Acta Cryst. (1998). D**54**, 799–804

The Use of *SnB* to Determine an Anomalous Scattering Substructure

G. David Smith,[a,b]* Bhushan Nagar,[c] James M. Rini,[c] Herbert A. Hauptman[a] and Robert H. Blessing[a]*

[a]*Hauptman–Woodward Medical Research Institute, 73 High Street, Buffalo, NY 14203, USA,* [b]*Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA, and* [c]*Departments of Molecular and Medical Genetics and Biochemistry, University of Toronto, Toronto, Ontario, Canada, M5S 1A8. E-mail: smith@hwi.buffalo.edu*

(*Received 25 August 1997; accepted 1 December 1997*)

Abstract

The positions of eight Se atoms in a selenomethionyl 35 kDa protein were determined at 2.0 and 2.5 Å resolution using the direct-methods program *SnB*. Data at the selenium peak, edge and remote wavelengths were measured and processed independently. Anomalous difference E magnitudes at each wavelength were derived by two different procedures: renormalized diff$E$ values were calculated according to the equation $\text{diff}E = \{\sum[(f + f')^2 + f''^2]||E_+| - |E_-||\}^{1/2}/2q(\sum f''^2)^{1/2}$, where q is a least-squares fitted renormalization function of $\sin\theta/\lambda$ such that $\langle\text{diff}E^2\rangle = 1.0$; and difference E magnitudes were calculated from ΔF^2. Locally normalized E magnitudes corresponding to $|F_A|$ were also derived from the combination of the data at all three wavelengths through the use of the *MADSYS* program suite. Each of the independent sets of anomalous difference E magnitudes was capable of producing the correct solution, as did the E data obtained from the F_A data. Higher success rates with *SnB* were observed for the 2.0 Å peak and edge diff$E$ data.

1. Introduction

During the past ten years, the use of MAD phasing to solve protein structures has increased considerably. This method relies on the presence of an atom or atoms that provide an anomalous signal at an appropriately chosen wavelength, and the use of a synchrotron allows the wavelength to be chosen to match the absorption edge of the anomalous scatterer. Typically, data are measured at three wavelengths: at minimum f' (edge), maximum f'' (peak) and at a wavelength remote from the edge and peak wavelengths. The *MADSYS* suite of programs (Hendrickson *et al.*, 1988; Hendrickson, 1991), based upon the algebraic approach derived by Karle (1980), has been developed to extract the amplitudes for normal scattering from the anomalous scattering substructure ($|F_A|$), the amplitudes for normal scattering due to all atoms ($|F_T|$) and the difference in phase between the two ($\Delta\varphi = \varphi_T - \varphi_A$). In favorable cases, a Patterson map based upon the squared values of $|F_A|$ can be used to obtain the positions of the anomalous substructure. These positions can then be used to calculate the values of φ_A, and hence values of φ_T. An electron-density map based upon the values of $|F_T|$ and φ_T would then, in principle, reveal the structure of the entire protein, and outstanding electron-density maps have been obtained (Burling *et al.*, 1996). This procedure, or the alternative approach whereby the MAD data are treated as a special case of the MIR formulation, can lead to high-quality electron-density maps in a straightforward fashion, provided that the anomalous scattering substructure can be determined from the Patterson map.

The introduction of selenium as selenomethionine into a protein has the advantage of introducing anomalously scattering atoms directly into the structure, thereby eliminating the need for heavy-atom screening. Furthermore, the K edge of selenium occurs at approximately 0.98 Å, a wavelength easily obtained at a synchrotron source. However, since the natural-occurrence abundance of methionine in proteins is about 5%, the number of Se atoms will in general increase as the size of the protein increases, making it more difficult to determine the positions of the anomalous substructure using Patterson methods. *MULTAN* (Germain *et al.*, 1971), in conjunction with E magnitudes based upon the derived values of $|F_A|$, has been used to determine the positions of four Se atoms (Athappilly & Hendrickson, 1995) and to verify positions obtained by Patterson methods (Yang *et al.*, 1990), but it is not obvious that these methods will be applicable to larger anomalous substructures.

