Acta Cryst. (1993). D**49**, 440–448

PRISM: Application to the Solution of Two Protein Structures

BY CHRISTOPHER BYSTROFF,*† DAVID BAKER,‡ ROBERT J. FLETTERICK* AND DAVID A. AGARD‡§

Howard Hughes Medical Institute and the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143-0448, *USA*

(*Received* 2 *October* 1992; *accepted* 21 *April* 1993)

Abstract

The previous paper described a phase-refinement strategy for protein crystallography which exploited the information that proteins consist of connected linear chains of atoms. Here the method is applied to a molecular-replacement problem, the structure of the protease inhibitor ecotin bound to trypsin, and a single isomorphous replacement problem, the structure of the N-terminal domain of apolipoprotein E. The starting phases for the ecotin–trypsin complex were based on a partial model (trypsin) containing 61% of the atoms in the complex. Iterative skeletonization gave better results than either solvent flattening or twofold non-crystallographic symmetry averaging as measured by the reduction in the free *R* factor [Brünger (1992). *Nature* (*London*), **355**, 472–474]. Protection of the trypsin density during the course of the refinement greatly improved the performance of both skeletonizing and solvent flattening. In the case of apolipoprotein E, previous attempts using solvent flattening had failed to improve the SIR phases to the point of obtaining an interpretable map. The combination of iterative skeletonization and solvent flattening decreased the phase error with respect to the final refined structure, significantly more than solvent flattening alone. The final maps generated by the skeletonization procedure for both the ecotin–trypsin complex and apolipoprotein E were readily interpretable.

Introduction

This is the second of two papers describing the continued development of the *PRISM* method (Wilson & Agard, 1993) for protein crystal structure solution. This method exploits the knowledge that proteins consist of linear and connected chains of atoms. In the previous paper using synthetic data we showed that *PRISM* was able to significantly reduce the phase error in a challenging molecular-replacement test case where initial phases were derived from a small fraction (about 33%) of the atoms of the known structure. The correlation of phase error with 'free *R* factor' (Brünger, 1992*a*) was established, and a general procedure was defined for using the free *R* factor to determine the optimal parameters for a run.

In this paper, we describe two applications of *PRISM* to solving protein structures. In both cases, *PRISM* converted an uninterpretable map to a map with well defined protein features that could be traced with minimal difficulty. In one case, the result was confirmed by independent means; in the other case, the final structure had been previously determined.

In the first case, the ecotin–rat trypsin complex, initial phases were derived from an incomplete molecular-replacement model (rat trypsin, 61% of the atoms). The initial map was untraceable in the ecotin region. Iterative skeletonization produced an interpretable map, better than that obtained by either solvent flattening using the Wang method (Wang, 1985) or twofold non-crystallographic symmetry (NCS) averaging (Fletterick & Steitz, 1976; Bricogne, 1976).

In the second case, the N-terminal receptor-binding domain of apolipoprotein E (apoE), the initial phases were based on the single heavy-atom isomorphous replacement (SIR) method. The structure of apolipoprotein E had been previously solved by combining SIR data with anomalous scattering data from the single mercury derivative (Wilson, Wardell, Weisgraber, Mahley & Agard, 1991). In the course of trying to solve the apoE structure, significant effort had been made to utilize SIR and solvent flattening. While the solvent flattening did improve the phases, the resultant map was not interpretable. By contrast, here we show that iterative skeletonization in combination with solvent flattening leads to a directly interpretable map.

Materials and methods

The *PRISM* algorithm is described in the previous paper (Baker, Bystroff, Fletterick & Agard, 1993). In brief, *PRISM* is a package of programs, some writ-

* Department of Biochemistry and Biophysics.

† Current address: Universidad Nacional de Ingeniería, Managua, Nicaragua.

‡ Howard Hughes Medical Institute and the Department of Biochemistry and Biophysics.

§ Author to whom correspondence should be addressed.

ISSN 0907-4449

ten by the authors and some borrowed from the *CCP*4 suite of programs (SERC Daresbury Laboratory, 1986) which perform skeletal density modification, solvent flattening and NCS density averaging. All of the programs used in the iterative procedure are available from the authors, as well as the VAX/VMS command files that run the package.

Based on the results presented in our previous paper, we monitored the free R factor as a measure of phase error to evaluate the progress of the density-modification procedures. The input parameters to *PRISM* (minden, maxden, epden, mingraph, β and solvent) were optimized as discussed previously. Because removing the 'free' set of observed reflections degrades the performance of density modification, we normally run *PRISM* again under the same conditions using the optimized parameters and with the free R factor option turned off.

The use of a 'mask' in *PRISM* refers to a two-step process by which known density is preserved from cycle to cycle. An envelope file is created which defines the known regions of the map, and is used in the first step to zero the known regions prior to skeletonization, flattening, or NCS averaging. In the second step, the structure factors for the known region are derived directly from the transformed atomic coordinates and are combined in reciprocal space with structure factors calculated from the transformed modified map (see Baker *et al.*, 1993).

In cases where there is non-crystallographic symmetry, the NCS correlation coefficient can be monitored as an independent measure of the phase error. The NCS correlation as calculated by *MAPCORREL* is defined as:

$$[\langle\rho(x)\rho(Mx+\mathbf{v})\rangle - \langle\rho(x)\rangle\langle\rho(Mx+\mathbf{v})\rangle]/$$
$$\{\langle(\rho(x)-\langle\rho(x)\rangle)^2\rangle\langle[\rho(Mx+\mathbf{v})-\langle\rho(Mx+\mathbf{v})\rangle]^2\rangle\}^{1/2} \quad (1)$$

where $\rho(x)$ is the density at x and matrix M and vector $\mathbf{v}$ describe the NCS operation. If the density is converging on the correct structure then the NCS correlation should increase to a limit of 1.00. As expected, the NCS correlation inversely parallels the free R factor. Note that, of course, the NCS correlation will not be a good measure of the phase error if the density-modification run includes NCS averaging.

