research papers

Taking MAD to the extreme: ultrafast protein structure determination

**Martin A. Walsh,[a] Irene
Dementieva,[a] Gwyndaf Evans,[a]†
Ruslan Sanishvili[a] and Andrzej
Joachimiak[a,b]***

[a]Building 202, Argonne National Laboratory,
9700 South Cass Avenue, Argonne IL 60439,
USA, and [b]Northwestern University, Depart-
ment of Biochemistry, Molecular Biology and
Cell Biology, Evanston IL 60208, USA

† Present address: MRC Laboratory of Molecular
Biology, Hills Road, Cambridge CB2 2QH,
England.

Correspondence e-mail: andrzejj@anl.gov

Multiwavelength anomalous diffraction data were measured
in 23 min from a 16 kDa selenomethionyl-substituted protein,
producing experimental phases to 2.25 Å resolution. The data
were collected on a mosaic 3 × 3 charge-coupled device using
undulator radiation from the Structural Biology Center 19ID
beamline at the Argonne National Laboratory's Advanced
Photon Source. The phases were independently obtained
semiautomatically by two crystallographic program suites,
*CCP*4 and *CNS*. The quality and speed of this data acquisition
exemplify the opportunities at third-generation synchrotron
sources for high-throughput protein crystal structure deter-
mination.

1. Introduction

As of September 1998, 7579 protein coordinate sets have been
deposited with the Protein Data Bank (Bernstein *et al.*, 1977).
Although this is a remarkable achievement, the number of
structures determined pales in comparison with the number of
proteins which are still being discovered. Many factors have
contributed to this disparity, such as the availability of stan-
dardized molecular-biology procedures and protocols in
protein purification and, in particular, the progress being
made in genomic analyses, the most ambitious of which is the
sequencing of the human genome (Collins & Galas, 1993;
Rowen *et al.*, 1997). In spite of the tremendous advances in the
structural biology field, rapid protein structure determination
is still hindered by two main stumbling blocks: protein crys-
tallization and initial phase determination. In order to bridge
the gap between genomics and structural biology, these
obstacles need to be urgently addressed.

Crystallization has seen significant developments in the past
decade, the most noteworthy being the use of incomplete
factorial analysis of conditions which may produce suitable
crystals for X-ray diffraction experiments (Carter, 1992;
Jancarik & Kim, 1991). Developments in phasing techniques
have also received significant attention, with much effort being
put into the improvement of phases obtained from multiple
isomorphous replacement, which remains the principal
method for *ab initio* phasing of protein structures. (For a
recent review, see Ke, 1997, and references therein.)

Here, we concentrate on recent developments which can
considerably reduce the time and effort required to obtain
phase information using the multiwavelength anomalous
diffraction (MAD) method, which has become popular in the
last several years owing to the work by Hendrickson and co-
workers (Hendrickson, 1991). The MAD method exploits
resonance or anomalous scattering, which occurs when the
energy of the incident X-rays approaches that of an electronic
transition in an atom. The effect is energy-dependent implying

that diffraction data be collected at a number of carefully selected X-ray energies. The successful use of anomalous scattering for protein structure determination requires measuring very small differences in diffraction intensities, and thus places heavy demands on the accuracy of the experiment. Yet the method is rapidly becoming more widespread, owing to the advent of dedicated synchrotron beamlines providing well collimated stable and tunable X-ray sources with high brightness (Smith *et al.*, 1996). Furthermore, cryo-cooling of protein crystals (Garman & Schneider, 1997; Rodgers, 1997) decreases the rate of radiation damage and often allows all data to be measured from a single crystal. This significantly reduces systematic errors in the experiment. The applicability of MAD has been extended by the ease of incorporation of Se atoms in the form of selenomethionine (Se-Met) into the protein, providing a source of measurable anomalous signal (Doublié, 1997; Hendrickson *et al.*, 1990). This technique eliminates the arduous search for suitable heavy-atom derivatives and the problems of non-isomorphism between native and derivative crystals, as all data are collected from the same crystal. As the Se atoms are covalently bound, problems with poorly occupied sites are eliminated, although disordered selenium sites are not uncommon.

