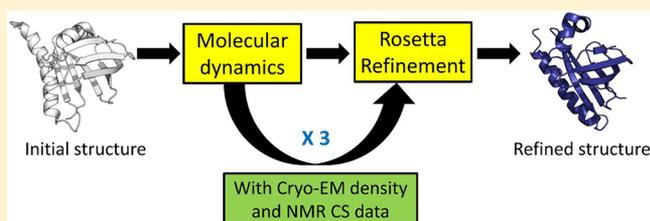# Using NMR Chemical Shifts and Cryo-EM Density Restraints in Iterative Rosetta-MD Protein Structure Refinement

Sumudu P. Leelananda and Steffen Lindert*

Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States

S *Supporting Information*

ABSTRACT: Cryo-EM has become one of the prime methods for protein structure elucidation, frequently yielding density maps with near-atomic or medium resolution. If protein structures cannot be deduced unambiguously from the density maps, computational structure refinement tools are needed to generate protein structural models. We have previously developed an iterative Rosetta-MDFF protocol that used cryo-EM densities to refine protein structures. Here we show that, in addition to cryo-EM densities, incorporation of other experimental restraints into the Rosetta-MDFF protocol further improved refined structures. We used NMR chemical shift (CS) data integrated with cryo-EM densities in our hybrid protocol in both the Rosetta step and the molecular dynamics (MD) simulations step. In 15 out of 18 cases for all MD rounds, the refinement results obtained when density maps and NMR chemical shift data were used in combination outperformed those of density map-only refinement. Notably, the improvement in refinement was highest when medium and low-resolution density maps were used. With our hybrid method, the RMSDs of final models obtained were always better than the RMSDs obtained by our previous protocol with just density refinement for both medium (6.9 Å) and low (9 Å) resolution maps. For all the six test proteins with medium resolution density maps (6.9 Å), the final refined structure RMSDs were lower for the hybrid method than for the cryo-EM only refinement. The final refined RMSDs were less than 1.5 Å when our hybrid protocol was used with 4 Å density maps. For four out of the six proteins the final RMSDs were even less than 1 Å. This study demonstrates that by using a combination of cryo-EM and NMR restraints, it is possible to refine structures to atomic resolution, outperforming single restraint refinement. This hybrid protocol will be a valuable tool when only low-resolution cryo-EM density data and NMR chemical shift data are available to refine structures.

INTRODUCTION

Being able to determine or predict protein structure is of paramount importance since protein structure determines protein function.[1,2] We also rely on macromolecular structure information in structure-based drug design.[3−5] With known catalytic site structural information, drugs can be designed to bind target proteins implicated in diseases to regulate function and relive symptoms.[4,6,7] Experimental methods of structure determination have made tremendous contributions to our understanding of protein structure. X-ray crystallography and NMR have historically been the two most popular and widely used experimental methods of structure determination.[8,9] The majority of the protein structures available in the PDB are X-ray crystal structures. In NMR spectroscopy, distance restraints of proximal atoms are obtained and have been utilized to elucidate almost 13 000 protein structures.[10] Cryo-electron microscopy (cryo-EM) is another increasingly popular technique to determine macromolecular structures.[11−15] Cryo-EM is continuing to revolutionize the field of structural biology and was also worthy of a recent noble prize award.[11,16,17]

Despite their successes, there are some caveats associated with experimental structure determination as well. Experimental methods of protein structure determination are expensive and time-consuming and have technical difficulties associated with them.[18] For example, in order to obtain X-ray diffraction patterns of a protein, that protein first needs to be in crystal form,[18] which can be challenging particularly for flexible proteins. NMR is mostly limited to small proteins although there are exceptions.[19] In cryo-EM, flexible protein regions are commonly not resolved in the densities and it is still oftentimes challenging to obtain side chain atom coordinates. For example, receptor binding pocket side-chain residues are frequently not explicitly defined in cryo-EM density maps.[20] Therefore, in order to supplement experimental structure determination, computational methods are valuable tools for prediction of protein structure and function. Computational tools are used routinely in understanding protein structure and in bridging the sequence—structure gap that is continuously increasing.[21]

Some computational methods are designed to predict structures solely based on sequence (*ab initio* predictions). Some of the software that is commonly used in *ab initio* structure prediction are Rosetta, I-TASSER, QUARK, and

MODELLER.[22−25] Experimental restraints are used extensively in protein structure prediction and refinement.[26−29] Many different types of experimental restraints have been used. Cryo-EM densities,[30−36] NMR chemical shifts,[37,38] Nuclear Overhauser Effect (NOE) data,[39,40] electron paramagnetic resonance (EPR),[41,42] small-angle X-ray scattering,[43] mass spectrometry,[44−48] and fluorescence energy transfer microscopy (FRET)[49] have successfully been used as restraints in structure prediction and refinement. Using sparse experimental data can significantly improve protein structure prediction and refinement methods.

Two methods that can use sparse experimental data in structure prediction and refinement are molecular dynamics simulations and Rosetta. Rosetta is a macromolecular structure modeling software[50,51] and one of the most popular tools used for structure refinement using cryo-EM experimental restraints.[36,52] It can use sparse experimental data to predict and refine protein structures.[42,53] Cryo-EM density data is used and the regions that least agree with the density maps are identified and rebuilt in an iterative fashion.[35,36,52,54] Notably, Rosetta can use NMR data in refinement of structures in conjunction with other types of experimental restraints in its structure refinement protocols.[55] Molecular dynamics (MD) simulations are another computational method frequently used in structure refinement.[56,57] Restraints obtained from experiments can be used in these calculations to bias the simulations toward native-like structures. One such experimental data used in molecular dynamics simulations is cryo-EM density data. Cryo-EM density data was first incorporated in molecular dynamics simulations by the Schulten group.[58] In these molecular dynamics simulations with flexible fitting (MDFF), the cryo-EM density map is converted to an external potential which is added in addition to the standard force field in molecular dynamics simulations. This additional potential function guides the model toward the density map structure. NMR chemical shift data has also been incorporated into molecular dynamics simulations, for example by PLUMED.[59] In PLUMED an additional force field term is introduced based on the agreement with a set of experimental chemical shift data.[59] Unlike in MDFF where the cryo-EM density is used as the additional force field in the molecular dynamics simulations, in PLUMED, the difference between the measured and the experimental chemical shift data is used as a linear bias potential to direct the simulations.

In our previous protein structure refinement protocols, we only used cryo-EM density data in refining protein structures.[60−62] Here, we explored the simultaneous use of cryo-EM density maps and NMR chemical shift data in refining a set of protein structures in an integrated fashion. Cryo-EM density maps and NMR chemical shifts contain complementary information. Previously, these two experimental methods have been integrated to derive atomic structures for the HIV-1 capsid protein C-terminal domain.[39] In this work, both of these experimental restraints were integrated and applied at the same time to molecular dynamics simulations. Here we went beyond the use of MD simulations by employing an iterative approach where we utilized cryo-EM density maps and NMR chemical shift data in Rosetta refinement and molecular dynamics simulations. Our findings show that the presence of multiple orthogonal experimental restraints further improved the quality of refined structures.

■ METHODS

***Ab Initio* Protein Model Building.** We selected six soluble proteins (5NPG, 2N5B, 2L8O, 2N2T, 2MZJ, and 5T1N) with NMR structures deposited in the PDB databank (Table 1).

Table 1. Proteins Used in This Work[a]

PDB ID	protein name	no. of residues	starting RMSD (Å)
5NPG	*Drosophila melanogaster* Loquacious dsRBD1	83	4.98
2N5B	Yeast Thioredoxin	103	4.93
2L8O	Chr148 from Cytophaga hutchinsonii	144	6.56
2N2T	de novo designed protein target OR303	84	5.22
2MZJ	Nop6-RBD from S. Cerevisiae	82	4.98
5T1N	phosphocarrier protein NPr	85	5.12

[a]The names of the proteins, the number of residues, and the ab initio starting model RMSDs are shown. The starting RMSDs are in the range of 4−7 Å. All proteins have less than 150 residues.

The starting structures that were used as input into the refinement protocol were generated using Rosetta *ab initio* model building (in the absence of any experimental restraints).[63] None of the structures had any missing residues. The number of residues in the starting structures ranged from 82 to 144. A total of 5000 models for each protein sequence were generated using Rosetta *ab initio* model building without using any experimental restraints. Rosetta *ab initio* builds protein structures starting from the sequence by assembling small protein fragments obtained from the PDB.[63−65] Out of the 5000 models generated for each protein sequence, the lowest Rosetta all atom energy model was selected as the input starting model for our refinement protocols.

