Acta Cryst. (1995). D**51**, 814–818

Phasing with Mercury at 1 Å Wavelength

BY CHRISTIAN DUMAS

Centre de Biochimie Structurale, U 414 *INSERM, UMR 9955 CNRS-Université Montpellier* I, 15 *Avenue Charles Flahault,* 34060-*Montpellier, France*

AND STÉPHANE DUQUERROY AND JOËL JANIN*

Laboratoire de Biologie Structurale, UMR 9920 *CNRS-Université Paris-Sud, Bâtiment* 34, 1 *Avenue de la Terrasse,* 91198-*Gif-sur-Yvette, France*

(*Received* 10 *November* 1994; *accepted* 13 *January* 1995)

Abstract

Synchrotron sources provide a continuously tunable X-ray beam which makes it possible to optimize the anomalous contribution to phase determination using heavy-atom replacement. This method was used to solve two protein structures, those of *Dictyostelium discoideum* nucleoside diphosphate kinase and of lobster enolase. The first had 17 kDa of protein in the asymmetric unit, the second, 47 kDa. In both cases, a single mercury derivative yielded single isomorphous replacement with anomalous-scattering phases from which an interpretable electron-density map was derived by solvent flattening. The efficient solution of the X-ray structure was largely due to the large anomalous scattering of mercury at a wavelength shorter than the L_{III} absorption edge.

Background

Anomalous scattering by heavy atoms is commonly used as an additional source of phase information when determining biological crystal structures by isomorphous replacement. In the well established SIRAS (single isomorphous replacement with anomalous scattering) method, the ambiguity of phases derived from a single heavy-atom derivative is broken by the comparison of Bijvoet pairs (Peerdeman, Bommel & Bijvoet, 1951; Blundell & Johnson, 1976). The availability of tunable synchrotron X-ray sources makes anomalous-phasing methods even more attractive. In the MAD (multiwavelength anomalous diffraction) method, which uses dispersive effects near absorption edges, all measurements carried out on the crystal that contains the anomalous scatterer, are free from the lack of isomorphism that affects phases in other methods (Kahn *et al.*, 1985; Hendrickson, 1991; Smith, 1991). Dispersive effects being small compared to those observed in isomorphous replacement, MAD phasing requires very accurate measurements of scattered intensities which are not feasible on all protein crystals. Thus, there is room for methods that combine the advantages of isomorphous replacement with those of a tunable X-ray source. SIRAS phasing with mercury is clearly one of them, and it has been exploited recently (Baker *et al.*, 1990). Values of the two components of the scattering factor for cysteinyl mercury were reported by Tesmer, Stemmler, Penner-Hahn, Jo Davisson & Smith (1994). Just below the L_{III} absorption edge at 1.0091 Å, the imaginary component f'' is 9.991 e, about seven times less than the real component Z less f'.

We report here the phase determination of two protein crystal structures from measurements performed near 1 Å on the wiggler line of the LURE synchrotron radiation center (Orsay, France). Nucleoside diphosphate kinase of *Dictyostelium discoideum* has 1130 non-H atoms (17 kDa) in its asymmetric unit; lobster enolase has 3270 (47 kDa). Substitution with mercury at a single site in the first case, one major and a few minor sites in the second, yielded useful initial phases. This confirms the value of optimizing anomalous scattering when determining phases with one heavy-atom derivative only.

Nucleoside diphosphate kinase

Nucleoside diphosphate kinase (NDP kinase) exchanges a γ-phosphate between nucleoside di- and triphosphates. The protein from the slime mould *D. discoideum* is a hexamer made of identical subunits with 155 residues (Lacombe, Wallet, Troll & Véron, 1990). It was crystallized in space group $P6_322$ by Dumas *et al.* (1991). Because of difficulties in finding heavy-atom derivatives in a protein devoid of cysteine residues, a sulfhydryl group was introduced by site-directed mutagenesis. The H122C mutant protein crystallized under the same conditions as the wild type and in the same space group, but non-isomorphously (Dumas *et al.*, 1992).

X-ray diffraction data were collected on the native H122C protein and on a mercury acetate derivative at the W32 station (Fourme *et al.*, 1992) of the LURE-DCI synchrotron center (Orsay, France) using an imaging-plate system with a diameter of 180 mm (MAR Research,

* To whom correspondence should be addressed.