Time is an important factor when collecting MAD data, as the stability of the system is crucial and also because at synchrotrons only a short period of time is assigned to the user (typically 48 h). Previous estimates of the time required to complete a typical MAD experiment at a synchrotron source have been 24–36 h (Smith *et al.*, 1996). This is because of the multiple data sets, experimental optimization and strategies needed to perform the experiment (Hendrickson, 1991). Finding ways to reduce MAD data-collection times without compromising data quality can only enhance the method.

In this paper, we describe the first MAD experiment at the Structural Biology Center 19ID undulator beamline using the configuration which is available to the crystallographic community. For this test experiment, we selected the apical domain of the chaperonin from *Thermus thermophilus*, as the protein was an ongoing project in our laboratory. Besides, it is small in size, crystallizes readily, is relatively stable to X-radiation and contains three methionines in 145 amino acids, providing an expected anomalous signal of approximately 6%, thus making it a good test case for a selenomethionine MAD experiment. The results show that high-quality MAD data can be collected in less than 30 min for a small protein. The protocol used to collect the data and the subsequent rapid determination of the protein's three-dimensional structure is described.

2. Materials and methods

2.1. Cloning, expression and purification.

The apical domain (AD) of *T. thermophilus* has been cloned and overexpressed in *Escherichia coli*. The amino-acid sequence contains three methinione residues in addition to the N-terminal one, which appears to be processed in *E. coli*. The

AD of *T. thermophilus* chaperonin (residues 192–336) was cloned by the polymerase chain reaction into polylinker sites of pET21 vector (Novagen). Two primers were designed (5′-AAGTTCCATATGGGGTACCAGTTTGACAAGGGGTAC and 5′-TAGAATTCATTAGCCCTTGCCGCCCACGATG-GTGGTC) to flank the respective chaperonin fragment with pET21 vector restriction-enzyme sites *Nde*I and *Eco*RI, yielding the AD expression vector pID3. The Se-Met labelled AD was expressed in the presence of selenomethionine using conditions known to inhibit methionine biosynthesis (Van Duyne *et al.*, 1993). The AD expression was induced with isopropyl-β-D-galactopyranoside (IPTG) to 1 mM final concentration. The cells were harvested 15 h after induction and re-suspended in 150 ml buffer A (50 mM Tris–HCl pH 8.5, 1 mM DTT, 0.5 mM EDTA, 50 mM NaCl). After sonication and centrifugation (30000g, 20 min) the soluble protein was loaded on a 25/35 Q Sepharose Fast Flow column (Pharmacia) equilibrated with buffer A. Unbound protein was concentrated using a 5 kDa molecular-weight cut-off membrane (Amicon) and then loaded on a 10/10 Mono S column (Pharmacia) equilibrated with buffer B (50 mM sodium acetate pH 4.7, 2 mM DTT, 0.5 mM EDTA, 50 mM NaCl). The AD was eluted using a linear gradient of 0–0.5 M NaCl in buffer B. The Se-Met incorporation was examined by amino-acid analysis and were found to be approximately 93% (W. M. Keck Foundation, Biotechnology Resource Laboratory, Yale University).