For the ecotin–trypsin case, the coefficients used for map calculation were $2wF_o - F_c$, α_c, where w is the Sim weight (Sim, 1960). For the apolipoprotein E case, the coefficients for map calculations were mF_o, α_{best}, where m is the figure of merit and the phases were generated by combination with the SIR phase probabilities.

Maps were traced using *INSIGHT*II (Biosym Technologies, 1991) on a Silicon Graphics Iris 4D/25, and by using *FRODO* (Jones, 1985) on an Evans and Sutherland PS 390. Skeletons were used as a guide in tracing the chain, but as yet we have no automatic conversion of the skeleton to protein coordinates. After placing the C^α atoms by hand, *MAXSPROUT* (Holm & Sander, 1991) was used to convert the trace to a complete backbone-atom coordinate set.

Results and discussion

Tryspin–ecotin complex phased using a partial model

Ecotin is a dimeric heat-stable protease inhibitor from *E. coli* (McGrath, Erpel, Browner & Fletterick, 1991). Its interest derives from its inhibition of such diverse pancreatic serine proteases as chymotrypsin, elastase and trypsin. Table 1 gives characteristics and X-ray data statistics for the ecotin–trypsin complex crystals. Two ecotin molecules and two trypsin molecules form the asymmetric unit. The ecotin–rat trypsin complex is an attractive candidate for the evaluation of this method for a number of reasons. First, the structure of ecotin is not known, nor is the structure of any homolog known. Secondly, molecular replacement using the known high-resolution structure of rat anionic trypsin (McGrath *et al.*, 1992) should provide an accurate estimate for about 61% of the total atoms in the complex. Thirdly, NCS could be monitored as an independent measure of the quality of the map, or could be used to improve the phases further. And fourth, the data have realistic levels of errors and completeness.

The complex structure was solved by molecular replacement using *X-PLOR* (Brünger, 1992*b*). Using 7–4 Å data, the rotation function peaks were 8.0 and 5.5 standard deviations above the mean value (σ) for the search. The translation function was a reciprocal space correlation of E^2's for 7–4 Å data, and produced one unambiguous peak for each rotation function solution (9.7 and 4.9σ). One additional search was necessary to find the relative position of the two molecules in Y and the relative choice of origin. The resulting two-trypsin model was refined against 2.8 Å data as two rigid bodies. The starting R factor for data to 2.8 Å was 44%.

Further details on the crystallization, data collection, molecular replacement, refinement and the analysis of the structure of ecotin will be presented elsewhere (Erpel, Bystroff, Fletterick & McGrath, 1993). In this study, we will concentrate on the density-modification step.

Predictably, in a $2F_o - F_c$ map, the density in the trypsin region of the starting map was easily interpretable but the density in the ecotin region was judged to be untraceable. It was possible to trace only a few short stretches of peptide. In the absence

Table 1. *Characteristics and X-ray data statistics for the ecotin–trypsin complex data from the MAR image-plate system at UCSD*

Space group	$P2_1$
Cell dimensions	$a = 80.66$, $b = 83.11$, $c = 62.46$ Å, $\beta = 97.48°$
Trypsin	223 residues, $M_r = 24000$, T
Ecotin	142 residues, $M_r = 16000$, E
Completeness to 2.8 Å	95%, 10118 reflections
Redundancy	2.8, 28438 observations
No. of molecules in asymmetric unit	2, E2T2

Summary of data collection

D_{min} (Å)	% Completeness	R_{sym}*	Average I	σ
8.05	87	0.047	878	79.7
5.69	87	0.050	424	38.5
4.65	90	0.046	646	56.8
4.02	93	0.050	686	60.2
3.60	94	0.062	444	44.1
3.29	95	0.086	308	40.9
3.04	96	0.112	207	36.3
2.85	96	0.154	133	31.3
2.68	97	0.233	84	27.1

* $R_{sym} = \Sigma(I - \langle I \rangle)/\Sigma I$.

of density-modification schemes, one would be forced to attempt to model the interpretable regions, refine the partial model, and look for additional interpretable density (for an example, see Grütter, Fendrich, Huber & Bode, 1988). This can be an extremely slow and laborious process.

Wang's method, as modified by Leslie (Wang, 1985; Leslie, 1987), was used to generate a solvent envelope from the starting map, but the envelope was found to misrepresent the ecotin region of the map as solvent, even if an unreasonably low value for the percent solvent was chosen. This was remedied by running a few cycles of *PRISM* (skeletonization) without using a solvent envelope. This placed sufficient density in the ecotin region to generate a good envelope. This envelope was used as the starting envelope for the following *PRISM* runs, including the optimization runs. Generally, the envelope was recalculated automatically every three to five cycles, but only minor changes were seen in the envelope after the initial two or three cycles.

To optimize the parameters minden, maxden, epden, mingraph, solvent and β, skeletonization was run for three to five cycles using the free R factor option. 3% of the data (546 reflections), chosen at random, were flagged in the input data as the 'free' set. Each of the above six parameters were varied in turn and the value of that parameter that gave the lowest free R factor was chosen, similar to the procedure used in our previous paper. In some cases, the parameters are interdependent (*i.e.* mingraph depends on maxden and minden) and needed to be re-optimized when one was changed. The optimum values for skeletonizing with a mask are as follows: minden = 1.2, maxden = 2.5, mingraph = 9, epden = 2.2, $\beta = 10.0$, solvent = 0.45. Fig. 1 shows the solvent-parameter optimization, which was also performed for the solvent-flattening control experiments.

