Acta Cryst. (1996), D**52**, 469–480

A Challenging Case for Protein Crystal Structure Determination: the Mating Pheromone Er-1 from *Euplotes raikovi*

DANIEL H. ANDERSON, MANFRED S. WEISS AND DAVID EISENBERG

UCLA-DOE Laboratory of Structural Biology and Molecular Medicine and Molecular Biology Institute, University of California, Los Angeles, Box 951570, Los Angeles, CA 90095-1570, USA. E-mail: david@ewald.mbi.ucla.edu

(Received 25 May 1995; accepted 13 October 1995)

Abstract

Four different phasing methods have been applied to the determination of the crystal structure of the 40 amino-acid mating pheromone of the unicellular ciliated protozoan *Euplotes raikovi*. The difficulties, failures and successes in attempts to solve the structure by: (1) molecular replacement, (2) direct phasing using the '*Shake and Bake*' algorithm, (3) isomorphous replacement, and (4) multiple-wavelength anomalous dispersion are described. The structure was first solved by molecular replacement, and then was the first successful structure determination by '*Shake and Bake*' without the direct involvement of its authors. A description of the current status of the high-resolution refinement of the structure is also given. The model is refined against 1 Å resolution data to an *R* factor of 12.9%, and includes H atoms and discretely disordered side chains.

1. Introduction

Despite technological and computational advances, protein crystal structure determination is still far from becoming a simple automated task. The protein crystallographer's repertoire for phase determination currently consists of four approaches: (i) single or multiple isomorphous replacement with or without anomalous dispersion (SIR, MIR), (ii) molecular replacement (MR), (iii) multiple wavelength anomalous dispersion (MAD), and (iv) direct methods. For a protein without a known homologous structure, the SIR and MIR methods (reviewed by Watenpaugh, 1985) are still by far the most widely used because they do not require a trip to a synchrotron, and are not biased by any assumptions about the structure. In the case where a homologous structure is known, the first approach to try is usually the MR method (reviewed by Lattman, 1985), which does not rely on the tedious search for heavy-atom derivatives. Especially in combination with Patterson correlation refinement (Brünger, 1993), MR has become very powerful. A third technique, originally developed by Hendrickson (1991) is MAD phasing. The presence of one heavy atom in the asymmetric unit can be sufficient to yield phases, when the data are measured at three properly chosen wavelengths. Direct phasing methods are widely used in small-molecule crystallography, but rarely used for proteins because they have been limited to about 100 atoms per asymmetric unit, and also because the methods require higher resolution data (about 1.1 Å) than is usually obtainable for most protein crystals.

The subject of this study is a 40 amino-acid peptide pheromone from the unicellular ciliated protozoan *Euplotes raikovi*. The biological function of this protein is context dependent. It can serve either as an autocrine growth factor, or as a paracrine signaling factor in the process of cell adhesion leading to mating-pair formation (Luporini, Vallesi, Miceli & Bradshaw, 1995; Ortenzi & Luporini, 1995).

The three-dimensional solution structure of Er-1 was solved by NMR (Mronga *et al.*, 1994). It consists of a three-helical bundle of up–down–up topology, linked by three disulfide bridges. The crystal form that was used for the structural investigations in this paper diffracts to 1 Å resolution, but the extreme packing density (V_M = 1.53 Å^3 Da^{-1}), and the near-spherical shape of the molecule hindered the structure determination. Structure determination has proceeded along four pathways: isomorphous replacement, multiple-wavelength anomalous dispersion (MAD), molecular replacement, and direct phasing. We present positive and negative results with each of these methods. The crystal structure and its biological significance is described in detail elsewhere (Weiss *et al.*, 1995).

2. Crystallization and the crystals

Er-1 has been crystallized in three different forms. Crystals of forms 1 and 2 diffract to about 2.8 Å resolution, and their growth has been described previously (Anderson, Raffioni, Luporini, Bradshaw & Eisenberg, 1990). Form 1 crystals are of space group $P6_1$ (or $P6_5$), with a = 86.6, c = 44.6 Å. Form 2 crystals are of space group $P4_12_12$ (or $P4_32_12$), with a = 64.7, c = 97.1 Å. The growth of crystal forms 1 and 2 was very unreliable. The form 3 crystals diffracted much better than forms 1 and 2. Although it was difficult to obtain single crystals of form 3, many of the diffraction patterns could be indexed, and at least one crystal grew in each drop.

© 1996 International Union of Crystallography
Printed in Great Britain – all rights reserved

Table 1. *X-ray data collected from native and derivatized form 3 crystals of Er-1*

$R_{sym} = 100 \sum_h |I_i - \langle I \rangle| / \sum_h \langle I \rangle$. $R_{scale} = 100 \sum_h ||F_1| - |F_2|| / \sum_h |F_1|$.

Crystal type	Wavelength	Detector*	No. Obs	No. unique	R_{sym} (%)†	R_{scale} (%)	Resolution (Å)	Completeness (%)
Native 99	2.291	MAR	5077	731	4.5		26–2.6	84.0
Native 99	0.711	MAR	101188	13461	5.4		8–1.0	92.4
Native 80	1.542	R-AXIS	18539	4187	3.9	6.1‡	26–1.5	96.7
Native 100	1.542	Multi	21793	1427	3.6	5.2‡	26–2.2	99.7
Composite§				13659			26–1.0	93.8
Hg98	1.542	Multi	7714	1253	4.0	16.2¶	26–2.2	87.0
Hg MAD λ_1	1.010	MAR	18658	4156	6.4		26–1.5	94.5
Hg MAD λ_2	1.006	MAR	18915	4158	7.2	25.2¶	26–1.5	94.5
Hg MAD λ_3	0.95	MAR	19156	4180	8.6		26–1.5	95.0
Native 70	1.542	R-AXIS	12660	3943	3.6		26–1.5	90.6
Hg86	1.542	R-AXIS	18248	4285	6.9	19.2¶	26–1.5	98.1

* The data sets were measured by and processed with the following equipment and software: Rigaku R-AXIS II imaging plate with Rigaku RU200 generator and MSC software; San Diego multiwire detectors (Multi), using a Rigaku RU200 generator, processed by Nielsen's software (Howard *et al.*, 1985). One data set from crystal 99 was measured with Cr $K\alpha$ radiation by Xiaoping Dai, with an Elliot GX-6-generator, and an 18 cm MAR detector. To correct for the strong absorption of Cr $K\alpha$ radiation, a Mo $K\alpha$ data set was measured from the same crystal with the same data-collection geometry, and processed with *CCP*4 (SERC Daresbury Laboratory, 1986) software. The Cr $K\alpha$ observations were then scaled to the corresponding Mo $K\alpha$ observations with a dynamic block scaling algorithm (Dai, 1995). The same crystal 99 was then used for a second data set measured with Mo $K\alpha$ radiation. This data set extended to 1 Å resolution and was used in the structure determination. The multiple-wavelength data for MAD phasing were measured at synchrotron beamline X12c at Brookhaven National Laboratory, using a 30 cm MAR imaging plate, and processed with *DENZO* (Otwinowski, 1993) and *CCP*4 (SERC Daresbury Laboratory, 1986) software. † Bijvoet-related reflections are merged. R_{sym} is lower when anomalous pairs are not merged. ‡ R factor for scaling the native data sets from crystals 80 and 100 to that from crystal 99. § The composition data set contains mostly data from crystal 99, with missing data filled in by crystals 80 and 100. ¶ Hg crystal 98 data set is scaled to native 100; Hg86 data set is scaled to native 70; Hg λ_2 MAD data are scaled to the composite native.

