View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Selenium-based MAD phasing: setting the sites on larger structures

AM Deacon and SE Ealick*

Address: Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA.

*Corresponding author.
E-mail: see3@cornell.edu

Structure July 1999, **7**:R161–R166
http://biomednet.com/elecref/09692126007R0161

© Elsevier Science Ltd ISSN 0969-2126

Introduction

Over the past few years the method of macromolecular structure determination known as multiwavelength anomalous diffraction (MAD) has been transformed from an esoteric technique, used in a few favorable cases, into a mature and generally applicable approach to solving the crystallographic phase problem [1]. Since 1993 there has been a dramatic increase in the number of structures solved by MAD [2] and one of the major contributing factors has been the production of selenomethionyl (SeMet) proteins. In most cases, methionine auxotrophic expression systems have provided an elegant mechanism for ensuring the incorporation of anomalous scattering atoms [3] and this technique seems to be broadly applicable to many macromolecular systems [4]. Additionally, the selenium K absorption edge (at 12.658 keV) is at an ideal energy for most synchrotron radiation sources. As a result, selenium has become the most popular anomalous scattering atom for MAD phasing and has been used in approximately two-thirds of all reported applications [2].

Over the same time period there has also been a significant increase in the number of synchrotron beamlines, including several third generation undulator beamlines, that offer tunable X-ray radiation suitable for MAD experiments. The high intensity X-ray beams, coupled with narrow energy bandwidth optics, allow reliable optimization of the small anomalous Bijvoet and dispersive MAD signals. Meanwhile, the efficiency of both new and existing beamlines is being further enhanced by the deployment of charge-coupled device (CCD) detectors [5]. These sensitive devices offer excellent data quality and a greatly reduced duty cycle, due to their rapid readout systems. Another important factor, which has influenced the entire field of macromolecular crystallography, has been the popularization of cryoprotection techniques [6]. Most MAD experiments can now be completed using a single frozen crystal. The inherent isomorphism of this approach, combined with the greatly reduced radiation damage, helps to minimize systematic errors in the data and facilitates the extraction of reliable phase information. In addition, the development of more sophisticated algorithms, embodied in several computer software systems, has served to improve the quality of experimental electron-density maps and decrease the time taken to obtain them. Many of these programs are now combined with intuitive, graphical user interfaces, which should make them instantly accessible to the user community. Finally, because anomalous scattering is largely independent of diffraction angle, the electron density maps obtained from MAD experiments show greater detail, compared to heavy-atom methods, and consequently the initial model is of higher quality leading to more rapid convergence of least-squares refinement procedures.

As a result of the advances summarized above, the structures of many macromolecules, particularly those of less than 100 kDa molecular weight, can be determined in a routine and automated fashion and investigators are now recognizing the potential of the MAD phasing approach. Given this recent success, it remains to be seen at what point the application of MAD to even larger structures will be limited by either theoretical or practical considerations. In this paper we discuss the application of SeMet MAD phasing to large macromolecular structures, including one example in which the asymmetric unit has a molecular weight of 370 kDa and contains 70 SeMet residues.

The suitability of MAD for larger structures

The possibility of using MAD phasing to solve larger structures is an attractive one. After all, it is these systems that often exhibit the most severe lack of isomorphism with heavy-atom derivatives and the largest cell-parameter shifts on freezing. In principle, the time-consuming, systematic searches for heavy-atom derivatives can be replaced by a single, well conceived multiwavelength experiment. In order to tackle these larger problems by MAD phasing it is necessary to sufficiently augment the anomalous scattering signal. This can be accomplished either by using heavier anomalous scattering atoms, to take advantage of their more pronounced L or M absorption edges, or by incorporating more anomalous scattering atoms. The latter alternative lends itself to the continued use of SeMet MAD phasing. In principle the natural occurrence of methionine as one in every 59 residues provides sufficient anomalous scattering signal (one selenium per 6.3 kDa). In practice, selenium concentrations as low as one selenium per 20 kDa have been reported [7]; however, a typical 500 kDa protein would contain about 100 selenium atoms. Therefore, for larger proteins and macromolecular complexes in the MDa range, it is clear that the ability to

locate the anomalous scattering atoms will be a major hurdle in the structure determination process.

