Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Application of molecular replacement to protein powder data from image plates

Jennifer A. Doebbler and Robert B. Von Dreele*

Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439, USA

Correspondence e-mail: vondreele@anl.gov

Received 3 October 2008
Accepted 31 January 2009

Macromolecular structures can be solved *via* molecular replacement from powder diffraction data collected not only on multi-analyzer diffractometers but also on image plates. Diffraction peaks recorded on image plates are generally broader than those collected using an array of crystal analyzer detectors, but the image-plate data often allow the use of powder data to lower *d*-spacings. Owing to the high incidence of overlaps in powder patterns, which is especially evident for larger structures, a multi-pattern Pawley refinement is necessary in order to distinguish intensity peaks. This work utilized various salt concentrations to produce small lattice distortions, which resulted in shifts of Bragg peak positions, in a suite of five powder patterns. Using reflection structure factors obtained from this combined refinement, the structure of hen egg-white lysozyme was determined by molecular replacement using the 60% identical human lysozyme (PDB code 1lz1) as the search model. This work also expands upon previous work by presenting a full-scale multi-species analysis combined with an investigation of the sensitivity with regard to discrimination between incorrect fold types. To test the limits of this technique, extension to higher molecular-weight structures is ongoing.

1. Introduction

X-ray powder diffraction techniques were developed in the early 1940s and, after the introduction of the Rietveld method (Rietveld, 1969), became an invaluable tool for crystal structure analysis of inorganic materials and small molecules. This technique is especially valuable for those materials that resist formation into single crystals suitable for diffraction. An ideal powder sample consists of myriad microcrystals randomly oriented in space. When incident radiation diffracts from this powder, all Laue conditions are satisfied at the same time. This collapses the three-dimensional pattern of Bragg peaks in reciprocal space onto a series of concentric rings. Each ring represents the set of reflections corresponding to a particular *d*-spacing for the given crystallite sample. Owing to this collapse, powder diffraction has an inherent overlap problem. All Friedel pairs and reflections with identical *d*-spacing values overlap exactly and cannot be resolved. Reflections with positions that differ by less than 10% of the full-width at half-maximum (FWHM) of the peak are also not distinguished. Mathematically, the former two overlap types cannot be resolved, but slight overlaps dependent upon the FWHM can be reduced by improving the detector resolution and sample characteristics. As a result, powder patterns do not provide a

© 2009 International Union of Crystallography
Printed in Singapore – all rights reserved

set of reflection intensities that can be treated as independent observables. However, one can obtain a set of 'observed intensities' from powder data *via* a Pawley refinement (Pawley, 1981), in which they are treated as independent variables in a Rietveld-like refinement (Rietveld, 1969). Because of overlaps between adjacent reflections, only the sum of the intensities is known *a priori* and the true partitioning of individual intensities within this sum is not known; this ambiguity is represented as a covariance between their values. The magnitude of the covariance between any pair of reflection intensities is unity if they exactly overlap and is zero when they are far apart; partially overlapped reflections have covariances within this range depending on the degree of overlap. Thus, any structure-solution technique that uses these intensities is handicapped in comparison with the use of intensities obtained from a single crystal.

In 1999, Von Dreele explored the possibility of examining protein structures using powder diffraction in a study of metmyoglobin. A full refinement of the macromolecule was performed using powder diffraction to a minimum d-spacing of 3.3 Å (Von Dreele, 1999). The success of this analysis led to proof-of-concept studies on the applicability of a standard single-crystal diffraction technique to the field of macromolecular powder diffraction: namely molecular replacement. Molecular replacement of $T_3R_3^f$ human insulin that had been found in a new crystalline phase was performed (Von Dreele *et al.*, 2000; Von Dreele, 2003). Five years later, Margiolaki *et al.* (2005) published a molecular-replacement solution of turkey egg-white lysozyme from a model of the monoclinic form with powder data collected from the hexagonal phase. This work established powder diffraction as a useful probe for protein polymorphs and demonstrated the robustness of molecular replacement using high-resolution powder data (*i.e.* sharp peaks) collected on diffractometers equipped with one or more crystal analyzer detectors. More recently, Margiolaki *et al.* (2007) determined the structure of the 67-residue second SH3 domain of ponsin using multi-analyser diffractometer data. This previously unknown structure was solved using a model with 40% sequence identity. This suggested that powder diffraction could not only be used for phase and derivative identification of known structures, but also for solving unknown protein structures. This is especially useful since microcrystalline powders are often overlooked as merely byproducts of attempts at growing suitable single crystals.

In the work presented here, the viability of using image-plate data to produce molecular-replacement solutions was examined. The peaks are broader for image plates than multianalyzer diffractometers owing to the detector resolution, but the entire d-spacing range can be recorded in one shot. This reduces the radiation damage sustained by the sample, which is highly beneficial for macromolecular studies. To this end, the structure of hen egg-white lysozyme (HEWL) was solved using the 60% identical human lysozyme (PDB code 1lz1; Artymiuk & Blake, 1981) as the starting model. Following the success with the SH3 domain of ponsin, an attempt was made to use the *MOLREP* routine (Vagin & Teplyakov, 1997) in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Two other Patterson-based searches available in *CCP*4, *AMoRe* (Navaza, 1994) and *Phaser* (McCoy *et al.*, 2007), were also tried, as well the *CCP*4 *Beast* routine (Read, 1999); the latter performs a Patterson search combined with an analysis of symmetry mates. Each of these routines was designed for single-crystal protein work and therefore contains no mechanism to deal with the ambiguous intensity values (arising from the overlap problem) when scoring the solutions.

