A Minimalist's Approach to the Phase Problem – Phasing Selenomethionyl Protein Structures Using Cu Kα Data

MARIUSZ JASKÓLSKI[a,b]* AND ALEXANDER WLODAWER[a]

[a]*Macromolecular Structure Laboratory, NCI–Frederick Cancer Research and Development Center, ABL – Basic Research Program, Frederick, MD21702, USA, and* [b]*Center for Biocrystallographic Research, Institute of Biorganic Chemistry, Polish Academy of Sciences and Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznań, Poland. E-mail: mariuszj@plpuaml1.amu.edu.pl*

Abstract

The feasibility of phasing protein structures through the use of the isomorphous and anomalous signal of selenomethionyl (Se-Met) derivative and diffraction data collected with a standard laboratory Cu Kα X-ray source has been investigated. Interpretable electron-density maps were obtained for the core domain of avian sarcoma virus integrase, a typical medium-sized protein having four Met residues in a sequence of 156 amino acids. The r.m.s. difference between 3.1 Å experimental phases obtained from Se-Met Cu Kα data and the final phases calculated from the refined model is 55 . A procedure combining single isomorphous replacement/single anomalous scattering phasing and solvent flattening for data based on a single Se-Met derivative and Cu Kα radiation has been tested on this and another protein. The results are encouraging enough to indicate that such procedures might be recommended when a synchrotron source is not readily available.

1. Introduction

The phase problem still remains one of the main obstacles in solving novel protein structures, in spite of the tremendous progress in theoretical background, experimental technique, and the accumulation of structural knowledge. The principal method to which the protein crystallographer resorts in such cases is the technique of isomorphous replacement, introduced by Bijvoet (1949) and extended to proteins through the pioneering work of Perutz (Green, Ingram & Perutz, 1954). This method, however, is tedious, time consuming and prone to failure because of unsuccessful derivatization, lack of isomorphism and disordered or numerous substitution sites. In many cases, the isomorphous signal present in a derivative can be augmented by the accompanying anomalous signal, but this does not remedy the general drawbacks of the method. In 1980, Karle developed an elegant algebraic formalism allowing precise calculation of the unknown phases from multiple anomalous diffraction data sets (Karle, 1980). Its general requirement for a tunable X-ray source can now be fulfilled at synchrotron sites. This method, known as multi-wavelength anomalous diffraction (MAD), was introduced mainly by Hendrickson (1985, 1991) and has become increasingly successful in solving new protein structures.

MAD is particularly useful because the anomalous scatterers do not have to be very heavy. The example of Se is of special interest because, through the use of selenomethionine instead of methionine, it can be easily incorporated into the protein under study, which can then be crystallized isomorphously with the unaltered protein. The production of selenomethionyl derivatives, also advocated by Hendrickson (Hendrickson, Horton & LeMaster, 1990), is currently a rather simple problem for proteins that are obtained by recombinant techniques. In the case of the *Escherichia coli* expression system, methionine-requiring auxotrophic strains of the bacterium are available. Therefore, crystallographers studying novel proteins may find it easier to produce a selenomethionyl derivative (chemical/biochemical effect) than to gain access to tunable X-ray radiation (the physical effect) at a remote synchrotron source. On the other hand, contemporary conventional X-ray techniques are capable of producing very high quality data (Jaskólski & Dauter, 1996). Therefore, a practical question to consider is whether the complete experimental cycle, founded on the principle of tunable X-ray, has an absolute requirement for synchrotron radiation.

In this paper, we demonstrate the feasibility of obtaining interpretable protein electron-density maps by using selenomethionyl derivatives and Cu Kα diffraction data. The protein used in the tests was the core domain of avian sarcoma virus integrase (ASV IN), which is 156 amino acids long and contains four Met residues in the native sequence (Kulkosky, Katz, Merkel & Skalka, 1995). In this respect, this is a much more typical protein than the much smaller crambin, the structure of which was solved previously

from anomalous scattering of S atoms (Hendrickson & Teeter, 1981).

