BROUDER, C. (1990). *J. Phys. Condens. Matter*, **2**, 701-738.
DMITRIENKO, V. E. (1983). *Acta Cryst.* A**39**, 29-35.
DMITRIENKO, V. E. (1984). *Acta Cryst.* A**40**, 89-95.
FANCHON, E. & HENDRICKSON, W. A. (1990). *Acta Cryst.* A**46**, 809-820.
HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70-75.
HENDRICKSON, W. A. (1985). *Trans. Am. Crystallogr. Assoc.* **21**, 11-21.
KARTHA, G. (1952). *Acta Cryst.* **5**, 845-846.
KIRFEL, A. & PETCOV, A. (1991). *Z. Kristallogr.* **195**, 1-15.
KIRFEL, A., PETCOV, A. & EICHHORN, K. (1991). *Acta Cryst.* A**47**, 180-195.
LUTZ, H. D., ALICI, E. & BUCHMEIER, W. (1985). *Z. Anorg. Allg. Chem.* **535**, 31-38.
SANDS, D. E. (1982). *Vectors and Tensors in Crystallography*, p. 100. Reading, MA: Addison-Wesley.
SCHWARZENBACH, D. & FLACK, H. D. (1989). *J. Appl. Cryst.* **22**, 601-605.
TEMPLETON, D. H. & TEMPLETON, L. K. (1980). *Acta Cryst.* A**36**, 237-241.
TEMPLETON, D. H. & TEMPLETON, L. K. (1985). *Acta Cryst.* A**41**, 133-142.
TEMPLETON, D. H. & TEMPLETON, L. K. (1986). *Acta Cryst.* A**42**, 478-481.
TEMPLETON, D. H. & TEMPLETON, L. K. (1987). *Acta Cryst.* A**43**, 573-574.
TEMPLETON, D. H. & TEMPLETON, L. K. (1988). *J. Appl. Cryst.* **21**, 151-153.
TEMPLETON, D. H. & TEMPLETON, L. K. (1991). *Acta Cryst.* A**47**, 414-420.
TEMPLETON, L. K. & TEMPLETON, D. H. (1989). *Acta Cryst.* C**45**, 672-673.

Acta Cryst. (1992). A**48**, 751-756

Improving Multiple Isomorphous Replacement Phasing by Heavy-Atom Refinement Using Solvent-Flattened Phases

BY MARK A. ROULD,* JOHN J. PERONA† AND THOMAS A. STEITZ

Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University, New Haven, CT 06511, *USA*

(*Received* 11 *September* 1991; *accepted* 20 *March* 1992)

Abstract

Solvent flattening of macromolecular MIR electron density maps is frequently used to improve the quality of the phases and the interpretability of resultant electron density maps. A new method is presented by which the heavy-atom parameters of isomorphous derivatives are refined against these same solvent-flattened phases and is shown to enhance convergence of the parameters by decoupling heavy-atom-parameter adjustment from parent-phase calculation. This approach is described here in the first example of its application in the solution of the glutaminyl-tRNA synthetase-$tRNA^{Gln}$-ATP co-crystal structure.

Introduction

While providing an unbiased view of a new macromolecular structure, the method of multiple isomorphous replacement often fails to yield interpretable electron density maps due to errors in the observed diffraction data and uncertainties in the estimation of heavy-atom parameters. The native and derivative amplitudes frequently contain errors arising from residual absorption and decay well beyond their estimated standard deviations and the heavy-atom coordinates and occupancies extracted from difference Patterson maps occasionally lie outside the radius of convergence of refinement. Additionally, low-occupancy sites can resist identification. Together these errors culminate in misestimation of the parent phases against which the parameters themselves must be refined. Since the 'lack of triangle closure' that is minimized in the least-squares refinement of heavy-atom parameters requires the parent phases to be calculated, heavy-atom-parameter refinement is coupled to parent-phase refinement. Adjustment of one of the two sets of variables is a dependent function of the other and must be either simultaneously or alternately performed with the other *en route* to convergence (Blow & Matthews, 1973). The problem of 'feedback' is particularly severe in situations where there are a small number of heavy-atom derivatives with closely related heavy-atom sites. Any constraints that can be applied to either the phases or parameters help to decouple the refinement and speed convergence.