Consequently, a version of the program *PSSP* (Pagola & Stephens, 2000) locally modified to accommodate protein structures was used to facilitate solution. In contrast to single-crystal codes (*e.g. MOLREP*), *PSSP* uses a cost function based on the severity of the overlaps (defined by their FWHM values and scattering angles) to evaluate a series of combined Monte Carlo/simulated-annealing (MC/SA) trials. The cost function (S) in *PSSP* is defined as

$$S = \frac{\sum_{h,h'} (I_{oh} - I_{ch}) \Phi_{hh'} (I_{oh'} - I_{ch'})}{\left(\sum_{h} I_{oh}\right)^2}, \qquad (1)$$

where the overlap function $\Phi_{hh'}$ is described by

$$\Phi_{hh'} = \sum_{2\Theta} f_h(2\Theta) f_{h'}(2\Theta) \qquad (2)$$

and f_h is the Gaussian peak-shape function (for a given reflection h), with I_o and I_c referring to observed (*i.e.* Pawley extracted) and calculated (*i.e.* generated from the model) intensities, respectively. $\Phi_{hh'}$ is dependent upon the FWHM values of the peak-shape function and therefore becomes large for overlapping reflections. This cost function is identical in form to that used by David *et al.* (2001), where the covariance between a pair of reflection intensities obtained from a Pawley refinement is used in place of $\Phi_{hh'}$. In essence, this cost function is sensitive to differences in the aggregate of intensities for overlapped reflections rather than for their individual values. The stability of this approach was then demonstrated by molecular-replacement experiments with several other lysozyme-like proteins. Finally, a variety of proteins with different folds showed the specificity of this technique for protein shape as well as the limits on choosing the correct starting model.

2. Materials and methods

The raw data analyzed here were used in a previous study (Von Dreele, 2007), in which the experimental setup and preprocessing were described in greater detail. The image plate records diffraction data from large to small d-spacings within a single rapidly collected image, rather than using a sequentially collected 2θ scan (as performed with a multi-analyzer diffractometer). This method virtually eliminates any angular variation in the recorded intensity arising from radiation damage compared with that from a 2θ scan. The macromolecular powder slurry in mother liquor was exposed for 30 s while spinning at 60 Hz. The data were collected with a MAR345 image plate on beamline 1-BM at the Advanced

Table 1
Results from *PSSP* for the human-to-hen lysozyme molecular replacement.

The x, y and z coordinates are in fractions of the tetragonal unit cell with respect to the geometric center of the protein.

Trial	S	α (°)	β (°)	γ (°)	x	y	z
1	0.4704	179.25	174.33	350.37	0.0021	0.2802	0.7266
2	0.4587	151.29	5.27	200.95	−0.0048	0.2782	0.5223
3	0.4588	153.23	5.22	199.21	−0.0061	0.2782	0.5234
4	0.4624	128.23	4.70	224.76	−0.0061	0.2762	0.5221
5	0.4600	135.66	4.84	217.35	−0.0060	0.2769	0.5226

Photon Source, Argonne National Laboratory with an incident radiation of 20 keV (λ = 0.620738 Å), yielding diffraction patterns to a minimum d-spacing of ∼2.0 Å with ∼0.03° FWHM. The patterns obtained from samples containing 0.05 *M* potassium hydrogen phthalate buffer pH 4.0 and 1.25, 1.00, 0.75, 0.50 and 0.25 *M* NaCl, respectively, were chosen for this investigation.

A multi-pattern Pawley refinement (Pawley, 1981) was carried out in *GSAS* (Larson & Von Dreele, 1994) to obtain a set of 4227 structure factors for each of the five data sets. The refinement spanned 0.6–14° (2θ), which corresponds to a 56–2.5 Å d-spacing range. The full extent of the collected data (56–2.0 Å, with ∼9000 reflections) was not used because the overlaps become more severe as the d-spacing decreases and thus the partitioning of the extracted intensities become less accurate. Since molecular-replacement routines do not require high-resolution data, the set was truncated in order to smooth the refinement.

A single diffraction image from an LaB_6 standard collected with the same instrument configuration was used to obtain suitable starting values for the peak-profile coefficients. An 18-term shifted Chebyshev function was manually described in *EXPGUI* (Toby, 2001) to initially define the background. After refinement of the Pawley intensities stabilized, the background was modified to an 18-term log-linear interpolation function, which uses more points in the low-2θ range and more easily describes the solvent-dominated background. An intensity extraction was performed during the Pawley refinement for each histogram, yielding a total of ∼21 900 extracted reflection intensities over all five patterns. These were merged using the *OVERLP* function (Von Dreele, 2007) in *GSAS* to yield ∼4380 averaged intensities for each histogram; the merging R value was 7.23%. *OVERLP* reduced the effective FWHM (required by the *PSSP* cost function) of each reflection to account for the extra information gained by the relative shifts in reflection positions.

The human lysozyme model (PDB code 1lz1) was modified with the *CCP*4 program *CHAINSAW* (Stein, 2008) to create an atomic sequence which more closely reflected that of the HEWL target. This algorithm takes information from a pairwise sequence alignment (created using *Jalview*; Clamp *et al.*, 2004) of the human lysozyme and HEWL and then truncates the human model. One extraneous residue, Gly48, was deleted from the β-hairpin of the human model and 51 residues were truncated to their last common atom. This resulted in a model that we believed best represented the main fold characteristics of our data while minimizing bias from inaccurate side-chain information.

