

# Protein NMR Structures Refined with Rosetta Have Higher Accuracy Relative to Corresponding X-ray Crystal Structures

Binchen Mao,<sup>†</sup> Roberto Tejero,<sup>†,‡</sup> David Baker,<sup>§</sup> and Gaetano T. Montelione<sup>\*,†</sup>

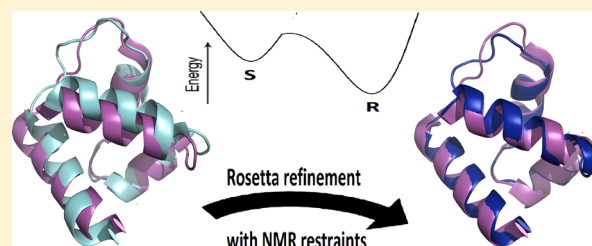
<sup>†</sup>Center for Advanced Biotechnology and Medicine, and Department of Molecular Biology and Biochemistry, and Department of Biochemistry and Molecular Biology of Robert Wood Johnson Medical School, and Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States

<sup>‡</sup>Departamento de Química Física, Universidad de Valencia, Dr. Moliner 50 46100-Burjassot, Valencia, Spain

<sup>§</sup>Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States

## S Supporting Information

**ABSTRACT:** We have found that refinement of protein NMR structures using Rosetta with experimental NMR restraints yields more accurate protein NMR structures than those that have been deposited in the PDB using standard refinement protocols. Using 40 pairs of NMR and X-ray crystal structures determined by the Northeast Structural Genomics Consortium, for proteins ranging in size from 5–22 kDa, restrained Rosetta refined structures fit better to the raw experimental data, are in better agreement with their X-ray counterparts, and have better phasing power compared to conventionally determined NMR structures. For 37 proteins for which NMR ensembles were available and which had similar structures in solution and in the crystal, all of the restrained Rosetta refined NMR structures were sufficiently accurate to be used for solving the corresponding X-ray crystal structures by molecular replacement. The protocol for restrained refinement of protein NMR structures was also compared with restrained CS-Rosetta calculations. For proteins smaller than 10 kDa, restrained CS-Rosetta, starting from extended conformations, provides slightly more accurate structures, while for proteins in the size range of 10–25 kDa the less CPU intensive restrained Rosetta refinement protocols provided equally or more accurate structures. The restrained Rosetta protocols described here can improve the accuracy of protein NMR structures and should find broad and general for studies of protein structure and function.



## INTRODUCTION

A protein's 3D structure provides the cornerstone for investigating its functions. The majority of the protein structures deposited in the Protein Data Bank are determined either by X-ray crystallography or solution-state nuclear magnetic resonance spectroscopy (NMR). While X-ray crystal structures are derived from electron density data and are often of higher accuracy, protein NMR structure determination in solution may more accurately reflect molecular dynamics and has the advantage of not requiring crystallization.

Solution NMR structure determination is generally based on three classes of experimental restraints: distance restraints, dihedral angle restraints, and orientation restraints. In combination with these restraints, different algorithms and force fields have been implemented to determine NMR structure using a variety of programs. Two groups of simulated annealing based programs are most commonly used by the NMR community: XPLOR/CNS<sup>1,2</sup> and DYANA/CYANA.<sup>3,4</sup> Aside from the accuracy and completeness of experimental data, the quality of NMR structures also depends on the programs utilized in structure calculation and structure refinement. In particular, as demonstrated by many studies, the quality of NMR structures can be improved by structure refinement in state-of-the-art force field with explicit or implicit

solvent.<sup>5–10</sup> Using such advanced refinement protocols, a few large-scale rerefinement studies have been done to improve the quality of NMR structures, especially for NMR structures determined prior to 2000.<sup>11–14</sup>

Protein NMR structure quality assessment metrics generally fall into two categories. One is the assessment of how well the structures fit with the experimental NMR data, including NOE-based distance restraint violations, dihedral angle restraint violations, NOE completeness,<sup>15</sup> and goodness-of-fit with NMR NOESY peak list<sup>16–18</sup> and RDC<sup>19,20</sup> data. The second class includes knowledge-based normality scores relative to high-resolution X-ray crystal structures, such as bond length, bond angle, backbone or side chain dihedral angle, and packing statistics.<sup>21–27</sup> Recent studies comparing various methods for automated analysis of NMR data and structure generation, such as the critical assessment of structure determination by NMR (CASD-NMR) study,<sup>28,29</sup> demonstrate that the algorithms and force fields utilized in NMR structure refinement can significantly improve these normality scores. For example, protein NMR structures refined by Rosetta without restraints generally have excellent knowledge-based stereochemical and

Received: September 29, 2013

Published: January 6, 2014

geometric quality scores, but sometimes have poorer fit to the original experimental data.<sup>28–31</sup>

The Rosetta molecular modeling program was first developed for *de novo* protein structure prediction,<sup>32,33</sup> homology modeling,<sup>34</sup> and protein design.<sup>35</sup> However, it has also been used in protein crystallography as part of improved protocols for determining crystallographic phases by molecular replacement<sup>36–40</sup> and for NMR structure determination and unrestrained NMR structure refinement.<sup>30,31,41–43</sup> Ramelot et al.<sup>30</sup> have shown that unrestrained Rosetta refinement can improve the phasing power of an NMR structure by moving it closer to its X-ray crystal structure counterpart. This observation has been corroborated for two additional NMR structures as part of a systematic investigation of using NMR structures in molecular replacement.<sup>31</sup> These results are intriguing, as they suggest that the force field of Rosetta may be even more accurate than the NMR data themselves in defining the protein structure. However, only one<sup>30</sup> or two<sup>31</sup> examples are reported in these two papers. In order to assess the generality of unrestrained Rosetta refinement, it is necessary to perform a systematic study using a much larger data set.

Another intriguing observation is that the number of restraint violations significantly increases after unrestrained Rosetta refinement,<sup>30,31</sup> which begs the question: Do those violated restraints reflect true structural differences between NMR structures and X-ray crystal structures? If that is the case, then would incorporating those NMR experimental restraints into Rosetta refinement drive the NMR structure away from its X-ray counterpart? More generally, what is the most efficient protocol for using Rosetta to improve the accuracy of protein NMR structures?

The Northeast Structural Genomics consortium (NESG; <http://www.nesg.org>) is one of several large-scale structure production centers of the Protein Structure Initiative (PSI). The NESG has contributed more than 500 NMR structures to the PDB over the past 12 years (summarized at <http://www.nesg.org/statistics.html>), representing some 5% of the ~10 000 NMR structures available in the PDB. Although most NESG structures have been solved by either NMR or X-ray crystallography, as of December, 2011 the NESG consortium had solved 41 pairs of protein structures for identical construct sequences using both X-ray crystallography and NMR methods. These 3D structures of proteins with identical sequences, together with the raw NMR and crystallography data available in the BioMagResBank (BMRB)<sup>44</sup> and Protein Data Bank (PDB),<sup>45</sup> are an extremely valuable composite data set available for studies directed at understanding structural variations between solution and crystal states and for new methods development.

In this study, we carried out a comprehensive and systematic study of both unrestrained and restrained Rosetta refinement for the NMR structures of 40 NESG NMR/X-ray structure pairs. NESG target GR4 was excluded from this study, since its deposited NMR structure is a single model and our protocol requires the input NMR structure as an ensemble of multiple models. For a subset of these pairs, we also assessed the value of the restrained CS-Rosetta method<sup>42,43,46</sup> carried out starting from extended conformations. The accuracy of (i) previously deposited PDB NMR structures, which were mostly refined using CNS with explicit solvent, (ii) unrestrained Rosetta refined structures, (iii) restrained Rosetta refined structures, and (iv) restrained CS-Rosetta structures generated with NMR restraints starting from extended conformations, were assessed

by various structural validation metrics, including restraint violation analysis, comparison against unassigned NOESY peak list data, convergence based on ensemble RMSD calculation, and various knowledge-based stereochemical and packing statistics. The Rosetta refined structures were further assessed based on their structural similarity with corresponding X-ray crystal structures and by analysis of how useful they are as molecular replacement (MR) templates for solving the corresponding X-ray crystal structure. This comprehensive study demonstrates the significant value of restrained Rosetta refinement of protein NMR structures, and provides efficient standard protocols for restrained Rosetta refinement that will be broadly useful to the protein NMR community.

## ■ MATERIALS AND METHODS

**Experimental Data.** Experimental data for this study were obtained by the NESG consortium for 40 proteins or protein domains solved by both solution NMR and X-ray crystallography and deposited in the PDB as of December 31, 2011. The structures range in size from 5 to 22 kDa and include 7 homodimers. Most of these NMR structures were refined using a standard NESG refinement protocol involving initial structure generation with CYANA<sup>3,4</sup> followed by structure refinement with CNS in explicit water solvent,<sup>7</sup> as described in detail at <http://www.nmr2.buffalo.edu/nesg.wiki/>. The coordinate files of both NMR structures and X-ray crystal structures were downloaded from the PDB database, along with the NMR restraint and X-ray structure factor files. Structure factor files, downloaded in mmCIF format, were converted to mtz format using the CCP4 program CIF2MTZ (Collaborative Computational Project, number 4, 1994). These protein data sets, together with citations to the corresponding PDB files, are summarized in Supplementary Table S1. Another CCP4 program *uniquefy* was used to standardize the mtz files and select reflections for free *R* calculation. The NMR restraints files are either in CYANA format or in Xplor/CNS format. PDBStat v5.9<sup>46</sup> was used to convert restraints into Rosetta restraint format. The extensive experimental data for the 40 NMR and X-ray structures have been organized in a single publicly available database ([http://psvs-1\\_4-dev.nesg.org/results/rosetta\\_MR/data\\_set.html](http://psvs-1_4-dev.nesg.org/results/rosetta_MR/data_set.html)). This compilation of NMR and crystallographic data, which was done as part of this study, will be valuable for future methods development projects.

**Rosetta Refinement Protocols.** A detailed protocol for restrained and unrestrained Rosetta refinement is included as Supporting Information. A brief summary is provided here. The Robetta fragment server<sup>47,48</sup> was used to generate a fragment library, based on the target protein sequence (excluding chemical shift data). Although this process could be done using chemical shift data in the fragment selection, for the refinement protocols developed in this work, chemical shift data were not used in the fragment generation for the Rosetta refinement calculations. Tests using chemical-shift-based fragment selection demonstrated no significant improvement in the refinement protocol, although there is no *a priori* reason not to use chemical shift data in the fragment selection for restrained Rosetta refinement.