Therefore, we proceeded with structure determination with the form 3 crystals. Er-1 form 3 crystals were grown by hanging-drop vapor diffusion using the following protocol. The lyophilized sodium salt of Er-1 was dissolved in high-performance liquid chromatography grade water then mixed with precipitant solution to an initial drop composition of 4.4 mg ml^{-1} Er-1, 0.068 *M* $(NH_4)_2SO_4$, 0.028 *M* Na citrate buffer, 3.33%(*v*/*v*) ethanol. The stock buffer solution was at pH 3.50 prior to mixing with the other drop components. This solution of Er-1 with precipitant was then centrifuged (10 000*g* for 5 min). Drops of 10 µl each, on glass cover slips siliconized with Aquasil (Pierce Chemical) were equilibrated with 1 ml reservoirs containing 0.12 to 0.15 *M* $(NH_4)_2SO_4$, 0.05 *M* Na citrate, pH 3.50, 6%(*v*/*v*) ethanol. The optimum temperature for crystallization was between 298 and 301 K. Crystals first appeared in about 5 d, but continued to grow for about 2 weeks, to a maximum size of about 0.6 × 0.4 × 0.3 mm. The ethanol is not required for crystal growth, but its presence encourages growth of single crystals, rather than clusters.

The space group and cell dimensions were determined independently by examination of precession photographs, by a Rigaku AFC5 diffractometer, and by each of the machines listed in Table 1. The space group is $C2$, with $a = 53.91$, $b = 23.08$, $c = 23.11$ Å, $\beta = 110.4°$. Assuming that the asymmetric unit contains a monomer of molecular weight 4410 Da, the Matthews parameter V_M (Matthews, 1968) is 1.53 Å^3 Da^{-1}, far smaller than the normal distribution of V_M for other hydrated protein crystals. The solvent content is less than 20%; there is enough space for only about 44 water molecules in the asymmetric unit.

The physical properties of Er-1 form 3 crystals are unlike any previously encountered in our laboratory. They are the most dense and hardest protein crystals in our experience. They sink in a saturated solution of $(NH_4)_2SO_4$, implying a density of at least 1.24 g ml^{-1}. The crystals frequently grow in clusters. We have never successfully separated individual crystals from a cluster. They remain intact despite aggressive probing with a glass needle, or they shatter into useless fragments. The diffraction lifetime in the X-ray beam is very long. One crystal (number 99) was used for the chromium and the molybdenum 1 Å data sets (see below), and was used until about 1 year later as a test crystal for new software and detectors in the laboratory of Dr Xuong at UCSD. Despite or because of the extremely dense packing and strength of the form 3 crystals, they crack or shatter with osmotic changes or addition of some heavy-atom reagents.

In order to verify that these unusual crystals contain protein, some of them were washed with 0.4 *M* $(NH_4)_2SO_4$, 0.05 *M* Na citrate, pH 3.5, then subjected to amino-acid analysis. The composition of the crystals is as expected for this amino-acid sequence (Simona Raffioni, personal communication).

3. Data collection

Diffraction data for form 3 Er-1 crystals were measured with four X-ray sources, five different detectors, and

processed by five software packages. Some statistics about the diffraction data used in this work are presented in Table 1. Almost all of the Er-1 data sets were of very high quality, as judged by R_{sym}. In the early stages of the structure analysis, data sets measured from individual crystals were used because the overall completeness was high for each (Fig.1). However, as we show later, the success of the structure determination depended on the use of a data set with even higher completeness. To achieve this, a composite data set was constructed from three native data sets. Most of the reflections in the composite were measured from crystal 99. Data sets from crystals 80 and 100 were selected to fill in the missing regions of reciprocal space by the quality of their anomalous Patterson functions: peaks representing the sulfur structure were similar to those obtained from the Cr $K\alpha$ data set measured from crystal 99. Crystal 80 was grown in $(NH_4)_2SeO_4$ instead of $(NH_4)_2SO_4$ in an attempt to produce an isomorphous derivative, but was at the time one of the best native data sets. Crystals 80 and 100 data sets were scaled to the crystal 99 1 Å data set with *ANISOSC* (SERC Daresbury Laboratory, 1986), then the crystal 100 data were read by *X-PLOR* (Brünger, 1993), followed by the crystal 99 data. Therefore, the crystal 99 data set displaced all reflections that were in common between the two. This resulted in a composite data set that is almost complete at low resolution, and also extends to very high resolution. Similarly, *X-PLOR* read the crystal 80 data set followed by the crystals 99+100 composite to produce the final composite used for molecular replacement and direct phasing. Statistics for this scaling and merging procedure are presented in Table 1, and Fig.1.

4. Initial phasing by isomorphous replacement

Our first attempt to solve the structure of Er-1 was by the multiple isomorphous replacement method (reviewed by Watenpaugh, 1985). Because of the extremely dense packing of the form 3 crystals, it was difficult to obtain suitable heavy-atom derivatives. Analogously, no derivatives could be found for the less densely packed crystals of crambin (Hendrickson & Teeter, 1981). Still, we were encouraged to try preparing derivatives by soaking in heavy-atom solutions because the crystals could be cracked by sudden osmotic changes; they must have solvent channels. About 27 different heavy-atom compounds were tried, some destroying the crystals, most not incorporating. For example, the crystals are simultaneously stained and cracked by exposure to $PtCl_4$, $K_2Pt(SCN)_4$, or KI_3. The crystals did not incorporate these reagents when soaked at concentrations low enough to avoid cracking. Also, no crystal grown in the presence of heavy-metal compounds incorporated the reagents. Data were collected from crystals that appeared to survive soaking in more than 17 heavy-atom reagents.

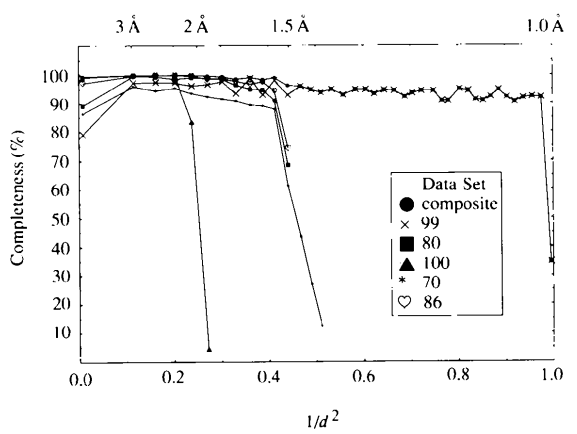


Fig. 1. Completeness of data sets 80, 99, 100, and of the resulting composite data set. Data sets 70 and 86 are also included. In the resolution range used for the rotation function, the composite data set has 48 more reflections than data set 99.

The only heavy-atom reagent found that could non-destructively bind the Er-1 crystals was dimethylmercury (DMM, Aldrich Chemicals). The vapor-phase soak was performed by the method of Ward Smith (personal communication, 1980). The crystal was mounted and dried in a thin-walled capillary, then a fragment of filter paper was inserted into the capillary. Using a 10 µl syringe in a fume hood, 1 µl of DMM was added to the filter paper, and the capillary was sealed with a rapid-hardening epoxy. The reagent is volatile and enters the crystals quickly: the transition in its fluorescence spectrum could be measured within 10 min (see *MAD*, below).

DMM data sets were measured from crystals 86 and 98, then scaled to native data sets from the same detectors (Table 1). The single Hg site [fractional coordinates (0.18, *y*, 0.0)] was determined by inspection of isomorphous and anomalous difference Patterson functions. Refinement and phasing were performed by *HEAVY* (Terwilliger & Eisenberg, 1983) for R-AXIS data, or *PHARE* (SERC Daresbury Laboratory, 1986) for multiwire data. The *PHARE* refinement statistics for all the multiwire data to 2.5 Å were [see Watenpaugh, 1985, for definitions; in that reference, 'phasing power' is $1/R$(modulus)]: $R_{Cullis} = 0.584$, figure of merit = 0.66. Phasing power is 2.2 for all but the eight reflections rejected from refinement. The average difference between the Hg SIR + anomalous and 1.6 Å refined model phases (Weiss *et al.*, 1995) is 68.4°, for all data from 26 to 2.5 Å [with the Hg site at fractional coordinates (0.180, 0.255, 0.495), see *Molecular replacement*, below]. The map shown in Fig. 2(*a*) resulted from multiwire data sets 100 and 98.

