

# Improving the Accuracy of Macromolecular Structure Refinement at 7 Å Resolution

Axel T. Brunger,<sup>1,\*</sup> Paul D. Adams,<sup>2</sup> Petra Fromme,<sup>3</sup> Raimund Fromme,<sup>3</sup> Michael Levitt,<sup>4</sup> and Gunnar F. Schröder<sup>5</sup>

<sup>1</sup>Howard Hughes Medical Institute and Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Fairchild Building, 299 Campus Drive, Stanford, CA 94305, USA

<sup>2</sup>Lawrence Berkeley National Laboratory, One Cyclotron Road, Building 64R0121, and Department of Bioengineering, University of California at Berkeley, Berkeley, CA 94720, USA

<sup>3</sup>Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA

<sup>4</sup>Department of Structural Biology, Stanford University School of Medicine, Fairchild Building, 299 Campus Drive, Stanford, CA 94305, USA

<sup>5</sup>Institute of Complex Systems (ICS-6), Forschungszentrum Jülich, 52425 Jülich, Germany

\*Correspondence: [brunger@stanford.edu](mailto:brunger@stanford.edu)

DOI 10.1016/j.str.2012.04.020

## SUMMARY

In X-ray crystallography, molecular replacement and subsequent refinement is challenging at low resolution. We compared refinement methods using synchrotron diffraction data of photosystem I at 7.4 Å resolution, starting from different initial models with increasing deviations from the known high-resolution structure. Standard refinement spoiled the initial models, moving them further away from the true structure and leading to high  $R_{\text{free}}$ -values. In contrast, DEN refinement improved even the most distant starting model as judged by  $R_{\text{free}}$ , atomic root-mean-square differences to the true structure, significance of features not included in the initial model, and connectivity of electron density. The best protocol was DEN refinement with initial segmented rigid-body refinement. For the most distant initial model, the fraction of atoms within 2 Å of the true structure improved from 24% to 60%. We also found a significant correlation between  $R_{\text{free}}$  values and the accuracy of the model, suggesting that  $R_{\text{free}}$  is useful even at low resolution.

## INTRODUCTION

While increasingly complex macromolecules or assemblies have been successfully crystallized, such crystals often diffract weakly due to limited crystal growth, high crystal mosaicity, or high sensitivity to radiation damage. Underlying causes can be inherent flexibility, inhomogeneity, or disordered solvent components that prove difficult to overcome. Nevertheless, the interpretation of low-resolution diffraction is often desirable as it provides information about the interaction of individual components in the system or insights about large-scale conformational changes between different states of the system. In addition, macromolecular data collection continues to evolve, notably with microdiffraction synchrotron facilities (Sanishvili et al., 2008) and hard X-ray free electron lasers (FEL) (Chapman et al., 2011).

It is a well-known principle in crystallography that the accuracy of determined atomic positions exceeds the resolution limit of the diffraction data. At atomic resolution (around 1.2 Å), this arises from the excluded volumes of atoms: electron cloud repulsion keeps the scattering objects further apart than half the wavelength of the X-ray radiation used (1–2 Å resolution), allowing the centroids of the atomic electron density to be typically determined to better than 0.1 Å accuracy. At moderate resolution (up to about 4 Å), knowledge of the stereochemistry of the system (bond lengths, bond angles, fixed torsion angles, chirality) allows this principle to be applied to the majority of macromolecular crystal structures. At even lower resolution (4–5 Å), DEN refinement (Schröder et al., 2007; Schröder et al., 2010) further extends this principle. New refinement methods based on physical energy functions such as Rosetta (DiMaio et al., 2011), are complementary to DEN refinement, and are expected to further improve the accuracy of low-resolution crystal structures. Other recent methods may also be useful at low resolution, including LSSR in Buster (Smart et al., 2012), external structure restraints or jelly body refinement in REFMAC (Murshudov et al., 2011), restraints in torsion angle space based on a reference model (Headd et al., 2012), and normal mode refinement (Kidera and Go, 1992; Delarue, 2008). It should be noted that the principle of achieving higher accuracy of positional information than the diffraction limit is referred to as “super-resolution” in optical microscopy (Moerner, 2007; Pertsinidis et al., 2010). We have therefore suggested adoption of the same term in X-ray crystallography (Schröder et al., 2010).

Here, we explore whether one can obtain more accurate structures than naively suggested by the minimum Bragg spacing of a crystal that diffracts to around 7 Å resolution. This resolution is close to the determinacy point for backbone torsion angles of protein crystal structures, i.e., it is the resolution at which the number of independent Bragg reflections is equal to the number of backbone torsion angles. This determinacy point relationship (for a derivation, see Table S1 available online; W. A. Hendrickson, personal communication) shows that it is reasonable to expect that the secondary structure and tertiary fold of a macromolecule can be determined at around 7 Å resolution. Furthermore, the average X-ray diffraction intensities of a typical macromolecular crystal structure have a characteristic resolution dependence with a local maximum between 6 and 15 Å that is determined by the fold of the molecule; at lower resolution,

the intensity distribution is dominated by the envelope of the crystallized molecular entity, and at higher resolution it is determined by the packing of atoms with a maximum at around 5 Å. Thus, the determinacy point for backbone torsion angles is close to the local maximum in X-ray diffraction intensity around 7 Å. The coincidence of high-diffraction intensity and determinacy of backbone torsion angles suggests that a reasonable degree of success might be achievable even at such low resolution.

DEN refinement consists of torsion angle refinement interspersed with *B*-factor refinement in the presence of a sparse set of distance restraints that are initially obtained from a reference model (Schröder et al., 2010). Typically, one randomly selected distance restraint is used per atom. The reference model can be simply the starting model for refinement, or it can be a homology or predicted model that provides external information. In this work, the reference model was the search model used for molecular replacement, and only an overall anisotropic *B*-factor refinement was performed as appropriate at very low resolution. During the process of torsion angle refinement with a slow-cooling simulated annealing scheme, the DEN distance restraints were slowly adjusted in order to fit the diffraction data. The magnitude of this adjustment of the initial distance restraints is controlled by an adjustable parameter,  $\gamma$ . The weight of the DEN distance restraints is controlled by another adjustable parameter,  $w_{\text{DEN}}$ . For the success of DEN refinement, it is essential to perform a global search for an optimum parameter pair ( $\gamma$ ,  $w_{\text{DEN}}$ ). Furthermore, for each adjustable parameter pair tested, multiple refinements should be performed with different initial random number seeds for the velocity assignments of the torsion angle molecular dynamics method and different randomly selected DEN distance restraints. The globally optimal model (in terms of minimum  $R_{\text{free}}$ ), possibly augmented by geometric validation criteria, is then used for further analysis. By default, the last two macrocycles of the DEN refinement protocol are performed without any DEN restraints. However, for the low-resolution refinements presented in this paper, the restraints were kept throughout the entire refinement process in keeping with a low ratio of number of observables to number of torsion angle degrees of freedom.

