THE RIGAKU JOURNAL
VOL. 23 / 2006, 3–12

Invited Papers

HIGH THROUGHPUT *DE NOVO* STRUCTURE DETERMINATION ON A HOME SOURCE USING QUICK SOAKS, ACTOR AND PARAMETER SPACE SCREENING

JOHN P. ROSE[a], ZHI-JIE LIU[a], LIRONG CHEN[a], DOOWON LEE, WOLFRAM TEMPEL, M. GARY NEWTON AND BI-CHENG WANG

Southeast Collaboratory for Structural Genomics, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, U.S.A.

Protein structure determination has benefited significantly from the focus on high throughput methods required by the NIH Protein Structure Initiative (PSI). As part of our efforts related to PSI activities, we have established a state-of-the-art X-ray diffraction facility at the University of Georgia (UGA) with a design goal of providing X-ray characterization (cell constants and diffraction limit) for at least 100 protein crystals per day. In addition, a novel parameter space screening approach for optimizing the phasing and structure determination process has been developed at UGA, which dramatically speeds up the structure determination process. Together these components form an extremely efficient and versatile structure determination pipeline.

The system described above has been used for *de novo* structure determination associated with ongoing structural genomics activities, including the determination of three structures in four days from crystals briefly soaked in gold and mercury solutions.

1. Introduction

From gene expression to the final structure, the determination of a protein structure is a lengthy and costly process. The Southeast Collaboratory for Structural Genomics (SECSG) [1] similar to other PSI-1 centers [2] has developed high throughput pipelines [3–6] for all aspects of the gene-to-structure process that have significantly reduced the time and cost associated with protein structure determination [7].

A key component of the SECSG pipeline is the Rigaku ACTOR (Automated Crystal Transfer, Orientation and Retrieval system) that allows for unattended crystal characterization (mosaicity, diffraction limit and unit cell parameters) of, and/or data collection on SECSG targets. The ACTOR, a refined version of the crystal mounter originally developed by Abbot Laboratories [8], is the first commercially available turnkey system for unattended crystal characterization and data collection. Since its introduction in 2002, both industrial and academic laboratories, including several synchrotron beamlines at the Advanced Photon Source (USA), SRS Daresbury (UK), Diamond Light Source (UK) and SOLEIL (France) have adopted the system.

There are several advantages provided by an efficient crystal mounting system capable of unattended operation, the most important being the ability to screen and characterize a large pool of crystals to find those of exceptional diffraction quality. Another important application is in drug discovery [9] where high throughput data collection on potential drug candidates complexed with their target protein is required. Finally, there is intense interest in the synchrotron community in using robotics and automation to increase experimental throughput by recovering beam time currently lost to manual crystal mounting and for remote data collection in conjunction with FedEx programs [10].

The ACTOR system as implemented at SECSG can characterize approximately 108 crystals per day in a normal 40 hr workweek and has proved to be invaluable for (1) evaluating hits from the initial crystallization trials, (2) evaluat-

[a] These authors contributed equally to the work.

Fig. 1. A photograph of the ACTOR/Saturn92 system at UGA used in the experiments described. Briefly, the system consists of an FR-D Cu rotating anode, Osmic CMF15-50Cu8 (MaxScreen) optic, AFC9 goniometer, ACTOR automounter and Saturn92 detector.

ing results from the crystal optimization process and (3) for determination of optimal cryoprotectant concentrations [11]. This reduces the time and effort required for data collection, both in-house and at the synchrotron, since data will be collected only on those crystals that have the 'best' chance of producing a structure.

In addition to crystal characterization, the SECSG ACTOR system has been used for SAD (single wavelength anomalous diffraction) structure determination from crystals briefly soaked in solutions containing millimolar amounts of bromine, gold, iodine, mercury and platinum compounds [12, 13]. In one high throughput study, three *de novo* structures were determined in four days [14] using SAD data collected using the ACTOR screening system and the SECSG SCA2Structure pipeline [4]. These experiments will be described in detail below.

2. Experimental Setup

The SECSG ACTOR system installed at the University of Georgia (Fig. 1) consists of a Rigaku FR-D rotating anode generator (0.15 mm cathode running at 50 kV and 80 mA), MaxScreen (CMF 15-50Cu8) confocal optics, AFC9 (quarter χ) goniometer, automated *xyz* goniometer head (Oceaneering Space Systems), Saturn92 CCD detector, X-Stream 2000 cryocooler and a six axis ACTOR sample mounter similar to the instrument installed at IMCA-CAT, Sector 17 APS. All screening/data collection operations were carried out using the CrystalClear/Director (Rigaku) program suite. The resulting data sets were indexed, integrated and scaled using HKL2000 (HKLResearch).

Structure determination was carried out using parameter space screening within the SECSG SCA2Structure structure determination pipeline [4]. The SCA2Structure pipeline (Fig. 2) utilizes a BioPerl [15] workflow management system and combination of crystallographic programs (SHELXD [16], SOLVE/RESOLVE [17], ISAS [18], DM [19], SOLOMON [20], ARP/wARP [21], and REFMAC [22]) to fine-tune the computational aspects of the structure determination process. SCA2structure spawns hundreds of jobs that run in parallel on the SECSG 128 node Linux cluster, with each job having a slightly different set of program input parameters (e.g. SOLVE resolution, RESOLVE resolution, number of sites, space group, etc). Upon completion of the SCA2Structure run a set of Web-based tools parses out the key data items (e.g. resolution used by SOLVE/RESOLVE, number of residues fitted, Z score etc.) from the hundreds of log files produced, ranks the results and presents them to the user in tabular form for review.