In an attempt to improve the phasing by maximizing the anomalous differences, the second wavelength MAD data (see below) were processed as a SIR + anomalous derivative, using the composite native data set and *PHARE* software. The phasing statistics were: $R_{Cullis} = 0.677$, figure of merit = 0.5, phasing

power = 1.22. The average difference between these optimized SIR + anomalous phases and the refined model phases was 74.7°

5. Initial phasing by MAD

Data at three wavelengths were collected from a dimethylmercury-containing crystal, in order to calculate phases by the MAD method (Hendrickson, 1991). The dimethylmercury vapor soak was performed as described above. Data were collected at beamline X12c at Brookhaven National Laboratory (Table 1). The wavelengths chosen for the transition, peak, and distant data were determined from a fluorescence spectrum around the L_{III} absorption edge, and were 1.00965, 1.00568 and 0.95 Å, respectively. MAD phasing was carried out with *HEAVY* software, which incorporates the algorithm described in Terwilliger (1994). The MAD phasing statistics for all data between 26 and 2.5 Å were: figure of merit = 0.43; R_{Cullis} = 0.76; phasing power = 0.61. The average difference between the MAD and the 1.6 Å refined phases (Weiss *et al.*, 1995) was 68.7° for all data from 26 to 2.5 Å. The MAD-phased map is shown in Fig. 2(*b*).

6. Initial phasing by molecular replacement

We tried to solve the Er-1 crystal structure by molecular replacement with four programs, seven search models, and four data sets. For an overview, see Table 2. In no attempt did we obtain an outstanding solution. Therefore, every suspected solution was checked against a (F_{Hg} – F_{native}) difference Fourier to see whether a peak could be found at a position consistent with the mercury isomorphous and anomalous difference Patterson functions. The mercury data set was from crystal 86 (Table 1), and the native sets were as specified below. The first search model that we used was measured from a stereo picture of the NMR structure of the related pheromone Er-10 (Brown *et al.*, 1993), using the program *STEREO* (Rossmann & Argos, 1980). The energy of this model was minimized with *X-PLOR* (Brünger, 1993), and it

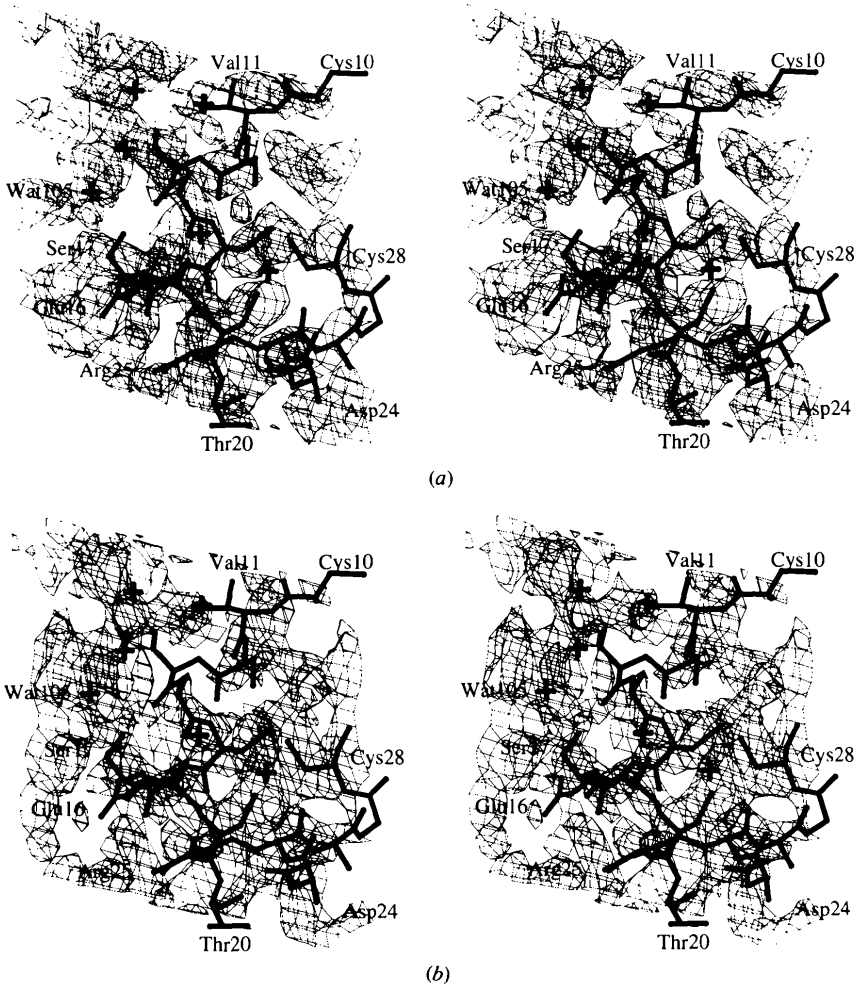


Fig. 2. Representative sections of electron-density maps, displayed in stereo, all in the same region, with phases from four sources. The stick figures represent atoms in the 1.0 Å refined model: the segment from Cys10 to Thr20, including two turns of helix 2, and a portion of helix 3, from Asp24 to Cys28. Water molecules are represented by + marks. Water 105 and the water adjacent to it are related by a twofold rotation. No other symmetry-related atoms are displayed. (*a*) Map phased by the dimethylmercury derivative with anomalous dispersion, using multiwire data sets 100 and 98, 10–2.5 Å resolution, contoured at +0.85σ. This view provides examples of how the many breaks and misconnections in the density rendered this map uninterpretable. If the contour level is set to 0.7σ, the helix becomes connected at Val11, but a misconnection occurs between Val11 and the symmetry-related molecule in the upper left corner (not displayed). A loop of density connects the side chains of Glu16 and Ser17. (*b*) MAD-phased map, 10–2.5 Å resolution, contoured at +0.85σ. The MAD map was even less interpretable than the SIR + anomalous map.

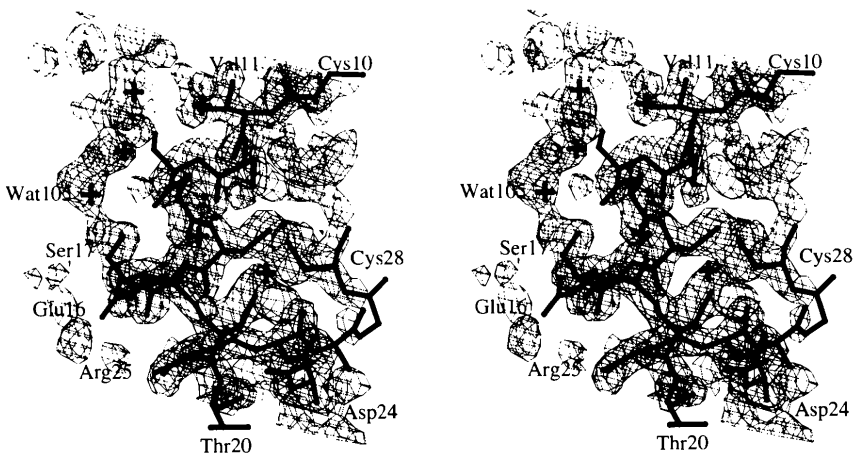
Table 2. *Summary of molecular-replacement trials*

Of the many tried, the successful combination of column entries was: Er-1 mean model, *X-PLOR* software, composite data.

