Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# A novel method for detection of selenomethionine incorporation in protein crystals *via* Raman microscopy

**Alessandro Vergara,[a,b,c] Antonello Merlino,[a] Elio Pizzo,[d] Giuseppe D'Alessio[d] and Lelio Mazzarella[a,b,c]***

[a]Dipartimento di Chimica, Università degli Studi di Napoli 'Federico II', Complesso Monte S. Angelo, Via Cinthia, I-80126 Napoli, Italy, [b]Istituto di Biostrutture e Bioimmagini, CNR, Via Mezzocannone 6, I-80134 Napoli, Italy, [c]Consorzio BioTekNet, Università di Napoli Federico II, Napoli, Italy, Complesso Monte S. Angelo, Via Cinthia, I-80126 Napoli, Italy, and [d]Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli Federico II, Napoli, Italy, Complesso Monte S. Angelo, Via Cinthia, I-80126 Napoli, Italy

Correspondence e-mail: lelio.mazzarella@unina.it

Received 31 July 2007
Accepted 13 November 2007

Multiwavelength anomalous dispersion (MAD) is the most widespread approach in structural biology for the determination of the crystal structure of a novel protein. Mass spectrometry is currently used to evaluate the selenomethionine (SeMet) content in solution, but excluding fluorescence spectroscopy at the absorption edge, no other routine method to check the SeMet incorporation and storage in the crystal state is yet available. Raman microscopy is an increasingly popular tool in physical biochemistry, with applications ranging from studies of ligand binding to secondary-structure analysis. Here, a novel methodological development is presented for the analysis *via* Raman microscopy of SeMet-labelled protein crystals to be used for MAD crystallography. The method is described and supported by validation and application to two novel proteins (a $\beta\gamma$-crystallin-like protein and a DNA-binding protein). Markers of the SeMet residues are in the range 570–600 $cm^{-1}$, where proteins do not usually show Raman bands.

## 1. Introduction

Raman microscopy is finding increasing application in structural biology. Ligand binding (Carey & Dong, 2004) and secondary-structure analysis (Carey, 2006) are the main applications of this versatile technique. SeMet derivatives are expressed by biocrystallographers for phase determination using multiwavelength anomalous dispersion (MAD) experiments (Cassetta *et al.*, 1999). MAD is a general method for determining novel crystal structures and is particularly valuable when a starting structural model of the protein is not available. Selenomethionine (SeMet) inclusion is a widespread approach for MAD experiments since Met residues are present with an average occurrence of one per 59 residues (Hendrickson *et al.*, 1990). In solution, mass spectrometry is a destructive technique to evaluate the SeMet content after expression. Only when crystals of an SeMet-derivative protein have been obtained can the absorption spectrum be used to reveal the presence of selenium for determination of the crystal structure. SeMet-derivative crystals should be stored in a reducing environment (usually by adding DTT) and diffraction experiments should be carried out using fresh crystals. It has been reported that a two-week storage of SeMet protein crystals can produce significant deterioration of the crystal diffraction power (Doublié, 1997). Since synchrotron beam time is not always available immediately after the growth of crystals, a tool to check the SeMet status in stored crystals would be beneficial.

**Table 1**
Data-collection statistics.

Values in parentheses are for the highest resolution shells.

| | Native geodin | SeMet geodin |
|---|---|---|
| Space group | $P2_1$ | $P2_1$ |
| Unit-cell parameters | | |
| $a$ (Å) | 27.976 | 27.518 |
| $b$ (Å) | 61.836 | 61.522 |
| $c$ (Å) | 50.574 | 50.027 |
| $\beta$ (°) | 97.209 | 97.380 |
| Data collection | | |
| Resolution limits (Å) | 20.00–1.90 (1.97–1.90) | 30.00–1.75 (1.81–1.75) |
| No. of observations | 93522 | 96450 |
| No. of unique reflections | 12795 | 31334 |
| Completeness (%) | 94.6 (91.1) | 95.3 (87.6) |
| $I/\sigma(I)$ | 30 (10) | 15 (6) |
| $R_{\rm merge}$ (%) | 4.9 (11.9) | 7.1 (9.7)† |
| Mosaicity | 0.17 | 0.80 |

† Averaged over Friedel pairs.

