RIVERA, J.-P., SPEZIALI, N. L., BERGER, H., AREND, H. & SCHMID, H. (1990). *Ferroelectrics*, **105**, 183–188.
SAWADA, S., YAMAGUCHI, T. & SUZUKI, H. (1985). *Ferroelectrics*, **63**, 3–11.
SIMMONS, E. H. & HEINE, V. (1987). *Acta Cryst.* A**43**, 626–635.
SPEZIALI, N. L. (1989). *Phase Transitions in* Cs_2CdBr_4 *and* $N(CH_3)_4HSO_4$: *a Crystallographic Study of Normal and Modulated Phases*. PhD Thesis, Univ. of Lausanne, Switzerland.
STEWART, J. M., KUNDELL, F. A. & BALDWIN, J. C. (1972). The *XRAY*72 system. Tech. Rep. TR-192. Computer Science Center, Univ. of Maryland, College Park, Maryland, USA. Modified by D. SCHWARZENBACH.
STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175–3187.
WOLFF, P. M. DE, JANSSEN, T. & JANNER, A. (1981). *Acta Cryst.* A**37**, 625–636.
YAMAMOTO, A. (1982). *Acta Cryst.* A**38**, 87–92.

Acta Cryst. (1991). B**47**, 766–771

Determination of the Structure of Papaya Protease Omega

BY RICHARD W. PICKERSGILL*

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

PIERRE RIZKALLAH

SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, England

AND GILLIAN W. HARRIS AND PETER W. GOODENOUGH

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

(*Received* 13 *September* 1990; *accepted* 11 *March* 1991)

Abstract

The structure of papaya protease omega (ppΩ) has been determined using Enraf–Nonius FAST TV data collected using the Daresbury Synchrotron Radiation Source (SRS). This is the first protein structure to be determined using the FAST/SRS combination and the first protein structure to be solved in space group $P3_112$. The structure has been refined to a crystallographic R factor of 0·1549 for all data in the range 10·0–1·8 Å.

1. Introduction

Papaya protease omega (ppΩ) is a cysteine protease from the latex of *Carica papaya*. ppΩ is homologous to papain (69% sequence identity; Dubois, Kleinschnidt, Schnek, Looze & Braunstzen, 1988) and the level of homology suggests that the structures of papain and ppΩ are similar and that the structure of ppΩ may therefore be solved by molecular replacement. The structure of papain was determined by multiple-isomorphous replacement and refined by Drenth and colleagues (Kamphuis, Kalk, Swarte & Drenth, 1984). A monoclinic ($P2_1$) crystal form has been solved by molecular replacement recently (Pickersgill, Harris & Garman, 1991) and refined to $R = 0{\cdot}1596$ using 10·0 to 1·60 Å data. This structure was used as the search model in molecular replacement.

The activity of ppΩ is similar to that of papain although subtle differences in specificity and stability have been reported.

ppΩ crystals diffract only weakly using a sealed-tube source and to a resolution limit of about 2·6 Å using X-rays from a rotating-anode source. At the Daresbury synchrotron radiation source (SRS) these crystals diffracted to 1·8 Å. This paper reports data collection, reduction and the use of these data to solve and refine the structure of ppΩ.

2. Experimental procedures

2.1. *Crystals and SRS/FAST data collection and reduction*

Protease Ω was eluted from a carboxymethyl-Sepharose column and dialysed against 100 mM sodium acetate at pH 5·0 containing 2·0 mM mercuric chloride (Pickersgill, Sumner & Goodenough, 1990). The mercuric chloride prevented autolysis of the protein. Before crystallization the protein was dialysed against 50 mM Tris buffer at pH 8·0 containing 250 mM sodium chloride to remove any

* Author to whom correspondence should be addressed.