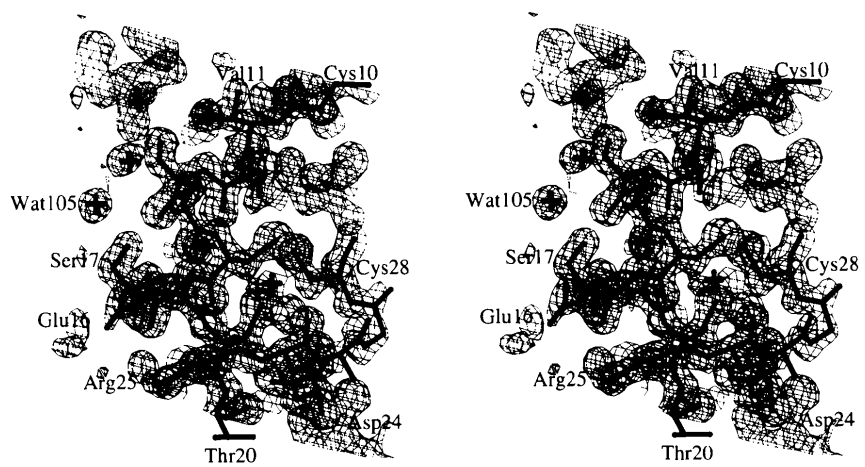
Search models	R.m.s. to final structure* (Å)	Programs†	Data sets‡
Er-10			
From stereo (energy minimized)	1.39 (36 C_α)	*ALMN*	Native 70
		TSEARCH	Native 99
Top of PDB list	1.26 (36 C_α)	*AMORE*	Composite of 80, 99, 100
Mean PDB model	1.30 (36 C_α)	*X-PLOR*	Native anomalous ΔF
		Manually	
Er-1 from Wüthrich			
Top of PDB list	1.26 (40 C_α)		
Mean PDB model	1.17 (40 C_α)		
Regular α-helix (Ala_8)	0.14 (8 C_α)		
7 S atoms	1.60 (7 S)		

* Final structure is the 1.6 Å refined model (*X-PLOR* round 5 in Table 4), described in Weiss *et al.* (1995). † Program references: *ALMN* and *AMORE* (Collaborative Computational Project, Number 4, 1994); *X-PLOR* (Brünger, 1993); *TSEARCH* (SERC Daresbury Laboratory, 1986); manually (see text). ‡ See Table 1.

was used for molecular replacement using data from crystal 70 (Table 1). The attempt failed, which is not surprising because of the rather large r.m.s. difference to the Er-1 structure, and also because of the incompleteness of the model (we were only able to retrieve main-chain coordinates from the picture). We later used Er-10 coordinates when they became available in the PDB (Bernstein *et al.*, 1977), but these attempts were also unsuccessful. Although the Er-10 NMR structures are a little closer to the final Er-1 structure than the



(*c*)



(*d*)

Fig. 2 (cont.) (*c*) 26–2 Å resolution map calculated from composite F_{obs}, phases from 302-atom *SnB* solution 143 (composition S_7N_{295}), after an origin shift, contoured at +1.0σ. Almost all of this map is easily interpretable. The misconnection from Ser17 to Wat105 would not generate confusion during model building because the density leads across a twofold to another Ser17. (*d*) 26–1.6 Å resolution ($2F_{obs} - F_{calc}$) map with *SIGMAA* weighting, calculated from composite F_{obs}, phases from the 1.6 Å refined model (Table 4 of this paper; Weiss *et al.*, 1995), contoured at +1.0σ.

coordinates from the stereo figure (1.30 Å r.m.s. rather than 1.39 Å), the r.m.s. difference is still too large to yield a clear solution in the rotation function. An attempt to use an ideal α-helix of just eight residues also failed, probably because of the incompleteness of the model. The r.m.s. difference to residues 2–9 of the final structure is only 0.14 Å, but with only 40 atoms this model barely comprises 1/8 of the content of the asymmetric unit.

When the group of Wüthrich solved the NMR structure of Er-1 we received the coordinates prior to their publication (Mronga *et al.*, 1994). After that, all our molecular-replacement attempts were focused on search models derived from the Er-1 NMR structure. At about the same time, Xiaoping Dai in the laboratory of Dr Xuong collected the native data sets from crystal 99: a 2.65 Å data set on a chromium anode, and a 1.0 Å data set on a molybdenum anode (Table 1). With the chromium data set, we tried to calculate a rotation function using as a probe only the S atoms of the NMR model, and the native anomalous differences as 'F_{obs}' in *X-PLOR*, again unsuccessfuly. 'Manual' molecular replacement using the chromium data set was also unsuccessful (see below). The rotation function against the very high quality 1 Å data set collected from crystal 99 did not yield a successful solution.

After more unsuccessful trials, we suspected that the unusually tight packing of the molecules in the crystal might require optimization of the parameters used in a molecular-replacement experiment. Therefore, a test case was constructed from a 'manual' model (see below) that was packed reasonably well into the unit cell. From this model, a set of F_{calc}'s was computed. As a search model, a coordinate set was generated whose 40 Cα positions had an r.m.s. difference of 0.8 Å from the target structure. Using the test probe and data, and the program *X-PLOR* (Brünger, 1993), the best parameters for the rotation function were found to be a resolution range of 7–3.25 Å, and a Patterson vector length range of 3.25–22 Å. A search model containing all atoms proved superior to just a main-chain model. Even then, the correct rotation-function solution turned out to be only 20% higher than the first wrong peak. In the translation function, using the correlation coefficient between E^2's and the same test model (but a resolution range of 10–3 Å), the correct peak was 70% higher than the first incorrect one. In the test case, the *X-PLOR* packing function could barely distinguish the correct from the highest incorrect solution (2.2σ *versus* 1.9σ). In the translation function, test models containing all atoms or only backbone, C_β, and S_γ atoms scored equally well.

Using the optimized molecular-replacement parameters with real Er-1 data, we were initially unable to find the correct solution. One major difference between the test and real cases was that the data set for the test case was complete, while the real case was based on data with about 90% completeness. Therefore, we combined the data sets of three crystals, in order to eliminate any possible effects from data incompleteness. In the resolution range optimized for the rotation function (7–3.25 Å), data from crystal 70 has 356 reflections, and data set 99 has 351. The composite data set constructed from data sets 80, 99 and 100 (see *Data*, above, and Table 1), contains 399 of the 400 possible reflections in the desired resolution range. Using the composite data set, optimized parameters, and searching with the mean Er-1 NMR model with artificial temperature factors derived from the r.m.s.d. of the atomic positions in the 20 NMR models [$B = 8\pi^2(\text{r.m.s.d.})^2$; constrained to be between 8 and 64 Å^2], the solution in the rotation function using *X-PLOR* was the eighth largest peak. Patterson correlation refinement (PC; Brünger, 1993) with the search model broken into five pieces, elevated the peak from 3.0σ above the mean to 3.2σ. The search model for the translation function contained only main chain, C_β, and cysteine S atoms from the oriented and PC-optimized model. The translation function (*X-PLOR*) gave the same result when either the correlation coefficient between E^2's or the packing function were used as target functions (peak heights 3.6σ and 1.9σ, respectively). The correct peak was not at the top of the list when the translation function was re-run with a search model containing all the atoms. A model-phased ($F_{Hg} - F_{native}$) difference Fourier showed a 6.7σ mercury peak in a position consistent with the Hg Patterson functions (the second largest peak was 3.4σ), at which point we were confident that we had found the correct solution. The mercury site consistent with the origin of the molecular-replacement solution and the subsequent 1.6 Å refined model (Weiss *et al.*, 1995) is at fractional coordinates (0.180, 0.255, 0.495).

In addition to using Patterson-based programs, molecular replacement was also performed 'manually'. Trial sulfur configurations were determined with *HASSP* (Terwilliger, Eisenberg & Kim, 1987) from native anomalous Patterson functions, usually calculated with the chromium data set from crystal 99 (Table 1). For each trial configuration, 27 unit cells of symmetry-related sulfur sites (and their mirror images) were displayed with *FRODO* (Jones, 1978), along with the first Er-1 NMR model from the list of 20 provided by K.Wüthrich. 'Manual' molecular replacement was performed by visual docking of the NMR model S atoms onto the crystallographically determined sulfur sites, while trying to fill the volume of the unit cell with minimal collisions. This approach probably failed because of the 1.6 Å r.m.s. difference between the sulfur structures of the NMR and crystal models. No plausible 'manual' packing resulted in a prominent Hg peak in model phased ($F_{Hg} - F_{native}$) difference Fourier functions. Because of the density of these crystals, any placement of the model that has minimal collisions will overlap the correct structure. At least two of these 'manual' models could be refined to deceptively low R factors, but R_{free} was random, and

plausible Hg peaks were weak and skipped between origins during refinement.