RMSDs of the starting structures generated by Rosetta were calculated using the BCL::Quality program[66] using all backbone atoms. The RMSDs of the models selected for each protein were in the 4−7 Å range from the native structure. These starting structures agreed with the native structures in their overall topology and in the arrangement of the secondary structure elements, however, some of the backbone and most of the side chain orientations were not accurate. This starting structure RMSD range is generally the same range that one would get using manual model building into low resolution cryo-EM density data. Alternatively, the starting structures could be generated by combining techniques such as Rosetta, SSEHunter, and EM-Fold.[30−32] The RMSDs of these initial models mimicked the models that are frequently built using other *ab initio* model building tools and tracing of cryo-EM density maps. The starting structure information and RMSDs are shown in Table 1.

Simulation of Density Maps. Density maps of all six proteins were generated using the Situs package.[67] Specifically, Pdb2vol was used to generate volumetric maps of the native protein structures. The voxel spacing used was one-third of the desired resolution of the simulated density map. A Gaussian smoothing kernel with amplitude 1 was used. Density maps at three different resolutions were generated (4, 6.9, and 9 Å) using the native NMR representative structures. Voxel spacings of 1.3, 2.3, and 3 Å, respectively, were used for 4, 6.9, and 9 Å resolution density maps. Higher resolution maps (4 Å) showed backbone details of structures as well as some of the side chain

details. Medium resolution (6.9 Å) and low-resolution maps (9 Å) did not resolve side chain and backbone coordinates.

Cryo-EM Density and Chemical Shift Guided Rosetta-MD (Hybrid Protocol). The initial starting models for the protocols were obtained using Rosetta *ab initio* model building without using any experimental restraints. In the new hybrid protocol three iterative rounds of Rosetta-MDFF were performed guided by both CS data and cryo-EM density data (Figure 1).
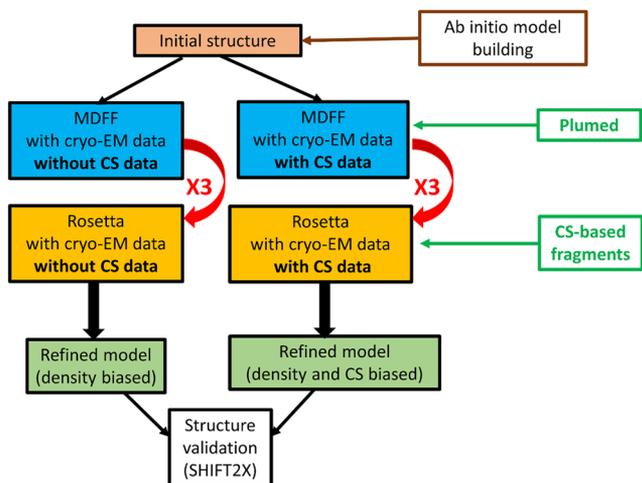

Figure 1. Schematic diagram illustrating the two structure refinement protocols used in this study. The initial models that were input to the protocols were obtained using Rosetta *ab initio* model building without using any experimental restraints. In the first protocol which is the same protocol used in our previous publication, three rounds of Rosetta-MDFF were run iteratively without using any chemical shift information. Only cryo-EM density data was used in this refinement. In the new hybrid protocol, three rounds of Rosetta-MDFF were performed using both chemical shift data and cryo-EM density data. The final models were obtained after the last round of Rosetta and validated by their chemical shift agreement using the ShiftX2 program. The chemical shift data was used in molecular dynamics simulations with the PLUMED program which was integrated into the MD flexible fitting algorithm. In the Rosetta steps of the protocol, chemical shifts were used in picking the fragments in the model building step.

After each round of MDFF simulation the final frame obtained was input into the next Rosetta refinement step. Out of all the models generated during the Rosetta step, a model was picked based on the fit of the model to the density map. This model was used as input into the next MDFF step. After three Rosetta-MDFF iterations, the final models were obtained after the third round of the Rosetta refinement step. CS-based fragments were used in the Rosetta model building steps. The final models obtained after the third round of Rosetta were validated by their chemical shift agreement using the ShiftX2 program. The chemical shift data was used in molecular dynamics simulations using the PLUMED program that was integrated into MD flexible fitting algorithm. Different MDFF simulations corresponding to different PLUMED weights were run in order to incorporate chemical shift data. From all the simulations corresponding to different PLUMED weights, the best fit-to-density final model out of all final frames was used in the next Rosetta round (explained in more detail in the Results and Discussion section). For comparison, the protocol was also

run without using any NMR chemical shift information exactly as in our previous publication. Only cryo-EM density data was used in this refinement.[60]

Experimental chemical shifts for backbone and side chain heavy atoms (C_α, C_β, C, and N) and backbone hydrogens ($C_\alpha-H$, $N-H$) were used. NMR chemical shift data for each of the proteins was obtained from the PDB databank. If experimental chemical shift data for a certain residue was missing, these residues were ignored when calculating chemical shifts and the data for these atoms were not considered. For chemical shift biasing, the simulations were guided by the reference experimental chemical shifts toward the experimental native reference structure. After each time step, the chemical shifts were calculated using *Almost* (CS2BACKBONE).[59] The difference between the measured and the experimental shift (current deviation) was applied as a linear potential with a slope (hereafter called the PLUMED weight) using PLUMED.[59]

The MD simulations were biased using both the experimental chemical shift data and the simulated cryo-EM density maps. The chemical shift biased molecular dynamic simulations were performed using PLUMED.[59] Cryo-EM density data was incorporated using molecular dynamics flexible fitting (MDFF).[68] In MDFF a potential corresponding to the density map is applied during the simulations to guide the structure toward the density map structure. In our hybrid protocol, chemical shift-based biasing and the potential corresponding to the density map were both integrated and applied at the same time. The density map was converted into a potential function that can be read by the MDFF simulations using *mdff griddx*. Particle Mesh Ewald (PME) was used with periodic boundary conditions in each simulation. A local interaction distance cutoff of 12 Å was used for van der Waals and electrostatic calculations. The simulation *pairlistdist* and *nonbondedFreq* were set to 13.5 and 1 Å. In our previous publications we showed that the default scaling factor of 0.2 does well overall for most of the test cases.[60] Therefore, we used the default density scaling factor of 0.2 in this study. For molecular dynamics simulations NAMD 2.10 with CHARMM22 force field was used. The MDFF package *ssrestraints* was used to apply restraints to enforce that secondary structures were maintained. All simulations were run at a 300 K simulation temperature for 1 ns.

The simulations were run at the following PLUMED bias weights: 0.0001, 0.0005, 0.001, 0.005, 0.01, and 0.05. After each simulation, the last frame was extracted. The fit-to-density of each of the last frames was calculated using Rosetta density-tools,[54] to measure the model-map agreement for all the different PLUMED bias weight simulations. In previous work we observed a correlation of high fit-to-density values and native-like structures.[60] The best fit-to-density model (corresponding to one PLUMED bias weight) was used in the subsequent Rosetta step in each case.

During the Rosetta refinement step, regions that least agree with the density map were identified and rebuilt. Amino acid sequence fragments (three-residue and nine-residue) were generated using the Robetta server based on the fasta sequence of the target protein.[69] The chemical shift information for the target protein was also used as input when generating fragment files.[70] These fragments were subsequently used in Rosetta ab initio structure prediction. Usage of chemical shift-filtered fragments led to a targeted reduction of the protein structure search space.

During each Rosetta step of the hybrid protocol a total of 5000 independent models were generated. RMSDs of each model were again calculated using BCL::Quality program. The fit-to-density score of each model was calculated using Rosetta density tools. This score gives an indication of how well the models agree with the density map of the native structures. After each Rosetta round the best fit-to-density model was obtained as input to the next MDFF round. Rosetta refinement was performed three times iteratively. After the third round of Rosetta the model picked was considered the final refined model. For comparison purposes, the initial structures were also refined using our previous cryo-EM density guided refinement protocol without using any NMR chemical shift data.[60]

■ RESULTS AND DISCUSSION

In this work, we used a combination of cryo-EM density maps and experimental NMR chemical shift (CS) data as restraints in computational protein structure refinement. Restraints from both methods were used in Rosetta and molecular dynamics simulation steps to refine the protein structures. We tested our iterative protocol on six proteins where experimental NMR chemical shift data was available and additionally guided the simulations by cryo-EM density maps of varying resolutions (4, 6.9, and 9 Å). We iteratively applied molecular dynamics and Rosetta refinement over the course of three rounds. Chemical shift data was incorporated into Rosetta in the fragment building step. CS data was also used in molecular dynamics simulations by integrating it directly into the MD flexible fitting algorithm. For each molecular dynamics simulation, two files were given as input for restraining the simulations; the cryo-EM density and NMR chemical shift data. For comparison, we also repeated the protocol without using any NMR chemical shift information, exactly as we did in our previous publication using only cryo-EM density.