Table 1. *Crystallographic data on the H122C NDP kinase mutant*

	Native	Hg derivative
Resolution (Å)	2.2	2.2
Observed intensities	82562	123562
Unique reflections	9012	8824
Completeness (%)	99.5	97.5
Reflections with $I > 3\sigma_I$ (%)	91.5	94.8
R_{merge}* (%)	4.6	4.2

* $R_{merge} = \sum_{h,i} |I(h)_i - \langle I(h)\rangle| / \sum_{hi} I(h)_i$.

Table 2. *SIRAS phasing statistics for the H122C NDP kinase mutant*

Phasing statistics from program *MLPHARE* (Collaborative Computational Project, Number 4, 1994). $\langle F_P\rangle$, $\langle F_H\rangle$ and $\langle F_H''\rangle$ are average values of the structure-factor amplitude for acentric reflections of the protein, the normal and anomalous components of the heavy atom; $\langle E\rangle$ and $\langle E''\rangle$ are the r.m.s. isomorphous and anomalous lack of closure for the same reflections. Mean phase errors are average absolute values of the angular difference from phases calculated with the refined atomic model of the H122C protein and the program *SIGMAA* (Collaborative Computational Project, Number 4, 1994).

Isomorphous pairs	8809
(acentric)	7045
Bijvoet pairs	6917
Mean isomorphous difference (%)	18.4
Mean anomalous difference (%)	6.0

Heavy-atom parameters

Site	Real occupancy	Anomalous	Fractional coordinates		
1	7.9	7.6	0.431	0.026	0.111

Resolution shell (Å)	Overall	20–6	6–3.8	3.8–2.8	2.8–2.2
Protein $\langle F_P\rangle$	4777	7308	8839	4891	3019
Heavy atom $\langle F_H\rangle$	1184	1769	1511	1216	952
$\langle F_H''\rangle$	164	285	210	171	139
R.m.s. lack of closure					
$\langle E\rangle$	490	727	708	602	393
$\langle E''\rangle$	174	210	178	181	165
Phasing power					
Isomorphous	2.43	2.43	2.13	2.02	2.42
Anomalous	0.95	1.35	1.18	0.94	0.84
Figure of merit	0.644	0.668	0.673	0.655	0.624
Mean phase error (°)					
SIRAS	42.4	40.8	37.5	41.6	44.6
Solvent flattening	31.6	27.7	32.5	33.9	38.7

Hamburg). The crystal-to-film distance was 171 mm and the crystal was kept at 277 K. One native and one derivative crystals were used for data collection to 2.2 Å resolution. The wavelength was set to 0.98 Å. 1° oscillations were taken with exposure times of 40 s for the native and 60 s for the derivative, covering 90° of rotation for the native data set and 130° for the derivative. Both data sets are nearly complete, highly redundant and of very good quality (Table 1).

Fig. 1 shows Harker sections of Patterson functions calculated with the isomorphous and anomalous differences. The strong anomalous signal originates from the single Hg site at Cys122. The Harker peaks are 22σ (standard deviation) high in the isomorphous-difference map, 18σ high in the anomalous-difference map. SIRAS phases using the Hg-atom position had a mean figure of merit of 0.65. Table 2 shows that it is rather uniformly distributed with resolution. The ratio $\langle F_H\rangle/\langle E\rangle$ measuring the phasing power of the isomorphous signal is near 2.4 over the whole resolution range; the ratio $\langle F_H''\rangle/\langle E''\rangle$, the phasing power of the anomalous signal, decreases from 1.3 to 0.8 at high resolution.

In the resulting SIRAS electron-density map, the protein hexamer centered on a threefold axis of the $P6_322$ unit cell, is clearly separated from the solvent region (Fig. 2). After applying solvent flattening (Wang, 1985; Leslie, 1987), we could trace 75% of the polypeptide chain in strong continuous density with the *BONES* option of *FRODO* (Jones, 1985). SIRAS phases were then combined with information from the partial model, yielding a map from which the entire amino-acid sequence was built (Dumas *et al.*, 1992).