Here, we propose Raman microscopy as a fast tool to detect SeMet incorporation in crystals grown for MAD experiments. Raman microscopy has been performed on single crystals of wild-type and SeMet $\beta\gamma$-crystallin from *Geodia cydonium* (geodin), the crystallization and preliminary X-ray diffraction of which are reported here. *G. cydonium* sponges are very primitive organisms that diverged more than 500 million years ago (some 300 million years earlier than mammals). A crystallin-like gene has been identified in the *G. cydonium* genome (Di Maro *et al.*, 2002). The recombinant protein has been expressed (molecular weight 17 780.8 Da) and preliminarily characterized *via* circular dichroism (Giancola *et al.*, 2005). Although the function of geodin in the sponge *G. cydonium* is still unknown, the structural characterization of geodin could provide information on how homologous monomeric proteins evolved (D'Alessio, 2002).

From comparison of Raman spectra performed on single crystals of wild-type and SeMet geodin, we identified the marker Raman bands of SeMet inclusion in the 570–600 $cm^{-1}$ region. The procedure has been validated by applying Raman microscopy to another SeMet-derivative crystal of a DNA-binding protein (BALQ, UniProtKB/TrEMBL entry Q97YG9), the crystal structure of which has been solved already *via* MAD (A. Di Fiore, R. M. Vitale, G. Fiorentino, P. Amodeo, R. Ronca, C. Pedone, S. Bartolucci and G. De Simone, paper in preparation).

## 2. Experimental section

### 2.1. Expression and purification

Wild-type recombinant geodin was prepared according to the procedure described in Giancola *et al.* (2005). SeMet geodin was produced by inhibition of the methionine-biosynthesis pathway according to procedures described elsewhere (Van Duyne *et al.*, 1993). *Escherichia coli* BL21 (DE3) cells transformed with the cDNA encoding geodin (Giancola *et al.*, 2005) were grown overnight at 310 K in LB medium containing 100 mg $l^{-1}$ ampicillin. Cells were spun down and resuspended in 20 ml minimum medium M9 containing a carbon source (glucose) at 4 g $l^{-1}$ and then added to 1 l of the same prewarmed medium. Cells were grown to mid-log phase before addition of the amino acids lysine, phenylalanine and threonine at 100 mg $l^{-1}$, isoleucine, leucine and valine at 50 mg $l^{-1}$ and L-selenomethionine at 60 mg $l^{-1}$. Induction of protein expression with 0.2 m*M* IPTG was performed 15 min after addition of the amino acids. Purification of selenomethionyl geodin was essentially similar to that of the wild-type recombinant protein (Giancola *et al.*, 2005). Furthermore, ES-MS analysis of SeMet geodin (17 921 Da, see supplementary material[1]) indicated that >95% substitution at the four methionine sites was achieved.

**2.1.1. Crystallization**. The geodin stock was 3 mg $ml^{-1}$ protein in 0.1 *M* ammonium acetate pH 4.5. Crystallization screening was performed using Hampton Research Crystal Screens 1 and 2. Small crystals were obtained with solution No. 35 from Crystal Screen 1. Optimization of the crystallization conditions was performed by changing the protein and precipitant concentrations, the pH and the temperature. Good-quality crystals of both wild-type and SeMet geodin were grown *via* the hanging-drop method by mixing 2 µl protein solution at 2.4 mg $ml^{-1}$ with 2 µl reservoir solution containing 0.2 *M* ammonium sulfate, 17% PEG 8000 and 2 m*M* dithiotreitol at pH 5.0 (50 m*M* acetate solution), 6.5 (50 m*M* cacodylate solution) or 7.5 (50 m*M* Tris) at 277 K. After a month, a very small number of well diffracting crystals (average dimensions 0.1 × 0.1 × 0.2 mm) appeared in the crystallization trials.

The DNA-binding protein (BALQ) used to validate the method was prepared and crystallized according to the procedure reported elsewhere (A. Di Fiore, R. M. Vitale, G. Fiorentino, P. Amodeo, R. Ronca, C. Pedone, S. Bartolucci and G. De Simone, paper in preparation).

**2.1.2. X-ray crystallography**. Diffraction data were collected using X-ray Cu $K\alpha$ radiation generated by a Nonius FR591 rotating-anode generator operating at 45 kV and 90 mA and equipped with a DIP 2030b imaging plate. Diffraction data from wild-type crystals were also collected at high resolution (1.5 Å) at the XRD1 beamline of the Elettra synchrotron.