The development of the minimal function (Debaerdemacker & Woolfson, 1983; Hauptman, 1988, 1991) and its use in the *SnB* program (Miller *et al.*, 1994) have met with considerable success in solving difficult small-molecule structures (Miller *et al.*, 1993). In the past several years, several small protein structures have been solved *ab initio* by *SnB*, including the redetermination of the gramicidin A dimer, crambin (Weeks *et al.*, 1995) and a scorpion protein toxin (Smith *et al.*, 1997) as well as several previously unknown structures, including Er-1 (Anderson *et al.*, 1996) and alpha-1 peptide (Prive *et al.*, 1995). However, atomic resolution data were available for all of these small proteins, allowing individual atoms to be selected and identified on the basis of the position

Table 1. *Data-measurement statistics*

	Edge	Peak	Remote
Wavelength (Å)	0.9794	0.9792	0.9678
Resolution (Å)	∞–2.0	∞–2.0	∞–2.0
Total data measured	125648	130972	128694
Unique data	40586	40753	41209
R_{merge}	0.054	0.055	0.052
Overall completeness (%)	89.4	89.9	91.3
$I \geq 3\sigma(I)$ (%)	90.5	88.4	89.8
Last resolution shell statistics (2.1–2.0 Å)			
Completeness (%)	52.2	54.5	59.2
$I \geq 3\sigma(I)$ (%)	72.6	66.6	72.4

Table 2. *E-value discrimination by diffE*

	Edge	Peak	Remote
Total number of pairs of data	20904	20972	21216
Number of pairs of data with†			
$\min[\lvert E_+\rvert/\sigma(E_+), \lvert E_-\rvert/\sigma(E_-)] \geq 3.0$	19291	18935	19484
$\lvert\lvert E_+\rvert - \lvert E_-\rvert\rvert/[\sigma^2(E_+) + \sigma^2(E_-)]^{1/2} \geq 1.0$	9589	9566	8335
diffE/σ(diff$E$) ≥ 3.0 (∞–2.0 Å)‡	2210	2393	1224
diffE/σ(diff$E$) ≥ 3.0 (∞–2.5 Å)‡	1900	2074	1085

† The following tests were applied sequentially, *i.e.* data which did not pass a particular test were omitted from further consideration. ‡ $\text{diff}E = \{\sum[(f+f')^2 + f''^2]\}\lvert\lvert E_+\rvert - \lvert E_-\rvert\rvert\}^{1/2}/2q(\sum f''^2)^{1/2}$ where q [$q = q_0 \exp(q_1 s^2 + q_2 s^4)$ and $s = \sin\theta/\lambda$] is a renormalization function such that $\langle \text{diff}E^2\rangle = 1.0$.

of the peaks obtained from a Fourier electron-density map.

The application of *SnB* to the determination of the positions of eight Se atoms in a selenomethionyl 35 kDa protein at 2.0 and 2.5 Å resolution is reported here. Three sets of anomalous difference E magnitudes were derived from the edge, peak and remote data, and independent analysis of each of the three sets of diff$E$ magnitudes produced the correct anomalous substructure. Using the three wavelength data sets jointly, this same substructure could also be determined from the E magnitudes derived from the F_A values obtained from *MADSYS* (Hendrickson *et al.*, 1988).