3. Diffraction Characterization

An important milestone in the structure deter-

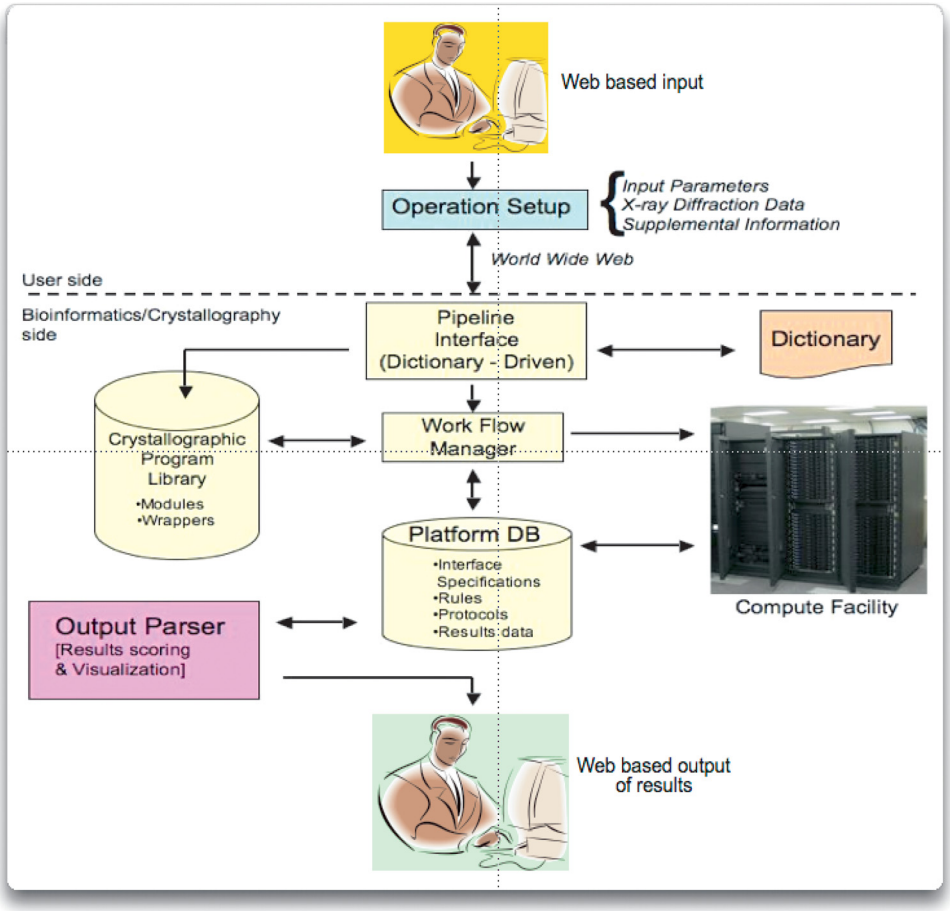


Fig. 2. A schematic diagram showing the components of the SCA2Structure pipeline. Input to the SCA2Structure pipeline is made via a dictionary generated Web input form. A Bioperl-based (http://www.bioperl.org) workflow manager manages the job creation/submission process using user supplied data and a library of crystallographic programs. The jobs then run in parallel on a multi-node Linux cluster. Upon completion of the run, a set of Web-based tools parses out the key data items needed by the crystallographer and presents them in tabular form that can be easily sorted or filtered via a web interface. For more details see Liu, et al., 2005 [4].

mination process is the production of diffraction quality crystals (crystals with diffraction that meets or exceeds the minimum resolution required by the experiment). In general, crystallization conditions are optimized to push the diffraction resolution as high as possible. This optimization process requires real-time characterization of crystals for their diffraction quality to be productive. In this area ACTOR excels. At SECSG, the ACTOR/DIRECTOR system described above routinely screens between 300–400 crystals during the week prior to a synchrotron data collection trip to identify the "best" crystal candidates for each target under study. For screening, the crystals are harvested using a loop of appropriate size [23], and flash frozen in LN_2 as they are loaded into the ACTOR magazines. The magazines (holding 12 crystals) are then loaded into the ACTOR Dewar for screening. During the screening process, the crystal under study is automatically retrieved from the Dewar, transferred to the goniometer and centered using a loop as guide. Two images are recorded at ω=0° and 90°. The crystal is then indexed based on the two images and returned to the Dewar. The cycle time for the entire mounting, centering, recording and indexing process is, on average, 10 minutes. Although recipes within DIRECTOR can be used to rank crystals, SECSG researchers prefer to visually inspect both the images and the indexing logs to arrive at a ranking.

ACTOR prescreening of crystals at SECSG has been instrumental in increasing the productivity of synchrotron trips where the time generally used to screen crystals in the past can be used for more productive experiments. For example, using prescreened crystals and the SCA2Structure pipeline, SECSG researchers were able to solve five *de novo* crystal structures on-site at

SER-CAT (Sector 22, Advanced Photon Source, Argonne National Laboratory) during one 24-hour data collection run [24] (details will be published elsewhere).

4. Cryoprotectant Optimization

Flash cooling techniques are widely used by crystallographers to the reduce radiation damage to proteins during X-ray data collection. The flash cooling technique usually involves the use of cryoprotectants, such as glycerol, to suppress ice formation during the flash cooling process. Studies by Garman [25] and others have shown that the optimization of cryoprotectant concentration and solution osmolarity matching can lead to better data quality in terms of mosaicity and resolution resulting in better structures. Again, ACTOR is ideally suited for automated cryoprotectant optimization; SECSG researchers have developed a semi-automated screening procedure for optimizing cryoprotectant concentration using the ACTOR system described above.