2.2. Crystallization, data collection and processing

Crystals of the AD of *T. thermophilus* chaperonin were grown from hanging drops initially containing 21 mg ml^{-1} protein, 4.5% PEG 6K, 50 mM HEPES pH 7.5, 2.5% MPD at 293 K and equilibrated against a reservoir consisting of 9% PEG 6K, 100 mM HEPES pH 7.5 and 5% MPD. These crystals belonged to the orthorhombic space group $C222_1$ with unit-cell dimensions $a = 62.7$, $b = 64.7$, $c = 74.2$ Å. Data were collected at 100 K from one crystal which had been flash-frozen with liquid nitrogen using glycerol as a cryoprotectant. The first step of the experiment involved taking a test diffraction image from which we assessed the optimal exposure time, oscillation range, orientation of the crystal and resolution of the data to be collected. The crystal was oriented by means of a mini-κ goniostat (G. Rosenbaum, unpublished results) so as to minimize systematic errors in the experiment by recording Bijvoet pairs on the same image. The X-ray flux to the crystal was adjusted by focusing the X-ray beam using the beamline optics and by the use of attenuators. This allowed the collection of diffraction images which resulted in little or no overloaded reflections. The selection of the X-ray flux was also governed by the desire to expose the crystal to the minimal number of X-ray photons and thus collect complete data sets without incurring any significant radiation damage to the crystal. An X-ray flux of 4×10^{11} photons s^{-1} produced a diffraction pattern with no visible overloaded reflections and usable data at the detector edge. An X-ray florescence spectrum was then recorded from the sample, which identifies the

Table 1
Summary of crystallographic analysis.

Data set	Number of observations	Number of unique reflections†	R_{merge}‡ (%), Bijvoet pairs taken as equivalent	R_{merge}§ (%), Bijvoet pairs taken as non-equivalent	Completeness§ (%)	Resolution (Å)
λ_1 (1.1271 Å)	52821	7078 (13136)	3.7 (15.5)	3.2 (13.9)	95.2 (74.0)	20.0–2.25
λ_2 (0.9793 Å)	52234	7176 (12815)	5.2 (11.1)	3.2 (9.6)	96.1 (90.0)	20.0–2.25
λ_3 (0.9791 Å)	59534	7283 (13506)	8.0 (18.3)	3.6 (14.2)	97.8 (94.7)	20.0–2.25
λ_4 (0.9465 Å)	58798	7225 (13412)	4.3 (9.6)	2.9 (8.0)	97.8 (98.0)	20.0–2.25

† Numbers in parentheses are for cases where anomalous Bijvoet pairs are considered non-equivalent.   ‡ $R_{merge} = \sum_{hkl} \sum_{i=1}^{N} |\langle I^{hkl} \rangle - I_i^{hkl}| / \sum_{hkl} \sum_{i=1}^{N} I_i^{hkl}$.   § The numbers in parentheses refer to data in the highest resolution shell.

precise energy of the selenium absorption edge in the protein crystal. These data were input into the program *CHOOCH* (Evans, 1998), which produces a plot of both f' and f''. Although, theoretically, data collected at only two energies are required to resolve the phase ambiguity, data at four energies

Figure 1
Harker section ($w = 0$) of the Patterson syntheses calculated using (*a*) the anomalous differences of the Se-Met in the λ_3 data and (*b*) using the three Se sites found by *CNS*. The peaks are contoured starting from 2σ and in increments of 1.5 and 2σ for the observed and calculated Pattersons, respectively.