Locating anomalous scattering atoms

Traditionally, the hand interpretation of Patterson maps has been the workhorse for locating anomalous scattering atoms in macromolecular structure determination. However, this approach is only practical for a few heavy-atom sites. The development of automated Patterson search programs, such as the Patterson superposition methods in SHELXS [8], have been successful in extending the application of Patterson techniques to the l0–20 site range. Beyond this, more sophisticated algorithms are called for. In the absence of any prior phase information two types of *ab initio* methods are available for large substructure determination: those based on more advanced Patterson searches and those based on direct methods.

Solving large substructures with Patterson searches

Two of the most successful and promising Patterson approaches form part of the complete structure solution packages of SOLVE [9] and CNS [10]. SOLVE has pioneered the development of completely automated MAD structure determination. It features a comprehensive data analysis procedure for MAD applications, which derives Bayesian estimates of F_A, the scattering factor amplitudes associated with the anomalous scattering substructure. SOLVE also provides valuable information on the correlations between dispersive and anomalous Patterson maps. It uses a Patterson map search procedure, HASSP [11], to generate two-site seed trials from an F_A Patterson map. An iterative difference Fourier analysis is then used to identify new sites. The procedure uses site-deletion to completely cross-validate each site in the substructure and a sophisticated scoring system is able to assess when the substructure is complete.

The CNS program is broad in scope, providing both structure determination and refinement algorithms. It includes a graphical user interface, task-oriented scripts and even a symbolic structure determination language. A new algorithm for locating anomalous scattering substructures has been implemented in this language. The method starts with a single site and uses an iterative translation search procedure to build up the anomalous scattering model. All trials are refined and a correlation coefficient between the observed and calculated squared structure-factor amplitudes is used to score them.

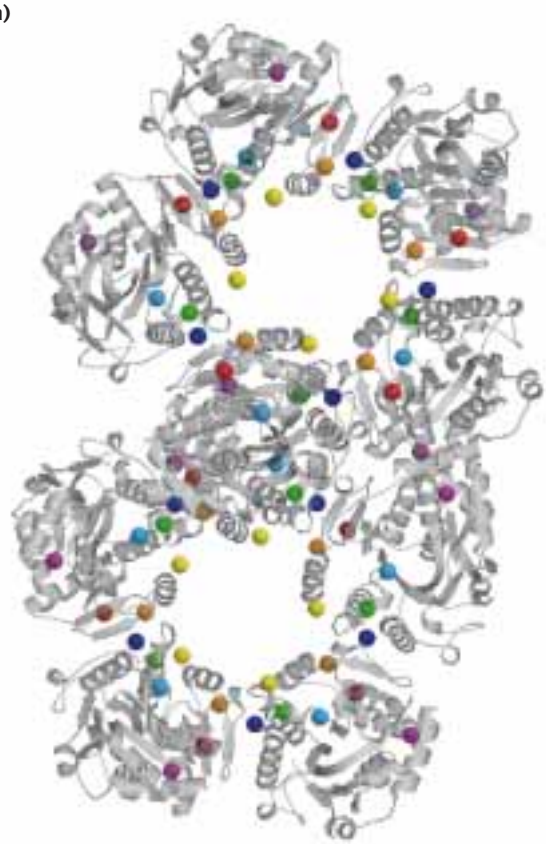
Solving large substructures with direct methods

