Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

HKL-3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes

Wladek Minor,[a]* Marcin Cymborowski,[a] Zbyszek Otwinowski[b] and Maksymilian Chruszcz[a]

[a]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22903, USA, and [b]Department of Biochemistry, UT Southwestern Medical Center at Dallas, Dallas, TX 75235, USA

Correspondence e-mail: wladek@iwonka.med.virginia.edu

Received 30 March 2006
Accepted 26 May 2006

A new approach that integrates data collection, data reduction, phasing and model building significantly accelerates the process of structure determination and on average minimizes the number of data sets and synchrotron time required for structure solution. Initial testing of the *HKL*-3000 system (the beta version was named *HKL*-2000_ph) with more than 140 novel structure determinations has proven its high value for MAD/SAD experiments. The heuristics for choosing the best computational strategy at different data resolution limits of phasing signal and crystal diffraction are being optimized. The typical end result is an interpretable electron-density map with a partially built structure and, in some cases, an almost complete refined model. The current development is oriented towards very fast structure solution in order to provide feedback during the diffraction experiment. Work is also proceeding towards improving the quality of phasing calculation and model building.

1. Introduction

The determination of a large macromolecular structure is a sophisticated multi-step process that usually requires a considerable amount of time and effort. The use of anomalous scattering (Hendrickson, 1991), tunable synchrotron radiation, selenomethionyl-labeled protein expression (Hendrickson *et al.*, 1990), powerful computers and increasingly sophisticated software has revolutionized macromolecular structure determination and permitted the routine use of high-throughput techniques (Chandonia & Brenner, 2006; Todd *et al.*, 2005; Walsh *et al.*, 1999). Nevertheless, structure solution is still very challenging, as even structures coming from very successful synchrotron beamlines (Holton, 2005) on average require about 50 data sets to produce a single PDB deposit. Further improvement in protein crystallography should come from reducing the ratio of sets collected per deposit rather than increasing the amount of collected data. The tool for achieving this goal should provide highly informative feedback during data collection and processing and all further stages of data analysis.

In recent years, several systems that merge different crystallographic computer programs into a structure-determination pipeline have been developed. The most popular or promising packages include *AUTOSHARP* (de La Fortelle & Bricogne, 1997; Vonrhein *et al.*, 2006), *ACrS* (Brunzelle *et al.*, 2003), *SGXPro* (Fu *et al.*, 2005), *ELVES* (Holton & Alber, 2004), *Auto-Rickshaw* (Panjikar *et al.*, 2005), *PHENIX* (Adams *et al.*, 2004) and *HKL2MAP* (Pape & Schneider, 2004). These packages were assembled with

doi:10.1107/S0907444906019949

different goals and degrees of built-in automation. We have developed a method for semi-automatic (or in some cases automatic) analysis of X-ray diffraction data that combines a number of existing macromolecular crystallographic computer programs and decision-making algorithms into a powerful expert system called *HKL*-3000. The beta version of *HKL*-3000 has been successfully used to determine *de novo* over 140 structures ranging from 9 to 273 kDa molecular weight in the crystallographic asymmetric unit and with data resolution limits varying between 1.1 and 3.4 Å. *HKL*-3000 is unique in the sense that it integrates all steps from data reduction to model building and in some custom versions may be integrated with a synchrotron beamline data-collection control system (Minor *et al.*, 2002).