For each target protein, fragments from the target protein itself were eliminated from the fragment library. Loop regions were then defined by the consensus of (i) secondary structure, (ii) “not-well-defined” residues identified by the PSVS server based on dihedral angle order parameter values,<sup>21,49</sup> and (iii) noncore residues determined by FindCore.<sup>50</sup> For these regions, loop rebuilding was done together with all-atom refinement of the entire structure using the loopmodeling application of Rosetta version 3.3, based on cyclic coordinate descent (CCD) and kinematic closure (KIC).<sup>51,52</sup>

In addition to the loop remodeling, the ‘fastrelax’ mode was used to allow the whole structure to relax in Rosetta all-atom force field. The process was used to sample side chain conformations of the well-defined regions and both backbone and side chain conformations of the loop and not-well-defined regions. The fast relax modes work by running many side chain repack and minimization cycles to locate a

Table 1. Summary of NMR Structure Quality Statistics for 40 NMR Structures Using Different Refinement Protocols

structural metric <sup>a</sup>	parameter range	PDB <sup>b</sup>	R3 <sup>b</sup>	R3rst <sup>b</sup>
NOE restraint violations per conformer (Å)	0.1–0.2 Å	5.0 ± 7.0	15.6 ± 8.7	6.6 ± 5.5
	0.2–0.5 Å	1.9 ± 4.5	31.7 ± 20.3	4.0 ± 4.1
	>0.5 Å	0.1 ± 0.2	74.3 ± 56.0	1.3 ± 1.2
dihedral restraint violations per conformer (°)	<10°	5.2 ± 6.9	7.9 ± 7.0	1.1 ± 1.5
	>10°	0.2 ± 0.6	5.9 ± 6.6	0.8 ± 1.2
	bb_ord	0.79 ± 0.69	1.05 ± 0.84	0.80 ± 0.81
Ensemble RMSD (Å) <sup>c</sup>	hvy_ord	1.19 ± 0.64	1.43 ± 0.80	1.10 ± 0.79
	bb_all	2.92 ± 1.85	3.38 ± 1.80	3.20 ± 1.70
	hvy_all	3.46 ± 1.85	3.90 ± 1.82	3.70 ± 1.70
	recall	0.94 ± 0.07	0.92 ± 0.07	0.94 ± 0.06
RPF scores <sup>d</sup>	precision	0.90 ± 0.06	0.90 ± 0.06	0.90 ± 0.06
	DP score	0.79 ± 0.08	0.76 ± 0.07	0.79 ± 0.08
	Verify3D	−2.11 ± 1.12	−1.28 ± 0.91	−1.44 ± 0.93
	Prosa	−0.57 ± 1.03	−0.18 ± 0.98	−0.26 ± 1.01
PSVS Z scores	Procheck_bb	−0.34 ± 1.68	0.14 ± 1.46	0.63 ± 1.58
	Procheck_all	−0.94 ± 1.85	1.23 ± 1.43	1.40 ± 1.60
	Molprobrity clash score	−2.10 ± 1.20	0.84 ± 0.38	0.55 ± 0.57

<sup>a</sup>Structure quality scores were analyzed by PSVS.<sup>21</sup> Constraint violations were calculated with the program PDBStat.<sup>46</sup> Knowledge-based statistics were calculated using the programs Verify3D,<sup>25</sup> ProsaII,<sup>24</sup> ProCheck,<sup>23</sup> and MolProbrity,<sup>26,27</sup> normalized to  $Z = 0$  for a set of 252 high-resolution X-ray crystal structures.<sup>21</sup> <sup>b</sup>Structure quality scores were calculated for the NMR structures available from the PDB (PDB) and for the unrestrained Rosetta (R3) and restrained Rosetta (R3rst) structures. For each statistic, the mean and standard deviation were computed across the 40 NMR structures and are formatted as mean ± sd. <sup>c</sup>Computed following superimposition of atoms with well-defined atomic positions, as determined by the dihedral angle order parameter method<sup>49</sup> as implemented in PSVS. bb\_ord, backbone atoms (N, C $\alpha$ , C') in well-ordered residues; hvy\_ord, all heavy (N, C, O, S) atoms in well-ordered residues; bb\_all, backbone atoms of all residues; hvy\_all, all heavy atoms in all residues. <sup>d</sup>RPF-DP scores were computed for 35 NMR structures for which NOESY peak list data is available, and provide a statistical assessment of the consistency of the 3D NMR structure ensemble with the NOESY peak list as provided by the RPF software.<sup>16,17</sup>

low-energy state for the input model. The structural divergence of the starting model to the relaxed model is determined by the resulting energy gap. The structure can change up to 2–3 Å from the starting conformation during the minimization cycles.

For restrained Rosetta refinement, Rosetta formatted distance restraints and dihedral angle restraints were generated using the PDBStat restraint converter software<sup>46</sup> and were merged into a single restraint file. Although dihedrals are restrained by Rosetta's energy terms, where chemical shift data provide reliable dihedral restraint data using Talos+,<sup>53</sup> these dihedral restraints were retained in the refinement process. These distance restraints were used in restrained Rosetta refinement with an upper-bound tolerance of 0.3 Å, to allow the structure to better relax energetically in the Rosetta force field. Details of the restraint violation penalty functions are provided in the Supporting Information. For each individual conformer of the NMR structure ensemble, the restrained Rosetta refinement was used to generate 10 decoys, and the one with lowest Rosetta energy was selected as the final Rosetta refined model for this specific conformer. For homodimeric NMR structures, a symmetry definition file, restraining the structures of protomers to be identical, is generated by Rosetta and used to guide Rosetta refinement, as outlined in Supporting Information. The other steps in the restrained Rosetta refinement were exactly the same as those outlined above for unrestrained Rosetta refinement.

Sophisticated methods could be used to define the relative weight,  $W$ , between knowledge-based Rosetta energy terms and experimental restraint terms. In this study, the relative weight  $W$  was set to 1.0. The rationale was that at this value of  $W$ , plots of total energy (Rosetta energy + restraint energy) vs  $W$  exhibit a minimum (as shown in Supplementary Table S3 and Figure S5).

**Restrained CS-Rosetta Protocols.** A detailed protocol for restrained CS-Rosetta (rCS-Rosetta) calculations, starting from extended conformations, is also included as Supporting Information. Chemical shift information is used in Rosetta fragment picking by MFR method<sup>54</sup> using the updated fragment library.<sup>55</sup> Restraints were converted to Rosetta format using PDBStat. Then, a total of 10 000 decoys were generated using the AbinitioRelax application of Rosetta

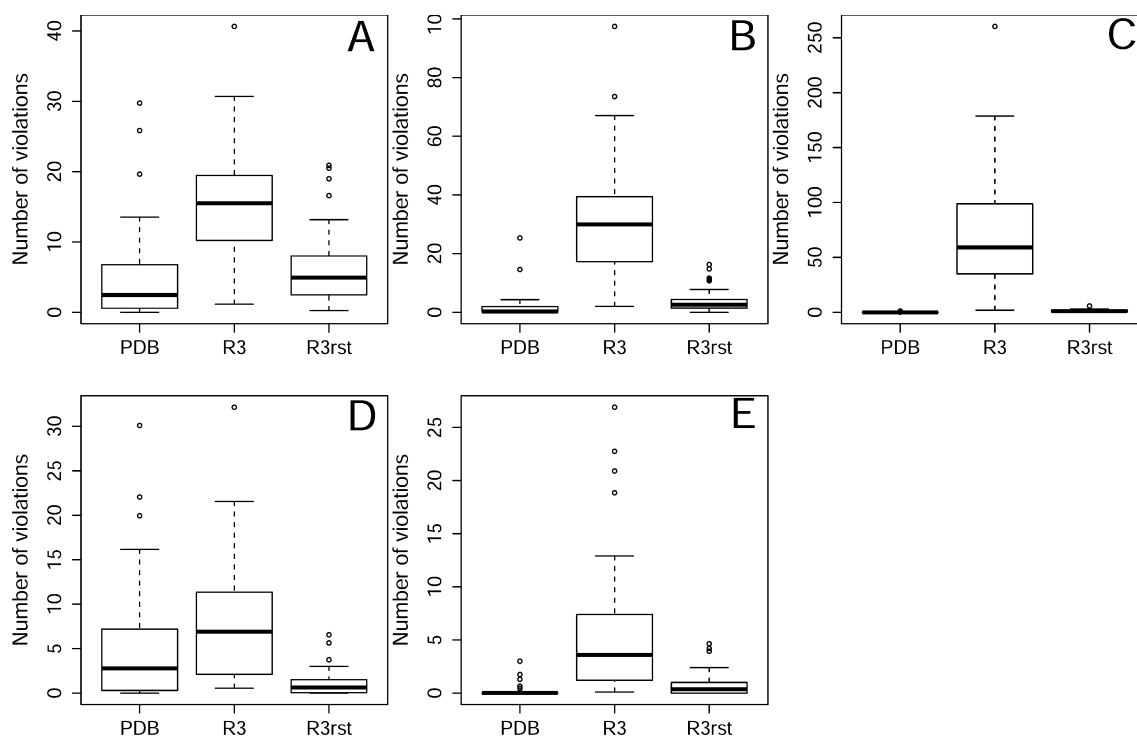
3.3 with the NMR restraints. For 9 targets in 10–22 kDa range, 20 000 decoys were generated for each target instead of 10 000 decoys. Chemical shift rescoring was then calculated for the 1000 lowest Rosetta energy decoys, and the 20 lowest rescore decoys are selected as the final rCS-Rosetta model.

**Global Distance Test Scores.** GDT.TS stands for global distance test total score, which measures the 3D similarity of two structures with identical amino acid sequences.<sup>56</sup> The global distance test performs many different sequence-independent superpositions of the model and the “gold standard” structure and calculates the percentage of structurally equivalent pairs of C $\alpha$  atoms that are within specified distance cutoffs  $d$ . The GDT.TS score is the arithmetic mean of four scores obtained with distance cutoffs of  $d = 1, 2, 4,$  and  $8 \text{ \AA}$ .