This study was motivated by the recent availability of low-resolution diffraction data of the Photosystem I (PSI) complex collected on a synchrotron light source (the Advanced Light Source, ALS at Lawrence Berkeley National Laboratory, LBL) (Chapman et al., 2011). The synchrotron data were collected on a single crystal and had a limiting resolution of 6 Å, making them comparable to diffraction data obtained at the first hard X-ray FEL light source (the Linac Coherent Light Source, LCLS, at the SLAC National Accelerator Laboratory) with a minimum Bragg spacing of 7.4 Å (limited in resolution by the wavelength of the FEL photons of 6.9 Å used in this study). The availability of a high-resolution ( $d_{\text{min}} = 2.5$  Å) crystal structure of PSI (PDB ID 1jb0) (Jordan et al., 2001) enabled an objective assessment of the accuracy of structures refined by various methods.

Here, we compared DEN refinement of PSI using the ALS diffraction data at 7.4 Å resolution to overall rigid-body refinement, segmented rigid-body refinement, standard refinement consisting of positional minimization, and torsion angle simulated annealing. We also tested combinations of segmented

rigid-body refinement with DEN refinement and with secondary structure and reference model restrained positional minimization. We assessed the performance of the refinements by (1)  $R_{\text{free}}$ , (2) the root mean square difference (rmsd) to the 2.5 Å resolution crystal structure of PSI, and (3) the significance of features observed in difference maps that were not part of the model used for molecular replacement and refinement. We generated a series of initial models with increasing distance to the 2.5 Å resolution crystal structure, all of which produced a molecular replacement solution. DEN refinement performed better than other methods for all initial models. The most powerful protocol was DEN refinement with initial segmented rigid-body refinement. We also found a good correlation between  $R_{\text{free}}$  and model accuracy among DEN refinements with different adjustable parameters, suggesting that cross-validation is useful even at such low resolution.

## RESULTS

### Molecular Replacement with Increasingly Distant Starting Models

We generated a series of starting models, designated “M1” to “M6,” in order to assess the sensitivity of molecular replacement phasing and subsequent refinement to the distance between starting and the 2.5 Å resolution crystal structure PSI (Protein Data Bank [PDB] ID 1jb0). Model M1 was the 2.5 Å resolution crystal structure of PSI itself. Models M2 through M6 were generated by molecular dynamics starting from M1 to give RMS displacements of C $\alpha$  backbone atoms from the 2.5 Å resolution crystal structure of PSI that ranged from 2.2 to 4.3 Å. We tested if these models produce the correct solution with molecular replacement phasing using the diffraction data of PSI collected at the ALS (Chapman et al., 2011) (Table 1). The ALS diffraction data were truncated to 7.4 Å resolution to make them comparable the limiting resolution of the first FEL (LCLS) data set of PSI (Chapman et al., 2011). We refer to these truncated data as the 7.4 Å diffraction data of PSI.

For all models, the correct solution emerged as the only solution produced by Phaser (McCoy et al., 2007) (see [Experimental Procedures](#)) (Figure S1). Thus, all models could have been used for molecular replacement against the 7.4 Å diffraction data of PSI, albeit with a nondefault parameter for Phaser for model M6 (see [Experimental Procedures](#)).

### Overall Comparison of Refinement Methods

The six initial models were subjected to four different refinement methods against the 7.4 Å diffraction data of PSI: (1) overall rigid-body refinement; (2) positional (Cartesian coordinate) minimization, referred to as “standard refinement”; (3) simulated annealing of torsion angles; and (4) DEN refinement as implemented in CNS v1.3 (Schröder et al., 2010). In addition, the most distant model (M6) was also subjected to segmented rigid-body refinement where the PSI protomer was broken up into 12 rigid-body segments that coincided with the 12 protein subunits and associated cofactors. The resulting segment-refined coordinates were further refined with standard refinement, torsion angle refinement, DEN refinement, and “restrained” refinement.

DEN refinement employed the default protocol that is available in CNS v1.3 (Brünger et al., 1998; Schröder et al., 2010), with the

**Table 1. Data and Refinement Statistics**

Space group	P6 <sub>3</sub>		
Unit cell parameters	a = 283.70 Å, b = 283.70 Å, c = 165.29 Å		
Data Collection			
Wavelength (Å)	1.00		
Resolution range (Å)	65.2-6.0		
Number of observations	110202		
Number of unique reflections	18989		
Completeness (%)	99.3 (100) <sup>a</sup>		
Mean I/σ(I)	3.5 (2.9) <sup>a</sup>		
R <sub>merge</sub> on I (%) <sup>b</sup>	44.7 (51.3) <sup>a</sup>		
R <sub>meas</sub> on I (%) <sup>c</sup>	49.4 (55.9) <sup>a</sup>		
Highest resolution shell (Å)	6.32-6.00		
Model and Refinement Statistics for DEN Refinement, Starting with Model M1			
Resolution range (Å)	49-7.4		
No. of reflections (total)	10004	Cutoff criteria	F  > 0
No. of reflections (test)	508	R <sub>cryst</sub>	0.260 <sup>d</sup>
Completeness (%)	99.5	R <sub>free</sub>	0.291 <sup>d</sup>
Ramachandran (% favored)	79		
Ramachandran (% outliers)	10.7		
Stereochemical Parameters			
Bond angle rmsd (°)	1.29		
Bond length rmsd (Å)	0.008		
Average protein isotropic B-factor (Å <sup>2</sup> )	120.9		
Protein	12 chains with a total of 2334 residues		
Chlorophyll	96 (95 Chlorophyll a, 1 Chlorophyll a')		
Beta-carotene	21		
Phylloquinone	2		
1,2-dipalmitoyl-phosphatidyl-glycerole	3		
1,2-distearoyl-monogalactosyl-diglyceride	1		
Ca <sup>2+</sup>	1		
[4Fe-4S] cluster	3 <sup>e</sup>		

<sup>a</sup>Highest resolution shell.<sup>b</sup> $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ .<sup>c</sup> $R_{\text{meas}}$  (redundancy-independent  $R_{\text{merge}}$ ) =  $\sum_{hkl} [(n/(n-1))^{1/2} \sum_j |I_j(hkl) - \langle I(hkl) \rangle|] / \sum_{hkl} \sum_j I_j(hkl)$  (Diederichs and Karplus, 1997).<sup>d</sup> $R = \sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$  where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure factor amplitudes, respectively.  $R_{\text{free}}$  as for  $R$ , but for 5% of the total reflections chosen at random and omitted from refinement.  $R_{\text{cryst}}$  as for  $R$ , but for the remaining 95% of the reflections.<sup>e</sup>Omitted in refined model for validation purposes.

following exceptions: only overall anisotropic  $B$ -factor refinement was carried out instead of restrained group  $B$ -factor refinement and the DEN restraints were kept throughout the process (see [Experimental Procedures](#) for more details). Restrained refinement included both secondary structure and reference model restraints (Headd et al., 2012) as implemented in the program *phenix.refine* (Afonine et al., 2012). We also tried to

refine model M6 with the jelly body method implemented in Refmac (Murshudov et al., 2011). However, our attempts did not result in improved  $R_{\text{free}}$ , and the gap between  $R_{\text{free}}$  and  $R$  significantly increased. Because we are uncertain whether we used the program optimally for this particular low-resolution crystal structure, we refrained from detailed comparisons with Refmac.

