The EMBO Journal vol.9 no.5 pp.1665 - 1672, 1990

Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure

Wayne A.Hendrickson[1,2], John R.Horton[2] and David M.LeMaster[3,4]

[1]Howard Hughes Medical Institute, [2]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, and [3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

[4]Present address: Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, IL 60208, USA

Communicated by G.N.Cohen

An expression system has been established for the incorporation of selenomethionine into recombinant proteins produced from plasmids in *Escherichia coli*. Replacement of methionine by selenomethionine is demonstrated at the level of 100% for both T4 and *E.coli* thioredoxins. The natural recombinant proteins and the selenomethionyl variants of both thioredoxins crystallize isomorphously. Anomalous scattering factors were deduced from synchrotron X-ray absorption measurements of crystals of the selenomethionyl proteins. Taken with reference to experience in the structural analysis of selenobiotinyl streptavidin by the method of multiwavelength anomalous diffraction (MAD), these data indicate that recombinant selenomethionyl proteins analyzed by MAD phasing offer a rather general means for the elucidation of atomic structures.

Key words: crystallography/protein structure/selenium/selenomethionine/thioredoxin

Introduction

The decisive molecular images obtained from crystal structures often provide powerful insight into biological activity. These images have traditionally been hard won, however, as macromolecular crystallography is generally arduous and time consuming. Recent advances in instrumentation and methodology have accelerated the pace substantially; but, nevertheless, two steps in the process remain especially problematic: crystallization and the phase problem. There is still more art than science in growing suitable crystals, although prospects are improved by the availability of clean protein stocks in abundance from recombinant expression systems. Molecular biology techniques can also aid in evaluating phase angles for the diffracted X-ray waves, and we report here on a new approach for introducing suitable diffraction labels into proteins.

One must know both the amplitude and phase for each of thousands of diffracted waves from a macromolecular crystal in order to reconstruct an image, but only amplitudes can be recovered from standard diffraction measurements. This presents the central conceptual obstacle in crystallography—the phase problem. Phase evaluation for new macromolecular structures traditionally has been based on the analysis of isomorphous replacements with heavy atoms, although other sources of phasing such as molecular averaging or anomalous scattering are often used in supplement. This, the method of multiple isomorphous replacement (MIR), entails the trial-and-error preparation of heavy-atom derivatives and it is frequently limited by lack of isomorphism.

Recently, an alternative approach has been devised for *de novo* phasing of macromolecular crystal structures. This is the method of multiwavelength anomalous diffraction (MAD) (Hendrickson, 1985) which exploits the scattering effects of resonance between X-rays and bound atomic orbitals as indicated in Figure 1. Qualitatively, MAD experiments can be thought of as *in situ* isomorphous replacements generated by the variation in scattering strength that accompanies change of wavelength. The MAD method does require the special properties of synchrotron radiation, but it has advantages in that isomorphism is perfect, all data can be measured from a single crystal, and the analysis is algebraically exact (Karle, 1980). Several recent applications demonstrate the effectiveness of MAD phasing (Guss *et al.*, 1988; Hendrickson *et al.*, 1988; Krishna Murthy *et al.*, 1988; Hendrickson *et al.*, 1989).

The MAD method requires the presence of a few heavier elements as resonance centers. The anomalous scattering from the lighter atoms (H, C, N and O) in proteins is inappreciable at typically achievable wavelengths (0.5–3.0 Å), and although sulfur anomalous scattering has been used for protein phasing in special cases (Hendrickson and Teeter, 1981) its resonance energy is ordinarily inaccessible for MAD experiments. Either intrinsic metal centers (e.g. Fe or Zn) or introduced heavy atoms (e.g. Hg, Au or Gd) are suitable and they can produce dramatic effects. However, metalloproteins or chemical derivatives are not always at hand. Systematic biological incorporation of selenomethionine in place of methionine residues in proteins offers the possibility of generality. Cowie and Cohen (1957) showed that a strain of *E.coli* made auxotrophic for methionine could grow for 100 generations in selenomethionine, suggesting that all proteins needed to sustain life in this organism can function well when fully substituted with selenomethionine. Our experiments in the MAD phasing of selenolanthionine (Hendrickson, 1985) and of selenobiotinyl streptavidin (Hendrickson *et al.*, 1989) demonstrate that selenium is very effective as a center for MAD phasing.

In this study, we have developed a system for expressing recombinant selenomethionyl proteins in *E.coli*. Selenomethionyl thioredoxins produced from this system have been crystallized and characterized. A theoretical evaluation based on anomalous scattering factors measured from selenomethionyl *E.coli* thioredoxin indicates that MAD analysis of selenomethionyl proteins should provide a broadly applicable solution to the phase problem.