We initially attempted to solve the structure by molecular replacement using the extracted and merged intensities from *GSAS* and conventional single-crystal diffraction software in the *CCP*4 suite. Trials involving *AMoRe* and *Phaser* either rendered nonsensical solutions (replete with severe symmetry-mate clashes) or simply failed to converge. A single trial using the *Beast* program ran for 2 d without converging. Conversely, the *MOLREP* routine ran to completion, with several seemingly worthwhile solutions of the rotation and translation functions: contrast values were around 95 (according to the manual a contrast above 2.5 is reported to be 'definitely a solution'). Upon further analysis, however, they were often found to be false positives: the solutions displayed severe clashes and/or overlaps with symmetry-related molecules. Occasionally *MOLREP* would produce the correct orientation, but with such high contrast scores for all of the results there was no discrimination between correct and incorrect solutions.

Clearly, the single-crystal diffraction software was ill-equipped to deal with image-plate powder data. This was in contrast to the results of the multianalyzer diffractometer studies, in which the higher instrumental resolution reduced the severity of the intensity ambiguity (thus more closely mimicking single-crystal data) and allowed the use of *MOLREP*. Therefore, we shifted to the *PSSP* program originally written by Pagola & Stephens (2000) and subsequently modified by us for analysis of protein structures.

3. Results and discussion

3.1. HEWL molecular-replacement solution

For our test with human lysozyme, the first 150 reflections ($56 > d > 8.4$ Å) from the multi-pattern Pawley refinement and the 934-atom coordinate set resulting from the truncation of human lysozyme (space group $P2_12_12_1$) in *CHAINSAW* were input to *PSSP*. Six parameters were varied over the course of the MC/SA protocol: the three coordinates of the geometric center of the starting model and three Eulerian rotations about that center. Rotations were performed over the z axis (ω: 0–360°), the x axis (χ: 0–180°) and the z axis again (φ: 0–360°). To simplify comparisons of results, a restricted asymmetric unit (−1/8 to 1/8, 0 to 1/2 and 1/4 to 3/4) was chosen for the coordinate ranges to force the molecular origin to be within the same asymmetric unit that is conventionally used to describe tetragonal HEWL. A search over the entire unit cell would have potentially yielded equivalent solutions over eight different origin choices and two sets of orientations, making comparisons difficult. (In our chosen search cell, two orientation choices are still possible since the full angular search range remains intact.) The simulated annealing began with an initial melt 'temperature' of 50 and adiabatically cooled in decrements of 20% until reaching a 'freeze' of 0.001, with 5000 trials at each step.

Table 2
Results from the multi-species molecular-replacement scheme.

The percentage identity was determined from a pairwise alignment in *Jalview* with the 194l HEWL structure. *PSSP* runs were performed for each species using the first 50, 100, 150, 200 or 300 reflections as input. The columns indicate the number of correct solutions (out of four trials) for each of the 15 respective sets of input data. The space-group and resolution information were obtained from the published references. 1el1a and 1el1b are the two noncrystallographic symmetry-related molecules for the canine milk protein.

PDB code	Lysozyme species	Identity (%)	Runs					Space group	Resolution (Å)	Reference
			50	100	150	200	300			
1a4v	*Homo sapiens* (α-lactalbumin)	36	3	3	3	2	3	$P2_12_12$	1.8	Chandra *et al.* (1998)
1el1a	*Canis familiaris*	53	3	2	4	4	4	$P3_2$	1.9	Koshiba *et al.* (2000)
1el1b	*C. familiaris*	53	3	4	4	4	4	$P3_2$	1.9	Koshiba *et al.* (2000)
1gd6	*Bombyx mori*	42	3	2	1	2	0	$P4_12_12$	2.5	Matsuura *et al.* (2002)
1hfx	*Cavia porcellus* (α-lactalbumin)	33	2	4	2	4	3	$P2_12_12_1$	1.9	Pike *et al.* (1996)
1iiz	*Antheraea mylitta*	38	1	3	2	4	4	$C222_1$	2.4	Jain *et al.* (2001)
1jhl	*Phasianus colchicus* (antibody complex)	92	3	4	4	4	4	$P6_5$	2.4	Chitarra *et al.* (1993)
1jsf	*H. sapiens*	60	1	4	4	4	4	$P2_12_12_1$	1.15	Harata *et al.* (1998)
1jug	*Tachyglossus aculeatus*	53	4	2	2	4	3	$P2_1$	1.9	Guss *et al.* (1997)
1lmq	*Oncorhynchus mykiss* (NAG_4 complex)	60	4	4	4	4	4	$P3_12_1$	1.6	Karlsen & Hough (1995)
1lsy	*Gallus gallus* (D52S mutation)	99	4	4	4	4	4	$P4_32_12$	1.9	Hadfield *et al.* (1994)
1op9	*H. sapiens* (antibody complex)	60	1	4	3	4	4	$P2_1$	1.86	Dumoulin *et al.* (2003)
7lyz	*G. gallus*	100	2	4	4	4	4	$P1$	2.5	Herzberg & Sussman (1983)
LYS4	*G. gallus* powder diffraction data	100	4	4	4	4	4	$P4_32_12$	—	Unpublished powder data refinement
—	Polyalanine chain	9	1	2	1	2	2	$P2_1$	—	Converted LYS4

