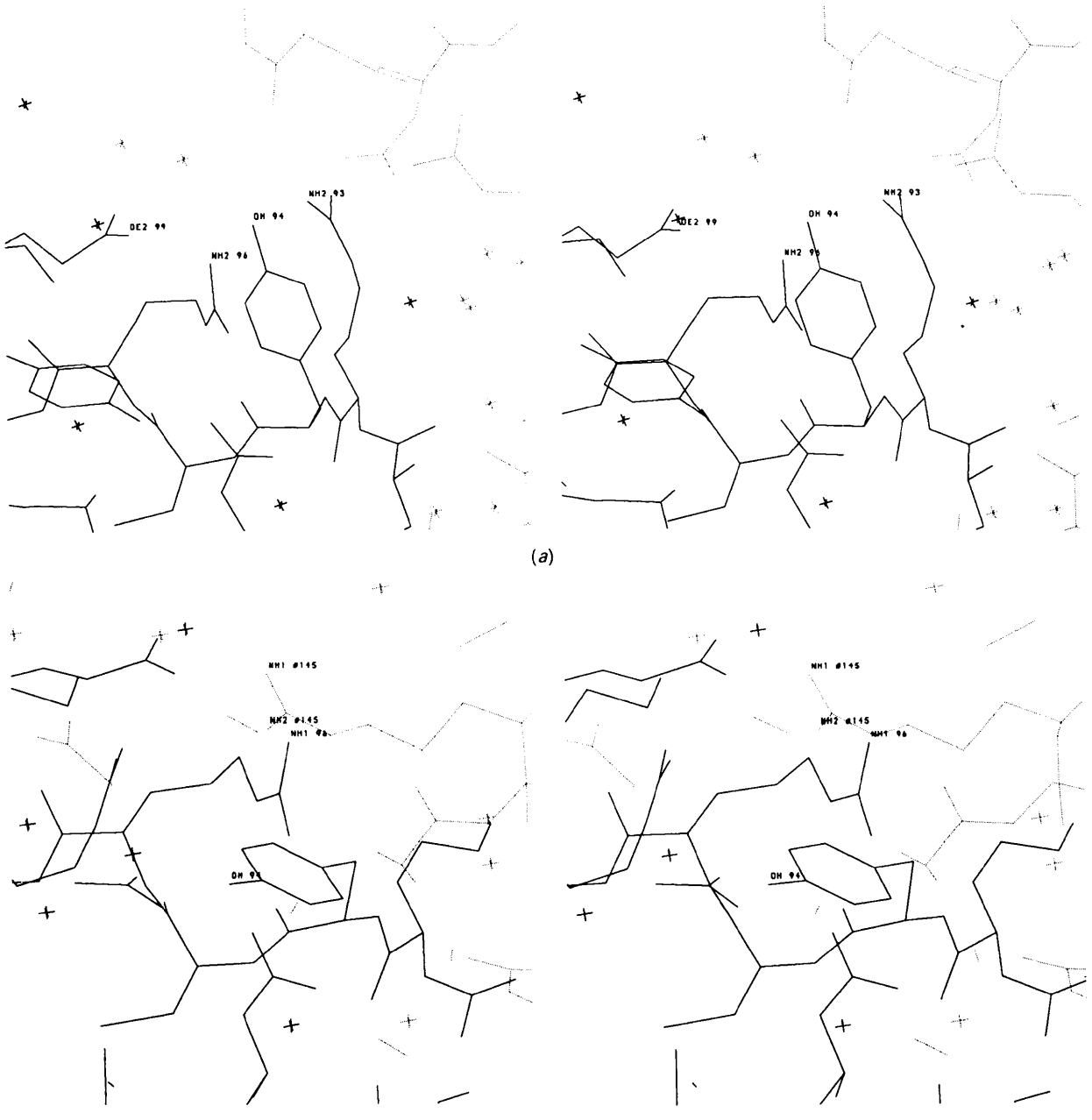

Fig. 4. (*a*) Monoclinic structure around Tyr 94 (full line) with symmetry-related molecule shown (broken line). (*b*) Orthorhombic structure around Tyr 94 (full line) with symmetry-related molecule shown (broken line).

Table 12. *Comparison of water structure in orthorhombic and monoclinic papain*

Distance between water molecules (Å)	No. of HOH	Mean MSDA for monoclinic structure (Å^2)	Mean MSDA for orthorhombic structure (Å^2)
0·0–0·5	50	0·213	0·244
0·5–1·0	43	0·472	0·497
1·0–1·5	32	0·523	0·612

(monoclinic) and 0·43 Å^2 (orthorhombic), the overall means are 0·48 and 0·54 Å^2 showing that the unmatched waters have high MSDAs and are consequently less well determined. There is no methanol in the orthorhombic structure corresponding to the single one observed in the monoclinic map.