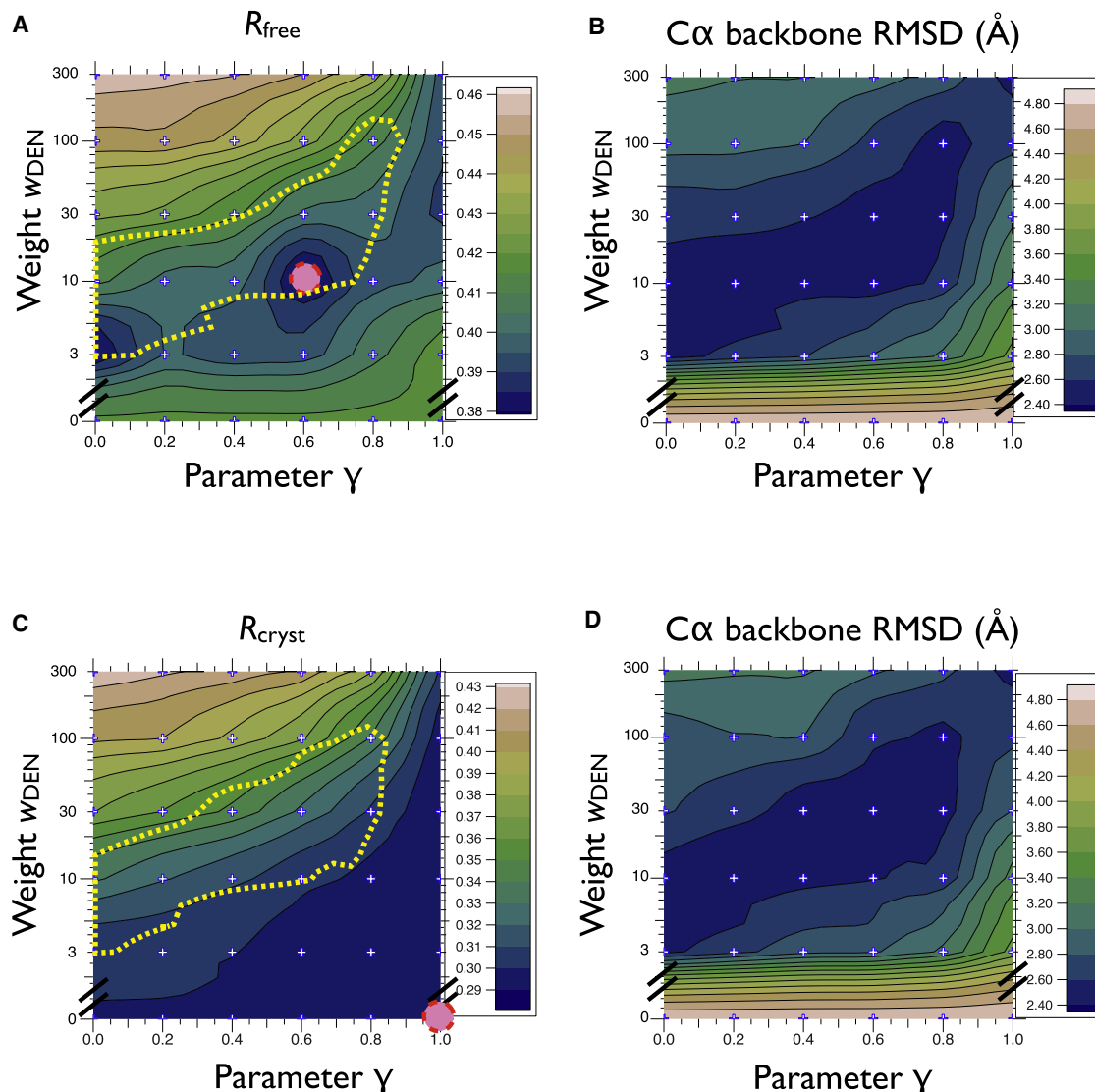
The quality and convergence of the refined models was assessed by  $R_{\text{free}}$  (where smaller values are better),  $C\alpha$  backbone, and chlorophyll  $\text{Mg}^{2+}$  rmsds to the 2.5 Å resolution crystal structure of PSI (smaller is better) and by  $\langle \sigma \rangle$ , the average Z-score (number of standard deviations above the mean of the difference electron density at the positions of the three omitted iron-sulfur clusters—larger is better). Of course, validation with rmsds and difference features was only possible because the high-resolution structure of PSI is known.

DEN refinement consistently performed better than any of the other methods tested as assessed by  $R_{\text{free}}$ , rmsd values, and  $\langle \sigma \rangle$  of the iron-sulfur cluster difference map peaks (Figure S2). The only exception was overall rigid-body refinement starting with model M1 which, by definition, produced rmsd values of zero, whereas the model moved away from M1 upon more extensive refinement, with DEN refinement (refinement statistics in Table 1) producing the smallest deviations (red lines in Figures S2C and S2D). The working  $R$  value ( $R_{\text{cryst}}$ ) was quite similar for all refinement methods that go beyond rigid-body refinement (Figure S2B). In contrast,  $R_{\text{free}}$  showed larger differences between the refined models (Figure S2A), with DEN refinements always achieving the lowest  $R_{\text{free}}$  values. Thus,  $R_{\text{free}}$  correctly indicated that the DEN refined models are generally the most accurate structures as is reflected in the rmsd values between the refined models and the 2.5 Å resolution crystal structure of PSI (Figures S2C and S2D). It should be noted that the relative  $R_{\text{free}}$  ranking of standard refinement and torsion angle simulated annealing is not well correlated with the rmsd values and  $\langle \sigma \rangle$  of the difference peaks. This discrepancy is related to the vastly different number of refined parameters in standard refinement and torsion angle refinement. Thus,  $R_{\text{free}}$  is most powerful when comparing different models using the same refinement method (see next section).

Because we achieved substantial improvements upon refinement of the most distant initial model (M6), we exclusively focus on refinements starting from this model in the following.

### Relation between $R_{\text{free}}$ and Model Accuracy

The relationship between  $R_{\text{free}}$  and model accuracy is shown in Figures 1A and 1B for structures that were refined with the same DEN refinement protocol, but different adjustable parameters ( $\gamma$ ,  $w_{\text{DEN}}$ ). All refinements started from model M6 and were refined against the 7.4 Å diffraction data of PSI. The  $R_{\text{free}}$  contour plot for the best DEN refinement repeats on a two-dimensional ( $\gamma$ ,  $w_{\text{DEN}}$ ) grid is similar to the corresponding contour plot of the  $C\alpha$  backbone rmsd to the 2.5 Å resolution crystal structure of PSI. In striking contrast, when the “best” refinement was selected by the working  $R$  value ( $R_{\text{cryst}}$ ), the resulting structure was very poor: in fact, the  $R_{\text{cryst}}$  and rmsd contour plots are approximately anticorrelated (Figures 1C and 1D). Thus, cross-validation (including  $R_{\text{free}}$ , but also applicable to other quantities such as the commonly used measure for model



**Figure 1.  $R_{free}$  and Corresponding Rmsd to the 2.5 Å Structure of PSI for DEN Refinements Performed against the 7.4 Å Diffraction Data of PSI, Starting from Model M6 with Initial Segmented Rigid-Body Refinement**

Note that the starting model is denoted M6+seg in Figure S2.

