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 Unpublished program.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

© 1992 International Union of Crystallography

tion on the anisotropic displacements of individual atoms in a protein structure X-ray refinement is limited to cases where the diffraction data is of exceptional resolution, *i.e.* greater than 1·0 Å. The poor data-to-parameter ratio for most X-ray protein structures does not allow for individual atomic anisotropic refinement of mean-square displacements.

It has recently been shown by Holbrook, Kim and co-workers (Holbrook & Kim, 1984; Holbrook, Dickerson & Kim, 1985; Holbrook, Wang, Rich & Kim, 1986) that the rigid-body **TLS** method of Schomaker & Trueblood (1968) can be applied to macromolecular structures to derive anisotropic displacements of rigid groups in DNA structures. Recently, the two domains of endothia pepsin have been refined using rigid-body **TLS** methods (Sali, Veerapandian, Cooper, Moss, Hofmann & Blundell, 1990). We have applied the rigid-body **TLS** method to the enzyme bovine pancreatic ribonuclease A (Howlin, Moss & Harris, 1989), obtaining information which correlates with the hydrogen bonding and solvent accessibility of the chosen rigid groups. Bovine ribonuclease A is a small protein (124 residues). As an extension of this work, we have applied the rigid-body **TLS** refinement method to the larger and more industrially relevant protein, papain. Papain has 212 residues, including five tryptophans, a residue type not represented in ribonuclease A.

In terms of the rigid-body **TLS** method, anisotropic displacements of relatively rigid groups are modelled by translation, libration and screw rotation tensors, the 20 components of which are refined, thus effecting a saving in the number of parameters that would be required for free anisotropic refinement. Individual anisotropic mean-square displacement amplitudes (MSDAs) were calculated from the refined **TLS** tensors. A fuller discussion of the method used and the theory of **TLS** refinement is given in Howlin, Moss & Harris (1989).

Experimental

The models used in the study were from the 1·60 Å isotropic refinement of monoclinic ($P2_1$) papain (Pickersgill, Harris & Garman, 1992) and the 1·45 Å **TLS** refinement of bovine pancreatic ribonuclease A (Howlin, Moss & Harris, 1989). Both proteins were refined using the **TLS** option in the restrained least-squares structure refinement program *RESTRAIN* (Driessen, Haneef, Harris, Howlin, Khan & Moss, 1989; Haneef, Moss, Stanford & Borkakoti, 1985), the anisotropic displacements of the selected rigid groups being refined by application of the rigid-body **TLS** method. Atoms outside the rigid groups are refined isotropically. The details of the **TLS** refinement for monoclinic papain are given in Table 1.

Table 1. *Crystal and refinement data for the* **TLS** *refinement of monoclinic papain*

Space group	$P2_1$
Unit-cell dimensions (Å and °)	a = 65·7 (2)
	b = 50·7 (2)
	c = 31·5 (2)
	β = 98·4 (2)
Resolution (Å)	1·60
Correlation coefficient	0·9613
R factor	0·1612 (all reflections)
Number of reflections	20172
Number of protein atoms	1659[a]
Number of water oxygen atoms	227[b]
Number of geometric restraints	4293
Number of **TLS** groups	68
Number of parameters	8537
Data:parameter ratio	2·87:1
Root-mean-square deviation from target geometry (Å)	
Bond lengths	0·026
Across bond angles	0·054
Peptide planes	0·020
Other planes (excluding peptide)	0·101

Notes: (*a*) including second site for Asn 169 (occupancy of 0·526); (*b*) plus one MeOH and one bound atom.

Table 2. **TLS** *groups in monoclinic papain and ribonuclease A*

	Number of groups	
Residue type	Papain	Ribonuclease A
Trp	6	
Tyr	16[a]	6
Phe	4	3
His	2	5[g]
Glu	8[b]	5
Gln	8[c]	7
Asp	6	5
Asn	11[d,f]	10
Arg	8[e]	4
Total	68	45

Notes: total of (*a*) 19, Tyr 61, Tyr 123 and Tyr 166 excluded; (*b*) 10, Glu 3 and Glu 99 excluded; (*c*) 10, Gln 73 and Gln 142 excluded; (*d*) 14, Asn 64, Asn 84 and Asn 212 excluded; (*e*) 12, Arg 41, Arg 98, Arg 111 and Arg 145 excluded; two sites for (*f*) Asn 169; (*g*) His 119, included.

The groups selected for the **TLS** refinement were: (1) indole side chains of the tryptophan (Trp) residues; (2) phenyl and imidazole side chains of phenylalanine (Phe), tyrosine (Tyr) and histidine (His) residues; (3) planar carboxylate, amide and guanidinium side chains of aspartic acid (Asp), glutamic acid (Glu), asparagine (Asn), glutamine (Gln) and arginine (Arg) side chains (see Table 2); and (4) the whole molecule. The side chains containing ring groups were included from the C*B* atom onwards. The planar side chains were four-atom rigid groups ('fishtails').