In this application, a series of cryoprotectant solutions are made with the cryoprotectant concentration ranging from 5% to 40% in 5% increments, using precipitant solution taken from the well that produced the crystal under study. A sample of each cryoprotectant solution is then harvested using the cryoloop technique and flash frozen in LN_2 as it is loaded into the ACTOR magazine. Once the series has been loaded, the cryoprotectant solutions are then screened for their diffraction characteristics (presence of ice rings and overall background in the image) using ACTOR (see Fig. 3). The optimum cryoprotectant mixture is judged to be the one that produced an image having no ice rings and the lowest background variation across the image [25]. Once the best conditions are determined from the initial screen, the conditions can be fine tuned as illustrated in Fig. 3.

Using this approach, UGA researchers have developed a cryoprotectant database against the eight commercial crystallization screens currently in use at SECSG [11] a paper describing the details of this work is in preparation.

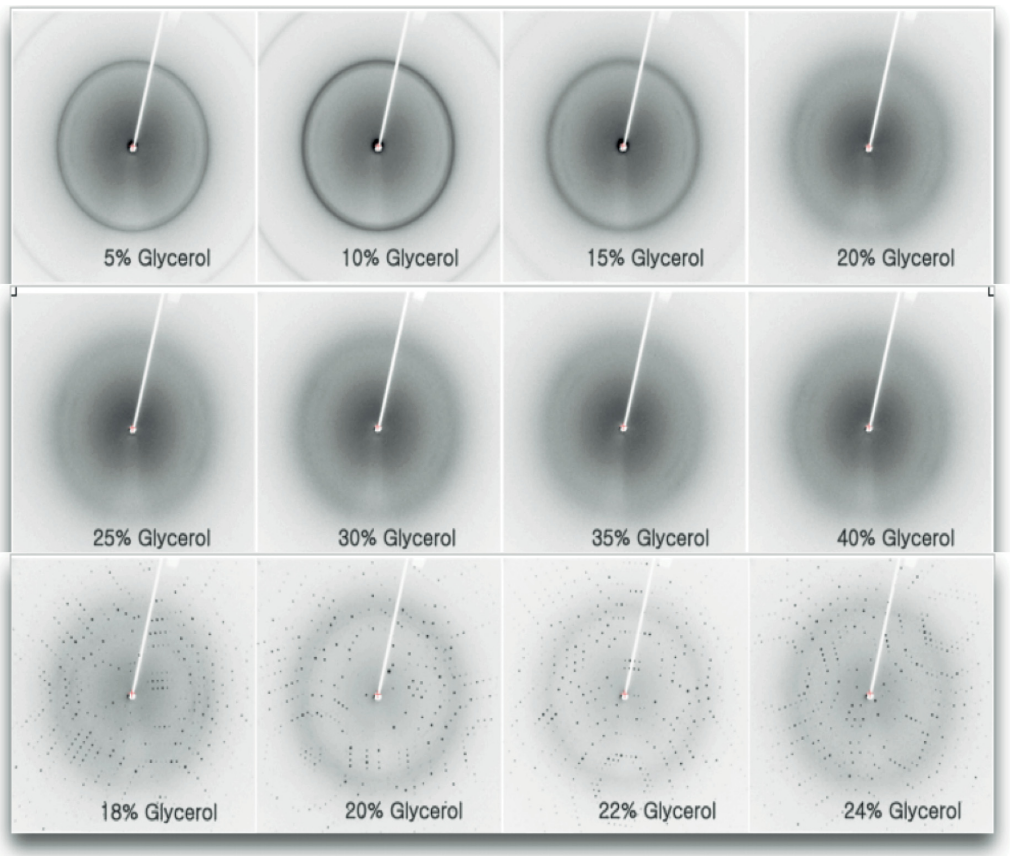


Fig. 3. A series of images collected during a cryoprotectant optimization experiment using glycerol concentrations ranging from 5% to 40%. Analysis of the average background across the image indicated that a glycerol concentration between 20–25% was optimal. The bottom row shows a fine screen within this range.

5. High throughput Structure Determination

Although well suited for rapid crystal screening, ACTOR can also serve as a high throughput structure determination platform both for industry [9] and academics [14]. To test the throughput of an automated in-house data collection system, SECSG researchers have used the instrument described above to carry out high throughput SAD structure determinations on a number of structural genomics targets. These tests include the determination (an easily interpretable electron density map with over 50% of the sequence fitted) of three (Cth-393, Cth-899 and Pfu-403030) *de novo* structures over a four-day period, which are described in more detail below.

Cth-393 – *Clostridium thermocellum* ORF 393 encodes a 12.7 kD protein initially annotated as a HIT family hydrolase. Cth-393 was cloned, expressed and purified using standard SECSG protocols [6, 26]. Crystals of Cth-393 were grown at 291 K by vapor diffusion (2 μL sitting drops) by mixing 1 μL of protein solution (protein concentration $\sim$10 mg/mL) with 1 μL of precipitant solution consisting of 100 mM Tris/HCl buffer pH 8.5 containing 500 mM potassium hydrogen phosphate and 100 mM ammonium hydrogen phosphate. For data collection, crystals were derivatized by the addition of one grain of $KAu(CN)_2$ to the crystallization drops and incubating the mixture for 30 minutes. The derivatized crystals were then harvested, cryoprotected (30% v/v glycerol in the precipitant solution), flash cooled in LN_2 and loaded into the ACTOR magazine. The crystals were then screened manually and ranked based on their diffraction characteristics. Data to 2.8 Å (see Fig. 4a) was then collected overnight on a crystal measuring 80$\times$80$\times$150 microns, which was deemed best from the initial X-ray characterization. The details of the data collection are given in Table 1.