As the structure of hexagonal NDP kinase crystals has now been refined to a R value of 19.6% at 1.8 Å resolution (Moréra *et al.*, 1994), we can evaluate the accuracy of the SIRAS phases. The final phases are 42.4° away on average from the SIRAS phases, or 31.6° from the same phases after solvent flattening (Table 2). Correlation coefficients were calculated between the electron density in several experimental maps and the

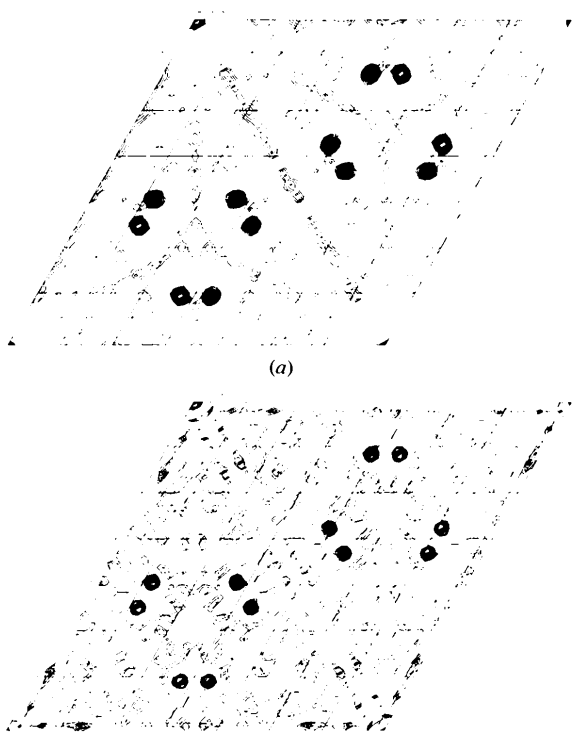

Fig. 1. Harker section (w = 0) of difference Patterson maps for the Hg derivative of H122C mutant NDP kinase: (*a*) isomorphous and (*b*) anomalous differences. Data between 10 and 2.2 Å resolution. Contours in increments of 1σ starting at 1σ.

Table 3. *Crystallographic data on lobster enolase*

	Native	Hg Derivative
Resolution (Å)	2.4	2.9
Observed intensities	78769	43760
Unique reflections	19309	11238
Completeness (%)	91.4	91.6
Reflections with $I > 3\sigma_I$ (%)	72.5	80.8
R_{merge} (%)*	7.7	9.1

* $R_{\text{merge}} = \sum_{hi,i} |I(h)_i - \langle I(h)\rangle| / \sum_{hi} I(h)_i$.

map obtained at 2.2 Å resolution from the refined atomic model. Their value was 0.71 for the SIRAS map and 0.89 for the solvent-flattened map. In contrast, SIR phases calculated by setting f'' to zero yielded a map with only 0.41 correlation to the reference map, demonstrating the importance of the anomalous information.

Enolase

We prepared enolase as a contaminant of arginine kinase from the lobster tail muscle, obtained crystals in trigonal space group $P3_121$, and solved the crystal structure even before we identified it as enolase (Duquerroy, LeBras & Janin, 1994). The protein is a dimer and the asymmetric unit contains a subunit with 433 residues (47 kDa).

A mercury derivative was obtained by soaking crystals for 4 h in 0.1 m*M* thimerosal (ethylmercury-thiosalicylate). X-ray diffraction data sets at a wavelength of 0.91 Å were collected on the W32 station as described above, using one native and one derivative crystal. The crystal-to-film distance was 200 mm and the resolution limit 2.4 Å for the native, 2.9 Å for the derivative; 90° of rotation were recorded by frames of 1.5° with 2 min exposure on each crystal. Statistics on data collection are reported in Table 3. The crystals did not diffract as strongly as for NDP kinase, the data were less complete and of lesser quality: the merging R factor is 8–9% instead of 4–5% and the signal-to-noise ratio much lower. This lesser quality of the enolase data is reflected in Harker sections of the Patterson maps (Fig. 3). The major Hg-atom site shows up clearly on an isomorphous-difference Patterson map as peaks 9σ high. The expected peaks are present in the anomalous map, but only 2.5σ high. We calculated phases with this Hg-atom site and located five others in a Fourier difference map (Table 4). Judging from the refined occupancies and the Cullis R factor for centric reflections (0.72 with two sites, 0.64 with six), the top two sites contribute most of the phase information.