Both iterative skeletonization and solvent flattening tend to degrade the density within the known (trypsin) region of the map. This is because of two effects: first, at this stage the phases and hence the electron-density map at any given cycle are significantly in error and this error will propagate into the 'known' region, and second, the density-modification step inevitably introduces some error (as shown in the preceding paper). The degeneration of the electron density in the trypsin region can be avoided by Fourier space recombination at each cycle of structure factors from the trypsin coordinates and structure factors from the latest skeletonized map (not including the trypsin region) as described in the previous paper.

Ten cycles of skeletonization were run with a mask to preserve the trypsin density using all available data to 2.8 Å. Missing reflections were replaced by F_c's according to the procedure described in the accompanying paper. The free R factor dropped to 38% within six cycles and thereafter did not change significantly. The NCS correlation rose from 0.37 for the initial map to 0.53 after ten cycles of skeletonization. The connectivity in the ecotin part of the map improved dramatically (Fig. 2) and the map could be traced with minimum difficulty. The overall β-sheet ('jelly roll') structure of the protein could be easily seen.

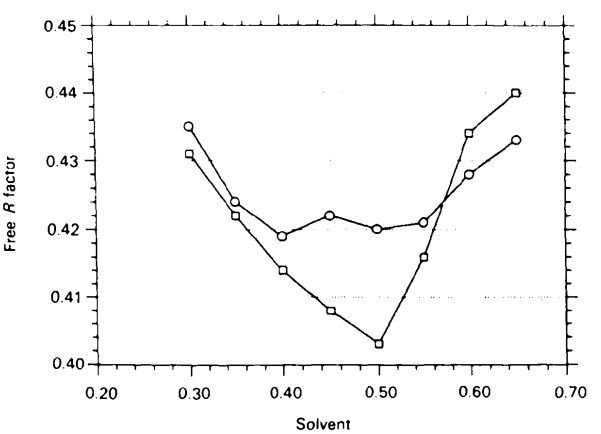

Fig. 1. Optimization of the percent solvent for the ecotin–trypsin complex by monitoring the free R factor. Circles denote free R factors obtained from runs of four cycles of solvent flattening with a mask. Squares denote minimum free R factors obtained from runs of four cycles of skeletonization with a mask. The mask in both cases protects the trypsin regions from density modification. A value slightly lower than the optimum was chosen for the extended runs in order to allow a little room for errors in the envelope. The calculated value for the solvent fraction is 52.5%. Optimizations of other *SKELETON* parameters (not shown are minden, maxden, mingraph, β) were carried out in a similar manner.

For comparison, solvent flattening was applied to the ecotin–trypsin map. Solvent flattening has been shown to improve the density of MIR maps but has not been generally used to improve molecular-replacement maps (an exception is Grütter *et al.*, 1988). The first experiment utilized ten cycles of solvent flattening without a mask to preserve the trypsin density. Solvent envelopes were calculated from the initial map, and thereafter every two cycles. The number used for the parameter solvent was determined by running five cycles of solvent flattening with a mask for solvent values ranging from 0.35 to 0.60. The optimal value was around 0.50, which is about the same as the calculated value (0.52) for the solvent fraction. The free *R* factor dropped to 53% in two cycles but then slowly rose, approaching a random value (55%). The resulting map was worse than the initial map; density in the trypsin region was broken and distorted. Density in the ecotin region was sparse and disconnected. Inspection of the solvent envelope calculated from the initial map, and all subsequent envelopes, showed that large portions of the ecotin region of the map were misassigned to the solvent.

To confirm that the erroneous envelope was the cause of the poor performance, solvent flattening was again run with a 'good' starting envelope calculated from the output map of the skeletonization run dicussed earlier. Other than the introduction of the good starting envelope, solvent flattening was performed in the same way as before. The free *R* factor dropped to about 45% in four cycles and thereafter rose slowly. The resulting map was improved from the original, but was still difficult to trace in the ecotin region, with large breaks, many wrong connections and disconnected density. Although labori-