(A) The panel shows the lowest  $R_{free}$  value for each parameter pair ( $\gamma$ ,  $w_{DEN}$ ) among 20 repeats; for each parameter pair, we performed 20 repeats of the DEN refinement protocol described in Experimental Procedures. The temperature of the slow-cooling simulated annealing scheme was 3000 K. The  $R_{free}$  value is contoured using values calculated on a  $6 \times 6$  grid (marked by small "+" signs) where the parameter  $\gamma$  was (0.0, 0.2, 0.4, 0.6, 0.8, 1.0) and  $w_{DEN}$  was (0, 3, 10, 30, 100, 300); the results for  $w_{DEN} = 0$  (i.e., torsion angle refinement without DEN restraints) are independent of  $\gamma$ , so the same value was used for all grid points with  $w_{DEN} = 0$ . The contour plot shows minima in the range  $30 \geq w_{DEN} \geq 3$ ; the absolute minimum is at  $w_{DEN} = 10$ ,  $\gamma = 0.6$  (red dashed circle), corresponding to an  $R_{free}$  value of 0.38. In contrast, the lowest  $R_{free}$  value for refinement without DEN restraints ( $w_{DEN} = 0$ ) is only 0.42. The yellow dashed line indicates the region of DEN-refined models with the smallest C $\alpha$  backbone rmsd to the 2.5 Å structure of PSI.

(B) The panel shows the C $\alpha$  backbone rmsd between the refinement repeat that produced the lowest  $R_{free}$  value and the 2.5 Å structure of PS for each of the parameter pairs ( $\gamma$ ,  $w_{DEN}$ ). Note the large rmsd for refinements without DEN restraints ( $w_{DEN} = 0$ ).

(C) The panel shows the lowest  $R_{cryst}$  value for each of the parameter pairs ( $\gamma$ ,  $w_{DEN}$ ) among 20 repeats; the absolute minimum is at  $w_{DEN} = 0$  (red dashed circle). The yellow dashed line indicates the region of DEN-refined models with the smallest C $\alpha$  backbone rmsd to the 2.5 Å structure of PSI.

(D) The panel shows the C $\alpha$  backbone rmsd between the refinement repeat that produced the lowest  $R_{cryst}$  value and the 2.5 Å structure of PSI for each of the parameter pairs ( $\gamma$ ,  $w_{DEN}$ ).  $R_{cryst}$  and the C $\alpha$  backbone rmsd are approximately anti-correlated.

See also Figure S1 and Table S1.

quality,  $\sigma_A$ ) (Read, 1986) produces measures that are indicative of the accuracy of the model if the true structure is yet unknown. In contrast, selection of refined models based on  $R_{cryst}$  can be grossly misleading due to extensive overfitting at low resolution.

As shown previously,  $R_{free}$  is a more objective measure of model quality than  $R_{cryst}$ . (Brünger, 1992), and the results presented in this paper show that this principle also applies to structures refined at around 7 Å resolution.



### Quality of Electron Density Maps

Electron density maps obtained from the different refinement methods are shown in Figure 2. All refinements started from model M6 and were refined against the 7.4 Å diffraction data of PSI. Both standard refinement (Figure 2C) and torsion angle simulated annealing (Figure 2B) moved away from the 2.5 Å resolution crystal structure of PSI, distorted the  $\alpha$  helices, and produced fragmented electron density maps; this poor performance correlated with relatively high  $R_{\text{free}}$  values for these refinements. In contrast, DEN refinement generally produced a well-connected electron density map, even for the rightmost  $\alpha$  helices shown in Figure 2D, demonstrating that electron density maps obtained by DEN refinement can be superior to those from other refinement methods, as has been demonstrated previously at higher resolution (Schröder et al., 2010; Brunger et al., 2012).

Segmented rigid-body refinement produced fragmented electron density maps that do not indicate how to improve the model (Figure 2E). Subsequent torsion angle simulated annealing refinement (Figure 2F) and standard refinement (Figure 2G) produced more-connected electron density maps, but these methods severely distorted the  $\alpha$  helix geometry, as also indicated by the poor Ramachandran statistics for these refinements (Figure S3). In contrast, restrained refinement with initial segmented rigid-body refinement maintained good Ramachandran statistics, but it did not correct the right-most  $\alpha$  helices (Figure 2H). The optimum method was DEN refinement with initial segmented rigid-body refinement; it generally produced a connected electron density map, even for the right-most  $\alpha$  helices, and good  $\alpha$ -helical geometry (Figure 2I).

### Accuracy of Refined Structures

The convergence (or divergence) of the various refined structures to the true structure becomes more apparent in the distribution of individual atomic rmsd values from the 2.5 Å resolution crystal structure of PSI (Figure 3A). The distribution is shifted to smaller values for DEN refinement alone and DEN refinement with initial segmented rigid-body refinement, with a pronounced maximum at 1.2 Å (Figure 3A, red solid lines), compared to the models after overall rigid-body refinement or segmented rigid-body refinement (blue lines). Remarkably, the fraction of atoms within 2 Å of the 2.5 Å resolution crystal structure of PSI improves from 12% to 60% for the combination of segmented rigid-body refinement and DEN refinement (Figure 3B). None of the other tested refinement methods reached this level of accuracy. This shift in the atomic rmsd deviations suggests that structures can be realistically refined beyond rigid-body methods even at around 7 Å resolution. Overall, DEN refinement with initial segmented rigid-body refinement performed best.

### Recovery of Larger Fragments

DEN refinement, and DEN refinement with initial segmented rigid-body refinement, produced structures that were closer to the 2.5 Å resolution crystal structure of PSI than other tested refinement methods and produced more significant difference peaks for the three iron-sulfur clusters, which were omitted for validation purposes (Figure S2E). We next asked the question of whether it would be possible to recover a larger fragment that was not part of the search model. We performed a series

of “omit” refinements against the 7.4 Å diffraction data of PSI with certain  $\alpha$  helices omitted. A particular example is shown in Figure 4, demonstrating that the omitted pair of  $\alpha$  helices (chain F, residues 103–126) is clearly visible in a  $mF_o - DF_o$  difference electron density map when model M1 is refined using DEN (Figure 4A). When the refinement was started from model M6, using DEN refinement with initial segmented rigid-body refinement, there were significant difference peaks in the regions occupied by the  $\alpha$  helices, although the electron density was somewhat fragmented (Figure 4B).