0108-7681/91/050766-06$03.00

Table 1. *SRS/FAST data collected*

Data	Dectector 2θ* ( )	ω ( )	SRS† (mA)	Ion chamber	Exposure frame (s)
Main region	−15·0	0·0–181·0	287	0·26	10·0
Blind region‡	15·0	0·0 31·0	177	0·20	13·0
Low resolution‡	0·0	0·0 41·0	163	0·17	10·0

* Crystal detector distance set to 100·0 mm.
† SRS at 2 GeV; multibunch mode.
‡ It was not necessary to translate the crystal along the needle axis for the blind-region and low-resolution collections.

unbound mercury. Crystals of ppΩ were grown by vapour diffusion of 200 µl of a 1·8% protein solution in 50 m*M* Tris buffer at pH 8·0 containing 250 m*M* sodium chloride against a reservoir of ethanol and water. The reservoir contained 82% ethanol and 182 m*M* sodium chloride in water. These needle-shaped crystals were trigonal, space group $P3_112$ (or $P3_212$) with $a = 74{\cdot}2$ (2), $c = 77{\cdot}9$ (2) Å and with one molecule in the asymmetric unit (Pickersgill, Sumner & Goodenough, 1990).

Data were collected from a single ppΩ crystal of size 0·8 × 0·2 × 0·2 mm using the Enraf–Nonius FAST TV detector on station 9.6 of the SERC Daresbury Synchrotron. The crystal was mounted with the needle c axis parallel to the capillary and spindle axes. The crystal-to-detector distance was set to 100·0 mm and the detector swung out to a 2θ angle of −15·0°, giving a maximum resolution of about 1·8 Å. Data were collected by oscillating the crystal about ω and a 0·1° oscillation range was collected per frame. With a 0·2 mm diameter collimator, the SRS source operating in multibunch mode at 2 GeV and 287 mA and with the wiggler magnet at 5 T and the wavelength at 0·904 Å diffraction could be recorded to the diffraction limit of the crystal using an exposure of 10 s per frame.

Data were collected in three batches: (1) 1·8 Å main data consisting of 181·0° of data, (2) 1·8 Å blind-region data with κ moved through 90·0° consisting of 31·0° of data and (3) low-resolution data, with the detector swing angle set to 0·0° (Table 1).

These data were processed using the *MADNES* software (Messerschmidt & Pflugrath, 1987) with modifications for the characteristics of the SRS source (Papiz & Andrews, 1987) and using the profile-fitting and correction options from the program *XDS* (Kabsch, 1988). These data were scaled and merged using *AGROVATA/ROTAVATA* (*CCP*4 program suite, Daresbury, England).

2.2. *Molecular replacement*

The structure of $P2_1$ papain (Pickersgill, Harris & Garman, 1991) was used as a search model in rotation-function calculations. The program *MERLOT* (Fitzgerald, 1988) was used for rotation-function calculations. The search model was centred in a cubic cell of side 80·0 Å, such that no intramolecular vectors shorter than 24·0 Å occurred, because this was the radius of integration used in the rotation function. Structure factors were calculated for the papain structure in the $P1$ cell using a temperature factor of 15 Å^2. These calculated structure factors and the observed structure-factor amplitudes (10·0 to 4·0 Å) for ppΩ were used in the Crowther fast rotation function (Crowther, 1972). The solution was refined to within ±1° using the Lattman rotation function (Lattman & Love, 1970).

The search model was rotated according to the refined molecular-replacement solution and used in the Crowther–Blow translation function (Crowther & Blow, 1967) and in R-value search calculations.

3. Results and discussion

3.1. *SRS/FAST data*

Data were collected to a much higher resolution limit than expected. On a rotating anode the resolution limit was about 2·6 Å but with SRS radiation this was increased to 1·8 Å. The increased resolution achieved is due to the combination of the high intensity and hence short data-collection time, low wavelength and hence low absorption and low radiation damage, and high-quality parallel-beam geometry from the synchrotron source. The X-ray wavelength was known to be 0·904 (1) Å as a result of refining the wavelength using a crystal of known cell dimensions. Refinement of the cell dimensions of ppΩ, crystal orientation and detector parameters gave $a = 74{\cdot}11$ (5), $c = 77{\cdot}81$ (2) Å, detector distance 98·497 (15) mm and swing angle −15·053 (1)°. The cell dimensions are in good agreement with the Cu $K\alpha$ measurements and the detector parameters with those set. Examination of the symmetry of the data revealed the point group to be 312 and not the more frequently observed 321. The data (Bijvoets not merged) consisting of a total of 200 612 reflections were scaled and merged in point group 312 using *AGROVATA/ROTAVATA* which gave a merging R_{sym}, on intensity, of 11·1% for 18 372 unique reflections to 1·8 Å (Table 2).