2. Experimental

Three sets of data for C3d, a fragment of complement component C3, were measured at station F2 at the Cornell High Energy Synchrotron Source. The molecule crystallizes in space group $P2_12_12_1$ and data were integrated, scaled and merged with *DENZO* and *SCALEPACK* (Otwinowski, 1993) to produce three sets of independent measurements at three different wavelengths, treating the Bijvoet pairs as independent measurements. Data-measurement statistics are given in Table 1. The anomalous substructure has been determined independently using two different procedures: (i) using classical Patterson and Fourier methods (Nagar *et al.*, 1998) employing the programs *HEAVY* (Terwilliger *et al.*, 1987) and *PHASES* (Furey & Swaminathan, 1997) and (ii) using the probabilistic direct-methods *SnB* program as described in this report. For the *SnB* phasing, the output of *SCALEPACK* for the data at the three different wavelengths (as shown in Table 1) were reprocessed with the *DREAR* program package (Blessing, 1989; Blessing *et al.*, 1998). Bayesian processing (French & Wilson, 1978) and local scaling (Matthews & Czerwinski, 1975; Blessing & Smith, 1998) were applied to each of the three sets of data. E magnitudes were obtained by a Wilson analysis that estimates the mean and the r.m.s. deviation from the mean of the unit-cell distribution of anisotropic atomic displacements (Blessing *et al.*, 1996). The equivalent values of $\langle B_{iso}\rangle$ were 19.9, 21.6 and 20.0 Å^2 for the edge, peak and remote data, respectively.

A normal probability plot (Howell & Smith, 1992) was constructed for each of the three sets of data, comparing $(F_+^2 - F_-^2)/[\sigma^2(F_+^2) + \sigma^2(F_-^2)]^{1/2}$ to the values expected for a normal distribution of errors. Each plot was a smooth sigmoidal curve, indicating that some of the differences in magnitude of the Bijvoet pairs were neither randomly nor normally distributed and hence were systematic and significant. However, the inner portions of each plot, which corresponded to approximately 50% of the data, were quite linear with slopes ranging from 2.0 to 2.5 and with intercepts of approximately zero. In principle, the slopes should be equal to unity; the increased slope suggested that either the standard deviations of the measurements were underestimated by a factor of approximately two (a not uncommon occurrence) or that a small but real difference did exist between those Bijvoet pairs.

2.1. ΔE *renormalization*

While normalized structure-factor differences corresponding to the anomalous substructure could be calculated easily ($\Delta E_{anom} = \lvert\lvert E_+\rvert - \lvert E_-\rvert\rvert$), the average square of such E-magnitude differences over the entire range of $\sin\theta/\lambda$, and in individual shells of $\sin\theta/\lambda$, would not equal unity and would have the effect of altering the value of $A \propto \lvert\Delta E_1 \Delta E_2 \Delta E_3\rvert$ for each triple. Furthermore, the differences in magnitude between Bijvoet pairs were found to be insignificant for approximately half of the data, as noted from the normal probability analysis. For these reasons, the difference E magnitudes were renormalized using the program *DiffE* (Blessing & Smith, 1998; Blessing *et al.*, 1998). Various tests were applied to the difference E magnitudes in order to exclude data which are not significantly different from zero and to select those data which exhibit the strongest reliable dispersive signals, Table 2. Three such tests were applied: (i) pairs of data were excluded for which the minimum of either $\lvert E_+\rvert/\sigma(E_+)$ or $\lvert E_-\rvert/\sigma(E_-)$ was less than a constant (3.0), which has the effect of including only those pairs of E data which are significantly different from zero, (ii) data were excluded for which the absolute difference between $\lvert E_+\rvert$ and $\lvert E_-\rvert$ was less

than some constant multiplied by the standard deviation of their differences, $\{||E_+| - |E_-||/[\sigma^2(E_+)+\sigma^2(E_-)]^{1/2} \leq 1.0\}$, thus excluding those difference magnitudes which are not significantly different from their standard deviations, and (iii) renormalized diff$E$ data† were excluded if the value of diff$E$ was less than a constant (3.0) times the standard deviation of diff$E$. Even with these stringent tests, it was important to examine the distribution of $\langle \text{diff}E^2 \rangle$ as a function of $\sin\theta/\lambda$. The distribution for the edge and remote data, illustrated in Figs. 1(*a*) and 1(*b*), was found to be relatively flat, as expected. However, the distribution for the peak data, shown in Fig. 1(*c*), was found to increase significantly beyond a resolution of 2.5 Å. This same behavior has been observed for data from other test cases, and at present no explanation has been formulated to explain the deviation of $\langle \text{diff}E^2 \rangle$ from unity. The non-linearity of $\langle \text{diff}E^2 \rangle$ as a function of $\sin\theta/\lambda$ may be due to small errors in f' and f''.