Although there are 21 unique components (6*T*, 6*L*, 9*S*) of the **TLS** tensor, the trace of *S* is indeterminate (when using only Bragg data) and hence a total of 20 independent parameters may be determined from Bragg reflections for a general rigid body. Numerical singularities will always appear if all the atoms of a rigid body lie on a conic section *i.e.* any coplanar arrangement of five or fewer atoms (Johnson & Levy, 1974). For the four-atom planar groups or 'fishtails', translational displacements perpendicular

to the plane can be described in terms of librations about two axes in the plane of the group and hence only 19 **TLS** components and linear combinations of components can be uniquely determined from Bragg data.

In the case of papain, certain of the above residues (see Table 2) were refined isotropically. With the exception of Tyr 166, the electron density for these residues was poor and the solvent accessibility large. For these residues with poor electron density, the harmonic approximation may break down (*e.g.* multimodal distribution of electron density). In such cases, the residue is better modelled as isotropic unless two or more distinct sites can be found, as for His 119 in ribonuclease A (Howlin, Moss & Harris, 1989) and Asn 169 in papain, where the **TLS** model can be applied to both positions of the side chain.

In Table 3 and throughout, the following definitions for the mean translations or librations of a group are used. For ring-containing groups, mean libration or translation is given by $\mathrm{tr}(T)/3$ or $\mathrm{tr}(L)/3$, where tr denotes the trace operation of the translation (T) or libration (L) matrix. For the 'fishtail' groups, owing to the indeterminancy of the four atom planes discussed above, only mean libration ($\mathbf{n}'L\mathbf{n}$) perpendicular to the plane and mean translation $\{\frac{1}{2}[\mathrm{tr}(T)-\mathbf{n}'T\mathbf{n}]\}$ parallel to the plane, where $\mathbf{n}$ is the unit vector perpendicular to the four-atom plane, tr denotes the trace operation and t denotes a matrix transpose, can be determined. The terms mean translation and mean libration, when used for Asp, Asn, Glu, Gln and Arg residues, are to be understood in this context. The average translation or average libration for a given residue type is the sum of the mean translations or librations for the residues of that type, divided by the number of residues of that type.

The computer program *TLSANL* (Howlin & Driessen, 1986) was used for the analysis of rigid-body parameters and the calculation of anisotropic MSDAs for individual atoms. Anisotropic ellipsoids were displayed with the program *ORTEP* (Johnson, 1965). The computer program *SOLVENT* (Lee & Richards, 1971) was used for solvent accessibility calculations. Mean solvent accessibilities were calculated for the **TLS** groups.

Results and discussion

Table 2 shows the number of each type of rigid group refined in each of the protein structures. There were 45 rigid groups in bovine ribonuclease A and 68 in papain. A comparison of the average librations and translations for the groups in the two proteins is given in Table 3.

The average libration for the tryptophan residues in papain is of the order of that found for small-molecule studies (Cruickshank, 1957), *i.e.* 14 deg^2, indicating that the tryptophans are tightly packed in the protein core and unable to undergo large librational displacement. This is confirmed by the generally low solvent accessibility (around 4 Å^2) of the tryptophans in papain.

Table 3. *Comparison of average libration and translation for the* **TLS** *groups in papain and ribonuclease A (Rnase)*

Residue type	Average libration (deg^2)		Average translation (Å^2)*	
	Rnase	Papain	Rnase	Papain
Trp	-	14	-	0·170[c]
Tyr	20	34	0·102	0·178[d]
Phe	10	22	0·130[a]	0·148
His	31	7·4	0·138[b]	0·096[e]
Glu	74	47	0·119	0·181
Gln	92	44	0·140	0·212[f]
Asp	98	28	0·153	0·157[g]
Asn	130	100	0·141	0·211[h]
Arg	85	40	0·185	0·236
Whole molecule	1·6	1·6	0·117	0·081

* For certain residues, the distance from the chosen residue origin (C*B* for Trp, Tyr, Phe, His, Asp and Asn, C*G* for Glu, Gln and N*E* for Arg) and the calculated centre of reaction is comparable to the dimensions of the molecule. This means that the group is effectively moving about a point which is outside the molecule. In such cases, the residue has an anomalously large mean translation value and has been omitted from the average translation calculation. Omitting (*a*) Phe 8; (*b*) His 48; (*c*) Trp 26; (*d*) Tyr 203; (*e*) His 81; (*f*) Gln 77, Gln 114; (*g*) Asp 158; (*h*) Asn 18, Asn 44 and Asn 184.

Generally the average libration of the other ring-containing groups is much larger for papain than for ribonuclease A (by about a factor of 2 for Phe and Tyr). The exception to this generalization is the histidine residues, for which the average libration in papain is 7·4 deg^2, compared to 31 deg^2 in ribonuclease A.

Conversely, the average libration of the 'fishtail' groups is larger in ribonuclease A than in papain, *e.g.* average libration for Glu residues is 74 deg^2 in ribonuclease A and 47 deg^2 in papain. To a first approximation, the mean libration can be thought of as describing the component of displacement which is due to local motion in the side chain, whereas the mean translation describes that part of the displacement which is more global and may be shared with the main chain or other side chains in the vicinity.

The mean translations in papain are generally larger than those in ribonuclease A. The reverse trends of the mean librations and translations for papain and ribonuclease A indicate that for papain, a greater component of the motion is due to collective motions of the protein rather than motions of individual side chains.