© Oxford University Press

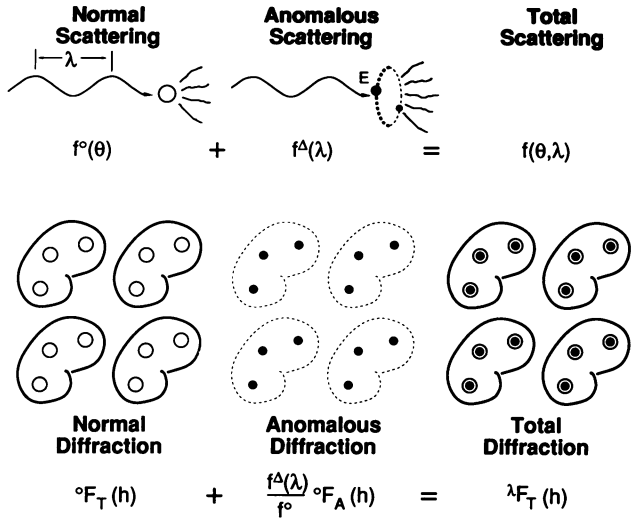

Fig. 1. Schematic description of anomalous scattering from atoms in macromolecular crystal structures. The upper portion illustrates the normal and anomalous components of atomic scattering. Normal scattering of X-rays by matter arises from radiation emitted by electrons that have been excited into free vibration by an incident X-ray wave. The amplitude of normal atomic scattering, $f^\circ(\theta)$ depends on scattering angle, θ, but it is independent of wavelength, λ. However, when the energy, E, associated with X-rays of wavelength λ approaches an energy level of electrons bound in atomic orbitals, the induced vibrations resonate with the intrinsic atomic oscillations. This affects the amplitude and the phase of scattering. The incremental anomalous scattering factor, $f^\Delta(\lambda)$, depends very strongly on wavelength but is essentially independent of scattering angle. The lower portion illustrates the components of diffraction from a crystalline array of molecules. Each molecule in a protein crystal might contain a few thousand atoms (such as C, N and O) that only scatter X-rays normally and few additional atoms (such as Se) that produce anomalous (resonant) scattering. The normal component of anomalous centers is symbolized by open circles and the anomalous component is represented by solid circles. The preponderant light atoms are not represented individually here. The total diffraction from such a crystal is described by the structure factor, $^\lambda F_T(h)$, for each reflection h as measured at a particular wavelength λ. The normal scattering component of this diffraction, which has components from all atoms in the structure, is given by $^\circ F_T(h)$. Knowledge of this wavelength independent component of diffraction, $^\circ F_T = |^\circ F_T| \exp(i^\circ\phi_T)$, including its phase, suffices by Fourier transformation to produce an image of the crystal structure. Only the heavier atoms, usually few in number, contribute to anomalous diffraction, and this component contains all wavelength dependence. In the case of a single kind of anomalous scatterer (as in selenomethionyl proteins) this component is related to the normal scattering contributions from these atoms, $^\circ F_A(h)$, by the ratio of anomalous to normal scattering factors, $f^\Delta(\lambda)/f^\circ(\theta)$.

Through an algebraic analysis of total diffraction data measured accurately at a few wavelengths (usually 3–5), it is possible to extract the essential components $|^\circ F_T|$, $|^\circ F_A|$ and $\Delta\phi = {}^\circ\phi_T - {}^\circ\phi_A$ from the set of $|^\lambda F_T|$ measurements (Karle, 1980; Hendrickson, 1985). This is done for each of the several thousand reflections. Then the $|^\circ F_A|$ values are used to deduce the structure of anomalous scattering centers, and from this structure $^\circ\phi_A$ values are computed. These together with the fitted $\Delta\phi$ values serve to specify $^\circ\phi_T$ and thereby solve the phase problem. The distributions of electron density are synthesized by Fourier transformation of $|^\circ F_T| \exp(i\ ^\circ\phi_T)$.

Results

Selenomethionine tolerant met⁻ strain

Different met⁻ strains of *E.coli* tolerate the substitution of selenomethionine in varying degrees. After testing the growth characteristics of a few methionine auxotrophs in selenomethionine medium, one of us (D.M.LeMaster) constructed a new auxotrophic strain, DL41, by introducing a lesion in the *metA* gene. This strain grows as vigorously as its parent MG1655 (Guyer *et al.*, 1981), a viable F^-, λ^- strain derived from *E.coli* K12. It can double in 80–85 min at 33°C on a defined glucose medium (LeMaster and Richards, 1985) supplemented by selenomethionine in place of methionine (as compared with 50–55 min in the methionine medium), and it will grow to an OD_{600} of >4.0 in a shaker bath. Cell growth in selenomethionine is somewhat variable, however. For example, the doubling time in shaker flasks at 31°C was found to be 160 min.

Recombinant selenomethionyl thioredoxin expression

Our initial tests in the production and crystallographic analysis of selenomethionyl proteins have been carried out on thioredoxins since a suitable expression system was already available. This system simultaneously produces both *E.coli* and bacteriophage T4 thioredoxins (LeMaster and Richards, 1988) and these molecules have concentrations of methionine suitable for MAD phasing: one residue in 108 for the *E.coli* protein (Holmgren, 1968) and three in 87 for the T4 protein (Sjöberg and Holmgren, 1972). The expression plasmid for this system, pDL59, is under control of the P_L promoter from bacteriophage λ, and it was constructed to be genetically portable for use in isotopic labeling of specific amino acids for NMR studies. Thus, the temperature-sensitive C_{I857} λ repressor gene and the antiterminator N gene were cloned into the plasmid along with the *trxA* gene for *E.coli* thioredoxin (Lunn *et al.*, 1984) and the *nrdC* gene for T4 phage thioredoxin (LeMaster, 1986). The pDL59 plasmid was introduced into the DL41 bacterial strain and this system was used to produce both selenomethionyl thioredoxins as well as the natural recombinant proteins. Starter cultures for selenomethionyl protein production were grown in methionine containing medium since stationary phase bacteria do not tolerate selenomethionine very well. The same purification procedures as described for the natural proteins (LeMaster and Richards, 1985) were used for purifying the selenomethionyl proteins. A yield of 65 mg of *E.coli* selenomethionyl thioredoxin and 20 mg of T4 selenomethionyl thioredoxin was obtained from 2.5 l of fermenter culture.

The level of selenomethionine substitution in the purified proteins was estimated by amino-acid composition analysis. Selenomethionine decomposes under the acid hydrolysis conditions used in the analysis (Shepherd and Huber, 1969), and thus it is the absence of methionine that is monitored. Comparable portions of the chromatograms for natural and selenomethionyl T4 thioredoxins are shown in Figure 2. Analyses made with normal loading (~10 μg) gave a yield of 2.5 and 1.0 residues of methionine per molecule from the natural T4 and *E.coli* proteins, respectively, whereas reproducibly there was no detectable methionine in the selenomethionyl proteins. When analyzed under overloaded conditions (30–40 μg), the T4 and *E.coli* selenomethionyl thioredoxins gave 0.04 and 0.02 residues of methionine per molecule (i.e. 1.3% and 2.0% of the methionine sites). These levels are well within the root mean square (r.m.s.) deviation of 0.09 residues between duplicates and a 0.61 residue r.m.s. discrepancy between analytical and theoretical compositions. Thus, within the limits of the sensitivity of our analysis the replacement of methionine by selenomethionine is complete in these experiments. Positive evidence for the selenomethionine substitution is provided by X-ray absorption spectra described below.