2. Experimental

2.1. Data collection and handling

An N,C-truncated construct containing only the central catalytic domain (residues 52–207) of ASV IN was expressed in *E. coli* as described by Kulkosky *et al.* (1995). Native crystals were grown at 277 K by the vapor-diffusion method using PEG/isopropanol as the precipitant (Bujacz *et al.*, 1995). The selenomethionyl derivative was produced in *E. coli* by G. Merkel (Bujacz *et al.*, 1995) as recommended by Hendrickson *et al.* (1990), using the auxotrophic strain DL41 provided by G. D. Markham. Mass spectrometry experiments (performed by R. Ogorzalek Loo) confirmed that the four Met sites were fully substituted by selenomethionine. The Se-Met protein was crystallized in a manner identical to the native protein, producing highly isomorphous crystals (Table 1), which were handled without any special precautions (Bujacz *et al.*, 1995). The crystals were stable in the beam and the two data sets, S-IN (native) and Se-IN (derivative), were collected at room temperature using a 300 mm MAR imaging plate scanner and graphite-monochromated Cu $K\alpha$ radiation generated by a rotating-anode source operated at 50 kV and 100 mA (Table 1). Both data sets were integrated and scaled using programs written by Otwinowski (1992). The two data sets had similar resolution (2.0 and 2.1 Å, respectively) and comparable quality (Table 1). The anomalous signal present in Se-IN was weak but visible $[R_{sym}^{iso}(I) = 0.093$, $R_{sym}^{ano}(I) = 0.083]$. Analysis of the isomorphous differences between the two data sets (Furey & Swaminathan, 1990) revealed that they were above 3σ to *ca* 3.7 Å resolution and above 2σ to *ca* 3.0 Å resolution (Table 2). Scaling of the two data sets in the 20.0–3.0 Å resolution range [where the errors in the individual sets were small, $R_{sym}(I)$ about 0.05 (overall) and 0.08 (last shell)] resulted in $R_{merge}(I)$ between 0.083 and 0.120 (overall 0.100).

2.2. Determination of the Se positions by direct methods

An isomorphous difference Patterson map calculated for the Se-Met and native data sets could be interpreted to reveal only one Se atom (Fig. 1). In retrospect, that map contained some Harker peaks corresponding to sites 2, 3 and 4, as well as some Se(1)–Se(*n*) cross vectors, which had heights and internal consistency reflecting the final Se *B* factors. However, because the Harker sections contained high spurious peaks and could not be satisfactorily interpreted in terms of all Se sites, we decided to use direct methods to locate all four Se atoms.

Table 1. *Crystallographic data and data-collection details for the native (S-IN) and Se-Met (Se-IN) protein*

Both experiments were carried out at room temperature (293 K) using Cu $K\alpha$ radiation (1.54178 Å). Data in parentheses correspond to processing with Bijvoet pairs unmerged.

	S-IN	Se-IN
Space group	$P4_32_12$	$P4_32_12$
a (Å)	66.4	66.4
c (Å)	81.4	81.4
Crystal size (mm)	$0.20 \times 0.25 \times 0.30$	$0.20 \times 0.20 \times 0.25$
Resolution (Å)	2.0	2.1
Completeness (%)	97.3	97.9 (92.3)
Completeness at maximum resolution (%)	97.0	97.2 (93.5)
No. of reflections	12521	10722 (18592)
$R_{sym}^{iso}(I)$	0.077	0.093 (0.083)

Table 2. *Isomorphous differences between Se-IN and S-IN*

Mean resolution (Å)	5.88	4.01	3.40	3.05	2.80	2.62	2.47	2.35	2.24	2.15
Mean difference (|FPH − FP|)	3.02	2.37	2.06	1.80	1.85	1.68	1.67	1.59	1.52	1.44
Mean σ (diff)	0.72	0.82	0.83	0.91	0.97	1.04	1.10	1.15	1.20	1.29