During each MD simulation step, six PLUMED weights (0.0001, 0.0005, 0.001, 0.005, 0.01, and 0.05) were used to run six different simulations. The cryo-EM density potential was always applied while running simulations. The simulations tended to get unstable when PLUMED weights of 0.1 or higher were used with the density potential. After each simulation, the fit-to-density of the final frames of the six simulations (corresponding to different PLUMED weights) were calculated. We picked the best model based on the fit-to-density score. Table 2 shows an example of the effect of

Table 2. RMSD and Fit-to-Density of the Final Frame Obtained When Different PLUMED Weights Were Used for 2L8O, MDFF3 Round[a]

PLUMED weight	RMSD (Å)	fit-to-density
0 (no PLUMED)	4.26	0.886
0.0001	1.71	0.923
0.0005	1.77	0.918
0.001	1.9	0.915
0.005	1.75	0.919
0.01	1.92	0.92
0.05	1.71	0.922
0.1	1.84	0.921

[a]The highest fit-to-density (0.922) corresponded to the lowest RMSD structure (1.71 Å) and showed a significant improvement compared to the density only refinement of 2L8O (RMSD 4.26 Å).
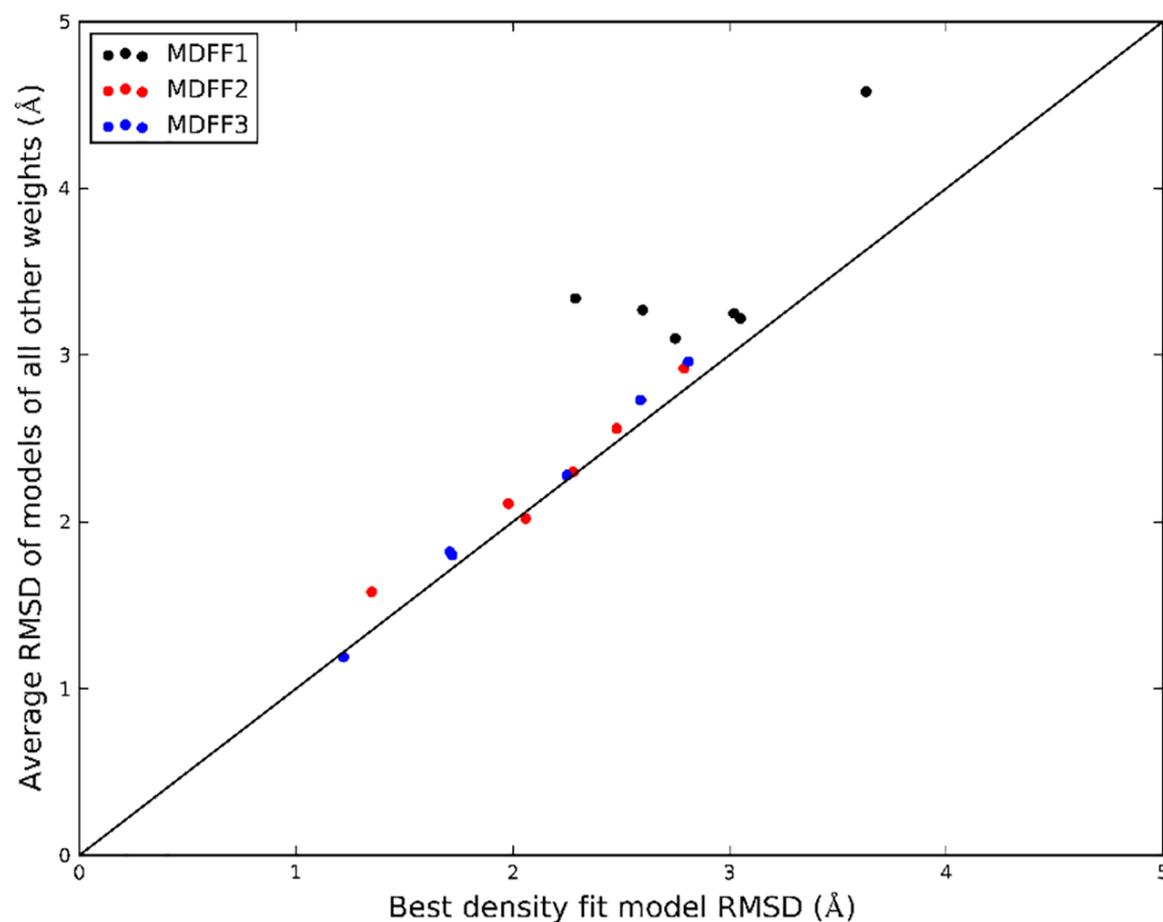
different PLUMED weights on protein RMSDs and the fit-to-density scores for the last frame in the final round of MDFF for 2L8O. At a PLUMED weight of 0.05, the highest fit-to-density of 0.92 corresponded to the lowest RMSD structure (1.71 Å). A significant improvement was obtained compared to the density-only refinement of 2L8O which yielded a model with an RMSD of 4.26 Å. The average RMSD of all other last frames (generated by nonzero PLUMED weights that did not yield the best fit-to-density) of the simulations corresponding to each PLUMED weight was compared to the best fit-to-density weight RMSD for all benchmark proteins (Figure 2). For MDFF1, MDFF2, and MDFF3 the final frame corresponding to the best fit-to-density was significantly better than the averaged model RMSD obtained for all other PLUMED weights. Out of all the MDFF rounds for all benchmark proteins, the best RMSD model was picked using the fit-to-density score in 89% of the time. There were only two cases for which the model selected by the best fit-to-density score had a slightly higher RMSD than the average of the other models. Even for these two cases this difference in RMSD was not significant. Hence, we used the fit-to-density to select the simulation (i.e., PLUMED weight) that best agreed with the NMR chemical shift data.

To test the effect of the combined NMR/cryo-EM guided refinement, we also performed three rounds of iterative refinement guided by cryo-EM density maps only for comparison. The RMSDs of the last frame of MDFF1, MDFF2, and MDFF3 simulations when only cryo-EM density maps were used in refinement and when cryo-EM density and NMR chemical shift data were both used in refinement were compared. In each of these steps, 5/6 times using NMR chemical shift data in addition to cryo-EM data yielded better RMSD models (Figure 3). On average, an RMSD improvement of 3.11 Å was observed when a dual NMR/cryo-EM-guided refinement was performed at a map resolution of 6.9 Å. We examined three different density map resolutions (4, 6.9, and 9 Å) and the ability of NMR chemical shift data to supplement the density map during refinement. Figure 4 shows the RMSDs for the final models for all three density map resolutions where the refinement was guided by either cryo-EM data or combined cryo-EM/NMR data.