The programs SHELXS [12] and RANTAN [13] have both been successfully applied to large substructure determinations. Perhaps the most promising direct methods programs for these applications, however, are based on the dual-space Shake-and-Bake algorithm. This algorithm has been implemented in two computer programs: SnB v2.0 [14], which has recently been released and includes an easy-to-use interface for substructure determination; and SHELXD [15], which is in the latter stages of testing and development. These two programs are similar in that they iterate cycles of phase refinement in reciprocal space, followed by density modification through peak-picking in real space. They differ in the precise algorithms used in each space [16]. Over the past few years these dual-space direct methods of structure determination have proved useful in extending the range of *ab initio* phasing of entire molecules to over 1000 atoms [17,18]. They are more robust and efficient than traditional direct methods programs, which suggests that they are appropriate for the task of extracting a large number of anomalous scattering positions from what may be a weak and potentially noisy MAD signal. These methods are also well matched to the average spacing of selenium atoms and the typical resolution of MAD data. Whereas traditional *ab initio* applications require atomic resolution (< 1.2 Å) diffraction data to locate atoms separated by 1–2 Å, substructure determinations using these new methods can be readily conducted at 3 Å resolution or even lower, because selenium atoms are sure to be separated by at least 4.0 Å.

Substructure determination with SnB involves a careful preprocessing of the anomalous differences at a single (usually the f″ maximum) wavelength [19]. This allows a subset of reflections corresponding to the most reliable anomalous differences to be selected. In the structure determination of ADP-L-glycero-D-mannoheptose 6-epimerase (AGM epimerase) the positions of 65 out of 70 selenium atoms were determined. The most reliable 13,797 out of 67,874 anomalous differences were selected, from which the largest 1400 normalized difference structure factor magnitudes (E values) were used in the SnB program. Despite the use of Bijvoet differences, which only approximate the true scattering contribution of the substructure, the height of the peaks returned from SnB were closely correlated to the temperature factors of the refined atomic coordinates (Figure 1). The five atoms not found in the SnB solution had high thermal parameters in the final refined model and were buried close to the noise level in the final SnB E maps. An alternative for SnB substructure determination is to use derived values of F_A. In principle these should more accurately characterize the anomalous scattering substructure. If the derived F_A values are noisy, however, then this approach can be less successful [20]. The SnB program can be run quickly, even for large unit-cell dimensions, by using a fairly coarse electron-density grid. The time taken can be a few minutes for small substructures, up to several hours for large substructures.

Direct methods can also be used to gradually build up a substructure, starting with a small portion of it. In this case, the known positions of anomalous scattering atoms

Figure 1



(b)

Methionine residue number	Molecule number 1	2	3	4	5	6	7	8	9	10
1	64	65	56	67	–	61	–	63	62	–
51	28	20	36	15	38	35	23	27	54	34
61	25	44	47	30	21	22	53	42	41	39
89	1	7	8	10	5	6	3	2	14	16
90	32	24	31	17	18	4	13	12	9	11
181	58	60	19	43	40	–	37	48	33	–
303	50	46	59	51	45	5	29	49	26	52

(c)

Methionine residue number	Molecule number 1	2	3	4	5	6	7	8	9	10
1	55	68	56	54	54	63	63	58	61	59
51	44	39	47	37	32	43	35	33	42	36
61	49	45	41	43	41	39	40	47	36	40
89	29	27	24	24	24	25	25	23	26	25
90	30	28	25	24	28	25	27	25	26	26
181	66	63	50	38	56	47	35	59	35	54
303	39	30	49	30	40	30	29	26	34	35