7. Initial phasing by direct methods

Phases for the Er-1 crystal were independently determined by the '*Shake and Bake*' (*SnB*) direct-phasing program (Miller, Gallo, Khalak & Weeks, 1994). We used *SnB* version 1.0.0, compiled and run on a Digital Equipment Corporation Alpha (operating system OSF/1, version 3.0). This version of *SnB* differs from that described in Miller *et al.* (1994) by the addition of an *E*-Fourier refinement of each output model. In our initial attempts, *SnB* failed to find solutions with the 1 Å data set from crystal 99. *SnB* was re-run with the composite data set constructed for molecular replacement (see Table 1, and *Data*, above). We ran *SnB* for 1000 random trial structures in order to identify promising solutions and to provide statistics for the success rate. All the input parameters except the number of peaks to pick were default values chosen by *SnB* (see Table 2 in Miller *et al.*, 1994). Each starting model contained 302 random atoms, but only 240 peaks were picked during the 'bake' cycle. Phase refinement during the 'shake' cycle was performed by the parameter-shift algorithm, with 90° steps. Each 155-cycle Shake-and-Bake calculation was followed by four cycles of *E*-Fourier refinement, picking 240, 260, 280 and then 302 peaks. The average computation time per trial structure was 23.0 min.

A histogram of the results of the 1000 trials is shown in Fig. 3. The *SnB* process appears to reduce the minimal-function values (R_{min}) to about 0.4, but the R_{min} of only a few structures did not diverge during the *E*-Fourier cycling. The phasing ability of the six solutions with the lowest minimal function values were analysed by four criteria: conventional *R* factor, phase difference to the refined phases, Hg peak height in a model phased ($F_{Hg} - F_{native}$) difference Fourier, and subjective map quality. The numerical results are shown in Table 3. For

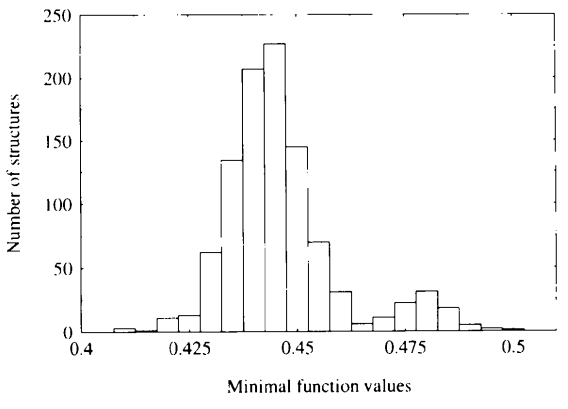


Fig. 3. Histogram of results from 1000 trial structures processed by *Shake and Bake*. Only the six structures with the lowest minimal function values were further analyzed, and the best four of them were found to be correct solutions.

Table 3. *Phasing statistics of the SnB models with the six lowest minimal function values*

R^*_{min}	Trial number	*R* factor† (%)	Hg peak (σ)‡	Phase†‡ difference	Hand of helices
0.412099	564	42.8	9.7	50.3	Left
0.412358	755	42.5	9.6	50.8	Left
0.416939	143	40.9	9.0	49.5	Right
0.417539	78	42.1	7.0	51.4	Left
0.422560	944	51.9	4.5	89.3	?§
0.423117	25	51.7	None¶	—	—

* Minimal function value R_{min} is that after *E*-Fourier refinement. The minimal function is described in Miller *et al.* (1993). †Although *SnB* phased only 3020 reflections from throughout the 26–1 Å resolution range, the *R* factor and the phase difference relative to the 1.6 Å refined model (Weiss *et al.*, 1995) are evaluated for all data from 26 to 2 Å. For calculation of F_{calc} and phase with *X-PLOR* , the *SnB* models were treated as seven S atoms, and the rest N atoms, all with equal temperature factors. ‡The ($F_{Hg} - F_{native}$) difference Fourier was calculated between 10 and 2.5 Å resolution, using Hg data from crystal 86, and the composite native data set. Prior to calculation of the phase difference and evaluation of map quality, the left-handed helices were inverted through a mirror at $y = 0$, and then the *SnB* models were placed on an origin equivalent to that of the refined structure so that phases calculated from all the models would produce coincident Hg peaks in difference Fourier maps [at fractional coordinates (0.180, 0.255, 0.495)]. § Using phases from trial 944, a difference Fourier peak near a plausible Hg site was third largest. The map calculated from composite native *F* and *SnB* model 944 phases had no recognizable connectivity, and the peak may have appeared near the right place by chance. ¶Model 25 did not produce a Hg difference peak and was not examined further.

phase calculation with *X-PLOR*, the first seven atoms in each *SnB* model (the seven highest peaks in the map) were assigned to be S atoms, and the rest N atoms. All the *SnB* atoms were assigned equal temperature factors. For evaluation of map quality, 26–2 Å resolution maps were calculated with F_{native} coefficients from the composite data set, and phases calculated from the *SnB* models. Maps with ($2F_{obs} - F_{calc}$) coefficients were calculated for *SnB* models 143, 564, and 755, but were worse than the F_{native} maps in almost all parts of the structure. All the maps were displayed and examined with *FRODO* (Jones, 1978), with and without the refined atoms displayed, to judge if they would be interpretable in the absence of the known structure. The *SnB* solutions with the four lowest R_{min} values gave subjectively very good quality maps (see Fig. 2*c*). Even with the *SnB* map on a different origin and different hand than the refined model, the helices and many side chains could be identified in a few minutes. Electron density was absent or fragmented for mobile atoms, such as the N-terminus, some side chains, and the region around Pro37-Pro38. Where there was a break in density in one *SnB* map, phases from another model were able to compensate. The map calculated from *SnB* model 143 was subjectively the most interpretable of the 302 atom models.

During the 1000-trial *SnB* calculation, the program searched the maps for 240 peaks at the 'bake' step, then recovered 302 atoms in the *E*-Fourier refinement. The

Table 4. *Er-1 refinement history and statistics*

Job	Description	Resolution (Å)	N_P*	N_H*	N_S*	N_{para}*	R†	R_{free}
1	*X-PLOR*‡ round 1	10–2.2	301	—	—	1204	27.7	39.0
2	*X-PLOR* round 2	10–2.0	295	—	—	1180	23.8	32.0
3	*X-PLOR* round 3	10–1.8	295	—	5	1200	21.4	28.6
4	*X-PLOR* round 4	10–1.6	303	—	9	1248	19.9	25.4
5	*X-PLOR* round 5	26–1.59	303	—	25	1310	19.6§	24.2
5*a*	*SHELXL*, repeat *X-PLOR* round 5	26-1.59	303	—	25	1311	17.8	23.4
6¶	*SHELXL*** no water	26–1.0	303	—	—	1213	24.20	26.35
7	*SHELXL*	26–1.0	303	—	4	1229	22.52	25.13
8	*SHELXL*	26–1.0	303	—	9	1249	22.12	24.68
9	*SHELXL*, add *SWAT*	26–1.0	303	—	13	1266	21.30	23.11
10	*SHELXL*, anisotropic refinement	26–1.0	303	—	17	2882	15.93	19.80
11	*SHELXL*, change *DEFS*	26–1.0	303	—	17	2882	14.96	19.20
12	*SHELXL*, 4 disordered sidechains	26–1.0	319	—	20	3048	13.92	18.25
13	*SHELXL*, add hydrogens	26–1.0	319	1029	20	3048	13.42	17.14
14	*SHELXL*, adjust model, more solvent	26–1.0	319	1039	30	3139	12.91	16.69
15	*SHELXL*, adjust model	26–1.0	319	1039	28	3121	12.74	16.58
16	*SHELXL*, adjust weighting	26–1.0	319	1039	28	3121	12.86	16.45
17	*SHELXL*, all data	26–1.0	319	1039	28	3121	12.92	—