Although the data was of marginal quality (Rsym=16.0%), the resulting structure factors and sequence were input to the SCA2Structure pipeline. Based on the 4 gold sites identified by SOLVE, RESOLVE was able to fit 131 of the 228 residues (57.4%) present in the asymmetric unit (see Fig. 4b). The initial RESOLVE model was then completed (XFIT [27]), refined (REFMAC [22] and validated (Procheck [28] and MolProbity [29]) in under a week (see Fig. 4c). A higher resolution 2.3 Å data set was later collected at SER-CAT and used for the final refinement. A manuscript describing the structure in more detail is in preparation.

Table 1. Experimental parameters for the three example datasets.

Protein	**Cth-393**	**Cth-833**	**Pfu-403030**
Figure	4	5	6
Molecular weight (kD)	12.7	13.4	20.2
Number of residues	114	114	176
Anomalous scatterers	Au	Hg	Hg
Crystal			
Space Group	$P4_122$	$P2_1$	$P2_1$
a (Å)	78.8	45.48	78.97
b (Å)	78.8	52.16	106.71
c (Å)	114.9	51.08	84.81
β (°)		92.4	117.5
Molecules per asymmetric unit	2	2	3
Data collection			
Step size (°)	0.5	0.5	0.33
Number of images	360	680	1200
Scan Range (°)	180	340	360
Mosaic spread (°)	0.69	0.68	1.02
Exposure time (sec)	20	15	10
Data reduction			
Data resolution (Å)	2.80	2.28	2.69
Completeness of data	99.8	88.7	96.5
Redundancy	13.9	6.5	6.2
$\langle I/\sigma_I \rangle$	20.7	31.59	23.48
Rsym (%)	16.0	5.7	8.5
Structure solution			
Number of sites	4	3	6
Phasing resolution	2.3	2.6	3.2
Modification resolution	2.8	2.4	2.8
Total time (min)	110	150	95
Refinement			
Resolution (Å)	2.3*	2.3	2.26*
R-factor	27.9	20.9	20.1
Rfree	32.6	26.7	24.6
R.M.S.D Bond lengths (Å)	0.017	0.015	0.018
R.M.S.D Bond angles (°)	1.329	1.200	1.461
PDB ID	1XQU	1XMA	1XX7

*Reported structure refined from higher resolution data collected at SER-CAT, Advanced Photon Source, Argonne National Laboratory.

Cth-833 – *Clostridium thermocellum* ORF 833 encodes a 13.4 kD protein initially annotated as a transcriptional regulator. Cth-833 was cloned, expressed and purified as described above. Crystals of Cth-833 were grown at 291 K by the microbatch under oil (Al's Oil) technique using 1 μL drops containing equal volumes of protein solution (protein concentration $\sim$10 mg/mL) and a precipitant solution containing 8.5% v/v isopropanol, 10% w/v PEG 4000, 15% v/v glycerol in 0.085M Tris/HCL pH 8.5. For data collection, crystals were derivatized by the addition of one grain of mersalyl acid (o-[(3-Hydroxymercuri-2-methoxypropyl)-carbamoyl]phenoxyacetic acid) to the crystallization drops and incubating the mixture for 2 hours. The derivatized crystals were then harvested, flash cooled in LN_2 (no cryoprotection was necessary), loaded into the ACTOR magazine and screened for their diffraction quality as described above. A data set to 2.3 Å resolution (see Fig. 5a) was then collected overnight on the best crystal from the screening