3.2. *Molecular replacement*

The sequences of papain and ppΩ are homologous (about 69% identity). Compared to papain, ppΩ has one insertion of four residues (at position 168) and 64 substitutions. The basic fold of these proteins is expected to be similar and molecular replacement was the method chosen to solve the structure of ppΩ. Structure-factor amplitudes in the range 10 to 4 Å were used in the rotation function and the solution $\alpha = 46$, $\beta = 90$, $\gamma = 35°$ was 6·93σ above background. This solution was refined to $\alpha = 45$ (1), $\beta = 88$ (1), γ

Table 2. *Internal agreement of data and percentage of data recorded as a function of resolution*

Resolution range (Å)	R_{sym}*	R_{sym} (cumulative)	% complete	% complete (cumulative)
∞ 8·03	11·7	11·7	100	100
8·03–5·68	9·6	10·4	100	100
5·68 4·64	9·0	9·7	100	100
4·64–4·02	9·2	9·5	100	100
4·02 3·60	9·5	9·5	100	100
3·60–3·29	9·6	9·5	100	100
3·29 3·04	9·1	9·4	98	100
3·04 2·85	9·1	9·4	100	100
2·85–2·68	9·3	9·4	100	100
2·68–2·55	10·2	9·4	100	100
2·55 2·43	11·2	9·6	90	99
2·43–2·32	12·0	9·7	80	96
2·32–2·23	12·5	9·8	84	95
2·23 2·15	14·1	9·9	79	93
2·15–2·08	15·4	10·1	80	92
2·08–2·01	17·7	10·3	69	90
2·01 1·95	20·9	10·5	69	88
1·95 1·90	26·3	10·7	77	87
1·90–1·85	31·2	10·9	68	86
1·85 1·80	33·4	11·1	50	83

* $R_{sym} = \Sigma_n[\Sigma_i^m{}_1(I_i - \bar{I})/\Sigma_i^m{}_1\bar{I}]$, where there are m equivalents with average intensity I and n sets of equivalent reflections.

= 37 (1)° using the Lattman rotation function. The ppΩ molecule in the crystal was expected to have a single mercury bound at the active-site cysteine and a mercury atom was added to the search model with a Cys25 SG to mercury distance of 2·54 Å. Mercury (Hg^{2+}) was included in *MERLOT*, *BRUTE* (Fujinaga & Read, 1987) and *XPLOR* (Brünger, 1988) programs.

The Crowther–Blow translation function, in $P3_112$ and $P3_212$, gave a number of peaks from which it was not easy to identify the correct solution, although some could be ruled out on packing considerations. The packing of molecules was examined using *PACK* running on an Evans and Sutherland PS390. *R*-value searches using the programs *BRUTE* and *XPLOR* gave a more convincing solution (shown in Fig. 1). The results of the *R*-value searches using *XPLOR* with a 1·5 Å grid step are given in Table 3. The *R* factor decreased from 0·51 to 0·47

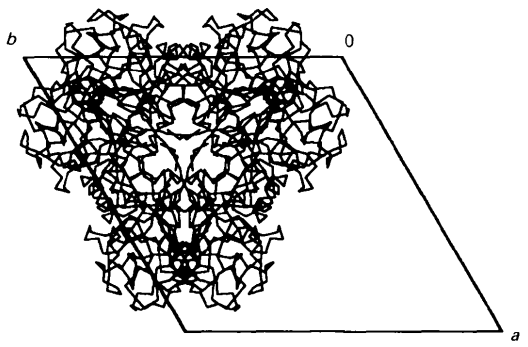

Fig. 1. The molecular replacement solution for ppΩ. $P2_1$ papain (Cα backbone) rotated and translated into the $P3_112$ cell of ppΩ. This projection is down the 3_1-screw axis, note the twofold axis in the plane ($c = 0$) is at 30° to the a axis.

