Acta Cryst. (1992). B48, 307–319

Structure and Molecular Model Refinement of *Rhizomucor miehei* Triacylglyceride Lipase: a Case Study of the Use of Simulated Annealing in Partial Model Refinement

BY ANDRZEJ M. BRZOZOWSKI,* ZYGMUNT S. DEREWENDA,† ELEANOR J. DODSON, GUY G. DODSON AND JOHAN P. TURKENBURG

Department of Chemistry, University of York, Heslington, York YO1 5DD, England

(*Received* 26 *June* 1991; *accepted* 9 *December* 1991)

Abstract

The method of crystallographic refinement with incorporated simulated annealing [Brunger, Kurijan & Karplus (1987). *Science*, **235**, 458–460] was used to refine a partial starting atomic model of *Rhizomucor miehei* triacylglyceride lipase derived by a combination of multiple isomorphous replacement and molecular replacement methods. Poly-Ala segments (amounting to almost 25% of the molecule) were introduced into the model whenever the original electron density map did not allow confident model building. A single cycle of the simulated annealing protocol resulted in a dramatic improvement of the quality of the electron density map. However, the progress of the refinement was erratic: substantial parts of the poly-Ala segments refined into the corresponding main-chain density, while others were unaffected by the procedure. The structure was refined further using 1.9 Å resolution data collected with a Xentronics (Siemens) area detector and employing a combination of simulated annealing and restrained crystallographic least-squares methods. The final crystallographic R factor is 0.129 [reflections within the 7.5–1.9 Å shell, for which $F > 2\sigma(F)$], with good stereochemistry (root-mean-square deviation from ideal bond lengths 0.015 Å, and from ideal planes 0.010 Å).

1. Introduction

Crystallographic least-squares refinement is a routine step in macromolecular structure analysis. The earlier difficulties associated with a low observations-to-parameters ratio are now normally resolved by the introduction of restraints.

In spite of this progress, traditional crystallographic least-squares refinement is subject to one particular limitation, *i.e.* the radius of convergence (the largest allowed displacement of a model atom from its target position, which can be corrected by the refinement procedure). In theory, the radius of convergence is given by approximately one-third of the maximum resolution; in practice an atom closer to its neighbour's target position than to its own is not likely to refine correctly. This restricts the radius of convergence to *ca* 0.7–0.8 Å or half the average bond length. Outside, or close, to this limit the least-squares procedures can converge at a local minimum. If the phase angles calculated from a current atomic model are sufficiently accurate, manual rebuilding can be carried out against an electron density map, usually calculated with the coefficients $mF_{obs} - nF_{calc}$. This procedure of escape from local minima has been routinely used in the course of refinement.

Recently, a new method was introduced to overcome energy barriers in search of the global minimum. It relies on the simulated-annealing (SA) approach (Kirkpatrick, Gellat & Vecchi, 1983) in which it is possible to overcome energy barriers by temporarily increasing the minimized target function. Molecular dynamics (MD) simulations (Karplus & McCammon, 1983) are used to introduce a 'temperature' factor which determines the amount of kinetic energy available to the system. The method has been tested using a number of known structures (*e.g.* Brunger, Karplus & Petsko, 1989; Fujinaga, Gros & van Gunsteren, 1989) and two independent software packages are available: *GROMOS* (van Gunsteren & Berendsen, 1987) and *XPLOR* (Brunger, 1988). This approach is much more efficient than standard least-squares algorithms, and has been employed in the refinement of a number of proteins [for an extensive overview of applications see Goodfellow, Henrick & Hubbard (1989)]. The reported shifts depend largely on the accuracy of the initial model, but movements in excess of 5 Å have been observed on a number of occasions.

In this paper we report an application of SA refinement to a partial model of *Rhizomucor* (former

* On leave from the Department of Crystallography, Institute of Chemistry, University of Łódź, 91 416 Łódź, Poland.

† Author to whom correspondence should be addressed. Present address: MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, 474 Medical Sciences Building, Edmonton, AB, Canada T6G 2H7.

0108-7681/92/030307-13$06.00

Mucor) *miehei* lipase [obtained by a combination of molecular replacement (MR) and multiple isomorphous replacement (MIR) methods], in which fairly large sections of the polypeptide chain were missing from the original model. These were modelled as poly-Ala, and allowed to find proper conformation during the SA run. To our knowledge, this is the first use of poly-Ala sequences with SA refinement.

2. *R. miehei* triacylglyceride lipase

The overall three-dimensional structure of this enzyme and a brief description of the crystallographic analysis has been published (Brady *et al.*, 1989; Fig. 1). The enzyme is a single polypeptide chain protein with 269 residues and a calculated molecular weight of 29472. Isolation and purification methods were given by Huge-Jensen, Gailuzzo Rubano & Jensen (1987) and Boel, Huge-Jensen, Christiansen, Lars & Fil (1988). The protein (15–16 mg ml^{-1}) dissolved in Tris.HCl buffer, pH 8.05, was crystallized using the hanging-drop technique with high concentrations of phosphate buffer (55–75% *v*/*v* of the saturated solution). The crystals were orthorhombic (space group $P2_12_12_1$, $a = 71.6$, $b = 75.0$, $c = 55.0$ Å), with a specific volume (Matthews, 1966) of 2.5 Å^3 dalton^{-1}. The estimated solvent content for these crystals is close to 50%.

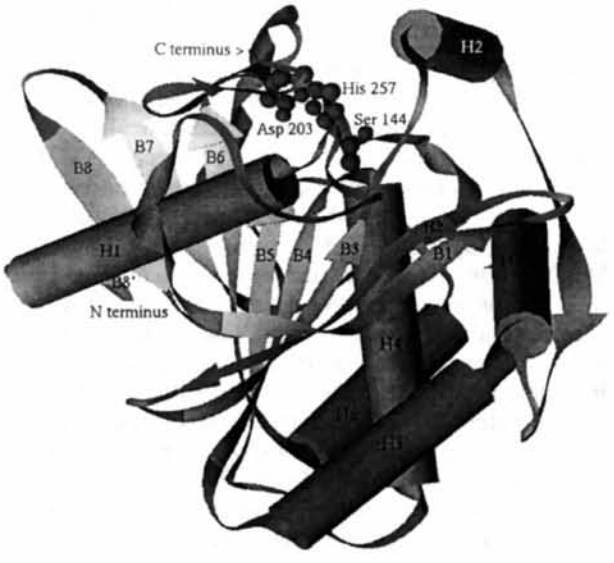

Fig. 1. Schematic representation of the molecule of *Rhizomucor miehei* triglyceride lipase. The secondary structure has been identified as follows (Kabsch & Sander, 1983): β-strands: $\beta 8'$, residues 6–8; $\beta 1$, 50–56; $\beta 2$, 63–69; $\beta 3$, 74–79; $\beta 4$, 137–143; $\beta 5$, 169–174; $\beta 6$, 195–200; $\beta 7$, 220–224; $\beta 8$, 231–235; α-helices: *H*1, residues 11–27; *H*2, 85–89; *H*3, 109–132; *H*4, 146–159; *H*5, 182–191; *H*6, 254–258. In addition, the catalytic triad is shown in full, and both the amino and the carboxyl termini are marked for clarity.

X-ray data were collected to 1.9 Å resolution using a Xentronics (Siemens) imaging proportional counter (area detector) mounted on a conventional X-ray source ($\lambda = 1.54178$ Å; graphite monochromator). The protocol used was that described by Derewenda & Helliwell (1989), with the exception that a fine focus tube was used. The chamber was placed 10 cm from the sample, and the generator was operated at 50 kV and 25 mA. Integration and data reduction were carried out using the VAX VMS version of the *XENGEN* software (Howard, Gilliland, Finzel, Poulos, Ohlendorf & Salemme, 1987). The details of the crystallographic data relating to the native enzyme crystal and to the heavy-atom derivatives are given in Table 1.

3. Multiple isomorphous replacement phasing

Among many heavy-atom derivatives tried, three were suitable for MIR phasing. These were: Hg (5 m*M* $HgCl_2$, soaking time 3 days), Pt (2.5 m*M* K_2PtCl_4, 2 days), and I (2 m*M* *N*-iodosuccinimide anhydride, 12 h). Two different Pt-derivative data sets collected from crystals with different metal occupancies (see Table 2) were used. The major sites were identified by difference Patterson methods, and subsequently refined using the F_{hle} method (Matthews, 1966; Dodson & Vijiyan, 1971). Minor sites were identified by difference Fourier methods based on phases calculated from major sites, and were subsequently introduced into the F_{hle} refinement. The procedure was cycled in an iterative manner until no further sites were found in any of the derivatives. The important crystallographic data describing the derivatives are given in Table 2. The phasing power is illustrated in Fig. 2.