**Structure Quality Assessment.** The Protein Structure Validation Software suite (PSVS)<sup>21</sup> (<http://psvs.nesg.org/>) was used for structure quality assessment analysis. PSVS provides Z scores for a variety of widely adopted structural quality measures, such as Procheck G factor,<sup>23</sup> MolProbrity clash score<sup>26,27</sup> and other structure quality assessment metrics. The Procheck all dihedral angle G factor is determined by the stereochemical quality of both backbone and side-chain dihedral angles of proteins, and the MolProbrity clash score is calculated by the program *probe* is a measure to reflect the number of high-energy contacts in a structure. Structure quality assessments also include ensemble RMSD analysis, restraint violations,<sup>46</sup> and RPF-DP<sup>16,17</sup> statistics. Z scores are computed relative to a set of 252 high-resolution X-ray structures and normalized so that more positive Z scores corresponding to better structure quality scores.<sup>21</sup>

To evaluate structural similarity between NMR structure models and their X-ray counterparts, we utilized the programs FindCore<sup>50</sup> and PDBStat v5.9<sup>46</sup> to calculate the RMSD of backbone atom and/or all heavy atom positions, for both well-defined residues and for all residues (including those that are ill-defined in the PDB NMR structures). We also used the TM-score<sup>57</sup> program to calculate the GDT.TS<sup>56</sup> global superimposition scores. To further determine RMSD for specific subset of atoms, such as side chain atoms of  $\alpha$ -helix residues, we used Pymol<sup>58</sup> to superimpose NMR structures with reference X-ray crystal structures, then calculated average RMSD based





**Figure 1.** The number of restraint violations is significantly reduced by incorporating NMR restraints into Rosetta refinement. Restraint violations were assessed using PDBStat<sup>46</sup>, across the complete set of 40 NESG NMR structures used in this study. (A) Boxplot of the number of distance restraint violations between 0.1 Å and 0.2 Å. (B) Boxplot of the number of distance restraint violations between 0.2 Å and 0.5 Å. (C) Boxplot of the number of distance restraint violations larger than 0.5 Å. (D) Boxplot of the number of dihedral angle restraint violations between 1 deg and 10 deg. (E) Boxplot of the number of dihedral angle restraint violations larger than 10 deg.

on the structural superimposition. The same procedures were performed to evaluate structural similarity between Rosetta refined structures and the corresponding X-ray crystal structures.

Well-ordered residues are defined by dihedral angle order parameters<sup>49</sup> with  $S(\varphi) + S(\psi) \geq 1.8$  units, and ‘core atoms’ were calculated using the FindCore program<sup>50</sup> based on interatomic distance variance matrices. The DSSP<sup>59,60</sup> program was utilized for annotating secondary structure elements, and solvent accessible areas of atoms were calculated by areaimol program in CCP4 package.<sup>61,62</sup>

**Molecular Replacement.** The program Phaser<sup>47</sup> (version 2.1) was used for estimating diffraction phases by molecular replacement. MR\_AUTO mode was adopted with RMS set to 1.5 units. The programs ARP/wARP<sup>63,64</sup> version 7.0 and/or Phenix.autobuild<sup>65</sup> were used for automatic model building, based on the Phaser MR solution. The ARP/wARP expert system mode was employed for automatic model building, and Refmac5<sup>66</sup> was used in refinement, starting from the positioned search model, and a maximum of 10 building cycles were allowed.

## RESULTS

Forty NESG NMR structures which have corresponding X-ray crystal structures were downloaded from the PDB and refined using unrestrained and restrained Rosetta protocols, as outlined in Supporting Information. The resulting structures were assessed with various structure quality assessment metrics. These results are summarized in Table S2. More comprehensive structure quality statistics for these structures are available on line at [http://psvs-1\\_4-dev.nesg.org/results/rosetta\\_MR/rosettaMR\\_PSVS\\_summary.html](http://psvs-1_4-dev.nesg.org/results/rosetta_MR/rosettaMR_PSVS_summary.html).

**Comparison of Restraint Violations Using Restrained and Unrestrained Rosetta Refinement Protocols.** We assessed distance restraint and dihedral angle restraint violations for the 40 protein NMR structures downloaded

from the PDB and for the corresponding unrestrained and restrained Rosetta refined structures. Restraint violations were assessed using the standardized methods of the PDBStat program,<sup>46</sup> against the original distance and dihedral restraint lists (i.e., not accounting for the 0.3 Å upper-bound tolerance used in the restrained Rosetta calculations). Distance restraint violations were divided into three categories based on the level of severity; i.e., distance restraint violations between 0.1 and 0.2 Å, between 0.2 and 0.5 Å, and higher than 0.5 Å. Dihedral angle restraint violations were divided into two categories: between 1° and 10° and higher than 10°. The mean and standard deviations of the number of restraint violations in each category were calculated. These restraint violation statistics for each NMR structure ensemble are summarized in Table S2, and the average violations per conformer of the 40 NMR structures assessed for each of the restraint violation categories and for each of three methods are presented in Table 1. These distributions of restraint violations obtained for these three data sets are also illustrated graphically in Figure 1.

As expected, unrestrained Rosetta refinement results in a significant number of restraint violations, especially for the most severe violation categories. However, in restrained Rosetta refined structures the number and distribution of restraint violations per conformer is similar to, though slightly higher than, those assessed for the NMR structure ensembles deposited in the PDB (Table 1 and Figure 1). From this analysis we conclude that protocols for incorporating NMR restraints into Rosetta refinement are effective in generating Rosetta refined NMR structures that satisfy the experimental distance and dihedral angle restraint data as well as the structures deposited in the PDB that have been refined by conventional methods.

### Restrained Rosetta Refinement Can Improve the Precision of Side Chain Heavy Atom Positions.

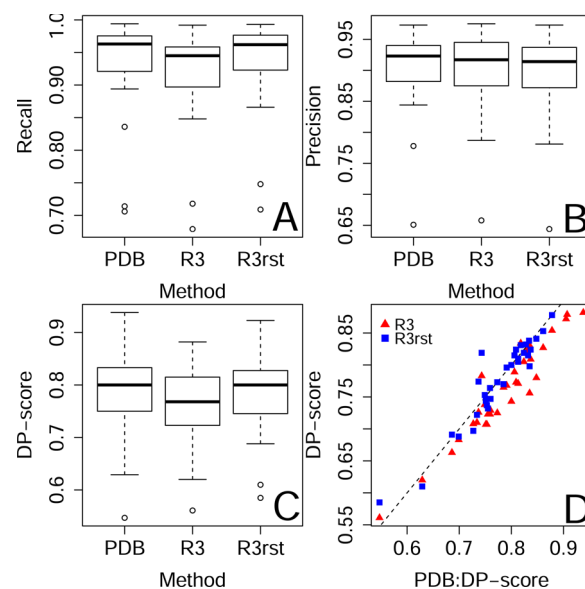
The resolution of electron density maps and atomic *B*-factors reflect the precision of X-ray crystal structures. However, there are no such experimental observables to define the precision of solution NMR structures. Usually, the RMSD of the ensemble of superimposed NMR conformers is considered to be a useful measure of its overall precision, although as discussed elsewhere this measure can be problematic if there are extensive intramolecular dynamics.<sup>67</sup> We calculated the ensemble RMSD of PDB NMR structures (PDB), unrestrained Rosetta refined structures (R3), and restrained Rosetta refined structures (R3rst) for each of the 40 protein NMR structure ensembles in our data set. Four categories of RMSD were calculated: (i) backbone RMSD of well-defined residues, defined by dihedral angle order parameters, (ii) backbone RMSD of all residues, (iii) heavy atom RMSD of well-defined residues, defined by dihedral angle order parameters, and (iv) heavy atom RMSD of all residues. The mean and standard deviations of RMSDs for backbone and for all heavy atoms in these well-defined residues are listed in Table 1, and the values for each of the 40 ensembles generated by each of the three protocols are plotted in Supplementary Figure S1. The RMSD's of unrestrained Rosetta-refined structures are higher than PDB NMR structures in all four categories, for both backbone atom and all-heavy atom classes (Figure S1A–D). Ignoring experimental restraints in Rosetta refinement (R3) generally increases structural uncertainty for all the backbone and side chain atoms of all residues. For restrained Rosetta refined structures (R3rst), in well-defined regions the average RMSD of backbone atoms is comparable with PDB NMR structures (Figure S1A), and the average RMSD of all heavy atoms is about 10% lower than PDB NMR structures (Supplementary Figure S1C). For most targets restrained Rosetta refined structures have lower ensemble RMSDs in well-defined regions for all heavy atoms (including both backbone and side chain atoms) than PDB NMR structures. These results demonstrate that restrained Rosetta refinement has the potential to improve the precision of side-chain atoms.

On the other hand, the average ensemble RMSD statistics for all residues (including atoms that are not well-defined in the original PDB NMR ensembles), for both unrestrained and restrained Rosetta refined structures, are higher than for the corresponding PDB NMR structures (Supplementary Figure S1B,D). This demonstrates that when restraints are included, the loop rebuilding process implemented in our Rosetta refinement protocol does a better job of sampling the wide range of conformations which are consistent with the experimental data.

**Restrained Rosetta Refined Structures Fit the NOESY Peak Lists Data Better than Unrestrained Rosetta Refined Structures.** RPF-DP<sup>16,17</sup> is a metric used to evaluate how well a protein NMR model fits the experimental unassigned NOESY peak list and resonance assignment data. The program calculates recall, precision, and DP scores of the match between short distances in the model and all possible NOESY crosspeak assignments. Recall is defined as the percentage of peaks in the NOESY peak list that are consistent with the interproton distances of the 3D structures. Precision is defined as the percentage of close distances (general set at <5 Å) between proton pairs in the query structures whose back calculated NOE cross peaks are also actually detected in NMR experiments. The DP score is a normalized *F*-score calculated

from the recall and precision to measure the overall fit between the query structure and the experimental data, with a freely rotating chain model and the quality of the NOESY data set defining the lower and upper bounds, respectively, of the *F*-measure.

NOESY peak list data are available for 35 of the 40 protein NMR/X-ray structure pairs used in this study. The mean and standard deviations of recall, precision, and DP score are listed Table 1, and boxplots of recall, precision, and DP score are shown in Figure 2. Unrestrained Rosetta refined structures



**Figure 2.** Rosetta-refined structures have RPF-DP scores, comparing the structure against the unassigned NOESY peak list, similar to those of structures deposited in the PDB. (A) Boxplots of Recall scores for structures deposited in the PDB or refined with Rosetta protocols. (B) Boxplots of Precision scores for structures deposited in the PDB or refined with Rosetta protocols. (C) Boxplots of DP-scores for structures deposited in the PDB or refined with Rosetta protocols. (D) DP-score scatterplot. DP-scores of the PDB NMR structures are plotted on the X-axis, while the DP-scores of both the unrestrained Rosetta refined structures represented by red solid triangle symbols (R3) and restrained Rosetta refined structures represented by blue solid rectangle symbols (R3rst) are plotted on the Y-axis. The black dashed line indicates  $y = x$ . Data are presented for 35 NMR structures for which NOESY peak list data are available.

generally have precision similar to PDB NMR structures but lower recall and DP scores, i.e., in general, the unrestrained Rosetta refined structure does not fit the NOESY peak list data as well as the PDB NMR structures. On the other hand, restrained Rosetta refined structures have recall, precision, and DP scores that are essential identical to those of the PDB NMR structures (Table 1 and Figure 2A–C). A scatter plot of these DP scores is shown in Figure 2D. While the majority of the NMR structures generated by unrestrained Rosetta refinement (R3) have DP scores lower than the PDB NMR structures, most structures refined by the restrained Rosetta protocol (R3rst) have DP scores similar to PDB NMR structures (Figure 2C,D). In a few cases, the restrained Rosetta refined NMR structures have significantly better DP scores compared with the corresponding PDB NMR structure (Figure 2D).