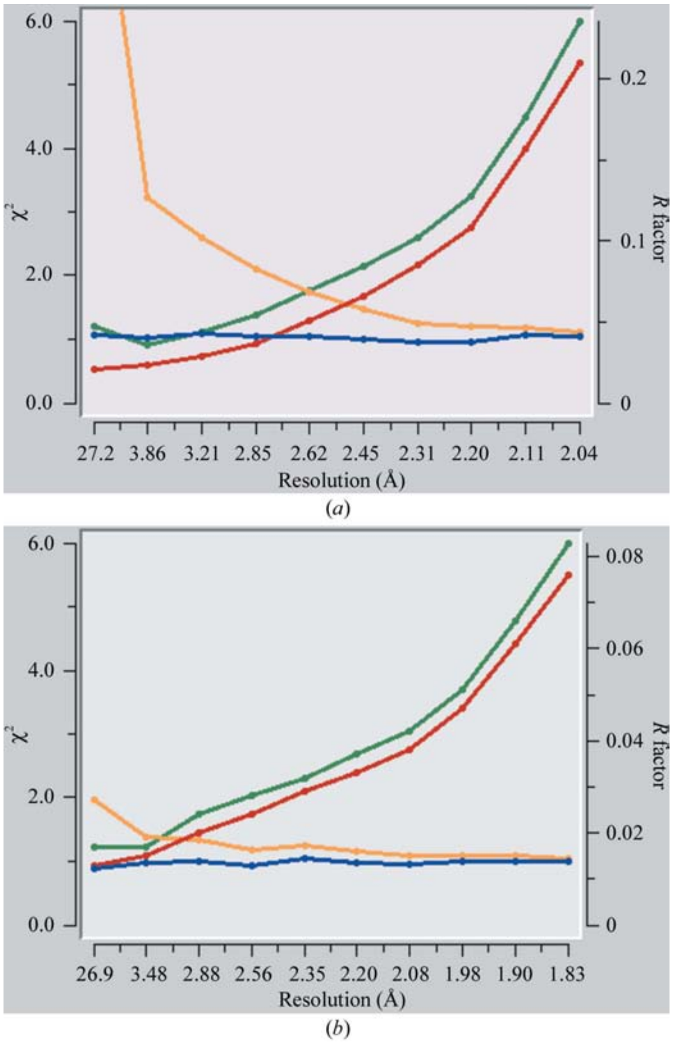


Figure 1
(*a*) The strong anomalous signal typically produced in SeMet data sets collected at the *K* absorption edge of selenium. The orange and dark blue lines correspond to χ^2 statistics with Friedel mates either merged together or considered separately, respectively. The difference between these two lines represents the significance of the anomalous signal at the level of individual observations. The comparison of the $R_{\rm merge}$ statistics for the merged or unmerged Friedel pairs, described by green and red curves, respectively, is also presented, but is less informative. (*b*) An example of a weak but significant anomalous signal produced by sulfur anomalous scattering at Cu *Kα* wavelength.

HKL-3000 is usually set up in a semi-automatic mode, but can also work in a fully automatic mode. The semi-automatic mode performs individual steps: data reduction and analysis, substructure solution, phasing and model building. For projects of known protein sequence, the system suggests the optimal input parameters for each step and provides sophisticated analysis of the outcome of each step. The analysis of results from each step is used to optimize input parameters for every subsequent step. In the case of an unsuccessful outcome for a particular step, the experimenter has the possibility to use a more sophisticated approach than that coded as a default in the system. The system has the ability to import partial solutions from external programs. Similarly, the experimenter has the ability to adjust hundreds of parameters; for example, the substructure-solution module is controlled by parameters such as the number of data sets, resolution limit, closest distance between sites, number of sites, number of cycles and treatment of special positions. Moreover, results can be sorted in several ways, one solution may be compared with others and any solution can be selected for subsequent steps. However, the authors find that these options and controls are useful more for system development than for use by experimenters, even for very difficult structures. The present package provides a complete structure solution pipeline for both SAD and MAD phasing.

2. Description of *HKL*-3000

2.1. X-ray data reduction and analysis

The first step in the process of structure determination is raw-image data reduction and analysis. This step is performed by the standard *HKL*-2000 package (Otwinowski & Minor, 1997). Preliminary experiences with pathological data resulted not only in a more sophisticated analysis (Borek *et al.*, 2003) in the scaling/merging step performed by *SCALEPACK*, but most critically an expanded repertoire of corrections during this step (Otwinowski *et al.*, 2003). The most important corrections include but are not limited to correction for

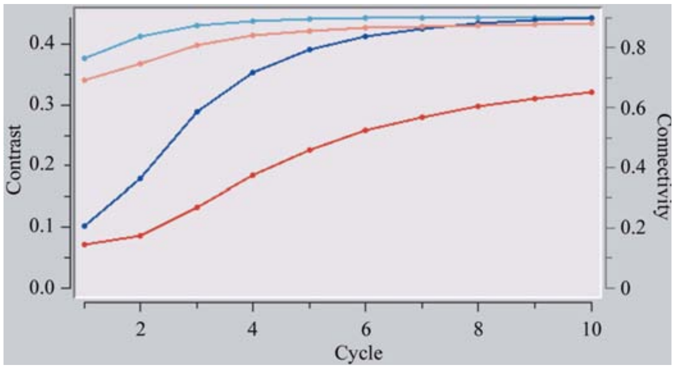
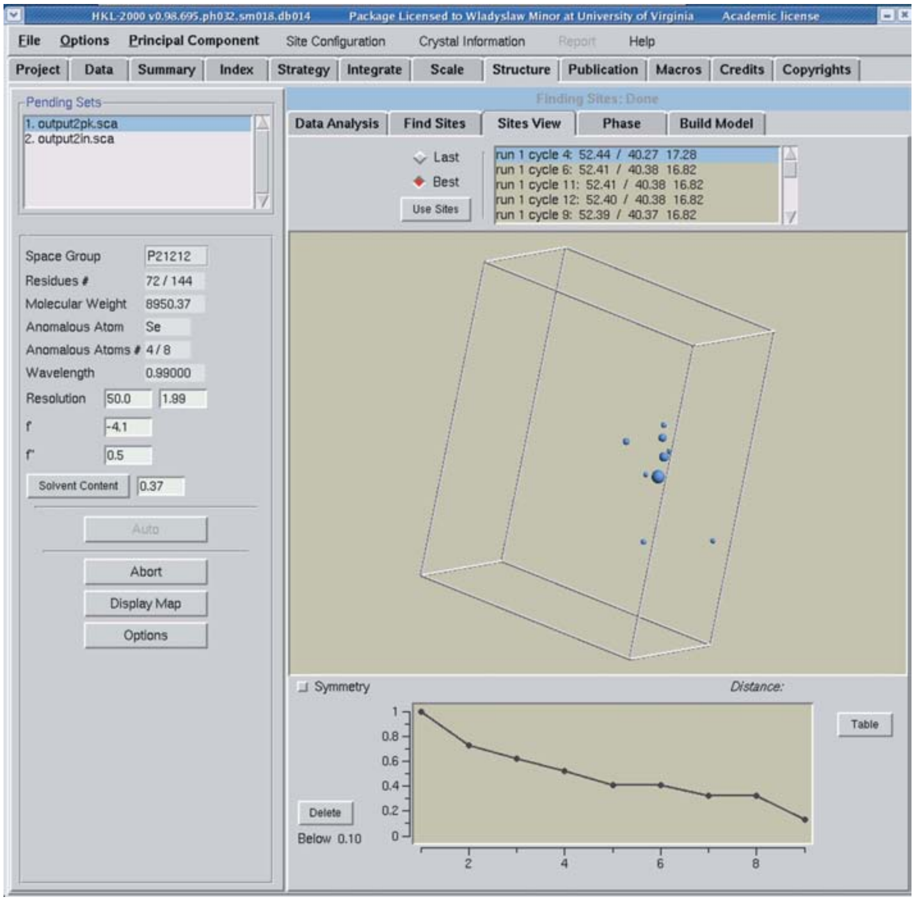