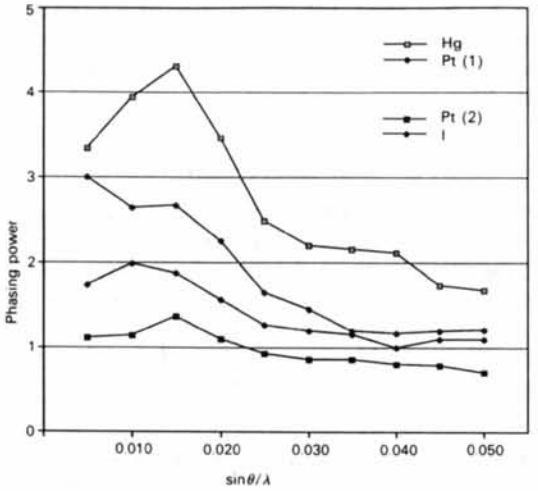

Fig. 2. Phasing power ($\langle F_h \rangle / \langle E \rangle$) calculated against $\sin\theta/\lambda$ for all heavy-atom derivatives used in this study (see text for further details). *E* is the lack of closure.

Table 1. *Crystallographic data details*

Resolution limit (Å)	No. of reflections Possible	Collected (%)	Average redundancy	Singly measured reflections (%)	$\langle I/\sigma(I)\rangle$	Reflections $>2\sigma$ (%)	R_{merge} (%)
Native data							
<3.45	4222	99.8	10.9	0.4	123.6	96.9	6.68
<2.74	4026	97.2	5.8	2.6	41.5	94.6	9.68
<2.39	3999	94.8	4.3	3.5	18.4	88.6	13.04
<2.17	3983	92.1	3.9	8.5	11.8	81.1	15.84
<2.02	3940	86.7	3.3	15.4	6.6	70.2	21.18
<1.90	3931	55.0	2.6	67.1	3.0	49.7	32.65
Total	24101	87.8	5.5	12.4	39.0	82.8	8.07
Hg derivative							
<5.79	949	91.5	5.2	6.5	89.1	93.4	5.34
<4.59	883	95.6	5.6	4.6	60.2	94.3	6.42
<4.01	874	97.6	5.6	3.9	47.1	93.8	6.90
<3.64	853	97.5	5.5	4.2	31.8	92.7	8.94
<3.38	852	91.4	4.9	8.8	21.8	88.7	11.52
<3.18	844	63.0	4.2	14.9	13.2	81.2	15.02
Total	5255	89.6	5.2	6.6	46.5	91.4	7.82
Pt derivative (1)							
<5.67	1012	87.1	6.6	5.1	84.0	95.2	5.93
<4.50	927	90.1	7.2	3.9	59.5	95.8	7.98
<3.93	927	92.5	7.2	3.6	45.7	93.5	9.57
<3.57	902	92.0	7.0	3.0	34.5	93.5	11.67
<3.31	913	77.0	6.4	5.0	26.1	90.8	14.06
<3.12	900	48.6	4.0	21.7	12.7	78.3	18.73
Total	5581	81.5	6.6	5.8	47.4	92.4	9.36
Pt derivative (2)							
<5.68	1015	71.8	3.5	21.1	83.2	96.0	4.65
<4.51	929	78.1	3.5	23.0	63.3	96.1	6.24
<3.94	929	79.2	3.4	23.0	51.5	95.8	7.08
<3.58	911	78.6	3.3	23.2	40.9	96.1	7.83
<3.32	917	69.1	3.0	30.3	31.2	94.3	8.79
<3.13	891	36.7	2.3	39.5	17.7	88.1	11.40
Total	5592	69.2	3.3	25.3	51.5	95.1	6.79
I derivative							
<5.66	1011	96.5	3.5	3.2	71.2	95.6	6.09
<4.49	928	98.6	3.8	3.2	57.4	96.7	7.28
<3.92	923	99.1	3.8	3.5	47.7	94.5	7.76
<3.56	909	97.7	3.7	5.1	37.5	93.9	9.05
<3.31	908	83.1	3.4	11.1	28.7	91.3	9.81
<3.11	898	46.0	2.6	23.9	17.2	87.9	13.20
Total	5577	87.2	3.6	6.6	46.8	94.0	7.95

Table 2. *Crystallographic details of the heavy-atom derivatives and refinement of the heavy-atom sites*

The heavy-atom parameters were refined using the F_{hle} method (Matthews, 1966); only positional and occupancy parameters were allowed to refine; temperature factors (B) were assigned and kept constant. The Pt sites marked '?' refer to a cavity between symmetry-related molecules, in which the stereochemistry of the coordination of the metal could not be identified. R_{iso} is defined as: $R_{iso} = \Sigma|F_{derivative} - F_{native}|/\Sigma F_{native}$. R_{Cullis} (R_C) is defined as: $R_C = \Sigma|F_{h(obs)} - F_{h(calc)}|/F_{h(obs)}$, where the summation extends over all reflections. g_D is the slope of $F_{h(calc)}$ against mod(Δ_{iso}). The mean figure of merit = 0.74.

Substitution site	No. of site	Site coordinates x	y	z	Occupancy	B	R_{iso}	R_C	g_D	No. of data
I derivative										
153C	9	−0.004	0.248	0.442	1.05	25	0.225	0.485	0.28	4049 > 5σ
55*W*		−0.160	0.183	0.416	0.90	25				
172*Y*		0.052	0.483	0.395	1.00	25				
60*Y*		−0.154	0.164	0.631	1.07	25				
133*Y*		0.006	0.239	0.130	0.61	25				
20*Y*		0.026	0.460	0.315	1.39	25				
187*Y*		0.191	0.253	0.580	0.71	25				
60*Y*		−0.160	0.230	0.700	1.00	25				
268C		0.265	−0.021	0.145	0.09	25				
Pt derivative (1)										
268C	3	0.268	0.954	0.139	1.20	25	0.193	0.580	0.1	3380 > 5σ
42H		0.304	0.909	0.120	0.60	25				
?		0.206	0.026	0.139	0.52	25				
Pt derivative (2)										
268C	3	0.270	0.951	0.144	0.68	25	0.117	0.700	0.06	3203 > 5σ
42H		0.322	0.909	0.128	0.20	25				
?		0.180	0.031	0.155	0.21	25				
Hg derivative										
153C	3	0.003	0.245	0.451	1.48	25	0.205	0.495	0.30	3618 > 2σ
17H		−0.170	0.681	0.087	0.58	25				
88*W*		−0.210	0.669	0.178	0.41	25				

4. Model building

Chronologically, the first calculated MIR electron density map was phased on two derivatives only (Hg and Pt). This map (calculated using the coefficients F_{obs} and α_{MIR}, with weights based on the figures of merit m) was only partly interpretable. However, *R. miehei* triacylglyceride lipase was not the first fungal lipase to be solved in our laboratory. We already had the knowledge of the overall fold for a related enzyme purified from *Humicola lanuginosa* (this is an unpublished study carried out in collaboration with Novo-Nordisk Research Laboratories in Copenhagen, Denmark). Much of the expected central β-sheet and some helical segments could be correlated with the *H. lanuginosa* model. In fact, a molecular model based on the *H. lanuginosa* structure alone, edited in regions of particularly poor sequence homology, was used in parallel MR calculations; one of the solutions displayed against the MIR electron density map proved to be generally correct (EJD, unpublished results). In order to improve the quality of the original MIR map, two other techniques were employed at this stage: solvent flattening (at 45% solvent level) (Wang, 1985), and phase combination, using the MIR and partial model phases. All three maps were displayed using *FRODO* software (Jones, 1978) on a PS 330 Evans & Sutherland computer-graphics system, and various manual modifications were introduced into the partial model. Restrained crystallographic least-squares refinement was attempted on the partial model, followed by phase combination. All these calculations resulted in little improvement with respect to the interpretability of the electron density map, and are not documented in this paper. At this stage a third suitable derivative (I) was found and used to obtain better MIR phases. These were immediately combined with the current model phases (see §7 for a discussion of the possible reasons why these original efforts failed to produce an easily refinable model).