As no distance restraints are enforced during the unrestrained Rosetta refinement process, the refined structures do not satisfy distance restraints as well as the PDB NMR

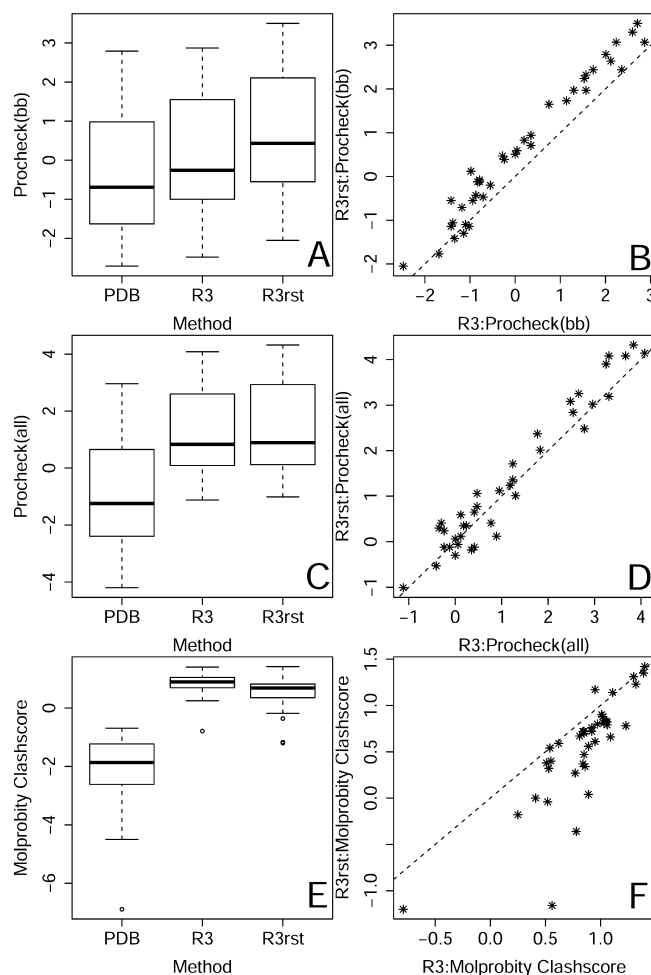
structures. They also do not fit as well to the unassigned NOESY peak lists data because the distance restraints are directly derived from NOESY peak lists. On the other hand, when distance restraints are incorporated into Rosetta refinement, the refined structures generally fit the NOESY peak list data as well or better than the PDB NMR structures that have been refined by conventional methods.

**Rosetta Refinement Consistently Improves Stereochemical Quality and Geometry of Protein NMR Structures.** We used PSVS to calculate a variety of knowledge-based structural quality Z scores, including Verify3D, Prosa, Procheck backbone G factor (Procheck\_bb), Procheck all dihedral angle G factor (Procheck\_all), and Molprobrity clash scores. These scores are normalized so that more positive Z scores correspond to better values of these knowledge-based metrics. The mean and standard deviation of those Z scores for the 40 NMR structures generated by the three methods (NMR PDB, unrestrained Rosetta, and restrained Rosetta) are summarized in Table 1. Both unrestrained and restrained Rosetta refined structures have better (i.e., more positive) Z scores for all the five measures, especially for Procheck all dihedral angle G factor and Molprobrity clash score Z scores. Boxplots of Procheck\_bb, Procheck\_all, Molprobrity clash score Z scores for these NMR PDB and unrestrained and restrained Rosetta refined structures are shown in Figure 3A,C,E. Rosetta refined structures consistently have improved Procheck\_bb, Procheck\_all, and Molprobrity clash score Z scores.

In order to further investigate the effect of incorporating experimental restraints into Rosetta refinement on these knowledge-based Z scores, we also made 2D scatter plots of Procheck\_bb, Procheck\_all, and Molprobrity clash score Z scores comparing unrestrained and restrained Rosetta refined structures. In Figure 3B,D,F the Z scores of unrestrained (R3) and restrained Rosetta refined structures (R3rst) are plotted on *x*- and *y*-axes respectively. The Procheck\_bb Z scores of restrained Rosetta refined structures are consistently better than unrestrained Rosetta refined structures (Figure 3A,B). This is attributable to the fact that the experimental dihedral angle restraints and local NOE data are very helpful in guiding Rosetta to generate decoys with more accurate backbone stereochemical quality. Procheck\_all Z scores, which include side chain dihedrals, are also marginally improved for the restrained Rosetta refined structures (Figure 3C,D). On the contrary, the Molprobrity clash score Z scores of unrestrained Rosetta refined structures are generally better than restrained Rosetta refined structures (Figure 3F). This is because some experimental restraints result in close contacts in the structure. However, while the unrestrained Rosetta refined structures have fewer Molprobrity clashes, they are less converged and sometimes underpacked relative to restrained Rosetta refined structures.

Overall, these data demonstrate that stereochemical quality and geometry of PDB NMR structures can be significantly improved by Rosetta refinement carried out with or without restraints. However, the restrained Rosetta refinement protocol provides structures that have both improved Z scores (Figure 3) and simultaneously fit well to both the experimental distance restraints (Figure 1) and the unassigned NOESY peak list data (Figure 2).

**Restrained Rosetta Refinement Consistently Moves NMR Structures Closer to Their X-ray Counterparts than Unrestrained Rosetta Refinement.** Theoretically, solution



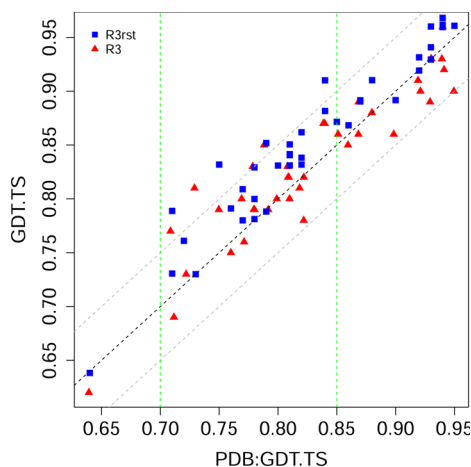
**Figure 3.** Knowledge-based structure quality scores are much improved after Rosetta refinement. (A) Boxplot of Procheck<sup>23</sup> backbone dihedral angle G-factor Z-scores for structures refined with different protocols. (B) Scatterplot of Procheck<sup>23</sup> backbone dihedral angle G-factor Z-scores. (C) Boxplot of Procheck<sup>23</sup> all dihedral angle G-factor Z-scores for structures refined with different protocols. (D) Scatterplot of Procheck<sup>23</sup> all dihedral angle G-factor Z-scores. (E) Boxplot of Molprobrity clashscore<sup>26,27</sup> Z-scores for structures refined with different protocols. (F) Scatterplot of Molprobrity clashscore<sup>26,27</sup> Z-scores. In the scatter plots (B), (D) and (F), the Z-scores of unrestrained Rosetta refined structures (R3) are plotted on the X-axis, while the Z-scores of restrained Rosetta refined structures (R3rst) are plotted on the Y-axis.

NMR structures need not necessarily be identical to X-ray crystal structures, which are determined in a crystalline environment. In addition, these crystal structures were determined with cryoprotection at  $\sim 77$  K, while the NMR structures were determined in solution at  $\sim 300$  K. In particular, crystal packing effects may stabilize conformers that do not predominate in solution. However, since X-ray structures are highly hydrated, with relatively few intermolecular contacts, one might expect that such effects are the exception rather than the rule and that the dominant structure in solution characterized by NMR should generally be very similar to the X-ray crystal structure that is obtained for the same protein construct. This conclusion is supported by comparisons of protein structures in different crystal forms, which generally agree within an RMSD of  $<0.5$  Å.<sup>68,69</sup> Based on these considerations, and assuming the X-ray structure to be an accurate representation of the



predominant solution structure, we also assessed whether or not Rosetta refinement, with or without experimental restraints, moves the PDB NMR structures closer to their X-ray counterparts.

For this assessment, we calculated the GDT.TS between (i) PDB NMR structures, (ii) unrestrained Rosetta refined structures, and (iii) restrained Rosetta refined structures, with their corresponding X-ray structures. NESG target DrR147D was left out of this analysis because its solution NMR structure is a monomer solved at pH 4.5, while its X-ray structure is a dimer solved at pH 6.0, and NMR studies demonstrate a significant structural change over this pH range (data not shown). These results for the remaining 39 NESG NMR/X-ray pairs are summarized in a GDT.TS scatterplot (Figure 4), with



**Figure 4.** Restrained Rosetta refined structures are more similar to their corresponding X-ray crystal structures than PDB NMR structures. GDT.TS values of PDB NMR structures to corresponding X-ray structures are plotted on the X-axis, and GDT.TS values of both unrestrained Rosetta refined structures (R3, represented by red solid triangle) and restrained Rosetta refined structures (R3rst, represented by blue solid rectangles) to their corresponding X-ray structures are plotted on the Y-axis. Data are summarized for 39 NESG NMR/X-ray pairs. The two green dash lines indicate GDT.TS of PDB NMR structures equal to 0.7 and 0.85 respectively. The black dash line indicates  $y = x$ , and the two gray dash lines indicate  $y = x + 0.05$  and  $y = x - 0.05$  respectively.

the GDT.TS of PDB NMR structures relative to the corresponding X-ray crystal structure on the  $x$ -axis and GDT.TS of the unrestrained or restrained Rosetta refined structures on the  $y$ -axis. Based on observations of previous studies<sup>30,31</sup> done with a much small number (i.e., 1 or 2) of protein targets, we expected unrestrained Rosetta refinement would generally move NMR structures closer to their X-ray counterparts. However, as illustrated in Figure 4, using this larger data set of 39 NMR/X-ray pairs, we observed that this is not the case. After unrestrained Rosetta refinement (R3), only 17 of 39 targets exhibit higher GDT.TS values, 6 targets remain about the same, and 16 of the protein ensembles have lower GDT.TS values than the NMR structures refined by conventional methods and deposited in the PDB. On average, unrestrained Rosetta refinement improved the GDT.TS by only 0.4%. On the other hand, as illustrated in Figure 4, restrained Rosetta refinement (R3rst) generally improved the GDT.TS score to the X-ray crystal structure, compared with the NMR structure deposited in the PDB; 32 of 39 targets have