For the determination of the Se positions, only reflections with *d* spacings between 20.0 and 3.7 Å were selected, for which the isomorphous differences were significant (Table 2) and the individual $R_{sym}(I)$ (below 0.05) values were about half the inter-set $R_{merge}(I)$. Each of the intensity data sets underwent careful data reduction, which included anisotropic scaling and normalization, using a program package developed by Blessing, Guo & Langs (1995). The two *E*-value sets were used to calculate the normalized structure factors for the difference structure on the assumption that it contained four 18-electron (Se-S) 'atoms'. The ΔE set consisted of 2026 values scaled by $\sum f_j^2(\langle E^2 \rangle = 1.110)$, of which 189 with $|E| \geq 1.76$ were selected for phase determination.

We carried out the phase-determination process by means of the minimal function (Hauptman & Hahn, 1993; Weeks, DeTitta, Miller & Hauptman, 1993; DeTitta, Weeks, Thuman, Miller & Hauptman, 1994; Weeks, DeTitta, Hauptman, Thuman & Miller, 1994), using a program written by Langs (Langs, Guo & Hauptman, 1995). In the minimal-function procedure, the program generates phases from random *N*-atom structures, applies phase shifts to minimize the phase invariant residual, $R(\varphi)$, calculates an *E* map and picks the *N* highest peaks to compute phases for the next cycle of refinement. Although the minimal-function procedure has been demonstrated to successfully determine heavy-atom sites (Pt, U) from macromole-
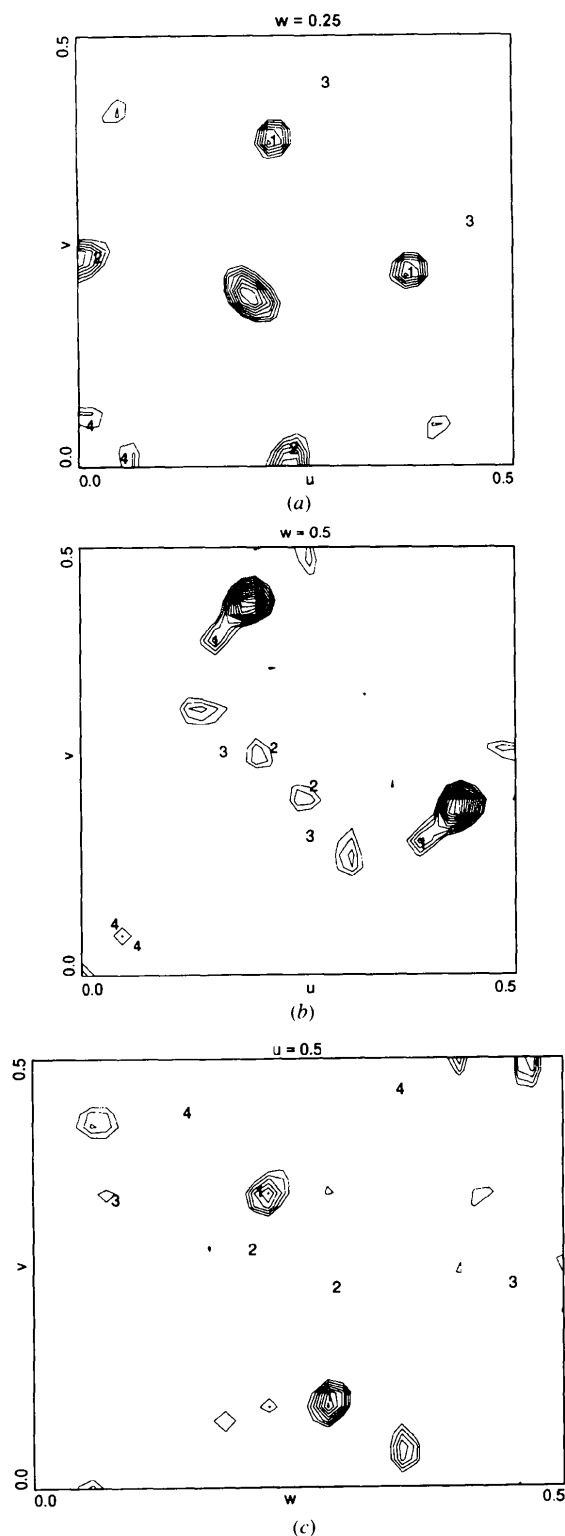
w = 0.25