At this stage the model had several substantial gaps and a number of side chains missing. Most importantly, residues 1–13, 30–51, 71–73, 159–169, 192–194, 205–212, 245–256 and 269 (70 out of 269 amino acids) could not be modelled satisfactorily, and were omitted. The resulting model was by far inferior to any of those used as a starting point in other applications of the SA refinement. Table 3 gives the exact details of the completeness of the atomic model at various stages of refinement.

5. Refinement of the atomic model

In spite of considerable efforts, no further improvement was introduced by repeated manual rebuilding and partial model refinement. It was decided to subject the model to SA refinement.

Table 3. *Completeness of the atomic model versus refinement stage*

Refinement stage	No. of atoms		Degree of completeness		
	Protein	Solvent	Protein	Solvent	All
Initial model	1773	0	0.85	0.00	0.76
First SA	1773	0	0.85	0.00	0.76
KH stage 1	1885	0	0.91	0.00	0.81
Second SA	1885	0	0.91	0.00	0.81
KH stage 2	1976	0	0.95	0.00	0.84
KH stage 3	1981	0	0.95	0.00	0.85
KH stage 4	2009	0	0.96	0.00	0.86
KH stage 5	2056	128	0.99	0.50	0.93
KH stage 6	2055	138	0.99	0.54	0.94
KH stage 7	2047	202	0.98	0.79	0.96
KH stage 8	2054	255	0.99	1.00	0.99

It was quite clear that the numerous gaps in the polypeptide chain at this stage could present serious difficulties during the simulation. We decided, therefore, to introduce poly-Ala segments for residues 30–51, 159–169, 192–194, 205–212 and 245–256, modelled into the MIR density as well as possible. In addition the terminal residues 1–13 and 269 were not modelled, and as it was found later, a gap between residues 71 and 73 was introduced by mistake. This model served as the starting point for the SA stage of refinement. Further stages are documented in Fig. 3 which monitors the decrease in the conventional crystallographic R factor, against the refinement stages. Table 3 shows the completeness of the model at various stages of the refinement, and the progress of the identification and inclusion of water molecules.

The details of the SA protocol used are shown in Fig. 4. This first simulation was followed by an examination of the resulting $2F_{obs} - F_{calc}$ electron density map and four subsequent cycles of standard restrained Konnert–Hendrickson (KH; Hendrickson,

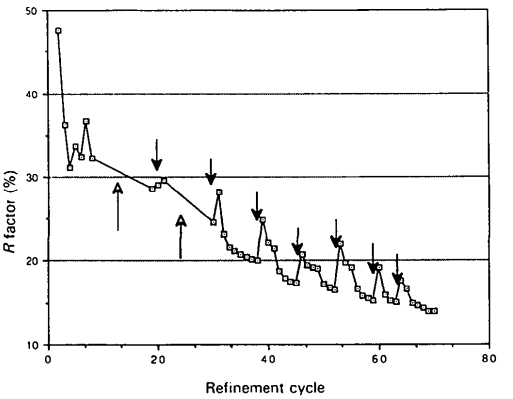

Fig. 3. The progress of the refinement shown as a plot of the conventional R factor against the refinement cycle. The two empty arrows indicate stages 1 and 2 of the KH refinement (see Table 3), each comprising of ten cycles which are not shown individually. The full arrows indicate modelling sessions prior to stages 2–8 of KH refinement.

1985) least-squares refinement with some additional manual interventions (in particular the addition of side chains into the stretches modelled originally as poly-Ala). A second simulation was carried out, followed by ten cycles of the KH procedure, and a manual examination of the electron density map, displayed using the *FRODO* software. The manually adjusted model was again refined for ten cycles. This model was subsequently found to contain only two major errors; the clarity of the electron density map now allowed an unambiguous tracing of the chain, thus producing a model practically completely within the radius of convergence of the KH method. Subsequent refinement and manual adjustment (consisting largely of minor charges and solvent addition) were repeated several times (see Fig 3).

The final adjustments to the model (not shown in Fig. 3), were carried out using *PROLSQ* with *FFT*, with both $(F_{obs} - F_{calc})$ and $(2F_{obs} - F_{calc})$ maps calculated and displayed several times. Water structure, the orientation of the Asn and Gln side chains, and a small number of surface side chains with bad electron density were the only areas of concern. Several rounds of refinement usually followed each such inspection, with different restraints in an attempt to find an ideal combination which would yield the lowest R factor and the best possible geometrical parameters. The refinement was finally terminated when no unaccounted significant difference density could be seen, and no unusual and/or unverified stereochemical parameters were present in the final output.

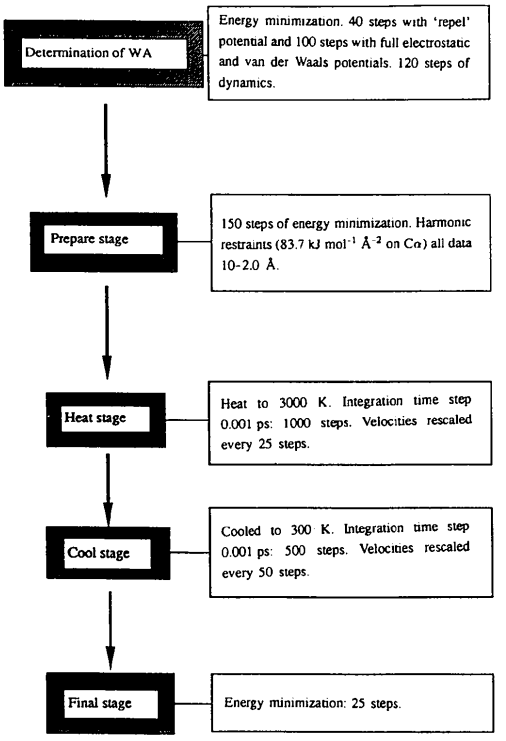

Fig. 4. A flow chart illustrating the details of the calculations within each SA refinement round.

6. Results: the accuracy of the model

The final atomic model yields an R factor of 0.129 in the 7.5–1.9 Å resolution range (an R factor of 0.153 was obtained for all available data). The lower cutoff was imposed owing to severe disagreement between low-resolution observed and calculated structure factors, as the latter take no account of bulk solvent flattening. Only observed data $[F > 2\sigma(F)]$ were used in the refinement, and all terms for which (F_{obs}/F_{calc}) or $(F_{calc}/F_{obs}) > 2.5$ were excluded from refinement and R-factor calculation. The latter condition was used to eliminate the 'rogue' observations, which are much more likely to occur among the area detector data than diffractometer data. 18 960 reflections (97.7%) out of a total of 19 404 reflections with non-negative intensities were used. The stereochemical parameters of the final atomic model are given in Table 4. They correspond well to the expected values. Fig. 5 gives the Luzzati (1952) plot which shows that the mean error in the atomic coordinates is close to 0.15 Å. The Ramachandran plot (Ramachandran & Sasisekharan, 1968) for the final structure is shown in Fig. 6. Only three non-Gly residues are outside the energetically favoured regions. They are: Cys 244, Ser 144 and Glu 201. We have analysed each of these amino acids against a difference $(F_{obs} - F_{calc})$ electron density map (Fig. 7). We found no indication of any residual errors in the atomic coordinates; the isotropic atomic temperature (B) factors were low, indicating that the atomic positions are well defined and that the refinement has converged. The strained conformation of Ser 144, which is a part of the catalytic triad and the lipase consensus sequence Gly-*X*-Ser-*X*-Gly, is discussed extensively elsewhere (Derewenda & Derewenda, 1992). Here we only wish to report the observation and emphasize the need for careful refinement and visual inspection of unusual features, as opposed to the brute force approach with rigid restraints enforcing the expected stereochemistry.

During the course of the refinement it was found that additional side-chain density was present for residue 156, listed as Gly by Boel *et al.* (1988) (codon GGT in cDNA). Asx could easily be built into that density, and the analysis of the hydrogen-bonding network and charge distribution among adjacent residues (see Fig. 8) strongly suggested the presence of Asp. In fact, Asp can be coded by GAT and it is easy to see that a cDNA sequence error was made with regard to only one purine base.