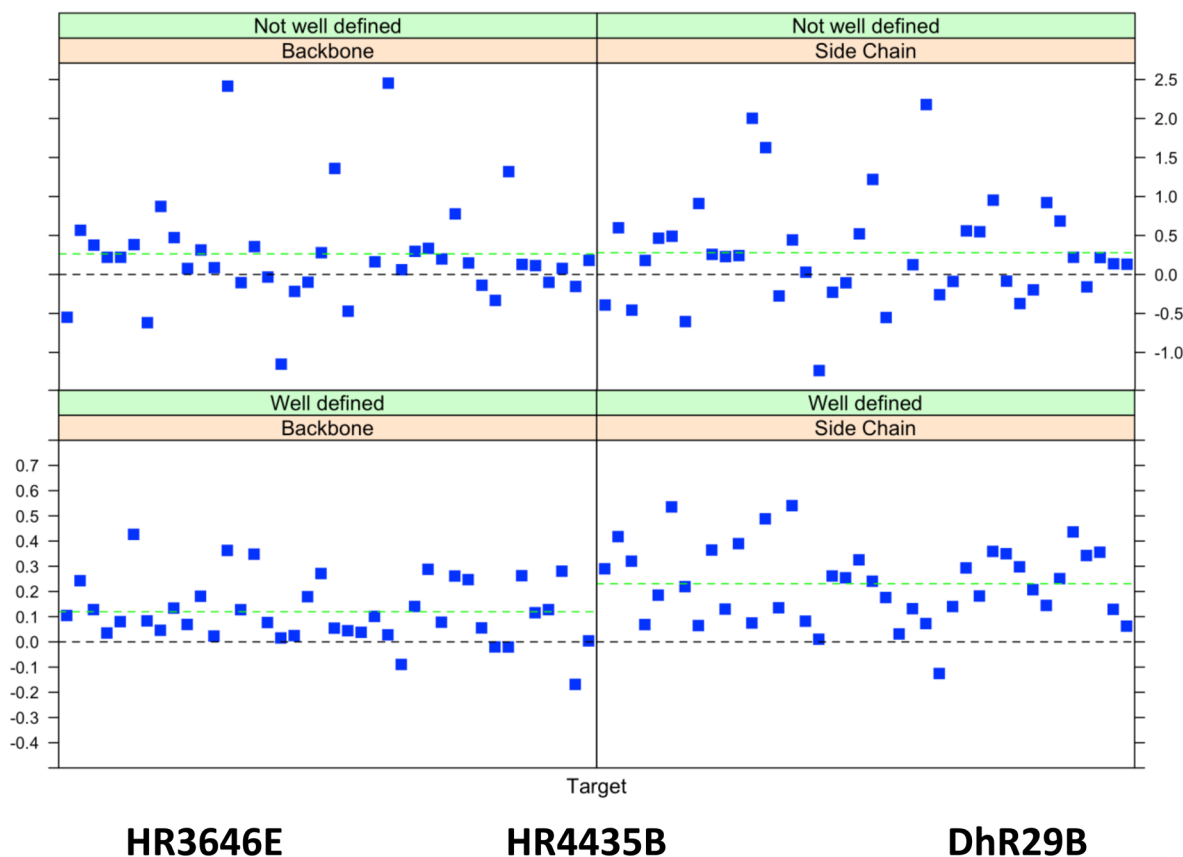
better GDT.TS values, 4 targets remain about the same, and only 3 targets have slightly lower GDT.TS values. On average, restrained Rosetta refinement improved the GDT.TS scores of the PDB NMR structures by 2.5%, with some increasing by as much as 10%.

Further analysis of the data of Figure 4 indicates that when the similarity between the conventionally-refined NMR structure and the corresponding X-ray crystal structure is moderate ( $0.7 \leq \text{GDT.TS} \leq 0.85$ ), more often than not, Rosetta refinement can move NMR structures closer to their X-ray counterparts when the experimental restraints are incorporated. However, in cases where the similarity between NMR and X-ray structures is initially high ( $\text{GDT.TS} > 0.85$ ), more often than not, unrestrained Rosetta refinement moves NMR structures further from their X-ray counterparts, while the improvement provided by restrained Rosetta refinement is less dramatic (Figure 4).

We further investigated in these data how restrained Rosetta refinement improves the similarity between NMR and X-ray structures by RMSD calculations. As illustrated in Figure 5 (top panel), restrained Rosetta refinement consistently improved the agreement between NMR and X-ray structures for both backbone and side chain atoms. The lower panel provides some comparisons between the mediod NMR conformer (i.e., the single conformer in the NMR ensemble most like all the other members of the ensemble)<sup>46,70</sup> before and after restrained Rosetta refinement, and the corresponding X-ray crystal structure coordinates. Typically, improvements in accuracy are the result of better packing between secondary structure elements. Often, this improves the accuracy of interhelical orientations, as shown for example in Figure 5 for NESG targets HR3646E and HR4435B. For DhR29B, in order to emphasize the structural changes, only the last two C-terminal  $\beta$  strands are plotted. In this case, the two-residue strand (76–77) of the NMR structure deposited in the PDB is extended to six residues long (76–81) after restrained Rosetta, which is more consistent with corresponding X-ray crystal structure.

On average, the improvement of structural similarity to corresponding X-ray structure resulting from restrained Rosetta refinement is modest. However, restrained Rosetta refinement drives some NMR structures significantly closer to their X-ray counterparts, as much as 0.45 and 0.55 Å RMSD, respectively, for backbone and side chain atoms in well-defined regions (Figure 5 top). Specific atom positions change by as much as 1–3 Å (as illustrated in some of the examples of superimposed structures shown in Figure 5 bottom). These changes may be biologically significant and can have significant effects on the phasing power of the structure, as illustrated below.

The fact that restrained Rosetta refinement consistently improves the accuracy of NMR structures relative to the corresponding X-ray crystal structure is a significant observation. In order to explore this in more detail, we compared the RMSD between either refined NMR or deposited NMR structures relative to X-ray crystal structures for several different classes of atoms. These included (i) atoms in well-defined residues (defined by dihedral angle order parameters), (ii) well-defined core atom sets calculated by FindCore program, (iii) atoms in buried residues, and (iv) atoms in regular secondary structure elements. Comparisons were made for both unrestrained Rosetta refinements (summarized in Supplementary Figure S2) and for restrained Rosetta refinements (summarized in Supplementary Figure S3). More often than not, restrained Rosetta refinement also improved the agreement between



**Figure 5.** The agreement between NMR structures and their X-ray counterparts are generally improved following restrained Rosetta refinement. Top: Plot of differences of RMSD to X-ray crystal structures before and after restrained Rosetta refinement. The NESG NMR/X-ray pair target index is plotted on the X-axis, and the differences between the RMSD of PDB NMR structures to their corresponding X-ray structures and the RMSD of restrained Rosetta refined structures to their corresponding X-ray structures are plotted on the Y-axis in units of Ångstroms. The four subpanels summarize data for well-defined (lower half) and not-well defined (upper half) residues, and for backbone (left) and sidechain (right) atoms. Well-defined vs not well-defined residues are defined by  $S(\phi)+S(\psi) \geq 1.8$ .<sup>21,49</sup> Data are summarized for 39 NESG NMR/X-ray pairs. Bottom: Superimposition of X-ray, NMR and restrained Rosetta refined structures. Left – HR3646E; middle – HR4435B ; right – DhR29B. The structures are color coded as: magenta- X-ray crystal structure; cyan – NMR structure deposited in PDB; blue – restrained Rosetta refined structure. For DhR29B, only the last two C-terminal beta strands are plotted.

NMR structures and X-ray structures for (i) not-well-defined regions, (ii) noncore residues, (iii) surface residues, and (iv) loop regions as well as for well-defined backbone and side-chain atoms. For classes of atoms in regions of the structure that are not-well-defined (i.e., less well converged) in the NMR ensemble, the improvement is often quite substantial, as illustrated in the top half of Figure 5 for some structures the accuracy relative to the corresponding crystal structure improves by 1.0–2.5 Å RMSD in loop regions. This reflects

the ability of Rosetta to accurately model regions of the protein structure, such as surface loops, that are under-restrained by the experimental NMR data.

**Restrained Rosetta Refinement Can Improve the Phasing Power of Poor NMR MR Templates.** Molecular replacement (MR) is widely used for addressing the phase problem in X-ray crystallography. Historically, the common notion in the structural biology community is that the quality of NMR structure is often not good enough for MR, even when

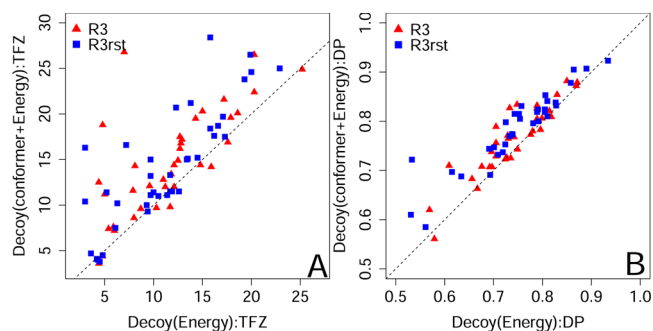


the sequence of the search model is identical to the target X-ray structure. However, as demonstrated by a recent study with 25 NESG NMR/X-ray pairs,<sup>31</sup> protein NMR structures prepared by excluding not-well-defined atom positions using an interatomic variance matrix-based protocol can generally be used successfully as MR templates. Additionally, the phasing power of NMR structures that failed to provide good MR solutions was observed to be improved by unrestrained Rosetta refinement in two cases. Using the extensive set of NMR/X-ray pairs, we critically assessed this hypothesis by comparing the phasing powers of the conventionally-refined PDB NMR structures with those of unrestrained and restrained Rosetta refined structures.

We prepared the MR starting models for PDB NMR structures, unrestrained Rosetta refined structures, and restrained Rosetta refined structures by first eliminating not-well-defined atoms using the 'FindCore' protocol.<sup>50</sup> Phaser was then used to search for MR solutions. Two targets (DrR147D, ER382A) were excluded in this study due to the following facts: The NMR structure of target ER382A (PDB ID: 2jn0) was solved as a monomer without a ligand, whereas its crystal structure counterpart (PDB ID: 3fif) has eight subunits in the asymmetric unit and was solved in complex with a heptapeptide ligand and appears to have a distinct structure, i.e., the  $C\alpha$  RMSD between the NMR structure and chain A of the crystal structure is 2.44 Å. As mentioned above, the NMR structure of target DrR147D (PDB ID: 2kcz) is a monomer solved at pH 4.5, while its crystal structure counterpart (PDB ID: 3ggn) is a dimer solved at pH 6.0, with significant structural changes in backbone structure due to pH-induced dissociation of the dimer. The remaining 38 NMR/X-ray pairs were used to assess the impact of restrained Rosetta refinement on the MR phasing power of the NMR ensemble.