Independent of whether or not chemical shift data was used for the refinement, the RMSDs of the final models obtained by the 4 Å density map refinement were the lowest and those obtained using the 9 Å density resolution were the highest (Figure 4). When NMR chemical shift restraints were used in addition to near-atomic resolution (4 Å) density data, the final structure RMSDs did not generally improve compared to the cryo-EM-only refinement. In fact, guiding the refinement by NMR chemical shift data frequently slightly worsened the final model RMSD when near-atomic resolution density maps were used. For 5NPG, 2N2T, 2MZJ, 2N5B, and 5T1N, we saw an average increase in RMSD of 0.3 Å. However, for 2L8O even with a 4 Å resolution density map, the final model was significantly improved when NMR data was used for the refinement. In the case of 2L8O, when NMR chemical shifts were not used, the alpha helices unwound during the refinement, causing RMSDs to the native structure of above 4 Å. However, when NMR chemical shifts were present, there was significant improvement with the final model for all density resolutions (Figure 4). When medium resolution (6.9 and 9 Å) density maps were used in the refinements, the final model RMSDs when both cryo-EM data and NMR chemical

Figure 2. RMSDs of the models picked based on the best fit-to-density of the model and the average model RMSDs of all other nonzero PLUMED weights for the 6 benchmark proteins. The RMSDs are shown for all the MDFF rounds (MDFF1, MDFF2, and MDFF3). Most of the time the best fit-to-density model is significantly better than the RMSDs corresponding all the other PLUMED weights.
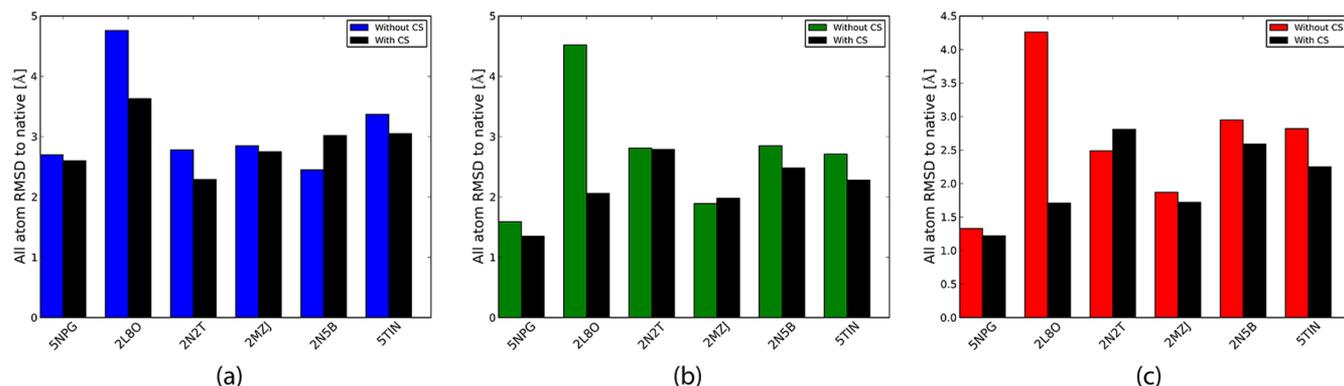

Figure 3. RMSDs of the last frame of MDFF1 (a), MDFF2 (b), and MDFF3 (c) simulations when only cryo-EM densities were used in refinement (blue, green, and red, respectively) and when cryo-EM density and NMR chemical shift data were both used in refinement (black). The density map resolution used was 6.9 Å. For each of these MDFF rounds, 5/6 times using NMR chemical shift data in addition to cryo-EM data helped identify a better RMSD model.

shifts were used (hybrid protocol) were better for all proteins than when only cryo-EM density maps were used. Using NMR chemical shift data in addition to the density maps yielded an average RMSD improvement of 0.55 Å for the 6.9 Å resolution density maps and an average RMSD improvement of 1.23 Å for the 9 Å resolution density maps.

The model RMSDs decreased gradually with each step of the protocol, as exemplified for 2N5B in Figure 5a and 2L8O in Figure 6a. A similar improvement of RMSDs was observed

with our earlier membrane protein structure refinement protocol as well.[60] Here, we observed that a combination of NMR chemical shift information with cryo-EM density maps, generally led to a greater improvement in RMSDs than compared with a cryo-EM-only refinement.

There was a clear drop in RMSDs after the first iteration for all six proteins regardless of the density map resolution for both the hybrid protocol and the cryo-EM only refinement. After the first iteration, the RMSDs further dropped slightly
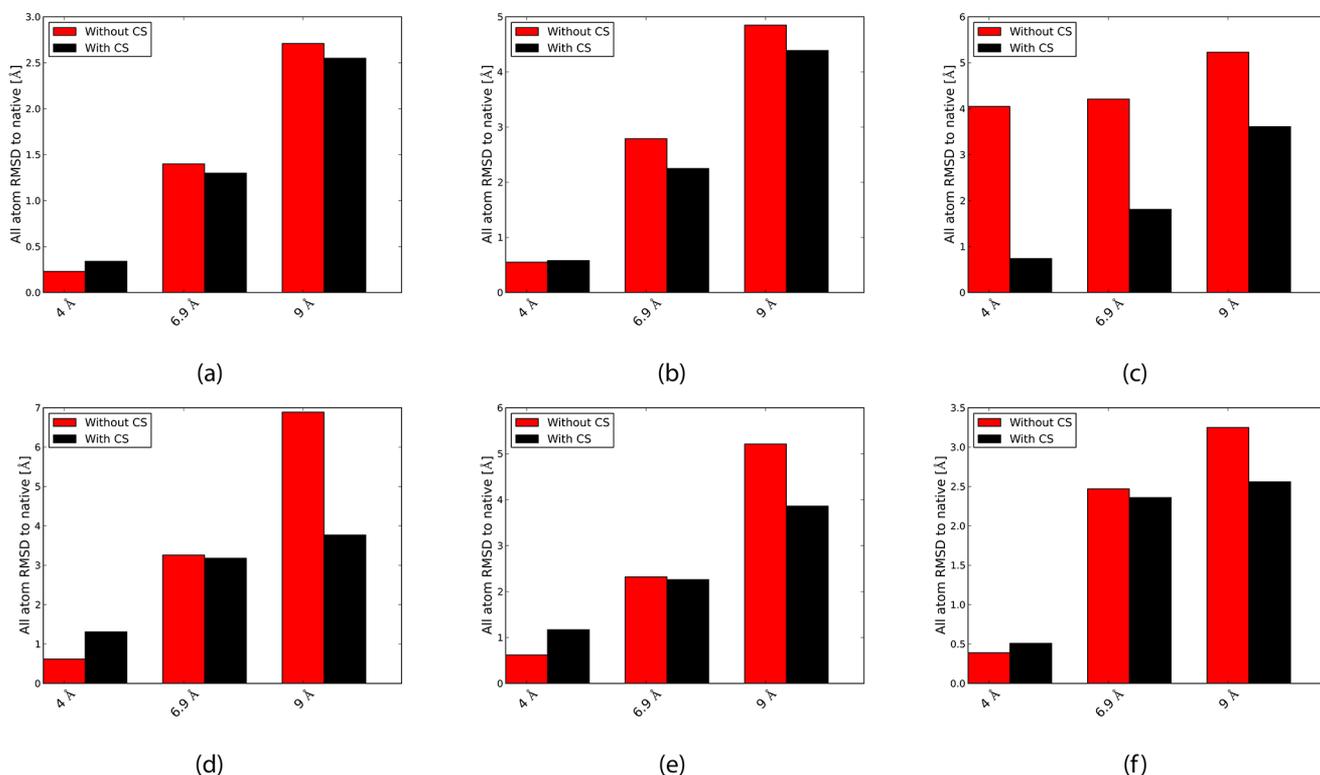
Figure 4. RMSDs after the final Rosetta round obtained for three different density map resolutions (4, 6.9, and 9 Å) are shown. (a) 5NPG. (b) 2N5B. (c) 2L8O. (d) 2N2T. (e) 2MZJ. (f) 5TIN. The results are shown for pure cryo-EM density refinement (red) and the integrated cryo-EM and NMR chemical shift refinement (black).
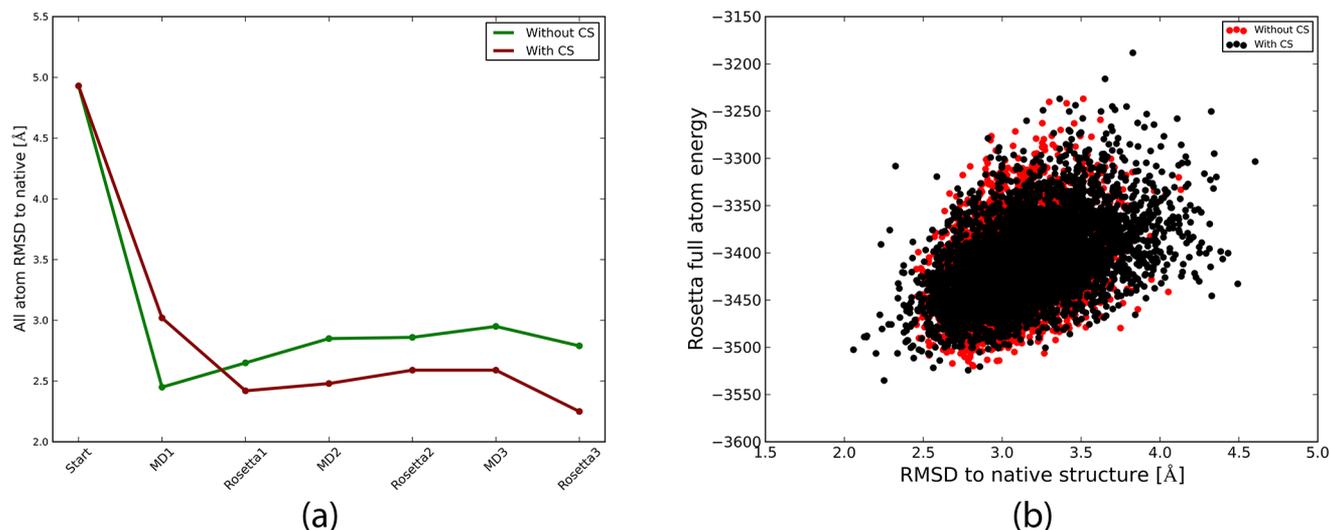