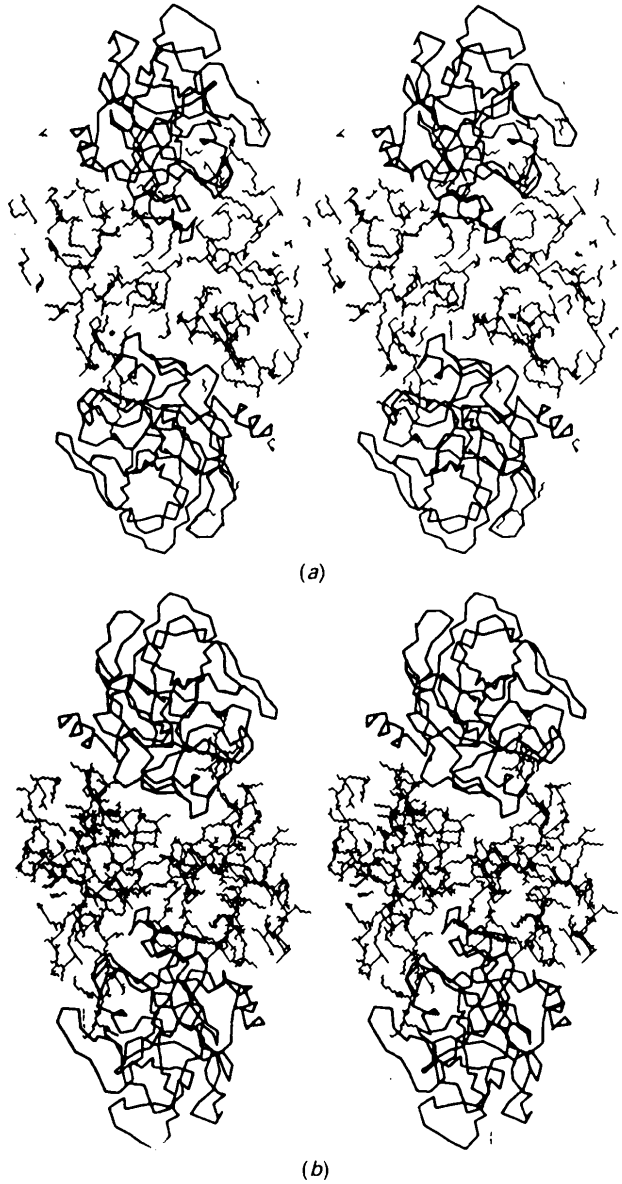

Fig. 2. Comparison of starting and final skeletons. Two trypsin monomers (thick lines) are shown alongside the skeleton (thin lines) of the initial 2.8 Å map (*a*) and the map after 20 cycles of skeletonizing and flattening with a mask (*b*). The view is approximately along the non-crystallographic twofold axis. The skeletons were generated using the same optimized minimum and maximum density cutoffs (minden and maxden) and the same minimum graph size (mingraph). The initial map contains mostly disconnected density. Twofold symmetrical features are visible in the final skeleton of the ecotin dimer. Iterative skeletonization has produced a much more connected map.

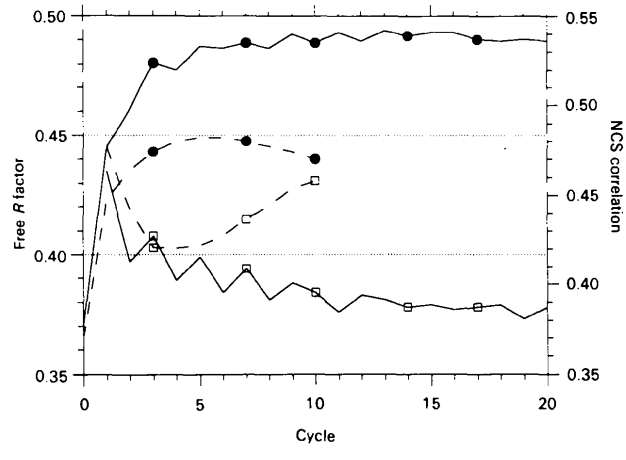

Fig. 3. Free *R* factor and NCS correlation for *PRISM* runs on the ecotin–trypsin complex. Ten cycles of solvent flattening with a mask (dashed lines); ten cycles each of alternating skeletonization and solvent flattening with a mask (solid lines); free *R* factor (open squares); NCS correlation (filled circles). The NCS correlation was measured only in the ecotin region. The mask protects the trypsin region from the density modification. Solvent flattening converges quickly and then overfits the data, as indicated by a rise in the free *R* factor. Skeletonization converges more slowly, and displays no rise in the free *R* factor. Skeletonization increases the NCS correlation more than flattening.

Table 2. *Summary of the results of various density-modification procedures on the ecotin–trypsin complex*

Unless otherwise stated, all available 2.8 Å data was used. In all cases, a solvent envelope was applied with 45% solvent. The free R factor was calculated from 3% of the observed data excluded from use in calculating the maps. The free R factor for no density modification is 1.00. The source of phases indicates the procedure described in the text that gave the above results. 'NCS' in this case refers to the non-crystallographic symmetry correlation [equation (1)] as calculated for the ecotin region of the maps. Note, the relatively high NCS value for standard flattening results from the fact that flat solvent regions are included within the envelope used for NCS averaging.

Source of phases	Free R	NCS	R factor
Starting map 2.8 Å	n/a	0.37	0.440
Skeletonizing with mask	0.377	0.53	0.312
Skeletonizing without mask	0.442	n/a	0.286
Flattening standard method	0.530	0.58	0.117
Flattening with good envelope	0.443	0.42	0.109
Flattening with mask	0.403	0.48	0.229
Flattening with mask and good envelope	0.404	0.48	0.228
NCS averaging without mask	0.399	0.85*	0.137
NCS averaging with mask	0.399	0.76*	0.226
Skeletonizing and averaging with mask	0.355	0.68*	0.323/0.283†
Skeletonizing and flattening with mask	0.373	0.54	0.334/0.283†
Skeletonizing and flattening with mask 2.4 Å	0.405	0.55	0.346/0.269†
Control with mask	0.531	0.33	0.076

* NCS correlation is not a measure of phase error in these cases because it is enforced.

† R factor oscillates. Higher R is after *SKELETON*.

ous, model building starting from this map would nonetheless have proceeded much more quickly than from the original $2F_o - F_c$ starting map. Of course, starting with a good solvent envelope is not an option in most solvent-flattening cases.

Preservation of the trypsin density through the use of a mask and Fourier space vector combination markedly improved the performance of solvent flattening. The free R factor dropped to about 40% in three cycles, and rose thereafter (see Fig. 3). The NCS correlation inversely paralleled the free R factor, rising from 0.37 to 0.48 and then dropping gradually. The maps from solvent flattening with a mask are much cleaner and more connected than those without and produced a map that could be traced in most places. In a separate run, simply replacing the density within the known region with the trypsin density at each cycle only led to a slight drop in the free R factor (Table 2).