The exceptions to this are the histidine side chains in papain, which have both reduced librational and translational character compared to ribonuclease A. This indicates that the motion of the histidine residues is severely restricted in papain. The solvent accessibility of both histidines (His 81 and His 159)

in papain is low (around 1 Å^2), His 81 is buried in the core of the protein and active-site residue His 159 makes a hydrogen bond to Asn 175. In ribonuclease A, the solvent accessibility is low for His 12 and His 48, but larger for His 105 and His 119 (which has two sites).

Plots of mean libration *versus* solvent accessibility for **TLS** residues (by residue type) – see for example Figs. 1 and 5 for tryptophan and glutamate groups respectively – show that, in general, residues with large mean librations have high solvent accessibility. These residues are generally on the surface of the protein where motion of flexible side chains is less restricted.

The mean translation can be taken as describing that portion of the motion which is more global, as a result of collective motions of main-chain or side-chain residues. Correlations of the mean translations for residues close together could indicate regions of secondary structure. In papain, there is a correlation of mean-translation values for Glu 50, Gln 51, Glu 52, Asp 55, Asp 57 (0·107, 0·170, 0·178, 0·155, 0·150 Å^2 respectively). This corresponds to the α-helix LII (residues 50 to 57). In ribonuclease A, there is a correlation of mean translation values for Asp 53, Gln 55 and Gln 60 (0·118, 0·123, 0·118 Å^2, respectively). This falls within the well defined α-helix HIII (from residue 50 to 60). There are only three **TLS** groups in the longest α-helix LI (residues 24–42) in papain: Trp 26, Phe 28 and Glu 35, with mean-translation values of 0·933, 0·106 and 0·103 Å^2, respectively. However, correlations of mean translation values are observed for residues which are close together, but not part of a secondary-structure element, for example, in papain, Arg 83, Tyr 86 and Tyr 88, with mean translation values of 0·159, 0·152 and 0·151 Å^2, respectively. For short α helices (up to ten residues), the correlations of mean translation values of rigid groups could indicate that these groups share the motion of the helix, while for larger helices the motion is likely to be less localized, the helix sharing the overall motion of the molecule (see later, whole molecule).

Results are presented for each type of amino-acid residue refined, as well as for the whole molecule. The trends and exceptions already noted are further discussed. Certain rigid groups have very large mean libration values [for example Tyr 197 in papain (228 deg^2) and Asp 83 in ribonuclease A (235 deg^2) and/or very large mean translation values (for example Phe 8 in ribonuclease A, see also footnote to Table 3)]. This could be an indication that the motion of these groups is too anharmonic and may be better modelled another way. Anomalously large mean values have been omitted from the averaging to avoid distorting the conclusions based on the data.

Tryptophans

For the tryptophan residues in papain, there is a general increase in the mean libration with mean solvent accessibility for the residue (see Fig. 1). Similar trends were found for the ring-containing groups in ribonuclease A.

Of note is Trp 26, which shows a low solvent accessibility (mean 0·17 Å^2) and a low mean libration (7·0 deg^2), with one of the largest mean translations (0·99 Å^2) of any residue, the mean translation for the other tryptophan residues being of the order of 0·1 Å^2. This buried residue is restricted in its libration and responds mainly to the more global motion of the protein. The calculated centre of reaction is about 30 Å from the origin of the group, indicating that the residue is librating about a point far from the residue. Residues that are undergoing significant librations tend to have their centres of reaction within an ångström of the origin for the rigid group [an exception is Asn 184 in papain, which has a large mean translation (2·3 Å^2), large mean libration (129 deg^2), high solvent accessibility (27·4 Å^2) and is librating about a point a distance of 11 Å away from the origin of the group].

Trp 26 is also unusual if the anisotropy of the MSDAs of the ring atoms is considered. The anisotropy is calculated as the ratio of the largest eigenvalue of the MSDA tensor over the smallest. For Trp 7, Trp 69 and Trp 177, the anisotropies of the individual atomic MSDAs increase along the side chain, going from 1·58 at the C*B* to 5·08 at the N*E*1 for Trp 177. Similar trends are found in the ring-containing groups in ribonuclease A. For Trp 26 and Trp 181, both buried residues, the anisotropy for each atom in the ring is about 2·0. Fig. 2 shows this for Trp 26; the atomic librations are small and almost isotropic.

Phenylalanines

The results for the phenylalanines in papain are very similar to those for ribonuclease A, *i.e.* the mean librations are of the order of 10 deg^2 for ribonuclease A and 20 deg^2 for papain, and the mean translations of the order of 0·1 Å^2 [the exception is Phe 8 in ribonuclease A which has a very large translational component (95 Å^2) and is librating about a point 177 Å distant]. These values are comparable to those found in small-molecule studies (Cruickshank, 1957) and are indicative of the phenylalanines sharing much of the motion of their local environment in the protein core. This is supported by the low solvent-accessibility values (less than 4 Å^2) of the phenylalanines.

The anisotropy of the MSDAs of the ring atoms increases moving out along the side chain, the largest anisotropy being found for the C*Z*, *e.g.* in Phe 28 in

papain the anisotropy of *CG* is 2·71 and that of *CZ* 7·25. The magnitudes of the anisotropies are similar in papain and ribonuclease A, reflecting the similarity in the MSDAs of the phenylalanine rings for the two proteins.