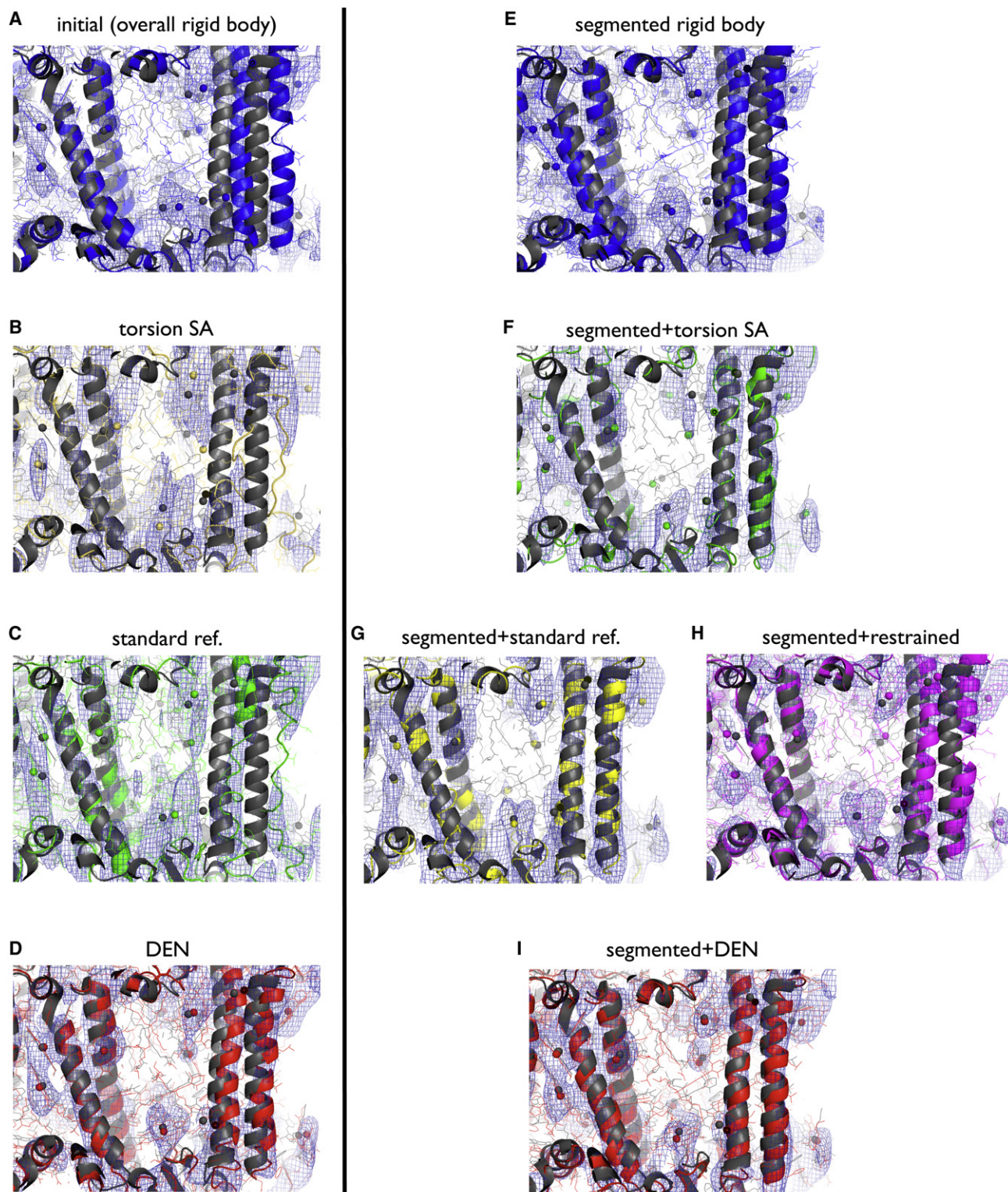
### DISCUSSION

Structure determination and refinement at low resolution remains a grand challenge for X-ray crystallography. The availability of high-flux microbeam synchrotron facilities and, potentially, hard X-ray FELs enables application of X-ray crystallography to ever more challenging biological systems. Such systems may not always give well-diffracting crystals, but may nevertheless provide important biological information even at low resolution. The challenge is to obtain an accurate model that makes use of all available information, including external information such as that from high-resolution structures of individual components of the system, as well as use of advanced physics-based energy functions that together make the problem well-determined. In this paper, we have explored the utility of recently developed reciprocal-space refinement methods, in particular DEN refinement (Schröder et al., 2010) and secondary-structure/reference model restrained refinement (Headd et al., 2012). We used an experimental diffraction data set of PSI at 7.4 Å resolution as the test case, collected at a synchrotron source (ALS).

We find that DEN refinement improves the accuracy of overall and segmented rigid-body refined models. It is remarkable that DEN refinement alone outperforms segmented rigid-body refinement (Figure 3B), although it is of course beneficial to precede DEN refinement with segmented rigid-body refinement. In that case, 60% of the atoms were within 2 Å of the 2.5 Å resolution crystal structure of PSI when the refinement was started from the most distant initial model (M6).

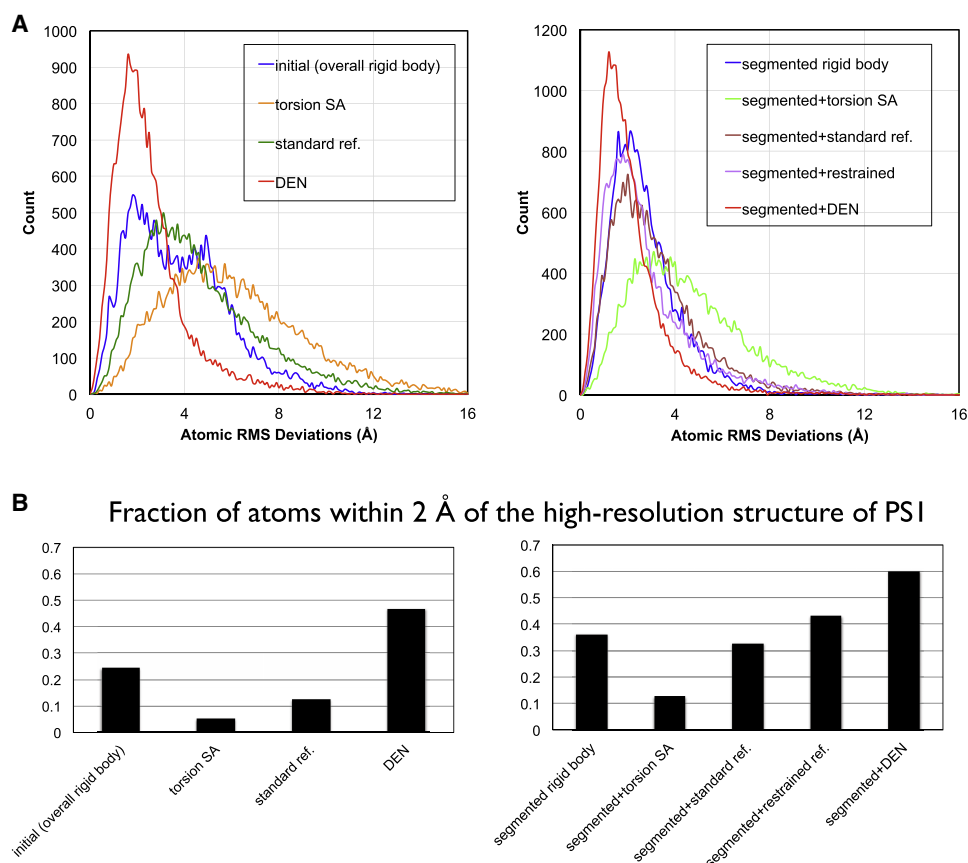
Secondary structure and reference model restrained refinement also led to some improvement when used after initial segmented rigid-body refinement (Figure 3B). However, this improvement was less than that observed for DEN refinement with initial rigid-body refinement. Still, it is interesting that this methodology actually improved the segmented rigid-body refined model in contrast to standard refinement (i.e., without such restraints) that significantly worsened the geometry of the model (Figure S3). Thus, one would expect that combinations of DEN refinement with secondary structure and reference model restrained refinement could lead to further improvements.