Figure 2
The map contrast and connectivity for the two possible enantiomorphs in consecutive solvent-flattening cycles in *SHELXE*. The substantial difference in map contrast between the original (dark red) and inverted (dark blue) solution indicates that the inverted substructure solution is correct. The map connectivity for the original and inverted solutions is in light red and blue colors, respectively.



HKL-2000 v0.98.695.ph032.sm018.db014 Package Licensed to Wladyslaw Minor at University of Virginia Academic license

Pending Sets
1. output2pk.sca
2. output2in.sca

Finding Sites: Done

run 1 cycle 4: 52.44 / 40.27 17.28
run 1 cycle 6: 52.41 / 40.38 16.82
run 1 cycle 11: 52.41 / 40.38 16.82
run 1 cycle 12: 52.40 / 40.38 16.82
run 1 cycle 9: 52.39 / 40.37 16.82

Space Group P21212
Residues # 72 / 144
Molecular Weight 8950.37
Anomalous Atom Se
Anomalous Atoms # 4 / 8
Wavelength 0.99000
Resolution 50.0 1.99
f' -4.1
f" 0.5
Solvent Content 0.37

□ Symmetry

Distance:

Below 0.10

Figure 3
A substructure solution visualized in a non-stereo three-dimensional widget that permits inspection and interactive manipulation of the heavy-atom positions.

absorption, spindle-axis misalignment, uneven speed of spindle-axis rotation and vibration of the cryogenic loop with the frozen crystal during data collection. A correction for crystal decay is currently being implemented. All these pathologies decrease data quality and substantially degrade the significance of the anomalous signal. In many SeMet experiments the magnitude of the anomalous signal is so high (Fig. 1*a*) that the degradation does not significantly affect substructure solution and the phasing procedure. The degradation sometimes affects automatic model-building procedures and subsequently leads to the model that requires more manual adjustments. In the case of a weak signal, such as the use of the sulfur anomalous scattering signal for phasing (Fig. 1*b*), the degradation of the anomalous signal could make substructure solution or phasing very difficult or even impossible.

Another pathology encountered frequently in many experimental data sets is the inability of the experimenter to collect complete low-resolution data. This incompleteness is most often caused by the presence of overloaded low-resolution reflections, as CCD detectors used on most synchrotron beamlines have a limited dynamic range or by an insufficiently small beamstop. The separate analysis of data completeness for low-resolution ranges gives adequate warning to the experimenter. In the case when scaling is performed during data collection, the experimenter has the opportunity to add a second, low-resolution pass with reduced exposure time and increased oscillation range per frame. Surprisingly, incomplete low-resolution data do not very strongly affect the ability to perform a substructure-solution search but significantly degrade the phasing process. The optional generation of missed reflections during the solvent-flattening process as implemented in *DM* (Cowtan, 2001; Cowtan & Main, 1998; Cowtan & Zhang, 1999) somewhat improves phases for structures with relatively high solvent content.