Table 4. *Stereochemical parameters for the refined atomic model*

	Standard deviation	σ
Bond distances (1–2) (Å)	0.015	0.010
Angle distances (1–3) (Å)	0.079	0.040
Distances (1–4) (Å)	0.097	0.060
Planarity (Å)	0.010	0.010
Chiral volumes (Å^3)	0.216	0.100
Non-bonded contacts (Å)		
Single-torsion contact	0.210	0.500
Multiple-torsion contact	0.303	0.500
Conformation torsion angles (°)		
Planar (0, 180°)	2.30	3.000
Staggered (±60, 180°)	20.43	90.000
Orthonormal (±90°)	33.38	90.000
Isotropic temperature factors (Å^2)		
Main-chain bond	5.23	1.000
Main-chain angle	5.88	1.500
Side-chain bond	9.37	1.500
Side-chain angle	11.75	2.000

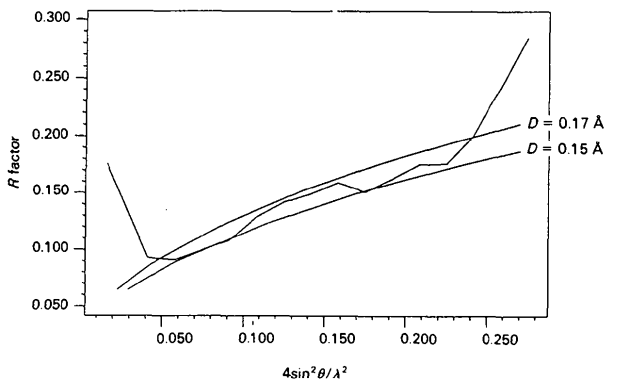

Fig. 5. The Luzzati (1952) plot showing the experimental curve against the theoretical curves calculated for two values of expected mean coordinate error (0.15 and 0.17 Å).

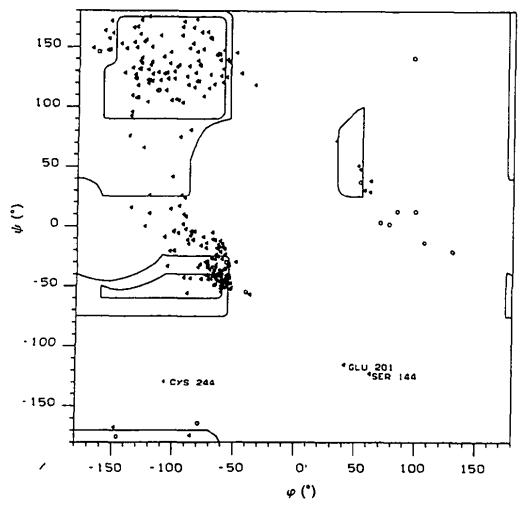

Fig. 6. The Ramachandran plot for the refined molecular model (Gly residues shown as open circles); the allowed regions of conformational space are shown, and the three non-Gly residues which occur outside these regions are labelled.

7. Discussion

Once a satisfactory atomic model is obtained, it can be instructive to look back at the early stages of model building and refinement. In our case we were primarily interested in the progress achieved by the early SA refinement of the partial model. We used the final calculated set of phase angles (α_{calc}) to

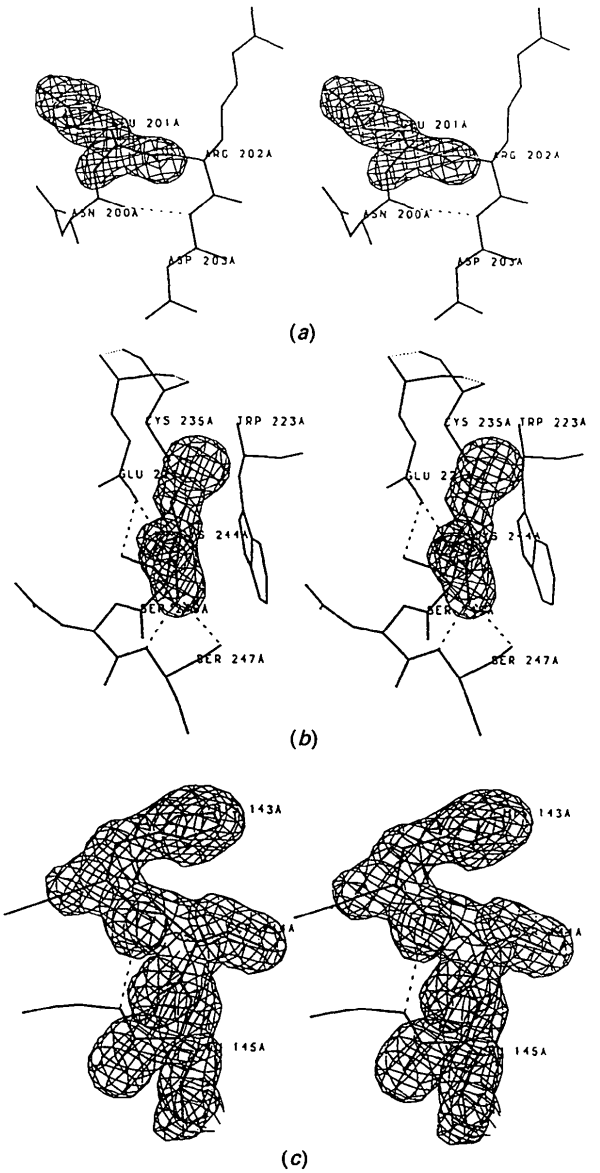

Fig. 7. The final difference ($F_{obs} - F_{calc}$) electron density for the three non-Gly residues which occur outside the allowed Ramachandran conformational space (see Fig. 6): (*a*) Glu 201, (*b*) Cys 244, (*c*) the tripeptide His-Ser-Leu containing the active Ser 144 residue; each difference map (calculated using the F_{calc} and α_{calc} obtained from a model from which the residue illustrated was omitted) was contoured arbitrarily at a single level; hydrogen bonds are denoted by broken lines.

monitor the accuracy of the initial refinement and the clarity of the electron density maps. In that respect we have also investigated the effect of phase combination, the causes of the original difficulties with the interpretation of the MIR map, and the ways in which these techniques could have been used more effectively.

7.1. *First SA simulation and the refinement of the poly-Ala segments*

Fig. 9 illustrates the distribution of error in the initial atomic model; it shows the r.m.s. difference (averaged per residue) between the initial and final model. The first cycle of the SA refinement resulted in a significant reduction in the R factor (from 0.48 to 0.36). Fig. 10 shows the reduction in error obtained after the first SA simulation (for main-chain atoms only). It is clear that a general improvement was achieved, with some areas undergoing particularly large changes. Three models (the initial input to *XPLOR*, the output from the first simulation and the final model) were displayed and analysed using *FRODO*.

7.1.1. *Refinement of the poly-Ala segments.* (*a*) Residues 30–51. This fragment constitutes the longest of the poly-Ala segments, and also a part of the structure least homologous to the *Humicola* enzyme. The MIR electron density was poor and the poly-Ala chain was fitted with considerable uncertainty. Fig. 11(*a*) compares this original model with both the SA output and the final atomic model in that region. Considerable improvement was achieved at both the N and C termini of the poly-Ala stretch, with residues 30–33 and 47–51 refining satisfactorily, *i.e.* with model atoms placed into the corresponding electron density. The latter part was modelled relatively well, and the largest shift observed during the SA refinement included a flip of peptide 50. In contrast, residues 30–33 were modelled badly, and the SA refinement performed very well reducing the error on average by *ca* 3 Å (refer to Fig. 10 for the evaluation of shifts). This progress was interrupted around residue 34, which happens to be a *cis*-Pro, while as an alanine it was strongly restrained to a *trans* configuration.

An obvious gross error was introduced when the main chain of the original model was built across the disulfide Cys 40–Cys 43. Owing to strong density of the disulfide bond, SA failed to correct this mistake. However, it is notable how the individual atoms were refined in an effort to account for the existing density. Fig. 11(*b*) shows in detail the region around the

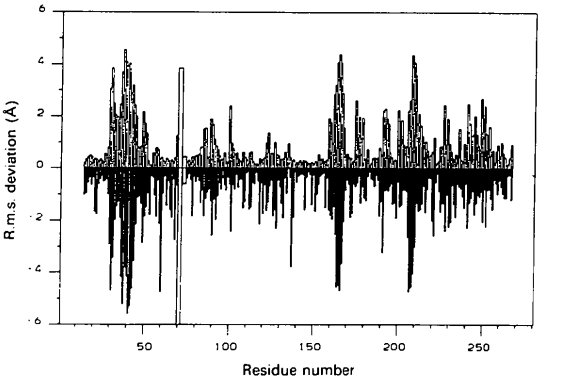

Fig. 9. Root-mean-square deviations in the initial model as compared to the final refined atomic model; the top part refers to main-chain atoms, the bottom part shows the side chains. The empty bars after residue 70 result from there being a gap in the model.