Tyrosines

For the five tyrosine residues in ribonuclease A, the anisotropy of the MSDAs of the ring atoms follows the same trend as the phenylalanine rings, *viz.* the anisotropy increases along the side chain, with the largest value for the OH. Of the 16 tyrosine groups in papain, half of the rings conform to this pattern (*viz.* Tyr 4, Tyr 67, Tyr 78, Tyr 88, Tyr 170, Tyr 186, Tyr 197, Tyr 208). A further two tyrosine rings show the anisotropy concentrated on the *CD*1, *CE*1 side of the ring (Tyr 94 and Tyr 116) and the remaining six show equal anisotropy (of the order of 2) around the ring (Tyr 48, Tyr 82, Tyr 86, Tyr 103, Tyr 144, Tyr 203). These latter tyrosines have low solvent accessibility.

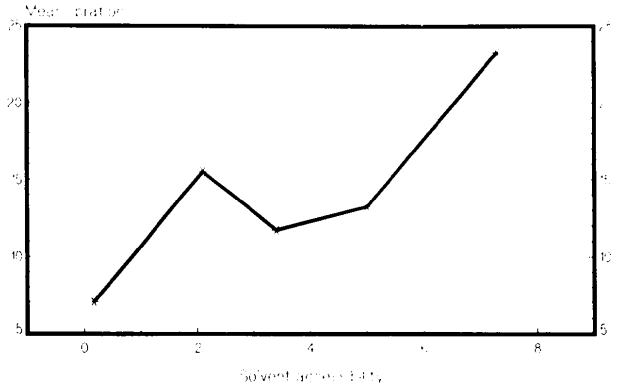

Fig. 1. Plot of mean libration (deg^2) *versus* solvent accessibility (Å^2) for the tryptophan groups in monoclinic papain.

Tyr 197 shows a very large mean libration of 228 deg^2 and some startling anisotropies, OH = 16·4, *CZ* = 10·9, *CE*1 = 10·8 (Fig. 3). The side chain nearest to the main chain is buried, whereas the other half of the ring is extremely exposed, Hence, the mean solvent accessibility for the group is 13·6 Å^2, but that of OH is 39 Å^2 and that of *CG* 0 Å^2.

The higher average libration of the tyrosine residues in papain (34 deg^2) compared to the average libration of the tyrosines in ribonuclease A (20 deg^2) is due to the distorting influence of Tyr 197. Omitting Tyr 197, the average libration for the remaining 15 tyrosines in papain is 21 deg^2, similar to that for ribonuclease A.

Tyr 94 has an average solvent accessibility of 16 Å^2, but a mean libration of 14 deg^2. The mean translation for this residue is relatively large, 0·42 Å^2. The anisotropies for the *CD*1, *CE*1 side of the ring (about 10) are much larger than those for the *CD*2, *CE*2 side (about 2). The residue appears to be undergoing a complicated motion involving larger displacements of one side of the ring.

Tyr 116 also shows a similar asymmetry of ring isotropy, although less pronounced (*CD*1 14, *CE*1 8; *CD*2, *CE*2 side 5). The mean solvent accessibility is 11·9 Å^2, the mean libration 28 deg^2, and the mean translation 0·17 Å^2. In contrast to Tyr 94, the whole of which is exposed, the solvent accessibility varies greatly around the ring, from 1 Å^2 for *CD*1 to 32 Å^2 for *CE*2. The largest anisotropy, that for *CD*1, corresponds to the lowest solvent accessibility. A similar example of this unexpected correlation is found for His 159 in papain (see later).

Tyr 203 in papain has a large mean translation (0·96 Å^2). This is another example of a residue with a centre of reaction at a distance (17 Å) from the group.

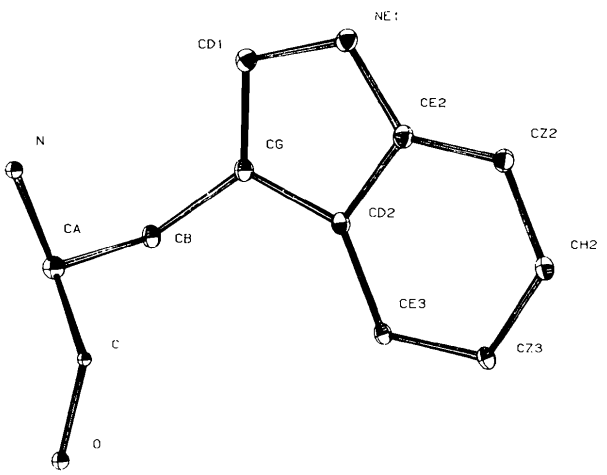

Fig. 2. *ORTEP* plot of Trp 26 in papain (at the 50% probability level).

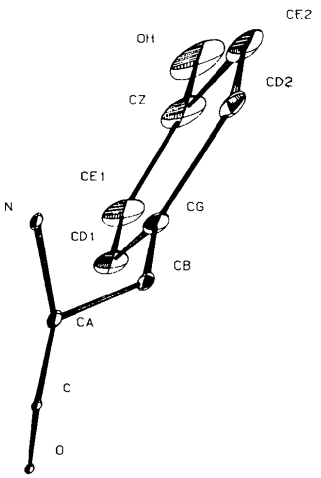

Fig. 3. *ORTEP* plot of Tyr 197 in papain (at the 22·5% probability level).