**2.1.3. Raman microscopy**. A confocal Raman microscope (Jasco, NRS-3100) was used to obtain the Raman spectra of single crystals. The 632.8 nm line of a He–Ne (25 mW) laser was injected into an integrated Olympus microscope and focused to a spot size of approximately 2 µm using a 20× objective. A holographic notch filter was used to reject the excitation laser line. The Raman backscattering was collected at 180° using a 0.1 or 0.2 mm slit and a 1200 grooves $mm^{-1}$ grating, corresponding to an average spectral resolution of 4 or 9 $cm^{-1}$, respectively. It took 200 s to collect a complete data set using a Peltier-cooled 1024 × 128 pixel CCD photon detector (Andor DU401BVI). Wavelength calibration was

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: SX5077). Services for accessing this material are given at the back of the journal.

performed by using polystyrene and carbon tetrachloride as standards at high and low frequency, respectively. Raman microscopy measurements on wild-type and SeMet-derivative geodin and BALQ samples were conducted on crystals in a drop of mother liquor, using a crystallization plate and the setup described in Carey & Dong (2004).

## 3. Results and discussion

### 3.1. X-ray crystallography

Using the conventional source, geodin crystals diffracted to 1.9 Å resolution. A summary of data-collection statistics is reported in Table 1. Preliminary analysis of these data indicated that the asymmetric unit corresponds to a monomer, giving an estimated solvent content of 50.6%. Attempts to solve the structure by molecular replacement were conducted using several different starting models and using different programs. No successful solutions were found. Therefore, the SeMet-derivative protein was expressed, purified and crystallized under the same conditions as the wild-type protein. Good-quality crystals were obtained in the pH range 5.0–7.5. Preliminary X-ray diffraction data were collected to 1.75 Å in-house for the SeMet derivative (diffraction statistics are reported in Table 1). Both fresh and several-month-old crystals diffracted well, contrary to other experiences with SeMet-derivative crystals (Doublié, 1997). The crystals of SeMet geodin were isomorphous to wild-type crystals. Synchrotron diffraction data ($\lambda$ = 1.00 Å) from SeMet-derivative crystals, recently collected at 0.99 Å resolution, allowed the crystallographic detection of SeMet. The geodin structure was solved *via* SAD phasing and is being refined.

### 3.2. Raman microscopy

A Raman microscopy study was conducted on isomorphous wild-type and SeMet geodin crystals. Spectra at low (400–1200 cm$^{-1}$) and high (1200–1800 cm$^{-1}$) frequencies are shown in Figs. 1(*a*) and 1(*b*), respectively. Usually, the Raman spectrum of a polypeptide is subdivided into three main regions of interest: (i) the range between 870 and 1150 cm$^{-1}$, which is associated with the vibrations of the backbone C$^{\alpha}$—C and C$^{\alpha}$—N bonds, (ii) the range between 1230 and 1350 cm$^{-1}$, which contains the amide III region vibrations associated with normal modes of various combinations of C$^{\alpha}$—H and N—H deformations together with C$^{\alpha}$—C and C$^{\alpha}$—N stretchings (Asher *et al.*, 2001), and (iii) the range between 1630 and 1700 cm$^{-1}$, which is associated with C═O stretching modes and is defined as the amide I region (Ngarize *et al.*, 2004). Furthermore, the lower and higher regions can also be informative: the conformation and detection of disulfide bridges can be investigated in the low-frequency (500–540 cm$^{-1}$) region (Kudryavtsev *et al.*, 1998) and hydrophobic interactions can be investigated by analysing the C—H stretching region (2800–3200 cm$^{-1}$; Chourpa *et al.*, 2006), while the S—H stretching region 2550–2600 cm$^{-1}$ can serve as a valuable probe of local dynamics (Thomas, 1999).

It is worth mentioning for the assignment of the primary structure features of a Raman band that geodin expressed in *E. coli* contains 163 residues (including the N-terminal Met), four SeMet residues (one N-terminal and three internal residues), three Trp residues, eight Phe residues, seven Tyr