Figure 5. (a) RMSD variation with each step of the iterative Rosetta-MD protocol for 2N5B when 6.9 Å density maps were used and (b) model Rosetta energy score vs the RMSD to the native for the last Rosetta round.

during the next two iterations of Rosetta-MDFF for the hybrid protocol that used CS data. This general trend was observed for all six test proteins. For example, for the last round of 2L8O Rosetta refinement with a 6.9 Å density map, RMSDs for the hybrid method were in the range of ~1.8−3.2 Å (Figure 6b), compared to an RMSD range of ~3.6−5.2 Å for the cryo-EM only refinement. For 2N5B, the difference in the range of RMSDs for the models generated by the hybrid method and the cryo-EM only refinement was not that distinguishable (Figure 5b). However, with the hybrid method, models with

RMSDs of less than 2.4 Å were generated, while this was not observed for the cryo-EM only refinement. In general, use of the hybrid method allowed us to build models with low RMSDs which were never generated with the cryo-EM only refinement.

Root mean square error (RMSE) of C_α chemical shifts at each time frame of the molecular dynamics simulations were calculated using shiftx2. For the six proteins, the final RMSEs were lower than the starting RMSEs when the hybrid protocol was used. For 5NPG, 2L8O, and 5T1N the average decrease in
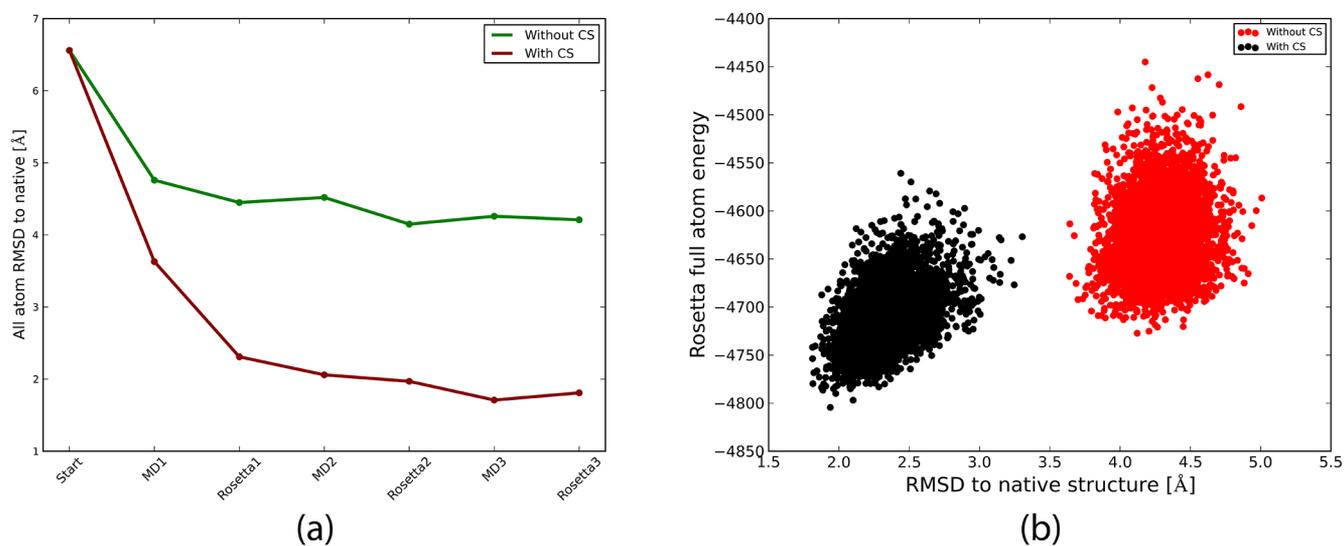
Figure 6. (a) RMSD variation with each step of the iterative Rosetta-MD protocol for 2L8O when 6.9 Å density maps were used and (b) model Rosetta energy score vs the RMSD to the native for the last Rosetta round.
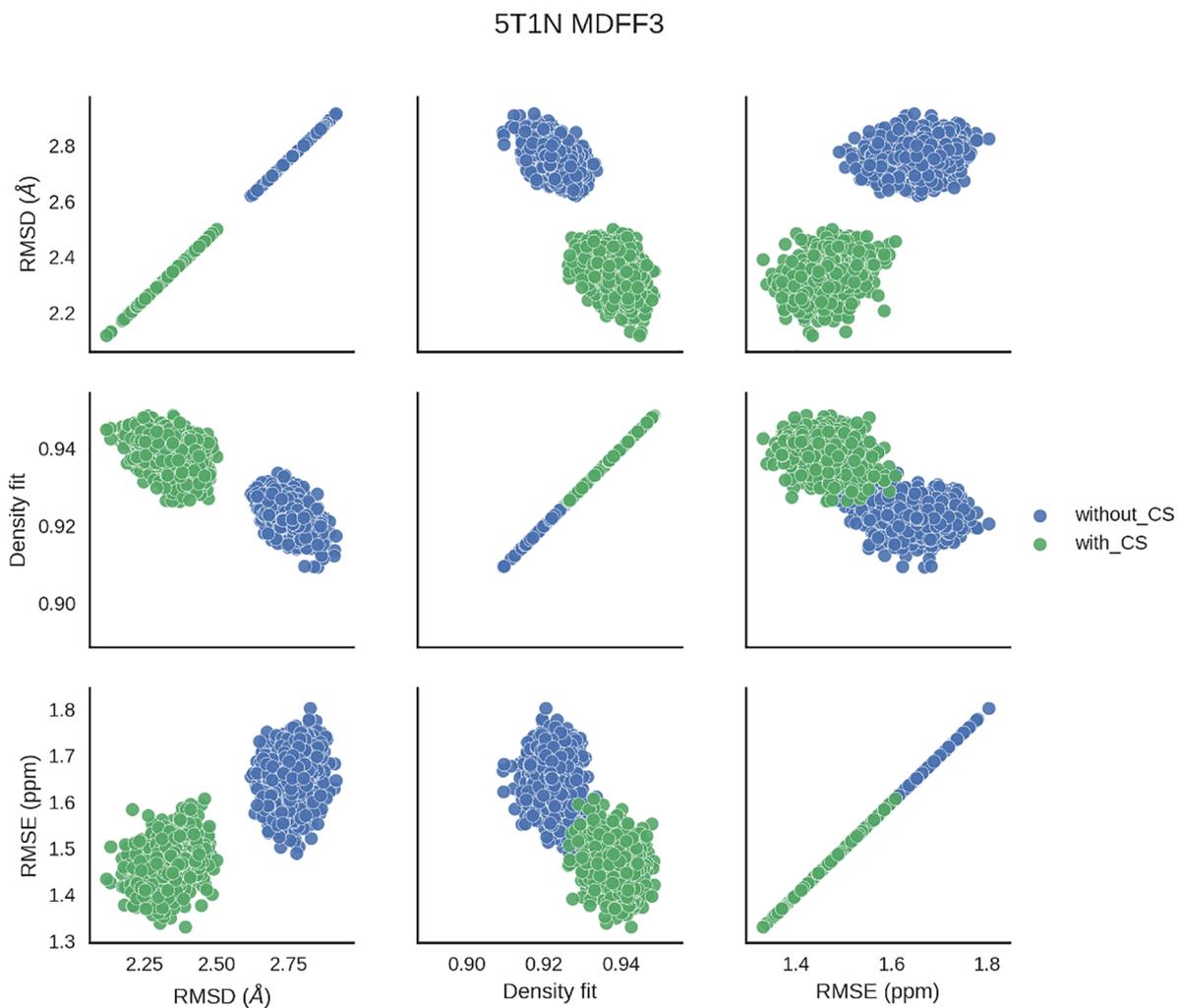

