Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Use of differentially substituted selenomethionine proteins in X-ray structure determination

Nadine C. Gassner and Brian W. Matthews*

Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, 1229 University of Oregon, Eugene, OR 97403-1229, USA

Correspondence e-mail: brian@uoxray.uoregon.edu

Received 6 July 1999
Accepted 15 October 1999

Using heavily methionine-substituted T4 lysozyme as an example, it is shown how the addition or deletion of a small number of methionines can simplify the location of selenium sites for use in MAD phasing. By comparing the X-ray data for a large number of singly substituted lysozymes, it is shown that the optimal amino acid to be substituted by methionine is leucine, followed, in order of preference, by phenylalanine, isoleucine and valine. The identification of leucine as the first choice agrees with the ranking suggested by the Dayhoff mutation probability, *i.e.* by the frequency of amino-acid substitutions in the sequences of related proteins. The ranking of the second and subsequent choices, however, differ significantly.

When a selenomethionine-substituted protein is used to facilitate X-ray structure determination (Hendrickson *et al.*, 1990; Hendrickson, 1991), the first requirement is to locate the positions of the Se atoms. Although increasingly powerful methods are becoming available (Weeks & Miller, 1999), this can become more difficult as the number of sites is increased and may be a limitation for larger proteins which contain many methionines or for cases in which there are several protein molecules in the asymmetric unit. The situation can be illustrated by a mutant of T4 lysozyme that contains a total of 12 methionines. It should be emphasized that the location of 12 Se atoms is by no means at the limit of present technology; a successful 70-site analysis has been reported (Deacon & Ealick, 1999). The lysozyme example is used because the data are available and serve to illustrate an approach which should be applicable in more demanding contexts. Five of the methionines present in the mutant lysozyme, Met1, Met6, Met102, Met106 and Met120, are present in the wild-type protein. The additional seven were introduced by the following site-directed substitutions: Leu84→Met, Leu91→Met, Leu99→Met, Leu118→Met, Leu121→Met, Leu133→Met and Phe153→Met (Gassner *et al.*, 1996). The Harker section of the Patterson map for this variant, calculated with the coefficients $(F_{\text{12-SeMet}} - F_{\text{12-Met}})^2$, is shown in Fig. 1(*a*). Although there is reasonably good agreement between the locations of the expected vector peaks and those observed, the interpretation of such a map is complicated by peak overlap and by the weakness of some of the expected peaks.

In such cases, one can obtain a much simpler starting point by engineering a related protein in which a small number of methionines have either been added or subtracted. This approach can be illustrated by consideration of the 12-methionine and the 10-methionine lysozymes which are identical except for the Leu133→Met and Phe153→Met replacements. By combining the diffraction data for crystals of these two proteins, one can calculate a Patterson map with coefficients $(F_{\text{12-SeMet}} - F_{\text{10-SeMet}})^2$ to give the result shown in Fig. 1(*b*). Here, the two highest peaks correspond to the expected vector peaks from the Se atoms at sites 133 and 153. The interpretation is straightforward and knowledge of these two sites can be used in the standard way to help locate the remaining Se atoms. Although the example uses isomorphous differences, the extension to use anomalous scattering or multiple-wavelength data is obvious. The approach can readily be generalized to mutants that include different numbers or combinations of methionine residues. Additionally, by using proteins that differ by a single methionine, the location of that amino acid in the unit call can be determined.

Leahy *et al.* (1994) have previously shown how methionines can be engineered into a protein in order to assist in the determination of its structure. They also discussed a possible strategy for the selection of suitable sites of substitution. In selecting sites for either addition or subtraction of methionines, one wants to choose replacements that will not interfere with the growth of crystals, will leave the crystal structure as isomorphous as possible with the parent and will result in well ordered (*i.e.* non-mobile) sites. The obvious residues that suggest themselves for replacement with methionine are leucine, isoleucine, phenylalanine and possibly valine. The first three have side-chain volumes of 76, 76 and 87 Å^3, respectively, which are comparable with that of methionine (76 Å^3). The value for valine

(57 Å^3) is somewhat lower (Creighton, 1993). Additionally, all such residues are likely to occupy internal sites within a protein (Rose *et al.*, 1985), minimizing the chance of interfering with crystal contacts and maximizing the likelihood that the methionine introduced will be well ordered.