The best results of density modification without using the NCS information were obtained by alternating skeletonizing and solvent flattening. The two procedures were alternated for 20 cycles, ending with a free R factor of 0.373, slightly better than skeletonizing without solvent flattening. A slightly improved NCS correlation also suggested that the phase error is lower than for skeletonizing alone (see Table 2). Parts of this map are shown in Figs. 4(*c*) and 5(*c*). This map was one of the three used in the actual tracing of the chain.

A side-by-side comparison of the masked/flattened map with the masked/skeletonized/flattened map showed the latter to have better connectivity and less disconnected density (Fig. 4). In a quick survey of connectivity, six places were found where solvent flattening left a significant break in the chain and skeletonizing did not (*i.e.* Fig. 5). No cases of the converse were found, although in a number of places both methods left a break. Skeletonizing tended to introduce false connections more often than flattening, but these are generally easy to detect by using knowledge of the sequence and secondary structure. In places where both maps had good density, the density in the skeletonized map was more evenly

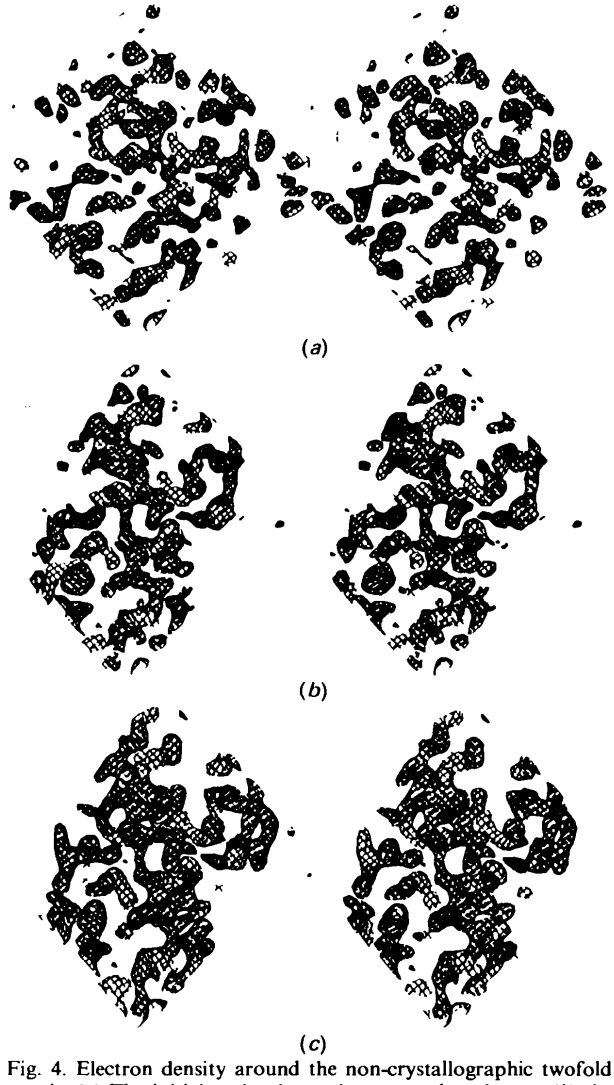

Fig. 4. Electron density around the non-crystallographic twofold axis. (*a*) The initial molecular-replacement phased map, (*b*) the map after ten cycles of solvent flattening with a mask, and (*c*) the map after ten cycles each of solvent flattening and skeletonization with a mask. The initial map shows disconnected density and very little indication of the symmetry. The flattened map has some twofold symmetry, but is still very disconnected. The skeletonized map is cleaner and more connected and shows more of the twofold symmetry.

distributed along the chain, while the density in the flattened map was relatively noisy, showing up as bulges and bottlenecks in the contours.

As expected from the comparison of the free R factor and phase error in the previous paper, the map quality in general seems to reflect the differences in the free R factors (37% for skeletonizing/flattening, 40% for flattening). Tracing the chain through the skeletonized map was much easier than through the solvent-flattened map. More dramatic differences between these two methods are expected when higher resolution data are used (see previous paper), and when the initial phase error is worse (see apolipoprotein E case).