Histidines

The anisotropy of the MSDAs of the ring atoms for the histidine residues in ribonuclease A follows the same pattern as that for the tyrosine and phenylalanine rings in ribonuclease A, *viz.* the anisotropy increases going out along the side chain (although for His 105 the anisotropy at *CB* is also large), with the exception of His 48, for which the anisotropy of the ring atoms is asymmetric, *e.g.* 11·07 for *CD*2, 7·05 for *NE*2, compared to 2·46 for *ND*1, 3·12 for *CE*1. For His 48, the mean translation is very large (10 Å^2), the centre of reaction being located at a distance of 51 Å from the origin of the group. The larger ellipsoids on one side of the ring could reflect the remoteness of the centre of reaction from the ring.

The histidines in ribonuclease A can be divided into two types: His 12 and His 48 are buried and have mean librations of the order of 17 deg^2, whereas His 105 and His 119 have higher mean solvent accessibility (about 10 Å^2) and larger mean librations (of the order of 40 deg^2). The two histidines in papain are buried and have low mean librations (3·2 deg^2 for His 81, 12 deg^2 for His 159), resembling the first type in ribonuclease A.

The two histidines in papain show different behaviour from each other and from the histidines in ribonuclease A. His 81 is another residue for which the centre of reaction is located away (20 Å) from the origin of the group; consequently, the mean libration is low (3·2 deg^2), and the mean translation is large (1·06 Å^2). This residue is almost completely buried, only *CD*2 showing any non-zero solvent accessibility and the anisotropies of the atoms are similar (about 2·9).

His 159, which is an active-site residue in papain, resembles His 48 in ribonuclease A with respect to the pattern of anisotropy of the ring atoms *viz.* 16·24 for *CD*2, 49·95 for *NE*2, compared to 4·38 for *ND*1, 6·13 for *CE*1. The mean libration of His 159 (12 deg^2) is similar to that of His 48 (15 deg^2), although the translation is much less (0·092 Å^2). Whereas His 48 is completely buried, for His 159, only *CD*2 and *NE*2 are completely inaccessible, the solvent accessibility being highest for *CB* and *ND*1. As was found for Tyr 116 in papain, the largest anisotropy corresponds to the lowest solvent accessibility for the ring atoms. For the active-site residue His 159 in papain, it is *ND*1, with the highest solvent accessibility (3·6 Å^2) and the lowest anistropy (4·38), which is involved in bonding to bound non-protein atoms in the active site. This is a contrast to the active-site histidine 119 in ribonuclease A, which has high solvent accessibility (9·5 Å^2), high mean libration (39 deg^2 for site *A*, 53 deg^2 for site *B*, 20% occupancy) and is highly mobile, being located in at least two separate positions in the protein. Hydrogen bonding, *e.g.* to Asp 121, may help to stabilize individual sites.

In monoclinic papain, the *NE*2 atom of the His 159 makes a hydrogen-bonding contact with *OD*1 of Asn 175 (see Fig. 4). The motion of the *NE*2 atom of His 159 is thus restricted as evidenced by the smaller ellipsoid. The anisotropy however is highest on the *NE*2 side of the ring (16·2 for *CD*2 and 49·9 for *NE*2), with the greatest librations perpendicular to the bond. The motion is more isotropic on the *ND*1 side of the ring, but the ellipsoids are larger. In monoclinic papain the *ND*1 of His 159 makes no contacts and is thus free to move within the restrictions imposed by the ring, resulting in a flapping motion of this side of the ring.

Aspartates and asparagines

For the Asp and Asn residues in both proteins, a general correlation is found between increasing mean libration and greater solvent accessibility. Asp 83 in ribonuclease A is an exception to the general rule, as the solvent accessibility is relatively low (6·4 Å^2), but the mean libration is 235 deg^2. Similar examples are Asn 155, Asn 195 (and Tyr 197) for papain.

As has been mentioned earlier, the mean librations for the 'fishtail' groups are much greater in ribonuclease A than in papain. This is especially noticeable for aspartates (even omitting Asp 83, the average libration for the Asp residues in ribonuclease is 64 deg^2, compared to 28 deg^2 for papain) and reflects the strong hydrogen bonding to other residues or waters of the Asp residues in papain. In ribonuclease A, Asp 14 and Asp 121, which have similar mean librations (33 deg^2), are also involved in hydrogen bonding.

In monoclinic papain, Asn 169 has at least two distinct sites. The mean libration for the second site

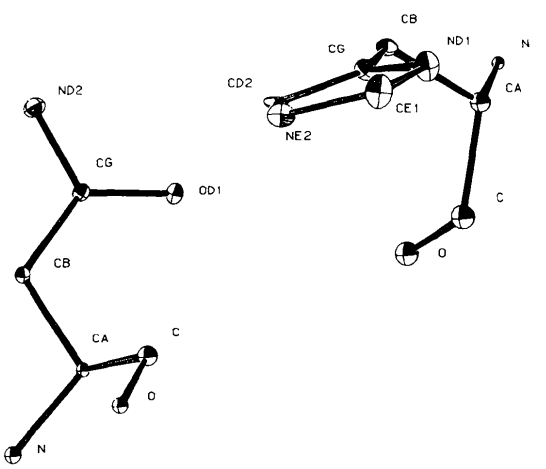

Fig. 4. *ORTEP* plot of Asn 175 and His 159 in papain (at the 50% probability level).

