Macromolecular powder diffraction: Structure solution via molecular replacement

J. A. Doebbler, R. B. von Dreele *

Advanced Photon Source, Argonne National Laboratory
* vondreele@anl.gov

Keywords: molecular replacement, protein, powder diffraction

Abstract. Macromolecular powder diffraction is a burgeoning technique for protein structure solution—ideally suited for cases where no suitable single crystals are available. Over the past seven years, pioneering work by Von Dreele et al. [1,2] and Margiolaki et al. [3,4] has demonstrated the viability of this approach for several protein structures. Among these initial powder studies, molecular replacement solutions of insulin and turkey lysozyme into alternate space groups were accomplished. Pressing the technique further, Margiolaki et al. [5] executed the first molecular replacement of an unknown protein structure: the SH3 domain of ponsin, using data from a multianalyzer diffractometer.

To demonstrate that cross-species molecular replacement using image plate data is also possible, we present the solution of hen egg white lysozyme using the 60% identical human lysozyme (PDB code: 1LZ1) as the search model. Due to the high incidence of overlaps in powder patterns, especially in more complex structures, we have used extracted intensities from five data sets taken at different salt concentrations in a multi-pattern Pawley refinement. The use of image plates severely increases the overlap problem due to lower detector resolution, but radiation damage effects are minimized with shorter exposure times and the fact that the entire pattern is obtained in a single exposure. This image plate solution establishes the robustness of powder molecular replacement resulting from different data collection techniques.

Introduction

X-ray powder diffraction techniques were developed in the 1940s and soon became an invaluable tool for phase identification and, more recently, determination of small-molecule crystal structures. At the turn of the 21st century, Von Dreele addressed the question of how to interpret the structure of a protein if suitable single crystals were not available [6]. The answer was macromolecular powder diffraction. Over the following eight years, the technique has emerged as a viable counterpart to single-crystal diffraction studies. In 2000, Von Dreele et al. demonstrated molecular replacement of T3R3f human insulin into a new crystalline phase [6]. This work was important because it demonstrated that if the structure of one
protein derivative was known from single-crystal studies, then the structural changes in a series of protein derivatives could be determined with powder data. Later, in 2005, Margiolaki et al. published a molecular replacement and subsequent refinement from powder data of the structure of turkey egg white lysozyme in the hexagonal phase [3]. These two studies affirmed that molecular replacement could work for powder data and established that powder diffraction is a useful probe for protein polymorphs. Most recently, Margiolaki et al. solved the structure of the 67-residue SH3 domain of ponsin by molecular replacement [5]. This previously unknown structure was solved using a model with 40% sequence identity, demonstrating that powder diffraction could be used not only for phase and derivative identification but for novel structural studies as well.

In this analysis, we survey a 14 kDa molecular weight structure and determine the robustness of molecular replacement solutions using powder data. The hen egg white lysozyme (HEWL) structure was solved by this method using the 60% identical human lysozyme (PDB code: 1LZ1) as the starting model. Data were collected on an image plate, which achieves high d-spacing resolution with a single image, thus virtually eliminating any changes resulting from radiation damage on a per-shot basis. However, the peaks are distinctly broader on an image plate than with the best multianalyzer diffractometers available, due to the resolution of detector, and it is therefore more difficult to resolve the peaks into individual observable intensities [7]. For large structures and high-symmetry space groups, the overlaps become even more problematic. Unlike multianalyzer diffractometers, the extracted intensities from image plates in the low d-spacing regime cannot be treated as if they were traditional single crystal diffraction data.

Materials and methods

The raw data used in this analysis are the same as used in a previous study, obtained on beamline 1-BM at the Advanced Photon Source, Argonne National Laboratory [8]. Radiation at 20 keV was focused to the surface of a MAR345 detector positioned ~715 mm away from and 6.5 cm vertically offset in relation to the sample. The macromolecular powder slurry in mother liquor was exposed for 30 s while spinning at 60 Hz. For this multi-pattern analysis, we focused on the samples prepared with 0.05 M potassium hydrogen phthalate (pH 4.0) and 1.25, 1.00, 0.75, 0.50, and 0.25 M NaCl, respectively.

A multi-pattern Pawley refinement [9] was carried out in GSAS [10] to obtain a set of structure factors for each of the five data sets. The Pawley refinement spanned 0.6–14.0° 2θ, which corresponds to a d-spacing range of 56–2.5 Å. Initially an 18-term Chebyshev function in EXPGUI [11] was used to describe the background and the preliminary profile coefficients were obtained from the analysis of a LaB₆ standard. After the Pawley intensities stabilized, the background was modified to a log linear interpolation function, which uses more points in the low 2θ range.

A combined refinement of data sets with differing salt concentrations offers a much better assignment of reflection intensities, because the resulting small anisotropic changes in lattice parameters alters the relative reflection positions, thus changing their overlaps. The 5 data sets were merged with the overlap [8] routine in GSAS which also adjusts the reflection full-width-at-half-maximum (FWHM) values subsequently required by the Monte-Carlo/Simulated Annealing (MC/SA) program (PSSP) described below. These adjustments
in FWHM reveal the additional information provided by the changes in reflection overlap from one pattern to another. In all, 4380 unique reflections were obtained from each profile; for the five data sets, this gave a total of 21,900 reflections and a merged R-value of 7.23%.