were collected, again providing redundant measurements for later use in the phasing protocol. The energies for the data collection were chosen to maximize the anomalous signal (f'') and the dispersive signal ($\Delta f'$). This was achieved by measuring data at the absorption-edge maximum of selenium (λ_3, Table 1), maximizing f'', the absorption-edge inflection point (λ_2, Table 1) minimizing f' and at two reference energies above and below the absorption edge (λ_1 and λ_4, Table 1), thus providing the largest dispersive signal between these data and the λ_2 data. At the time of the experiment, the minimum undulator gap setting was restricted. At this gap setting, the intensity of the first harmonic of the undulator spectrum falls off drastically above approximately 13.5 keV. Therefore, λ_4 remote data were measured only 500 eV above the Se *K* absorption edge. This latter point encouraged the measurement of the low-energy remote (λ_1) data set. Data at each of the four energies were then collected using $2°$ oscillation per image with an exposure time of 3 s. The absorption peak data λ_3 were collected first, followed by λ_2, λ_1 and λ_4. A total of $120°$ of data were collected at each wavelength in a single pass. The data were processed and scaled using *HKL*2000 (Otwinowski & Minor, 1997).
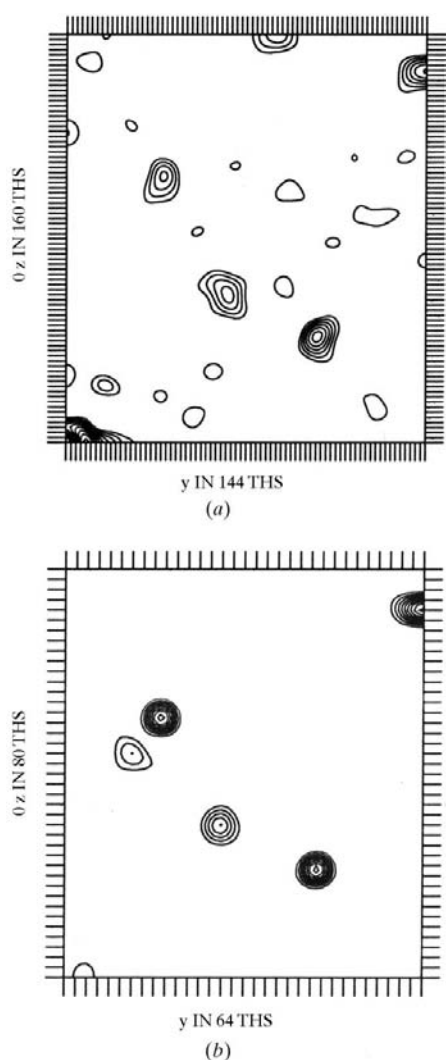
2.3. Structure solution and refinement

Two representative crystallographic software suites, *CCP*4 (Collaborative Computational Project, Number 4, 1994) and *CNS* (Brünger *et al.*, 1998), were used in the structure determination for an independent assessment of the results obtained. A summary of the crystallographic analysis is given in Tables 1 and 2. The AD of *T. thermophilus* chaperonin contains three methionines in its sequence and hence three potential selenium sites. In the *MLPHARE* case, as the data are treated as a special case of MIR, the λ_2 data were treated as native and the λ_1, λ_3, λ_4 data as derivatives (Table 2). This is a convenient choice for the native data set, as it results in a positive dispersive difference between the other data sets. The positions of the Se atoms were found automatically using a real-space Patterson peak algorithm implemented in the program *RSPS* in the *CCP*4 suite. This successfully found two out of the three Se atoms using an anomalous difference Patterson. These sites were refined with the program *MLPHARE* (Collaborative Computational Project, Number 4, 1994) and produced maps of high quality. In the *CNS* case, all three sites were found automatically within the program

using a Patterson search method. An anomalous Patterson map output from *CNS* calculated using the anomalous difference in the λ_3 data is shown in Fig. 1(*a*) and the predicted Patterson using the Se-atom positions found in the automatic Patterson search implemented in *CNS* is shown in Fig. 1(*b*). These Patterson maps show that the third selenium site partially overlaps with the first selenium site (Fig. 1) and provide an explanation of why *RSPS* failed to locate it. Furthermore, this selenium site turns out to be disordered in the refined structure, its *B* factor refining to 62.4 Å^2, and thus contributes little to the phases obtained from *MLPHARE* and *CNS*.

The maps obtained from the two analyses are qualitatively and quantitatively similar, having overall figures of merit (FOMs) of 0.73 in both cases, when all three sites and four wavelengths were used. A section of the experimental electron density calculated using these phases is shown in Fig. 2(*a*), which can be compared with the final map calculated from the model refined against the λ_4 2.25 Å data (Fig. 2*b*).