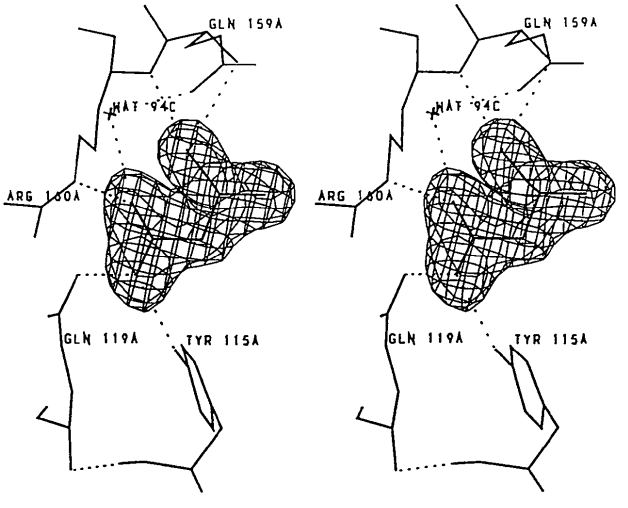

Fig. 8. The final difference ($F_{obs} - F_{calc}$) electron density for Asp 156, which was previously reported as Gly (Boel *et al.*, 1988). Other details are as described for Fig. 7.

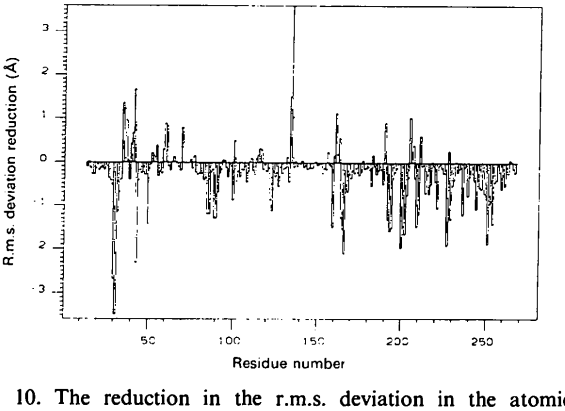

Fig. 10. The reduction in the r.m.s. deviation in the atomic coordinates (averaged values per residue for the main-chain atoms only) following the first SA simulation. These values have been obtained by subtracting the mean r.m.s. deviation in the original model from that in the one output from the SA refinement: a negative value represents improvement, and a positive value indicates an increased error.

disulfide (with all the side-chain atoms) and the way in which SA compensated for the errors.

The described error forced a translation in the density of the model polypeptide chain, and as a result residues 37–39 are out of step, *i.e.* the main-chain atoms have been refined into the electron densities corresponding to analogous atoms of the residue further down the sequence. The availability of an extra residue created by this process meant that a bulge was introduced between residues 34 and 37 making any further progress impossible.

(*b*) Residues 159–169. While the original MIR density in this region was poor, the knowledge of the *Humicola* enzyme allowed an approximate positioning of the 11-residue-long poly-Ala insertion. Fig. 11(*c*) shows that the model loop Glu 161–Ser 165 was built too short, and residues 163 to 166 are out of step. In contrast, the loop Ser 165–Asn

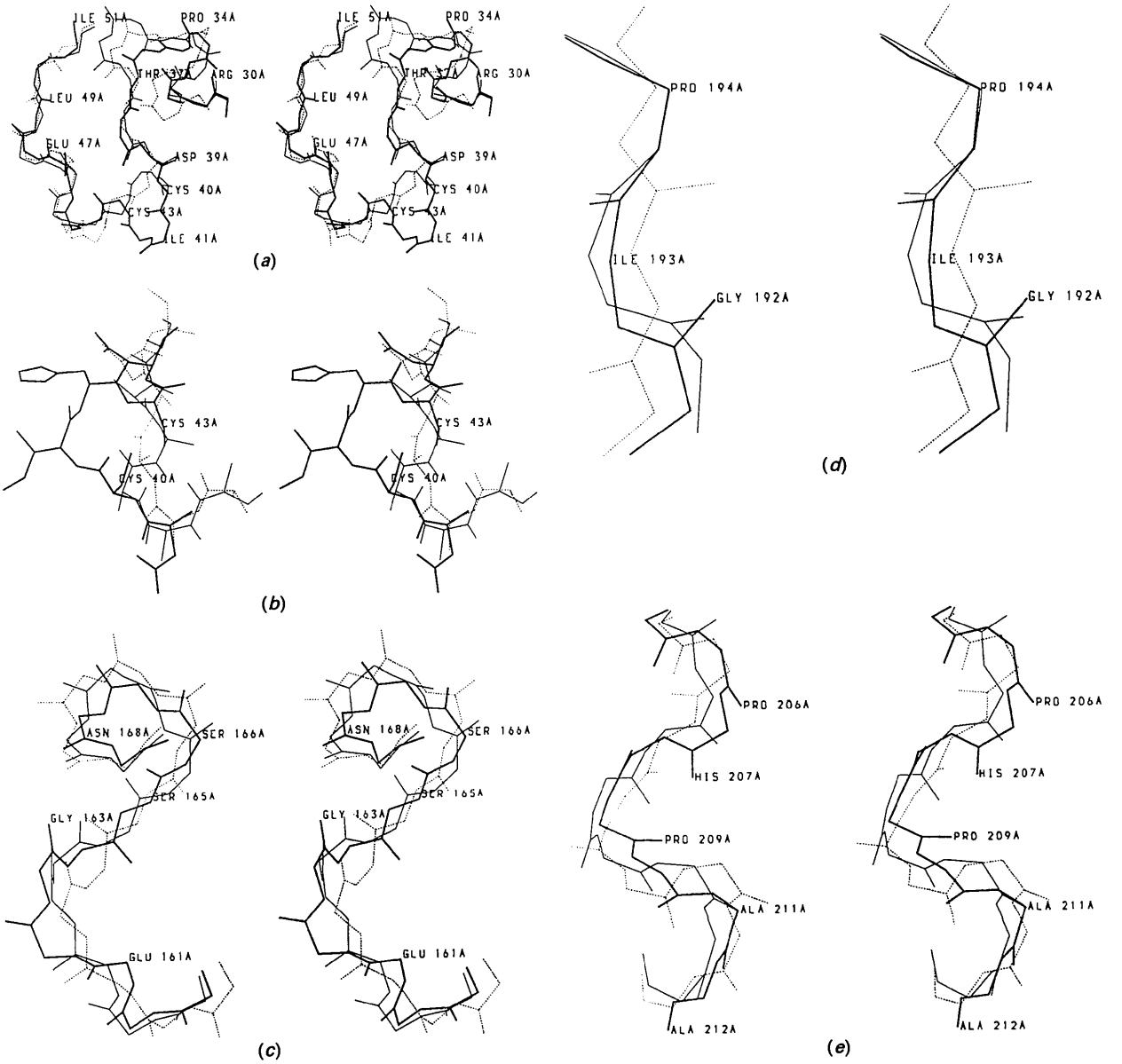

Fig. 11. The progress of the automated SA refinement of the five poly-Ala segments. The dashed lines represent the original poly-Ala stretch, the thin solid lines show the output of *XPLOR*, whereas the thick solid lines show the final structure of the polypeptide main-chain atoms. The chain identifier *A* indicates a single-chain protein and has no further meaning. The same conventions are used throughout. Other details are given in the text. (*a*) Residues 34–51. (*b*) A fragment of the previous poly-Ala stretch illustrating the refinement of main-chain atoms into the density of the disulfide bridge. All atoms of the final model are shown in the figure. (*c*) Residues 159–169. (*d*) Residues 192–194. (*e*) Residues 205–212.