For the initial Rosetta refinement protocol, the decoys are picked solely based on Rosetta energy, that is, we picked the top 20 decoys with the lowest Rosetta energy from the entire pool of decoys generated from all the conformers in NMR structure ensemble. It was observed, however, that frequently those 20 decoys originated from the same one or two similar conformers in the unrefined NMR ensemble; thus the structural variance information within the NMR ensemble is lost using this simple decoy picking process. In order to preserve the conformational variability information within the NMR ensemble, we adopted a protocol in which the one lowest-energy Rosetta decoy was selected from the ensemble of decoys generated from each NMR conformer. As shown in Figure 6A,B, the resulting Rosetta ensembles are much better MR templates and also fit the NOESY peak list data better than the Rosetta ensembles generated by our initial protocol, as manifested by the significantly improved TFZ and DP scores for the majority of the targets. These ensembles of Rosetta refined (with or without restraints) conformers were then trimmed to exclude not-well-defined atoms using FindCore and used as templates for Phaser as described previously.<sup>31</sup>

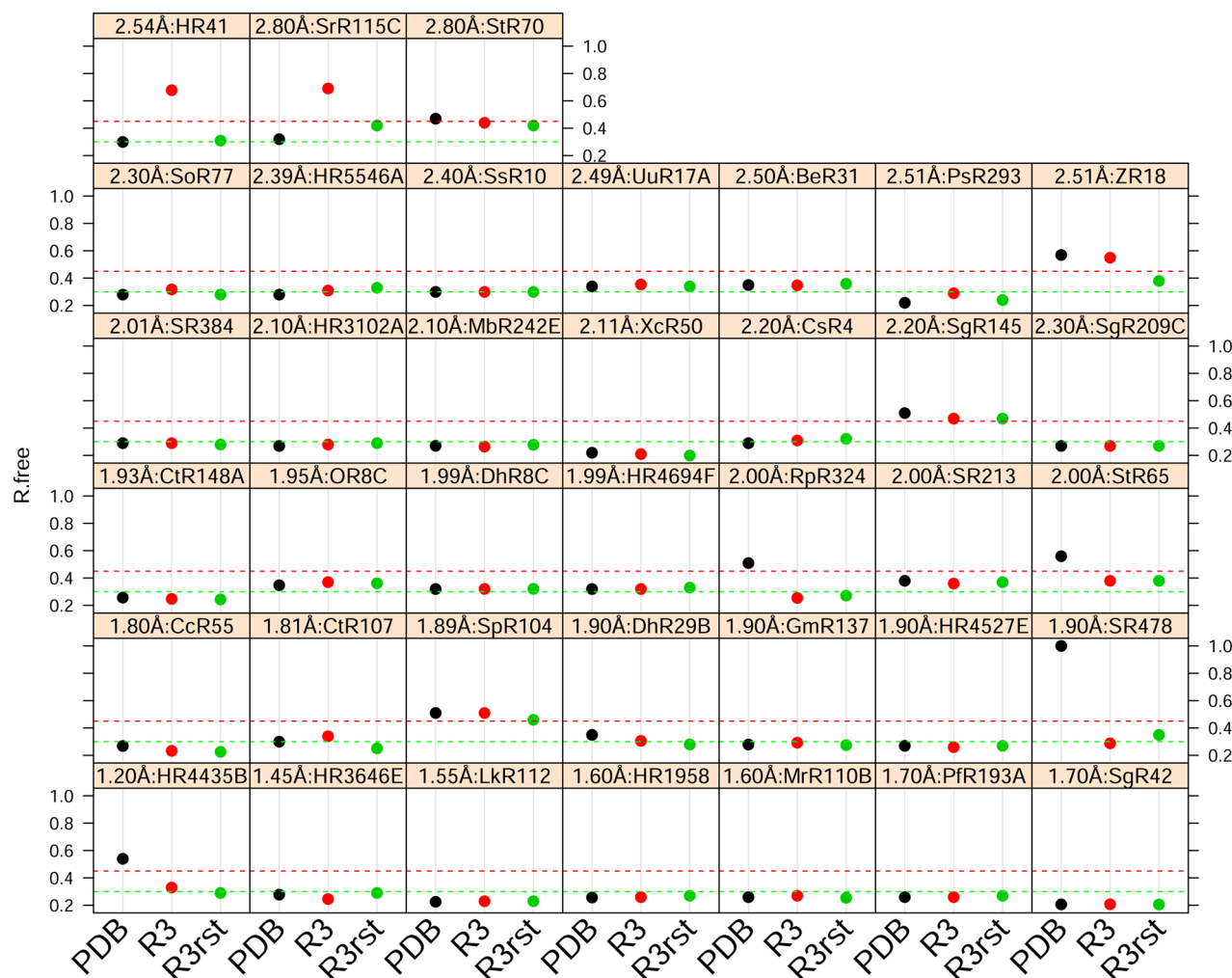
**Structures Determined Using Restrained Rosetta NMR Structures as MR Templates Are More Accurate.** Starting from Phaser MR solutions obtained by three methods [NMR PDB, unrestrained Rosetta refinement (R3), and restrained Rosetta refinement (R3rst)] for 38 NMR structures, we utilized Phenix and Arp/Warp for automatic model rebuilding and refinement. Models generated by either software with the lowest  $R_{\text{free}}$  values were chosen as the final structures solved by MR. Hence 114 crystal structures were determined



**Figure 6.** For Rosetta refinement of NMR structure, preserving ensemble information is beneficial for MR success. Scatterplot of Phaser<sup>47</sup> TFZ scores (A) and DP-scores<sup>16,17</sup> (B) for two different protocols for selecting models for MR. Decoy(Energy) Rosetta-refined structure ensembles are composed of the 20 lowest Rosetta energy decoys from the entire pool of decoys generated from all the NMR conformers. Decoy(Conformer+Energy) Rosetta-refined structure ensembles are composed of each lowest Rosetta energy decoy generated from each NMR conformer. The scores of structures picked by Decoy(Energy) protocol are plotted on the X-axis, and the scores of structures picked by Decoy(Conformer+Energy) protocol are plotted on the Y-axis. Unrestrained Rosetta refined structures are represented by red solid triangles and restrained Rosetta refined structures are represented by blue solid rectangles. Data are summarized for 38 NESG NMR/X-ray pairs used in the crystallographic MR study.

from the NMR structures and compared with the corresponding X-ray crystal structure available in the PDB. For each target, the  $R_{\text{free}}$  values of the final MR structures are plotted against the sources of their templates in Figure 7. Structures generated using the ensembles of PDB NMR structures, unrestrained Rosetta refined structures, and restrained Rosetta refined structures are represented by black, red, and green dots, respectively. The green dashed line indicates  $R_{\text{free}} = 0.3$ , and the red dashed line indicates  $R_{\text{free}} = 0.45$ . Data points above the red dashed line ( $R_{\text{free}} > 0.45$ ) are considered as failed MR solutions. Starting from NMR structures deposited in the PDB as MR templates, seven targets (ZR18, SgR145, RpR324, StR65, SpR104, SR478, HR4435B) failed to provide valid MR solutions. Four of these (RpR324, StR65, SR478, HR4435B) provided good MR solutions after Rosetta refinement with or without experimental restraints. One target (ZR18) provided a good MR solution and another target (SpR104) a borderline acceptable MR solution (GDT.TS between MR structure and X-ray structure is 0.875) only after restrained Rosetta refinement. Two targets (HR41, SrR115C), which originally provided valid MR solutions when using their PDB NMR structures as MR templates, failed to provide valid MR solutions after unrestrained Rosetta refinement but could be solved after restrained Rosetta refinement. Only one target (SgR145) failed to provide good MR solutions with any of the protocols, even after restrained Rosetta refinement. SgR145 is a sparse-restraint NMR structure,<sup>43</sup> and its  $C\alpha$  RMSD to the corresponding X-ray structure is relatively large (3.1 Å).

The same conclusion can be drawn from Figure S4, comparing the GDT.TS of the X-ray crystal structure models phased using the PDB NMR structures or Rosetta refined structures as MR templates, and autotraced, compared to the corresponding X-ray crystal structures deposited in the PDB. Most of these crystal structures deposited in the PDB were solved by anomalous dispersion (SAD or MAD) methods.



**Figure 7.** Restrained Rosetta refined NMR structures provide better templates for MR, and generally yield crystal structures with better  $R_{\text{free}}$  scores. Dotplot of  $R_{\text{free}}$  values of MR structures using MR templates for 38 NESG NMR/X-ray pairs deposited in the PDB or refined by Rosetta. The MR structures were solved either by Phenix<sup>65</sup> or Arp/WARP.<sup>63,64</sup> The  $R_{\text{free}}$  values are plotted on the Y-axis. PDB NMR structures (PDB), unrestrained Rosetta refined structures (R3) and restrained Rosetta refined structures (R3rst) are colored black, red and green respectively. Each subpanel represents one NESG target, and the subpanels are organized in ascending order of the resolution of its X-ray crystal structure from bottom left corner to top right corner.

These data further demonstrate that when the NMR structures available from the PDB are poor MR templates to start with, with  $\text{GDT.TS} < 0.8$ , their phasing power and the quality of the resulting crystal structure solution were generally significantly improved by restrained Rosetta refinement.

**Unrestrained Rosetta Refinement Can Deteriorate the Phasing Power of Good NMR MR Templates.** As is also illustrated in Figure 7 for targets SrR115C, PsR293, and HR41, if the initial NMR structures are good MR templates, their phasing power can potentially deteriorate by unrestrained Rosetta refinement. Therefore, despite the fact that unrestrained Rosetta refinement has been reported to sometimes improve the phasing power of NMR structures,<sup>30,31</sup> ignoring the experimental restraints is not recommended when preparing NMR structures for use in phasing crystallographic data by molecular replacement.