Table 3. *R-value search and rigid-body refinement using XPLOR*

Space group	Resolution (Å)	Peak height	Map mean	Height/ σ	Solution
$P3_212$	5 2	35	26	6·9	0·86, 0·90, 0·44
$P3_112$*	5 2	38	26	9·5	0·06, 0·70, 0·10†
$P3_112$	6 3	31	13	10·3	0·06, 0·70, 0·10†
$P3_112$‡	10 4	27	10	7·3	0·72, 0·03, 0·10

* 42 cycles of rigid-body refinement resulted in an r.m.s. shift of 0·73 Å and a reduction in the *R* factor from 0·511 to 0·466.

† Correct solution.

‡ 49 cycles of rigid-body refinement did not reduce the *R* factor from 0·5.

Table 4. *Progress of the restrained least-squares refinement of ppΩ*

Number of cycles	Resolution (Å)	*R*	Comment
10	3·0 10·0	0·2982	Rebuild 1, refine overall MSDA
5	2·5–10·0	0·3222	Extend resolution
10	2·0–10·0	0·3206	Extend resolution
5	2·0 10·0	0·2644	Rebuild 2, 74 waters
15	2·0–10·0	0·2143	Refine isotropic MSDA's
12	2·0–10·0	0·1620	Rebuild 3, 160 waters
34	1·9 10·0	0·1515	Rebuild 4, 214 waters Occupancy of Hg: 0·485
40	1·8 10·0	0·1556	After rebuild 5, 150 waters Occupancy of Hg: 0·526
36	1·8 10·0	0·1549	After rebuild 6, 133 waters Occupancy of Hg: 0·547

during rigid-body refinement against 5·0 to 2·0 Å data. A $2F_o - F_c$ map with F_c based on this solution and calculated with data to 3·0 Å resolution confirmed that the molecule had been correctly located in the unit cell. At this stage the structure was modified from that of papain to that of ppΩ. Substituted (64) and inserted (4) residues were built into the density present in the $2F_o - F_c$ map (rebuild 1).

3.3. *Least-squares refinement*

The course of the least-squares refinement using the least-squares refinement program *RESTRAIN* (Driessen, Haneef, Harris, Howlin, Khan & Moss, 1989) is summarized in Table 4. Restrain is a least-squares refinement program minimizing a function containing terms involving structure amplitudes, phases and stereochemical or pseudo-energy restraints. The latter consist of restraints involving interatomic bonded distances, distances between two non-bonded atoms and planarity restraints. Isotropic U values were refined for all atoms, including the solvent atoms, in the later stages of the refinement. Six rounds of rebuilding and refinement reduced the *R* factor from 0·466 (5·0–2·0 Å data) to 0·1549 for all data in the range 10·0 to 1·8 Å. Each round of refinement following a rebuild involved five cycles of geometric regularization followed by cycles of structure-factor least-squares refinement. In the first round following rebuild 1 regularization was followed by least squares at 3·0, 2·5 and 2·0 Å using an

Table 5. *Details of least-squares refinement of ppΩ at 1·8 Å*

Number of reflections	18226
High resolution cut-off (Å)	1·80
Low resolution cut-off (Å)	10·00
Number of protein atoms	1641
Number of bound atoms (Hg^{2+})	1
Number of water molecules	133
Residual, R*	0·1549
Weighted residual, wR†	0·1761
Correlation coefficient, C‡	0·9497
Number of positional parameters	5325
Number of thermal parameters	1775
Number of overall scaling parameters	2
Number of occupancies refined (Hg^{2+})	1
Total number of parameters refined	7103

*$R = \Sigma||F_o| - G|F_c||/\Sigma|F_o|$.