2.2. ΔF^2 *normalization*

In a second data-processing procedure, the absolute differences between Bijvoet pair amplitudes ($\Delta F^2 = |F_+^2 - F_-^2|$) were calculated for all three sets of data. Following the application of Bayesian statistics to the differences, normalized structure factors were calculated using Wilson statistics as described above. Difference E magnitudes were rejected from further use if $|E| < 3\sigma(E)$.

2.3. $|F_A|$ *normalization*

In a third set of calculations, the values of $|F_A|$ were obtained from the *MADSYS* program package using the data at three different wavelengths. E magnitudes were obtained from the F_A values at 2.0 Å resolution through a simple local normalization procedure.

2.4. *SnB phasing*

Given appropriate values for the normalized structure factors corresponding to the anomalous substructure, the determination of the atomic positions for the anomalous scatterers is straightforward using Version 1.5.0 of *SnB* (Miller *et al.*, 1994). In each case, the 500 largest diff$E$ values, E values from the normalized ΔF^2 values, or the locally normalized F_A values were used to generate a maximum of 5000 triple invariants. Eight random-atom positions were generated for each of 100 trial structures, and each trial structure was subjected to 20 *SnB* cycles, each cycle consisting of phase refinement and Fourier filtering. Fourier filtering consisted of picking the 16 largest peaks from the E map for further consideration, while each phase-refinement cycle

† $\text{diff}E = \{\sum[(f+f')^2 + f''^2]||E_+| - |E_-||\}^{1/2}/2q(\sum f''^2)^{1/2}$, where q is a least-squares-fitted renormalization function, $q = q_0\exp(q_1 s^2 + q_2 s^4)$ and $s = \sin\theta/\lambda$ (Blessing *et al.*, 1998), such that $\langle \text{diff}E^2 \rangle = 1.0$.