(a)

w = 0.5

(b)

u = 0.5

(c)

Fig. 1. Three Harker sections of the Se-IN − S-IN difference Patterson map (3.5 Å resolution): (a) $uv\frac{1}{4}$; (b) $uv\frac{1}{2}$; (c) $\frac{1}{2}vw$. The positions expected for the Harker vectors of the four Se sites listed in Table 3 are indicated by site numbers. First contour level 2σ, contour step 0.5σ.

Table 3. *Positions and peak heights/B(iso) ($Å^2$) of the four Se atoms*

First row: from direct methods (293 K, 3.7 Å). Second row: after lack-of-closure phase refinement for Se-IN − S-IN (293 K, 3.7 Å). Third row: after lack-of-closure phase refinement based on isomorphous Se synchrotron data, 'derivative' − 'native', (Bujacz *et al.*, 1995) (108 K, 2.2 Å). The last column shows occupancy (refined). Fourth row: in least-squares refined structure (Bujacz *et al.*, 1995) (293 K, 2.1 Å).

Site	x	y	z	Relative height or B(iso)
Se1	0.69756	0.5770	0.51582	100
(Met193)	0.69655	0.57792	0.51626	1.18/7.0*
	0.69334	0.57746	0.51693	1.00
	0.69667	0.57803	0.51627	7.0†
Se2	0.86518	0.61502	0.51922	71
(Met71)	0.86591	0.61127	0.51988	1.06/15.0*
	0.86391	0.61895	0.51689	0.98
	0.86575	0.61152	0.51986	15.0†
Se3	0.63908	0.42382	0.67512	55
(Met155)	0.63215	0.41760	0.66571	0.71/43.0*
	0.62873	0.42049	0.67154	0.67
	0.63208	0.41743	0.66578	50.0†
Se4	1.03177	0.48138	0.44746	40
(Met177)	1.03167	0.48116	0.44975	0.79/34.0*
	1.02833	0.47771	0.44790	0.70
	1.03186	0.48102	0.44971	37.0†

* Occupancy (refined)/B(iso) (fixed).   † B(iso).

cular single isomorphous replacement (SIR) data (Langs *et al.*, 1995), it has never been used in a case as marginal as Se-Met derivatization and Cu $K\alpha$ data.

In performing several runs of the program, we found that although it never produced solutions with clear-cut figures of merit [$R(\varphi)$; average cosine invariant, Cosav], in all cases there were a small number of 'solutions' with better-than-random figures of merit [$R(\sigma)$ 0.45–0.50, Cosav 0.51–0.57], which appeared with a success rate of about 4% and which invariably corresponded to the same three to four atoms in the E maps. A typical run corresponded to 1000 trial sets with $N = 3$ (the program considers $N + 3$ E-map peaks with decreasing fractional weights for the extra peaks) and 20 minimal-function refinement cycles, and produced a solution [$R(\sigma)=0.49$, Cosav $= 0.52$] with four Se positions having relative peak heights of 100, 71, 55, 40 (Table 3) (the next, spurious, peak had a height of 27).

2.3. Phase determination and extension

Initially, we determined the protein phases at 3.7 Å resolution from the isomorphous differences between Se-IN and S-IN and from the anomalous signal in Se-IN (SIRSAS method), using the *PHASIT* program written by Furey & Swaminathan (1990). The values given in the *International Tables for X-ray Crystallography* (Vol. IV, 1974) for Se and Cu $K\alpha$ radiation were used as the scattering factors for the 'derivative atom'. This approach is not exactly correct, as the physical effect