The results of the five runs of *PSSP* are shown in Table 1. Similar values of the cost function *S* (which is a 'goodness-of-fit' indicator; smaller *S* values denote better solutions) were observed for all five runs, although runs 2–5 exhibited slightly lower *S* values than that obtained for run 1. Runs 2–5 superimposed well with the previously published HEWL structure (PDB code 194l; Vaney *et al.*, 1996) and all displayed essentially identical orientations in the unit cell. The *x*, *y* and *z* coordinates of their centroids had a maximum variation of 0.0013, 0.0020 and 0.0011 from the best solution, which corresponded to deviations of 0.10, 0.16 and 0.04 Å, respectively, from the known structure of HEWL. Compared with the 8.4 Å resolution of the data, these differences are trivial. The first solution, however, differed from the 194l structure by a 180° rotation, which signified a different choice of origin rather than an erroneous solution, since symmetry-mate contacts remained the same. (Unlike many single-crystal molecular-replacement packages, *PSSP* does not check for atomic clashes or symmetry-related solutions within the output solutions; all of this analysis must be performed externally.) *PSSP*, therefore, rendered five out of five correct solutions and provides a robust method for molecular replacement of proteins using powder diffraction data.

3.2. Multi-species runs

After the successful molecular replacement of the human model from 1lz1 into the HEWL tetragonal space group, the method was re-examined using 15 structures from nine different species along with a polyalanine chain (Table 2). This was performed to demonstrate that the technique would work for starting models with a broad range of percentage identities to HEWL. All 15 of these structures were evaluated using the same HEWL data from the previous multi-pattern run. In an attempt to reduce the impact of different choices of molecular orientation (which would make straight comparisons difficult),

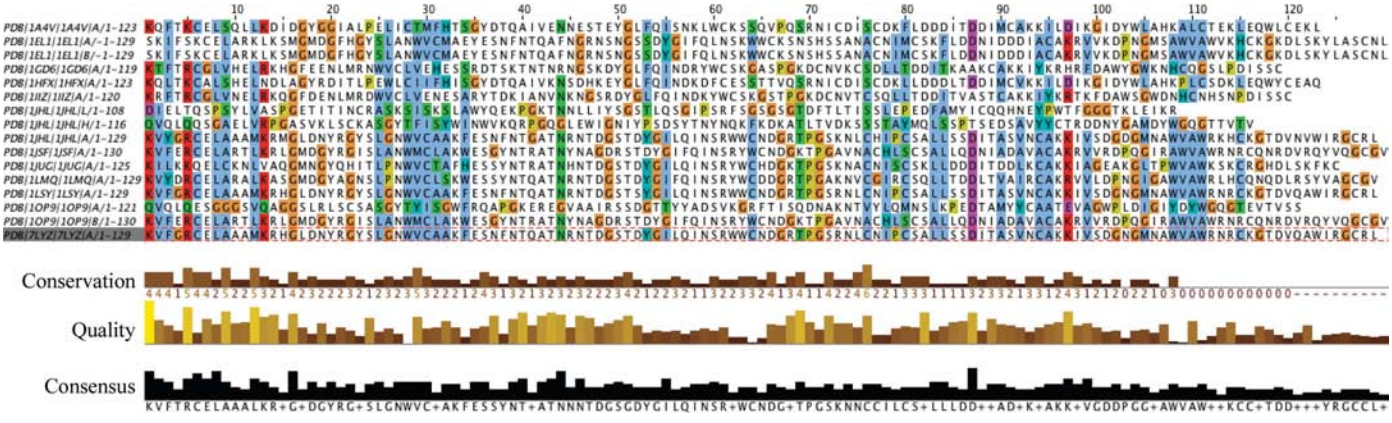

Figure 1
Multiple sequence-alignment results from *Jalview* for the multi-species runs. The A and B tags after the PDB code refer to the chain ID. The LYS4 and polyalanine chains were omitted from the alignment. The coloring was added according to the *ClustalW* algorithm (Larkin *et al.*, 2007).

a superposition of 14 of the structures onto the *Tachyglossus aculeatus* protein (PDB code 1jug) was carried out. This placed all initial atomic coordinates in the same orientation as found for this monoclinic protein. The same restricted asymmetric unit was used as for the initial molecular replacement. All waters, inhibitors and complexed proteins were removed from each of the coordinate files before running *PSSP*. The chosen species were lysozyme proteins with identities to HEWL in the range 9–100% (Fig. 1). The exceptions were 1hfx and 1a4v; these proteins are α-lactalbumins, which maintain lysozyme-like folds.

PSSP runs were begun for each species using the first 50, 100, 150, 200 or 300 reflections (corresponding to a minimum *d*-spacing of 12.4, 9.6, 8.4, 7.6 and 6.5 Å, respectively) obtained from the Pawley refinement. Each run consisted of four separate trials for a total of 300 molecular-replacement experiments. The results did not indicate an optimal number of reflections to use, but generally 200 reflections ($d_{\rm min}$ of $\sim$7.6 Å) consistently gave the highest percentage of correct solutions. The *S* values for each trial (not presented here) varied depending upon the number of reflections input. Lower *S* values were observed on average when a higher number of intensity points were evaluated. However, the *S* values were a good indicator of a correct solution only on a per-run basis. Of the four trials performed for each set of reflections, the wrong solutions consistently had higher *S* values than the correct solutions. For example, the *Phasianus colchicus* protein (PDB code 1jhl; run with an input of 50 reflections) gave *S* values of 0.8695, 0.8696 and 0.8695 for the correct answer and 1.5130 for an incorrect orientation.

The 50- and 100-reflection runs contained more outliers (*i.e.* incorrect solutions) in terms of Euler angles and *xyz* coordinates than the 200- and 300-reflection runs, but overall the correct solutions were still obtained (Fig. 2). This could be because the Babinet's principle solvent correction internal to *PSSP* was inadequate to describe the lowest order reflection intensities obtained from powder diffraction data. They would be most affected by the solvent, which could also explain why runs with larger numbers of reflections input were on average more efficient at finding the correct solutions.