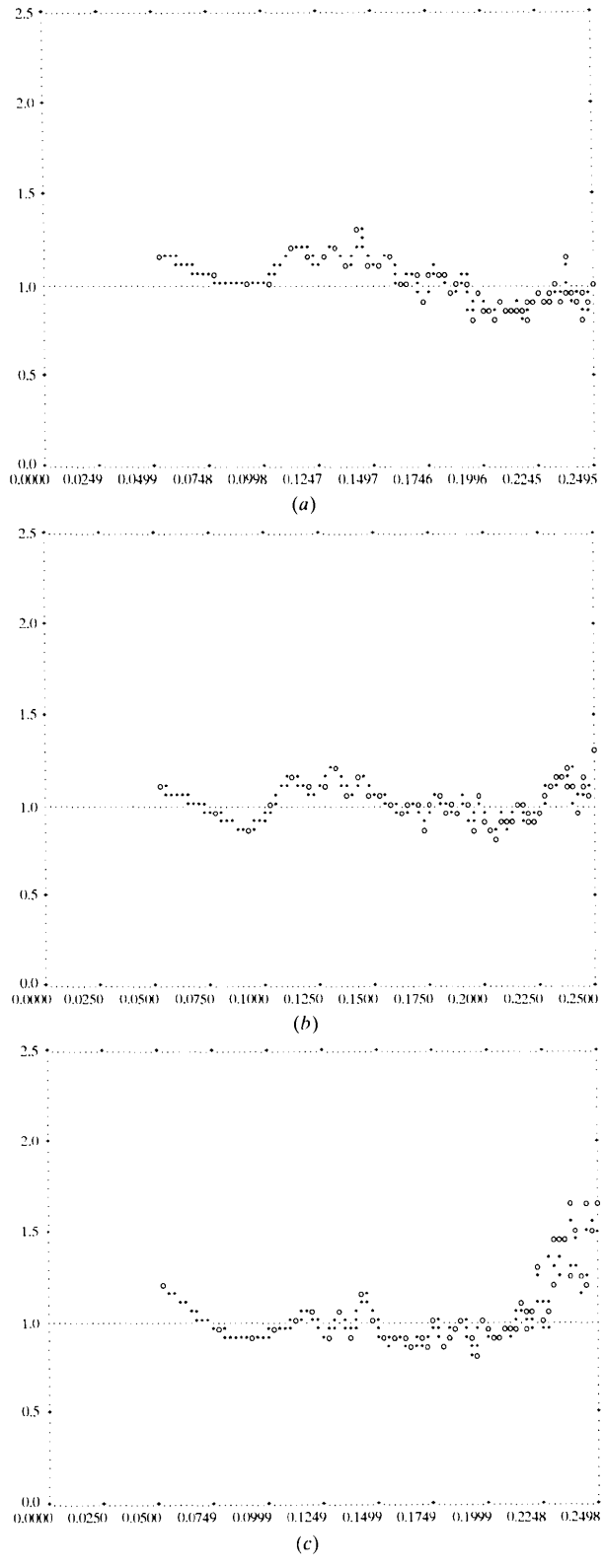

Fig. 1. Plots of $\langle E^2 \rangle$ *versus* $\sin\theta/\lambda$ for the (*a*) edge, (*b*) remote and (*c*) peak diff$E$ data. ○, local average values; ∗, spline-interpolated values.

Table 3. *SnB solution results*

Wavelength	Normalization method†	Resolution	Number of solutions	R_{min} range	Number of correct peaks less than 0.5 Å from Se	Number of correct peaks less than 1.0 Å from Se	Mean displacement, best R_{min}	R.m.s. displacement, best R_{min}
Edge	ΔE	2.0	6	0.246–0.250	8	8	0.135	0.147
			7	0.389–0.435	0	0		
			1	0.512	2	3		
Peak	ΔE	2.0	12	0.218–0.219	8	8	0.124	0.140
			6	0.345–0.348	0	0		
Remote	ΔE	2.0	9	0.216–0.218	8	8	0.179	0.214
			1	0.217	7	8		
			1	0.237	7	7		
			8	0.350–0.391	0	0		
			2	0.482–0.520	0	0		
			1	0.520	0	1		
Edge	ΔE	2.5	7	0.258–0.260	8	8	0.129	0.136
			1	0.280	7	7		
			4	0.397–0.402	0	0		
Peak	ΔE	2.5	7	0.234–0.235	8	8	0.186	0.193
			7	0.378–0.390	0	0		
Remote	ΔE	2.5	5	0.226	8	8	0.192	0.225
			4	0.225–0.226	7	8		
			1	0.233	7	7		
			4	0.352–0.362	0	0		
Edge	ΔF^2	2.0	6	0.333–0.334	8	8	0.125	0.148
			4	0.351–0.357	7	7		
			2	0.527–0.557	0	1		
Peak	ΔF^2	2.0	2	0.320–0.327	7	7	0.973	2.447
			1	0.331	6	6		
			1	0.336	7	7		
			7	0.442–0.511	0	1		
Remote	ΔF^2	2.0	2	0.384–0.385	8	8	0.161	0.180
			6	0.387–0.404	7	7		
			1	0.511	0	0		
			3	0.523–0.554	0	1		
Three wavelength	F_A^2	2.0	4	0.278–0.280	8	8	0.155	0.177
			2	0.282–0.293	7	7		
			1	0.358	0	0		

† For ΔE renormalization, Bijvoet pairs were processed independently and ΔE was derived from the difference between $|E_+|$ and $|E_-|$ (Blessing & Smith, 1998; Blessing *et al.*, 1998); for ΔF^2 normalization, the difference in the squared amplitudes was normalized; and for F_A^2 normalization, F_A values derived from *MADSYS* were locally normalized.

consisted of three cycles of parameter-shift refinement, with a maximum of a 90° phase shift for each phase. At the end of the 20 cycles, two additional cycles of Fourier refinement were performed, accepting the eight largest peaks in the E map. The edge, peak and remote diff$E$ magnitudes at 2.0 and 2.5 Å resolution, the magnitudes obtained from normalization of ΔF^2 at 2.0 Å resolution and the locally normalized F_A values at 2.0 Å resolution were used in the procedure described above to generate solutions for the anomalous substructure.