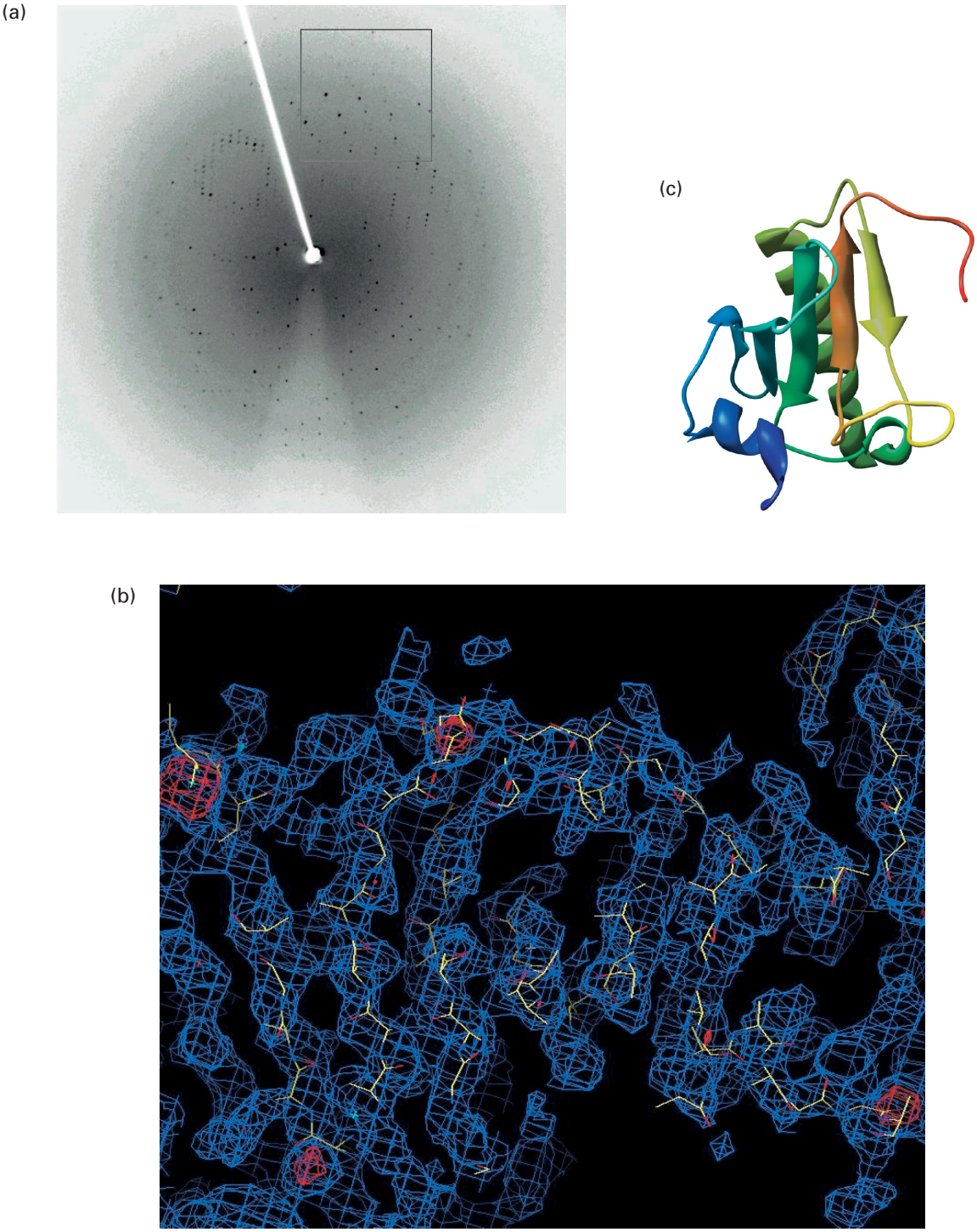


Fig. 4. (a) A representative diffraction image from the Cth-393 data collection showing the marginal quality of the diffraction pattern. (b) A section of the 2.8 Å experimental ISAS-phased electron density map for Cth-393 contoured at 1α with the refined coordinates for the structure superimposed and (c) a ribbon drawing of the Cth-393 structure (PDB entry 1XQU) colored (blue to red) by sequence number.

that measured 100×100×100 microns. The details of the data collection are given in Table 1.

Based on the 4 mercury sites identified (SOLVE) during the SCA2Structure run, RESOLVE was able to fit 186 of the 228 residues (81.8%) present in the asymmetric unit (see Fig. 5b). The initial RESOLVE model was then quickly completed, refined and validated to give the final model (see Fig. 5c) as described previously. A manuscript describing the structure in