3.3. Fold discrimination

Finally, an analysis of whether *PSSP* was sensitive to the protein fold of the starting model was undertaken. Ideally, the molecular-replacement starting model would belong to the same fold family as the target structure. The molecular-replacement algorithm treats the model as a rigid body and therefore chain movements to optimize the model within the search are not possible. Several different fold families with

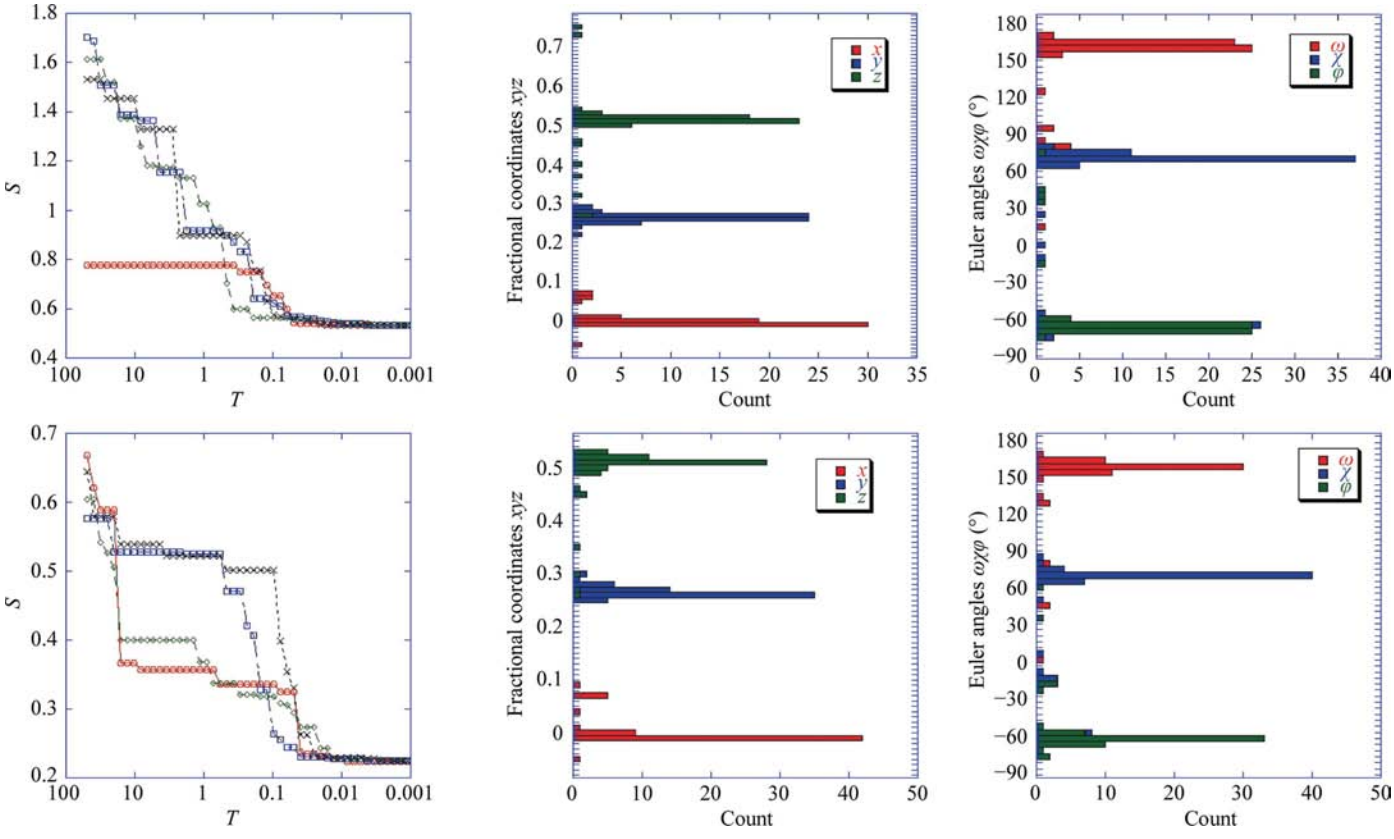

Figure 2
Representative results from the multi-species *PSSP* runs. The two graphs on the left display the reduction of the *S* value as a function of temperature (*T*) for each of the four trials of the 1jhl model performed using the first 100 (top) or 300 (bottom) reflection intensities. The four graphs on the right are the combined results from all 300 multi-species trials. They show a histogram of the resulting Euler angles and *xyz* coordinates as a function of the number of total results with those values (counts). The upper two are the outcomes of the 100-reflection runs, while the lower two are from the 300-reflection runs.

molecular weights ranging from 13.7 to 14.8 kDa were analyzed (Fig. 3). The 194l and 1a4v structures were used as controls representing the lysozyme-like family. Ribonuclease A (PDB code 1rbx; Dunbar *et al.*, 1997) and GABA RAP (PDB code 1gnu; Knight *et al.*, 2002) were constituents of the $\alpha\beta$-roll group. Hemoglobin I (PDB code 1b0b; Bolognesi *et al.*, 1999) represented orthogonal bundles. The sandwich fold was analyzed using profilin (PDB code 1a0k; Thorn *et al.*, 1997) and NKP44 (PDB code 1hkf; Cantoni *et al.*, 2003). The final family evaluated was the up–down bundle represented by histidine kinase (PDB code 1a0b; Kato *et al.*, 1997) and phospholipase (PDB code 1bp2; Dijkstra *et al.*, 1981).