Phases produced by back-transforming a MIR electron density map in which the solvent-filled interstices of the crystal have been set to a constant electron density (solvent-flattened) are significantly less biased by the heavy-atom parameters. To the extent that the macromolecular crystal is composed of solvent and

* Present address: Department of Biology/Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

† Present address: Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446, USA.

0108-7673/92/050751-06$06.00

that the solvent boundary can be located, that portion of the unit cell can be considered 'solved', and this information used to constrain the phases (Bricogne, 1976; Schevitz, Podjarny, Zwick, Hughes & Sigler, 1981). To determine the solvent envelope, an electron density map based on the initial MIR phases is scanned for regions of density higher than the average noise (Bhat & Blow, 1982; Wang, 1985). The better the initial phases, the more likely these regions correspond to macromolecular material. Phases obtained by setting the putative solvent regions of the electron density map to some constant value (*i.e.* flattening) and back Fourier transforming can be combined with the initial phases to yield a better estimate of the parent phases. Several iterations of mask determination, solvent flattening and phase recombination improve electron density maps when the solvent content is high. As these improved maps result from phase improvement independent of the heavy-atom parameters, these phases can be used in the further refinement of the heavy-atom parameters.

Procedure for heavy-atom refinement using solvent-flattened phases

A protocol for implementing this procedure is given in flowchart form in Fig. 1. Conventional refinement is performed on each derivative separately, with parent phases initially determined as MIR 'best' phases (Blow & Crick, 1959) using the crude parameters of all the derivatives together. Occupancy and heavy-atom *B* factors are refined alternately as usual. After each derivative has been refined, new parent MIR phases are recalculated and conventional refinement repeated until the parameters do not change. If possible, an anisotropic *B*-factor matrix for each heavy atom is refined in the latter stages. When conventional refinement is complete, MIR phases and phase-probability coefficients (Hendrickson & Lattman, 1970) are calculated using all derivatives. The macromolecular envelope (solvent boundary) is determined and the solvent-flattening constraints are applied to the MIR phases (Wang, 1985).* Examination of electron density maps created using the native amplitudes and the new solvent-flattened MIR (SF-MIR) phases should show improvement over earlier maps using the MIR phases; if not, the variables pertaining to the solvent content and the solvent density used in solvent flattening require optimization. Enforcement of the non-negativity and solvent content of the electron density is accomplished by proper choice of these parameters. In our case, the best maps resulted when parameters were chosen so that 55% of the unit cell was flattened (estimated true solvent content is about 70%) and less than 0.5% of the non-solvent density was truncated as negative. In the fifth step, difference Fourier maps between the heavy-atom derivative and the native data calculated using the SF-MIR phases are used to locate new heavy-atom sites. These maps are superior to the usual cross-difference Fourier maps, in which the derivative whose amplitudes are included in the difference term is omitted from MIR phasing, since the solvent-flattening process greatly reduces this phase bias. After addition of any new heavy atoms, each derivative is again refined using the SF-MIR phases in place of the MIR phases in conventional refinement.* The heavy-atom parameters are refined to convergence without recalculating the parent phases during refinement. When all of the derivatives have been refined to convergence, new MIR best phases are generated and the cycle is repeated by applying solvent-flattening constraints to the new MIR phases. A lack of shift in heavy-atom parameters or of significant difference in the MIR phases signals that the procedure is complete. The last native map using SF-MIR phases provides a view of the macromolecule unbiased by any atomic model.

Application to the synthetase–tRNA structure

The procedure described above was applied to the structure determination of the glutaminyl-tRNA synthetase-$tRNA^{Gln}$-ATP complex† (Rould, Perona, Soll & Steitz, 1989); see Fig. 2. While the diffraction data were of good quality, all four derivatives {$K[AuCl_4]$, di(nitromercurio)acetate (DMA), ethylmercuriophosphate (EMP) and *p*-chloromercuriobenzoate (PCMB)} shared the same heavy-metal binding sites and differed from each other mainly in the relative occupancies of these common sites (see Table 1 of Rould, Perona, Soll & Steitz, 1989). This made the use of cross-difference Fourier maps tedious

* A short program bypasses the routine provided with Wang's solvent-flattening package for combining the phase profiles obtained from each separate derivative and instead directly loads the MIR phase-probability coefficients determined by the refinement program into the program that determines the molecular envelope. It is not clear why the resultant phase-probability profiles are different, but in our experience this gave significantly better results.