Figure 7. Variations of RMSD, RMSE, and fit-to-density of models for the last round of MDFF for 5T1N. Models obtained from the hybrid refinement protocol are shown in green, while models that were refined using only the cryo-EM density map are shown in blue.
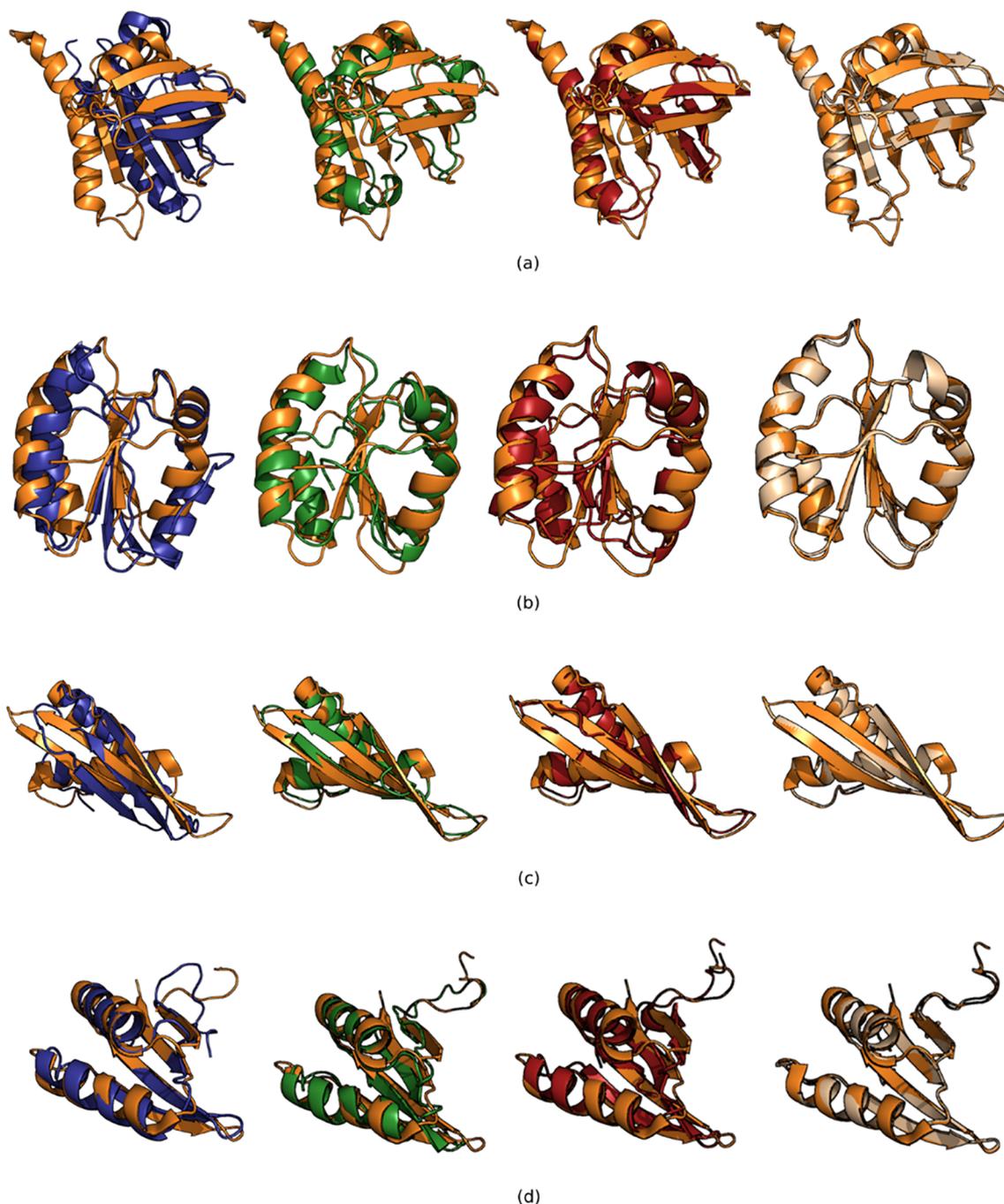
Figure 8. Native structure (orange), starting structure (blue), refined structure without NMR chemical shift data at 6.9 Å resolution density maps (green), refined structure with both NMR chemical shift data and cryo-EM density maps at 6.9 Å resolution (red) and refined structure with both NMR chemical shift data and cryo-EM density maps at 4 Å resolution (tan). (a) 2L8O. (b) 2N5B. (c) 5NPG. (d) 2N2T.

C_α chemical shift RMSE of the final MDFF3 simulations for the hybrid protocol was 0.45 ppm (Figure S1). However, in the case of refinement by just cryo-EM density data the final RMSEs did not show the trend seen for the hybrid refinement. An average increase of 0.51 ppm was observed. For 5NPG, 2L8O, 2N2T, 2MZJ, and 5TIN, the RMSEs increased with the simulation time. For 2N5B, the RMSE slightly decreased but was still considerably higher than what was obtained for the hybrid protocol (1.61 vs 1.29 ppm).

The variation of RMSE, fit-to-density and RMSD were further analyzed for final MDFF round models. We found strong correlations between RMSE, fit-to-density, and RMSD. For example, low RMSD models exhibited low chemical shift RMSEs, suggesting that those agreed with the experimental shifts of the native structures (Figure 7). The low RMSD models also showed high fit-to-density scores, indicating these models agreed with the experimental cryo-EM density data better. All three qualifiers were generally better for models obtained from the hybrid refinement protocol. Figure 7 exemplifies these trends for 5T1N models with the hybrid method that used cryo-EM and chemical shift data and the cryo-EM only protocol.

We calculated the RMSEs separately for each of the secondary structure classes (helix, sheet, and loop) using DSSP.[71] The DSSP database includes the secondary structure assignment for the proteins deposited in the PDB. The RMSEs of each structural class were lower when chemical shift data was incorporated than when only cryo-EM density maps were used in structure refinement (Figure S2). The average RMSEs for each class showed that helices were the best refined in terms of NMR chemical shift agreement. The average C_α chemical shift RMSE for helices with and without CS data was 0.78 and 0.93 ppm, respectively. The highest RMSEs were observed for sheets. The average C_α chemical shift RMSEs for sheets with and without NMR chemical shift refinement were 1.22 and 1.31 ppm, respectively (Table S1). Additionally, the largest improvement of chemical shifts upon usage of the hybrid method was observed for helices.

The final refined structures with the hybrid protocol showed better agreement with the native structures than structures generated with cryo-EM density data as the only refinement restraint (Figure 8). When medium resolution density maps (6.9 Å) guided the refinement for all the six proteins, the final refined structure RMSDs were lower for the hybrid method than for the cryo-EM only refinement. The final model RMSDs for 5NPG, 2L8O, 2N2T, 2MZJ, 2N5B, and 5TIN were 1.30, 1.81, 3.18, 2.26, 2.25, and 2.36 Å, respectively, when the hybrid method with medium resolution density data was used. For comparison, when only cryo-EM density data was used in the refinement, final model RMSDs were 1.40, 4.21, 3.26, 2.32, 2.79, 2.47 Å, respectively. The best final RMSDs were obtained for the 4 Å refinement. The final refined model RMSDs for the hybrid protocol using density maps of 4 Å resolution for 5NPG, 2L8O, 2N2T, 2MZJ, 2N5B, and 5TIN were 0.34, 0.74, 1.31, 1.17, 0.58, and 0.51 Å, respectively. For four out of the six proteins the final model RMSDs were less than 1 Å. For the nonhybrid method the average RMSD improvement for 4, 6.9, and 9 Å refinement was 4.22, 2.55, and 0.61 Å, respectively. With the hybrid method the average RMSD improvement for 4, 6.9, and 9 Å refinement was 4.52, 3.11, and 1.84 Å, respectively. Therefore, the addition of CS data had the greatest impact when 9 Å resolution density maps were used.