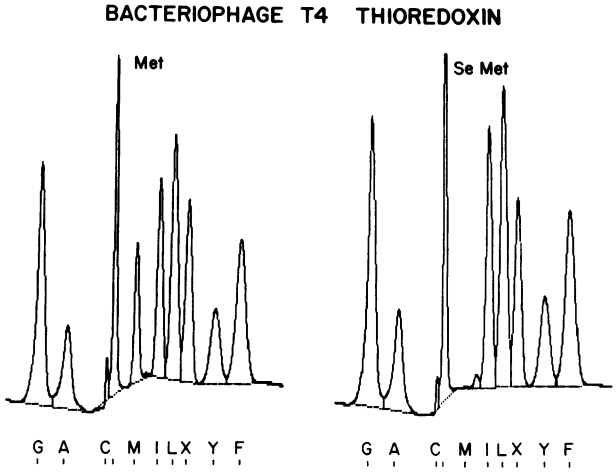

Fig. 2. Portions of the amino acid composition chromatograms for natural and selenomethionyl T4 thioredoxin. The respective elution profiles of absorbances of ninhydrin derivatives of the amino acids are shown with positions marked by the one-letter code for amino acids. Valine is in the unmarked peak following cysteine and X indicates the location of the norleucine standard.

Crystallization of selenomethionyl thioredoxins

Crystal structure analyses have been reported previously both for *E.coli* thioredoxin (Holmgren *et al.*, 1975) and for T4 thioredoxin (Söderberg *et al.*, 1978), and the structure of the *E.coli* protein has recently been refined at high resolution (Katti *et al.*, 1990). In this study we have attempted to crystallize the selenomethionyl proteins under conditions similar to those used with the natural methionyl thioredoxins. Pertinent characteristics of resulting crystals are summarized in Table I. T4 thioredoxin was crystallized at 4°C from ethanol in the presence of cadmium, but by vapor diffusion in hanging drops rather than in microdiffusion capillaries as before (Sjöberg and Söderberg, 1976). Nevertheless, the natural and selenomethionyl crystals are nearly isomorphous. Crystals of *E.coli* thioredoxin were grown in the presence of copper, but using polyethylene glycol (PEG) rather than 2-methyl-2,4-pentanediol (MPD) as the precipitating agent and at 20°C rather than 4°C. The PEG crystals differ markedly from the MPD crystals although they are clearly related. The natural recombinant protein and its selenomethionyl counterpart crystallize isomorphously in vapor diffusion against the same PEG concentration. This is evident from crystal morphology and X-ray diffraction patterns (Figure 3). The most noticeable difference in the crystallization of the two *E.coli* proteins is that crystals appear sooner in the selenomethionine case. This may reflect the somewhat reduced solubility of the variant protein—under the initial hanging-drop conditions (4% PEG 3350 and 1 mM Cu^{2+} at pH 4.2), a slight precipitate can be observed with selenomethionyl *E.coli* thioredoxin at 3.9 mg/ml, whereas under the same conditions but with the natural *E.coli* thioredoxin at 6.3 mg/ml the initial solution is clear.

Anomalous scattering factors

Since the efficiency of the MAD method derives from the variation of scattering strength with X-ray wavelength, it is essential that the anomalous scattering factors be known. These can be obtained from X-ray absorption spectra measured near the resonant orbital energy with synchrotron radiation and analyzed in relation to theoretical values (Cromer, 1983) that are computed for the elemental state. X-ray absorption spectra were measured from crystals of both *E.coli* and T4 thioredoxin, and these data were analyzed as before (Hendrickson *et al.*, 1988) to produce scattering factors. The resulting spectra are similar to those from other organic selenides including DL selenomethionine (unpublished results), selenolanthionine (Templeton and Templeton, 1988) and selenobiotinyl streptavidin (Hendrickson *et al.*, 1989). In particular, the selenium anomalous scattering is anisotropic which demonstrates that the selenomethionine is incorporated into thioredoxin in an ordered manner.

Anomalous scattering factors measured from one orientation of a selenomethionyl *E.coli* thioredoxin crystal are shown in Figure 4. These data were obtained with the high energy resolution of a Si111 double crystal monochromator and they exhibit a pronounced 'white line' of edge absorption. The strength of these sharp edge features enhances the phasing potential of the MAD method. Table II gives the locations and values of extrema in three nearly orthogonal orientations with respect to the polarized X-ray beam. A crystal of selenomethionyl T4 thioredoxin examined with radiation from a Ge111 monochromator gave a spectrum similar to that shown in Figure 4, but with somewhat lower resolution.

Theoretical feasibility for MAD phasing of selenomethionyl proteins

The ultimate test of the idea of direct determination of protein crystal structures by MAD phasing from selenomethionine is the actual analysis. We have experiments in progress on *E.coli* thioredoxin and other proteins. However, it is already clear by analogy with selenobiotinyl streptavidin that the MAD phasing of selenomethionyl proteins will be feasible provided that the methionine residues are ordered and appropriately abundant. The average frequency of occurrence of methionine residues in proteins is 1 in 59 residues (Dayhoff, 1978) and this compares favorably with the single selenium site per 126 residues in selenobiotinyl streptavidin.

Our analysis of anomalous diffraction data involves a least-squares optimization (Hendrickson, 1985) based on an exact algebraic analysis (Karle, 1980). An analysis of the prospective accuracy in a particular application of this procedure is difficult; however, the information content resides in diffracted intensity differences for which expected values can be estimated. Two such differences are pertinent and give mutually orthogonal phase information. These are the Bijvoet difference, $\Delta F_{\pm \mathbf{h}} = |^{\lambda}F(\mathbf{h})| - |^{\lambda}F(-\mathbf{h})|$, and the dispersive difference, $\Delta F_{\Delta\lambda} = |^{\lambda i}F| - |^{\lambda j}F|$ where the structure factors, F, are as defined in the legend for Figure 1. Given the number of protein atoms, the number of anomalous centers, and the anomalous scattering factors it is possible to determine expected values for the respective diffraction ratios and thereby to estimate the relative phasing strength of a candidate for MAD phasing (Hendrickson *et al.*, 1985). Such estimates are presented in Table III for a number of pertinent actual and hypothetical situations. It is obvious that MAD phasing of prospective selenomethionyl proteins should be readily feasible since diffraction ratios are in excess of those that have proved adequate for related problems.