In order to give possible guidance in the choice between the above alternatives, we have collected together in Table 1 the relevant statistics for all T4 lysozyme mutants in which methionines have been substituted at different sites in the protein. In all 19 examples, the side chain of the residue is at least 90% buried.

The difference in structure amplitudes between wild-type lysozyme and a given methionine mutant, *i.e.* the isomorphous difference, will arise from three factors: (i) the difference in scattering between the methionine and the amino acid it replaces; (ii) non-isomorphism introduced by the substitution and (iii) errors in the data. Because each of the side chains involved scatters fairly weakly and, in any case, should be roughly comparable, at least for methionine, leucine, isoleucine and valine, we assume that the first term can be ignored. The errors in the data can be estimated from $R_{\rm merge}$. Taking the average value of $R_{\rm merge}$ (5.5%) and bearing in mind that it is calculated from intensities rather than structure amplitudes, the average contribution to the isomorphous difference from errors in the data can be estimated as about 0.055. Actually this is, if anything, an overestimate, since $R_{\rm merge}$ is calculated to the limit of the data, whereas the isomorphous difference is calculated at lower resolution (Table 1). This suggests that a mutant such as Leu99Met, for which the isomorphous difference is low (0.081 at 2.5 Å resolution), is highly isomorphous with wild type. Indeed, the Leu99Met structure has been shown to be extremely similar to wild type, with shifts in the backbone not exceeding about 0.2 Å (Eriksson *et al.*, 1993). Conversely, mutants for which the isomorphous difference is large (*e.g.* 0.217 for Val111Met) would be expected to be rather non-isomorphous. This is also the case: the Val111→Met mutation causes some backbone shifts of up to 2 Å (Gassner, 1998).