CONCLUSIONS

In our previously reported iterative Rosetta-MDFF protocol, we used cryo-EM density maps for refinement of protein structures. Here we showed that using additional experimental restraints in conjunction with cryo-EM density maps we can further improve the refined structures. As an additional experimental restraint we used NMR chemical shift data integrated with cryo-EM density maps in our hybrid protocol in both the Rosetta and MD steps. In the density-guided Rosetta refinement steps, the NMR chemical shift data was employed in the selection of protein fragments for the model building. In density-guided molecular dynamics (MDFF) refinement, the chemical shift data were incorporated to guide the simulations using the PLUMED program. These reference chemical shifts were used to guide the simulations toward the experimental native reference structure. After each time step, the chemical shifts were calculated using *Almost* (CS2BACKBONE). The difference between the measured and the experimental shift (current deviation) was applied as a linear potential with a slope.

In 15 out of 18 cases for all MDFF rounds, the refinement results obtained when both density maps and NMR chemical shift data were used outperformed those of the density map-only refinement. The improvement in refinement was highest when maps of higher resolutions were used. For medium (6.9 Å) and low (9 Å) resolution maps, with our hybrid method, the RMSDs of the final models were always lower than the RMSDs obtained by our previous protocol using just density-guided refinement. When we used near-atomic resolution density maps (4 Å) for the refinement, we frequently built sub 2 Å models without using the NMR chemical shift data. Additionally, at this stage, addition of the chemical shifts did not improve the refinements and actually decreased the model quality for almost all the test cases except 2L8O (Figure 4). This suggests that cryo-EM density maps at around 4 Å resolution contain sufficient structural restraints for a successful refinement even without NMR chemical shift data. Addition of NMR chemical shift data improved the refinement in cases where the cryo-EM density maps did not contain sufficient structural restraints for an unambiguous refinement to atomic resolution. As such, a combination of NMR chemical shift data and medium resolution cryo-EM density maps proved to be most useful.

This study demonstrates that by using a combination of cryo-EM and NMR restraints it is possible to refine structures to atomistic resolution that is not attainable using only one type of restraint. This hybrid protocol will be a valuable tool when only low to medium resolution cryo-EM density data and NMR chemical shift data are available to refine structures. Instructions for running the protocols are provided in the Supporting Information.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b00932.

Instructions on the protocol and how to generate simulated density maps are provided. The variation of RMSE with each time step of the final round of MDFF (MDFF3) for each of the six proteins when the hybrid protocol with 6.9 Å density data was used is shown. The NMR chemical shift RMSEs for each secondary class type (helix, sheet, and loop) for the final MDFF round for the test proteins and the average C_α chemical shift RMSEs for each of the secondary structure classes are shown (PDF)

Chemical shift data files (CAshifts.dat, CBshifts.dat, Cshifts.dat, HAshifts.dat, Hshifts.dat, Nshifts.dat), parameter files to calculate chemical shifts (camshift.db, par_all36_cgenff.prm) and the pdb coordinates of the example protein (template.pdb) (ZIP)

AUTHOR INFORMATION

Corresponding Author

*Mailing address: Department of Chemistry and Biochemistry, Ohio State University, 2114 Newman & Wolfrom Laboratory, 100 W. 18th Avenue, Columbus, OH 43210. Office: 614-292-8284. Fax: 614-292-1685. E-mail: lindert.1@osu.edu.

ORCID Ⓘ

Steffen Lindert: 0000-0002-3976-3473

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Abbot, E. S. The Causal Relations between Structure and Function in Biology. *Am. J. Psychol.* **1916**, 27 (2), 245−250.

(2) Lee, D.; Redfern, O.; Orengo, C. Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell Biol.* **2007**, 8, 995.

(3) Leelananda, S. P.; Lindert, S. Computational methods in drug discovery. *Beilstein J. Org. Chem.* **2016**, 12, 2694−2718.

(4) Schmidt, T.; Bergner, A.; Schwede, T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today* **2014**, 19 (7), 890−897.

(5) Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **2003**, 10 (9), 787−797.

(6) Brooijmans, N.; Kuntz, I. D. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, 32 (1), 335−373.

(7) Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S. H. Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. *AAPS J.* **2012**, 14 (1), 133−141.

(8) Ilari, A.; Savino, C. Protein Structure Determination by X-Ray Crystallography. In *Bioinformatics: Data, Sequence Analysis and Evolution*, Keith, J. M., Ed.; Humana Press: Totowa, NJ, 2008; pp 63−87.

(9) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* **2010**, 18 (8), 923−933.

(10) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104 (23), 9615.

(11) Cheng, Y. Single-particle cryo-EM at crystallographic resolution. *Cell* **2015**, 161 (3), 450−457.

(12) Egelman, E. H. The Current Revolution in Cryo-EM. *Biophys. J.* **2016**, 110 (5), 1008−1012.

(13) Murata, K.; Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim. Biophys. Acta, Gen. Subj.* **2018**, 1862 (2), 324−334.

(14) Carroni, M.; Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods (Amsterdam, Neth.)* **2016**, 95, 78−85.

(15) Scott, H.; Huang, W.; Bann, J. G.; Taylor, D. J. Advances in structure determination by cryo-EM to unravel membrane-spanning pore formation. *Protein Sci.* **2018**, 27 (9), 1544−1556.

(16) Bai, X.-c.; McMullan, G.; Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **2015**, 40 (1), 49−57.

(17) Raunser, S. Cryo-EM Revolutionizes the Structure Determination of Biomolecules. *Angew. Chem., Int. Ed.* **2017**, 56 (52), 16450−16452.

(18) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem., Int. Ed.* **2003**, 42 (24), 2718−2736.

(19) Früh, D. P.; Goodrich, A.; Mishra, S.; Nichols, S. NMR methods for structural studies of large monomeric and multimeric proteins. *Curr. Opin. Struct. Biol.* **2013**, 23 (5), 734−739.

(20) García-Nafría, J.; Lee, Y.; Bai, X.; Carpenter, B.; Tate, C. G. Cryo-EM structure of the adenosine A(2A) receptor coupled to an engineered heterotrimeric G protein. *eLife* **2018**, 7, No. e35946.

(21) Rost, B.; Sander, C. Bridging the Protein Sequence-Structure Gap by Structure Predictions. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, 25 (1), 113−136.

(22) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. y.; Pieper, U.; Sali, A. *Current protocols in bioinformatics* **2006**, 15, 5.6.1.

(23) Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinf.* **2008**, DOI: 10.1186/1471-2105-9-40.

(24) Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **2004**, 383, 66−93.

(25) Xu, D.; Zhang, Y. Toward optimal fragment generations for ab initio protein structure assembly. *Proteins: Struct., Funct., Genet.* **2013**, 81 (2), 229−239.

(26) Mierke, D. F.; Huber, T.; Kessler, H. Coupling constants again: Experimental restraints in structure refinement. *J. Comput.-Aided Mol. Des.* **1994**, 8 (1), 29−40.

(27) Li, W.; Zhang, Y.; Kihara, D.; Huang, Y. J.; Zheng, D.; Montelione, G. T.; Kolinski, A.; Skolnick, J. TOUCHSTONEX: Protein structure prediction with sparse NMR data. *Proteins: Struct., Funct., Genet.* **2003**, 53 (2), 290−306.

(28) Kovalevskiy, O.; Nicholls, R. A.; Long, F.; Carlon, A.; Murshudov, G. N. Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta crystallographica. Section D, Structural biology* **2018**, 74 (3), 215−227.

(29) Lindert, S.; Stewart, P. L.; Meiler, J. Hybrid approaches: applying computational methods in cryo-electron microscopy. *Curr. Opin. Struct. Biol.* **2009**, 19 (2), 218−225.

(30) Lindert, S.; Staritzbichler, R.; Wötzel, N.; Karakaş, M.; Stewart, P. L.; Meiler, J. EM-Fold: De Novo Folding of α-Helical Proteins Guided by Intermediate-Resolution Electron Microscopy Density Maps. *Structure* **2009**, 17 (7), 990−1003.

(31) Lindert, S.; Alexander, N.; Wötzel, N.; Karakaş, M.; Stewart, P. L.; Meiler, J. EM-Fold: De Novo Atomic-Detail Protein Structure Determination from Medium-Resolution Density Maps. *Structure* **2012**, 20 (3), 464−478.

(32) Baker, M. L.; Ju, T.; Chiu, W. Identification of Secondary Structure Elements in Intermediate-Resolution Density Maps. *Structure* **2007**, 15 (1), 7−19.