Table I. Characteristics of thioredoxin crystals

Molecule	Precipitant	Space group	Unit cell parameters				References
			a	b	c	β	
Natural *E.coli* thioredoxin	MPD, Cu^{2+}	C2	89.7 Å	51.1 Å	60.3 Å	113.5°	Holmgren *et al.* (1975)
Natural *E.coli* thioredoxin	PEG, Cu^{2+}	I2[a]	88.8	49.2	116.0	100.2	This work
Se met *E.coli* thioredoxin	PEG, Cu^{2+}	I2	89.2	49.6	115.6	100.2	This work
Natural T4 thioredoxin	Ethanol, Cd^{2+}	$P2_1$	54.1	45.9	40.8	99.4	Sjöberg and Söderberg (1976)
Se met T4 thioredoxin	Ethanol, Cd^{2+}	$P2_1$	53.9	45.8	39.0	100.4	This work

[a]The crystals are described here in the unconventional space group of I2 so that relatedness to the previously reported MPD crystals is evident. This unit cell is doubled in the c-direction but the corresponding additional reflections are relatively weak and are missing at low angles (d spacings >8 Å). The corresponding conventional cell in space group C2 has unit cell dimensions of a = 133.1 Å, b = 49.2 Å, c = 88.8 Å, β = 120.9°.

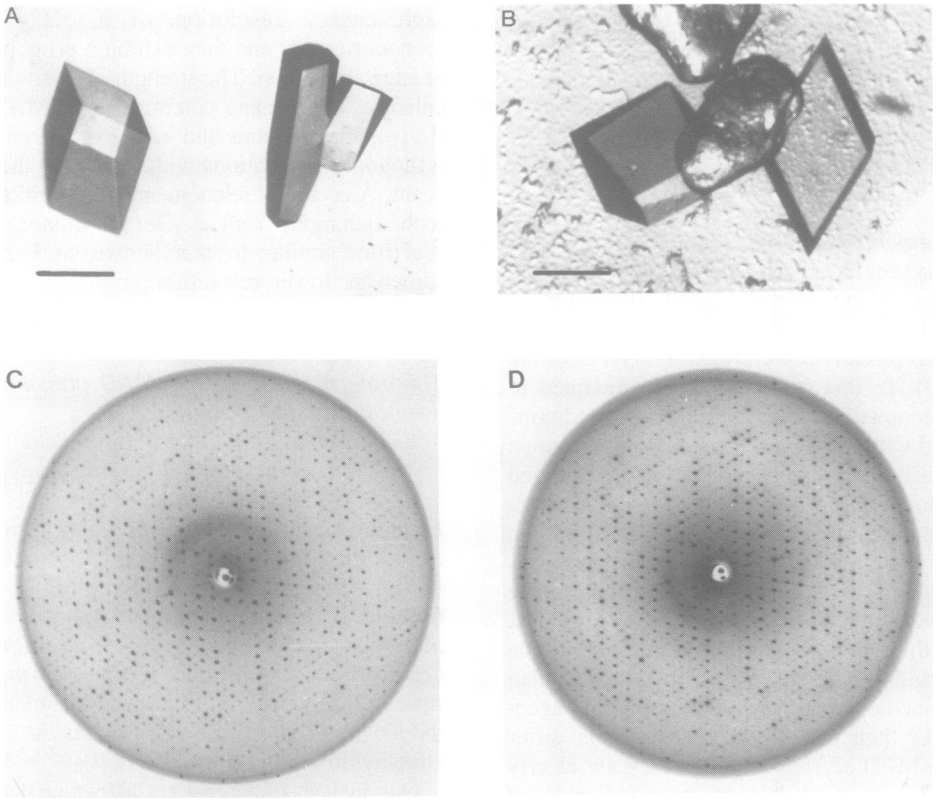

Fig. 3. Photographs of crystals and diffraction patterns of natural and selenomethionyl *E.coli* thioredoxin. (**A**) Type II crystals of the natural recombinant protein grown from PEG in the presence of Cu. The bar corresponds to 0.2 mm in length. (**B**) Type II crystals of the selenomethionyl protein. Protein used for this crystallization contained 10% methionine, but the crystals were essentially identical with those reported in Table I. The rounded object is a grain of sand used to induce nucleation. The bar corresponds to 0.2 mm. (**C**) A 15° precession photograph of the h01 zone of the diffraction pattern from a Type II crystal of natural thioredoxin. (**D**) A 15° precession photograph of the h01 pattern from a Type II crystal of selenomethionyl thioredoxin.

Discussion

Selenium chemistry and metabolism

The chemistry of selenium is similar to that of other elements in the oxygen series with which it shares the same outer electronic configuration. In this it has properties noticeably closer to sulfur than to its other neighbor tellurium. Indeed, selenium sufficiently resembles sulfur that it substitutes in a wide variety of organic compounds. Organic selenium chemistry is however distinctive (Odom, 1983). In particular, (i) single C−Se bonds are typically 0.14 Å longer than C−S bonds, (ii) organic selenium compounds are generally more reactive than their sulfur counterparts, owing presumably to lower bond strength, (iii) nucleophilic attack occurs more readily at selenium than at sulfur atoms, and (iv) selenium in comparison with sulfur is easier to oxidize from valence state (II) to (IV) but it is more difficult to oxidize from (IV) to (VI). Thus, selenides (R-Se-R′) and selenols (R-SeH) go more readily to selenoxides (R-SeO-R′) and seleninic acids (R-Se(O)OH) but less well on to selenones (R-SeO_2-R′) and selenonic acids (R-Se(O_2)OH) than happens for the sulfur analogs. In general, compounds with terminal selenium atoms (such as in selenols or selenoketones) degrade in air; whereas compounds with selenium bonded to two atoms (as in selenides such as selenomethionine) are relatively stable.