The solvent-flattened map produced by the program *DM* (Collaborative Computational Project, Number 4, 1994) from phases output by *MLPHARE* were used as input into the *wARP* procedure, which extended and improved phases against data collected to 1.7 Å resolution at room temperature from a separate crystal. The *wARP* protocol generated pseudo-protein models consisting of free atoms which were placed in significant density by the program. The procedure is described in detail elsewhere (Perrakis *et al.*, 1997; Van Asselt *et al.*, 1998). This free-atom model was refined against the data at 1.7 Å resolution with *ARP* (Lamzin & Wilson, 1997) and averaged as described by Perrakis *et al.* (1997). The resulting map was then used to automatically build an improved free-atom model. This model was then utilized to automatically trace all possible polypeptides (V. S. Lamzin and A. Perrakis, unpublished results). Then 200 cycles of restrained refinement with *ARP*/*REFMAC* were run, in which the model was updated every cycle and retraced from scratch every ten cycles. At the end of this procedure, 142 peptides were traced in two chains consisting of 125 and 17 amino-acid residues. Residues 209, 210 and the C-terminal glycine were missing from this model. Of the 145 side-chains, 75 (52%) were automatically built into the model. The procedure was fully automatic, with the only required input being the experimental amplitudes and phases. The protocol took approximately 16 h of computer time on an R10000 Silicon Graphics workstation. Completion of the refinement was straightforward using the program *REFMAC* (Murshudov *et al.*, 1997).

Manual model adjustment was performed in the program *O* (Jones *et al.*, 1991). The model refined against the λ_4 data at 2.25 Å has an *R* factor of 19.7% for all data (7224 reflections) and an R_{free} value (5% of data) of 27.8%. The r.m.s. deviations in bond lengths and angles from ideality are 0.018 and 0.049 Å, respectively. The final model refined against the data collected to 1.7 Å resolution at room temperature has an *R* factor of 19.3% for all data (15 676 reflections) and an R_{free} value (5% of data) of 25.7%. The r.m.s. deviations in bond lengths and angles from
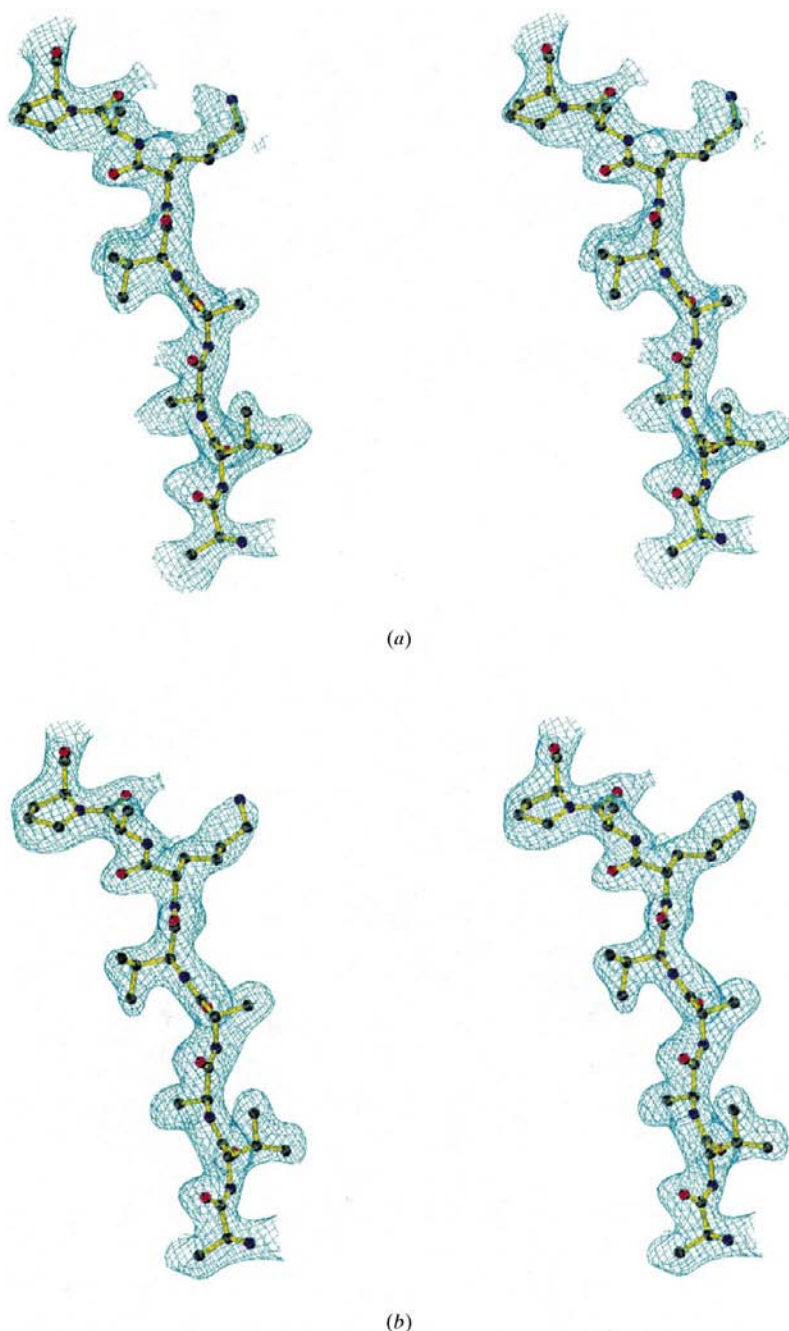