†$wR = [\Sigma w(|F_o| - G|F_c|)^2/\Sigma w|F_o|^2]^{1/2}$.

‡$C = n\Sigma(|F_o||G||F_c|) - \Sigma|F_o|\Sigma|G||F_c|/\{[n\Sigma|F_o|^2 - (\Sigma|F_o|)^2][n\Sigma(G|F_c|)^2 - (\Sigma|G||F_c|)^2]\}^{1/2}$, where n is the number of amplitudes used, G is the scale applied to F_c, and constant weight, $w = 0{\cdot}1 \times 10^{-3}$, is given to all reflections.

Table 6. *Details of the final structure of ppΩ*

(*a*) Crystallographic R factor as a function of resolution

Resolution (Å)	R	No. of structure amplitudes
10·00–3·24	0·1461	3734
3·24–2·59	0·1339	3605
2·59–2·26	0·1472	3184
2·26–2·06	0·1583	2829
2·06–1·91	0·1869	2634
1·91–1·80	0·2410	2240
Total	0·1549	18226

(*b*) Statistical information on the geometry

	Number	R.m.s. deviation (Å) from ideal values
$D < 2{\cdot}12$	1626	0·025
$2{\cdot}12 < D < 2{\cdot}62$	2166	0·050
$D > 2{\cdot}62$	27	0·069
Planes		
Type 1 (peptide)	216	0·020
Type 2 (other)	65	0·011
Chiral centres*	114	0·028

*Chiral restraints were applied as distance restraints along the edges of chiral tetrahedra with $d_i < 2{\cdot}12$ Å.

overall MDSA (mean-square-displacement amplitude or **U** value). A $2F_o - F_c$ map showed substantial errors in the regions of Pro103 to Thr107 and Pro68 to Pro69; other minor errors were corrected and 74 water molecules were added. Following rebuild 2 isotropic MDSA's were refined for all atoms and further water molecules were added. Refinement of rebuild 3 using data from 10·0 to 2·0 Å gave an R factor to 0·1620. At this stage a $3F_o - 2F_c$ map was calculated to provide an even greater degree of correction in the electron density map. Pro197, Gly101, Gly102 and Pro103 were rebuilt substantially and other relatively minor adjustments made. An example of the substantial rebuilding involved in irregular regions is shown in Fig. 2. The chirality of Thr162, Thr214 and Ile104 was corrected and additional water molecules added (rebuild 4). This rebuild was refined to an R value of 0·1515 using data from 10·0 to 1·9 Å; this refinement included the occupancy of the mercury atom bound to the active-site cysteine. The last two rebuilds involved Ser169, Gly170, Gly171 and Ser136, and the solvent structure was critically evaluated and water molecules with high MSDA's and poor density deleted. The final model (rebuild 6, Table 5) contains 133 water molecules compared to the 214 included in rebuild 4. The crystallographic R factor as a function of resolution for the final structure is given in Table 6(*a*).* The stereochemistry of this final model is good as judged from the deviations from ideality (Table 6*b*).

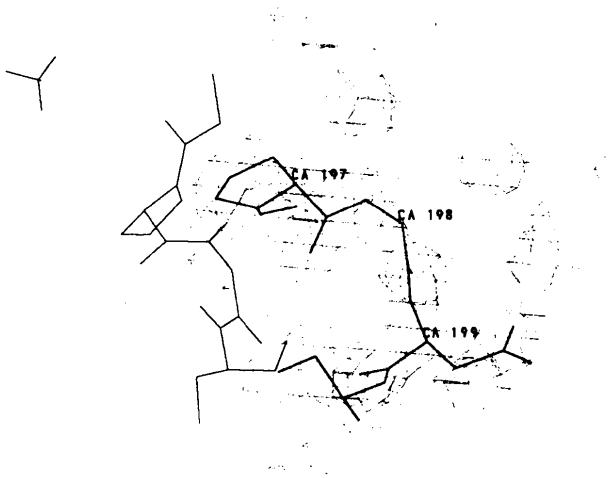

Fig. 2. Electron density map 6 contoured at 3·5% of the peak (mercury) density with the final fit to the electron density (R = 0·1549) and the rigid-body-refined model (R = 0·466) based on the structure of $P2_1$ papain shown. Note the significant changes during rebuilding and refinement.