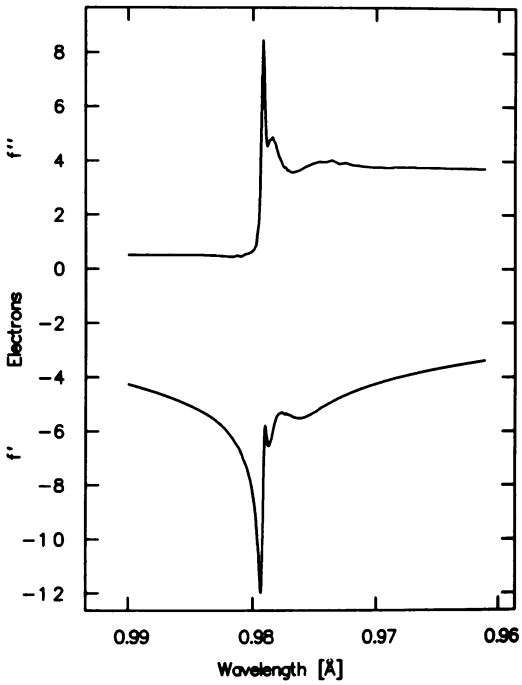

Fig. 4. Anomalous scattering factors near the absorption edge of selenium measured from a crystal of selenomethionyl *E.coli* thioredoxin. Here the electric vector **E** of the synchrotron X-ray beam was parallel with the b axis of the crystal.

Table II. Extrema of anomalous scattering from *E.coli* selenomethionyl thioredoxin

Feature	**E** || **a***	**E** || **b**	**E** || **c***	Average
Energy of f'_{min}	12 660 eV	12 659 eV	12 660 eV	12 659 eV
Minimal f'	−11.08	−11.83	−10.99	−11.20
Energy of f''_{max}	12 663 eV	12 662 eV	12 663 eV	12 662 eV
Maximal f''	7.82	8.18	7.93	7.88

The direction of X-ray polarization in the synchrotron beam is designated as **E** and direction in the crystal are along crystallographic axes in the real (**b**) and reciprocal lattice (**a*** and **c***). The correspondence between energy levels and X-ray wavelength is $\lambda = 12.398/E$ (keV). Thus, 12 659 eV corresponds to 0.9794 Å and 12 662 eV corresponds to 0.9792 Å.

Selenium is both toxic and an essential element for most animal and bacterial life. The requirement for selenium is by virtue of enzymes that exploit selenocysteine as a strong nucleophile and through the occurrence of selenium in certain bacterial tRNA species (Stadtman, 1987). Its toxicity results from the facile substitution of selenium in sulfur metabolism with untoward effects due to the exceptional reactivity of some products. Sulfur metabolism is quite different in bacteria than it is in higher organisms. In bacteria, sulfur compounds derive from inorganic sulfur, typically sulfate, which is first used to produce cysteine and then methionine and other sulfur metabolites. In mammals, methionine ingested in foodstuffs is the sulfur source and cysteine derives from methionine. Yeast and fungi have pathways operating in both directions.

Selenomethionine is remarkably benign as a substitute for methionine. Cowie and Cohen (1957) interpreted their experiments, based on the strict absence of alternative sources, as demonstrating that 100% of the methionine in *E.coli* could be replaced by selenomethionine. However, direct physical assays were not performed, and assays for selenomethionyl azurin produced by *Pseudomonas aeruginosa* showed incorporation to be at most 94% complete (Frank *et al.*, 1985). Here we show specifically that in the case of recombinant thioredoxins the methionine residues can indeed be fully replaced. A number of early *in vitro* studies make this a plausible and expected result. First, the aminoacylation of methionyl tRNA by the methionyl-tRNA synthetases from *Sarcina lutea* (Hahn and Brown, 1967), *E.coli* and rat liver (McConnell and Hoffman, 1972) is indistinguishable when selenomethionine replaces methionine as a substrate. Secondly, and most stringently, selenomethionine is as good or better as a substrate for activation by S-adenosylmethionine synthetases from yeast and rat liver (Mudd and Cantoni, 1957), and the selenium analog is as effective as a methyl donor as is S-adenosylmethionine itself in choline biosynthesis by rat liver microsomes (Bremer and Natori, 1960). Finally, the activity and properties of selenomethionyl and natural proteins are found to be very similar, if not identical, as was first shown for β-galactosidase (Huber and Criddle, 1967).

Although selenomethionine is a faithful mimic of methionine, the same cannot be said for selenium metabolism generally. The growth of *E.coli* with selenite replacing sulfite as a sulfur source arrests after depletion of sulfur reserves (Cowie and Cohen, 1957). Moreover, the β-galactosidase produced by *E.coli* grown on a selenate medium proved to have selenomethionine incorporated at only the 53% level (Huber and Criddle, 1967). Interestingly, this purified protein contained no selenocysteine despite the necessary existence of selenocysteine as an intermediate in selenomethionine synthesis. Probably, selenocysteine (or an oxidized product) is not recognized by the cysteinyl tRNA synthetase.

Attributes of selenomethionine as a general phase vehicle

Until now the nearest thing to a general phasing vehicle for protein crystallography has been in the use of cysteinyl residues as targets for the preparation of isomorphous mercurial derivatives. In contemplating new opportunities for generality with the introduction of MAD methodology, several qualities for an ideal phasing label were considered. These considerations include both biochemical properties such as convenience of labeling production, degree of isomorphism with the parent molecule, and durability of the derivative product, and also diffraction concerns such as the spectral position of the anomalous element, prospective phasing power, and the relative abundance and mobility of target sites. The properties of selenomethionyl proteins relate quite favorably to these considerations.

A most significant attribute to selenomethionine is that biological incorporation into proteins obviates the complexities of chemical labeling. Equally important, the resulting selenomethionyl proteins appear to be essentially isostructural with their natural methionyl counterparts. That this is so could be anticipated from the isomorphism between crystal structures of D,L methionine and D,L selenomethionine (Rajeswaran and Parthasarathy, 1984) and the remarkable tolerance of protein structure to site-directed point mutation. The isomorphism between natural and selenomethionyl crystal structures that is demonstrated here for T4 and *E.coli* thioredoxins has now been extended to interleukin 1α (IL-1α)