SIRAS phases calculated with six Hg-atom sites had a low mean figure of merit: 0.39 at 3 Å. A comparison of Tables 2 and 4 indicates that this is largely due to the smaller contribution of the heavy atom to the scattered amplitudes. Enolase has one main site per 47 kDa, NDP kinase has one per 17 kDa. In addition, the apparent occupancy of Hg sites in enolase is less for the imaginary than for the real component of the heavy-atom structure factor, whereas the two are nearly equal at the NDP kinase site. The phasing power of the anomalous differences is still significant at medium resolution, and the SIRAS phases proved to be a sufficient starting point for structure determination. The 3 Å SIRAS electron-density map (Fig. 4) shows a recognizeable protein–solvent boundary when calculated in space group $P3_121$, with features inside the protein region that could be interpreted as α-helices and β-strands. A map calculated in the enantiomorph space

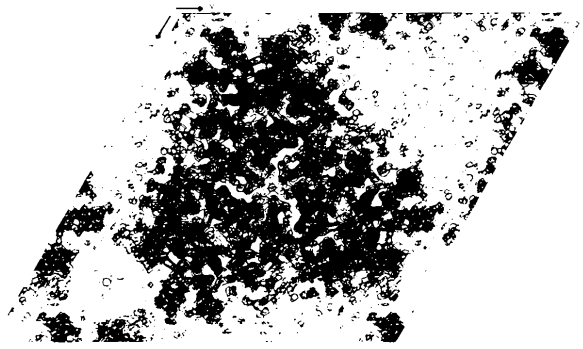

Fig. 2. The 2.2 Å SIRAS electron-density map of NDP kinase. The 20 Å thick slab is cut orthogonal to the c axis of the $P6_322$ cell. Contours in increments of 1σ starting at 1.5σ.

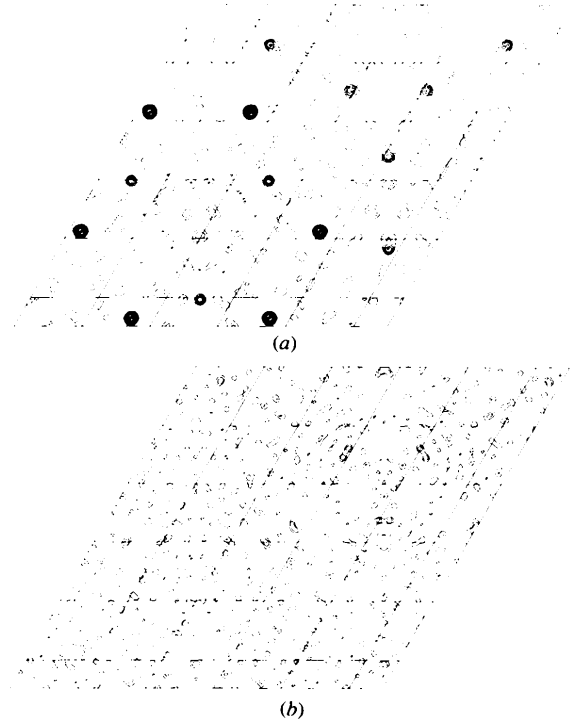

Fig. 3. Harker section (w = 1/3) of difference Patterson maps for the thimerosal derivative of lobster enolase: (*a*) isomorphous and (*b*) anomalous differences. Data between 10 and 3 Å resolution. Contours in increments of 1σ starting at 1σ.

group $P3_221$ was not interpretable, demonstrating the role of the anomalous signal. Nevertheless, the SIRAS map of enolase could not be used as such to build an atomic model. It was subjected to iterative solvent-flattening procedure, beginning at 4.5 Å resolution and progressively extending the resolution to 3.0 Å by adding 15% more reflections, calculating a new envelope and running five cycles of solvent flattening at each step. This had a dramatic effect on the figure of merit which increased from 0.37 to 0.87, and significantly improved the electron-density map as judged from the connectivity in the polypeptide chain.