The trace of the minimal function, R_{min}, for correct solutions was found to exhibit a dramatic decrease in the first several *SnB* cycles and then remained constant at its final minimal value. Previous experience has shown this behavior to be typical for a correct solution. In contrast, R_{min} for non-solutions showed no decrease as a function of cycle number. The bimodal distribution of the R_{min} histograms could also be used to predict which solutions were correct. Table 3 contains the number of correct solutions for values of R_{min} less than 0.5, which are those which are outside the major distribution of non-solutions; also included are the mean and r.m.s. displacements of the correct atomic positions from the peaks obtained from the solution with the lowest value of R_{min}. The top eight peak positions, in conjunction with the edge, peak and remote data and the *MADSYS* data at 2.0 Å resolution, were used to calculate phases for the largest 500 E values for each, and to generate an E map. A typical portion of the E maps in the vicinity of the Se atom with the smallest occupancy is illustrated in Fig. 2. *SnB* ranking of peaks, peak heights and rankings from the E maps, along with the refined thermal parameters and occupancies are given in Table 4.

3. Discussion

Version 1.5.0 of *SnB* was originally designed for use with atomic resolution data, but since the selenium positions in a selenomethionyl protein are resolved even at a

Table 4. *SnB rankings of peaks and E-map peaks*

Atom†	B_{iso}	G‡	*SnB* rank				E density§	Map rank	P density	Map rank	R density	Map rank	F_A density	Map rank
			E¶	P¶	R¶	F_A								
Se1	20.4	0.50	8	8	8	8	9.75	8	11.29	8	8.33	9	10.90	8
Se22	17.0	1.02	2	1	2	3	15.36	4	16.41	3	16.86	1	16.07	3
Se25	17.4	1.00	4	4	6	4	17.51	2	16.59	2	14.08	6	15.83	5
Se136	16.6	1.06	3	2	1	1	16.61	3	16.43	4	15.55	5	19.39	1
Se148	17.0	0.91	5	5	4	2	15.10	5	15.17	5	15.65	3	18.76	2
Se188	17.3	0.82	7	6	5	6	13.19	6	14.85	6	16.47	2	15.05	6
Se206	21.6	0.57	6	7	7	7	11.61	7	11.64	7	11.28	7	12.93	7
Se281	16.7	1.03	1	3	3	5	18.99	1	16.92	1	15.61	4	15.98	4
							7.67	9	9.37	9	9.67	8	8.55	9
							6.99	10	6.52	10	7.69	10	6.06	10

† The number following Se is the residue number of the selenomethionyl residue. ‡ G is the occupancy. § Edge, peak and remote densities are reported as the number of standard deviations from the mean. ¶ E, P and R refer to edge, peak and remote wavelengths, respectively.

resolution of 3.0 Å, it is appropriate to select peak positions in the real-space portion of each *SnB* cycle.

Table 3 clearly shows that the success rate of *SnB* is considerably better for the diff$E$ magnitudes as compared to the E values generated from the ΔF^2 differences or the E values derived from the F_A data. R_{min} for a correct solution varies from 0.216 to 0.258 for the 2.0 and 2.5 Å resolution data using the diff$E$ values; in contrast, R_{min} ranges between 0.320 and 0.384 for the normalized difference amplitudes and between 0.278 and 0.280 for the E values derived from the F_A data. At 2.0 Å resolution only eight complete correct solutions are obtained from the normalized difference amplitudes for the edge and remote data, and positions for only seven of the eight atoms are obtained for the peak data in two solutions. Only four correct and complete solutions are obtained from the *MADSYS* treatment of the data. In contrast, at 2.0 Å resolution the diff$E$ magnitudes provide a total of 27 complete and correct solutions for the three different wavelengths, while at 2.5 Å resolution the number of solutions is reduced to 19.