Table III. MAD phasing strength from anomalous diffraction ratios

Molecule	Residues	N_A	N_P	Maximal Bijvoet ratio	Maximal dispersive ratio	Source of scattering factors
E.coli thioredoxin	108	1 Se	830	5.8%	3.5%	This work
T4 thioredoxin	87	3	670	11.1%	6.7%	This work
Hypothetical[a]	300	1	2310	3.5%	2.1%	This work
Hypothetical[a]	300	5	2310	7.7%	4.7%	This work
Hypothetical[a]	300	10	2310	10.9%	6.6%	This work
Hypothetical[b]	300	5	2310	3.4%	3.7%	Hendrickson *et al.* (1989)
Streptavidin	252	2	1850	2.5%	2.3%	Hendrickson *et al.* (1989)
Crambin	46	6 S	350	1.5%		Hendrickson and Teeter (1981)

The Bijvoet diffraction ratio is computed from $<|\Delta F_{\Delta h}|> = q \cdot 2f''$ and the dispersive diffraction ratio is $<|\Delta F_{\pm\Delta\lambda}|> = q \cdot |f'(\lambda_i) - f'(\lambda_j)|$ where $q = \sqrt{N_A/2N_P}/Z_{eff}$ (Hendrickson *et al.*, 1985). Here N_P, the number of non-hydrogen atoms in the protein molecule, is calculated assuming 7.7 atoms per amino acid residue and Z_{eff}, the effective normal atomic scattering at zero scattering angle, is 6.7 electrons. N_A is the number of anomalous scatterers per protein molecule.

[a]These hypothetical molecules have methionine concentrations at 1/5, 1 and 2 times average, and diffraction ratios are calculated assuming measurements made at wavelengths corresponding to the extrema of the average spectrum of *E.coli* thioredoxin as given in Table I. A remote wavelength of 0.90 Å ($f' = -1.62$) is assumed for the dispersive difference.

[b]This hypothetical model assumes methionine concentration at the average and scattering factors with low energy resolution as in the streptavidin experiment.

(Graves *et al.*, 1990), to IL-2 (M.H.Hatada and B.J.Graves, personal communication) and to ribonuclease H (W.Yang, unpublished results). Clearly, the selenomethionyl proteins are sufficiently stable to permit the necessary diffraction measurements.

The scattering properties of selenium are also very favourable. The energy level of the selenium K-shell (1s) orbital corresponds to an X-ray wavelength (0.98 Å) that is readily accessible with synchrotron radiation. This also has advantages of limited radiation damage and sample absorption. The relative abundance of methionine residues in proteins is such that adequate diffraction signals can be expected in typical cases (Table III). Of course, an atom can only contribute coherently to the diffraction if it is well ordered, and thus atomic mobility is a concern. Selenomethionine is a good label in this regard. A survey of crystallographic B values (atomic mobility parameters) reported for 189 protein structures in the Protein Data Bank (Bernstein *et al.*, 1977) shows that on average the atomic mobility at the Sγ position of a methionine is just 4% above the average for all side-chain atoms in the molecule. For comparison, mobilities at the Cγ positions in leucine average 16% less, and those at Cγ in glutamine, arginine and lysine average 33%, 16% and 42% higher, respectively (J.R.Horton, unpublished results). A complication can arise in the diffraction analysis of large proteins where the number of selenium sites may be so large that this substructure presents a difficult problem. However, a simulation based on the structure of citrate synthase (Wiegand *et al.*, 1984), which with 30 had the most methionines per asymmetric unit among proteins in a particular release from the Protein Data Bank, shows that this selenium substructure could readily be solved by direct methods with the program MULTAN (Germain *et al.*, 1970) even when 20% random error was applied (J.R.Horton, unpublished results).

Possible complications in selenomethionyl protein expression

While met$^-$ strains of *E.coli* can clearly be grown on selenomethionine, it is also evident that care must be taken to achieve complete replacement. The major complication arises from the difficulty in bringing cells out of stationary phase in the presence of selenomethionine, whereas methionine brought forward with the inoculum for log phase growth may be preferentially incorporated. Thus, when we produced thioredoxins with excess methionine present in the starter culture, only 90% selenomethionine incorporation was obtained. Similar complications also limited the production of selenomethionyl IL-1α to an 86% level (Graves *et al.*, 1990). On the other hand, we have produced recombinant *E.coli* ribonuclease H with 100% selenomethionine incorporation by using a defined selenomethionyl starter medium supplemented by 5% Luria-Bertani medium (W.Yang, unpublished results) while in other experiments incorporation was limited to ~90%.

The vitality of the DL41 strain, which derives directly in three steps from the ancestral K12 stock, is an advantage for selenomethionine replacement. Other auxotrophs tested did not tolerate selenomethionine as well. However, the lambda P_L expression system is probably a disadvantage. The potential temperature sensitivity of selenomethionyl proteins can complicate induction of the recombinant plasmid by temperature jump.

Another potential complication is that, as with selenomethionine itself (Shepherd and Huber, 1969), the solubilities of selenomethionyl proteins are somewhat reduced in comparison with their sulfurous counterparts. Although this has not yet prevented us from obtaining crystals of the selenomethionyl products under conditions similar to those used for the natural proteins, this might not always be the case. A more serious complication might arise from susceptibility of selenomethionine to oxidation. We have noted a loss of birefringence and diffraction (despite retention of clean morphology) from crystals of *E.coli* thioredoxin kept for two months in a stabilizing medium. The cause of this change has not been investigated but storage in an oxygen-free environment is obviously indicated.

Finally, although procedures for the production of selenomethionyl proteins in *E.coli* are now well developed, little has been done in eukaryotic cells. As discussed above, mammals are naturally auxotrophic for methionine. However, the greater complexity of mammalian cells tends to make them more sensitive to selenomethionine. Nevertheless, we have achieved 87% incorporation of

selenomethionine into a recombinant soluble CD4 (Deen *et al.*, 1988) produced in CHO cells (P.D.Kwong, unpublished results). Since yeasts contain pathways for production of methionine from cysteine as well as the reverse pathways as found in mammals, appropriate mutants will probably be required for the expression of selenomethionyl proteins in yeast cells.