* N_P, N_H and N_S are the number of protein, H and solvent atoms, respectively. Starting at job 14, the model includes two half-occupied ethanol molecules. N_{para} is the number of refineable parameters. † $R = 100 \cdot \sum_h ||F_{obs}| - |F_{calc}|| / \sum_h |F_{obs}|$. R_{free} is the same, but only evaluated for a test set of 10% of the data that were excluded from the refinement. ‡ A typical round of *X-PLOR* refinement consisted of sequential positional, slow cool, positional, individual restrained *B*-factor and positional refinements. § The *R* factor for all data including the test set was 19.9% (Weiss *et al.*, 1995). ¶ The starting model for *SHELXL* job 6 was the result of *X-PLOR* round 5, not job 5*a*. ** A round of *SHELXL* refinement consisted of 20 cycles of conjugate-gradient refinement.

refined Er-1 models contain 303 protein atoms and about 20–25 water molecules. *SnB* found some of the water molecules with low temperature factors, but because it was restricted to a maximum of 302 peaks, it did not find some of the protein atoms. We tried to improve the *SnB* structures with the four lowest R_{min} values by picking ten more atoms during each of nine cycles of the *E*-Fourier refinement: 240, 250, 260, . . . , 320 atoms. On average, picking more peaks (without increasing the temperature factors of the new atoms) slightly increased the phase errors and *R* factors of the *SnB* models, but also marginally increased the Hg peak heights. Some side-chain density was improved by the 320 atom models. Subjectively, *SnB* model 564 (average phase difference 48.1°) resulted in the best of the 320 atom maps, but was not much more interpretable than the 302-atom model 143 map.

8. Refinement of the model

The Er-1 model has been refined against 1.6 Å data with *X-PLOR* (Brünger, 1993) to an *R* factor of 19.6% (described in Weiss *et al.*, 1995), and then refined against 1.0 Å data with *SHELXL* (Sheldrick & Schneider, 1996) to an *R* factor of 12.92%. The history and results of these refinements are summarized in Table 4. The progress of the *X-PLOR* refinement was monitored by *R* factor and R_{free}, and by the height of the Hg peak in a model phased $(F_{Hg} - F_{native})$ difference Fourier function. In the early stages of refinement, fluctuations in R_{free} resulting from the small number of reflections limited its usefulness as an indicator of the progress of refinement. Complete cross-validation was not undertaken. The mercury data set for the difference Fourier calculation was from crystal 86, and the native was the composite data set (see Table 1). All the maps used for displaying and rebuilding the model during *X-PLOR* refinement were calculated from *SIGMAA*-weighted $(2F_{obs} - F_{calc})$ and $(F_{obs} - F_{calc})$ coefficients (Read, 1986), and displayed using the program *FRODO* (Jones, 1978). The maps from *SHELXL* refinement were unweighted.

Molecular replacement succeeded before '*Shake and Bake*', and refinement was, therefore, initiated with the correctly oriented and placed mean Er-1 NMR structure, without the carboxy terminal O atoms, but with the same artificial temperature factors calculated for the rotation-function search model (see above). Using phases from the molecular-replacement model, an $(F_{Hg} - F_{native})$ difference Fourier revealed a 6.7σ peak at the expected mercury site. Rigid-body refinement using data in the resolution range 10–3 Å reduced the *R* factor from 56.7 to 54.4% (R_{free} from 49.4 to 48.8%). The resolution was gradually increased to 2.2 Å during the first refinement round, which consisted of a total of 240 cycles of least-squares refinement, a slow-cooling stage from 2000 to 100 K, 100 cycles of least-squares refinement, ten cycles of restrained temperature-factor refinement, and again 200 cycles of least-squares refinement. After this round the *R* factor was 27.7%, R_{free} was 39.0%, and the mercury peak in the $(F_{Hg} - F_{native})$ difference Fourier increased to 9.1σ (next highest was 4.3σ). In the next four refinement rounds, the resolution was increased to 1.59 Å, all the low-resolution data were included, both termini of the structure, as well as several side chains were rebuilt, and a total of 25 water molecules were inserted according to three criteria: there was at

least 1σ density in the $(2F_{obs} - F_{calc})$ map, at least 3σ density in the $(F_{obs} - F_{calc})$ map, and plausible hydrogen-bond partners were reasonably close. The structure after *X-PLOR* round 5 (Table 4) was described in Weiss *et al.* (1995).

Since the model was built and refined without the use of the anomalous-dispersion data, we used an anomalous-difference Fourier as a further independent check for correctness of the model (Strahs & Kraut, 1968). The coefficients were $(F^+ - F^-)$ from the Cr/MAR native data set (crystal 99), the phases were the 1.6 Å refined model phases minus 90°. The Fourier coefficients (as *A* and *B*) were generated by the *MAPS* function of the *HEAVY* software package (Terwilliger, unpublished work), then the Fourier summation was carried out with *CCP*4 (SERC Daresbury Laboratory, 1986) programs. An entire asymmetric unit of the map, contoured at $+3\sigma$, is shown in Fig. 4, along with all the atoms in the model. Peaks appear only at or near the positions of the seven S atoms, showing that the model S atoms are correctly placed.

For high-resolution refinement, we switched to the program *SHELXL* (Sheldrick & Schneider, 1996), which was originally developed for the refinement of small molecules, but can also be used for proteins when high-resolution data are available. Sheldrick & Schneider (1996) discuss the advantages of *SHELXL* over other programs. At 1.6 Å, the *R* factor decreased by 1.8% simply by repeating *X-PLOR* round 5 using *SHELXL* (see Table 4). To use the automated water-placement facility of *SHELXL*, we re-started the high-resolution refinement with only the protein structure resulting from *X-PLOR* round 5.

As shown in Table 4, in rounds 6–8 of *SHELXL* refinement, all temperature factors were kept isotropic, and a few water molecules were added at a time, monitoring the progress by R_{free}. In round 9 we added the parameter *SWAT* to model the diffuse solvent region. Despite the small unbound solvent region, this one extra parameter appreciably lowered R_{free} in the subsequent refinements. After that round it was evident from the $(F_{obs} - F_{calc})$ difference Fourier that we had to switch to anisotropic refinement. The seven largest features in the difference map were found too close to S-atom positions to represent anything but anisotropy. The switch to restrained anisotropic refinement was also justified by a concomitant drop in the *R* factor (about 5%), and in R_{free} (about 3%). A few more water molecules were added, and the default standard deviation parameters *DEFS* were increased to 0.03, 0.2, 0.03, and 0.10, as recommended in the manual. This change also lowered R_{free}, and was retained, therefore. By the time of *SHELXL* job 12, clear density had been consistently seen for alternate conformations of the side chains of Glu21, Asp24, Met30 and Ser34. Additionally, water 131 is present only during one conformation of Asp24. A further drop in the *R* factor of about 0.5% was achieved by adding H atoms to the model in *SHELXL* job 13. Note that this does not increase the number of refined parameters because the H atoms 'ride' on their supporting atoms. In *SHELXL* job 14, more water molecules were added, and some side-chain conformations were adjusted. The water molecules near Ser34 were replaced with two half-occupied ethanol molecules whose orientation may be coupled to the conformations of Ser34 OH. In *SHELXL* job 15, more side-chain conformations were slighly adjusted, and finally in job 16, the weighting parameter was systematically varied to optimize R_{free}. The optimum weight was 0.3, the same as the value recommended by the program to produce constant variance of the reflections as a function of the magnitude of F_{calc} (Sheldrick & Schneider, 1996).