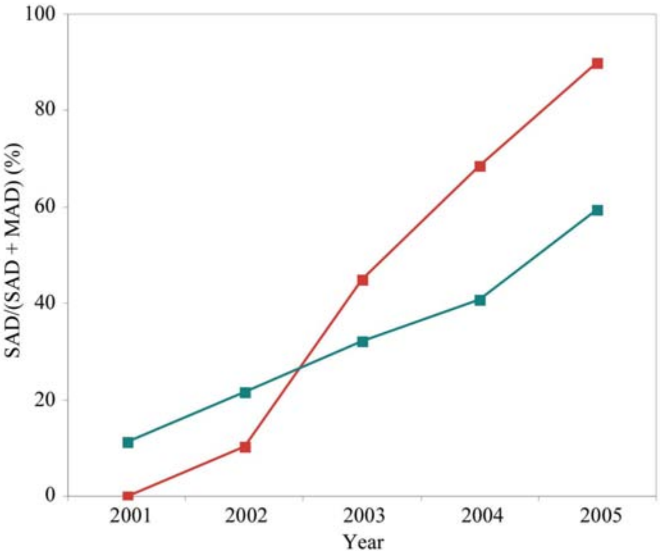


Figure 4
The fraction of structures solved by SAD techniques among all structures solved by MAD and SAD. The green line represents all structures deposited in the PDB between 2001 and 2005 and the red line represents MCSG structures only.

2.2. Substructure solution and enantiomorph elucidation

The parameters for the structure-solution routine are derived from the magnitude and resolution limit of the anomalous signal. Additional information is derived from the protein sequence and the type of atoms most significant for anomalous signal generation. The substructure solution is performed by *SHELXD* (Schneider & Sheldrick, 2002). The progress is analyzed on the fly and real-time plots displaying correlation coefficients (CC) *versus* the Patterson figure of merit (PATFOM) and the number of equivalent solutions are displayed. For any particular solution, the heavy-atom sites and the symmetry-equivalent site positions can be analyzed in an interactive three-dimensional window. The average occupancy for a particular set of sites can also be monitored. The procedure of substructure search is automatically accomplished when a pre-defined CC is obtained. To avoid sub-optimal substructure solutions, the number of trials cannot be smaller than (number of sites) × 2 + 2, even if a very high CC is obtained in the first trial of *SHELXD*. The experimenter can also interrupt a substructure search at any point or continue to search indefinitely. The heavy-atom search is followed by ten cycles of solvent flattening as implemented in *SHELXE* (Sheldrick, 2002). Two parallel runs for two possible enantiomorphs are performed and the map contrast *versus* cycle number of solvent flattening is displayed and analyzed, so that the enantiomorph is assigned automatically. The large difference in map contrast between the two enantiomorphs (Fig. 2) strongly indicates that the substructure solution is correct. In the case of a small difference, the experimenter may return to the substructure-solution procedure or perform additional analysis of the data, such as searching for the possible presence of merohedral twinning. Owing to time constraints during the synchrotron experiments, these additional analyses are performed only on request, usually when substructure solution or phasing fails. In the case of space-group determination ambiguity, the substructure solution should be performed for all possible space groups.