The human lysozyme model was altered with the CCP4 program chainsaw [12] in preparation for the molecular replacement trials. This algorithm took information from a Jalview [13] pair-wise sequence alignment of the human and hen egg white lysozymes and truncated the human model. One extra residue was deleted and 51 residues were truncated to their last common atom. This ensured that the model retained minimal bias toward inaccurate side chain information.

We initially attempted to solve the molecular replacement problem using the extracted intensities and conventional single-crystal diffraction software in the CCP4 suite [12]. Trials involving the amore [14] and phaser [15] algorithms either rendered nonsensical solutions (replete with severe symmetry mate clashes) or simply failed to converge. A single trial using the beast [16] program ran for two days without converging. Conversely, the molrep [17] function ran successfully, with several seemingly worthwhile solutions of rotation and translation functions—contrast values were around 95, where a contrast above 2.5 is reported to be definitely a solution. Upon further analysis, most were found to be false positives. The solutions displayed severe clashes with crystallographic symmetry mates, which is geometrically unfeasible. Occasionally molrep would produce the correct solution, but with such high contrast scores for all solutions, there was no discrimination between correct and incorrect solutions.

Thus, it was determined that the single-crystal diffraction software was not equipped to deal with intensities, which become increasingly ambiguous with increasing 2θ (resulting from the overlaps inherent to powder data). This was in contrast to previous studies utilizing multi-analyzer diffractometer data, which can be better treated as single-crystal data at low resolution. Therefore, we shifted focus to a locally modified version of the PSSP program [18], which employs a cost function to account for the poor intensity assignments.

**Results and discussion**

PSSP was run using the first 150 reflections (56>d>8Å) from the multi-pattern Pawley refinement and the 934-atom result from the chainsaw truncation of the human model. Six parameters were varied over the course of the simulated annealing protocol—three Euler angles and three coordinates. Rotations were performed over the z-axis (0–360°), the x-axis (0–180°), and the z-axis again (0–360°) corresponding to the Eulerian angles ω, χ, and φ, respectively. A restricted asymmetric unit was chosen to preferentially define an origin point in the tetragonal space group. The parameters were allowed to vary from -1/8 to 1/8, 0 to 1/2, and 1/4 to 3/4 of the unit cell dimensions for the x, y, and z axes, respectively. The simulated annealing began with the “temperature” at 50 and “cooled” in steps of 20% to a final temperature of 0.001°, with 5000 cycles at each temperature. The results of the five trials are shown in table 1.

Similar S values were observed for all five runs, though trial 2 exhibited the lowest S value, and therefore the best solution. Trials 2–5 superimposed nicely with the previously published 194L structure, and all four displayed similar orientations in the unit cell (see figure 1). The first solution only differed from the 194L structure by a 180° rotation. This signified
a different choice of origin (due to the full angular range searched) rather than an erroneous solution, since symmetry contacts remained the same.

Table 1. Human to Hen PSSP results. The S-value is the score for the individual trial. The $\omega$, $\chi$, and $\phi$ are the Euler angles in degrees and the $x$, $y$, and $z$ are the coordinates of the center of mass of the model with the highest score for the trial.

<table>
<thead>
<tr>
<th>Trial</th>
<th>S value</th>
<th>$\omega$</th>
<th>$\chi$</th>
<th>$\phi$</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4704</td>
<td>179.25</td>
<td>174.32</td>
<td>350.37</td>
<td>0.0021</td>
<td>0.2802</td>
<td>0.7266</td>
</tr>
<tr>
<td>2</td>
<td>0.4587</td>
<td>151.29</td>
<td>5.27</td>
<td>200.95</td>
<td>-0.0048</td>
<td>0.2782</td>
<td>0.5223</td>
</tr>
<tr>
<td>3</td>
<td>0.4588</td>
<td>153.23</td>
<td>5.21</td>
<td>199.21</td>
<td>-0.0061</td>
<td>0.2782</td>
<td>0.5234</td>
</tr>
<tr>
<td>4</td>
<td>0.4624</td>
<td>128.23</td>
<td>4.70</td>
<td>224.76</td>
<td>-0.0061</td>
<td>0.2762</td>
<td>0.5221</td>
</tr>
<tr>
<td>5</td>
<td>0.4600</td>
<td>135.66</td>
<td>4.83</td>
<td>217.35</td>
<td>-0.0060</td>
<td>0.2769</td>
<td>0.5226</td>
</tr>
</tbody>
</table>

Figure 1. Superposition of the result of PSSP trial two (black) and the previously published HEWL structure (grey).
Conclusion

The use of the PSSP algorithm combined with the overlap intensity extraction protocol of GSAS allowed us to extend the powder diffraction molecular replacement technique to larger proteins and enabled the use of image plate data. PSSP rendered five out of five correct solutions and provided a robust method for molecular replacement of proteins using powder diffraction data. Further research into larger molecular weight proteins is ongoing.

References


Acknowledgements. Use of the APS was supported by the DOE/OS/BES under contract number W-31-109-ENG-38.