Structure

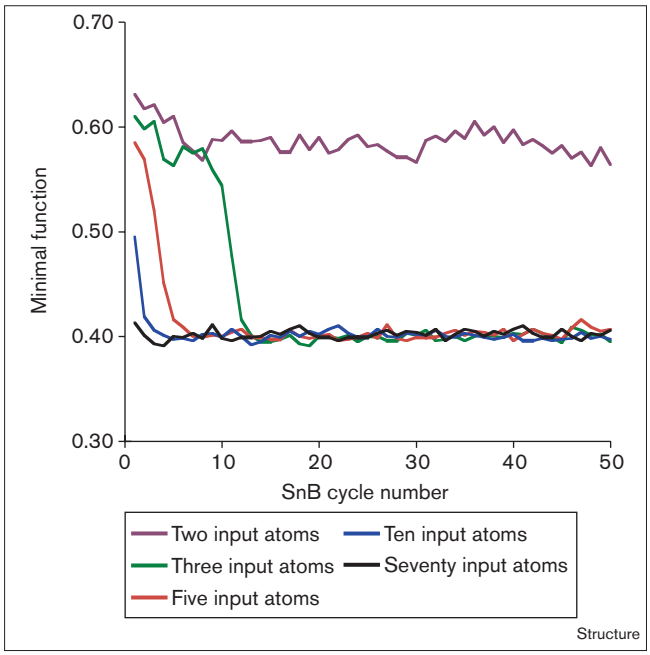
The 70 selenium atom substructure of AGM epimerase. There are ten molecules in the asymmetric unit and seven selenium sites per molecule. **(a)** Ribbon diagram of AGM epimerase with selenium atoms superimposed. **(b)** Peak height order from SnB (1 is the highest peak, '–' indicates peak not found). **(c)** Temperature factor (Å^2) of selenium sites in the refined model (sites that were missing from the SnB solution are underlined). The color coding in parts (b) and (c) corresponds to the colors used in (a).

can be used to estimate initial phase angles, which are further improved by the direct methods algorithm. In addition, direct methods can utilize Patterson space information. The SHELXD program provides a hybrid of the Patterson and direct methods approaches in which random starting trials for the direct methods portion of the program are selected on the basis of their consistency with the Patterson function (typically an F_A Patterson). This allows a rapid screening of trials and the selection of only the most promising ones. The potential benefits of this approach are illustrated in Figure 2. When given a starting trial that consists of at least three correct selenium positions for AGM epimerase, the SnB program very quickly converges (in a matter of seconds) on a complete substructure of 70 independent selenium atoms.

Recent examples of large substructure determinations

Over the past 12 months investigators have started to turn to large selenium substructures as an effective phasing tool for MAD. Over this short time period there has been a substantial increase in the size of substructures that have successfully been exploited, leading to the recent structure determination of AGM epimerase, which has a substructure of 70 selenium atoms (AMD, Y Ni, WG Coleman Jr and SEE). The data presented in Table 1 serves to emphasize both the wide choice of beamlines and computer programs that are now available for tackling these problems, and the size of substructures that appear to be routinely attainable. It is clear that the limits have not yet been reached. The inclusion of two structures in space group P1 indicates that highly redundant data sets are not required in order to solve these large substructures, although increased redundancy can clearly have an important role in compensating for weak signals. External phase information has been used in only two of these cases. In both of these studies other phasing methods — molecular isomorphous replacement (MIR) and molecular replacement (MR) — had failed to yield an adequate structure solution and the use of MAD was the key to obtaining readily interpretable electron-density maps.

Figure 2



Minimal function versus cycle number in SnB for a 70 selenium atom substructure determination. Each curve represents a single trial starting from a number of known selenium coordinates (as indicated). Trials starting from at least three known selenium positions converge rapidly to the complete 70-atom substructure.

The sensitive residual maps calculated using the program SHARP [21] were employed in both these cases to extract the heavy-atom positions. Even without any prior phases, it seems likely that these selenium substructures could have been determined *de novo*.

Calculating phase angles

Once the selenium atom positions are known, phase angles for larger MAD structures can be calculated as would be the case for any MAD structure. The use of MLphare [22], which essentially treats MAD as a special case of MIR [23], remains a very popular approach for heavy-atom refinement and phase determination. The excellent quality of the resulting electron-density maps, speed of execution and the familiarity of the program to the crystallographic community are certainly the overriding factors. The more MAD-specific approaches implemented in other programs will no doubt increase in popularity as they become more familiar to the community. It is also worth noting that although several of the largest structures listed in Table 2 crystallize with multiple molecules present in the asymmetric unit, the use of non-crystallographic symmetry has not generally been needed to obtain high quality electron-density maps (Figure 3).