Table 4. *Statistics of the phase refinement at 3.7Å (1778 reflections phased)*

Mean resolution (Å)		8.0	6.3	5.5	5.1	4.7	4.4	4.2	4.0	3.9	3.8	Overall	
													0.57*
FOM*		0.79	0.74	0.75	0.72	0.77	0.72	0.69	0.68	0.69	0.69	0.72	0.76†
													0.43‡
Phasing	iso*	3.84	3.92	4.02	3.77	3.33	3.06	2.71	2.79	3.50	3.02	3.37	
power	ano‡	2.59	2.84	2.43	2.15	2.37	1.87	2.04	1.93	1.74	2.03	2.16	

* Isomorphous data set (Se-IN − S-IN). † FOM for the 376 centric reflections in the isomorphous data set, for which $R_{Cullis} = 0.378$.
‡ Anomalous data set (Se-IN).

Table 5. *Phase extension through solvent flattening*

Resolution (Å)	R	Correlation coefficient	FOM	No. of reflections
3.70	0.234	0.964	0.916	1778*
3.40	0.216	0.970	0.965	2596
3.30	0.212	0.972	0.967	2830
3.24	0.217	0.969	0.974	2985
3.20	0.216	0.970	0.981	3087
3.16	0.217	0.970	0.983	3198
3.12	0.215	0.971	0.984	3322
3.10	0.215	0.971	0.988	3380

* Before phase extension.

corresponds to an 'atom' which is a difference between Se and S. However, for the normal component of the scattering factor and for low-resolution data, this is merely a scaling question. The parameters optimized during the lack-of-closure refinement included the atomic coordinates and occupancies, whereas the isotropic displacement parameters were adjusted (and fixed) to reflect the peak heights produced by the direct methods.

After several cycles, the lack-of-closure phase refinement converged with the statistics reported in Table 4. The phasing power (3.37 and 2.16 for isomorphous and anomalous data, respectively), overall figure of merit (FOM = 0.724), and R_{Cullis} for centric reflections (0.378) confirmed that the information content was sufficient for productive phasing.

Since the anomalous signal was low, in the next step we carried out a solvent-flattening procedure (Wang, 1985) to resolve the phase ambiguity and to improve the phases. In the solvent-flattening calculations (performed in the *PHASES* package; Furey & Swaminathan, 1990), the solvent content was progressively increased from cycle to cycle, starting at 35% and reaching 42%, which is 10% less than the real solvent content estimated from the Matthews volume (Matthews, 1968) as 52%. The phase improvement by solvent flattening at 3.7Å converged with a map-inversion R factor of 0.234, correlation coefficient of 0.964 and mean FOM of 0.916 for 1778 phased reflections (phase set 87% complete).

Next, the solvent-flattening procedure was used to carefully extend the resolution to 3.1Å. The phase extension was carried out very slowly, in seven convergent steps. It nearly doubled the number of

phased reflections (94% complete set of 3380 phases) and improved the statistical criteria of the phasing process as shown in Table 5. The phase-extension process not only extended phases beyond 3.7Å but also completed the original 3.7Å set (to 95%).

2.4. Electron-density maps

An $[|F_o|, \varphi(exp)]$ electron-density map calculated using the 3.1Å phases was inspected visually using the program *O* (Jones & Kjeldgaard, 1994). The map contains a relatively low noise level and is generally interpretable, as illustrated in Fig. 2, which shows the electron density (contoured at 1σ level) in a fragment of an extended β-sheet. The corresponding final refined model (Bujacz et al., 1995) is superimposed. Although the side chains are poorly visible, because of the limited resolution and phasing errors, the course of the main chain is generally unambiguous. There is, however, a more serious drawback consisting of truncation of electron density corresponding to surface loops connecting the relatively well represented fragments with canonical secondary structure. Obviously, this problem is related to the extensive use of solvent flattening which is known to mask out electron density on molecular surface. Although annoying, this drawback is probably not prohibitive to successful interpretation of electron-density maps and might also be minimized by more cautious use of solvent flattening.