168 had an extra residue thus creating a bulge. On first glance no substantial improvement was achieved by the use of SA. However, it is easy to see from Fig. 11(*c*) that the polypeptide chain was significantly shifted along its length thus bringing the amino acids much closer to their respective positions. It is possible that a different SA protocol (higher temperature, longer heating stage, slow cooling) could have been more effective.

(*c*) Residues 192–194. This was a short stretch of uncertain conformation. The automatic SA refinement was very successful in reorienting all three of the peptide links (Fig. 11*d*).

(*d*) Residues 205–212. This case was similar to (*b*) as the general fold in this region was identifiable, though the detail was obscure. Fig. 11(*e*) shows the progress of the SA refinement in this region. It is clear that only limited improvement was achieved because two of the residues in this region are *cis*-prolines (206 and 209). In a way identical to that described above for Pro 34, the restrictions imposed on Ala have prevented proper progress of the refinement. In order to impose the *trans* configuration on Pro 206 the model peptide 205 was flipped from the originally correctly modelled structure to an incorrect one.

(*e*) Residues 244–255. This is a polypeptide stretch, close to the C terminus of the molecule, which was not readily interpretable from the first maps. The refinement of this region was particularly successful. The two larger errors are the peptides of residue 250 (yet again a *cis*-Pro!) and that of residue 246.

It can be concluded that the refinement of the partial model was generally successful in spite of the difficulties encountered with *cis*-prolines or whenever the model poly-Ala stretch refined 'out of step'.

7.1.2. *Other parts of the model.* The remainder of the model molecule behaved in a complex way during the simulation, although a clear overall improvement has been achieved as can be seen from Fig. 10. In some cases we observed complete refinement of both main- and side-chain atoms. Perhaps the most spectacular example in this category is residues 200–204 (Fig. 12). The most common result was that of an overall improvement (expressed in terms of the reduction of the mean r.m.s. deviation) but incomplete, with the final model still outside or barely inside the radius of convergence of a classical least-squares refinement. Typical examples here are fragments 225–232 and 240–243 (Fig. 13*a*), both originally modelled incorrectly. In a number of cases we observed no improvement. The incorrectly built and unimproved peptides of residues Gly 175, Pro 177 serve as examples of this category. In a few limited cases an increase in the overall error, either for main- or side-chain atoms, or all, was observed. The particular examples are: Tyr 60, with its side chain refined into the density of Glu 117, and Pro 101, badly distorted by the adjacent unused electron density – later identified as that of originally excluded Phe 251 from a molecule related by crystallographic symmetry (Fig. 14). SA refinement has been shown to be particularly sensitive to unused electron density. Weis, Brunger, Shekel & Wiley

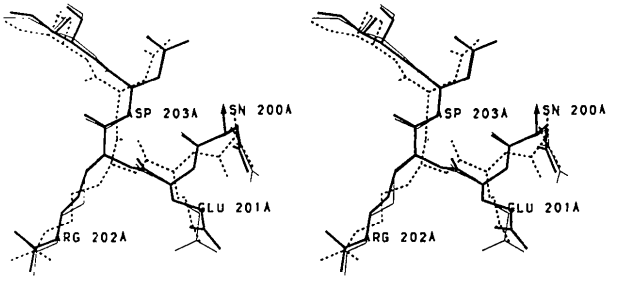

Fig. 12. A region of the molecular model refined succesfully following a single SA simulation: residues 201–204; note complete reorientation of two peptides (other details as in the legend to Fig. 11).

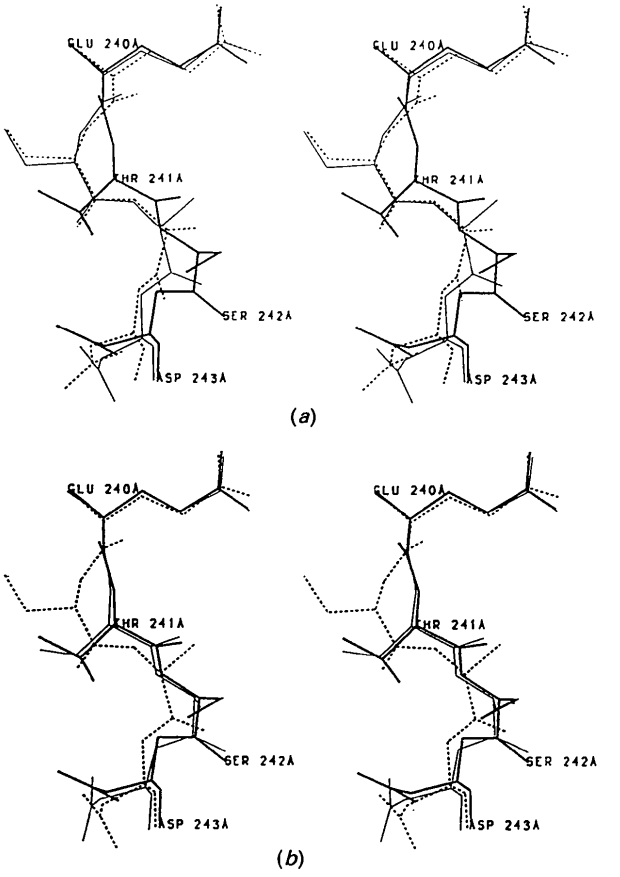

Fig. 13. Residues 240–243: (*a*) following the first SA simulation and (*b*) the second SA refinement cycle. The imput model in (*b*) is the output one in (*a*). Other details as in the legend to Fig. 11 and in the text.

(1990) report one such example which occurred during their refinement of the influenza virus haemagluttinin, when a tryptophan side chain was erroneously placed into the N-linked carbohydrate density.

The first SA refinement cycle resulted in a dramatic improvement in accuracy of the phasing. Subsequent electron density maps (calculated with coefficients $2F_{obs} - F_{calc}$) resolved many of the uncertainties of the MIR map and allowed us to build in or correct much of the model. Figs. 15(*a*) and 15(*b*) compare the MIR map in the region of the 40–43 disulfide with a $2F_{obs} - F_{calc}$ map phased on the complete model after the first SA simulation.

7.2. *Second SA simulation*

Following the first SA simulation the resulting model was taken through four cycles of standard KH refinement with intermittent modelling sessions; the latter were very limited, and amounted mainly to the introduction of side chains into all loops modelled as poly-Ala, with the exception of 245–256. In addition, the loops containing residues 30–51, 155–173 and 200–215 were corrected with regard to both main- and side-chain stereochemistry according to the improved electron density map. The resulting model was used as the starting point of the second SA simulation; the same protocol as described earlier (Fig. 4) was used.

Fig. 16 illustrates the reduction of error (main-chain only) achieved during the second SA simulation. Again, overall improvement was achieved, although the mean shift was about half of that in the first round. In order to analyse the changes in detail, we have again compared three sets of coordinates, *i.e.* the input model, the output and the final refined structure.

The changes were now confined to fewer regions, in particular those that were remodelled manually

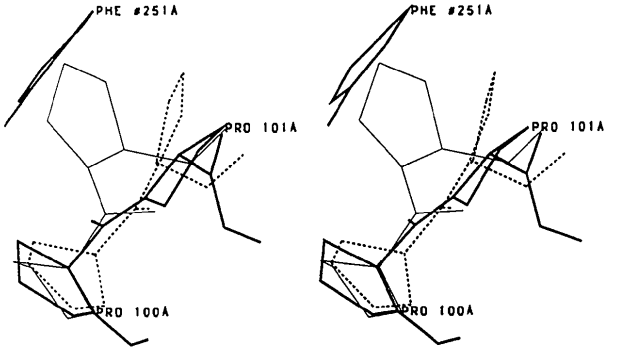

Fig. 14. An error introduced by the first SA simulation (thin lines) when the side chain of Pro 101 was moved into the density which should have been accounted for by Phe 251 from the symmetry-related molecule (other conventions as listed in the legend to Fig. 11).