Figure 2
Stereoview of residues 272–279 showing (*a*) the experimental electron-density map (before density modification) and (*b*) the $2F_o - F_c$ electron-density map calculated using phases calculated from final model refined against λ_4 data. Both maps are contoured at the 1σ level and calculated using data in the 20–2.25 Å resolution range. The figure was produced with the program *BOBSCRIPT* (Esnouf, 1997).

Table 2
MAD statistics.

Observed diffraction ratios†

	λ_1	λ_2	λ_3	λ_4
λ_1	0.0312	0.0948	0.0745	0.0454
λ_2		0.0576	0.0711	0.0904
λ_3			0.1020	0.0581
λ_4				0.0458

CNS‡

	Phasing power§	Generalized R_{Cullis}¶	R_{Kraut}††
$\lambda_4 \rightarrow \lambda_4^-$	1.02	0.83	0.024
$\lambda_4 \rightarrow \lambda_1^+$	1.36	0.74	0.027
$\lambda_4 \rightarrow \lambda_1^-$	1.91	0.58	0.041
$\lambda_4 \rightarrow \lambda_2^+$	2.57	0.50	0.036
$\lambda_4 \rightarrow \lambda_2^-$	1.66	0.63	0.030
$\lambda_4 \rightarrow \lambda_3^+$	3.01	0.46	0.030
$\lambda_4 \rightarrow \lambda_3^-$	2.56	0.51	0.020

MLPHARE‡

	Phasing power‡‡	R_{Cullis},§§ acentric	R_{Cullis}, centric	R_{Cullis}, anomalous
λ_1	1.96	0.62	0.57	0.97
λ_2¶¶	0.00	—	—	0.61
λ_3	1.12	0.8	0.83	0.55
λ_4	1.90	0.63	0.58	0.63