* Since *PROTEIN* did not allow parameter refinement using externally supplied parent phases, the program *PHARE* of the Crystallographic Computing Package *CCP*4 (SERC Daresbury Laboratory, 1986) was used for further refinement. *PHARE* does not accept weights, *e.g.* figures of merit, reflecting the reliability of the externally-supplied phases; only standard lack-of-closure error weighting was used in the refinement. All parent phases were used in the refinement, regardless of the figure of merit.

† The *E. coli* glutaminyl-tRNA synthetase-$tRNA^{Gln}$-ATP complex crystallizes in space group $C222_1$ with unit-cell parameters $a = 243$, $b = 94$, $c = 116$ Å. There is one complex per asymmetric unit. The refined model from which calculated structure factors are derived has an *R* factor of 19.6% for all data in the resolution range 6–2.8 Å. The r.m.s. deviation from ideality is 0.012 Å for bond lengths and 2.6° for bond angles. Tightly restrained individual *B* factors are applied to each non-hydrogen atom.

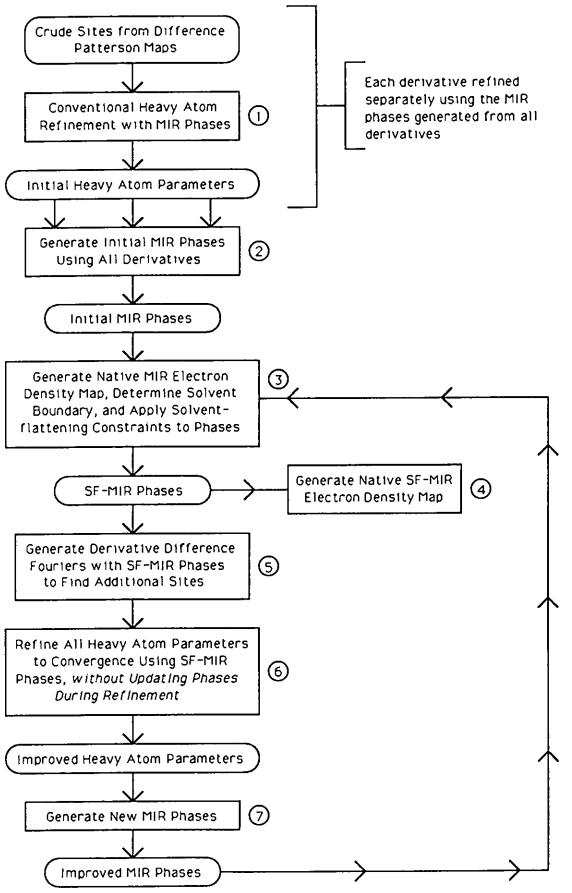

Fig. 1. Protocol for refining heavy-atom parameters against phases constrained by solvent flattening. See text for discussion of each step.

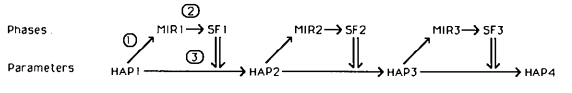

Fig. 2. Evolution of the heavy-atom parameters and phases during the development of an interpretable electron density map for the glutaminyl-tRNA synthetase-$tRNA^{Gln}$-ATP complex structure. The process begins with the set of heavy-atom parameters (HAP1) that give the best statistics after conventional refinement. MIR 'best' phases, MIR1, with corresponding Hendrickson-Lattman coefficients are first generated in step 1, then Fourier transformed with the native amplitudes for use in determination of the solvent region of the unit cell by an automated procedure (Wang, 1985). This second step is completed when iterative recombination of the original phase-probability distributions with those obtained by back-transforming solvent-flattened electron density maps converges, yielding solvent-flattened phases, SF1, with updated probability profiles. These phases are used to generate derivative-native difference Fourier maps for the extraction of new heavy-atom sites. In the final step of one cycle of this protocol, the parameters from all of the heavy-atom sites of all derivatives are refined using only the solvent-flattened phases as the parent phases, resulting in an improved parameter set *HAP*2. Three passes through this protocol were performed. The final refinement yielding parameter set *HAP*4 served only to confirm convergence of the parameters.