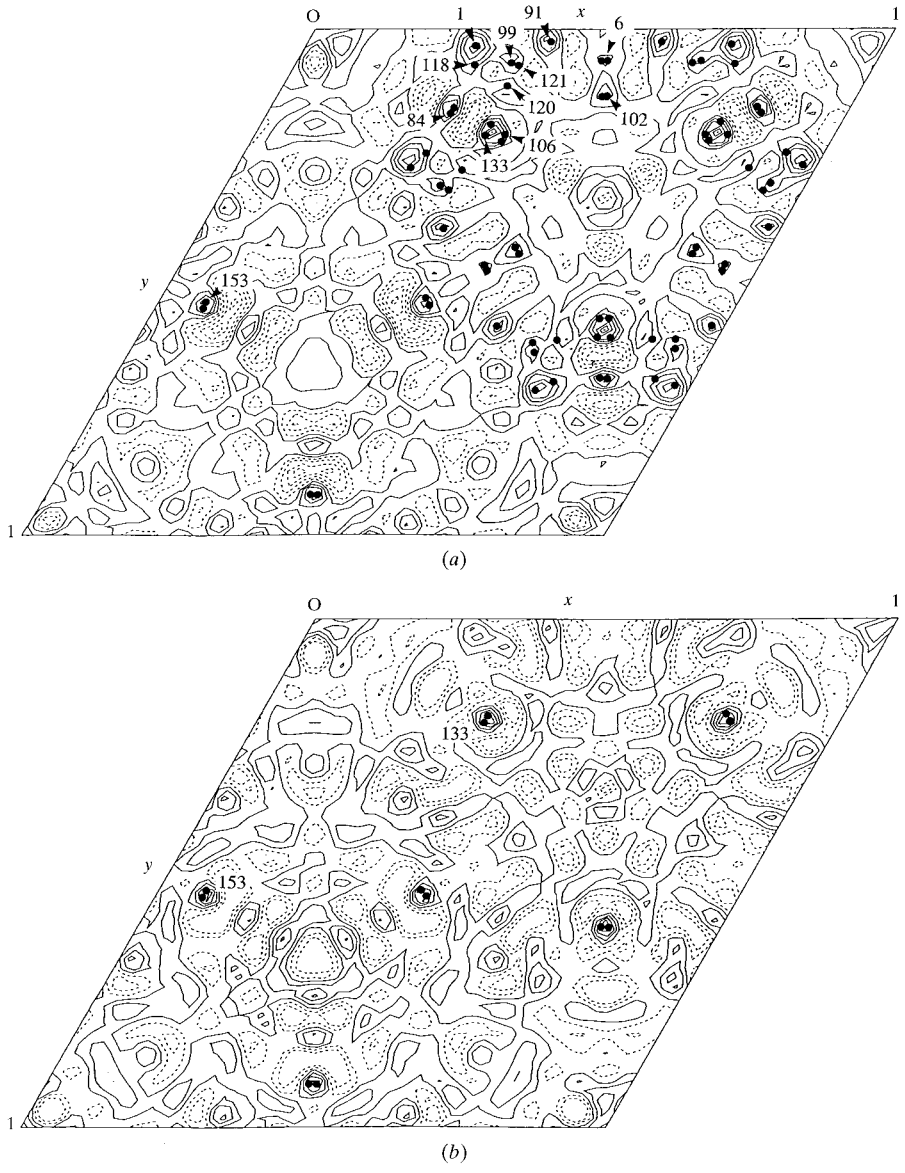

Figure 1
(*a*) Section z = 1/3 of a Patterson map calculated with coefficients $(F_{\text{12-SeMet}} - F_{\text{12-Met}})^2$, where $F_{\text{12-SeMet}}$ is the observed structure amplitude for lysozyme crystals containing 12 selenomethionines and $F_{\text{12-Met}}$ is the observed structure amplitude for the related crystals containing 12 methionines (Table 1). The resolution is 3.5 Å and the map is contoured at $\pm 0.6\sigma$, where σ is the root-mean-square density throughout the unit cell. The space group of T4 lysozyme is $P3_221$ (Weaver & Matthews, 1987) and the positions of the expected selenium–selenium vectors, indicated by solid circles, were calculated from the refined structure of the 12-selenomethionine protein. For simplicity, only one set of unique vector peaks is identified, the numbers corresponding to the Se atoms at sites 1, 6, 84, 91, 99, 102, 106, 118, 120, 121, 133 and 153. (*b*) Section z = 1/3 of the Patterson function calculated with amplitudes $(F_{\text{12-SeMet}} - F_{\text{10-SeMet}})^2$, where $F_{\text{12-SeMet}}$ and $F_{\text{10-SeMet}}$ are the observed structure amplitudes for T4 lysozymes containing 12 selenomethionines and ten selenomethionines, respectively (Table 2). The proteins are identical except that the 12-selenomethionine variant has selenomethionine at sites 133 and 153, whereas the 10-selenomethionine protein has leucine and phenylalanine, respectively, at these sites. The Harker peaks that correspond to the Se atoms at these two sites are labeled (*cf.* Fig. 1*a*). The resolution of the map is 3.5 Å and it is contoured at $\pm 0.6\sigma$, where σ is the root-mean-square density throughout the unit cell.

As suggested by Table 2, the second-best substitution is Phe→Met rather than Ile→Met, although it should be emphasized that this is based on a small number of examples. Val→Met appears to be least favorable, reflecting the fact that the larger volume and different shape of a methionine side chain do not usually allow it to replace that of valine without some disruption of the structure. This ranking does not, however,

Table 1
X-ray data-collection statistics for different methionine substitutions in T4 lysozyme.