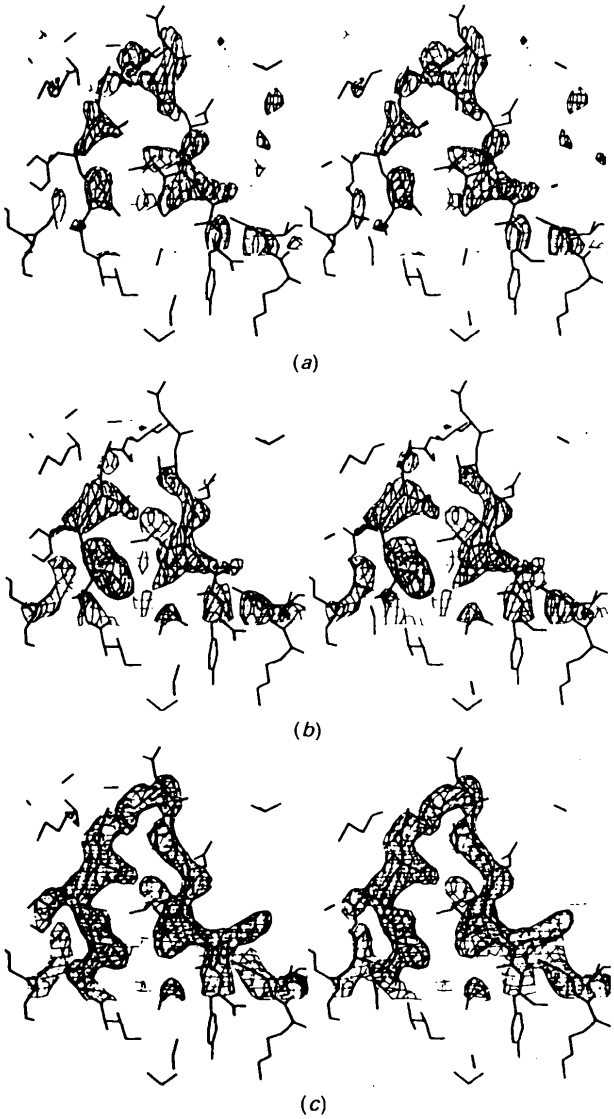

Fig. 5. A close-up of a surface loop in ecotin. The coordinates are from a partially refined model. (*a*) The starting map shows weak and disconnected density. (*b*) After solvent flattening, the density for the loop is stronger, but still disconnected and difficult to interpret. (*c*) After skeletonization and flattening, the density is connected and easy to trace. Lowering the contour level does not improve the connectivity for (*a*) and (*b*).

The conventional way to solve the ecotin–trypsin structure would be by density averaging using the twofold non-crystallographic symmetry. Application of NCS averaging dropped the free R factor to about 40%. The map showed mostly connected density with some breaks and some erroneous connections, but tracing the chain was possible. Preservation of the density in the trypsin region did not significantly alter the results (Table 2), presumably because the twofold averaging greatly dampens the accumulation of errors.

To see if skeletonizing could augment the power of NCS averaging, we alternated cycles of skeletonizing with averaging, and followed the free R factor. The optimum parameter settings were found to be slightly different than for skeletonization alone. The parameters maxden = 4.0, epden = 2.8 and mingraph = 15 were used in this run. These values (higher maxden, higher mingraph) more tightly impose the topology constraint when skeletonizing. Perhaps because averaging improved the map somewhat, the topology constraint could be more tightly imposed without introducing errors.

After 20 cycles of alternating skeletonizing with averaging, the free R factor converged at about 0.355. This is the best free R factor we have obtained from density modification of ecotin–trypsin. The map reflected this improvement, showing fewer breaks in the chain and fewer wrong connections.

The skeletonized NCS-averaged map was used to trace the chain and for the initial building of the side chains before refinement (Erpel *et al.*, 1993). The skeletons were used as a guide during the chain tracing. The chain was traced from residues 9 to 139 (except for residues 88 to 94) in about 3 days. Residues known to be in the trypsin active site were helpful in starting the side-chain assignments as the side-chain densities were often ambiguous. In the refinement process only one major mistake in the trace was detected: a shortcut taken through a β-turn, leaving out two residues. No mistakes in chain direction or connectivity were made, despite the presence of a fairly controversial fold (discussed in Erpel *et al.*, 1993).

In this case study, we have shown that introducing a topological constraint on the density map can improve the phases, as measured by the free R factor and map interpretability, for a structure phased by a partial model at 2.8 Å resolution. The phase improvement is better than that obtained by solvent flattening or NCS averaging, and the method is complimentary to both. Skeletonization alternated with solvent flattening gave the greatest improvement in the free R factor if the non-crystallographic

symmetry was ignored. Skeletonization alternated with NCS averaging provided an improvement on either averaging or skeletonizing alone.

The use of a mask to preserve the density of the partial model, and the concomitant two-parameter scaling of structure factors from the partial model and structure factors from the modified map to the observed amplitudes, was found to be necessary for the convergence of solvent flattening and skeletonization but not for NCS averaging.

The unexpected result that skeletonizing worked better than averaging when a mask is employed shows that the topology constraint is stronger than the NCS constraint for the unknown ecotin density. That is to say, skeletonizing adds more information than twofold NCS averaging when the initial density is poor.

Apolipoprotein E with SIR starting phases

In this and the accompanying paper we have shown the power of the *PRISM* iterative skeletonization procedure for crystallographic problems in which partial structure information is available. The other major source of initial phase information for macromolecules is of course isomorphous replacement. In the case of a single isomorphous heavy-atom derivative, the phase probability distributions are bimodal. Solvent flattening has proven to be a powerful means of reducing this phase ambiguity. Here we investigate the power of iterative skeletonization in improving the phase probability distributions.

The structure of the LDL receptor binding domain of apolipoprotein E was originally solved using a combination of single isomorphous replacement and anomalous-scattering data on a dimethyl mercury derivative (Wilson *et al.*, 1991). The phasing power of the SIR data (100% complete from 10.0 to 2.5 Å resolution) was 1.55, and the mean figure of merit 0.41. Maps generated using only the SIR data were improved by solvent flattening, but not to the point of interpretability. Wilson *et al.* (1991) concluded '(after solvent flattening) noise in the map was significantly reduced, (but) the electron density remained too ambiguous to allow modeling of the structure.' This provides a good test for the power of the skeletonization procedure: Would it have allowed solution of the structure without the anomalous-scattering data?

To address this question, a map was generated using the SIR data and the standard solvent-flattening protocol was applied until convergence was reached. The weighted phase difference with respect to the transform of the final 2.25 Å refined model (1LPE) was 60.5° before and 53.5° after solvent flattening. The solvent-flattened map was then subjected to alternating cycles of solvent flattening and skeletonization, with the envelope being recalculated periodically. At each cycle, updated phase-probability distributions were calculated using the formula (Sim, 1960):

$$P_{\text{new}}(\varphi) = P_o(\varphi)\exp[2(|F_o||F_c|/|F_o^2 - F_c^2|)\cos(\varphi - \varphi_c)],$$

where P_o is the original (bimodal) SIR phase probability distribution and F_c and φ_c are structure-factor amplitudes and phases calculated from the modified map. New maps were generated using the observed amplitudes weighted by the figure of merit and the best (centroid) phases. Because of fluctuations in the progress of refinement, two runs with slightly different envelope recalculation schedules were performed and are shown in Fig. 6. The phase error relative to the final refined model rapidly dropped to 49°, and then gradually leveled out at around 46°. The oscillations evident after the first ten cycles reflect the alternation between skeletonization and solvent flattening – the phase error increases slightly after skeletonization, and then falls to a level slightly below that of the previous cycle after solvent flattening.