The 1lz1 (human lysozyme) and 1a4v (α-lactalbumin) models were the only sequences of the nine selected which aligned pairwise with the 194l (HEWL) model for a significant portion of the sequence. 1lz1 exhibited 60% sequence identity to 194l for a 130-residue alignment, while 1a4v was 36% identical over 128 residues. The seven other protein models and fold types produced pairwise alignments of only 2–20 residues in length. Based on the *Jalview* analysis, we determined that the truncation method used in *CHAINSAW* could only be applied to the 1lz1 and 1a4v proteins. These two sequences were modified and the rest of the proteins were processed with *PSSP* once all water molecules had been removed. All of the multi-species *PSSP* runs used a 50–0.001 annealing schedule and 5000 cycles at each step, as for the previous *PSSP* studies, with the exception of 1rbx and 1hkf, which were run for both 5000 and 1000 cycles.

The 194l and 1lz1 structures both served as controls and gave five out of five correct solutions. The 1a4v sequence that resulted from the *CHAINSAW* truncation failed to give any solutions aligned with the 194l structure. This sequence was then rerun without truncation. As before, three out of four correct solutions were obtained (for a 150-reflection run). The use of *CHAINSAW* improved the fit for the 60% identical 1lz1 human lysozyme, but was detrimental to 1a4v. We surmise that sequence truncations will only improve the fit in cases where the percentage identity is quite high; thus, trials should therefore be run without truncation initially and only modified as a secondary step to probe further.

All other fold types only produced solutions with major clashes no matter the number of reflections that were input

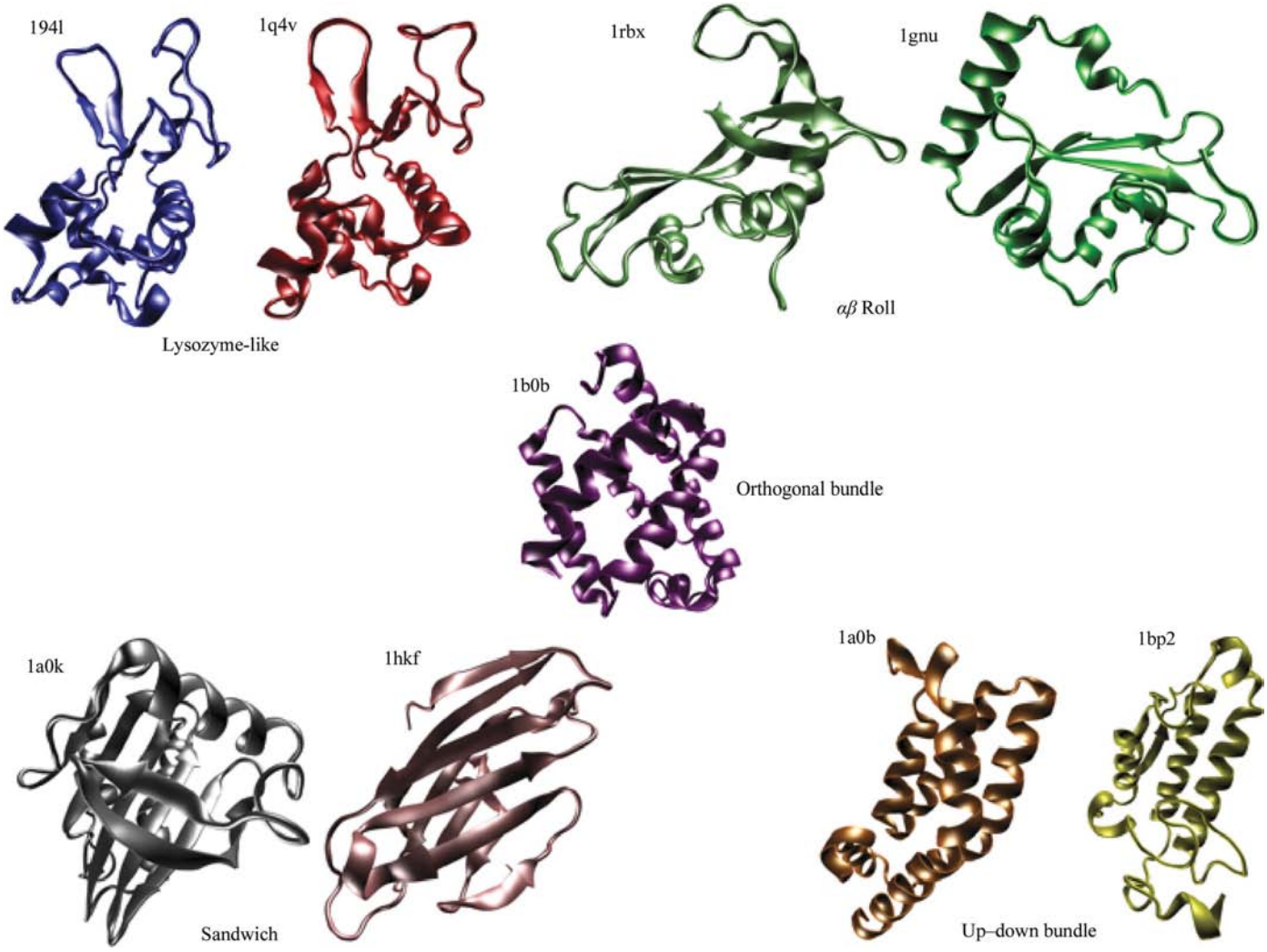

Figure 3
Ribbon diagrams of the nine proteins used in the fold-matching analysis. Diagrams were created using *VMD* (Humphrey *et al.*, 1996).