The average MSDA's for main-chain atoms, side-chain atoms and all protein atoms are 0·134, 0·227 and 0·178 Å^2 respectively. Contacts within a distance of 3·2 Å between protein molecules in the crystal are given in Table 7. The NZ atom of Lys64 is 2·88 Å from the mercury atom of a symmetry-related molecule and this contact may be important in determining the arrangement of molecules and hence the space group. Intermolecular contacts less than 2·7 Å tend to have higher MSDA values associated with one or both of the atoms involved, and these are the result of errors in positioning flexible parts of the structure. One exception is the O of Gly93 to NH2 of Arg139 contact of 2·36 Å; both atoms have low

* 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 37044 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 7. *Intermolecular contacts in ppΩ crystals* (< 3·2 Å)

Atom 1	Atom 2	d (Å)	Symmetry operation*	U_1 (Å^2)	U_2 (Å^2)
4 *CG* Asn	170 *CA* Gly	3·13	4	0·434	0·549
4 *CG* Asn	170 C Gly	3·07	4	0·434	0·953
4 *ND2* Asn	170 *CA* Gly	2·50	4	0·270	0·549
4 *ND2* Asn	170 C Gly	3·05	4	0·270	0·953
55 O Asp	139 NH2 Arg	2·72	5	0·227	0·199
58 *CD* Arg	139 *CD* Arg	2·64	5	0·957	0·284
58 *CD* Arg	139 *NE* Arg	2·70	5	0·957	0·424
61 O His	142 *OE1* Gln	2·88	5	0·104	0·147
63 N Cys	142 *NE2* Gln	2·94	5	0·081	0·000
64 *CD* Lys	136 *OG* Ser	3·18	5	0·103	0·050
64 *NZ* Lys	217 *HG* Mer	2·88	5	0·238	0·101
92 O Gln	143 *CD2* Leu	3·14	5	0·193	0·196
93 O Gly	139 NH2 Arg	2·36	5	0·185	0·199
108 *OG* Ser	123 *OD1* Asn	3·15	6	0·246	0·533
108 O Ser	114 N Gln	2·96	6	0·131	0·130
110 N Val	112 O Arg	2·99	6	0·075	0·134
110 O Val	112 N Arg	2·78	6	0·138	0·014
123 *CG* Asn	216 *OXT* Asn	2·66	6	0·550	0·963
123 *OD1* Asn	216 N Asn	3·10	6	0·553	0·960
123 *OD1* Asn	216 *OXT* Asn	2·63	6	0·533	0·963
123 *ND2* Asn	216 *OXT* Asn	2·58	6	0·551	0·963
168 O Lys	170 N Gly	3·14	4	0·302	0·448
170 N Gly	174 OH Tyr	2·84	4	0·448	0·327

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

MSDA values (Table 7). Comparison of the water structure of ppΩ with that of papain shows that 26 water molecules occupy similar positions (within 0·5 Å) in the two structures. These 26 water molecules have an MSDA of 0·175 Å^2 compared to the average for the 133 water molecules of 0·370 Å^2.

4. Concluding remarks

Data collected using the SRS/FAST combination have been used to solve and refine the structure of ppΩ. These data extend to much greater resolution than could be measured using a conventional rotating-anode source and were of sufficient quality to enable the structure to be refined to an R factor of 0·1549 (10·0 to 1·8 Å data).