Since we already knew the structural model when we calculated the Se/Cu $K\alpha$-phased maps, we could not objectively assess the difficulty in building a model into them. To have a more objective estimate of the quality of the above $[|F_o|, \varphi(exp)]$ map, we have calculated a correlation coefficient between the model and the map using the procedure described by Jones, Zou, Cowan & Kjeldgaard (1991) as implemented in *O* (Jones & Kjeldgaard, 1994). In these calculations, the program computes a number (between −1 and 1) for each residue, which characterizes how well a five-residue fragment (centered around the residue in question) of a model fits its electron density. In our calculations, which are summarized in Table 6, all model atoms (including side chains) were considered. The overall fit between the Se/Cu $K\alpha$-phased 3.1Å map (the present test) and the corresponding structural model refined to $R = 0.148$ at 2.1Å resolution (Bujacz et al., 1995) is

Table 6. *Correlation coefficients between models and maps*

Model	Map	Map resolution (Å)	Correlation coefficient*									
			D64†	D121†	E157†	M193	M71	M177	M155	T80‡	Max.	Overall
SERT§	Present test	3.1	0.65	0.46	0.32	0.47	0.57	0.71	0.02	0.75	0.76	0.51
SERT§	$2F_o - F_c$	2.1	0.93	0.88	0.83	0.37	0.33	0.53	0.52	0.92	0.96	0.84
SER¶	MIR based on synchrotron Se anomalous data	2.2	0.91	0.87	0.78	0.88	0.88	0.93	0.92	0.80	0.95	0.85

* Correlation coefficients were calculated according to Jones, Zou, Cowan & Kjeldgaard (1991), using *O* (Jones & Kjeldgaard, 1994).
† Active-site residue.   ‡ Residue within protein core.   § Se-Met protein, structure refined from Se-IN data ($R = 0.148$, resolution $= 2.1$ Å) (Bujacz *et al.*, 1995).   ¶ Se-Met protein, structure refined from low-temperature synchrotron data ($R = 0.139$, resolution $= 1.95$ Å) (Bujacz *et al.*, 1995).

0.51. This result indicates a very significant degree of correlation, particularly in view of an analogous comparison based on the final 2.1 Å $2F_o - F_c$ map, for which the correlation coefficient is 0.84. (As an interesting by-product of our comparison, we point out that the latter correlation is at the same level as that calculated for the original MIR map based on synchrotron Se anomalous data, from which the structure was originally determined; Bujacz *et al.*, 1995.) However, the present correlation of 0.51 is not representative of the interpretable regions of the map, because it is severely influenced by the surface features of the model missing in the map (see above). This is illustrated by the poor fit at Se-Met155, which immediately follows a surface-loop segment, and by the reduced fit (in comparison with the other two active-site residues) at E157. This residue has a long side chain extending from the center of the active site (between D64 and D121), which itself is quite exposed, to the protein surface (Bujacz *et al.*, 1995). At residues embedded in the protein core, the correlation is high and only somewhat lower than in the other two (excellent) cases illustrated in Table 6.