Selenocysteine as a prospect for MAD phasing

Those few proteins that naturally contain selenocysteine are obvious candidates for analysis by MAD phasing. For example, in light of results in Table III, there is little doubt that the structure of glutathionine peroxidase with one selenocysteine in each 182 residue chain (Epp *et al.*, 1983) could have been solved in this way. Unfortunately, the failure of *E.coli* grown on selenate to incorporate selenocysteine into β-galactosidase (Huber and Criddle, 1967) makes a general replacement scheme such as used for selenomethionine seem unlikely. On the other hand, the recent discovery of utilization of the UGA termination codon in the biosynthesis of selenocysteinyl proteins (Chambers *et al.*, 1986; Zinoni *et al.*, 1986) suggests a general approach through the engineering of selenocysteinyl mutants. This, however, is complicated by the apparent involvement of the nucleotide context of the UGA codon in the signal of selenocysteine incorporation. An alternative possibility suggested by Jeffrey Miller (personal communication) would be to construct and use an amber suppressor gene (Normanly *et al.*, 1986) starting from the *E.coli selC* gene which codes for the special tRNA species leading to selenocysteine insertion (Leinfelder *et al.*, 1988). Both plans suffer from possible adverse effects of the susceptibility of selenocysteine to oxidation. Epp *et al.* (1983) found that the selenocysteine of glutathione peroxidase is oxidized to seleninic acid. It might be possible, however, to capitalize on this proclivity toward oxidation to form especially stable diselenide selenosulfide analogs of disulfide bridges.

Materials and methods

Methionine auxotroph transduction

The DL41 strain was constructed by phage P1 transduction of mal E: Tn*10* from strain TST1 (CGSC strain no. 6137) into MG1655 (Guyer *et al.*, 1981) followed by transduction of *metA* from AB1932 (Howard-Flanders *et al.*, 1966) by selection for maltose utilization and screening for methionine auxotrophy.

Thioredoxin overproduction and purification

A starter culture of the DL41 strain containing the pDL59 expression plasmid (LeMaster and Richards, 1988) was grown overnight at 33°C in a defined medium that contains 40–50 mg/l methionine. This starter culture was then diluted 1/200 (v/v) into prewarmed medium devoid of methionine but containing 40–50 mg/l D,L selenomethionine (Sigma). This leads to only a 0.5% residue of sulfurous methionine in the growth medium; it is critical to minimize this methionine content. Cells in this medium were grown at 33°C to OD_{600} of 1.0 , at which point transcription from the plasmid was induced by abrupt shift of temperature to 43°C. Expression was allowed to continue for 2.5 h.

Purification of the *E.coli* and T4 thioredoxins expressed in these fermentations was carried out as previously described (LeMaster and Richards, 1988). In brief, the debris from sonicated cells was pelleted and the supernatant was treated by streptomycin sulfate and ammonium sulfate fractionation . The thioredoxin-containing fraction was then resuspended, further fractionated on a sizing column, and then separated on a DE52 column into the two constituent thioredoxins. Finally, the samples were lyophilized after dialysis against 25 mM NH_4HCO_3.

Amino acid analysis

Analyses for amino acid composition were performed on a Beckman 6300A analyzer after hydrolysis *in vacuo* with 6 N HCl at 110°C for 24h.

Crystallization

Crystallizations were carried out with the hanging-drop vapor diffusion technique in 24-well Linbro cell culture plates. All chemicals were obtained from Sigma Chemical Co., St Louis, MO unless otherwise noted.

T4 thioredoxin

Lyophilized natural and selenomethionyl T4 thioredoxins were dissolved in buffer containing 40 mM bis-tris propane, pH 6.8, to a concentration of ~13 mg/ml. Droplets (3 μl) of this saturated solution were placed on siliconized glass coverslips which were then inverted over wells containing a 0.5 ml solution of 12–28% ethanol, the bis-tris propane buffer and 2 mM cadmium acetate (Aldrich) and sealed with silicone vacuum grease. The plates were stored at 4°C for 5–7 days at which time 1 μl of the well solution was added to the drops. Clusters of small crystals appeared to grow from amorphous material in some of the 16–18% ethanol setups ~5 days later.

E.coli thioredoxin

Starting droplets for crystallization of the *E.coli* proteins were constituted from 5 μl of protein solution (~10 mg/ml), 5 μl of a pH 4.2 buffer including 10 mM sodium acetate and 2 mM cupric acetate, and 1 μl of 40% (400 mg/100 ml) polyethylene glycol 3350 (PEG 3350). Such droplets formed on siliconized coverslips were inverted and sealed over 1 ml reservoirs containing 12–24% PEG 3350, 5 mM acetate buffer (pH 4.2), and 1 mM cupric acetate. These experiments were carried out in an incubator held at 20°C. Crystals of selenomethionyl *E.coli* thioredoxin first appeared after 4–5 days whereas those of the native protein appeared in 10–14 days. The best crystals grew over wells of 16–20% PEG without appreciable distinction between the two proteins. However, crystals only grew in some setups which suggested to us that nucleation was a rate limiting process. The yield of usable crystals was improved by microseeding. Grains of washed and dried sand (Baxter, Edison, NJ) that were added to some droplets also appeared to increase the yield of crystals.

The crystals used to determine unit cell dimensions (Table I) and at SSRL for X-ray absorption spectroscopy (Figure 4) and synchrotron data collection were grown from material produced as described above from protein solutions of 11 mg/ml over wells of 16% PEG. The crystals and diffraction patterns shown in Figure 3 were obtained from protein produced in subsequent fermentations. These crystallizations were carried out over 20% PEG wells starting from stock solutions of 13.9 and 8.6 mg/ml for the natural and selenomethionyl proteins respectively. The selenomethionyl thioredoxin in this case proved to have only 90% substitution in this instance due to an excess of methionine in the starter culture.

Crystal characterization

Crystal growth and morphology were monitored by examination with a Zeiss stereomicroscope. Crystals were mounted in thin-walled glass capillaries for diffraction analysis. Symmetry properties were determined by photography on a Huber precession camera. Unit cell dimensions were determined from ~20 reflections centered on a Rigaku AFC-5 diffractometer. X-rays used in these experiments were emitted from the copper anode of a Rigaku RU-200 generator.

X-ray absorption spectroscopy

X-ray absorption spectra were measured by scintillation-counter detected X-ray fluorescence using the area detector system (Phizackerley *et al.*, 1986) at the Stanford Synchrotron Radiation Laboratory. The method of measurement and the procedures for reduction of the normalized fluorescence spectrum to scattering factor curves were those described in Hendrickson *et al.* (1988).