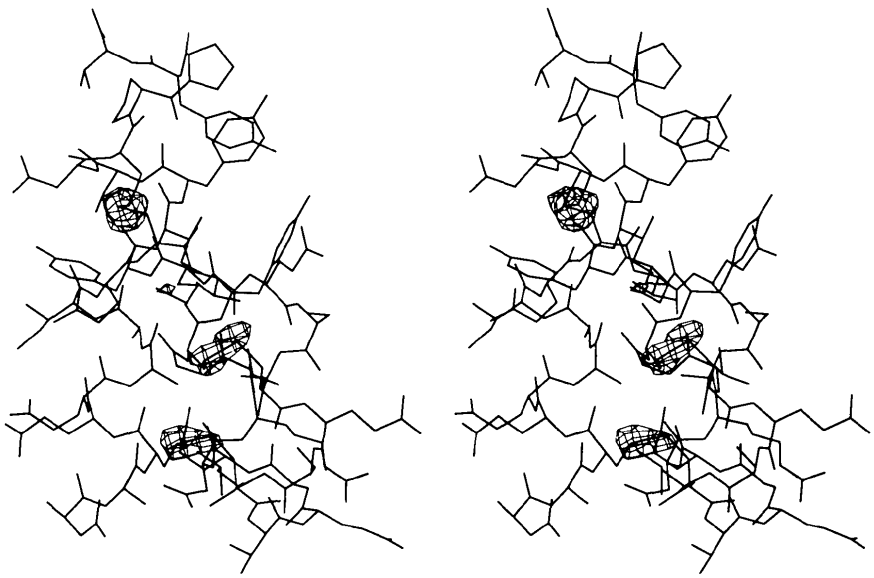


Fig. 4. Stereoview of all the protein atoms in the model of Er-1, with the anomalous Fourier superimposed (contoured at $+3\sigma$). There are peaks only at the locations of S atoms, showing that the S atoms are correctly placed. The view is similar to that in Fig. 1(*a*) of Weiss *et al.* (1995). This figure was prepared with the program *FRODO* (Jones, 1978).

Table 5. *Largest residual peaks in the final* $(F_{obs} - F_{calc})$ *map*

Peak rank	Peak height (e Å^{-3})	(σ)	Nearest non-H atom	Next nearest
1	0.51	5.1	Gln9 NE2 (2.27 Å)	Ala2 N* (2.95 Å)
2	0.50	5.0	Pro38 CA (2.12 Å)	Pro37 CA (2.26 Å)
3	0.47	4.7	Asn35 CG (1.73 Å)	Asn35 CB (1.74 Å)
4	0.46	4.6	Ser34 CA (1.11 Å)	Ser34 N (1.46 Å)
5	0.43	4.3	Water 108 (2.08 Å)	Leu18 O (2.79 Å)

* This Ala2 N atom is related to the model by the symmetry operator $(-x, y, 1 - z)$.

The last refinement round was then repeated with all data. The model refined to a final R factor of 12.92%, and the resulting residual map showed that the model is essentially complete; none of the peaks can be incorporated into the model in a straightforward way, except for the highest peak, which could be an additional water molecule. The five largest features of the $(F_{obs} - F_{calc})$ map are listed in Table 5. The R_{free} value increases when the peak between Gln9 NE2 and a symmetry-related Ala2 N is filled with an additional water molecule, and, therefore, we refrained from doing so. The other positive peaks are too close to protein atoms to allow addition of more atoms.

9. Discussion

9.1. *Monitor functions*

During this work, we used four checks for the correctness of the models. For molecular replacement, the R factor was a poor diagnostic. Whether a solution was essentially correct, or completely wrong, it could be refined to about the same R factor. The crystal is so dense that any packing of the model that has few collisions overlaps the correct structure. R_{free} only became a good monitor of model improvement when the resolution was increased enough to dampen its random fluctuations. For example, R_{free} did not reliably indicate improvement during rigid-body refinement after molecular-replacement because the test set contained only about 50 reflections. In retrospect, complete cross-validation might have improved the diagnostic power of R_{free} at low resolution. The phasing power of each trial MR or *SnB* model was checked by a model phased $(F_{Hg} - F_{native})$ difference Fourier function: a prominent Hg peak meant that the model was essentially correct, and a substantial increase in peak height meant that the phases had improved. However, the Hg site overlaps the side chain of Tyr29 in its native conformation, and, therefore, the Hg peak height depends both on overlap of atoms, and on the phasing power of the model. The native anomalous difference Fourier had significant peaks only on the disulfides and methionine, and served as a further independent check of the refined model.

9.2. *Isomorphous replacement and MAD*

Multiple isomorphous replacement failed for Er-1 form 3 crystals because the packing is so dense that few compounds diffused into them, and those that did either damaged the crystals, or did not bind at all. Dimethylmercury entered the crystals, bound at a single site, and the heavy-atom structure, therefore, contains a center of symmetry. The centrosymmetric Hg structure results in SIR maps with a mirror through the Hg site that is broken only by the anomalous dispersion of the Hg atom. The maps resulting from this derivative are not fully interpretable, therefore. We later learned that the DMM bound by displacing the side chain of Tyr29, the a cell dimension increased by 0.5 Å, and the assumption of isomorphism was not as valid as in other cases, therefore. Refinement of the Er-1 model against crystal 86 Hg data (Table 1) revealed in retrospect that the Hg occupancy is about 50%, and that the Tyr29 conformation is 50% native, 50% alternate.

The best parts of the SIR + anomalous map are the helical regions from residues 13–20 and 25–33. The best part of the 'optimized' SIR + anomalous electron-density map calculated with λ_2 MAD data is the region from residues 26 to 32. With the refined model superimposed on the maps, the meaning of the electron density becomes more evident. But when the eye is not guided by the atoms, tracing the maps is not possible. Even these best parts of the maps become uninterpretable when the model display is turned off.

The MAD-phased electron-density map was subjectively much worse than the SIR + anomalous map, despite the similar average phase errors. The MAD phase errors were larger at low than at high resolution. There are several possible reasons for this poor performance of MAD phasing. Judged by R_{sym}, the MAD data quality is not as good as for other Er-1 data sets (Table 1). The signal strengths as measured by Patterson peaks are apparently not enough for good MAD phasing. The λ_2 anomalous Patterson shows a single non-origin peak of 6σ, and the $(F\lambda_3 - F\lambda_1)^2$ dispersive difference Patterson peak was 4σ. The transition in the fluorescence spectrum of mercury is inherently very broad, limiting the dispersive difference that can be used for MAD phasing. Also, in this case, the heavy-atom structure contains a center of symmetry. In recent cases of successful MAD phasing with Hg, the Hg structures were not centrosymmetric (Hubbard, Wei, Ellis & Hendrickson, 1994; Krishna, Kong, Gary, Burgers & Kuriyan, 1994).

In our experience, experimental phases should be within about 70° of the eventual refined phases to produce an interpretable map. In the case of Er-1, the crystals are so dense that the phase-accuracy requirement is more stringent than usual. The accumulated phase error results in broken density within the molecule, and strong connections between side chains, to water molecules, and to neighboring molecules because of the

almost non-existent boundary regions in the complementary surfaces between symmetry mates. In a less dense crystal, the molecular boundary would focus model building attention on connections and re-connections within the molecule. Improvement of the experimental phases by the usually powerful method of solvent flattening was not possible in this case because about half of the solvent molecules are immobilized on the protein surface, and the remaining solvent volume is very small. Even the additional phase improvement provided by simultaneous histogram matching and Sayre's equation using the *SQUASH* program (Zhang & Main, 1990) were not powerful enough to break the misconnections in the maps.

9.3. *Molecular replacement*

The structure of E*r*-1 was first solved by molecular replacement, although it was a marginal case, and we encountered difficulties at almost every step of the procedure. First, the search model derived from an NMR structure determination had an r.m.s. deviation to the final refined structure of about 1.2 Å for 40 C_α positions. That the energy-minimized mean model derived from the 20 NMR conformers was closer to the final structure than was the top model of the 20 conformers suggests that some among the 20 coordinate sets were even closer. This possibility, however, was not explored.

Because the search model with all atoms included was nearly spherical, it did not help to include the low-resolution reflections that would usually match the overall shape of the search model to the crystal structure. During the systematic optimization of the MR conditions with a test case, we found that reflections below 7 Å had to be omitted to yield the clearest signal. The data sets used in our early MR trials were about 90% complete in the desired resolution range, a completeness usually sufficient to yield a clear solution. In our case, however, near-100% completeness was crucial (399 of 400 reflections in the resolution range 7–3.25 Å). The Patterson correlation refinement in *X-PLOR* was also necessary for the success of this MR project. PC refinement helped to identify the solution, although it only marginally elevated the correct peak (and then only when the search model was broken into five fragments).