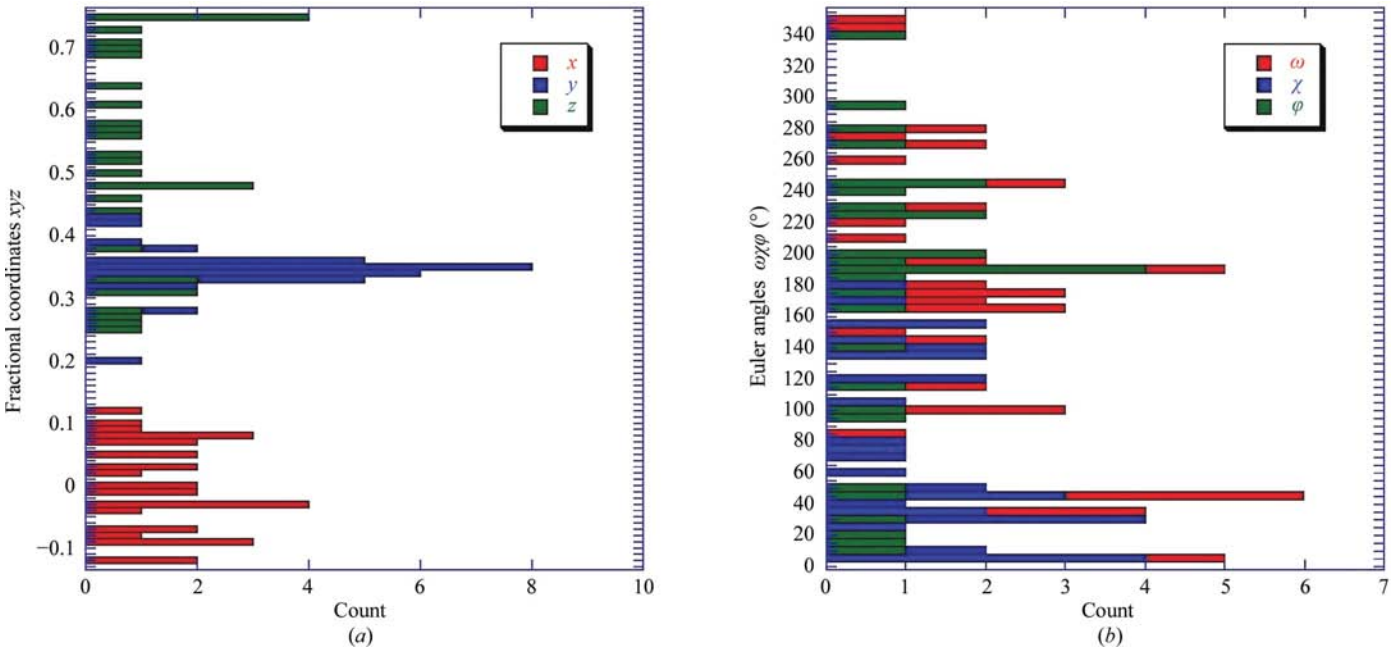

Figure 4
Histograms of the results of the fold-discrimination output. One set of outputs from seven of the fold types are plotted to reveal the randomness of the results. The fractional coordinate solutions are shown in (*a*), while the angular results are plotted *versus* the number of solutions at each value in (*b*).

(Fig. 4). The *y* coordinates of the results cluster around a common value (incidentally not the correct *y* value), but the *x*- and *z*-coordinate values, like the Euler angles, span the full range of the allowed values. This was not surprising because non-lysozyme-like folds should not properly fit into a lysozyme-like unit cell because of packing restrictions unique to the size and shape of the protein. Moreover, the non-lysozyme fold types yield calculated structure factors that must be sufficiently different from those of a lysozyme fold, so that the computed cost function (*S*) for solutions with the correct placement in the cell is not at a minimum and is thus not a 'reasonable' solution. The clashes were significant across the board: globular proteins did not differ from sandwich configurations. The angular coordinates had a wide spread of values that gave correct solutions for the 1lz1 and 1a4v proteins; however, the *x*, *y* and *z* coordinates lined up within a narrower range. For the non-lysozyme-like proteins, both the angular positions and the centroid coordinates varied greatly between trials. With the center not properly aligned, even globular proteins (which could theoretically pack in a tetragonal cell) were not optimally configured and thus produced severe clashes.

4. Conclusion

The single human lysozyme to HEWL hen-egg white lysozyme experiment and the multi-species trials demonstrated the reproducibility of molecular-replacement solutions using image-plate data. The lower instrument resolution of the image-plate data rendered single-crystal molecular-replacement software unusable, which had not been the case for the crystal analyzer experiments. It is likely that a combination of multianalyser diffractometer and image-plate data would lead to better peak assignment while maintaining measurable diffraction out to low *d*-spacings. Future developments in detector design could also reduce the impact of the overlaps and allow powder diffraction of larger protein structures. However, as it stands, multianalyser diffractometer and image-plate data can both facilitate structure solution of proteins in at least the 15–20 kDa range. The multi-pattern intensity extraction was sufficient to discriminate between fold types that otherwise would have been possible molecular-replacement candidates based on molecular size. As an overall strategy, several runs with various numbers of reflection intensities as input should be undertaken to confirm the solutions. It should be noted that the *S* values within each run (containing multiple individual trials) were found to signify the best solution for that run, but cross-run correlations with different protein models were not possible even on the basis of percentage identity.