The active site. The orthorhombic crystals contain an inactive form of the enzyme with oxidized cysteine 25. This cysteine has three O atoms attached to its Sγ. The electron density map of the monoclinic crystals also contains enzyme with covalently modified cysteine 25 (Fig. 5) (no precautions were taken to prevent modification of the active-site cysteine which occurred during preparation and crystallization). Density for an additional atom was seen in all maps examined. At a contour level of 0·65 e Å^{-3}, a strong density for the atom attached to the cysteine is seen; an extension to this density is seen at 0·45 e Å^{-3}. In the final map more additional density suggests that a small thiol-containing molecule is covalently bound. However, the density is poor which suggests that the molecule has multiple binding modes, is mobile or has low occupancy. The presence of this covalently attached species has implications for the detailed geometry of the active-site groups. The atom bound to the Sγ of 25, presumably an S atom, makes no hydrogen bonds with neighbouring atoms. In contrast in the molecules in the orthorhombic crystals the oxygens attached to Cys 25 make a series of hydrogen bonds. The Oδ3 oxygen makes a hydrogen bond with the Nδ1 atom of His 159 and as a consequence the Sγ to Nδ1 distance is 3·65 Å. In the monoclinic structure the Sγ and Nδ1 atoms are further apart, 3·80 Å, because there is no hydrogen bond to the Nδ1 atom of His 159. The histidine is slightly rotated about Cα—Cβ, translated relative to its position in the orthorhombic structure and makes a stronger hydrogen bond with Oδ1 of Asn 175. Rotation of the histidine about the Cα—Cβ bond is essential for protonation of the substrate as originally proposed by Drenth, Kalk & Swen (1976). The Oδ1 oxygen bound to the cysteine in the orthorhombic structure hydrogen bonds the main-chain amide of 25 (2·93 Å) and more weakly the Nε2 of Gln 19, but these groups do not hydrogen bond in the monoclinic structure.

The additional density attached to cysteine 25 in the monoclinic map extends so the attached molecule would lie between His 159 and the carbonyl of 158 (Fig. 5). A refinement including β-mercaptoethanol attached to the cysteine resulted in no significant change in R factor, and an overall occupancy of 0·862 and mean U of 0·498 Å^2 for the β-mercaptoethanol molecule. The mode of binding is not that of the chloromethyl derivatives which bind in the active-site cleft (Drenth *et al.*, 1976), but approximately perpendicular to the cleft.

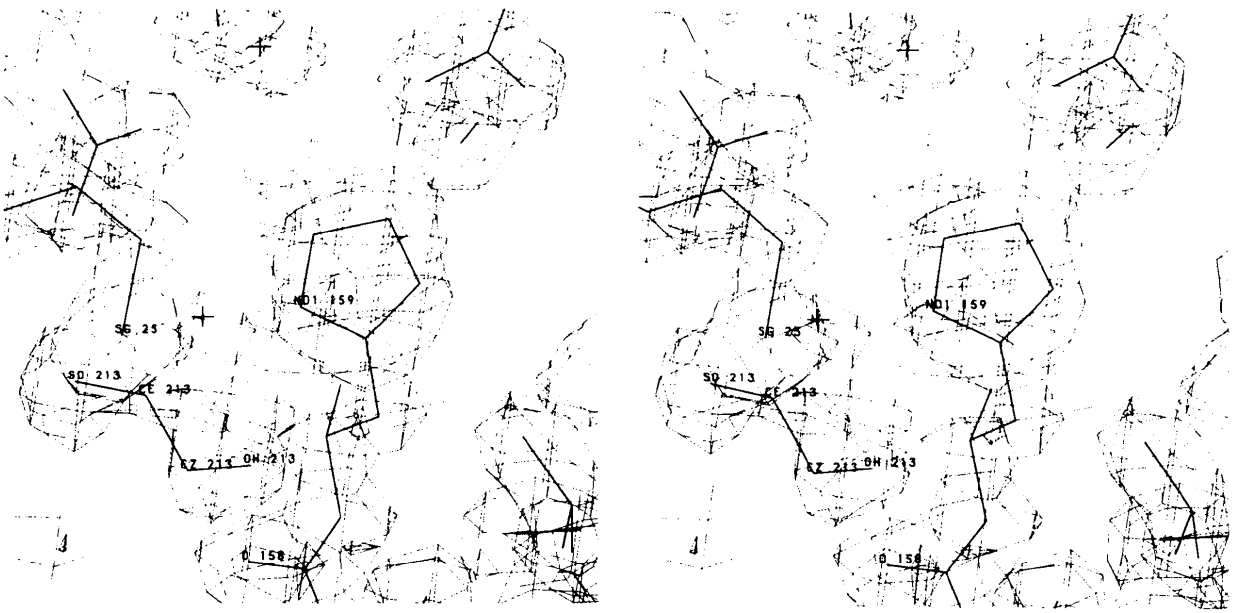

Fig. 5. Electron density around active-site groups in monoclinic papain showing strong density for the atom attached to the cysteine and the weaker extension to this density which the β-mercaptoethanol occupies. Contour level 0·45 e Å^{-3}.