Acknowledgements

We thank Ray Salemme for suggesting thioredoxin as a suitable test system for selenomethionyl protein production, Paul Phizackerley and Ethan Merritt for help with synchrotron experiments at Stanford, Bob Batterman for measuring the absorption spectrum from selenomethionine at the Cornell synchrotron, Marcos Hatada and Brad Graves for sharing results on selenomethionyl interleukins, Mary Ann Gawinowitz for amino acid analyses, Jeffrey Miller for discussions about selenocysteine, and our colleagues Peter Kwong, Tom Peat and Wei Yang for unpublished results and helpful discussions. This work was supported in part by a grant (GM34102) from the National Institutes of Health, and the synchrotron facilities used at Stanford are supported by the Department of Energy and by the National Institutes of Health.

References

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice, M.D., Rogers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Bremer,J. and Natori,Y. (1960) *Biochim. Biophys. Acta*, **44**, 367–370.

Chambers,I., Frampton,J., Goldfarb,P., Affara,N., McBain,W. and Harrison,P.R. (1986) *EMBO J.*, **5**, 1221–1227.

Cowie,D.B. and Cohen,G.N. (1957) *Biochim. Biophys. Acta.*, **26**, 252–261.

Cromer,D.T. (1983) *J. Appl. Crystallogr.*, **16**, 437.

Dayhoff,M.O. (1978) *Atlas of Protein Sequence and Structure.* National Biomedical Research Foundation, Washington, D.C., Vol. 5, Suppl. 3, p. 363.

Deen, K. C., McDougal, J.S., Inacker, R., Folena-Wasserman, G., Arthos,J., Rosenberg, J., Maddon, P. J., Axel, R. and Sweet,R.W. (1988) *Nature*, **331**, 82–84.

Epp,O., Ladenstein,R. and Wendel,A. (1983) *Eur. J. Biochem.*, **133**, 51–69.

Frank,P., Licht,A., Tullius,T.D., Hodgson,K.O. and Pecht,I. (1985) *J. Biol. Chem.*, **260**, 5518–5525.

Germain,G., Main,P. and Woolfson,M.M. (1970) *Acta Crystallogr. B*, **26**, 274–285.

Graves,B.J., Hatada,M.H., Hendrickson,W.A., Miller,J.K., Madison,V.S. and Satow,Y. (1990) *Biochemistry*, in press.

Guss,J.M., Merritt,E.A., Phizackerley,R.P., Hedman,B., Murata,M., Hodgson,K.O. and Freeman,H.C. (1988) *Science*, **241**, 806–811.

Guyer,M., Reed,R.R., Steitz,J.A. and Low,K.B. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 135–140.

Hahn,G.A. and Brown,J.W. (1967) *Biochim. Biophys. Acta*, **146**, 264–271.

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

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

Hendrickson,W.A., Smith,J.L. and Sheriff,S. (1985) *Methods Enzymol.*, **115**, 41–55.

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

Hendrickson,W.A., Pähler,A., Smith,J.L., Satow,Y., Merritt,E.A. and Phizackerley,R.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2190–2194.

Holmgren,A. (1968) *Eur. J. Biochem.*, **6**, 475–484.

Holmgren,A., Söderberg,B.-O., Eklund,H. and Brändén,C.-I. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 2305–2309.

Howard-Flanders,P., Boyce,R.P. and Theriot,L. (1966) *Genetics*, **53**, 1119–1136.

Huber,R.E. and Criddle,R.S. (1967) *Biochim. Biophys. Acta*, **141**, 587–599.

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

Katti,S.K., LeMaster,D.M. and Eklund,H. (1990) *J. Mol. Biol.*, in press.

Krishna Murthy,H.M., Hendrickson,W.A., Orme-Johnson,W.H., Merritt,E.A. and Phizackerley,R.P. (1988) *J. Biol. Chem.*, **263**, 18430–18436.

Leinfelder,W., Zehelein,E., Mandrand-Berthelot,M.-A. and Böck,A. (1988) *Nature*, **331**, 723–725.

LeMaster,D.M. (1986) *J. Virol.*, **59**, 759–760.

LeMaster,D.M. and Richards,F.M. (1985) *Biochemistry*, **24**, 7263–7268.

LeMaster,D.M. and Richards,F.M. (1988) *Biochemistry*, **27**, 142–150.

Lunn,C.A., Kathju,S., Wallace,B.J, Kushner,S.R. and Pigiet,V. (1984) *J. Biol. Chem.*, **259**, 10469–10474.

McConnell,K.P. and Hoffman,J.L. (1972) *FEBS Lett.*, **24**, 60–62.

Mudd,S.H. and Cantoni,G.L. (1957) *Nature*, **180**, 1052.

Normanly,J., Masson,J.-M., Kleina,L.G., Abelson,J. and Miller,J.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6548–6552.

Odom,J.D. (1983) *Structure and Bonding*, **54**, 1–26.

Phizackerley,R.P., Cork,C.W. and Merritt,E.A. (1986) *Nucl. Instrum. Methods A*, **246**, 579–595.

Rajeswaran,M. and Parthasarathy,R. (1984) *Acta Crystallogr. C*, **40**, 647–650.

Shepherd,L. and Huber,R.E. (1969) *Can. J. Biochem.*, **47**, 877–881.

Sjöberg,B.-M. and Holmgren,A. (1972) *J. Biol. Chem.*, **247**, 8063–8068.

Sjöberg,B.-M. and Söderberg,B.-O. (1976) *J. Mol. Biol.*, **100**, 415–419.

Söderberg,B.-O., Sjöberg,B.-M., Sonnerstam,U. and Brändén,C.-I. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 5827–5830.

Stadtman,T.C. (1987) *FASEB J.*, **1**, 375–379.

Templeton,L.K. and Templeton,D.H. (1988) *Acta Crystallogr. A*, **44**, 1045–1051.

Wiegand,G., Remington,S., Deisenhofer,J. and Huber,R. (1984) *J. Mol. Biol.*, **174**, 205–219.

Zinoni,F., Birkman,A., Stadtman,T.C. and Böck,A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4650–4654.

Received on January 5, 1990