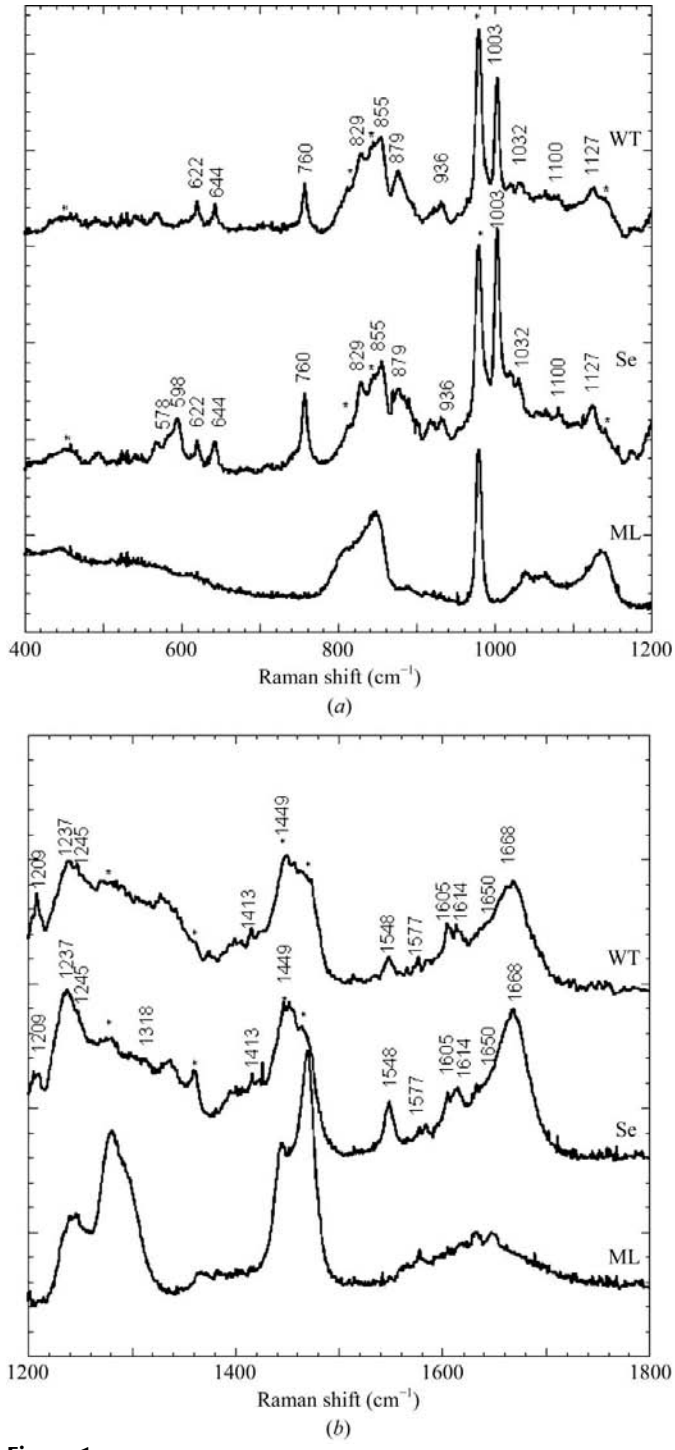


**Figure 1**
Low-frequency (*a*) and high-frequency (*b*) Raman spectra of wild-type geodin crystals (WT), SeMet-labelled crystals (Se) and the mother liquor from which both types of crystals grew (ML). In spectra (*a*) and (*b*) the signals attributed to the mother liquor are indicated by a star. The spectral resolution is 4 cm$^{-1}$.

**Table 2**
Tentative assignment of characteristic Raman bands measured for the SeMet-derivative crystals of the crystallin-like protein from *G. cydonium*.

| Band frequency ($cm^{-1}$) | Tentative assignment of vibration mode | Primary structure | Secondary structure |
|---|---|---|---|
| 1668 | C—O stretching | | Amide I, $\beta$-sheet and random coil |
| 1650, shoulder | | | Amide I, $\alpha$-helix |
| 1614 | Ring stretching, side chain | Tyr-Phe | |
| 1605 | | | |
| 1577 | Ring stretching | Trp | |
| 1548 | Ring stretching | Trp | |
| 1449 | C—$H_2$ scissoring | | |
| 1400 | $COO^-$ stretching | | |
| 1318 | N—H, C—H and $CH_2$ bending | | Amide III, $\alpha$-helix |
| 1245 | | | Amide III, random coil |
| 1237 | | | Amide III, $\beta$-sheet |
| 1205 | C—C stretching | Tyr-Phe | |
| 1159 | | | |
| 1127 | C—C stretching | | |
| 1100 | C—C stretching | | |
| 1032 | C—C stretching | Phe | |
| 1003 | Ring breathing | Phe | |
| 936 | C—C, skeletal stretching | | |
| 889 | C—C, C—O deformations | Trp | |
| 855 | Fermi resonance doublet | Tyr | |
| 829 | | | |
| 760 | C—C, C—O deformations | Trp | |
| 644 | Ring bending | Tyr | |
| 622 | Ring bending | Phe | |
| 598 | Symmetric C—Se stretching | SeMet | |
| 578, shoulder | Asymmetric C—Se stretching | SeMet | |

residues and no Cys residues (Di Maro *et al.*, 2002). Signals corresponding to the mother liquor (which consisted of 17% PEG 8000, 0.2 *M* ammonium sulfate in 50 m*M* Tris–HCl pH 7.5) are marked by a star in the wild-type and SeMet-derivative Raman spectra. The frequency of the major bands and their tentative assignment, proposed in agreement with previous studies (Li *et al.*, 1990), are reported in Table 2.