3. Discussion

The only successful attempt to date to phase a protein structure by using anomalous scattering of exclusively light atoms and data collected with a laboratory X-ray source resulted in the structure of crambin (Hendrickson & Teeter, 1981). In that case, however, the data extended to 1.5 Å, and the high resolution was crucial to the successful phasing. Other structure determination efforts relied on the use of Se-Met data collected with Cu $K\alpha$ radiation to provide partial phasing information, in addition to data from traditional heavy-atom derivatives. An example of such phasing is the structure of a complex of human immunophilin FKBP-12 with the drug FK506, in which Se-Met contributed significant information together with Hg and Pt derivatives (Van Duyne, Standaert, Karplus, Schreiber & Clardy, 1993). A less significant contribution of Se-Met data to phasing (phasing power 0.31) was reported for the structure of the GreA transcript cleavage factor (Stebbins *et al.*, 1995). To the best of our knowledge, our work is the first attempt to show the possibility of phasing a protein structure by using exclusively the limited-resolution Se-
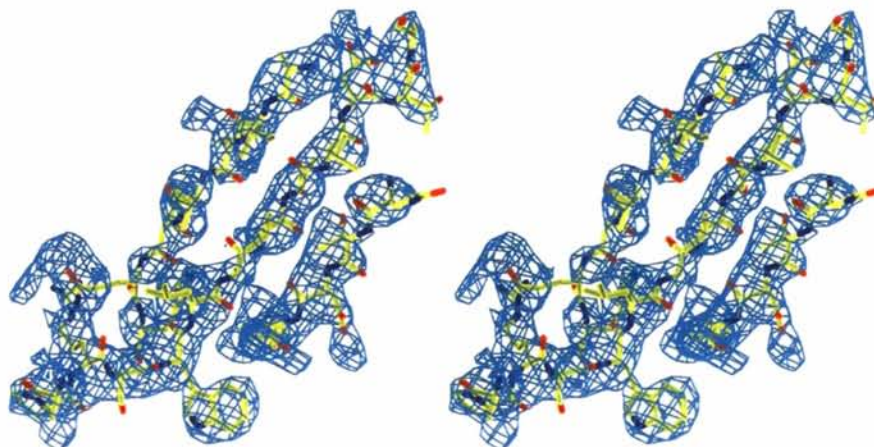

Fig. 2. Stereoview of the electron-density map based on experimental phases extended to 3.1 Å. The map is contoured at the 1σ level and is superimposed on the refined structural model (Bujacz *et al.*, 1995) in a β-sheet region (residues 62–65, 72–84, 87–93). The figure was generated with *O* (Jones & Kjeldgaard, 1994).

Met data that can be collected with standard laboratory equipment.

The maps resulting from the phase calculations based on experimental data collected without recourse to tunable X-ray sources clearly show that with good quality data the isomorphous and anomalous signal can provide productive phasing even for relatively light atoms, such as Se. The laboratory X-ray source cannot be a substitute for a synchrotron-based data-collection facility. However, since access to synchrotron sources usually involves considerable delays and is not always practical, it is important to recognize that under certain conditions is is not absolutely necessary. The quality of the experimental map is such that we would have no difficulty in fitting to it a crude model of a target protein, such as a homology-based model, whereas a *de novo* fit would be more problematic. We stress, however, that the data used here were not optimized for the purpose of phasing as described above, and that these crystals were not the best diffracting ones. We could expect better results in a properly designed experiment. For example, even clearer maps were obtained in our attempt to phase the structure of the FKBP-12/FK506 complex (mentioned above) using the Cu $K\alpha$ Se-Met data provided by A. Karplus. In this case, we phased the data extending to 3 Å, with the resulting $R_{Cullis} = 0.303$, phasing power of 2.47 and FOM of 0.578. After solvent flattening, the map-inversion R factor was 0.196, the correlation coeffcient was 0.977 and FOM was 0.0873 for 2075 reflections. The phases resulting from these claculations differed by about 45° from the final phases calculated from the refined model (A. Karplus, personal communication).

In the present tests, the r.m.s. deviation between the Se/Cu $K\alpha$-derived and the final model (Bujacz *et al.*, 1995) phases in the resolution range 10.0–3.1 Å (used for generating the map in Fig. 2) is about 60° (Table 7). This value is somewhat higher than the value for FKBP-12/FK506, but is still very encouraging and indicates successful phasing. For reflections with d spacings greater than 3.7 Å (original SIRSAS phase determination) the r.m.s.$(\Delta\varphi)$ value is about 50°, in very good agreement with the number of FKBP-12/FK506. At higher resolution the r.m.s.$(\Delta\varphi)$ rises to about 80° (still an acceptable value) indicating that the quality of phases derived by phase extension is not as good as of those determined directly. For comparison, the r.m.s.$(\Delta\varphi)$ values for the SIRSAS phases (to 2.2 Å resolution) calculated from the MAD data collected using synchrotron radiation (Bujacz *et al.*, 1995) are in the order of 35° (before solvent flattening) to 27° (after solvent flattening) (to 3.7 Å, respectively, 32 and 26°), attesting to the unquestionable power of tunable anomalous diffraction.