The mutants were purified and crystallized using methods described previously (Eriksson *et al.*, 1993; Gassner *et al.*, 1996, 1997; Gassner, Baase, Lindstrom *et al.*, 1999). Except for mutant Ile78Met, for which the data were measured with an R-AXIS IV detector, all X-ray data were collected using a San Diego Multiwire Area Detector System (Hamlin, 1985). $R_{\rm merge} = 100\sum |I_{hkl} - \bar{I}_{hkl}|/\sum \bar{I}_{hkl}$, where the I_{hkl} are independent measurements of the intensity of a given reflection and $\bar{I}_{hkl}$ is the average. $R_{\rm merge}$ is calculated to the limit of the measured data. The isomorphous difference, $R_{\rm iso}$, is calculated to two representative resolutions, 2.5 and 3.5 Å, and is defined as $R_{\rm iso} = \sum |F_{\rm Mut} - F_{\rm WT*}|/\sum F_{\rm WT*}$, where $F_{\rm Mut}$ is the structure amplitude for the mutant crystal and $F_{\rm WT*}$ is that for pseudo wild-type lysozyme, which is the reference protein from which all the mutants were constructed. WT* lysozyme crystallizes in space group $P3_221$ with unit-cell parameters $a = b = 60.9$, $c = 96.9$ Å. Data for mutants Leu99Met and Phe153Met are from Eriksson *et al.* (1993) and for mutant Ala129Met from Baldwin *et al.* (1996). Protein Data Bank access codes are given.

	Unit-cell parameters					20–2.5 Å		20–3.5 Å	
Mutant	a, b (Å)	c (Å)	$R_{\rm merge}$ (%)	Limiting resolution (Å)	PDB code	Completeness of data (%)	Isomorphous difference	Completeness of data (%)	Isomorphous difference
Leucine to methionine substitutions									
Leu66Met	61.1	97.3	7.9	1.97	1d3j	85	0.133	85	0.121
Leu84Met	61.1	96.9	4.1	1.85	1cu2	84	0.154	84	0.140
Leu91Met	61.0	97.1	6.7	2.05	1cu5	83	0.120	84	0.099
Leu99Met	61.2	96.9	5.2	2.00	1l93	83	0.081	84	0.074
Leu118Met	61.1	97.1	4.1	1.80	1cv4	82	0.118	84	0.106
Leu121Met	61.0	96.9	3.6	1.80	1cv3	83	0.116	82	0.104
Leu133Met	61.1	97.0	3.4	1.87	1cv5	84	0.134	84	0.120
Average isomorphous difference							0.122		0.109
Isoleucine to methionine substitutions									
Ile27Met	60.9	97.4	8.8	2.08	1d2w	81	0.164	81	0.144
Ile50Met	61.1	97.4	7.8	2.06	1d2y	82	0.146	81	0.133
Ile58Met	61.0	97.1	6.6	2.05	1d3f	85	0.127	85	0.114
Ile78Met	60.5	97.7	7.2	2.20	1cuo	84	0.196	84	0.164
Ile100Met	60.9	97.1	4.2	1.89	1cup	81	0.106	81	0.094
Average isomorphous difference							0.148		0.130
Phenylalanine to methionine substitutions									
Phe153Met	61.0	95.9	4.6	2.10	1l88	82	0.156	83	0.143
Phe104Met	60.9	97.2	4.9	2.12	1cvo	82	0.112	81	0.094
Average isomorphous difference							0.134		0.119
Valine to methionine substitutions									
Val87Met	61.0	97.2	5.9	2.12	1cu3	79	0.162	80	0.149
Val103Met	61.0	97.4	5.9	2.05	1cuq	81	0.181	81	0.169
Val111Met	61.0	97.4	5.5	1.90	1cv1	85	0.217	85	0.201
Val149Met	60.9	97.3	6.1	1.90	1cv6	82	0.115	81	0.099
Average isomorphous difference							0.169		0.155
Alanine to methionine substitution									
Ala129Met	61.2	96.0	4.6	2.30	196l	80	0.198	83	0.187

Table 2
X-ray data-collection statistics.