Raman amide I (Apetri *et al.*, 2006) and amide III (Mikhonin *et al.*, 2006) frequencies correlate with $\varphi$ and $\psi$ angles and hydrogen bonding. Therefore, these bands reveal secondary-structural features. In the amide III region (1230–1320 $cm^{-1}$) of both wild-type and SeMet geodin crystals, the prominent shorter frequencies are assigned to $\beta$-sheets ($\sim$1237 $cm^{-1}$) and random coils ($\sim$1245 $cm^{-1}$), while the very minor higher frequencies ($\sim$1318 $cm^{-1}$) are attributed to $\alpha$-helices. Consistently, in the amide I region of geodin crystals the spectrum indicates the position of major components assignable to $\beta$-sheet and random-coil conformations around 1668 $cm^{-1}$, with traces of $\alpha$-helix (shoulder at 1650 $cm^{-1}$). Altogether, the amide I maximum at 1668 $cm^{-1}$ and amide III bands (with a maximum at 1237 $cm^{-1}$ and shoulder at 1245 $cm^{-1}$) indicate $\beta$-strands/random coils to be the major secondary structures in geodin. These results are in agreement with literature CD data on geodin solutions, which reported a relatively low $\alpha$-helical content (Giancola *et al.*, 2005). No disulfide-bridge stretching bands (500–540 $cm^{-1}$) are observed, consistent with the absence of Cys residues in the primary structure of geodin (Di Maro *et al.*, 2002).

The Raman spectra collected on the isomorphous crystals of wild-type and SeMet geodin crystals reveal the same secondary-structure features. These findings suggest that the presence of SeMet does not alter the structure of geodin, as observed for most proteins. The main difference between the two spectra (excluding the slightly different intensity of the mother-liquor signals) is in one narrow region at low frequency. Indeed, the bands in the 570–600 $cm^{-1}$ region are present only in the SeMet geodin crystal and not in the spectra of the mother liquor or in the wild-type geodin crystal (Table 2). This spectral feature can be confidently assigned to the C—Se stretching, in agreement with previous Raman studies on selenium-containing organic compounds (Hamada & Morishita, 1977; Paetzold *et al.*, 1967) and on the amino acid selenomethionine (Zainal & Wolf, 1995; Lopez *et al.*, 1981).

The procedure developed using geodin crystals has been validated by collecting the Raman spectra of another SeMet-derivative protein (BALQ). Again, the SeMet band at 598 $cm^{-1}$ is clearly only observed in the SeMet derivative of BALQ and not in the wild-type protein or the mother liquor (see supplementary material). Spectra were collected at a lower resolution (9 $cm^{-1}$ resolution) in order to show that high resolution is not required by the method. Indeed, no other protein or mother-liquor bands overlap the SeMet spectral region (the 570–600 $cm^{-1}$ region). Since mother liquor might in principle interfere with identification of the SeMet bands, Raman features of the typical components of commercial crystallization kits (Hampton Research Crystal Screens 1 and 2) have also been analysed. The main interference may come from the use of cacodylate (but not Tris, acetate, phosphateor citrate) as a buffer and from the precipitant MPD (but not PEGs, ammonium sulfate, ethanol or dioxane). Interestingly, glycerol (a typical cryoprotectant) and DTT and EDTA (typical additives for the storage of SeMet derivatives) do not Raman scatter in the SeMet region. Carey and others have shown that difference Raman spectra are sensitive to changes in protein folding and inhibitor binding (Carey, 2006). Fortunately, the majority of these transitions are not in the 570–600 $cm^{-1}$ region.

Regarding the effect of the storage of SeMet crystals, it is known that SeMet oxidation generates an additional band at 789 $cm^{-1}$ assigned to Se—O stretching (Lopez *et al.*, 1981). This mode might be a marker of the SeMet oxidation state as a function of the storage time. No additional band at 789 $cm^{-1}$ was present for crystals of the SeMet derivatives of geodin and BALQ that had been stored for a month, indicating that this storage did not produce significant oxidation in the two cases under study.