The solution in the translation function (*X-PLOR*) was more distinctive than in the rotation function with PC refinement. The translation-function peak was the same when using as target function the correlation between E^2 values, or the packing function, suggesting that the solution was correct. The *R* factors after the translation function were very high, but the $(F_{Hg} - F_{native})$ difference Fourier map showed a clear peak at a plausible Hg position, further supporting the correctness of the solution.

Data sets of much poorer quality and lesser completeness are usually sufficient for MR. The differences between our search model and the crystal structure were not much worse than in other cases. It was probably because of the density of the crystals that these problems impeded structure solution.

As mentioned above, the *R* factor and R_{free} were not clear indicators of a correct solution by molecular replacement. In our case, a much better indicator was the $(F_{Hg} - F_{native})$ difference Fourier map. For an MR problem without the availability of a heavy-atom derivative, a distinctive feature omitted from the search model should appear in an $(F_{obs} - F_{calc})$ difference Fourier when a solution is correct.

9.4. *Direct methods*

The E*r*-1 structure was also solved by the '*Shake and Bake*' algorithm. For generation of initial phases, *SnB* worked much better than SIR + anomalous and MAD. The phase errors from *SnB* were small (about 50°), the resulting electron-density maps were easily interpretable, and the diagnostic Hg peaks were unambiguous. In each correct *SnB* solution, the first six *SnB* atoms (the six highest peaks) were correctly assigned to disulfide S atoms. In no *SnB* model was the seventh atom assigned to the S atom of Met30; because of the temperature factors, other protein atoms have more scattering power. The only disadvantages of *SnB* were the requirement for very high resolution data, and the long computation time (about 16 d on our DEC Alpha). The high completeness of the composite data set was required for successful phasing by *SnB*, as it was for molecular replacement. Although the R_{min} value may have reached minima for the best of the correct solutions, the *E*-Fourier refinement in version 1.0.0 of *SnB* was crucial for clearly distinguishing solutions from non-solutions. We did not investigate the possibility that higher R_{min} structures than those listed in Table 3 may also be solutions of lesser quality.

Direct phasing was the one method that was not impeded by the density of the crystals. A possible reason for the surprisingly high *SnB* success rate of at least 4/1000 could be that a random distribution of atoms can be a good approximation to the mass distribution in such a densely packed crystal. The method would be much less likely to find solutions if the random structure generator placed half of the atoms into the solvent regions of a more typical protein crystal.

9.5. *Refinement*

The molecular-replacement solution has been refined with *X-PLOR* to 1.6 Å and then refined to 1 Å resolution with *SHELXL*. We switched to refinement with *SHELXL* for a variety of reasons, including its automatic water placement, its treatment of the diffuse solvent region, because it refines against F^2, and it provides easy modeling of anisotropy and discrete disorder. During preliminary *X-PLOR* refinement, the resolution was gradually increased to 1 Å to allow re-convergence with every new shell of data. *SHELXL* could incorporate all

new data from 1.6 to 1 Å at once. Also, the *R* factor from *SHELXL* was lower than from *X-PLOR*. Finally, evidence of anisotropy drove us toward modeling with *SHELXL*.

9.6. *Conclusions*

We have described our attempts to solve the structure of the Er-1 form 3 crystals by four methods. Two of them failed, and two succeeded. Good quality images of the molecule resulted independently from molecular replacement and '*Shake and Bake*' direct phasing. Confirming images of lesser quality were produced by SIR + anomalous and MAD phasing. The structure solution by the two most commonly used methods was impeded by the extreme density of these crystals. The lessons that are applicable to less dense crystals are as follows. (i) Dimethylmercury seems to be able to enter crystals through very small openings. Since it does not react with the protein, but merely binds to a hydrophobic depression on the protein surface, it works very gently and should be tried as one of the first candidates in a heavy-atom search. (ii) The phasing powers of mercury derivatives in MAD phasing are not as strong as the phasing powers of equivalent derivatives with other elements because of the broad transition in the Hg fluorescence spectrum. (iii) In a marginal molecular-replacement case, it seems that both a complete data set and systematically optimized MR conditions can pull a solution out of the noise level. (iv) '*Shake and Bake*' is very powerful, and should be tried if the required data are available.

This work has been assisted by the following people and institutions:. Lesa Beamer (critical reading of the manuscript); Robert Blessing (preparation of *E* values); Xioaping Dai (Cr and Mo data from crystal 99); Jennifer Kelly (introduction to *SnB*); Russ Miller (*SnB* help); Simona Raffioni (materials, encouragment, discussion, *etc.*); Bob Sweet (MAD data); Tom Terwilliger (MAD phasing); Kurt Wüthrich (NMR model coordinates). MSW was supported by EMBO, and the overall project was supported by GM-31299. Data from crystal 99 were collected at the University of California, San Diego Research Resource for Protein Crystallography, supported by NIH grant RR01644.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 2ERL, R2ERLSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AM0028).

References

Anderson, D. H., Raffioni, S., Luporini, P., Bradshaw, R. A. & Eisenberg, D. (1990). *J. Mol. Biol.* **216**, 1–2.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.

Brown, L. R., Mronga, S., Bradshaw, R. A., Ortenzi, C., Luporini, P. & Wüthrich, K. (1993). *J. Mol. Biol.* **231**, 800–816.

Brünger, A.T. (1993). *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR.* New Haven, CT, USA: Yale University Press.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dai, X. (1995). PhD thesis, University of California, San Diego, California, USA.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London),* **290**, 107–113.

Hendrickson, W. A. (1991). *Science,* **254**, 55–58.

Howard, A. J., Nielsen, C. & Xuong, N. H. (1985). *Methods Enzymol.* **114**, 452–472.

Hubbard, S. R., Wei, L., Ellis, L. & Hendrickson, W. A. (1994). *Nature (London),* **372**, 746–754.

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

Strahs, G. & Kraut, J. (1968). *J. Mol. Biol.* **35**, 503–512.

Krishna, T. S. R., Kong, X.-P., Gary, S., Burgers, P. M. & Kuriyan, J. (1994). *Cell,* **79**(7), 1233–1243.

Lattman, E. (1985). *Methods Enzymol.* **115**, 55–77.

Luporini, P., Vallesi, A., Miceli, C. & Bradshaw, R. A. (1995). *J. Eukaryotic Microbiol.* **42**(3), 208–212.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.

Mronga, S., Luginbühl, P., Brown, L. R., Ortenzi, C., Luporini, P., Bradshaw, R. A. & Wüthrich, K. (1994). *Protein Sci.* **3**(9), 1527–1536.

Ortenzi, C. & Luporini, P. (1995). *J. Eukaryotic Microbiol.* **42**(3), 242–248.

Otwinowski, Z. (1993). *Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Rossmann, M. G. & Argos, P. (1980). *Acta Cryst.* B**36**, 819–823.

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

Sheldrick, G. M. & Schneider, T. R. (1996). *Methods Enzymol.* In the press.

Terwilliger, T. C. & Eisenberg, D. (1983). *Acta Cryst.* A**39**, 813–817.

Terwilliger, T. C., Eisenberg, D. & Kim, S.-H. (1987). *Acta Cryst.* A**43**, 1–5.

Terwilliger, T. C. (1994). *Acta Cryst.* D**50**, 17–23.

Watenpaugh, K. D. (1985). *Methods Enzymol.* **115**, 3–15.

Weiss, M. S., Anderson, D. H., Raffioni, S., Bradshaw, R. A., Ortenzi, C., Luporini, P. & Eisenberg, D. (1995). *Proc. Natl Acad. Sci. USA,* **92**, 10172–10176

Zhang, K. & Main, P. (1990). *Acta Cryst.*, A**46**, 377–381.