† Values for the observed anomalous diffraction ratios are $\langle\Delta|F|^2\rangle^{1/2}/\langle|F|^2\rangle^{1/2}$, where $\Delta|F|$ is the absolute value of the Bijvoet differences at one wavelength (along the diagonal) or the dispersive difference between intersecting wavelengths (off-diagonal elements). Values were calculated at 20.0–2.25 Å resolution within *CNS*. ‡ Overall figure of merit (FOM) in both *CNS* and *MLPHARE* was 0.73. § MAD phasing power is defined as $\{\langle|F_{h4}-F_{hi}|^2\rangle / \int_\varphi P_{\lambda4\rightarrow\lambda i}(\varphi)[||F_{\lambda4}|\exp(i\varphi)+F_{hi}-F_{h4}|-|F_{\lambda i}|]^2 d\varphi\}^{1/2}$ for individual lack-of-closure expressions between the reflections of the reference data set λ_4, its Friedel mate (indicated by λ_4^-) and the Friedel pairs measured at the other energies (indicated by λ_i^+ and λ_i^-). F_{hi} are the heavy-atom structure factors and $P_{\lambda4\rightarrow\lambda i}(\varphi)$ is the corresponding phase probability distribution (Phillips & Hodgson, 1980; Burling *et al.*, 1996). ¶ Generalized $R_{\text{Cullis}} = [\langle(\text{LOC})^2\rangle]^{1/2}/(\langle|\Delta F|^2\rangle)^{1/2}$, where LOC is the lack-of-closure error †† $R_{\text{Kraut}} = \sum ||F_p+F_h|-|F_p||/\sum |F_p|$. ‡‡ Phasing power = (F_h/LOC). §§ $R_{\text{Cullis}} = \sum ||F_{ph}(\text{obs})\pm|F_p(\text{obs})||-F_h(\text{calc})/\sum ||F_{ph}(\text{obs})| \pm|F_p(\text{obs})||$. ¶¶ λ_2 assigned as native data set in *MLPHARE* analysis.

ideality are 0.018 and 0.038 Å, respectively. A residue-by-residue assessment of the quality of the experimental maps is plotted in Fig. 3.

3. Conclusions

Anomalous diffraction data were measured at four X-ray energies at the Structural Biology Center's 19ID undulator beamline at the Advanced Photon Source. The protein analysed was the apical domain of the chaperonin from *T. thermophilus*. The time required to collect each of the four independent data sets was 345 s; 23 min in total. The sample was exposed to X-rays for 12 min. The whole experiment took 45 min to complete. This included the mounting and orienting of the crystal on the crystal goniostat, acquisition of an X-ray fluorescence spectrum to determine the position of the selenium absorption edge in the crystal, tuning and optimization of the four X-ray energies used and collection of the diffraction data. This reduction in data-collection time is in itself

noteworthy, but the data quality is more remarkable considering the stringent requirements which the MAD method places on the accuracy of the measurements.

The very short time span (less than 1 h) and acquisition of these high-quality and redundant data are desirable assets when determining protein crystal structures by the MAD method. The ability to complete a MAD experiment at these speeds circumvents a number of potential problems which can be encountered when using synchrotron radiation. Problems with beam instability such as beam dumps and energy drifts (Evans & Pettifer, 1996; Weis *et al.*, 1991) are dramatically reduced or eliminated. From a user viewpoint, it is now feasible to collect a number of MAD experiments in one day, analyze the quality of the data collected and in many cases successfully determine the structure shortly after performing the MAD experiment. New data-collection strategies also become viable, such as the collection of highly redundant data at two or three energies (Evans & Wilson, 1999), and, in general, compromises made in the past owing to time constraints when collecting X-ray data at a synchrotron are eliminated. All these contribute to the success rate of determining a protein structure using the MAD technique. Although it has long been predicted that MAD would become the method of choice in the determination of protein structure (Hendrickson, 1991), our results again go beyond what was previously envisioned and allow rapid throughput of X-ray data as described here. In addition, with the appearance of auto-tracing programs such as *wARP* (Perrakis *et al.*, 1997; Van Asselt *et al.*, 1998), model building is vastly accelerated when data is available to a resolution of 2.0 Å or better. Therefore, providing the crystal quality permits, an ideal MAD experiment would entail collection of accurate MAD phases at medium resolution ($\leq$3.0 Å), followed by the collection of high-resolution data (>2.0 Å) from the same crystal or from an additional crystal after collection of the MAD data. Extension of the MAD phases against these high-resolution data using programs such as *DM* provides a far
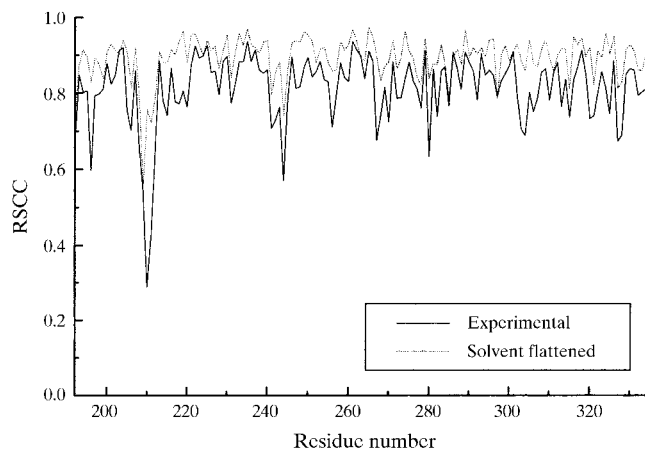