The solvent-flattened map shows pairs of closely associated protein subunits related by crystallographic twofold axes. In the absence of an amino-acid sequence, we were able to trace the polypeptide chain over 350 residues with few ambiguous connections, using the *BONES* option in the graphic program *O* (Jones, Zou, Cowan & Kjeldgaard, 1991) and checking the original SIRAS map to detect regions which might have been wrongly attributed to the solvent. As reported by Duquerroy *et al.* (1994), it was at this stage that the protein was identified by sequencing two short peptides. Yeast enolase was the only one for which an X-ray structure was available (Lebioda, Stec & Brewer, 1989) and the lobster sequence was unknown. Our data confirm that the two proteins are closely related. A model of yeast enolase (file 3ENL) manually positioned in the lobster electron-density map, confirmed the chain tracing. Rigid-body adjustment yielded a model that had an R factor of 40% against our 3 Å diffraction data. The fit to the electron density was good over most of the molecule, with obvious differences where side chains were not conserved. The determination of the lobster enolase gene sequence (to be reported elsewhere) eventually confirmed the validity of this model and enabled refinement. The current model has an R factor of 20% at 2.4 Å. It yields phases that differ by 50° on average from the SIRAS phases at 3.5 Å resolution and 44° from the solvent-flattened phases.

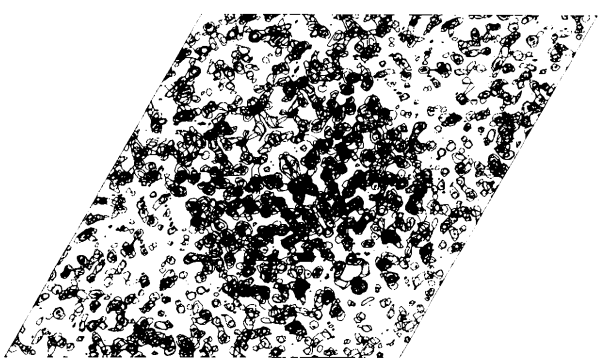

Fig. 4. The 3 Å SIRAS electron-density map of lobster enolase. The 20 Å thick slab is cut orthogonal to the c axis of the $P3_121$ cell. Contours in increments of 1σ starting at 1.5σ.

Table 4. *SIRAS phasing statistics for lobster enolase*

Phasing statistics from program *MLPHARE* (Collaborative Computational Project, Number 4, 1994). $\langle F_P \rangle$, $\langle F_H \rangle$ and $\langle F_H'' \rangle$ are average values of the structure-factor amplitude for acentric reflections of the protein, the normal anomalous components of the heavy atom; $\langle E \rangle$ and $\langle E'' \rangle$ are the r.m.s. isomorphous and anomalous lack of closure for the same reflections.

Isomorphous pairs	6115
(acentric)	5592
Bijvoet pairs	5238
Mean isomorphous difference (%)	14.8
Mean anomalous difference (%)	9.1

Heavy-atom parameters

Site	Real Occupancy	Anomalous	Fractional coordinates		
1	6.2	2.8	0.55	0.12	0.01
2	3.2	1.2	0.40	0.33	0.27
3	1.9	1.0	0.27	0.96	0.26
4	1.6	0.3	0.36	0.27	0.24
5	1.1	0.8	0.48	0.27	0.11
6	1.0	0.3	0.53	0.78	0.11

Resolution shell (Å)	Overall	20–7.8	7.8–5	5–3.7	3.7–2.9
Protein $\langle F_P \rangle$	4184	7266	4760	5135	3145
Heavy atom $\langle F_H \rangle$	731	1108	985	773	595
$\langle F_H'' \rangle$	39	64	52	42	32
R.m.s. lack of closure					
$\langle E \rangle$	436	683	393	428	425
$\langle E'' \rangle$	380	251	256	312	462
Phasing power					
Isomorphous	1.67	1.62	2.50	1.80	1.40
Anomalous	0.10	0.25	0.20	0.14	0.07
Figure of merit	0.37	0.45	0.53	0.40	0.30

Discussion

Monochromatic beams of high intensity near 1 Å are available at most synchrotron radiation centers from bending magnets or insertion devices. They are well adapted to data collection on crystals of biological macromolecules. This region of the X-ray range has the advantage of a reduced radiation damage and absorption correction, and this contributes to the better quality of data taken at 1 Å compared to Cu $K\alpha$ radiation. In addition, the wavelength can be chosen below the L_{III} absorption edge of heavy elements used for isomorphous replacement, making the anomalous component larger (Baker *et al.*, 1990). Mercury is attractive not only because reactive sulfhydryl groups are common, but also because they can be engineered by site-directed mutagenesis: a cysteine can be introduced as we did for NDP kinase, or it can be removed if the reaction with mercury perturbs the protein structure.