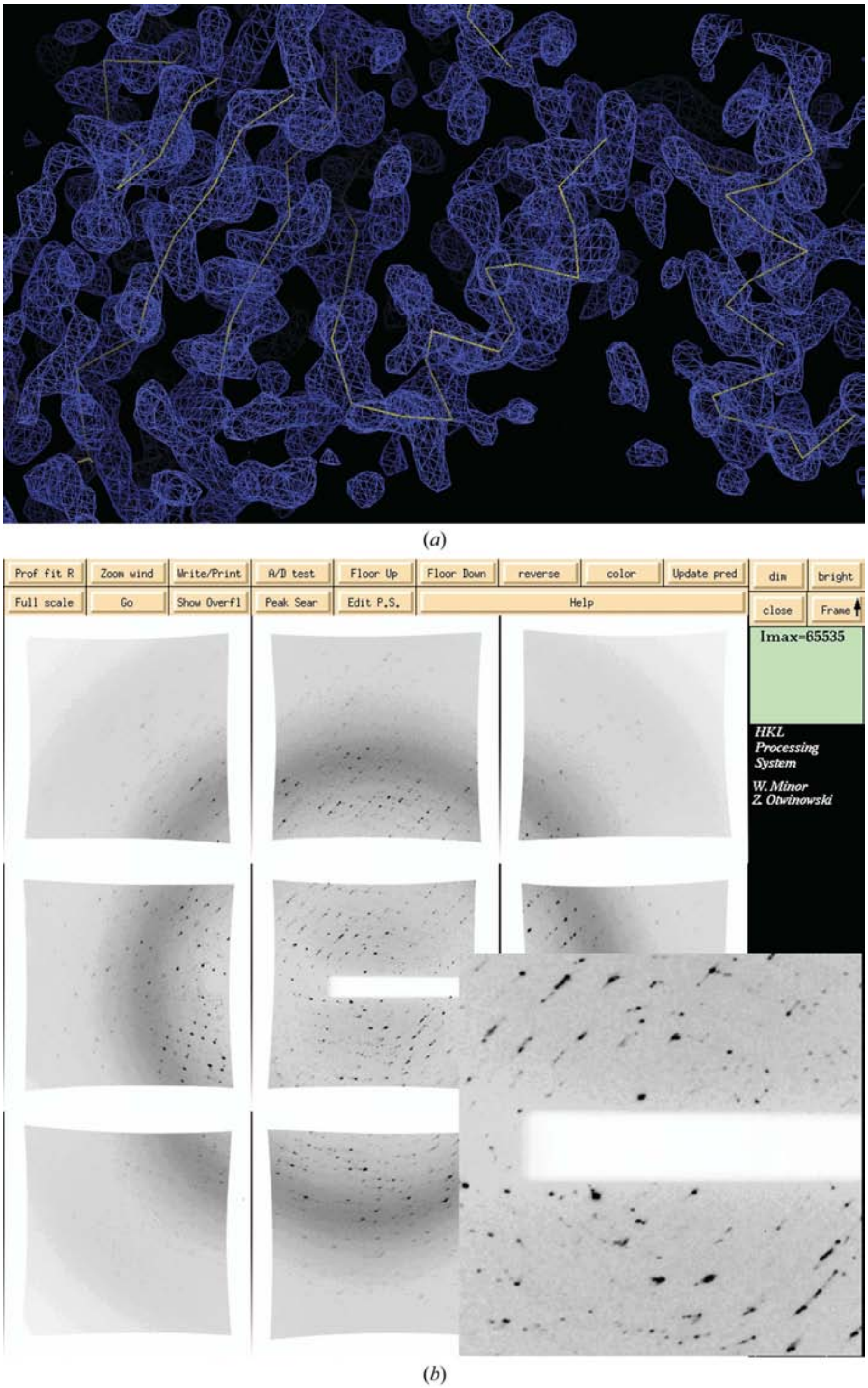


Figure 5
The experimental electron-density map (*a*) and corresponding diffraction pattern (*b*) of a large crystal with good microscopic order compensating for poor macroscopic order: high mosaicity and multiple crystals. Data correspond to PDB entry 1xvi. The initial model was built by *RESOLVE* in the fast mode. For picture clarity, only a C^{α} trace is presented. It is important to note that a poor-quality crystal was used because it was the only crystal available for structure solution.

The heavy-atom sites with low occupancy could be rejected automatically at this stage, but the preferred (default) path is to analyze them with the help of visual tools (Fig. 3). Our experience shows that in the case of SeMet experiments one can observe double conformations of SeMet side chains. The use of both conformations (Table 1) improves phasing substantially. Further analysis of heavy atoms is performed automatically during the phasing procedure, as described in the next paragraph.

2.3. Phasing and initial phase improvement

Phasing is performed by multiple successive runs of *MLPHARE* (Otwinowski, 1991) and *DM* with sophisticated automatic analysis of each run. In the first run, only positional and occupancy refinement of sites is performed. Subsequent runs refine positions and temperature factors of sites. In the final run, the anisotropic temperature factor is refined for sites with a significant ratio of occupancy to its uncertainty. A similar criterion is used for the automatic removal of weak sites during the phasing procedure. The phasing procedure has an optional ability to add new sites that appear in the anomalous difference map. This option is only performed after a visual inspection of new site positions in the difference map. The *CCP*4 set of programs is used to calculate mtz-format files, maps and some other auxiliary calculations (Collaborative Computational Project, Number 4, 1994). The maps and models are displayed by the programs *O* (Jones, 2004) or *Coot* (Emsley & Cowtan, 2004), which are called directly from *HKL*-3000.

Phase improvement and extension is performed by *DM.* During multiple runs of density modification, the solvent content is modified in order to optimize the molecular envelope. The full solvent-content optimization is not employed during this step, but rather in the model-building process. Optionally, the experimenter may try to employ non-crystallographic symmetry (NCS) if more than one molecule in the asymmetric unit is expected. At present, NCS can be established only by using heavy-atom sites and can recognize only the most frequent case of NCS described by a single rotational axis. *HKL*-3000 uses *SOLVE* for this task, but graphical analysis and constraints related to the number of molecules in the asymmetric unit are added. As the impact of NCS averaging is very high, algorithms that can identify more complex NCS cases will be implemented.

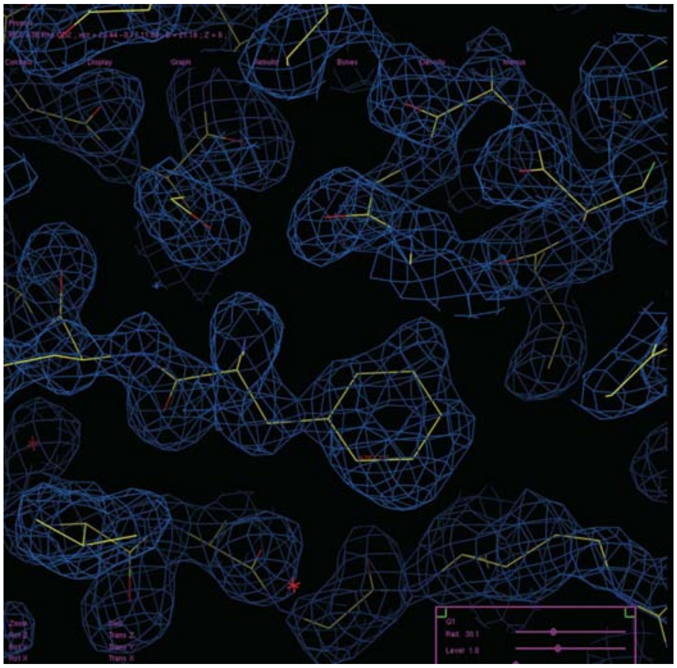


Figure 6
The 1.8 Å experimental electron-density map obtained from a SAD sulfur experiment. The data were collected at the home laboratory from a lysozyme crystal. The data processing, scaling, structure solution and model building took 14 min. 128 out of 129 residues were built in fast mode. The anomalous signal was relatively weak, as presented in Fig. 1(*b*).