A comparison of the number of correct atoms and correct solutions in a given R_{min} range suggests that it might be difficult to identify the correct solution. This is not the case as complete and correct solutions for the peak and edge data lie in a very narrow range of R_{min}. For example, 12 complete and correct solutions lie in an R_{min} range of 0.218–0.219 for the peak data at 2.0 Å. The next range of R_{min}, all of which correspond to non-solutions, varies from 0.345 to 0.348, giving a clear indication of which solutions are correct. Consistency in the peak positions should provide another indication of which sets of solutions are correct, but in this case the peak positions from the above six incorrect solutions from the peak data have mean displacements from each other of less than 0.05 Å.

The results obtained from remote data are similar to that of the edge and peak data, but it is clear that the 2.0 Å resolution data is superior to the 2.5 Å resolution data. At 2.0 Å resolution, there are ten solutions in an R_{min} range of 0.216–0.218, and one solution has only seven peaks that are less than 0.5 Å from a Se-atom position. Thus, in this case the internal consistency in the positions of the nine correct solutions would allow these solutions to be easily identified. At 2.5 Å resolution nine solutions have an R_{min} range of 0.225 to 0.226, but four of these have a single peak which lies between 0.5 and 1.0 Å from a correct Se-atom position. Thus, greater difficulty would be encountered in identifying the five correct solutions. However, in all of the above cases, the single peak position which is displaced from a correct atom position corresponds to the Se atom with the lowest occupancy and highest thermal parameters. The poorer behavior of the remote data, as compared to the peak or edge data, is most likely to be a result of the weaker anomalous signal at the remote wavelength.

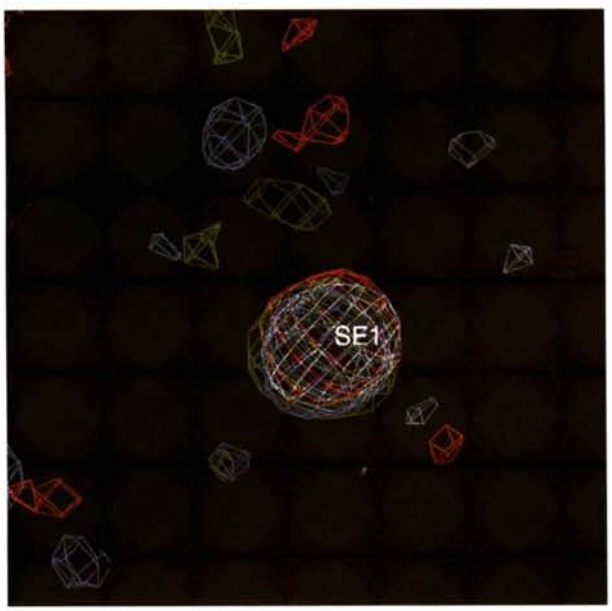

Fig. 2. E maps contoured at 3σ for the edge diff$E$ data (blue), peak diff$E$ data (green), remote diff$E$ data (red), and the locally normalized F_A data (cyan) in the vicinity of the N-terminal selenomethionyl residue which contains the Se atom (Se1) with the lowest occupancy and highest thermal parameter.

Some difficulty was encountered in choosing the various cutoffs used in the *MADSYS* package. The F_A values were found to be quite noisy and it was not uncommon to obtain many unreasonably large values of $|F_A|$. Thus, the choice of a maximum threshold for $|F_A|$ and an $|F_A|/\sigma(F_A)$ cutoff was found to be critical, as noted previously (Athappilly & Hendrickson, 1995). For the successful calculations described above, a maximum threshold of 200 was applied and $|F_A| \geq 5\sigma(F_A)$. It is interesting to note that the use of a larger value for the maximum threshold of $|F_A|$ resulted in E magnitudes that did not generate the 5000 requested triples and gave incorrect solutions. As the $|F_A|$ threshold was decreased and the ratio of $|F_A|$ to $\sigma(F_A)$ was increased, the number of triples also increased and, therefore, the ability to generate triples was a good indicator of potential success.