The optimized wavelength made structure solution efficient in the two cases reported here. With NDP kinase, data were collected, phases determined and most of the polypeptide chain traced within a week of the day when crystals of the H122C mutant were obtained. With enolase, a partial chain tracing was available even before we had chemically identified the protein. The quality of the diffraction data obtained with NDP kinase was above average, and the mercury signal relatively strong: one mercury per 1130 protein atoms is expected to yield an average anomalous difference equal to 5.9%

of the average amplitude, significantly more than the estimated error on F and close to the observed value of 6.0%. The resulting SIRAS phases were good over the whole resolution range, the electron-density map was excellent. With enolase, the quality of the diffraction data was average or below average and the mercury signal weaker. The expected anomalous difference is only 3.5% of the average amplitude for one major site per 3270 protein atoms. Though this is less than the standard deviation of the measurements, SIRAS phases gave an electron-density map which, after solvent flattening, proved interpretable in the absence of an amino-acid sequence. As not all protein crystals yield data of the quality we had with NDP kinase, the success with enolase confirms that the method is generally applicable as long as the mercury derivative is isomorphous.*

We thank Professor R. Fourme, Professor J. P. Benoît and the staff of the LURE synchrotron radiation center (Orsay, France) for making station W32 available to us, for help and advice in data collection and processing, and for discussion.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1NDK, 1NPK). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: PA0304).

References

BAKER, P. J., FARRANTS, G. W., STILLMAN, T. J., BRITTON, K. L., HELLIWELL, J. R. & RICE, D. W. (1990). *Acta Cryst.* A**46**, 721–725.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. London: Academic Press.

COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4 (1994). *Acta Cryst.* D**50**, 760–763.

DUMAS, C., LASCU, I., MORÉRA, S., GLASER, P., FOURME, R., WALLET, V., LACOMBE, M. L., VÉRON, M. & JANIN, J. (1992). *EMBO J.* **11**, 3203–3208.

DUMAS, C., LEBRAS, G., WALLET, V., LACOMBE, M. L., VÉRON, M. & JANIN, J. (1991). *J. Mol. Biol.* **217**, 849–859.

DUQUERROY, S., LEBRAS, G. & JANIN, J. (1994). *Proteins*, **18**, 390–393.

FOURME, R., DHEZ, P., BENOÎT, J. P., KAHN, R., DUBUISSON, J. M., BESSON, P. & FROUIN, J. (1992). *Rev. Sci. Instrum.* **63**, 982–987.

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

JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

JONES, T. A., ZOU, J. Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110–119.

KAHN, R., FOURME, R., BOSSHARD, R. , CHIADMI, M., RISLER, J. L., DIDEBERG, O. & WRY, J. P. (1985). *FEBS Lett.* **179**, 133–137.

LACOMBE, M. L., WALLET, V., TROLL, H. & VÉRON, M. (1990). *J. Biol. Chem.* **265**, 10012–10018.

LEBIODA, L., STEC, B. & BREWER, J. M. (1989). *J. Biol. Chem.* **264**, 3685–3693.

LESLIE, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.

MORÉRA, S., LEBRAS, G., LASCU, I., LACOMBE, M. L., VÉRON, M. & JANIN, J. (1994). *J. Mol. Biol.* **243**, 873–890.

PEERDEMAN, A. F., BOMMEL, A. J. V. & BIJVOET, J. M. (1951). *Nature (London)*, **165**, 271.

SMITH, J. L. (1991). *Curr. Opin. Struct. Biol.* **1**, 1002–1011.

TESMER, J. J. G., STEMMLER, T. L., PENNER-HAHN, J. E., JO DAVISSON, V. & SMITH, J. L. (1994). *Proteins*, **18**, 394–403.

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