Concluding remarks

The structures and dynamics of papain determined from orthorhombic and monoclinc crystals are very similar despite the different packing arrangements, pH values used for crystallization and covalent modifications of the active-site cysteines. Comparison of the main-chain structures of these models suggests that the well determined parts of the structure, those parts with MSDAs of 0·15 Å^2 or less, are determined to an accuracy of better than 0·14 Å. Several regions of the main chain differ by more than this and reflect differences due to the different crystal packing. The α-carbon of residue 94 moves by about 1·0 Å in response to changes in packing of the side chain of Tyr 94 (Fig. 4). Another cause of differences is high MSDA and correspondingly weak electron density, *e.g.* the carbonyl oxygen of residue 21. About half of the water molecules occupy similar positions (displaced by less than 1·5 Å) whilst about 20% (50 water molecules) were within 0·5 Å separation. The water molecules that occupy similar positions have low MSDAs whilst those that differ have high MSDAs. The positions of water molecules with MSDAs greater than 0·60 Å^2 should be viewed with considerable caution.

References

COHEN, L. W., COGHLAN, V. M. & DIHEL, L. C. (1986). *Gene*, **48**, 219–227.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMAN, pp. 173–178. New York: Gordon & Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

DRENTH, J., JANSONIUS, J. N., KOEKOEK, R., SWEN, H. M. & WALTHERS, B. G. (1968). *Nature (London)*, **218**, 929–932.

DRENTH, J., KALK, K. H. & SWEN, H. M. (1976). *Biochemistry*, **15**, 3731–3738.

DRIESSEN, H. P., HANEEF, M. I. J., HARRIS, G. W., HOWLIN, B., KHAN, G. & MOSS, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

HANEEF, I., MOSS, D. S., STANFORD, M. J. & BORKAKOTI, N. (1985). *Acta Cryst.* A**41**, 426–433.

HARRIS, G. W., BORKAKOTI, N., MOSS, D. S., PALMER, R. A. & HOWLIN, B. (1987). *Biochem. Biophys. Acta*, **912**, 348–356.

JONES, J. A. (1985). *J. Appl. Cryst.* **11**, 268–272.

KAMPHUIS, J. G., KALK, K. H., SWARTE, M. B. A. & DRENTH, J. (1984). *J. Mol. Biol.* **179**, 233–256.

LATTMAN, E. E. & LOVE, W. E. (1970). *Acta Cryst.* B**26**, 1854–1857.

MOSS, D. S., HOWLIN, B. & HANEEF, M. I. J. (1985). *EULER*. Unpublished program. Birkbeck College, London, England.

PICKERSGILL, R. W., SUMNER, I. G., COLLINS, M. E. & GOODENOUGH, P. W. (1989). *Biochem. J.* **257**, 310–312.

VAN GUNSTEREN, W. F. C., BERENDSEN, H. J. C. & FUJINAGA, M. (1988). *GROMOS: MD/X-ray Refinement Program (MDXDREF)*. BIOMOS b.v., Groningen, The Netherlands.

Acta Cryst. (1992). B**48**, 67–75

The Segmented Anisotropic Refinement of Monoclinic Papain by the Application of the Rigid-Body TLS Model and Comparison to Bovine Ribonuclease A

BY GILLIAN W. HARRIS AND RICHARD W. PICKERSGILL

Department of Protein Engineering, AFRC Institute of Food Research, Reading Laboratory, Shinfield, Reading RG2 9AT, England

BRENDAN HOWLIN

Department of Chemistry, University of Surrey, Guildford GU2 5XH, England

AND DAVID S. MOSS

Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England

(*Received* 29 *October* 1990; *accepted* 6 *June* 1991)

Abstract

The anisotropic displacements of selected rigid groups in monoclinic papain have been refined from X-ray diffraction data by application of the rigid-body **TLS** model. The rigid groups chosen were the aromatic side chains of tryptophan, tyrosine, histidine and phenylalanine, and the planar carboxylic and guanidinium side chains of aspartic acid, glutamic acid, glutamine, asparagine and arginine. The derived translation and libration tensors have been compared with those previously derived for bovine ribonuclease A and provide evidence for different modes and anisotropies of displacement over the two proteins.

Introduction

In view of the flexible nature of biological macromolecules, the motion of side-chain groups within proteins is likely to be highly anisotropic. Informa-

0108-7681/92/010067-09$03.00