The authors would like to thank Peter Lee for help with the data collection and Bob Blessing for fruitful conversations. Use of the APS was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract No. DE-AC02-06CH11357.

References

Artymiuk, P. J. & Blake, C. C. (1981). *J. Mol. Biol.* **152**, 737–762.
Bolognesi, M., Rosano, C., Losso, R., Borassi, A., Rizzi, M., Wittenberg, J. B., Boffi, A. & Ascenzi, P. (1999). *Biophys. J.* **77**, 1093–1099.
Cantoni, C., Ponassi, M., Biassoni, R., Conte, R., Spallarossa, A., Moretta, A., Moretta, L., Bolognesi, M. & Bordo, D. (2003). *Structure*, **11**, 725–734.

Chandra, N., Brew, K. & Acharya, K. R. (1998). *Biochemistry*, **37**, 4767–4772.
Chitarra, V., Alzari, P. M., Bentley, G. A., Bhat, T. N., Eisele, J. L., Houdusse, A., Lescar, J., Souchon, H. & Poljak, R. J. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 7711–7715.
Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004). *Bioinformatics*, **20**, 426–427.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
David, W. I. F., Shankland, K., Cole, J., Maginn, S., Motherwell, W. D. S. & Taylor, R. (2001). *DASH User Manual.* Cambridge Crystallographic Data Centre, Cambridge, England.
Dijkstra, B. W., Kalk, K. H., Hol, W. G. & Drenth, J. (1981). *J. Mol. Biol.* **147**, 97–123.
Dumoulin, M., Last, A. M., Desmyter, A., Decanniere, K., Canet, D., Larsson, G., Spencer, A., Archer, D. B., Sasse, J., Muyldermans, S., Wyns, L., Redfield, C., Matagne, A., Robinson, C. V. & Dobson, C. M. (2003). *Nature (London)*, **424**, 783–788.
Dunbar, J., Yennawar, H. P., Banerjee, S., Luo, J. & Farber, G. K. (1997). *Protein Sci.* **6**, 1727–1733.
Guss, J. M., Messer, M., Costello, M., Hardy, K. & Kumar, V. (1997). *Acta Cryst.* D**53**, 355–363.
Hadfield, A. T., Harvey, D. J., Archer, D. B., MacKenzie, D. A., Jeenes, D. J., Radford, S. E., Lowe, G., Dobson, C. M. & Johnson, L. N. (1994). *J. Mol. Biol.* **243**, 856–872.
Harata, K., Abe, Y. & Muraki, M. (1998). *Proteins*, **30**, 232–243.
Herzberg, O. & Sussman, J. L. (1983). *J. Appl. Cryst.* **16**, 144–150.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Jain, D., Nair, D. T., Swaminathan, G. J., Abraham, E. G., Nagaraju, J. & Salunke, D. M. (2001). *J. Biol. Chem.* **276**, 41377–41382.
Karlsen, S. & Hough, E. (1995). *Acta Cryst.* D**51**, 962–978.
Kato, M., Mizuno, T., Shimizu, T. & Hakoshima, T. (1997). *Cell*, **88**, 717–723.
Knight, D., Harris, R., McAlister, M. S., Phelan, J. P., Geddes, S., Moss, S. J., Driscoll, P. C. & Keep, N. H. (2002). *J. Biol. Chem.* **277**, 5556–5561.
Koshiba, T., Yao, M., Kobashigawa, Y., Demura, M., Nakagawa, A., Tanaka, I., Kuwajima, K. & Nitta, K. (2000). *Biochemistry*, **39**, 3248–3257.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Larson, A. C. & Von Dreele, R. B. (1994). *General Structure Analysis System (GSAS).* Los Alamos National Laboratory Report. Los Alamos National Laboratory, Los Alamos, USA.
Margiolaki, I., Wright, J. P., Fitch, A. N., Fox, G. C. & Von Dreele, R. B. (2005). *Acta Cryst.* D**61**, 423–432.
Margiolaki, I., Wright, J. P., Wilmanns, M., Fitch, A. N. & Pinotsis, N. (2007). *J. Am. Chem. Soc.* **129**, 11865–11871.
Matsuura, A., Yao, M., Aizawa, T., Koganesawa, N., Masaki, K., Miyazawa, M., Demura, M., Tanaka, I., Kawano, K. & Nitta, K. (2002). *Biochemistry*, **41**, 12086–12092.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Pagola, S. & Stephens, P. W. (2000). *Mater. Sci. Forum*, **321–324**, 40–45.
Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.
Pike, A. C., Brew, K. & Acharya, K. R. (1996). *Structure*, **4**, 691–703.
Read, R. J. (1999). *Acta Cryst.* D**55**, 1759–1764.
Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
Thorn, K. S., Christensen, H. E., Shigeta, R., Huddler, D., Shalaby, L., Lindberg, U., Chua, N. H. & Schutt, C. E. (1997). *Structure*, **5**, 19–32.
Toby, B. H. (2001). *J. Appl. Cryst.* **34**, 210–213.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.
Von Dreele, R. B. (1999). *J. Appl. Cryst.* **32**, 1084–1089.
Von Dreele, R. B. (2003). *Methods Enzymol.* **368**, 254–267.
Von Dreele, R. B. (2007). *J. Appl. Cryst.* **40**, 133–143.
Von Dreele, R. B., Stephens, P. W., Smith, G. D. & Blessing, R. H. (2000). *Acta Cryst.* D**56**, 1549–1553.