Table 1. *Heavy-atom-parameter statistics at each cycle of the protocol shown in Fig.* 2

For each derivative is listed the number of sites used at each cycle, the phasing power of the derivative expressed as the r.m.s. amplitude of the structure factor arising from the heavy atoms divided by the lack-of-closure error (fh/E), the r.m.s. shift in Å of all sites (from the previous parameter set) and the largest shift of any of the sites in the derivative. For the EMP and PCMB derivatives, the high- and low-resolution data sets were collected from different crystal batches and showed differences in degrees of heavy-metal substitution. EMP1, PCM1 and AUC1 extended to 3.5 Å; EMP2 and PCM2 to 2.8 Å; DMA1 to 3.0 Å.

	Derivative	HAP1	HAP2	HAP3	HAP4
AUC1	Number of sites	Not used	5	6	6
	fh/E		0.81	0.93	0.95
	R.m.s. shift			0.32	0.08
	Max. shift			0.65	0.16
DMA1	Number of sites	Not used	7	8	8
	fh/E		1.40	1.62	1.67
	R.m.s. shift			0.26	0.19
	Max. shift			0.60	0.47
EMP1	Number of sites	2	6	7	7
	fh/E	1.50	1.90	2.18	2.29
	R.m.s. shift		0.13	0.12	0.13
	Max. shift		0.17	0.18	0.35
EMP2	Number of sites	2	6	7	7
	fh/E	1.14	1.17	1.43	1.51
	R.m.s. shift		0.14	0.08	0.04
	Max. shift		0.19	0.13	0.10
PCM1	Number of sites	4	5	5	5
	fh/E	1.55	1.77	1.96	2.04
	R.m.s. shift		0.12	0.27	0.06
	Max. shift		0.17	0.60	0.13
PCM2	Number of sites	4	5	5	5
	fh/E	1.78	1.96	2.10	2.21
	R.m.s. shift		0.07	0.05	0.03
	Max. shift		0.11	0.10	0.06
All derivatives					
	R.m.s. shift		0.11	0.21	0.11
	Figure of merit	0.51	0.55	0.58	0.59

and less accurate, since analysis of a given heavy-atom site in one derivative required the omission of that site from the other derivatives used in phasing (lest the errors in occupancy and *B* factor for the same site in the derivatives used in phasing adversely influence that site in the derivative being analyzed). For this same reason, heavy-atom parameter refinement of one derivative using parent phases derived only from the other derivatives failed to converge; the derivatives were simply too similar to each other. The common method of refinement using only the centric reflections (whose phases are exactly known when the native and derivative amplitudes are known without error, except for crossover reflections) failed to converge in our case.

The best MIR phasing statistics were generated using only two of the derivatives (Table 1), EMP with only two sites and PCMB with four sites, using the *PROTEIN* crystallographic analysis package (Steigeman & Huber, 1982) for the heavy-atom refinement. The weak anomalous-scattering signal from the mercury atoms assisted the refinement and anisotropic *B* factors were necessary to fit the elongated sites.