DEN refinement works by guiding the refinement path and increasing the chances of obtaining a better model than with standard refinement, and so the imposition of additional information might make the search for a minimum in  $R_{\text{free}}$  even more efficient. However, the imposition of secondary structure restraints is only advisable if the secondary structural elements are conserved between the initial model and the true structure. In fact, this was not the case for the examples studied here: for



**Figure 2. Models and Corresponding  $m2F_o - DF_c$  Electron Density Maps for Specified Refinements against the 7.4 Å Diffraction Data of PSI, Starting from Model M6**

The electron density maps (blue mesh) were calculated with phases from the corresponding refined model and contoured at 1.5  $\sigma$ . The 2.5 Å structure of PSI (PDB ID 1jb0) is shown in dark gray in each of the panels. Spheres indicate  $Mg^{2+}$  ions at the center of the chlorin rings. All nonhydrogen atoms are shown (lines) along with a cartoon representation. The region shown in the figure includes four  $\alpha$  helices (residues 54–100, 155–181, 669–694, and 720–750 of chain A) along with their protein environment and associated cofactors.



**Figure 3. Individual Atomic RMS Deviations to the 2.5 Å Structure of PSI for Specified Refinements against the 7.4 Å Diffraction Data of PSI, Starting from the Model M6**

(A) Histogram of individual atomic RMS deviations between the model refined by the specified method and the 2.5 Å structure of PSI (PDB ID 1jb0).

(B) Fraction of atoms that show RMS deviations less than 2 Å from the 2.5 Å structure of PSI.

See also Figure S3.

example, the right-most  $\alpha$  helix shown in Figure 2 for model M6 has a break that secondary structure restrained refinement cannot overcome (Figure 2H), whereas DEN refinement moves the two  $\alpha$ -helical fragments together so as to converge to the true structure (Figure 2I). This particular example is especially interesting because the DEN restraints have no knowledge of the secondary structure of the high-resolution crystal structure of PSI, so the convergence of this  $\alpha$  helix to the true structure is a consequence of the conformational search that occurs during DEN refinement against the low-resolution diffraction data rather than imposition of some external information. This example is a further demonstration that DEN refinement is

a more general method than rigid-body refinement (or, presumably, normal mode refinement) because, at least in principle, it can achieve any type of conformational change. Clearly, there is room for extension of the method by allowing more general coordinate transformations than the relatively simple interpolation scheme currently used in DEN refinement (Schröder et al., 2010).

Our results show that even at low resolution, around 7 Å, the cross-validation  $R$  value ( $R_{\text{free}}$ ) has predictive power: PSI structures that refine to low  $R_{\text{free}}$  values generally have better accuracy than structures with a high  $R_{\text{free}}$ . In contrast, structures that refine to low working  $R$  values ( $R_{\text{cryst}}$ ) were further away

(A) Initial, overall rigid-body refined model (blue).

(B) Model obtained by torsion angle simulated annealing (yellow).

(C) Model obtained by standard refinement (green).

(D) Model obtained by DEN refinement (red).

(E) Model obtained by segmented rigid-body refinement (blue).

(F) Model obtained by torsion angle simulated annealing with initial segmented rigid-body refinement (green).

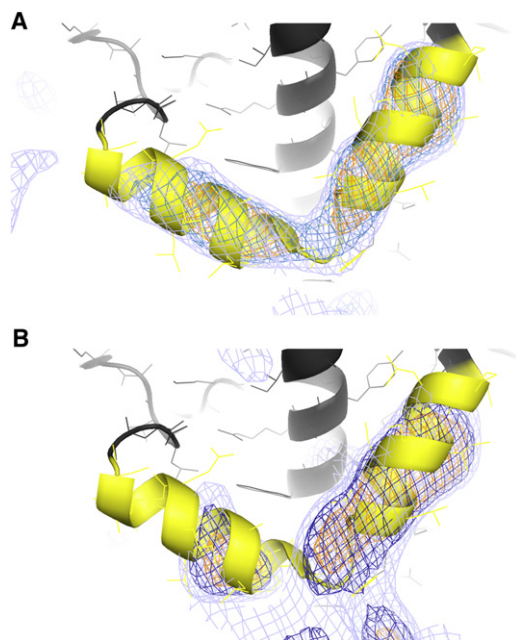
(G) Model obtained by standard refinement with initial segmented rigid-body refinement (yellow).

(H) Model obtained by refinement with secondary structure and reference restraints with *phenix.refine* with initial segmented rigid-body refinement (magenta).

(I) Model obtained by DEN refinement with initial segmented rigid-body refinement (red). Refinement protocols are described in Experimental Procedures.

See also Figure S2.





**Figure 4. Omit DEN Refinement against the 7.4 Å Diffraction Data of PSI**

(A) The initial model was model M1, i.e., the 2.5 Å structure of PSI (PDB ID 1jb0), with a pair of  $\alpha$  helices omitted (chain F, residues 103:126). Shown are  $mF_o-DF_c$  electron density maps at 3  $\sigma$  (orange), 2.5  $\sigma$  (blue), and 2  $\sigma$  (light blue). Note that these two  $\alpha$  helices are located at the detergent-exposed periphery of the PSI complex.

(B) DEN refinement with initial segmented rigid body refinement starting from model M6, with the same  $\alpha$  helix pair omitted, against the 7.4 Å diffraction data of PSI. Shown are  $mF_o-DF_c$  electron density maps at 3  $\sigma$  (orange), 2.5  $\sigma$  (blue), and 2  $\sigma$  (light blue).

from the 2.5 Å resolution crystal structure of PSI (Figure 1). Of course, cross-validation relies on the availability of a sufficient number of reflections that can be omitted for the test set (at least 1,000 reflections are generally advisable) (Brünger, 1997). However, this should not be a problem, because most of the systems that will be studied at low resolution comprise large unit cells and hence have a large number of reflections even at low resolution. We also note that the applicability of  $R_{\text{free}}$  to low-resolution structures suggests that the accuracy of several alternate models (e.g., obtained by different sequence alignments during homology modeling) could be tested by refinement of these candidate models using the same refinement protocol.

In summary, we showed that it is possible to refine structures at around 7 Å resolution using DEN refinement or secondary structure/reference model restrained refinement. In both cases, better convergence to true structure was achieved than possible with segmented rigid-body refinement alone (Figure 3B). For the test case presented here, the optimum protocol is DEN refinement with initial segmented rigid-body refinement.

## EXPERIMENTAL PROCEDURES

### 7.4 Å Diffraction Data of PSI

Synchrotron diffraction data of PSI single crystals were obtained at beam line 8.2.2 at the ALS as described previously (Chapman et al., 2011); these diffraction data were used in that work for comparison to the diffraction data

collected at the LCLS FEL. The synchrotron diffraction data were collected from a single crystal (0.5 × 1 mm) of PSI to about 6 Å resolution at 100 K. The data statistics are provided in Table 1. In order to use a limiting resolution comparable to that of the LCLS data of PSI, the synchrotron diffraction data were truncated to 7.4 Å resolution for molecular replacement and refinement. The maximum likelihood estimate of the overall isotropic component of the  $B$ -factor tensor was 66.5 Å<sup>2</sup> for the synchrotron diffraction data, as obtained by the program *phenix.xtriage* (Zwart et al., 2005). The actual overall isotropic component of the  $B$ -factor tensor upon model refinement was 120.9 Å<sup>2</sup>.