These results show that it is possible to obtain the coordinates of the anomalous substructure from a single-wavelength set of data, even at the remote wavelength. As three or more sets of independent data are usually available from a MAD experiment as well as the derived F_A values, the values of $R_{\min}$ and the positions obtained from the solutions from each set of data can be used to verify the correct solutions. It should be noted that equivalent solutions may differ with respect to origin and enantiomorph, and care must be taken to recognize consistent equivalent solutions. Patterson methods and heavy-atom refinement can be used to confirm the correctness of the solutions, but an Se atom with low occupancy and high thermal motion may not be obvious from the map. The use of the data-processing procedure described above, in conjunction with the direct methods program *SnB*, readily yielded the positions of the eight Se atoms in this structure. Very recently, these same procedures provided a straightforward determination of the positions of all 30 Se atoms in the 95 kDa asymmetric unit of a multimeric protein of previously unknown structure (Turner *et al.*, 1998). To the best of our knowledge, this latter result represents the record for the size of a structure determined by anomalous dispersion methods.

This research was supported by NIH grant GM-46733 and MRC MT-12458.

References

Anderson, D. S., Weiss, M. S. & Eisenberg, D. (1996). *Acta Cryst.* D**52**, 469–480.

Athappilly, F. K. & Hendrickson, W. A. (1995). *Structure*, **3**, 1407–1419.

Blessing, R. H. (1989). *J. Appl. Cryst.* **22**, 396–397.

Blessing, R. H. (1997). *J. Appl. Cryst.* **30**, 176–178.

Blessing, R. H. & Smith, G. D. (1998). *Acta Cryst.* D**54**, Submitted.

Blessing, R. H., Guo, D. Y. & Langs, D. A. (1996). *Acta Cryst.* D**52**, 257–266.

Blessing, R. H., Guo, D. Y. & Langs, D. A. (1998). *Direct Methods for Solving Macromolecular Structures.* NATO ASI Series, Vol. 507, edited by S. Fortier. pp. 47–71. Dordrecht: Kluwer Academic Publishers.

Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.

Debaerdemaeker, T. & Woolfson, M. M. (1983). *Acta Cryst.* A**39**, 193–196.

French, S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.

Furey, W. & Swaminathan, S. (1997). *Methods Enzymol.* **277**, 590–620.

Germain, G., Main, P. & Woolfson, M. M. (1971). *Acta Cryst.* A**27**, 368–376.

Hauptman, H. A. (1988). Am. Crystallogr. Assoc. Meet., Abstract R4.

Hauptman, H. A. (1991). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 324–332. Oxford: IUCr/Oxford University Press.

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

Hendrickson, W. A., Smith, J. A., Phizackerley, R. P. & Merritt, E. A. (1988). *Proteins*, **4**, 77–88.

Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81–86.

Karle, J. (1980). *Int. J. Quantum Chem.* **7**, 357–367.

Matthews, B. W. & Czerwinski, E. W. (1975). *Acta Cryst.* A**31**, 480–497.

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.

Nagar, B., Jones, R. G., Diefenbach, R. J, Isenman, D. E. & Rini, J. M. (1998). *Science*, **280**, 1277–1281.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Prive, G., Ogihara, N., Wesson, L., Cascio, D. & Eisenberg, D. (1995). Am. Crystallogr. Assoc. Meet., Abstract W008.

Smith, G. D., Blessing, R. H., Ealick, S. E., Fontecilla-Camps, J. C., Hauptman, H. A., Housset, D., Langs, D. A. & Miller, R. (1997). *Acta Cryst.* D**53**, 551–557.

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

Turner, M. A., Yuan, C.-S., Borchardt, R. T., Hershfield, M., Smith, G. D. & Howell, P. L. (1998). *Nature Struct. Biol.* **5**, 369–376.

Weeks, C. M., Hauptman, H. A., Smith, G. D., Blessing, R. H., Teeter, M. M. & Miller, R. (1995). *Acta Cryst.* D**51**, 33–38.

Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). *Science*, **249**, 1398–1405.