This work shows that ppΩ has a similar structure to that of papain (Fig. 3), but with an additional four-residue insertion (similar to that in actinidin) and with other differences in main-chain conformation resulting from proline substitutions. Like papain, however, ppΩ has a *cis* Pro at position 152. In the first of these proline substitutions, Pro102 (papain) is substituted by glycine (ppΩ) and the adjacent Tyr103 (papain) by a proline (ppΩ). This results in a significant difference in main-chain conformation (Fig. 3). The second region is around residue 197 which is proline in ppΩ and this substitution results in a substantially different main-chain conformation (Fig. 2). However, proline substitutions do not always result in a different main-chain conformation as illustrated by the substitution of proline (ppΩ) for Tyr69 (papain). This substitution is accepted without major rearrangement of the main chain. The active-site Cys25, His159 and adjacent Asp179 (175 in papain) are in identical positions in ppΩ and papain. The additional mercury present in ppΩ does not therefore affect the conformation of the protein. This mercury was refined to an occupancy of 0·547 indicating that a mercury atom is bound to approximately half of the molecules in the crystal. This is not surprising since no precautions were taken to ensure that the reactive Cys25 did not undergo oxidation and some degree of oxidation must have occurred in about half of the molecules, the other half binding mercury.

There are five examples of buried charge pairs (Arg192–Glu187, Arg195–Glu118, Asp6–Arg8,

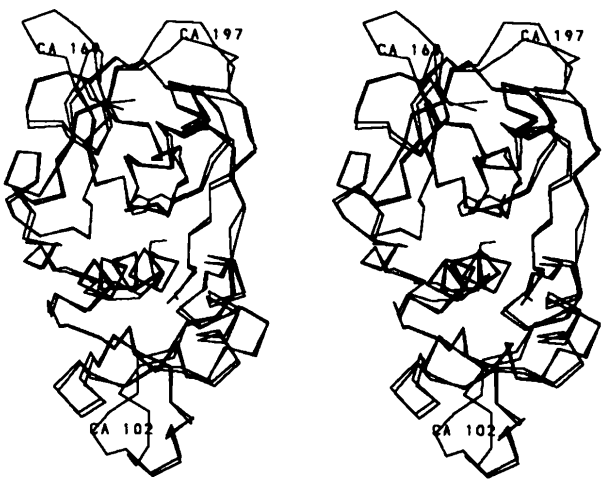

Fig. 3. Superposition of $P2_1$ papain and the final model of ppΩ ($C\alpha$ backbones). The active-site cysteine and histidine are shown and the regions of greatest difference (around 102, 169 and 197) are labelled.

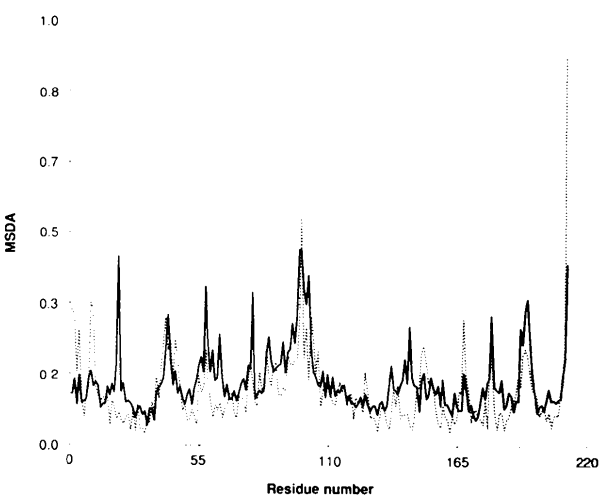

Fig. 4. Plot of average main-chain MSDA as a function of residue number. Papain is shown in full line and ppΩ in broken line. Note ppΩ has systematically lower MSDA values.

Asp158–Lys137 and His81–Glu52) and a charge network involving Glu35, Arg17, Glu50 and Arg83. In addition the following buried charge groups are adequately compensated for by non-charge group interactions or are partially solvent accessible; Asp55, Glu52, Glu57, Glu135, Lys178 and Lys215. The rings of Trp181, Tyr86 and Tyr207 have unusual χ_2 values, judged by comparison with those observed by Ponder & Richards (1987), owing to steric hindrance of the rings. Lys64 has an unusual conformation (χ_1) determined by its contacts with the symmetry-related molecule and Thr42 also has a somewhat unusual χ_1 owing to hydrogen bonding. Finally, it is interesting to note that ppΩ is a less-flexible molecule than papain (Fig. 4) despite a similar number of crystal contacts and this observation may be related to its higher stability than that of papain.