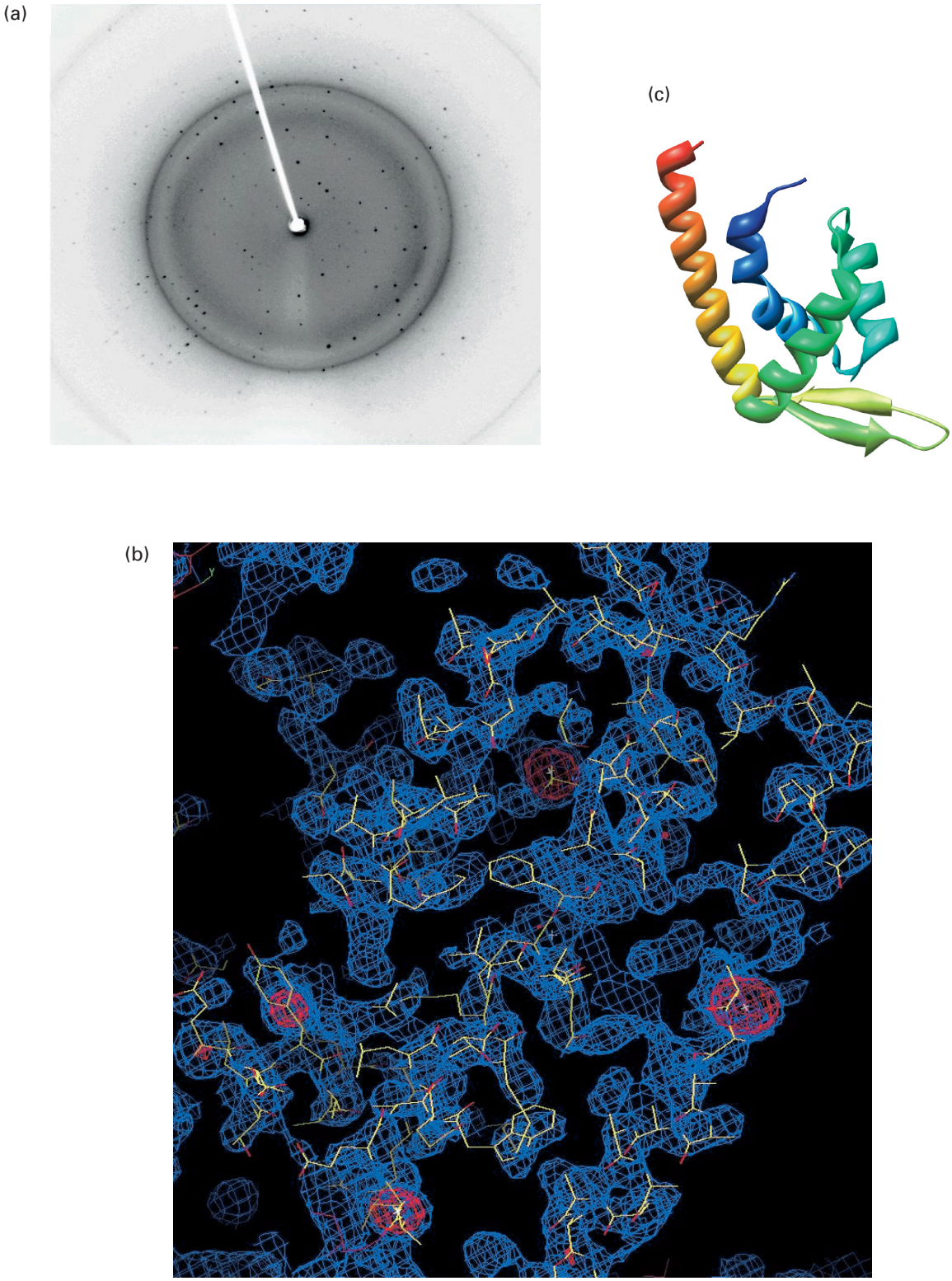
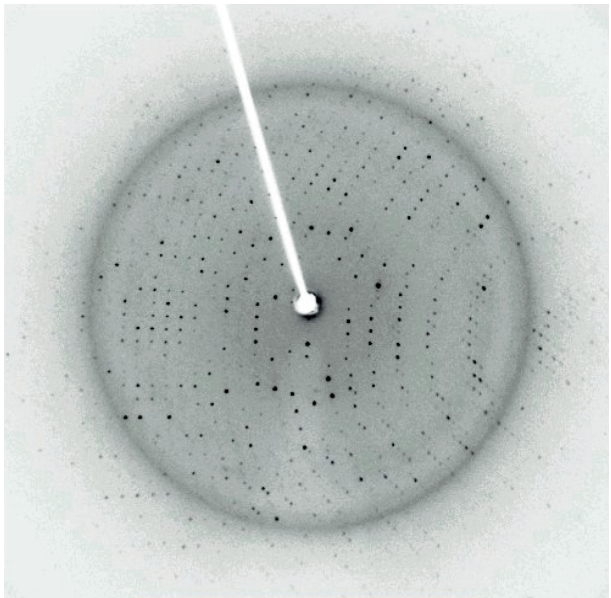
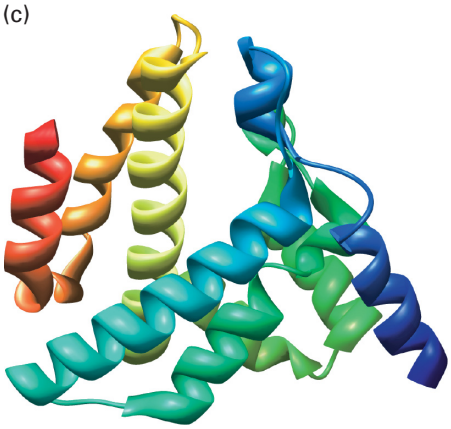


Fig. 5. (a) A representative diffraction image from the Cth-833 data collection showing the quality of the diffraction pattern. (b) A section of the 2.4 Å experimental RESOLVE-phased electron density map for Cth-833 contoured at 1α with the refined coordinates for the structure superimposed and (c) a ribbon drawing of the refined Cth-833 structure (PDB entry 1XMA) colored (blue to red) by sequence number.

(a)



(c)



(b)