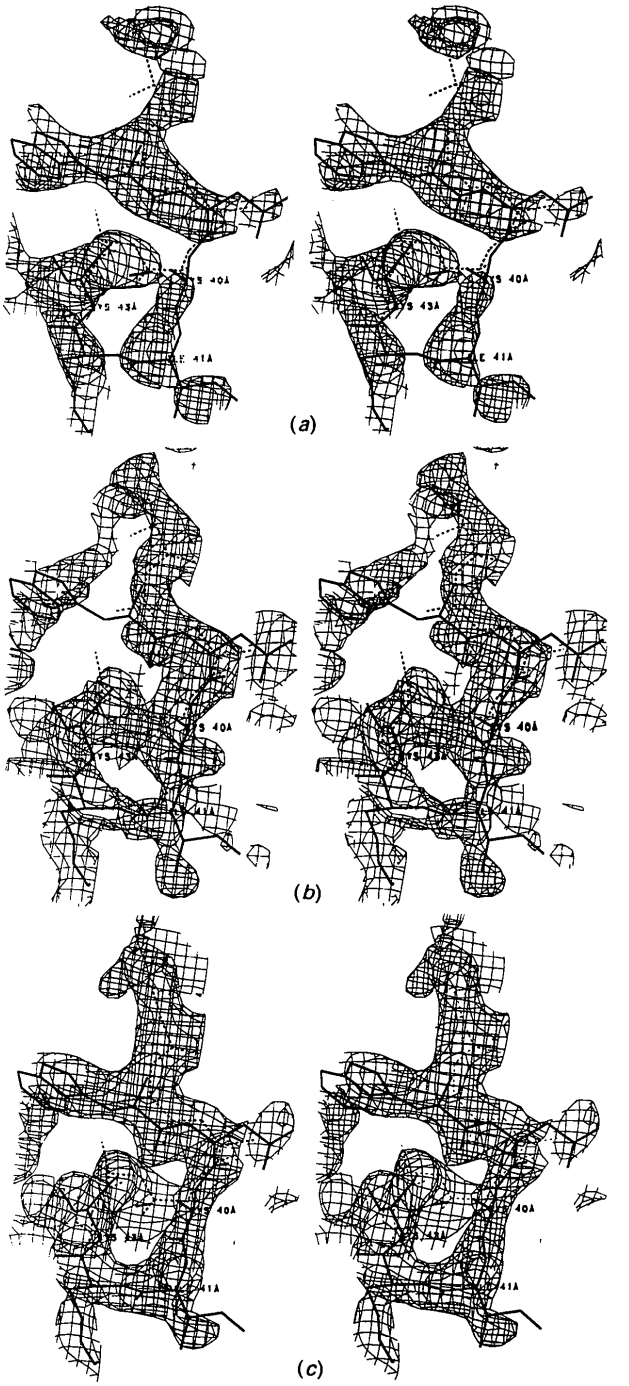

Fig. 15. The improvement in the quality of the electron density maps in the region around the disulfide bond (Cys 40–Cys 43) which was erroneously built in the original model. (*a*) MIR electron density map; only the initial and final structures are shown for clarity; (*b*) $2F_{obs} - F_{calc}$ electron density map (calculated at 2.0 Å resolution) based on the model following the first SA simulation; note the bias towards the model as well as the indication of the correct structure; (*c*) an $F_{obs}\alpha_{comb}$ electron density map following the first SA simulation; phase combination was carried out for all reflections for which the MIR phases were available, and phases were extended to 2.5 Å resolution.

after the first round of SA. However, occasional additional improvement was observed in a limited number of model fragments where the first simulation failed to reach the target structure, but which were not adjusted manually. In particular, peptides 172 and 179 were now flipped, though peptide 177 remained in error. Residues 240–243, practically unchanged in the first simulation, have now completely and successfully refined (Fig. 13*b*).

7.3. *Possible enhancement of SA refinement efficiency*

The results discussed in §7.1 and 7.2 illustrate that improved protocols may take the refinement automation beyond its present limits. In order to assess some possible means of enhancing the effectiveness the method we carried out two additional calculations.

Firstly, the output from the first SA simulation was subjected directly to another cycle of *XPLOR*. We found that some further refinement was achieved. In fact, most of the more pronounced changes are identical to those obtained during the second SA simulation. This means that the manual rebuilding which took place between the simulations had little effect on the course of the SA second refinement cycle.

Secondly, we investigated the question of the 'temperature' parameter in the heating stage. We used 4000 K and repeated the SA simulation with the output of the first cycle of *XPLOR*. There was little change between this run and the one at 3000 K although, surprisingly, in some areas there was less improvement than previously.

7.4. *Analysis of the accuracy of phasing during initial model building and partial refinement*

In the analysis of a protein crystal structure the initial success of chain tracing and model building depends critically on the accuracy of the phase determination, and the resulting quality of the electron density map.

In this study the mean difference between the original MIR phases and the final phases (*i.e.* mean phase error) was 50.5° (4139 acentric terms to *ca* 3 Å resolution); out of 892 centric terms 543 (61%) were phased correctly. This is somewhat worse than could be expected from the figure of merit – 0.74 indicative of a mean error of 42.3°. We have investigated the subsequent progress of phase refinement and the resulting quality of the electron density maps at the critical early stages of model building and partial structure refinement. Fig. 17 shows the results of these calculations.

The calculated phase angles based on the initial partial model yielded calculated phases with a mean error of 45° in the 3 Å sphere, and 55.3° in the 1.9 Å sphere. Neither the 3 Å resolution electron density map calculated with coefficients $mF_{obs}\alpha_{MIR}$ nor the one calculated at 2 Å with coefficients $2F_{obs} - F_{calc}$ (based on the partial model) were clear enough to trace the polypeptide chain in a significant proportion of loop regions, and even some of the helices. However, following a phase combination procedure, the mean phase error was reduced to 38.9° in the 3 Å sphere. Our original attempt to interpret a map calculated to 2 Å resolution, using combined phases between infinity and 3 Å, with extended model phases to 2 Å, proved to be a mistake. The 13 000 high-resolution terms increased the overall mean error to 54.8°, and in this case obscured the improvement achieved at medium resolution by phase combination.

Following the first SA simulation a decrease in phase error of 10.6° (from a mean error of 55.3–

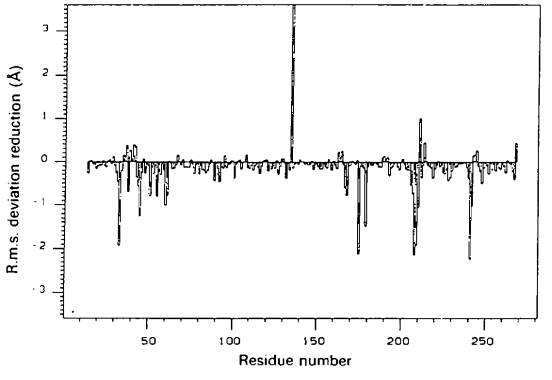

Fig. 16 Reduction in mean coordinate deviation (averaged per residue – main-chain atoms only) following the second SA simulation. For other details see Fig. 10.

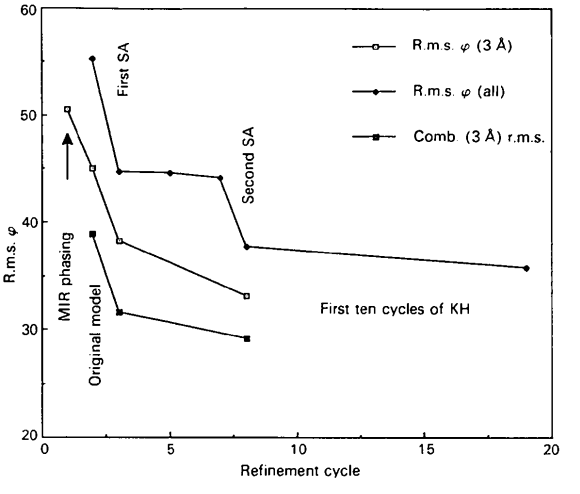

Fig. 17. Analysis of mean phase error at different refinement stages; note the relative insensitivity of phase error to the KH restrained refinement protocol.

44.7°) could be observed for acentric data. Interestingly, only modest improvement in the centric data can be observed, with 1381 reflections (out of a total of 2212) phased correctly (an increase of 58).

We have also investigated how far into the refinement one can benefit from combining the MIR phases with those calculated from a current atomic model. Fig. 17 shows the details. The results indicate, that even after the second SA simulation (when most of the $2F_{obs} - F_{calc}$ map was readily interpretable) further improvement could have been obtained by phase combination.

Fig. 15(*c*) illustrates marked improvement in the quality of the electron density map achieved by phase combination and phase extension following the first SA simulation.

8. Concluding remarks

We have successfully solved and refined the molecular model of *R. miehei* triacylglyceride lipase using both conventional MIR methodology and partial model refinement using simulated annealing. The original progress of the SA refinement, and its subsequent detailed analysis results in the following conclusions.