The strength of the practical proof of the method described above is somewhat weakened by the fact that during the present tests the structure was determined

Table 7. *R.m.s. difference (°) between Se/Cu $K\alpha$-derived (present test) and final model phases (Bujacz et al., 1995) versus resolution (according to Brünger, 1992)*

Resolution range (Å)	No. of reflections	Phase difference	Accumulative
6.17–10.0	353	55.0	55.0
5.11–6.17	337	47.5	51.3
4.53–5.11	344	51.7	51.4
4.15–4.53	334	53.2	51.9
3.87–4.15	330	52.3	52.0
3.65–3.87	331	59.8	53.2
3.48–3.65	306	66.6	55.0
3.33–3.48	317	74.1	57.3
3.21–3.33	318	82.8	60.0
3.10–3.21	309	81.1	62.0

through the use of synchrotron MAD/SIRSAS phasing (Bujacz *et al.*, 1995, 1996). However, the prior knowledge of the structure has never been used when testing the Se-Met/Cu $K\alpha$ method in a way other than to select the space group enantiomorph and for comparing the results with reality. In retrospect, the reason why we had not pursued structure solution by using the Cu $K\alpha$ data alone was that we were not fully convinced how far to trust the Se locations determined by direct methods and ultimately managed to gain access to a synchrotron source in a timely manner. In conclusion, we not only demonstrate a successful phasing of Se-Met protein Cu $K\alpha$ data but also want to encourage protein crystallographers to put more faith in the quality of the diffraction data that become routinely available through the use of modern equipment and processing methods.

References

Bijvoet, J. M. (1949). *Proc. R. Acad. Amsterdam*, **52**, 313–314.

Blessing, R. H., Guo, D. Y. & Langs, D. A. (1995). *Acta Cryst.* D**51**, 1020–1024.

Brünger, A. (1992). *X-PLOR: A system for X-ray crystallography and NMR.* Yale University Press, New Haven, USA.

Bujacz, G., Jaskólski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1995). *J. Mol. Biol.* **253**, 333–346.

Bujacz, G., Jaskólski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1996). *Structure,* **4**, 89–96.

DeTitta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Acta Cryst.* A**50**, 203–210.

Furey, W. & Swaminathan, S. (1990). *PHASES - A program package for the processing and analysis of diffraction data sets for macromolecules.* Am. Crystallogr. Assoc. Meet. Program Abstract 18, 73.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. Ser. A,* **225**, 287–307.

Hauptman, H. A. & Hahn, F. (1993). *Acta Cryst.* D**49**, 3–8.

Hendrickson, W. A. (1985). *Trans Am. Crystallogr. Assoc.* **21**, 11–21.

Hendrickson, W. A. (1991). *Science,* **254**, 51–58.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.

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

Jaskólski, M. & Dauter, Z. (1996). *Pol. J. Chem.* In the press.

Jones, T. A. & Kjeldgaard, M. (1994). *O - The Manual.* Uppsala University, Sweden.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Karle, J. (1980). *Int. Quant. Chem.* **7**, 356–367.

Kulkosky, J., Katz, R. A., Merkel, G. & Skalka, A. M. (1995). *Virology,* **205**, 448–456.

Langs, D. A., Guo, D. Y. & Hauptman, H. A. (1995). *Acta Cryst.* D**51**, 1020–1024.

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

Otwinowski, Z. (1992). *An Oscillation Data Processing Suite for Macromolecular Crystallography.* Yale University, New Haven, USA.

Stebbins, C. E., Borukhov, S., Orlova, M., Polyakov, A., Goldfarb, A. & Darst, S. A. (1995). *Nature (London),* **373**, 636–640.

Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.

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

Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.

Weeks, C. M., DeTitta, G. T., Miller, R. & Hauptman, H. A. (1993) *Acta Cryst.* D**49**, 179–181.