Concluding remarks

Visiting a synchrotron to conduct a MAD experiment requires substantial planning and sample preparation. The data collection itself can also require a significant amount of effort, however, the potential benefits are great. It is already becoming feasible to return home with interpretable electron-density maps in hand. In the future, the amount of effort required to conduct these experiments will be greatly

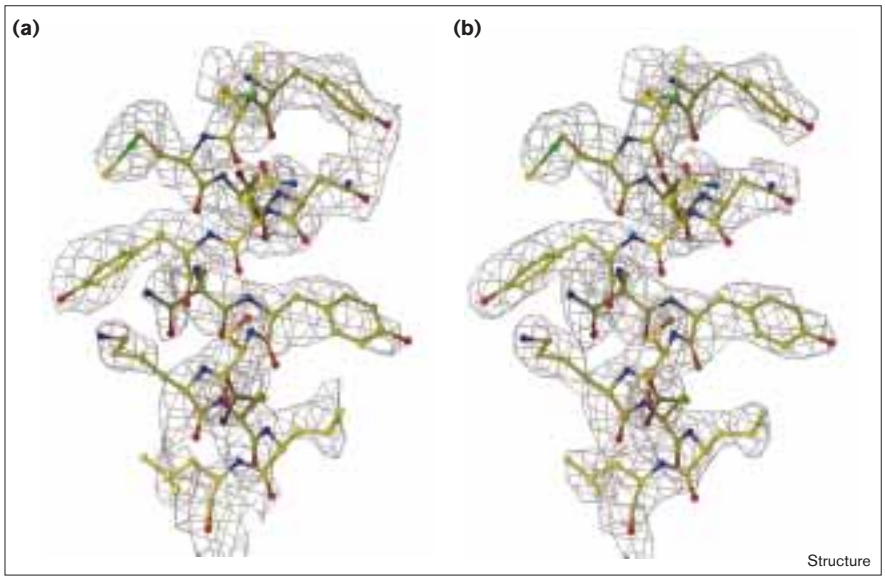
Table 1

New structures solved by MAD phasing from large selenium substructures.

Structure	Synchrotron source	Space group	Longest cell (Å)	D_{min} (Å)	No. Se atoms	MW (kDa)	kDa /Se	Program (Se atoms)	Program (phasing)	Ref
E1b	ESRF, BM14	14_122	382	2.4	22	82	3.7	MR/SHARP*	SHARP	[25]
AdoMet decarboxylase	CHESS, F2	$P2_1$	90	2.25	24	76	3.2	SnB/SOLVE	MLphare	[26]
Napthalene-1,2-dioxygenase	ESRF, BM14	R32	323	3.5	26	145	5.6	SIR/SHARP†	SHARP	[27]
PurR	ESRF, BM14	P1	83	2.3	28	126	4.0	SHELXS‡	MLphare	[28]
AIR synthetase	CHESS, F2	$P2_12_12_1$	212	3.0	28	148	5.3	SnB	MLphare	[29]
FTHF synthetase	APS, SBC	R32	256	3.0	28	120	4.3	SHELXD	MLphare	§
FA hydrolase	NSLS, X12-C	$P2_1$	110	1.9	30	92	3.1	SOLVE	SOLVE/MLphare	#
AdoHcy hydrolase	NSLS, X12-C	C222	168	2.8	30	96	3.2	SnB	Phases	[30]
Cyanase	APS, SBC	P1	82	2.4	40	170	4.2	CNS/RANTAN	MLphare	¶
EphB2–SAM	ALS, 5.0.2	$P4_1$	105	1.95	48	78	1.6	SnB	MLphare	[31]
Not reported	APS, IMCA	$P2_12_12_1$	174	2.4	52	183	3.5	SOLVE	SOLVE	¥
AGM epimerase	CHESS, F2	$P2_1$	180	3.0	70	370	5.3	SnB	MLphare	**

*Molecular replacement phases and Fourier techniques were used. †Single isomorphous replacement phases and Fourier techniques were used. ‡Four sites were picked by hand from a Patterson map and the rest were obtained by tangent expansion in SHELXS. §L Lebioda, R Radfar, R Shin, GM Sheldrick, JD Odom and RB Dunlap, personal communication. #DE Timm, personal communication. ¶MA Walsh, Z Otwinowski, A Perrakis, PM Anderson and A Joachimiak, personal communication. ¥CA Janson, WW Smith and TC Terwilliger, personal communication. **AMD, Y Ni, WG Coleman Jr and SEE, unpublished data.