Table 1
Experimental electron-density map correlation *versus* final model.

Each data set was phased with and without double sites. The correlation coefficient CC was calculated between the final refined model and the experimental density map obtained from density modification without NCS averaging.

PDB code	Total No. of sites	No. of double sites	Solvent content (fraction)	Resolution (Å)	CC for double sites (main/side chain)	CC for single sites (main/side chain)
1wq6	5	3	0.37	2.0	0.76/0.67	0.70/0.61
2g3b	4	1	0.45	2.0	0.57/0.50	0.51/0.43
2g7u	16	4	0.53	2.3	0.81/0.64	0.80/0.63

The computational time of phasing and initial phase improvement depends very strongly on the number of sites. For 1–8 sites, the time between substructure solution and the final map takes usually less than 5 min on a standard desktop or even notebook computer. For a large number of sites, the phasing time can be extended to hours, especially when many of them are strong enough to trigger anisotropic temperature-factor refinement.

An alternative path for phasing and phase improvement can be performed with the use of *SOLVE*/*RESOLVE* (Terwilliger, 2002). This option is particularly useful to compare results from various procedures.

2.4. Preliminary model building

Currently, the preliminary model building is performed with the use of the fast option of *RESOLVE* (Terwilliger, 2004), which usually takes less than 5 min for a 150-residue protein. For a reasonable resolution of about 2.3 Å (diffraction limit, not anomalous signal limit) and high-quality SeMet data, about 70% of the model can be built in the fast mode. Another aspect of *HKL*-3000 is the ability to automatically build the most complete and accurate model. Extensive calculations can produce a fairly accurate and complete model and we investigated the trade-off between computational time and the quality of the result. There is a rather complex dependence of the ability of a particular algorithm to automatically build a model on resolution, solvent content and the use of NCS. In some cases, a rather complete model can be built even with 3 Å data. Statistical model building, which derives a composite model from several independent *RESOLVE* or *ARP*/*wARP* (Perrakis *et al.*, 1999) runs, is particularly promising.

3. Results and discussion

The initial goal of the *HKL*-3000 system was to evaluate very quickly the results of synchrotron SAD or MAD experiments and to allow the experimenter to decide whether the experiment had been successfully finalized and the crystal could be removed from the goniostat. In many cases, the described system was able to solve the structure even before finishing a one-wavelength data collection. Having a rapid preliminary structure solution provides tremendous value for managing limited resources, in particular crystal lifetime and beamline time. The advantage and growing popularity of single-wavelength SAD experiments (Fig. 4) is related not only to the simplicity of the measurements but most critically to the minimization of the effect of radiation damage on the phasing procedure. There is no guarantee that a SAD experiment will produce high-quality maps, especially for crystals with relatively low (below 40%) solvent content, as the power of solvent-flattening techniques used to resolve the phase ambiguity depends both on data resolution and solvent content. The presence of NCS significantly improves the