Program *PACK* was written by Dr J. Warwicker and the authors acknowledge his help in using this program.

References

BRÜNGER, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

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

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

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

DUBOIS, T., KLEINSCHNIDT, T., SCHNEK, A. G., LOOZE, Y. & BRAUNSTZEN, G. (1988). *Biol. Chem.* **369**, 741–754.

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

FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.

KABSCH, W. (1988). *J. Appl. Cryst.* **21**, 916–924.

KAMPHUIS, I., 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.* **B26**, 1854–1857.

MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.

PAPIZ, M. Z. & ANDREWS, S. J. (1987). *Computational Aspects of Protein Crystal Data Analysis*. In *Proceedings of the Daresbury Study Weekend*, 23–24 *January*, edited by J. R. HELLIWELL, P. A. MACHIN & M. Z. PAPIZ. Warrington: SERC Daresbury Laboratory.

PICKERSGILL, R. W., HARRIS, G. W. & GARMAN, E. (1991). *Acta Cryst.* In the press.

PICKERSGILL, R. W., SUMNER, I. G. & GOODENOUGH, P. W. (1990). *Eur. J. Biochem.* **190**, 443–444.

PONDER, J. W. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 775–791.

Acta Cryst. (1991). B**47**, 771–775

Studies on 1,3-Diaryltriazene Analogues of Berenil: Molecules with Potential GC Base-Pair Selectivity

BY ADRIAL R. WALTON, TERENCE C. JENKINS AND STEPHEN NEIDLE*

Cancer Research Campaign Biomolecular Structure Unit, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, England

(*Received* 17 *July* 1990; *accepted* 15 *April* 1991)

Abstract

1,3-Bis(4-acetylphenyl)triazene (II): $C_{16}H_{15}N_3O_2$, M_r = 281·32, monoclinic, $P2_1/c$, a = 14·002 (5), b = 12·359 (3), c = 8·457 (3) Å, β = 96·35 (2)°, V = 1454·5 Å^3, Z = 4, D_x = 1·287 Mg m^{-3}, Cu $K\alpha$, λ = 1·54178 Å, μ = 0·672 mm^{-1}, $F(000)$ = 592, T = 294 K, final R = 0·089 for 1890 unique observed reflections. 1,3-Bis{4-[2-(dimethylamino)ethoxycarbonyl]phenyl}triazene monohydrate (III): $C_{22}H_{29}N_5O_4.H_2O$, M_r = 445·52, triclinic, $P\bar{1}$, a = 9·500 (2), b = 11·753 (3), c = 13·328 (2) Å, α = 62·84 (1), β = 66·60 (2), γ = 77·58 (2)°, V = 1214·1 Å^3, Z = 2, D_x = 1·172 Mg m^{-3}, Cu $K\alpha$, λ = 1·54178 Å, μ = 0·640 mm^{-1}, $F(000)$ = 476, T = 294 K, final R = 0·063 for 1100 unique observed reflections. Both crystal structures have extended conformations for the 1,3-diaryltriazene groups, with a *cisoid* arrangement of phenyl rings. The terminal N—N bonds of each triazene are non-equivalent, and a hydrogen atom has been located in the N=N—NH moiety. Extensive molecular-orbital (MNDO) calculations on the model core 1,3-diphenyltriazene system have confirmed that this geometry is energetically favoured, and have revealed the shape of the energy surface for rotation about the N—NH bond in the triazene linkage.

* To whom correspondence should be addressed.

0108-7681/91/050771-05$03.00

Introduction

The recognition of specific nucleotide sequences in DNA by small molecules has recently been the subject of considerable study. Drugs such as netropsin