Representative portions of the original SIR map, the map after extensive solvent flattening, and the map after combined skeletonization and solvent-flattening are shown in Fig. 7. The original SIR map is quite noisy and fairly disconnected. Solvent flattening significantly reduces the noise, but the map

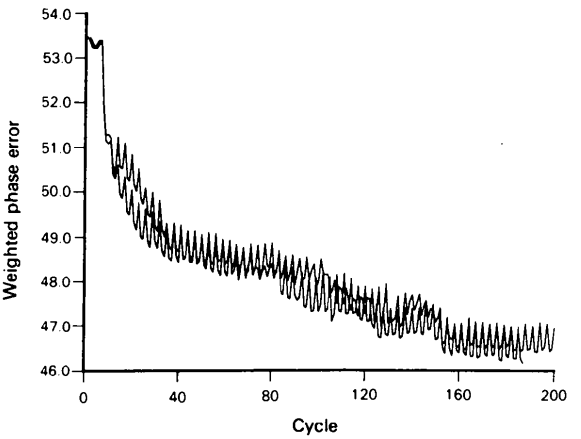

Fig. 6. Application of *PRISM* to an SIR problem. The SIR phase probability distributions described in Wilson *et al.* (1991) were improved by solvent flattening until convergence was reached. The resulting map was then subjected to alternating cycles of skeletonization and solvent flattening (two cycles of solvent flattening per cycle of skeletonization). The figure shows the weighted phase error relative to the refined model of apolipoprotein E during the last six cycles of solvent flattening and the subsequent 200 cycles of skeletonization/flattening. The solvent envelope was recalculated every ten cycles beginning with the fifth cycle for one of the traces and the sixth cycle for the other.

remains disconnected. Considerable improvement is seen after the combined skeletonization and solvent-flattening procedure. This final map is easily interpretable.

It is not difficult to understand the synergism between solvent flattening and skeletonization. Solvent flattening is a relatively mild procedure and is readily trapped in local minima. The skeletonization procedure is potentially much more powerful since it enforces much stronger density constraints, but it also is susceptible to larger errors. These errors have two sources: first, the solvent in the crystal is effectively ignored, and second, the algorithm inevitably makes mistakes in tracing the polypeptide chain. The combination of the two methods is powerful since skeletonization can keep solvent flattening from converging on incorrect local minima, while solvent flattening is well suited to removing errors introduced by skeletonization. The solvent region is reasonably well represented by setting the density at gridpoints outside the envelope to their mean value, and at least of a portion of the errors made by the skeletonizer inside the envelope will be filtered out by the flattening procedure in the next cycle.

Future developments

The success of *PRISM* in these two cases suggests that protein structures may now be solved with considerably less initial phase information. For instance, the solution of a hetero-dimer structure may now be possible starting with phases derived from only one monomer using no additional experimental phase information (*i.e.* a protein complexed with an antibody should be solvable starting with phases from the antibody). Similarly, structures may now more easily be solved using only a single heavy-atom derivative. Combination of *PRISM* with other methods for phase improvement, notably the implementation of Sayre's equation and histogram matching (*SQUASH*) described by Zhang & Main (1990), may result in a still more powerful phase-improvement method. *PRISM* and *SQUASH* exploit