X-ray data for the 12-Met lysozyme were measured with a San Diego Area Detector System as described previously (Gassner *et al.*, 1996). Data for the other variants were collected in the same way. The selenomethionine-containing variants were expressed in RR1 (Muchmore *et al.*, 1989), a strain of bacteria not auxotrophic for methionine, using an adaptation of the procedure of Van Duyne *et al.* (1993) (Gassner, 1998; Gassner, Baase, Hausrath *et al.*, 1999). Protein Data Bank access codes are included.

	Unit-cell parameters					20–2.5 Å		20–3.5 Å	
Mutant	a, b (Å)	c (Å)	$R_{\rm merge}$ (%)	Limiting resolution (Å)	PDB code	Completeness of data (%)	Isomorphous difference	Completeness of data (%)	Isomorphous difference
WT*	60.9	96.9			1l63				
12-Met	61.4	96.4	4.4	1.90	1cx7	92	0.24	92	0.22
12-SeMet	61.4	96.7	7.1	2.01	1cx6	95	0.31	93	0.29
10-Met	61.2	96.6	5.8	2.12	1d3m	94	0.18	93	0.16
10-SeMet	61.3	96.6	6.0	2.00	1d3n	92	0.22	92	0.21
12-SeMet *versus*							0.16		0.16
12-Met						95		93	
12-SeMet *versus*							0.19	93	
10-SeMet						95			0.18

agree with that based on the frequency of amino-acid substitutions in related proteins. Rather than the overall ranking of Leu, Phe, Ile and Val seen here, sequence comparison suggests Leu, Ile, Val, Thr and Phe (Jones *et al.*, 1992; Leahy *et al.*, 1994)

If it is desired to remove methionines from a protein without disruption of the structure, amino-acid sequence comparison suggests that the optimal replacements, in order of choice, are Ile, Leu, Val, Thr and Phe (Jones *et al.*, 1992; Leahy *et al.*, 1994). We do not have direct experimental data to test this, but have collected together in Table 3 the data that are available, namely for the replacement of the methionines in T4 lysozyme with leucine and with alanine. The former mutants were constructed as a possible way to stabilize the protein (Hurley *et al.*, 1992; Lipscomb *et al.*, 1998), while the latter were generated primarily to investigate the hydrophobic stabilization of proteins (Blaber *et al.*, 1995; Baldwin *et al.*, 1998; Xu *et al.*, 1998). It will be noticed that, as a group, the replacements of methionine with leucine have larger isomorphous differences than do the replacements with alanine. This is because, in the case of Met6→Leu for example, there are large structural perturbations [and, in addition, the protein is substantially destabilized (by 2.8 kcal mol^{-1} or 11.7 kJ mol^{-1})]. The replacement of methionine with alanine occurs with relatively low frequency in the Dayhoff mutation probability matrix presumably because such substitutions, if at internal sites, are likely to reduce protein stability owing to decreased hydrophobic stabilization and possible cavity formation. In terms of maintaining isomorphism, however, Table 3 suggests that there may be an advantage in simply truncating a methionine to alanine, rather than replacing it with an amino acid of different shape and incurring the risk of structural non-isomorphism induced by steric clash.

We thank Hong Xiao and Leslie Gay for excellent help in purifying and crystallizing the mutant lysozymes, Blaine Mooers and Wendy Breyer for technical assistance, and Eric Anderson, Richard Kingston, Ingo Korndoerfer, Dale Tronrud, Michael Quillin and Larry Weaver for helpful discussions. This work was supported in part by NIH grant GM21967 to BWM.

Table 3
X-ray data-collection statistics for the replacement of methionines in T4 lysozyme with leucine or alanine.

Statistics quoted in the table are as defined in Table 1. The percentage of the side chain accessible to solvent is based on the WT* structure and was calculated as described by Alber *et al.* (1987). Mutants Met6Leu, Met106Leu and Met120Leu are from Lipscomb *et al.* (1998), Met102Leu is from Hurley *et al.* (1992), Met102Ala is from Baldwin *et al.* (1998), Met6Ala and Met106Ala are from Xu *et al.* (1998) and Met102Ala is from Blaber *et al.* (1995).