(occupancy 53%) is three times that of the first site (76 deg^2 compared to 26 deg^2), although the mean translations are similar (about 0·25 Å^2). This mobility of the side chain is indicative of several positions which cannot be resolved by X-ray analysis.

The mean translations are large for Asn 18, Asn 44, Asn 184 and Asp 158 in papain. All four of these residues have the centre of reaction at a distance (of 6–14 Å) from the origin of the group.

Asn 175, which hydrogen bonds to His 159 in papain, is shown in Fig. 4. For Asn 175, the ellipsoids for the side-chain atoms are perpendicular to the bond and thus represent real libration rather than disorder. Similar anisotropy, with the amplitudes of the ellipsoids being smaller in the directions of covalent bonds, has been observed in a **TLS** study of avian polypeptide (Glover *et al.*, 1985).

The mean libration for Asn 175 is 16 deg^2. In ribonuclease A, the Asn groups (Asn 44, Asn 67, Asn 71) that are involved in hydrogen bonding also have lower mean librations (< 55 deg^2).

Glutamates and glutamines

The glutamines for both proteins follow the general trend of increasing mean libration with greater solvent accessibility, with the magnitude of the mean librations for ribonuclease A on average twice as large as for papain. The mean translations are large for Gln 77 and Gln 114 in papain. These two groups are further examples of residues having remote centres of reaction.

For the glutamates, the general correlation between larger mean libration and higher solvent accessibility is shown very clearly in a plot of mean libration *versus* mean solvent accessibility for the Glu's of ribonuclease A and papain plotted on the same graph (Fig. 5), forming an almost continuous plot with an upward trend, the Glu's in papain being in the lower left portion (lower mean libration, smaller solvent accessibility) of the graph, and the Glu's in ribonuclease A being in the upper right portion (higher mean libration, greater solvent accessibility). The mean translations for the Glu residues are similar in both proteins.

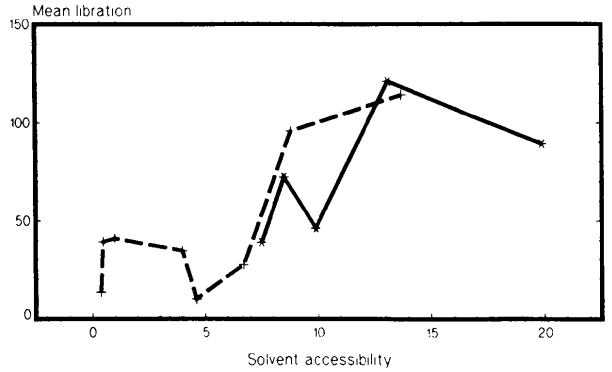

Fig. 5. Plot of mean libration (deg^2) *versus* mean solvent accessibility (Å^2) for the glutamate groups in papain (dotted line) and ribonuclease A (solid line).

Arginines

The four arginines in ribonuclease A fall into two groups, Arg 10 and Arg 33 have low solvent accessibility (average 9 Å^2), low mean libration (average 24 deg^2) and low mean translation (average 0·12 Å^2), while Arg 39 and Arg 85 have high solvent accessibility (average 20 Å^2), high mean libration (average 146 deg^2) and high mean translation (average 0·26 Å^2). These resemble the two types of histidines in ribonuclease A (see earlier). The arginine groups in papain are more similar to the first group of arginines in ribonuclease A, with solvent accessibility of 13 Å^2, average libration of 40 deg^2, although the average translation of 0·24 Å^2 is more comparable to the second group, reflecting the generally higher mean translations found in papain.

The whole molecule

The results for the refinement treating the whole protein molecule as a rigid body (Table 3) indicate that the mean librations are very similar (1·6 deg^2) for both proteins and the molecules are undergoing small librational displacements. The components of the **L** tensor are fairly isotropic (diagonal components 1·9, 1·4, 1·6) for papain, more anisotropic for ribonuclease A (diagonal components 1·3, 2·1, 1·4).

The mean translation is slightly smaller for papain than for ribonuclease A, but of the same order of magnitude (0·1 Å^2). The **T** tensor is almost isotropic for both proteins.

Since the protein molecule is not a pseudo rigid body, the tensors do not have a simple interpretation in terms of translation and libration (see Howlin, Moss & Harris, 1989). However, the results indicate that the overall molecular motion of both proteins is isotropic and is mainly translational in character.

Refinement of the whole-molecule **TLS** parameters used the molecular centre of mass as origin for the **TLS** calculation. Fig. 6 shows the papain molecule and its principal axes of libration. It is of interest to note that one of the principal axes of libration aligns approximately with the helix axis of α-helix LI (residues 24 to 42), the longest α-helix in papain.