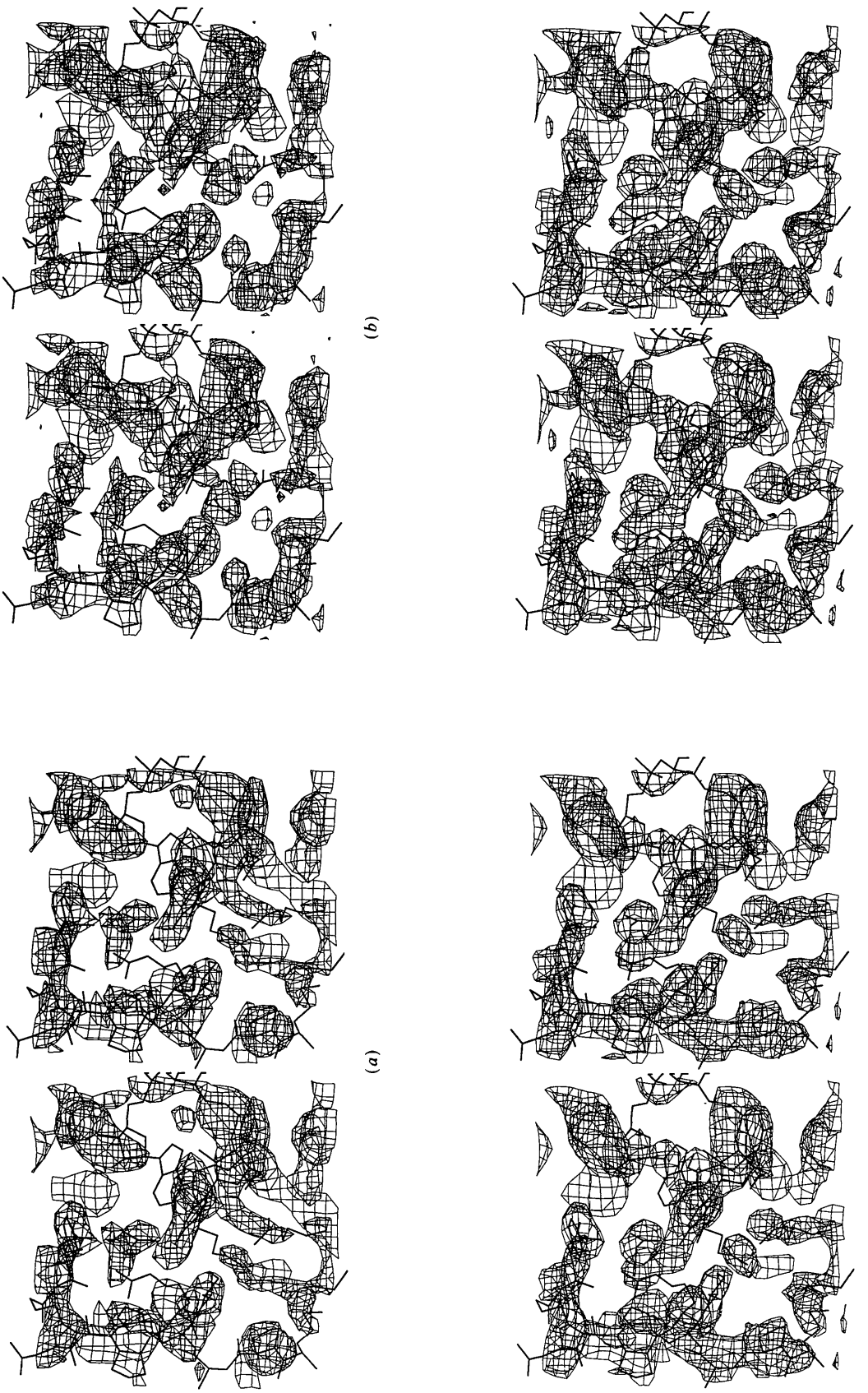

Fig. 3. Electron density maps of the tRNA anticodon nucleotide 35 and its binding pocket in the enzyme, showing the real-space effects of the process described in Fig. 2, superimposed on the refined model in this region. All maps use the same native data over the same resolution range of 10–2.8 Å, with figure-of-merit weighting applied to each term in the Fourier summation, and are contoured at 1.25σ. Each map thus differs only in the phases and figures of merit used: (*a*) MIR1, (*b*) SF1, (*c*) MIR3 and (*d*) SF3. While some improvement in the electron density was seen after the first round of solvent flattening, two rounds of heavy-atom refinement using solvent-flattened phases resulted in clear definition of the interacting protein and RNA residues, even before solvent flattening of the final MIR phases.

Table 2. *Mean phase differences between and among the MIR and SF-MIR phase sets shown in Fig. 2 as a function of resolution*

Convergence of MIR phasing was essentially complete after one cycle through the procedure outlined in Fig. 1.

Resolution (Å)	Number of reflexions	MIR1 MIR2	MIR2 MIR3	MIR1 MIR3	SF1 SF2	SF2 SF3	SF1 SF3	MIR1 SF1	MIR2 SF2	MIR3 SF3
10–4.8	5787	26	6	29	34	5	35	38	10	9
4.8–3.8	6512	37	8	40	44	9	45	54	23	20
3.8–3.3	6790	44	10	47	54	14	54	62	36	33
3.3–3.0	6406	49	12	52	59	18	59	67	49	45
3.0–2.8	5574	48	12	52	65	22	66	71	59	55
10–2.8	31 069	41	10	44	51	14	52	58	35	32

While the first MIR electron density maps were too noisy to interpret (Fig. 3*a*), maps generated using solvent-flattened phases showed substantial improvement (Fig. 3*b*). Most importantly, difference Fourier maps calculated using these solvent-flattened phases allowed more reliable identification of the minor heavy-atom binding sites. Some of the elongated sites resolved into two sites in such maps.