## 4. Conclusions

Crystallization experiments yielded crystals of the $\beta\gamma$-crystallin-like protein from *G. cydonium* (geodin) and its SeMet derivative suitable for X-ray diffraction analysis. A Raman microscopy study performed on these crystals suggests that this technique could be a valuable tool to crystallographers, not only to evaluate secondary-structural features and to follow ligand binding *via* soaking but also to check for SeMet inclusion and storage in crystals grown for multi-wavelength anomalous dispersion experiments. Raman spectra of isomorphous crystals of wild-type and SeMet-derivative geodin have been collected *in situ* in a typical hanging-drop plate. Bands in the 570–600 $cm^{-1}$ range are shown to be good markers of SeMet incorporation into the crystals since they do not overlap other protein or additive (such as glycerol) Raman signals. This work suggests that Raman microscopy is as valuable as X-ray fluorescence in detection of the presence of SeMet in protein crystals. Moreover, the multiple applications of Raman microscopy in biocrystallography suggest that this technique may become standard for the preliminary and routine analyses of crystals in structural biology laboratories.

This work was financially supported by Regione Campania. We acknowledge the ELETTRA Synchrotron for providing the synchrotron-radiation facilities and we thank the beamline staff for their assistance during data collection. G. Sorrentino and M. Amendola are acknowledged for their technical assistance. We acknowledge Drs G. De Simone, A. Di Fiore and R. Berisio for kindly providing SeMet-derivative crystals, Dr A. Di Maro for measuring mass spectra and Dr L. Vitagliano for valuable discussion.

## References

Apetri, M. M., Maiti, N. C., Zagorski, M. G., Carey, P. R. & Anderson, V. E. (2006). *J. Mol. Biol.* **355**, 63–71.
Asher, S. A., Ianoul, A., Mix, G., Boyden, M. N., Karnoup, A., Diem, M. & Schweitzer-Stenner, R. (2001). *J. Am. Chem. Soc.* **123**, 11775–11781.
Carey, P. R. (2006). *Annu. Rev. Phys. Chem.* **57**, 527–554.
Carey, P. R. & Dong, J. (2004). *Biochemistry*, **43**, 8885–8893.
Cassetta, A., Deacon, A. M., Ealick, S. E., Helliwell, J. R. & Thompson, A. W. (1999). *J. Synchrotron Rad.* **6**, 822–833.
Chourpa, I., Ducel, V., Richard, J., Dubois, P. & Boury, F. (2006). *Biomacromolecules*, **7**, 2616–2623.
D'Alessio, G. (2002). *Eur. J. Biochem.* **269**, 3122–3130.
Di Maro, A., Pizzo, E., Cubellis, M. V. & D'Alessio, G. (2002). *Gene*, **299**, 79–82.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Giancola, C., Pizzo, E., Di Maro, A., Cubellis, M. V. & D'Alessio, G. (2005). *FEBS J.* **272**, 1023–1035.
Hamada, K. & Morishita, H. (1977). *Spectrosc. Lett.* **10**, 367.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Kudryavtsev, A. B., Mirov, S. B., DeLucas, L. J., Nicolete, C., van der Woerd, M., Bray, T. L. & Basiev, T. T. (1998). *Acta Cryst.* D**54**, 1216–1229.
Li, T. S., Chen, Z. G., Johnson, J. E. & Thomas, G. J. J. (1990). *Biochemistry*, **29**, 5018–5026.
Lopez, J., Jao, T. C. & Rudzinski, W. E. (1981). *J. Inorg. Biochem.* **14**, 177–188.
Mikhonin, A. V., Bykov, S. V., Myshakina, N. S. & Asher, S. A. (2006). *J. Phys. Chem. B*, **110**, 1928–1943.
Ngarize, S., Herman, H., Adams, A. & Howell, N. (2004). *J. Agric. Food Chem.* **52**, 6470–6477.
Paetzold, V. R., Lindner, U., Bochmann, G. & Reich, P. (1967). *Z. Anorg. Allg. Chem.* **352**, 295.
Thomas, G. J. J. (1999). *Annu. Rev. Biophys. Biomol. Struct.* **28**, 1–27.
Van Duyne, G. D., Standaert, P. A., Karplus, S. L., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–125.
Zainal, H. A. & Wolf, W. R. (1995). *Transition Metal Chem.* **20**, 225–227.