The average MSDAs for the main-chain and side-chain atoms (Table 4) for papain and ribonuclease A show that the displacements of the side-chain atoms are larger than those of the main-chain atoms, *e.g.* for papain, average side-chain MSDA = 0·236 Å^2, average main-chain MSDA = 0·150 Å^2. The MSDAs

Table 4. *Comparison of average MSDAs (Å²) for monoclinic papain and ribonuclease A*

	Papain	Ribonuclease A	
	1·6 Å resolution	1·4 Å resolution	1·6 Å resolution
Main-chain atoms	0·150	0·168	0·165
Side-chain atoms	0·236	0·277	0·276
All protein atoms	0·192	0·220	0·218

are of the same order of magnitude for both proteins, with those for ribonuclease A (at 1·6 Å) being some 12–17% larger.

The MSDAs in Table 4 have been corrected for the difference in resolution between the papain and ribonuclease A structures by comparing the MSDAs for a refinement of ribonuclease A at 1·6 Å (Harris & Moss, 1992). The MSDAs for ribonuclease A are only slightly smaller at 1·6 Å than at 1·4 Å (*i.e.* average MSDA for main-chain atoms is 0·165 Å at 1·6 Å and 0·168 Å at 1·4 Å).

Concluding remarks

We have previously shown by comparison of MSDA values for monoclinic and orthorhombic crystals of papain that the MSDAs reflect true molecular motion rather than some property of the crystal (Pickersgill, Harris & Garman, 1992).

The motion of the side chains in papain appears to be of a significantly different nature to that of ribonuclease A, *i.e.* there is a tendency in papain for less librational motion and more tanslational motion (with the exception of Tyr and Phe residues). The displacements in papain thus appear to reflect more global motion of the protein rather than localized motion of individual residues. For example, the mean librations of the 'fishtail' groups are on average

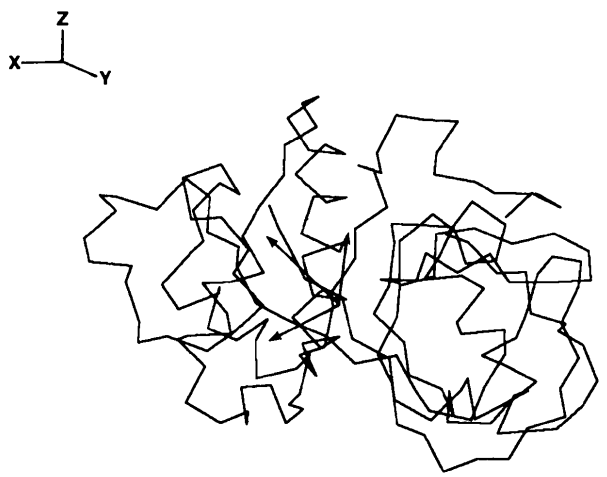

Fig. 6. Projection of the backbone structure of papain. The arrows show the principal libration axes drawn to the scale of the mean-square libration amplitudes. The origin of the axes is at the centre of reaction.

Table 5. *Average libration and translation for the* **TLS** *groups in* (I) *papain and* (II) *ribonuclease A for* (*i*) *solvent-accessible* (*mean solvent accessibility* >5 Å²) *and* (*ii*) *buried* (*mean solvent accessibility* <5 Å²) *residues*

Residue type	Number of groups	Average solvent accessibility (Å²)	Average libration (deg²)	Average translation (Å²)
(I) Papain				
(i) Solvent-accessible groups				
Trp	2	6·2	18	0·160
Tyr	8	12·1	25[a]	0·221
Phe	0			
His	0			
Glu	3	9·7	79	0·254
Gln	5	19·6	53	0·291[b]
Asp	4	9·6	31	0·160[c]
Asn	9	16·5	112	0·227[d]
Arg	7	14·7	35	0·237
Total	38		57	0·226
(ii) Buried groups				
Trp	3	1·89	11	0·180[e]
Tyr	8	1·88	18	0·129[f]
Phe	4	1·61	22	0·148
His	2	0·96	7	0·096[g]
Glu	5	2·09	28	0·137
Gln	3	2·03	27	0·132
Asp	2	0·55	22	0·153
Asn	1	0·03	16	0·080
Arg	1	3·50	72	0·224
Total	29		22	0·140
(II) Ribonuclease A				
(i) Solvent-accessible groups				
Trp	0			
Tyr	3	11·5	23	0·103
Phe	0			
His	2	9·7	35	0·157
Glu	5	11·8	74	0·119
Gln	6	13·7	94	0·142
Asp	3	16·2	141	0·180
Asn	9	17·7	141	0·145
Arg	4	14·3	85	0·185
Total	32		97	0·145
(ii) Buried groups				
Trp	0			
Tyr	3	1·81	16	0·101
Phe	3	0·72	10	0·13[h]
His	2	0·20	17	0·10[i]
Glu	0			
Gln	1	2·50	79	0·127
Asp	2	0·89	33	0·113
Asn	1	0·11	32	0·101
Arg	0			
Total	12		24	0·112

Notes: Average libration excluding (*a*) Tyr 197 (227·9 deg²). Average translation excluding residues with remote centres of reaction (*b*) Gln 77, Gln 114; (*c*) Asp 158; (*d*) Asn 18, Asn 44 and Asn 184; (*e*) Trp 26; (*f*) Tyr 203; (*g*) His 81; (*h*) Phe 8; (*i*) His 48.

twice as large in ribonuclease A as in papain. Papain has about twice as many amino-acid residues as ribonuclease A and therefore has a greater percentage of buried residues (43% as against 27%). As a consequence residues in ribonuclease are more likely to undergo librational motion as their motion is less restricted. The behaviour of the buried-ring residues in papain resembles that of ring groups in a close-packed small-molecule environment, with small mean librations and translations.