It is possible to improve the accuracy of partial models (either built into an MIR map or obtained by MR) by including poly-Ala sequence stretches and subjecting them to SA refinement. While this generally resulted in success and significantly helped the process of structure solution and model refinement, the progress of the refinement was still limited.

The problem of *cis*-prolines can easily be traced to the way in which the Ala peptide bond is heavily constrained by *XPLOR* (419 kJ mol^{-1}) which uses a potential function with a single minimum for the *trans* configuration. The proline imide groups are treated differently using a two-minima potential function and a much lower energy constraint, which leads to a relatively high rate constant for the *cis/trans* isomerization. *R. miehei* triacylglyceride lipase is characterized by an unusually high percentage of prolines in the *cis* configuration [35%, as opposed to the average 6.5% as seen in a survey of 1534 *X*-Pro imide bonds in the Protein Data Bank (Stewart, Sarkar & Wampler, 1990)]. It could be argued that in a routine case *cis*-prolines are less frequent, and the risk associated with the introduction of more flexible energy constraints would outweigh the benefits.

The second serious problem identified in our study is a more generally applicable situation in which a stretch of poly-Ala becomes out of step causing a 'bulge' and a 'shortcut' elsewhere. Although this clearly prevents further progress of refinement by constraining the 'out of step' alanines in the position of neighbouring residues, the benefits resulting from improved phasing accuracy are very significant. After all, from the point of view of phasing it is irrelevant which alanine is positioned where, as long as the electron densities are correctly accounted for.

We have no doubt, however, that our study indicates further potential in the *XPLOR* technique. The present limitations stem from the fact that there is no assessment of the progress of the refinement (other than the monitoring of the conventional *R* factor). It is conceivable, especially when well defined parts of the structure are being refined, that an automated procedure may be employed instead. One possible way is to use a real-space *R* factor (Branden & Jones, 1990) correlating the model with the current electron density map, with the MIR phases used somewhat longer than is traditionally accepted. In this fashion, a well behaved residue located between two obvious errors may be indentified as being possibly 'out of step'. An automatic procedure of building in a sequence into the well refined poly-Ala segments may also be considered.

In spite of its documented superiority over traditional least-squares crystallographic refinement, the performance of SA refinement is somewhat unpredictable. In this study we have frequently observed that smaller errors remained unaffected while much larger conformational changes took place during a single simulation. In several cases regions unaffected in the first cycle quickly refined during the second round, although we found no rational explanation for such behaviour. These observations indicate that SA refinement may be optimized further and that the problem of local energy barriers referred to by Fujinaga *et al.* (1989) may not be the main reason behind the unpredictability of the success of SA refinement. It is possible that the course of the SA simulation depends critically on the exact set of initial conditions. More studies are clearly required to optimize the current protocols.

Our data shows – yet again – that the MIR experimental phases should not be abandoned too early during the refinement. On the contrary, they should be included every time an electron density map is calculated *via* a phase combination step until the model is refined close to or below *ca* 0.30.

The atomic coordinates and structure factors of the model described above have been deposited with the Protein Data Bank.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 3TGL, R3TGLSF), 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 37059 (as microfiche). Copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Unless otherwise specified, all the programs referred to in this paper have been used in their VAX VMS versions of the *CCP*4 suite of crystallographic programs (SERC Daresbury Laboratory, 1986). *XPLOR* Version 1.5 as obtained from A. T. Brunger was used on a CRAY X-MP/416 supercomputer (SERC Rutherford Appleton Laboratory).

Novo-Nordisk Industri A/S, Copenhagen, Denmark, and the European Economic Community (Bridge Program) are thanked for financial support; the Protein Crystallography Group in York is also financed by a consolidated grant from the SERC (UK), who are also acknowledged for the provision of the CRAY computing time. JPT gratefully acknowledges a studentship under the EEC Biotechnology Action Program. The preparation of the final version of this paper took place at the University of Alberta, Edmonton, where ZSD is financed by an MRC (Canada) grant to the Group in Protein Structure and Function.

References

BOEL, E., HUGE-JENSEN, B., CHRISTIANSEN, M., LARS, T. & FIL, N. P. (1988). *Lipids*, **23**, 701–706.

BRADY, L., BRZOZOWSKI, A. M., DEREWENDA, Z. S., DODSON, E. J., DODSON, G., TOLLEY, S., TURKENBURG, J. P., CHRISTIANSEN, L., HUGE-JENSEN, B., NORSKOV, L., THIM, L. & MENGE, U. (1989). *Nature (London)*, **343**, 767–770.

BRANDEN, C.-I. & JONES, T. A. (1990). *Nature (London)*, **343**, 687–689.

BRUNGER, A. T. (1988). *XPLOR Manual*. Yale Univ., USA.

BRUNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* A**45**, 50–61.

BRUNGER, A. T., KURIJAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.

DEREWENDA, Z. S. & DEREWENDA, U. (1992). *Biochem. Cell Biol.* **69**, 842–851.

DEREWENDA, Z. S. & HELLIWELL, J. R. (1989). *J. Appl. Cryst.* **22**, 123–137.

DODSON, E. J. & VIJIYAN, M. (1971). *Acta Cryst.* B**27**, 2402–2411.

FUJINAGA, M., GROS, P. & VAN GUNSTEREN, W. F. (1989). *J. Appl. Cryst.* **22**, 1–8.

GOODFELLOW, J., HENRICK, K. & HUBBARD, R. (1989). Editor. *Molecular Simulation and Protein Crystallography, Proceedings of the Joint CCP4/CCP5 Study Weekend.* Warrington: SERC Daresbury Laboratory.

GUNSTEREN, W. F. VAN & BERENDSEN, H. J. C. (1987). *GROMOS, Groningen Molecular Simulation Library.* BIOMOS bv, Groningen, The Netherlands.

HENDRICKSON, W. A. (1985). *Methods. Enzymol.* **115**, 252–270.

HOWARD, A. J., GILLILAND, G. L., FINZEL, B. C., POULOS, T. L., OHLENDORF, D. H. & SALEMME, F. (1987). *J. Appl. Cryst.* **20**, 383–387.

HUGE-JENSEN, B., GAILUZZO RUBANO, D. & JENSEN, R. G. (1987). *Lipids*, **22**, 559–565.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.

KARPLUS, M. & MCCAMMON, J. A. (1983). *Annu. Rev. Biochem.* **52**, 263–300.

KIRKPATRICK, S., GELLAT, C. D. JR & VECCHI, M. P. (1983). *Science*, **220**, 671–680.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

MATTHEWS, B. W. (1966). *Acta Cryst.* **20**, 230–239.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **23**, 283–437.

SERC Daresbury Laboratory (1986). *CCP4. A Suite of Programs for Protein Crystallography.* SERC Daresbury Laboratory, Warrington, England.

STEWART, D. E., SARKAR, A. & WAMPLER, J. E. (1990). *J. Mol. Biol.* **214**, 253–260.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

WEIS, W. I., BRUNGER, A. T., SHEKEL, J. J. & WILEY, D. C. (1990). *J. Mol. Biol.* **212**, 737–761.

Acta Cryst. (1992). B**48**, 319–324

Model Deformation Density Studies of Tetraethylthiuram Disulfide

BY SHOW-KEI YEH AND YU WANG*

Department of Chemistry, National Taiwan University, Taipei, Taiwan

(*Received* 21 *October* 1991; *accepted* 13 *January* 1992)

Abstract

The title compound, tetraethylthiuram disulfide, $C_{10}H_{20}N_2S_4$, was studied by X-ray diffraction at 105 K. Crystal data: $P2_1/c$, $a = 10.922\,(3)$, $b = 15.946\,(4)$, $c = 8.444\,(1)$ Å, $\beta = 91.92\,(1)°$, $Z = 4$. The bonding character of this molecule was examined by means of the deformation density distribution both in a multipole expansion model and in a theoretical calculation according to the extended Hückel molecular orbital method. The deformation density distribution, after multipole refinements, shows clear features in both the bonding and lone-pair electron regions. Positive deformation density is found along the C—C, C—N and C—S bonds. There is very little density along the S—S bonds. The net atomic charges obtained from both methods are discussed.

* To whom all correspondence should be addressed.

0108-7681/92/030319-06$06.00