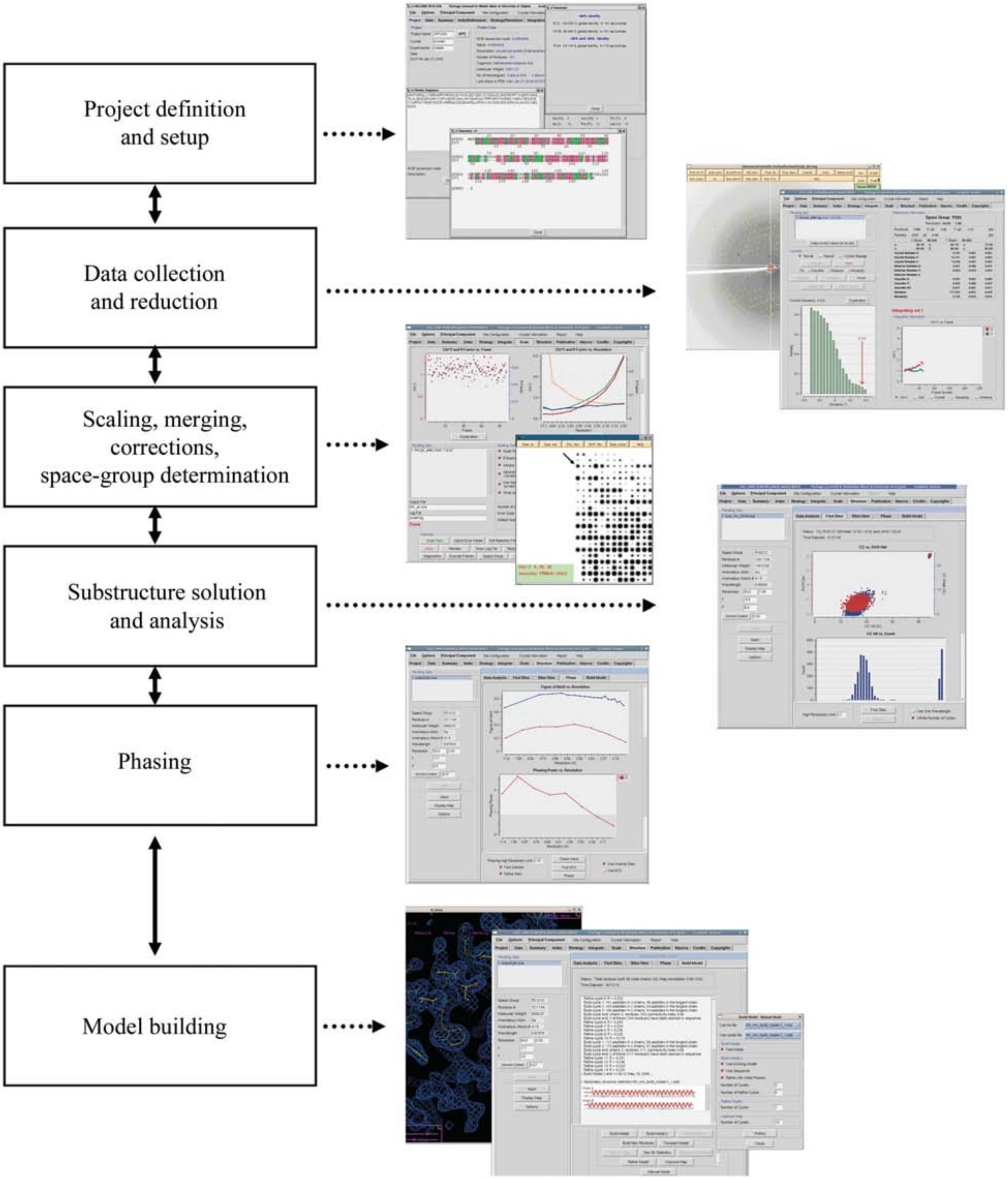

Project definition and setup
Data collection and reduction
Scaling, merging, corrections, space-group determination
Substructure solution and analysis
Phasing
Model building

Figure 7
General layout of *HKL*-3000. The experimenter is not limited to movement between neighboring blocks. For example, using the feedback of model building one can decide to collect more data (for example, a second wavelength or low-resolution pass) and jump directly back to the block 'Data collection and reduction'.

Table 2
Experimental electron-density map correlation *versus* final map for SAD (peak) and MAD (peak and inflection) phasing.

The same optimal sites were used. The comparison does not address the issue of relative advantage of SAD *versus* MAD strategies when the same exposure level is used.

PDB code	Solvent content (fraction)	Resolution (Å)	CC (main/ side chain), SAD	CC (main/ side chain), MAD
1wq6	0.37	2.0	0.76/0.67	0.81/0.72
2g7u	0.53	2.3	0.81/0.64	0.84/0.66
2fdo	0.46	2.5	0.58/0.43	0.66/0.48

chance of a successful SAD/MAD experiment. The concurrent data collection, processing and almost instantaneous preliminary structure solution provides an opportunity for ultimate verification of the X-ray experiment and allows one to change the data-collection strategy when the crystal is still in the cryoloop at the goniostat. When the quality of the SAD map is not satisfactory, the experimenter has the option to collect an additional second-wavelength data set and subsequently use the dispersion differences to improve phasing (Table 2). The addition of a third-wavelength data set usually does not produce substantial improvement.

The Bijvoet differences have a much higher signal-to-error ratio than dispersive (f'-dependent) differences (Minor *et al.*, 2000). Depending on the data-collection strategy and the possible influence of radiation damage, the non-isomorphism between different wavelengths may result in more problems than the small amount of phase information derived from dispersive differences. The benefit/cost ratio of measuring additional wavelengths is quite low in such a case and it becomes clear that most effort should be spent on optimal data collection at the absorption peak.

The current beta version of *HKL*-3000 is available for users at the Structural Biology Center, sector 19 at the Advanced Photon Source at Argonne National Laboratory (Rosenbaum *et al.*, 2006). The system is also routinely used for work related to the Midwest Center for Structural Genomics (MCSG) projects. The performance of the system is being evaluated and parameters optimized on the basis of the structures included in the MCSG database (http://www.mcsg.anl.gov). Of the 146 SAD/MAD structures solved by the MCSG since 1 January 2005, *HKL*-3000 has been used for the solution of 72 and 50% of SAD and MAD structures, respectively.