	Unit-cell parameters					20–2.5 Å		20–3.5 Å	
Mutant	*a*, *b* (Å)	*c* (Å)	R_{merge} (%)	Limiting resolution (Å)	Side-chain solvent accessibility (%)	Completeness of data (%)	Isomorphous difference	Completeness of data (%)	Isomorphous difference
Methionine to leucine substitutions									
Met6Leu	60.87	97.20	7.7	1.90		88	0.208	88	0.182
Met102Leu	61.00	96.30	6.8	2.05	3	81	0.142	71	0.113
Met106Leu	60.89	97.17	7.5	1.90	24	91	0.123	91	0.102
Met120Leu	60.93	97.30	6.5	1.90	21	92	0.120	92	0.088
Average isomorphous difference							0.148		0.121
Methionine to alanine substitutions									
Met6Ala	60.90	96.60	3.0	1.80		93	0.155	92	0.131
Met102Ala	60.90	97.20	3.3	1.85	3	74	0.130	72	0.111
Met106Ala	60.90	96.50	4.7	1.90	24	78	0.096	75	0.086
Met120Ala	60.99	96.57	3.2	1.80	21	87	0.128	86	0.111
Average isomorphous difference							0.127		0.110

References

Alber, T., Dao-pin, S., Nye, J. A., Muchmore, D. C. & Matthews, B. W. (1987). *Biochemistry*, **26**, 3754–3758.

Baldwin, E., Baase, W. A., Zhang, X.-J., Feher, V. & Matthews, B. W. (1998). *J. Mol. Biol.* **277**, 467–485.

Baldwin, E., Xu, J., Hajiseyedjavadi, O., Baase, W. A. & Matthews, B. W. (1996). *J. Mol. Biol.* **259**, 542–559.

Blaber, M., Baase, W. A., Gassner, N. & Matthews, B. W. (1995). *J. Mol. Biol.* **246**, 317–330.

Creighton, T. E. (1993). *Proteins, Structures and Molecular Properties*, 2nd ed. New York: W. H. Freeman and Co.

Deacon, A. M. & Ealick, S. E. (1999). *Structure*, **7**, R161–R166.

Eriksson, A. E., Baase, W. A. & Matthews, B. W. (1993). *J. Mol. Biol.* **229**, 747–769.

Gassner, N. C. (1998). PhD thesis, University of Oregon, USA.

Gassner, N. C., Baase, W. A., Hausrath, A. C. & Matthews, B. W. (1999). In the press.

Gassner, N. C., Baase, W. A., Lindstrom, J., Lu, J., Dahlquist, F. W. & Matthews, B. W. (1999). In the press.

Gassner, N. C., Baase, W. A., Lindstrom, J. D., Shoichet, B. K. & Matthews, B. W. (1997). *Techniques in Protein Chemistry VIII*, edited by D. Marshak, pp. 851–863. New York: Academic Press.

Gassner, N. C., Baase, W. A. & Matthews, B. W. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 12155–12158.

Hamlin, R. (1985). *Methods Enzymol.* **114**, 416–452.

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

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

Hurley, J. H., Baase, W. A. & Matthews, B. W. (1992). *J. Mol. Biol.* **224**, 1143–1159.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). *Comput. Appl. Biosci.* **8**, 275–282.

Leahy, D. J., Erickson, H. P., Aukhil, I., Joshi, P. & Hendrickson, W. A. (1994). *Proteins*, **19**, 48–54.

Lipscomb, L. A., Gassner, N. C., Snow, S. D., Eldridge, A. M., Baase, W. A., Drew, D. L. & Matthews, B. W. (1998). *Protein Sci.* **7**, 765–773.

Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. (1989). *Methods Enzymol.* **177**, 44–73.

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985). *Science*, **229**, 834–838.

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

Weaver, L. H. & Matthews, B. W. (1987). *J. Mol. Biol.* **193**, 189–199.

Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.

Xu, J., Baase, W. A., Baldwin, E. & Matthews, B. W. (1998). *Protein Sci.* **7**, 158–177.