Table 5 shows the **TLS** residues in papain and ribonuclease A divided into solvent accessible and buried residues. For the buried groups in the two proteins the average librations are similar in papain

(22 deg^2) and ribonuclease A (24 deg^2), and the average translation in papain (0·14 Å^2) is similar to that for ribonuclease A (0·11 Å^2).

For the solvent-accessible residues, the average libration is considerably greater for ribonuclease A (97 deg^2) than for papain (57 deg^2), whereas the average translation is greater for papain (0·23 Å^2) than for ribonuclease A (0·14 Å^2). The mean librations appear similar for the buried residues in both proteins as might be expected from the close-packed nature of these residues. The mean librations are greater for the solvent-accessible residues in ribonuclease A than for those in papain and thus the larger average librations found in ribonuclease A are accounted for by the greater percentage of surface groups having higher librations.

The mean translations are generally higher in papain than in ribonuclease A, especially the solvent-accessible residues. Thus there is more motion in papain of a translational nature, whereas for ribonuclease A, the motion is comprised in the main of surface-group librations.

Some of the motions described here may be of biological significance. For instance, in papain His 159 rocks about a single conformation; this may be of importance for catalysis because the histidine is expected to rock to protonate the proposed tetrahedral intermediate. His 119 in ribonuclease occupies two positions and multiple positions may be important for catalysis by this enzyme.

References

CRUICKSHANK, D. W. J. (1957). *Acta Cryst.* **10**, 504–508.

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

GLOVER, I. D., MOSS, D. S., TICKLE, I. J., PITTS, J. E., HANEEF, I., WOOD, S. P. & BLUNDELL, T. L. (1985). *Adv. Biophys.* **20**, 1–12.

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

HARRIS, G. W. & MOSS, D. S. (1992). *Acta Cryst.* A**48**, 42–45.

HOLBROOK, S. R., DICKERSON, R. E. & KIM, S. H. (1985). *Acta Cryst.* B**41**, 255–262.

HOLBROCK, S. R. & KIM, S. H. (1984). *J. Mol. Biol.* **173**, 361–388.

HOLBROOK, S. R., WANG, A. H. J., RICH, A. & KIM, S. H. (1986). *J. Mol. Biol.* **187**, 429–440.

HOWLIN, B. & DRIESSEN, H. (1986). *TLSANL*. Unpublished computer program.

HOWLIN, B., MOSS, D. S. & HARRIS, G. W. (1989). *Acta Cryst.* A**45**, 851–861.

JOHNSON, C. K. (1965). *ORTEP*. Report ORNL-3794. Oak Ridge National Laboratory, Tennessee, USA.

JOHNSON, C. K. & LEVY, H. A. (1974). *International Tables for X-ray Crystallography*, Vol. IV, edited by J. A. IBERS & W. C. HAMILTON, pp. 320–322. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

LEE, B. & RICHARDS, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

PICKERSGILL, R. W., HARRIS, G. W. & GARMAN, E. (1992). *Acta Cryst.* B**48**, 59–67.

SALI, A., VEERAPANDIAN, B., COOPER, J. B., MOSS, D. S., HOFMANN, T. & BLUNDELL, T. L. (1990). *J. Mol. Biol.* In the press.

SCHOMAKER, V. & TRUEBLOOD, K. N. (1968). *Acta Cryst.* B**24**, 63–76.

Acta Cryst. (1992). B**48**, 75–88

Structure Determination of Monoclinic Canine Parvovirus

BY JUN TSAO,* MICHAEL S. CHAPMAN, HAO WU, MAVIS AGBANDJE, WALTER KELLER AND MICHAEL G. ROSSMANN

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, *USA*

(*Received* 30 *April* 1991; *accepted* 2 *August* 1991)

Abstract

The three-dimensional structure of the single-stranded DNA canine parvovirus has been determined to 3.25 Å resolution. Monoclinic crystals belonging to space group $P2_1$ (a = 263.1, b = 348.9, c = 267.2 Å, β = 90.82°) were selected for data collection using primarily the Cornell High Energy Synchrotron Source and oscillation photography. There was one icosahedral particle per crystallographic asymmetric unit, giving 60-fold redundancy. The particle orientations in the unit cell were determined with a rotation function. The rough positions of the particles in the unit cell were estimated by considering the packing of spheres into the $P2_1$ crystal cell. More accurate particle centers were determined from Harker peaks in a Patterson function. Hollow-shell models were used to compute phases to 20 Å resolution. The radii of the models were based on packing considerations, the fit of spherical shells to the low-resolution X-ray data and

* Present address: Department of Microbiology, Center for Macromolecular Crystallography, University of Alabama, Birmingham, Alabama 35294, USA.

0108-7681/92/010075-14$03.00

© 1992 International Union of Crystallography