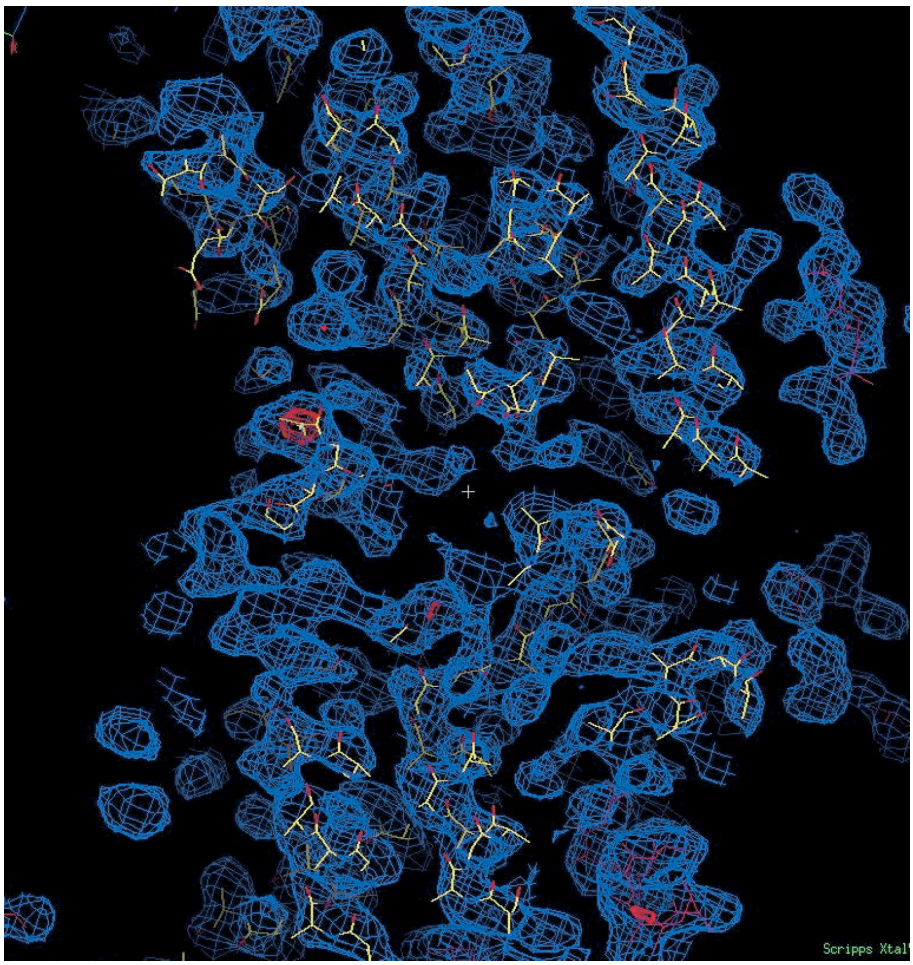


Fig. 6. (a) A representative diffraction image from the Pfu-403030 data collection showing the quality of the diffraction pattern. (b) A section of the 2.8 Å experimental RESOLVE-phased electron density map for Pfu-403030 contoured at 1 σ with the refined coordinates for the structure superimposed and (c) a ribbon drawing of the refined Pfu-403030 structure (PDB entry 1XX7) colored (blue to red) by sequence number.

more detail is in preparation.

Pfu-403030 – *Pyrococcus furiosus* ORF 403030 encodes a 20.2 kD protein which is annotated as a conserved hypothetical protein. Pfu-403030 was cloned, expressed and purified as described above. Crystals of Pfu-403030 were grown at 291 K by the vapor diffusion method using 2 μL sitting drops made by mixing 1 μL of protein solution (protein concentration $\sim$10 mg/mL) with 1 μL of precipitant solution consisting of 30% v/v PEG 2000 MME pH 4.6 containing 0.2 M ammonium sulfate and 0.1 M ammonium acetate. For data collection, crystals were derivatized by the addition of one grain of mersalyl acid to the crystallization drops and incubating the mixture for 3 hours. The derivatized crystals were then harvested, flash cooled in LN_2 (no cryoprotection was necessary) and loaded into the ACTOR magazine and screened for their diffraction quality as described above. A data set to 2.69 Å resolution (see Fig. 6a) was then collected overnight on the best crystal found during screening that measured 150$\times$150$\times$150 microns, see Table 1).

Using the SCA2Structure pipeline 6 mercury sites were identified by SOLVE and based on these sites RESOLVE was able to fit 498 of the 528 residues (94.3%) present in the asymmetric unit (see Fig. 6b). This model was then quickly completed, refined and validated (see Fig. 6c) as described above. Data were later collected to higher resolution at SER-CAT resulting in the reported 2.26 Å structure. A manuscript describing the structure in more detail is in preparation.

6. Conclusions

The ACTOR system at UGA has proven to be a valuable tool for high throughput crystal structure determination. In its role in crystal characterization, it has screened thousands of crystals over the last three years resulting in considerable savings in synchrotron beam time since only those crystals that cannot be characterized in-house are screened at the synchrotron. In addition, although beamline automation is becoming more common, which makes efficient screening at the beamline a real possibility, the disconnect between harvesting crystals at home, getting them to the beamline, and waiting your turn in the queue makes this approach less attractive WHEN compared to a dedicated in-house system, such as the ACTOR. The same can be said for screening for heavy atoms/anomalous scatterer incorporation since these activities generally require real-time feedback to be successful.

As described above, the ACTOR has also proven to be a useful tool for the optimization of cryoprotectant conditions, which can lead to better quality data. The ability to easily setup the cryoprotectant screen and have ACTOR automatically carry out the data collection makes this approach very attractive not only for high throughput studies but also for screening and optimizing cryoprotectants for those crystals that are sensitive to cryoprotectant choice and/or cryoprotectant concentration.

We have also shown that the ACTOR system can be used for high throughput *de novo* structure determination using SAD data when coupled with a high performance structure determination engine such as SCA2Structure. The throughput in the study reported here was limited by the fact that the manual ranking was used and that crystals were available from only three targets. It should also be noted that we generally separate the data collection for phasing purposes from data collection for high-resolution refinement, which is generally carried out at SER-CAT. This approach avoids making compromises that either effect signal-to-noise ratio of the phasing data or in the recording of reflections at high-resolution.

Acknowledgements

The samples used in theses studies were provided by the SECSG protein production and crystallomics groups. Work supported in part from funds from the National Institute of General Medical Sciences (GM62407), The Georgia Research Alliance and The University of Georgia Research Foundation.

References

[1] M. W. Adams, H. A. Dailey, L. J. DeLucas, M. Luo, J. H. Prestegard, J. P. Rose and B. C. Wang: *Acc Chem Res*, **36** (2003), 191–198.

[2] J. C. Norvell and A. Z. Machalek: *Nat Struct Biol*, **7** (2000), 931.

[3] Z.-J. Liu, W. Tempel, J. D. Ng, D. Lin, A. K. Shah, L. Chen, P. S. Horanyi, J. E. Habel, I. A. Kataeva, H. Xu, H. Yang, J. C. Chang, L. Huang, S.-H. Chang, W. Zhou, D. Lee, J. L. Praissman, H. Zhang, M. G. Newton, J. P. Rose, J. S. Richardson, D. C. Richardson and B.-C. Wang: *Acta Cryst.*, **D61** (2005), 679–684.