### Generation of Initial Models

Water molecules were removed from the 2.5 Å resolution crystal structure of PSI (PDB ID 1jb0). In addition, the three iron-sulfur clusters were removed from this model for validation purposes. All other cofactors were included (see Table 1 for a list of the cofactors). The resulting model is designated “M1.” This model also serves as the high-resolution comparison model in order to evaluate the performance of the refinements. Five different models were generated by performing simulated annealing molecular dynamics in torsion angle space, using slow-cooling simulated annealing starting at 1800, 2200, 2600, 3000, and 3400 K using a cooling rate of 24 fsec per 50 K. These molecular dynamics calculations included crystal symmetry, but the crystallographic diffraction data were not used. We also included randomly selected pair-wise local distance restraints (about 1 per atom, between 3 and 15 Å) to prevent large excursions, because the molecular dynamics calculations were performed in vacuum at relatively high temperature. The resulting five models are designated “M2,” “M3,” “M4,” “M5,” and “M6.” The resulting C $\alpha$  backbone rmsds to the 2.5 Å resolution crystal structure of PSI were between 2.24 and 4.28 Å.

### Molecular Replacement

Molecular replacement phasing using Phaser (McCoy et al., 2007) was performed starting from the six initial models, M1 through M6, with  $B$ -factors transferred from the 1jb0 crystal structure. The truncated 7.4 Å diffraction data of PSI were used (Table 1). Default settings were used for models M1–M5. In each of these cases a unique solution emerged that coincided with the position and orientation of the high-resolution structure of PSI (taking into account different origin choices). In order to obtain a solution for model M6, the rotation function clustering was turned off. A unique solution then emerged, matching the 1jb0 crystal structure of PSI. For the subsequent refinements, the  $B$ -factors of the corrected placed and oriented models were set to a uniform value of 50 Å<sup>2</sup>. These models served as starting points for all subsequent refinements, respectively.

### Refinement Target Functions

The MLF target function (Pannu and Read, 1996) was used for all refinements. Electron density maps were calculated using  $\sigma_A$  weighting. Maximum likelihood target functions were used as implemented in both CNS and *phenix.refine*.

### Overall Rigid-Body Refinement

Overall rigid-body refinement was performed with CNS v1.3 for each of the six starting models. Eight cycles with 20 steps of conjugate gradient minimization (Powell, 1971) were performed.

### Segmented Rigid-Body Refinement

Each of the 12 protein chains and associated cofactors of a PSI protomer were defined as individual rigid bodies. Eight cycles with 100 steps of conjugate gradient minimization (Powell, 1971) were performed with CNS v1.3. The rigid-body refinement method implemented in *phenix.refine*, which uses a L-BFGS optimization method (Nocedal, 1980), produced similar results; however, it was necessary to use a single resolution zone, i.e., *rigid\_body.number\_of\_zones* was set to 1. The result of the segmented rigid-body refinement was used as a starting point for DEN refinement, standard refinement, torsion angle simulated annealing refinement, and restrained refinement.

### DEN Refinement

The particular initial model was used as both the starting and reference model for DEN refinement (Schröder et al., 2010). For the cases where the initial



model was first subjected to segmented rigid-body refinement, the resulting refined model was used as both the starting and reference model for DEN refinement. The refinement protocol was similar to previous work (Schröder et al., 2010) (as also described in the tutorial for DEN refinement in CNS v1.3, <http://cns-online.org/v1.3/>), with the following non-default settings: only overall anisotropic *B*-factor refinement was carried out instead of restrained group *B*-factor refinement and the DEN restraints were kept throughout the process. In the default protocol, the DEN restraints are turned off during the last two macrocycles. Specifically, eight macrocycles of torsion angle refinement with a slow-cooling simulated annealing scheme were performed in which the first cycle always used  $\gamma = 0$  and the following seven cycles used a specified value for  $\gamma$  (see below).

DEN distance restraints were generated from *N* randomly selected pairs of atoms in the reference model that were separated by 3–15 Å in space; no sequence selection criterion was used. Therefore, distances were drawn from any pair of atoms between any protein chain and cofactor. The value of *N* was chosen to be equal to the number of atoms, so the set of distance restraints was relatively sparse with an average of one restraint per atom. The minimum of the initial DEN potential was set to the coordinates of the particular starting model. We determined the optimum values of the  $\gamma$  and  $w_{\text{DEN}}$  parameters of DEN refinement by a global two-dimensional grid search. At each grid point, twenty refinement repeats were performed with different random initial velocities and different randomly selected DEN distances. We used thirty combinations of six  $\gamma$  values (0.0, 0.2, 0.4, 0.6, 0.8, 1.0) and five  $w_{\text{DEN}}$  values (3, 10, 30, 100, 300). In addition, six different temperatures for the slow-cooling simulated annealing scheme were tested (300, 600, 1000, 1500, 2000, and 3000 K) except in cases of DEN refinement with initial segmented rigid-body refinement, where only 3000 K was used. A representative example of the results of the grid search is shown in Figure 2A. The SBGrid DEN refinement portal (<http://www.sbgrid.org>) was used for most of these refinements. Out of all these resulting models, the one with the lowest  $R_{\text{free}}$  value was used for subsequent analysis.

### Torsion Angle Simulated Annealing

As a control, we performed twenty repeats with  $w_{\text{DEN}} = 0$  at 3000 K. This corresponded to using the refinement protocol without DEN restraints, with results being independent of  $\gamma$ . Out of the resulting models, the one with the lowest  $R_{\text{free}}$  value was used for subsequent analysis.

### Standard Refinement

As a further control, eight macrocycles of 200 steps of conjugate gradient minimization using the L-BFGS optimizer implemented in CNS v1.3 were performed starting from the same models that were used for the DEN refinements. These refinements did not employ DEN restraints.

### Secondary Structure and Reference Restrained Refinement

As an additional control, we performed secondary structure and reference model (Headd et al., 2012) restrained refinement with *phenix.refine* (Afonine et al., 2012). A simulated annealing refinement scheme was used with default control parameters with the exception that a single group *B*-factor was refined for the entire model and no individual atomic displacement parameters were refined and a starting temperature of 5000 K was used for the simulated annealing stage. Additionally, secondary structure restraints (Headd et al., 2012) were automatically determined from the starting model and applied during refinement. Reference model restraints (Headd et al., 2012) were generated from the starting model and used to restrain the model during refinement. A total of three macrocycles of refinement were performed, with simulated annealing performed only in the second macrocycle. The weight on the X-ray term in the refinement ( $w_{\text{xc\_scale}}$ ) was reduced by a factor of two, i.e., the weight was 0.25. Geometric restraints for the ligands in the structure were generated using *phenix.elbow* (Moriarty et al., 2009). Manual modifications were made to the chlorophyll restraints to maintain a planar porphyrin ring geometry.