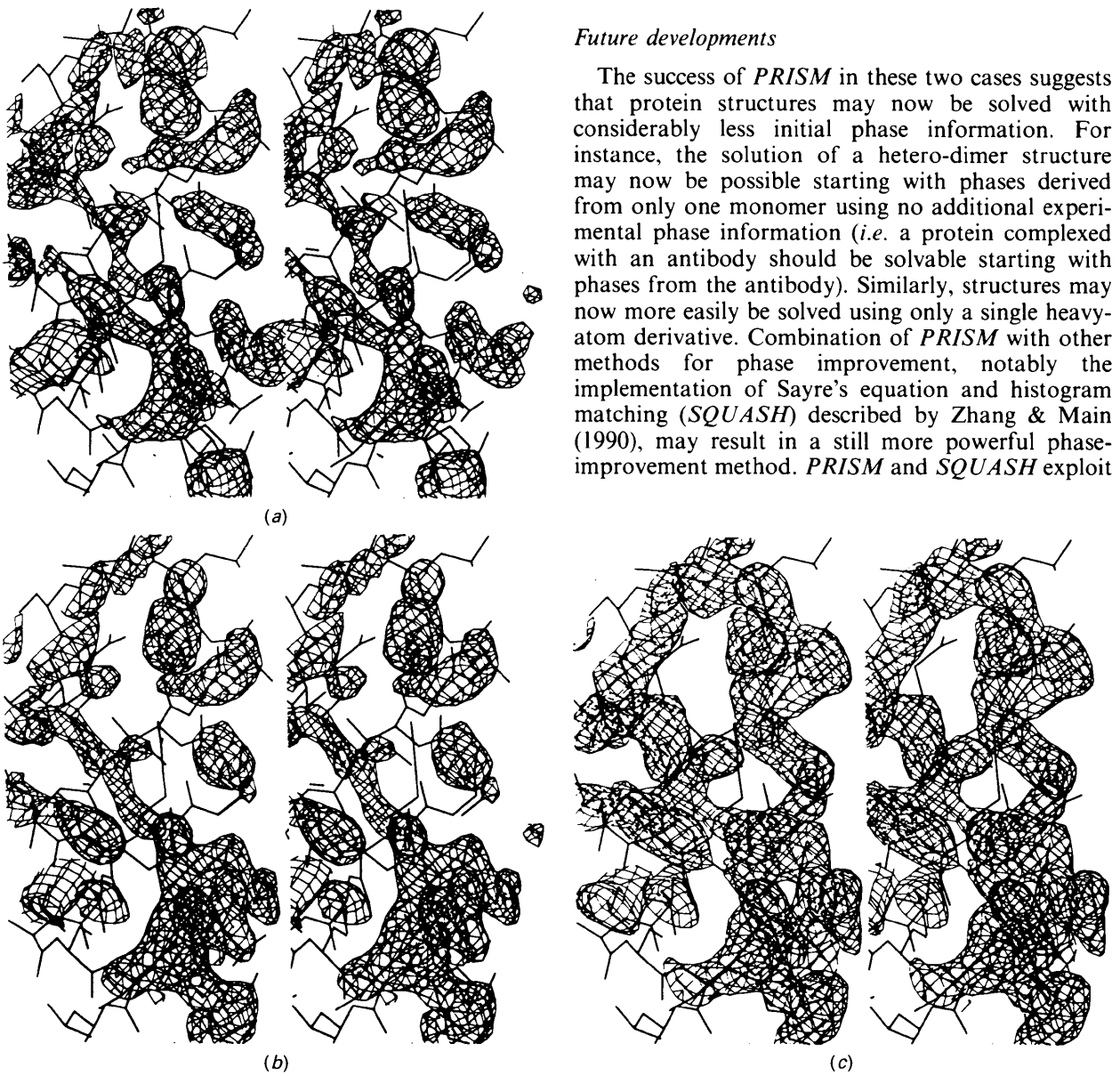

Fig. 7. Density around one of the α-helices in apolipoprotein E. (*a*) The initial SIR-phased map, (*b*) the map after solvent flattening to convergence and (*c*) the map after 200 cycles of skeletonization and solvent flattening. The density in the skeletonized map is well connected and easy to trace, whereas the solvent-flattened map is not interpretable.

different chemical information and hence should be at least in part orthogonal.

The *PRISM* method depends critically on the accuracy of skeletonization. When there are two possible pathways through the density, the skeletonizer will break the one with the weakest link, most often the longer pathway. A relatively simple improvement to the algorithm would be to select the pathway with the higher average density. A more sophisticated algorithm might insert segments of idealized helix, turn or strand into likely regions of density. The increase in chemical information employed by such a pattern recognizer should significantly increase the radius of convergence of the method. A final step would be the automatic interpretation of the skeleton as protein (for steps in this direction, see Jones, Zou, Cowan & Kjeldgaard, 1991).

In molecular-replacement problems with an incomplete and error-containing starting model it may be useful to alternate between *PRISM* refinement with the density corresponding to the model fixed, and conventional atomic positional refinement with partial structure factors derived from a fixed skeletonized map. Such a procedure would eliminate the tendency of an incomplete atomic model to inappropriately adjust during positional refinement to compensate for the missing scattering density.

We thank Shell Chen for programming assistance. Thanks to Mary McGrath and Thorsten Erpel for trusting their molecule to our hands. And thanks to Virginia Rath, Cecilia Schiffer and Partho Ghosh for trying out the unfinished product. This work was made possible by funding from the Howard Hughes Medical Institute (DB and DAA), NIH grants DK26081 and DK39304 (CB and RJF). DB is an HHMI LSRF Postdoctoral Fellow.

References

BAKER, D., BYSTROFF, C., FLETTERICK, R. J. & AGARD, D. A. (1993). *Acta Cryst.* D**49**, 429–439.
Biosym Technologies (1991). *INSIGHTII*. Biosym Technologies, San Diego, California, USA.
BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.
BRÜNGER, A. T. (1992*a*). *Nature (London)*, **355**, 472–474.
BRÜNGER A. T. (1992*b*). *X-PLOR*. Version 3.0. Yale Univ., New Haven, USA.
ERPEL, T., BYSTROFF, C., FLETTERICK, R. J. & MCGRATH, M. E. 91993). Submitted.
FLETTERICK, R. J. & STEITZ, T. Z. (1976). *Acta Cryst.* A**32**, 125–132.
GRÜTTER, M. G., FENDRICH, G., HUBER, R. & BODE, W. (1988). *EMBO J.* **7**, 345–351.
HOLM, L. & SANDER, C. (1991). *J. Mol. Biol.* **218**, 183–194.
JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–170.
JONES, T. A., ZOU, J.-Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110–119.
LESLIE, A. W. (1987). *Acta Cryst.* A**43**, 134–135.
MCGRATH, M. E., ERPEL, T., BROWNER, M. F. & FLETTERICK, R. J. (1991). *J. Mol. Biol.* **222**, 139–142.
MCGRATH, M. E., VASQUEZ, J. R., CRAIK, C. S., YANG, A. S., HONIG, B. & FLETTERICK, R. J. (1992). *Biochemistry*, **31**, 3059–3064.
SERC Daresbury Laboratory (1986). *CCP*4. *A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington, England.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.
WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–111.
WILSON, C. & AGARD, D. A. (1993). *Acta Cryst.* A**49**, 97–104.
WILSON, C., WARDELL, M., WEISGRABER, K. H., MAHLEY, R. W. & AGARD, D. A. (1991). *Science*, **252**, 1817–1822.
ZHANG, K. Y. J. & MAIN, P. (1990). *Acta Cryst.* A**46**, 377–381.