The heavy-atom parameters were then refined using these solvent-flattened parent phases, as described above. A total of three such cycles of heavy-atom refinement, MIR phasing, solvent flattening (with redetermination of the solvent envelope) and map generation resulted in convergence of the heavy-atom parameters and a readily interpretable electron density map (Fig. 3*d*).

Examination of Table 2 shows that after one cycle through the procedure the MIR phases had converged. Native maps using either MIR2 or MIR3 phase sets are nearly identical, even though the phasing power of the derivatives improved (Fig. 2 and Table 1). After the first cycle there was a small increase (about 10%) in the occupancies and about half that increase in subsequent cycles. Proper refinement of the heavy-atom occupancies was crucial in this case where all derivatives share the same sites and is one of the strengths of this method (Cura, Krishnaswamy & Podjarny, 1992). That the occupancies were approaching their correct values was reflected in the reduced artifacts at the heavy-atom sites in subsequent native MIR electron density maps. The absence of these artifacts in the solvent-flattened MIR maps indicates that heavy-atom bias is substantially reduced in the SF-MIR phases. Interestingly, the mean phase change upon solvent flattening of the MIR phases of the final cycle is about half that upon solvent flattening the MIR phases of the first cycle. Reflecting this, the final MIR map (Fig. 3*c*) improves little after solvent flattening (Fig. 3*d*).

Compared with the best phases obtainable before solvent flattening, the overall figure of merit to 2.8 Å for the final MIR phases improved from 0.51 to 0.59; accordingly, these final MIR phases more closely resemble phases calculated from a refined model of the complex (Table 3). The overall phase improvement of 20° (66 to 46° *versus* calculated phases) is substantial (compare with Cura, Krishnaswamy & Podjarny, 1992). Concomitantly, the phasing power of the individual derivatives, expressed as the r.m.s. heavy-atom amplitude signal divided by the r.m.s. lack-of-closure error, improved by an average of 36%. Several sites with occupancies less than 10% of the most substituted site were identified, with all sites found to be ligands to cysteine or histidine residues after model building.

Table 3. *Mean phase differences between phases calculated from a refined model of the complex and MIR or SF-MIR phases as a function of resolution*

Despite major improvement in the interpretability of the final MIR *versus* the initial MIR electron density maps, the mean |MIR − calc.| phase difference changes only modestly. Since the mean figure of merit of a phase set is approximately equal to the cosine of the mean expected discrepancy between the phase set and the true phases, the figure of merit for the final solvent-flattened phases is thus estimated to be 0.7 (assuming the calculated phases to be close to the true phases), in contrast to the value of 0.8 given by the solvent-flattening programs.

Resolution (Å)	MIR1	MIR2	MIR3	SF1	SF2	SF3
10–4.8	49	37	37	45	34	34
4.8–3.8	60	46	46	51	36	36
3.8–3.3	68	57	56	62	44	44
3.3–3.0	75	67	67	68	56	55
3.0–2.8	79	75	75	76	65	65
10–2.8	66	57	56	60	46	46

In examining these tables, it might be argued that the improvement in the quality of the phases may have simply been due to the addition of two derivatives (AUC1 and DMA1) and to the addition of new sites to the original derivatives. Prior to refinement against the solvent-flattened phases, the AUC1 derivative could not be refined well and the new sites in the original derivatives were in the noise level of the MIR-only maps. It must not be overlooked that each of these derivatives was refined to convergence in essentially one cycle by using solvent-flattened MIR phases from a rather poorly determined set of only the two original derivatives (EMP1/EMP2 and PCM1/PCM2). The additional sites were extracted using derivative–native difference Fourier maps using the solvent-flattened phases, SF1. This method may be most powerful in the SIR case, where solvent flattening biases the bimodal phase distribution toward the correct parent phase for use in further refinement.

Subsequently, other groups have applied this protocol to other difficult structure determinations, with similar success; for example, the RecA protein of *E. coli* (Story, Weber & Steitz, 1992) and the engrailed homeodomain/DNA complex (Kissinger, Liu, Martin-Blanco, Kornberg & Pabo, 1990). Other more recent examples and a critical examination of the application of this method is presented by Cura, Krishnaswamy & Podjarny (1992).