[4] Z. J. Liu, , D. Lin, W. Tempel, J. L. Praissman, J. P. Rose and B. C. Wang: *Acta Cryst.*, **D61** (2005), 520–527.

[5] Z. J. Liu, W. Tempel, J. D. Ng, D. Lin, A. K. Shah, L. Chen, P. S. Horanyi, J. E. Habel, I. A. Kataeva, H. Xu, H. Yang, J. C. Chang, L. Huang, S. H. Chang, W. Zhou, D. Lee, J. L. Praissman, H. Zhang, M. G. Newton, J. P. Rose, J. S. Richardson, D. C. Richardson, and B. C. Wang: *Acta Cryst.*, **D61** (2005), 679–684.

[6] B. C. Wang, M. W. Adams, H. Dailey, L. DeLucas, M. Luo, J. Rose, R. Bunzel, T. Dailey, J. Habel, P. Horanyi, F. E. Jenney, Jr., I. Kataeva, H. S. Lee, S. Li, T. Li, D. Lin, Z. J. Liu, C. H. Luan, M. Mayer, L. Nagy, M. G. Newton, J. Ng, F. L. Poole, 2nd, A. Shah, C. Shah, F. J. Sugar and H. Xu: *J Struct Funct Genomics*, **6** (2005), 233–243.

[7] J. M. Chandonia and S. E. Brenner: *Science,*. **311** (2006), 347–351.

[8] S. W. Muchmore, J. Olson, R. Jones, J. Pan, M. Blum, J. Greer, S. M. Merrick, P. Magdalinos and V. L. Nienaber: *Structure Fold Des*, **8** (2000), R243–246.

[9] A. J. Sharff: *Rigaku J.*, **19** (2003), 5–10.

[10] J. P. Rose, J. Chrzas, Z. Jin, J. Fait, V. Babson and B. C. Wang: The SER-CAT remote user participation program. in Annual meeting of the American Crystallographic Association. 2006. Honolulu, HI.

[11] D. Lee, J. P. Rose, Z. J. Liu, M. G. Newton and B. C. Wang: P190, Rapid optimization of cryoprotectant concentration using a robotic sample mounting system, in American Crystallographic Association 2005, American Crystallographic Association Orlando, FL.

[12] P. D. Sun, and S. Radaev: *Acta Cryst.*, **D58** (2002), 1099–1103.

[13] Z. Dauter, M. Dauter and K. R. Rajashankar: *Acta Cryst.* **D56** (2000) 232–7.

[14] L. Chen, Z.-J. Liu, D. Lee, W. Zhou, W. Tempel, J. P. Rose and B. C. Wang: SAS Phasing Using Home Source and Synchrotron X-rays. in 2nd Annual SER-CAT Symposium. 2005. St. Judes Childrens Research Hospital, Memphis, TN.

[15] J. E. Stajich, D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz, C. Dagdigian, G. Fuellen, J. G. Gilbert, I. Korf, H. Lapp, H. Lehvaslaiho, C. Matsalla, C. J. Mungall, B. I. Osborne, M. R. Pocock, P. Schattner, M. Senger, L. D. Stein, E. Stupka, M. D. Wilkinson and E. Birney: *Genome Res*, **12** (2002), 1611–1618.

[16] T. R. Schneider and G. M. Sheldrick: *Acta Cryst.*, **D58** (2002), 1772–1779.

[17] T. C. Terwilliger: *Acta Cryst.*, **D58** (2002), 1937–1940.

[18] B. C. Wang: *Methods Enzymol*, **115** (1985), 90–112.

[19] K. Cowtan: *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31** (1994), 34–38.

[20] S. Bailey: *Acta Cryst.*, **D50** (1994), 760–763.

[21] A. Perrakis, M. Harkiolaki, K. S. Wilson and V. S. Lamzin: *Acta Cryst.*, **D57** (2001), 1445–1450.

[22] G. N. Murshudov: *Acta Cryst.*, **D53** (1997), 240-255.

[23] T. Y. Teng: *J. Appl. Cryst.*, **23** (1990), 387–391.

[24] J. P. Rose, Z.-J. Liu and B. C. Wang: MS23.25.3 HT structure determination at SER-CAT: five Structures in 23 hours. in XX Congress of the International Union of Crystallography. 2005. Florence, Italy: International Union of Crystallography.

[25] E. F. Garman and S. Doublie: *Methods Enzymol*, **368** (2003), 188–216.

[26] Z. J. Liu, , A. K. Shah, J. E. Habel, J. D. Ng, I. Kataeva, H. Xu, P. Horanyi, H. Yang, J. Chang, M. Zhao, L. Huang, S. Chang, W. Tempel, L. Chen, W. Zhou, D. Lee, D. Lin, H. Zhang, M. G. Newton, J. Rose and B. C. Wang: *J Struct Funct Genomics*, **6** (2005), 121–127.

[27] D. E. McRee: *J Struct Biol*, **125** (1999), 156–165.

[28] B. A. Laskowski, M. W. MacArther, D. S. Moss and J. M. Thornton: *J. Appl. Cryst.*, **26** (1993), 283–291.

[29] I. W. Davis, L. W. Murray, J. S. Richardson and D. C. Richardson: *Nucleic Acids Res*, **32** (2004), W615–619.