### Assessment of the Quality and Accuracy of the Refined Models

The various refinement methods were assessed by three criteria:  $R_{\text{free}}$ , rmsd to the 2.5 Å resolution crystal structure of PSI (PDB ID 1Jb0), and the significance of the difference peaks for the three iron-sulfur clusters that were

omitted in the refinement. The  $R_{\text{free}}$  value was used to provide a model-free assessment of the quality of the refined model. The refined models were compared to the 2.5 Å resolution crystal structure of PSI by computing the rmsd for all C $\alpha$  backbone atoms and the rmsd for the Mg<sup>2+</sup> ions of the 96 chlorophyll cofactors; prior to computing the rmsd, the models were least-squares superimposed using the backbone C $\alpha$  atoms to account for possible translation of the model in the z-direction since space group P6<sub>3</sub> has an arbitrary origin choice in the z-direction. For each refined model,  $mF_o - DF_c$  difference maps were computed. For each of the three iron-sulfur clusters,  $\sigma$ , the Z-score (standard deviation above the mean) of the difference electron density was determined and the average of the three  $\sigma$  values calculated as  $\langle \sigma \rangle$ . Because in some cases the refinements had moved, some of the side chains of the four coordinating cysteine residues into the difference density, the CB and SG atoms of these residues were excluded in the calculation of the phases for the difference electron density maps. For the better performing refinements, clear peaks emerged in the difference density maps within the extent of the iron-sulfur clusters; the  $\sigma$  values at these peak positions were used. For some of the poorer performing refinements, no clear peak in the difference density map was found within the extent of an iron-sulfur cluster. In these cases, the significance of the corresponding difference density was estimated by the value of the difference electron density map at the center of the cluster. These procedures were uniformly applied to all refinements.

### Computer Programs Used

MOSFLM (Leslie, 2006) was used for the indexing and integration of the ALS data of PSI. The analysis of diffraction data was performed with the *phenix.xtriage* program (Zwart et al., 2005). The Crystallography and NMR System (CNS) (Brünger et al., 1998) v1.3 was used for DEN refinement, standard (positional minimization) refinement, and torsion angle simulated annealing refinement. *phenix.refine* was used for secondary structure and reference model restrained refinement (Adams et al., 2010; Afonine et al., 2012). PyMOL (DeLano, 2002) was used for molecular illustrations, structure, and electron density map superposition. Molprobity (Chen et al., 2010) was used to calculate the Ramachandran statistics.

### ACCESSION NUMBERS

The low resolution diffraction data set of PSI has been deposited in the PDB (PDB ID 4fe1).

### SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and one table and can be found with this article online at doi:10.1016/j.str.2012.04.020.

### ACKNOWLEDGMENTS

We thank Thomas White and Henry Chapman for stimulating discussions and critical reading of the manuscript, and Corie Ralston for support at beamline 8.2.2 at ALS. A.T.B. acknowledges support by HHMI, M.L. is supported by award GM063817 from NIH, P.D.A. acknowledges support by the US Department of Energy under contract DE-AC03-76SF00098 and NIH/ NIGMS grant P01GM063210, and R.F. and P.F. acknowledge support by the Center for Bio-Inspired Solar Fuel Production, an Energy Frontier Research Center funded by the Department of Energy (DOE), Office of Basic Energy Sciences (award DE-SC0001016). Experiments were carried out the Advanced Light Source, a National User Facilities operated, respectively, by Stanford University and the University of California on behalf of the DOE, Office of Basic Energy Sciences. A.T.B. and P.D.A. performed calculations, analyzed the results, and wrote the paper. R.F. measured and processed the data at beam line 8.2.2 at ALS. G.F.S., M.L., P.F., and R.F. analyzed the results and wrote the paper.

Received: February 17, 2012

Revised: April 5, 2012

Accepted: April 29, 2012

Published: June 5, 2012

## REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* 68, 352–367.
- Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.
- Brünger, A.T. (1997). Free R value: cross-validation in crystallography. *Methods Enzymol.* 277, 366–396.
- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921.
- Brunger, A.T., Das, D., Deacon, A.M., Grant, J., Terwilliger, T.C., Read, R.J., Adams, P.D., Levitt, M., and Schröder, G.F. (2012). Application of DEN refinement and automated model building to a difficult case of molecular-replacement phasing: the structure of a putative succinyl-diaminopimelate desuccinylase from *Corynebacterium glutamicum*. *Acta Crystallogr. D Biol. Crystallogr.* 68, 391–403.
- Chapman, H.N., Fromme, P., Barty, A., White, T.A., Kirian, R.A., Aquila, A., Hunter, M.S., Schulz, J., DePonte, D.P., Weierstall, U., et al. (2011). Femtosecond X-ray protein nanocrystallography. *Nature* 470, 73–77.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21.
- DeLano, W.L. (2002). The Pymol Molecular Graphics System on World Wide Web <http://www.pymol.org>.
- Delarue, M. (2008). Dealing with structural variability in molecular replacement and crystallographic refinement through normal-mode analysis. *Acta Crystallogr. D Biol. Crystallogr.* 64, 40–48.
- Diederichs, K., and Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269–275.
- DiMaio, F., Terwilliger, T.C., Read, R.J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H.L., et al. (2011). Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* 473, 540–543.
- Headd, J.J., Echols, N., Afonine, P.V., Grosse-Kunstleve, R.W., Chen, V.B., Moriarty, N.W., Richardson, D.C., Richardson, J.S., and Adams, P.D. (2012). Knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. D Biol. Crystallogr.* 68, 381–390.
- Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., and Krauss, N. (2001). Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* 411, 909–917.
- Kidera, A., and Go, N. (1992). Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. *J. Mol. Biol.* 225, 457–475.
- Leslie, A.G.W. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* 62, 48–57.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674.
- Moerner, W.E. (2007). New directions in single-molecule imaging and analysis. *Proc. Natl. Acad. Sci. USA* 104, 12596–12602.
- Moriarty, N.W., Grosse-Kunstleve, R.W., and Adams, P.D. (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* 65, 1074–1080.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* 67, 355–367.
- Nocedal, J. (1980). Updating quasi-newton matrices with limited storage. *Math. Comput.* 35, 773–782.
- Pannu, N.S., and Read, R.J. (1996). Improved structure refinement through maximum likelihood. *Acta Crystallogr. A* 52, 659–668.
- Pertsinidis, A., Zhang, Y., and Chu, S. (2010). Subnanometre single-molecule localization, registration and distance measurements. *Nature* 466, 647–651.
- Powell, M.J.D. (1971). On the convergence of a variable metric algorithm. *J. Inst. Math. Appl.* 7, 21–36.
- Read, R.J. (1986). Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. *Acta Crystallogr. A* 42, 140–149.
- Sanishvili, R., Nagarajan, V., Yoder, D., Becker, M., Xu, S., Corcoran, S., Akey, D.L., Smith, J.L., and Fischetti, R.F. (2008). A 7 μm mini-beam improves diffraction data from small or imperfect crystals of macromolecules. *Acta Crystallogr. D Biol. Crystallogr.* 64, 425–435.
- Schröder, G.F., Brunger, A.T., and Levitt, M. (2007). Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* 15, 1630–1641.
- Schröder, G.F., Levitt, M., and Brunger, A.T. (2010). Super-resolution biomolecular crystallography with low-resolution data. *Nature* 464, 1218–1222.
- Smart, O.S., Womack, T.O., Flensburg, C., Keller, P., Paciorek, P., Sharff, A., Vornheim, C., and Bricogne, G. (2012). Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* 68, 368–380.
- Zwart, P.H., Grosse Kunstleve, R.W., and Adams, P.D. (2005). Characterization of X-ray data sets. *CCP4 Newsletter* 42, contribution 8.