Figure 3
Residue-based real-space map correlation coefficient (RSCC; Brändén & Jones, 1990) for the experimental and *DM* solvent-flattened electron-density maps for the AD of *T. thermophilus* chaperonin at 2.25 Å resolution.

easier route to initial high-resolution phases free of model bias than performing a MAD experiment at high resolution.

At the time of writing, several additional ultrafast MAD experiments have been successfully completed at the 19ID SBC beamline using the above protocol. These experiments were far more demanding projects, being carried out on proteins with higher molecular weights, and hence requiring a higher number of heavy-atom sites in order to obtain a reasonable anomalous signal. The largest structure successfully tackled thus far has a molecular weight exceeding 170 kDa in the asymmetric unit, with 40 potential selenium sites (M. A. Walsh *et al.*, in preparation).

In summary, these successful results are made possible by the use of the 3 × 3 mosaic CCD detector (Westbrook & Naday, 1997) with its large area, fast readout time and accuracy, in combination with beamline optics which produce X-rays of high brilliance and collimation. Taken together with the current crystallographic software suites available for data analyses, these developments allow a crystallographer to perform a MAD experiment and obtain initial phase information in less than 1 d. This data, in addition to other experiments carried out to date at the SBC 19ID undulator beamline, validates the rapid approach to protein MAD structure determination. This method will complement the advances being made in molecular biology and will benefit genome-characterization efforts in progress.

References

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

Brändén, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

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

Carter, C. W. Jr (1992). *Crystallization of Proteins and Nucleic Acids, A Practical Approach*, edited by R. Giegé & A. Ducruix, pp. 47–71. Oxford: IRL Press/Oxford University Press.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Collins, F. & Galas, D. (1993). *Science*, **262**, 43–46.

Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.

Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.

Evans, G. (1998). *CHOOCH Download Page.* http://lagrange.mrc-lmb.cam.ac.uk/doc/gwyndaf/Chooch.html.

Evans, G. & Pettifer, R. F. (1996). *Rev. Sci. Instrum.* **67**, 3428–3433.

Evans, G. & Wilson, K. S. (1999). *Acta Cryst.* D**55**, 67–76.

Garman, E. F. & Schneider, T. R. (1997). *J. Appl. Cryst.* **30**, 211–237.

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

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

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Ke, H. (1997). *Methods Enzymol.* **276**, 448–461.

Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.

Phillips, J. C. & Hodgson, K. O. (1980). *Acta Cryst.* A**36**, 856–864.

Rodgers, D. W. (1997). *Methods Enzymol.* **276**, 183–203.

Rowen, L., Mahairas, G. & Hood, L. (1997). *Science*, **278**, 605–607.

Smith, J. L., Thompson, A. & Ogata, C. M. (1996). *Synchrotron Rad. News*, **9**, 12–14.

Van Asselt, E. J., Perrakis, A., Kalk, K. H., Lamzin, V. S. & Dijkstra, B. W. (1998). *Acta Cryst.* D**54**, 58–73.

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

Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. (1991). *Science*, **254**, 1608–1615.

Westbrook, E. M. & Naday, I. (1997). *Methods Enzymol.* **276**, 233–243.