Figure 3



Comparison of the experimental electron-density maps for AGM epimerase. **(a)** The map obtained using MLphare and DM (no averaging); figure of merit 0.80. **(b)** MLphare and DM (tenfold averaging); figure of merit 0.85.

reduced as automatic data collection procedures become a routine part of MAD instrumentation. Experimental parameters will be optimized and monitored by computers, which will in turn be able to execute and guide the various steps in the structure determination process. The direct methods approaches for substructure determination are in a position to join SOLVE in the world of rapid and automatic structure determination. Using these methods, large anomalous scattering substructures can be determined as soon as the peak wavelength data collection has been completed. Attempts to solve the entire structure can be conducted once the second wavelength has been collected and the need for a third wavelength can be assessed. The recent development of automatic tracing and refinement strategies [24] offer the possibility of proceeding quickly from initial phases to a fully traced and refined model. These capabilities are sure to open the door for none specialists to solve structures by MAD phasing and will also serve to further increase the throughput of synchrotron beamlines.

In principle, these same procedures will be scaleable and applicable to larger macromolecules. It should be noted, however, that the question of radiation damage is starting to reappear, especially as investigators are turning to brilliant undulator-based synchrotron sources for their most challenging MAD phasing applications. It may be necessary to strike a balance between resolution and crystal lifetime, especially in low-symmetry space groups where the amount of data required is quite substantial. In cases where the problem is severe it may become necessary to resort to narrow segment data collection strategies, where data at all wavelengths are collected from the same crystal at approximately the same time.

Nevertheless, several algorithms for MAD phasing from large selenium substructures are now in place. It remains to be seen what limits of application are encountered as they are applied to larger and more challenging biological systems.

Acknowledgements

We would like to acknowledge the generous support of grant GM-46733 from the National Institutes of Health. Research at CHESS is supported by grant DMR-9311772 from the National Science Foundation and the MacCHESS research resource is funded by grant RR-01646 from the National Institutes of Health.

References

1. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.
2. Ogata, C.M. (1998). MAD phasing grows up. *Nat. Struct. Biol.* **5**, 638-640.
3. Hendrickson, W.A., Horton, J.R. & LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multi-wavelength anomalous diffraction (MAD) – a vehicle for direct determination of 3-dimensional structure. *EMBO J.* **9**, 1665-1672.
4. Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523-530.
5. Gruner, S.M. & Ealick, S.E. (1995). Charge coupled device X-ray detectors for macromolecular crystallography. *Structure* **3**, 13-15.
6. Rodgers, D.W. (1994). Cryocrystallography. *Structure* **2**, 1135-1140.
7. Thiel, D.J., *et al.,* & Ealick, S.E. (1995). MAD phasing of the human interferon-γ/receptor complex. *Fifth International Conference on Biophysics and Synchrotron Radiation*. Grenoble, France, Abstract 01/38.
8. Sheldrick, G.M. (1997). Patterson superposition and *ab initio* phasing. *Methods Enzymol.* **276**, 628-641.
9. Terwilliger, T.C. & Berendzen, J. (1999). Automated structure solution for MAD and MIR. *Acta Crystallogr. D* **55**, 849-861.
10. Brünger, A.T., *et al.*, & Warren, G.L. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.
11. Terwilliger, T.C., Kim, S.-H. & Eisenberg, D. (1987). Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallogr. A* **43**, 1-5.