(33) Lindert, S.; Hofmann, T.; Wötzel, N.; Karakaş, M.; Stewart, P. L.; Meiler, J. Ab initio protein modeling into CryoEM density maps using EM-Fold. *Biopolymers* **2012**, 97 (9), 669−677.

(34) Frenz, B.; Walls, A. C.; Egelman, E. H.; Veesler, D.; DiMaio, F. RosettaES: a sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nat. Methods* **2017**, 14 (8), 797−800.

(35) Wang, R. Y.-R.; Song, Y.; Barad, B. A.; Cheng, Y.; Fraser, J. S.; DiMaio, F. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **2016**, 5, No. e17219.

(36) DiMaio, F.; Song, Y.; Li, X.; Brunner, M. J.; Xu, C.; Conticello, V.; Egelman, E.; Marlovits, T. C.; Cheng, Y.; Baker, D. Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **2015**, 12 (4), 361−365.

(37) Weiner, B. E.; Alexander, N.; Akin, L. R.; Woetzel, N.; Karakas, M.; Meiler, J. BCL::Fold−protein topology determination from limited NMR restraints. *Proteins: Struct., Funct., Genet.* **2014**, 82 (4), 587−595.

(38) Meiler, J.; Baker, D. Rapid protein fold determination using unassigned NMR data. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100 (26), 15404−15409.

(39) Perilla, J. R.; Zhao, G.; Lu, M.; Ning, J.; Hou, G.; Byeon, I.-J. L.; Gronenborn, A. M.; Polenova, T.; Zhang, P. CryoEM Structure Refinement by Integrating NMR Chemical Shifts with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2017**, 121, 3853.

(40) Papaleo, E.; Camilloni, C.; Teilum, K.; Vendruscolo, M.; Lindorff-Larsen, K. Molecular dynamics ensemble refinement of the heterogeneous native state of NCBD using chemical shifts and NOEs. *PeerJ* **2018**, 6, No. e5125.

(41) Alexander, N.; Bortolus, M.; Al-Mestarihi, A.; McHaourab, H.; Meiler, J. De Novo High-Resolution Protein Structure Determination

from Sparse Spin labeling EPR Data. *Structure (Oxford, U. K.)* **2008**, *16* (2), 181−195.

(42) Hirst, S. J.; Alexander, N.; McHaourab, H. S.; Meiler, J. RosettaEPR: An Integrated Tool for Protein Structure Determination from Sparse EPR Data. *J. Struct. Biol.* **2011**, *173* (3), 506−514.

(43) Shevchuk, R.; Hub, J. S. Bayesian refinement of protein structures and ensembles against SAXS data using molecular dynamics. *PLoS Comput. Biol.* **2017**, *13* (10), No. e1005800.

(44) Kahraman, A.; Herzog, F.; Leitner, A.; Rosenberger, G.; Aebersold, R.; Malmström, L. Cross-link guided molecular modeling with ROSETTA. *PLoS One* **2013**, *8* (9), No. e73411.

(45) Aprahamian, M. L.; Lindert, S. Utility of Covalent Labeling Mass Spectrometry Data in Protein Structure Prediction with Rosetta. *J. Chem. Theory Comput.* **2019**, *15* (5), 3410−3424.

(46) Harvey, S. R.; Seffernick, J. T.; Quintyn, R. S.; Song, Y.; Ju, Y.; Yan, J.; Sahasrabuddhe, A. N.; Norris, A.; Zhou, M.; Behrman, E. J.; Lindert, S.; Wysocki, V. H. Relative interfacial cleavage energetics of protein complexes revealed by surface collisions. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (17), 8143.

(47) Aprahamian, M. L.; Chea, E. E.; Jones, L. M.; Lindert, S. Rosetta Protein Structure Prediction from Hydroxyl Radical Protein Footprinting Mass Spectrometry Data. *Anal. Chem.* **2018**, *90* (12), 7721−7729.

(48) Seffernick, J. T.; Harvey, S. R.; Wysocki, V. H.; Lindert, S. Predicting Protein Complex Structure from Surface-Induced Dissociation Mass Spectrometry Data. *ACS Cent. Sci.* **2019**, *5*, 1330.

(49) Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **2012**, *9* (12), 1218−1225.

(50) Kaufmann, K. W.; Lemmon, G. H.; DeLuca, S. L.; Sheehan, J. H.; Meiler, J. Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You. *Biochemistry* **2010**, *49* (14), 2987−2998.

(51) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In *Methods in Enzymology*; Johnson, M. L., Brand, L., Eds.; Academic Press: 2011; Vol. 487, Chapter 19, pp 545−574.

(52) DiMaio, F.; Tyka, M. D.; Baker, M. L.; Chiu, W.; Baker, D. Refinement of Protein Structures into Low-Resolution Density Maps using Rosetta. *J. Mol. Biol.* **2009**, *392* (1), 181−190.

(53) Pilla, K. B.; Leman, J. K.; Otting, G.; Huber, T. Capturing Conformational States in Proteins Using Sparse Paramagnetic NMR Data. *PLoS One* **2015**, *10* (5), No. e0127053.

(54) DiMaio, F.; Chiu, W. Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps. *Methods Enzymol.* **2016**, *579*, 255−276.

(55) Ovchinnikov, S.; Park, H.; Kim, D. E.; Liu, Y.; Wang, R. Y.-R.; Baker, D. Structure prediction using sparse simulated NOE restraints with Rosetta in CASP11. *Proteins: Struct., Funct., Genet.* **2016**, *84* (S1), 181−188.

(56) Feig, M.; Mirjalili, V. Protein structure refinement via molecular-dynamics simulations: What works and what does not? *Proteins: Struct., Funct., Genet.* **2016**, *84*, 282−292.

(57) Lupala, C. S.; Rasaeifar, B.; Gomez-Gutierrez, P.; Perez, J. J. Using molecular dynamics for the refinement of atomistic models of GPCRs by homology modeling. *J. Biomol. Struct. Dyn.* **2018**, *36*, 2436.

(58) Trabuco, L. G.; Villa, E.; Schreiner, E.; Harrison, C. B.; Schulten, K. Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **2009**, *49* (2), 174−180.

(59) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604−613.

(60) Leelananda, S. P.; Lindert, S. Iterative Molecular Dynamics−Rosetta Membrane Protein Structure Refinement Guided by Cryo-EM Densities. *J. Chem. Theory Comput.* **2017**, *13* (10), 5131−5145.

(61) Lindert, S.; Meiler, J.; McCammon, J. A. Iterative Molecular Dynamics- Rosetta Protein Structure Refinement Protocol to Improve Model Quality. *J. Chem. Theory Comput.* **2013**, *9* (8), 3843−3847.

(62) Lindert, S.; McCammon, J. A. Improved cryoEM-Guided Iterative Molecular Dynamics−Rosetta Protein Structure Refinement Protocol for High Precision Protein Structure Prediction. *J. Chem. Theory Comput.* **2015**, *11* (3), 1337−1346.

(63) Bonneau, R.; Tsai, J.; Ruczinski, I.; Chivian, D.; Rohl, C.; Strauss, C.; Baker, D. Rosetta in CASP4: Progress in ab initio protein structure prediction. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 119−26.

(64) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions1. *J. Mol. Biol.* **1997**, *268* (1), 209−225.

(65) Bonneau, R.; Strauss, C. E. M.; Rohl, C. A.; Chivian, D.; Bradley, P.; Malmström, L.; Robertson, T.; Baker, D. De Novo Prediction of Three-dimensional Structures for Major Protein Families. *J. Mol. Biol.* **2002**, *322* (1), 65−78.

(66) Woetzel, N.; Karakaş, M.; Staritzbichler, R.; Müller, R.; Weiner, B. E.; Meiler, J. BCL::Score—Knowledge Based Energy Potentials for Ranking Protein Models Represented by Idealized Secondary Structure Elements. *PLoS One* **2012**, *7* (11), No. e49242.

(67) Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68* (4), 344−351.

(68) McGreevy, R.; Teo, I.; Singharoy, A.; Schulten, K. Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods* **2016**, *100*, 50−60.

(69) Kim, D. E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004**, *32* (suppl 2), W526−W531.

(70) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; et al. Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (12), 4685−4690.

(71) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22* (12), 2577−2637.

(72) Ohio Supercomputer Center. 1987.