**Restrained CS-Rosetta.** For thirteen NMR structures with  $\text{GDT.TS} < 0.85$  relative to their X-ray counterparts, we also carried out restrained CS-Rosetta (rCS-Rosetta) calculations,<sup>42,43,46</sup> starting from extended conformations, using the same restraints used in the corresponding restrained Rosetta

refinement calculations. rCS-Rosetta calculations are much more CPU intensive than the restrained Rosetta refinement protocol outlined above, requiring 4–5 times more CPU time for proteins in the size range of 5–10 kDa and exponentially longer times for larger proteins. These results are summarized in Table 2. For proteins <10 kDa, the restrained CS-Rosetta structures (CS-Rrst) were slightly closer to the X-ray structure than the corresponding restrained Rosetta refined structures (R3rst), especially for targets ER382A and ZR18. On the other hand, for proteins in the 10 – 22 kDa range, the faster R3rst restrained Rosetta refinement protocol provides structures with accuracy, relative to their X-ray counterparts, similar to the computationally-intensive CS-Rrst protocol. Indeed, for the 19 kDa target HR41, the faster R3rst protocol provided a more accurate structure than the restrained CS-Rosetta protocol. These results demonstrate that the two restrained Rosetta protocols described in this work, R3rst which refines a structure initially modeled with other methods and CS-Rst which generates a structure starting from an extended conformation, are well suited for small to medium sized proteins, of up to about 25 kDa. While restrained CS-Rosetta (CS-Rst) can

**Table 2. Comparison of Unrestrained (R3) and Restrained (R3rst) Rosetta Refinement of Monomeric NMR Structures with Results of Restrained CS-Rosetta, Based on GDT.TS to Corresponding X-ray Crystal Structures**

target	length <sup>a</sup>	M <sub>w</sub> <sup>b</sup>	PDB <sup>c</sup>	R3 <sup>d</sup>	R3rst <sup>e</sup>	CS-Rrst <sup>f</sup>
ER382A	53	6.0	0.77	0.80	0.81	0.91
HR4435B	53	6.1	0.71	0.77	0.79	0.80
GmR137	70	7.5	0.78	0.79	0.80	0.84
ZR18	83	9.4	0.77	0.76	0.78	0.90
UuR17A	101	11.9	0.72	0.73	0.76	0.75
HR3646E	111	12.3	0.75	0.79	0.83	0.82
PsR293	117	13.7	0.81	0.82	0.83	0.83
SR213	123	14.5	0.81	0.83	0.85	0.86
HR5546A	133	14.6	0.79	0.79	0.79	0.80
StR70	134	15.1	0.76	0.75	0.79	0.84
SgR209C	147	16.8	0.81	0.80	0.84	0.87
HR41	167	19.5	0.82	0.78	0.84	0.74
SgR145	194	21.3	0.64	0.62	0.64	0.65

<sup>a</sup>Number of residues, excluding short disordered purification tags.

<sup>b</sup>Molecular weight (kDa). <sup>c</sup>GDT.TS score for NMR structures deposited in PDB. <sup>d</sup>GDT.TS score for unrestrained Rosetta refined structures.

<sup>e</sup>GDT.TS score for restrained Rosetta refined structures.

<sup>f</sup>GDT.TS scores for restrained CS-Rosetta structures generated from extended structures.

provide slightly more accurate structures, the improvements relative to the restrained CS-Rosetta (R3rst) results are often marginal relative to the much longer CPU times required.

## DISCUSSION

The quality of solution NMR structures is mainly determined by two factors: the accuracy and completeness of experimental data and the algorithm and energy force field used in structure calculation and refinement. In the past few years, several papers have demonstrated that unrestrained Rosetta refinement can improve the stereochemical quality of NMR structures and move NMR structures closer to X-ray crystal structures.<sup>30,31</sup> These observations may be explained by an interesting hypothesis: once the protein conformation has been placed in a near-native structure using experimental restraints, energy minimization by all-atom relaxation in the Rosetta energy field without restraints can produce a more accurate structure than is obtained using the restraints. In this interpretation, the small errors in the NMR experimental restraints, which are in conflict with the X-ray structure, can be circumvented or corrected using the unrestrained energy force field. In this study, we tested this hypothesis in a large-scale investigation of the impact of Rosetta refinement on NMR structure accuracy, and the significance of experimental restraints in Rosetta refinement. This analysis has allowed us to design a protocol for using Rosetta to improve the quality of protein NMR structures with tractable computational CPU requirements.

**Restrained Rosetta Refinement of NMR Structures.** As would be expected, restrained Rosetta refinement of NMR structures produces models with much fewer restraint violations than models generated by unrestrained Rosetta refinement. This result is significant in that it demonstrates that our restrained Rosetta refinement protocol is self-consistent with respect to a large number of experimental restraints, validating the accurate implementation of restraint conversions by PDBStat software and interpretations of these restraints by the Rosetta program. The weights of both distance and dihedral

angle restraints are set to 1. We observed that if the relative weights on restraints are too high, the final Rosetta refined models are over restrained and often end up with poor Rosetta energies. On the other hand, if the weights of restraints are too low, the final Rosetta refined models exhibit a large number of restraint violations, and the restraint information is not fully utilized.

**The X-ray Crystal Structure as a Proxy for Structural Accuracy.** An important concern in assessing NMR method development regards which structure to used as the “gold standard” of accuracy. Although the natural choice is the corresponding X-ray crystal structure, this issue has been controversial insofar as the crystal structure may be influenced by the structural and energetic requirements of intermolecular packing. For example, the crystal lattice may select for one of multiple conformational states of the protein structure. Moreover, protein X-ray crystal structures are often determined using cryoprotected crystals at ~77 K, while NMR structures are generally determined at 20–40 °C. None the less, as we have demonstrated elsewhere,<sup>31</sup> except under special circumstances, the solution NMR structure is generally quite similar to the crystal structure and can be used for phasing by molecular replacement methods. Hence, we contend that the crystal structure is an excellent proxy for NMR structure accuracy. The availability of these 40 NMR/X-ray pairs together with extensive raw experimental NMR and diffraction data (summarized in Supplemental Tables S1 and S2) will greatly facilitate the testing and development of new methods for protein NMR structure refinement and analyses of subtle structural differences between crystal and solution NMR structures.

**Assessment of Structures Resulting from Restrained Rosetta Refinement.** Judged by ensemble RMSD analysis, unrestrained Rosetta refinement generally decreases the precision of NMR structures, while restrained Rosetta refinement can increase the precision of the side chain heavy atoms of otherwise well-defined residues. Additionally, restrained Rosetta refined structures fit the unassigned NOESY peak list data significantly better than unrestrained Rosetta refined structures. Rosetta refinement can generally improve the stereochemical quality and geometry of NMR structures. More specifically, the experimental backbone dihedral angle restraints can guide Rosetta to generate models with even better backbone structures than is achieved without restraints. In most cases, restrained Rosetta refinement will move protein NMR structures closer to their X-ray counterparts, while unrestrained Rosetta refinement often fails to do so, especially when the structural similarity between the NMR and X-ray structures is high (GDT.TS > 0.85). For NMR structures with poor phasing power, Rosetta refinement can often be used to generate MR templates which are better able to guide phasing software, such as Phaser, to identify correct MR solutions. The phasing power of the template and the accuracy of the resulting crystal structures are better when experimental restraints are utilized in Rosetta refinement. Indeed, unrestrained Rosetta refinement can sometimes make NMR structures less useful MR templates, even when they are good MR templates to start with.

With respect to our hypothesis regarding final-stage unrestrained Rosetta refinement providing more accurate structures than can be achieved using all of the experimental restraints, this comprehensive study with 40 NMR/X-ray pairs demonstrates that the majority of NMR experimental restraints are



completely consistent with the corresponding X-ray structures. While in some cases, a few inaccurate restraints may be identified using an unrestrained Rosetta refinement protocol as proposed by Ramelot et al.,<sup>30</sup> the most accurate structures, with the highest phasing power, were obtained by combining the experimental restraints with the sophisticated algorithms and the more advanced force field of Rosetta.

We also computed Rosetta energies of “relaxed X-ray” and “relaxed NMR” structures. The structures are first idealized using Rosetta idealization application and then are relaxed in Rosetta all-atom energy field. The relaxed X-ray structures generally have lower Rosetta energy per residue, while the relaxed NMR structures, the R3-refined NMR structures, and the R3rst-refined NMR structures generally have slightly higher, similar Rosetta energies. This suggests that even our R3rst-refined NMR structures have some room for improvement in terms of their energies.

**Residual Restraint Violations in Restrained Rosetta Refined Protein NMR Structures.** While they are more accurate relative to the corresponding X-ray crystal structure and generally satisfy the experimental restraint data, restrained Rosetta refined structures have modestly more small distance restraint violations than the NMR structures from which they are derived. As discussed in detail elsewhere,<sup>46</sup> these small restraint violations associated with Rosetta refined NMR structures may reflect inaccuracies in the interpretation of upper-bound distance restraints from NOESY data due various effects, including relaxation-modulation of NOE intensities in heteronuclear filtered NOESY data and the effects of dynamic averaging.

It is generally not possible to satisfy all of the experimental restraints in a R3rst refinement. This is evidenced first from the fact that the lowest-energy Rosetta structures generated in this study [i.e., the unrestrained Rosetta refined (R3) structures] have poorer agreement with the NOESY peak list data (i.e., lower DP scores summarized in Figure 2) than restrained Rosetta refined (R3rst) structures. The R3 structures also often diverge from the X-ray structures (Figure 4), our best proxy for a ‘gold standard’. The inconsistency between the NOE-derived distance restraints and Rosetta energy terms is also illustrated in analyses presented in Supplementary Figure S5 and Table S3; increasing the relative weight on restraint terms allows excellent satisfaction of restraints but results in structures with higher Rosetta conformational energies. Accordingly, there is some fundamental inconsistency between our minimum Rosetta energy structures, the NOESY peak list data, the NMR restraints, and the “gold standard” X-ray crystal structures.

In order to further investigate potential inaccuracies on the NOE-derived restraints themselves, we also assessed how well the corresponding X-ray crystal structures fit these NMR restraint data. Hydrogen atoms were added to the X-ray structure coordinates using Rosetta idealization application, which rebuilds molecules using ideal bond lengths, bond angles, and torsion angles. All those resulting idealized X-ray structures have quite a few restraint violations, and the number of restraint violations varies from target to target. No attempt was made to further adjust the X-ray crystal structures to better match the NMR restraint data. These results are summarized in Supplementary Table S4.

Inconsistencies between the NMR restraints, the X-ray crystal structures, and the Rosetta energy function arise from several sources, including (i) the crystal lattice, which may stabilize a subset of the conformations that are present in the

solution NMR experiments and that contribute to the NOESY data, (ii) the NOESY data arise from ensemble averages of dynamic distributions which are not captured by the methods used in these studies to model restraints and NMR structures, and (iii) there may be inaccuracies in the Rosetta potential energy function. However, aside from these fundamental challenges in modeling protein structures from NMR data, it is not surprising that the lowest-energy models do not perfectly satisfy the NOE-derived distance restraints. These distance restraints are interpreted from NOESY spectra assuming a simple two-spin approximation, single isotropic rotational correlation time, uniform linewidths, identical relaxation in filtering through bound C and N atoms, and many other assumptions that are simply not correct. The details of how upper bound distance restraint violations were defined are different as various laboratories across the NESG use somewhat different methods for calibrating these distances. Although NOESY peak lists (usually providing resonance intensities) are available for many of the data sets, the issues of linewidths and differential relaxation in different X-filtered NOESY spectra cannot be addressed with the available data. It could be interesting to compare simulated spectra generated for R3 and R3rst models using full-relaxation matrix analysis with relaxation-corrected, integrated, NOESY spectra data, but such an analysis is beyond the scope of the current work.