While solvent flattening provided the phase constraints for decoupling parent-phase generation from heavy-atom refinement, other sources may be exploited, for example noncrystallographic symmetry (Bricogne, 1976; Cura, Krishnaswamy & Podjarny, 1992), translational noncrystallographic symmetry (Agard & Stroud, 1982) and entropy maximization (Prince, Sjölin & Alenljung, 1988). These and other constraints are reviewed by Tulinsky (1985) and Podjarny, Bhatt & Zwick (1987).

References

AGARD, D. A. & STROUD, R. M. (1982). *Acta Cryst.* A**38**, 186-194.

BHAT, T. N. & BLOW, D. M. (1982). *Acta Cryst.* A**38**, 21-29.

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794-802.

BLOW, D. M. & MATTHEWS, B. W. (1973). *Acta Cryst.* A**29**, 56-62.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832-847.

CURA, V., KRISHNASWAMY, S. & PODJARNY, A. D. (1992). *Acta Cryst.* A**48**, 756-764.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136-143.

KISSINGER, C. R., LIU, B., MARTIN-BLANCO, E., KORNBERG, T. B. & PABO, C. O. (1990). *Cell*, **63**, 579-590.

PODJARNY, A. D., BHATT, T. N. & ZWICK, M. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 351-373.

PRINCE, E., SJÖLIN, L. & ALENLJUNG, R. (1988). *Acta Cryst.* A**44**, 216-222.

ROULD, M. A., PERONA, J. J., SOLL, D. & STEITZ, T. A. (1989). *Science*, **246**, 1135-1142.

SCHEVITZ, R. W., PODJARNY, A. D., ZWICK, M., HUGHES, J. J. & SIGLER, P. B. (1981). *Acta Cryst.* A**37**, 669-677.

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

STEIGEMANN, W. & HUBER, R. (1982). *PROTEIN Crystallographic Program Suite.* Max-Planck Institut für Biochemie, Martinsried, Germany,

STORY, R. M., WEBER, I. T. & STEITZ, T. A. (1992). *Nature (London)*, **335**, 318-325.

TULINSKY, A. (1985). *Methods Enzymol.* **115**, 77-89.

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

Acta Cryst. (1992). A**48**, 756-764

Heavy-Atom Refinement Against Solvent-Flattened Phases

BY V. CURA, S. KRISHNASWAMY* AND A. D. PODJARNY†

Laboratoire de Cristallographie Biologique, IBMC du CNRS, 15 rue Descartes, 67084 Strasbourg, France

(*Received* 11 *September* 1991; *accepted* 20 *March* 1992)

Abstract

A new algorithm for refinement of heavy-atom parameters is defined by an iterative procedure where external phases are provided by density modification. This algorithm is applied to two cases, $tRNA^{Asp}$ and the complex between $tRNA^{Asp}$ and aspartyl-tRNA synthetase. In the first case, where the structure was solved by multiple isomorphous replacement (MIR) methods, it was found that the new method gives accurate values for the native-derivative scale and for occupancy of heavy-atom sites. Position refinement was more delicate and it needed to be handled in a restricted resolution range. In the second case, where a similar method was used in the early stages of solving the phase problem, it slightly decreased the phase error. It was followed by an improvement of the density-modification masks, which led to better maps at higher resolution.

Introduction

Heavy-atom isomorphous replacement (Blow & Crick, 1959) is widely used for phasing a new molecular structure. However, its application is rarely straightforward (Philips, 1988; Blow, Henrick & Vrielink, 1988) and new developments such as improvements of the error analysis (Read, 1991) or of the underlying theory (Bricogne, 1992; Otwinowski, 1992) are welcome.

The refinement of heavy-atom parameters relies on the proper estimate of either the heavy-atom amplitude or the protein phase. The first option is adequate if enough centric reflections or good-quality derivative anomalous-dispersion data are available. If this is not the case, it is necessary to estimate the protein phase independently of the heavy-atom parameters under refinement (Dodson, 1976, and references

* Present address: Madurai Kamaraj University, Madurai, India.

† To whom correspondence should be addressed.

0108-7673/92/050756-09$06.00