12. Sheldrick, G.M. (1990). Phase annealing in SHELX-90 – direct methods for larger structures. *Acta Crystallogr. A* **46**, 467-473.
13. Yao, J-X. (1983). On the application of phase relationships to complex structures. XVIII. RANTAN – random MULTAN. *Acta Crystallogr. A* **37**, 35-37.
14. Weeks, C.M. & Miller, R. (1999). The design and implementation of SnB version 2.0. *J. Appl. Crystallogr.* **32**, 120-124.
15. Sheldrick, G.M. (1998). *SHELX*: *Applications to Macromolecules. Direct Methods for Solving Structures*. (Fortier, S., ed.), pp. 401–411, Kluwer Academic Publishers, Dordecht, Germany.
16. Weeks, C.M., Sheldrick, G.M., Miller, R., Usón, I. & Hauptman, H.A. (1999). *Ab-initio* phasing by dual-space direct methods. *Bulletin of the Czech and Solvak Crystallographic Association*, in press.
17. Deacon, A.M., Weeks, C.M., Miller, R. & Ealick, S.E. (1998). The Shake-and-Bake structure determination of triclinic lysozyme. *Proc. Natl Acad. Sci.* **95**, 9284-9289.
18. Frazão, C., Sieker, L., Sheldrick, G.M., Lamzin, V., LeGall, J. & Carrondo, M.A. (1999). *Ab initio* structure solution of a dimeric cytochrome c_3 from *Desulfovibrio gigas* containing disulfide bridges. *J. Biol. Inorg. Chem.* **4**, 162-165.
19. Blessing, R.H. & Smith, G.D. (1999). Difference structure factor normalization for heavy atom or anomalous scattering substructure determinations. *J. Appl. Crystallogr.,* in press.
20. Smith, G.D., Nagar, B., Rini, J.M., Hauptman, H.A. & Blessing, R.H. (1998). The use of SnB to determine an anomalous scattering substructure. *Acta Crystallogr. D* **54**, 799-804.
21. De La Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494.
22. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering*. (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80-86, SERC, Daresbury Laboratory, Warrington, UK.
23. Ramakrishnan, V. & Biou, V. (1997). Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. *Methods Enzymol.* **276**, 538-557.
24. Perrakis, A., Morris, R. & Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol.* **6**, 458-463.
25. Ævarsson, A., Seger, K., Turley, S., Sokatch, J.R. & Hol, W.G.J. (1999). First picture of all the major components of 2-oxo acid dehydrogenase multienzyme complexes obtained by the crystal structure of 2-oxoisovalerate dehydrogenase. *Nat. Struct. Biol.,* in press.
26. Ekstrom, J.L., Mathews, I.I., Stanley, B.A., Pegg, A.E. & Ealick, S.E. (1999). The crystal structure of human S-adenosylmethionine decarboxylase at 2.25 Å resolution reveals a novel fold. *Structure* **7**, 583-595.
27. Kauppi, B., Lee, K., Carredano, E., Parales, R.E., Gibson, D.T., Eklund, H. & Ramaswamy, S. (1998). Structure of an aromatic-ring-hydroxylating dioxygenase-napthalene 1,2-dioxygenase. *Structure* **6**, 571-586.
28. Sinha, S., Krahn, J.M., Shin, B.S., Tomchick, D.R., Zalkin, H. & Smith, J.L. (1998). The structure of the purine operon repressor from *B. subtilis*, solved by SeMet MAD. *Abstract 11.02.04*, ACA Meeting, Arlington, July 18–23.
29. Li, C., Kappock, T.J., Stubbe, J. & Ealick, S.E. (1999). X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM) from the *E. coli* purine biosynthetic pathway at 2.5 Å resolution. *Structure*, in press.
30. Turner, M.A., Yuan C.S., Borchardt, R.T., Hershfield, M.S., Smith, G.D. & Howell, P.L. (1998). Structure determination of selenomethionyl *S*-adenosylhomocysteine hydrolase using data at a single wavelength. *Nat. Struct. Biol.* **5**, 369-376.
31. Thanos, C.D., Goodwill, K.E. & Bowie, J.U. (1999). Oligomeric structure of the human EphB2 receptor SAM domain. *Science* **282**, 833-836.