The continuous advancement of the decision-making procedures within *HKL*-3000 made it a system of choice for MCSG projects. A very quick path of 10–15 min from raw images to solved structure with 70% of a model built is no longer a surprise, but is a routine operation for data that diffract to 2.3 Å or better. The main goal of current development is to expand the applicability of the system to more difficult cases. The difficulties could be related to poorly diffracting crystals, large asymmetric unit size or high mosaicity. Highly mosaic non-perfect crystals (Fig. 5) do not pose difficulty for phasing as long as tools to refine spot size and multi-frame processing are used (Otwinowski & Minor, 2000). Similarly, a weak sulfur anomalous signal (Fig. 1*b*) does not preclude an excellent electron-density map (Fig. 6) as long as all pathologies are properly corrected. For our purposes, we define a difficult structure as one that has pathologies that are not recognizable by the current version of *HKL*-3000.

Low-resolution data, *i.e.* worse than 3 Å, do not present serious difficulty for substructure-solution or phasing procedures but often require extensive effort to complete the model building and refinement. A relatively high success rate with the SAD method with *HKL*-3000 (Figs. 4 and 6) is a consequence of the ability to recognize at data-collection time (Dauter, 2002) that a single-wavelength data set is often enough to solve the structure. SAD seems to be an effective strategy even for low-resolution data when the solvent fraction is high enough to resolve the phase ambiguity in the solvent-flattening procedure.

Despite the fact that *HKL*-3000 is under constant development, the general concept is already established, as presented in Fig. 7. Further tests of the program are planned with a web-based server to broaden the diversity of crystallographic projects tackled by *HKL*-3000.

We would like to thank the MCSG and SBC groups and Zbigniew Dauter for many test cases and feedback from the use of the early version of the program. We would also like to thank Tom Terwilliger and George Sheldrick for their extensive help with interfacing their programs and improving the stability of the system. We would like to thank Dominika Borek for continuous testing and critical analysis of the system behavior. We would also like to thank Andrzej Joachimiak, Diana Tomchick, Micha Machius, Alex Wlodawer and Matt Zimmerman for discussions and critical comments. We are grateful to Youngchang Kim for providing the 1xvi data used to create Fig. 5. We would like to thank the National Institutes of Health for supporting this work with grants GM53163 and GM62414. Part of the work was also supported by contract GI11496 from HKL Research.

References

Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. & Terwilliger, T. C. (2004). *J. Synchrotron Rad.* **11**, 53–55.

Borek, D., Minor, W. & Otwinowski, Z. (2003). *Acta Cryst.* D**59**, 2031–2038.

Brunzelle, J. S., Shafaee, P., Yang, X., Weigand, S., Ren, Z. & Anderson, W. F. (2003). *Acta Cryst.* D**59**, 1138–1144.

Chandonia, J. M. & Brenner, S. E. (2006). *Science*, **311**, 347–351.

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

Cowtan, K. (2001). *Acta Cryst.* D**57**, 1435–1444.

Cowtan, K. & Main, P. (1998). *Acta Cryst.* D**54**, 487–493.

Cowtan, K. D. & Zhang, K. Y. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.

Dauter, Z. (2002). *Acta Cryst.* D**58**, 1958–1967.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fu, Z. Q., Rose, J. & Wang, B.-C. (2005). *Acta Cryst.* D**61**, 951–959.

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

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

Holton, J. (2005). *Annual Meeting of the American Crystallographic Association.* Abstract W-0308.

Holton, J. & Alber, T. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 1537–1542.

Jones, T. A. (2004). *Acta Cryst.* D**60**, 2115–2125.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

Minor, W., Cymborowski, M. & Otwinowski, Z. (2002). *Acta Phys. Pol. A*, **101**, 613–619.

Minor, W., Tomchick, D. & Otwinowski, Z. (2000). *Structure*, **8**, R105–R110.

Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.

Otwinowski, Z., Borek, D., Majewski, W. & Minor, W. (2003). *Acta Cryst.* A**59**, 228–234.

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

Otwinowski, Z. & Minor, W. (2000). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 226–235. Dordrecht: Kluwer Academic Publishers.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.

Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Rosenbaum, G., Alkire, R. W., Evans, G., Rotella, F. J., Lazarski, K., Zhang, R. G., Ginell, S. L., Duke, N., Naday, I., Lazarz, J., Molitsky, M. J., Keefe, L., Gonczy, J., Rock, L., Sanishvili, R., Walsh, M. A., Westbrook, E. & Joachimiak, A. (2006). *J. Synchrotron Rad.* **13**, 30–45.

Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.

Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.

Terwilliger, T. C. (2004). *J. Synchrotron Rad.* **11**, 49–52.

Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1937–1940.

Todd, A. E., Marsden, R. L., Thornton, J. M. & Orengo, C. A. (2005). *J. Mol. Biol.* **348**, 1235–1260.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2006). In *Crystallographic Methods*, edited by S. Doublié. Totowa, NJ, USA: Humana Press.

Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimiak, A. (1999). *Acta Cryst.* D**55**, 1168–1173.