#### Comparison with Restrained CS-Rosetta Calculations.

An alternative method for incorporating the Rosetta force field into the NMR structure determination process is restrained CS-Rosetta (rCS-Rosetta) in which structures are generated starting from extended conformations with NMR restraints and the CS-Rosetta protocol. Generally, for a 100-residue protein rCS-Rosetta calculations require about 5–10 times more CPU time to generate each decoy than restrained Rosetta refinement. The difference in CPU time becomes even larger as the size of protein increases. rCS-Rosetta calculations generally require tens of thousands of decoys in order to ensure convergence, compared to only hundreds of decoys required for restrained Rosetta refinements which begin with native-like conformations as the starting point. Hence, the restrained Rosetta refinement protocols used here are some 200–500 (or more) times faster than restrained CS-Rosetta methods. For example, for a 100 residue protein, rCS-Rosetta calculations required about 4000 min to generate 10 000 decoys using 20 2.5 GHz processors (0.4 min per decoy). For the same size protein, Cyana structure generation followed by restrained Rosetta refinement requires about 10–20 min per for an ensemble of 20 conformers. The RASAC iterative CS-Rosetta protocol<sup>43,71,72</sup> may sometimes provide more accurate structures using restrained CS-Rosetta, but it is even more CPU intensive. Therefore, for proteins of more than 10 kDa, a good practice is to use traditional methods for NMR structure generation, followed by restrained Rosetta refinement.

Although Rosetta refinement can modify the input conformation to some extent, Rosetta refined structures will not deviate significantly from the input structure because the Rosetta refinement protocol samples only conformations that are close to the initial NMR structure. If the NMR structures are highly inaccurate to begin with, these severe structural differences cannot be corrected by the restrained Rosetta refinement protocol alone. Moreover, for sparse-restraint NMR structures, such as the SgR145 target, additional information, such as evolutionary restraints,<sup>60</sup> or more advanced sampling

techniques, such as RASREC Rosetta,<sup>43,69,70</sup> may also be required to obtain the most accurate NMR structures.

## CONCLUSIONS

In comparison with NMR structures refined by traditional methods, restrained Rosetta refined structures fit the experimental NMR data equally well and are of significantly better stereochemical and geometric quality. Rosetta refinement drives NMR structures to be more similar to their X-ray counterparts, thus increasing their phasing power. Despite the fact that they are more accurate relative to the corresponding X-ray crystal structure, restrained Rosetta refined structures tend to have slightly higher distance restraint violations. This may reflect inaccuracies in the interpretation of NMR data in terms of upper bound restraints, providing guidance to the experimentalist to confirm and possibly refine these interpretations of the raw experimental data. The restrained Rosetta refinement protocols described here utilize NMR structures initially determined by more conventional methods as input. They are much less CPU intensive than restrained CS-Rosetta methods, which generate NMR structures from extended starting structures, and provide comparable or better results.

**Data Deposition.** All of the NMR and crystallographic experimental data used in this project are available on line at: [http://psvs-1\\_4-dev.nesg.org/results/rosetta\\_MR/data\\_set.html](http://psvs-1_4-dev.nesg.org/results/rosetta_MR/data_set.html). Coordinates of the unrestrained and restrained Rosetta refined structures, together with structure quality assessment reports, are available on line at: [http://psvs-1\\_4-dev.nesg.org/results/rosetta\\_MR/rosettaMR\\_PSVS\\_summary.html](http://psvs-1_4-dev.nesg.org/results/rosetta_MR/rosettaMR_PSVS_summary.html)

## ASSOCIATED CONTENT

### Supporting Information

Detailed protocols for restrained Rosetta refinement, unrestrained Rosetta refinement, and restrained CS-Rosetta; restraint conversion and implementation in restrained Rosetta calculations and determination of the relative weights of restraint violations and Rosetta energy terms; figures comparing NMR ensemble RMSD scatterplots, RMSDs within the NMR structure ensembles deposited in the PDB with those refined with unrestrained or restrained Rosetta protocols, RMSD to corresponding X-ray crystal structure for NMR structures deposited in the PDB and the same structures following unrestrained Rosetta refinement, and RMSD to corresponding X-ray crystal structure for NMR structures deposited in the PDB and the same structures following restrained Rosetta refinement; quality assessment of X-ray crystal structures solved by MR using NMR structures or Rosetta refined NMR structures as templates; assessment of optimum restraint weight  $W$  determining the relative contribution of Restraint and Rosetta energies to the total energy target function; tables with experimental protein NMR and X-ray crystallography data sets, NMR and X-ray crystal structures, structure factor files, NOESY peak list, and NMR resonance assignments determined by the Northeast Structure Genomics Consortium ([www.nesg.org](http://www.nesg.org)) and deposited in the PDB, structure quality statistics for NMR structures, benchmark data used for optimization of  $W$ , and NMR upper-bound restraint violations present regularized X-ray crystal structures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

[guy@cabm.rutgers.edu](mailto:guy@cabm.rutgers.edu)

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank all the members of the several NMR and X-ray crystallography groups of the Northeast Structural Genomics Consortium who contributed constructive criticisms and the test data sets used in this work. Experimental data and structures used in this work were provided from the laboratories of C. Arrowsmith, J. Hunt, R. Powers, L. Tong, T. Szyperki, and J. Prestegard as well as from our own laboratory. We also thank J. Aramini, R. Guan, Y. J. Huang, G. Liu, Y. Tang, and G. V. T. Swapna for helpful discussions. This work was supported by a grant from the Protein Structure Initiative of the National Institutes of Health grant U54-GM094597. R.T. also acknowledges support from CONSOLIDER INGENIO CSD2010-00065 and Generalitat Valenciana PROMETEO 2011/008.

## REFERENCES

- (1) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65.
- (2) Brunger, 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.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 905.
- (3) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283.
- (4) Güntert, P. *Methods Mol. Biol.* **2004**, *278*, 353.
- (5) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
- (6) Xia, B.; Tsui, V.; Case, D. A.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **2002**, *22*, 317.
- (7) Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. *Proteins* **2003**, *50*, 496.
- (8) Feig, M.; Brooks, C. L., III *Curr. Opin. Struct. Biol.* **2004**, *14*, 217.
- (9) Chen, J.; Im, W.; Brooks, C. L., III *J. Am. Chem. Soc.* **2004**, *126*, 16038.
- (10) Chen, J.; Brooks, C. L., III; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140.
- (11) Nederveen, A. J.; Doreleijers, J. F.; Vranken, W.; Miller, Z.; Spronk, C. A.; Nabuurs, S. B.; Guntert, P.; Livny, M.; Markley, J. L.; Nilges, M.; Ulrich, E. L.; Kaptein, R.; Bonvin, A. M. *Proteins* **2005**, *59*, 662.
- (12) Nabuurs, S. B.; Nederveen, A. J.; Vranken, W.; Doreleijers, J. F.; Bonvin, A. M.; Vuister, G. W.; Vriend, G.; Spronk, C. A. *Proteins* **2004**, *55*, 483.
- (13) Lee, S. Y.; Zhang, Y.; Skolnick, J. *Proteins* **2006**, *63*, 451.
- (14) Yang, J. S.; Kim, J. H.; Oh, S.; Han, G.; Lee, S.; Lee, J. *Nucleic Acids Res.* **2012**, *40*, D525.
- (15) Doreleijers, J. F.; Raves, M. L.; Rullmann, T.; Kaptein, R. *J. Biomol. NMR* **1999**, *14*, 123.
- (16) Huang, Y. J.; Powers, R.; Montelione, G. T. *J. Am. Chem. Soc.* **2005**, *127*, 1665.
- (17) Huang, Y. J.; Rosato, A.; Singh, G.; Montelione, G. T. *Nucleic Acids Res.* **2012**, *40*, W542.
- (18) Bagaria, A.; Jaravine, V.; Huang, Y. J.; Montelione, G. T.; Guntert, P. *Protein Sci.* **2012**, *21*, 229.
- (19) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923.
- (20) Valafar, H.; Prestegard, J. H. *J. Magn. Reson.* **2004**, *167*, 228.
- (21) Bhattacharya, A.; Tejero, R.; Montelione, G. T. *Proteins* **2007**, *66*, 778.
- (22) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. *Proteins* **1992**, *12*, 345.
- (23) Laskowski, R. A.; Rullmann, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477.
- (24) Sippl, M. J. *Proteins* **1993**, *17*, 355.

- (25) Luthy, R.; Bowie, J. U.; Eisenberg, D. *Nature* **1992**, *356*, 83.
- (26) Davis, I. W.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res.* **2004**, *32*, W615.
- (27) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12.
- (28) Rosato, A.; Bagaria, A.; Baker, D.; Bardiaux, B.; Cavalli, A.; Doreleijers, J. F.; Giachetti, A.; Guerry, P.; Guntert, P.; Herrmann, T.; Huang, Y. J.; Jonker, H. R.; Mao, B.; Malliavin, T. E.; Montelione, G. T.; Nilges, M.; Raman, S.; van der Schot, G.; Vranken, W. F.; Vuister, G. W.; Bonvin, A. M. *Nat. Methods* **2009**, *6*, 625.
- (29) Rosato, A.; Aramini, J. M.; Arrowsmith, C.; Bagaria, A.; Baker, D.; Cavalli, A.; Doreleijers, J. F.; Eletsky, A.; Giachetti, A.; Guerry, P.; Gutmanas, A.; Guntert, P.; He, Y.; Herrmann, T.; Huang, Y. J.; Jaravine, V.; Jonker, H. R.; Kennedy, M. A.; Lange, O. F.; Liu, G.; Malliavin, T. E.; Mani, R.; Mao, B.; Montelione, G. T.; Nilges, M.; Rossi, P.; van der Schot, G.; Schwalbe, H.; Szyperski, T. A.; Vendruscolo, M.; Vernon, R.; Vranken, W. F.; Vries, S.; Vuister, G. W.; Wu, B.; Yang, Y.; Bonvin, A. M. *Structure* **2012**, *20*, 227.
- (30) Ramelot, T. A.; Raman, S.; Kuzin, A. P.; Xiao, R.; Ma, L. C.; Acton, T. B.; Hunt, J. F.; Montelione, G. T.; Baker, D.; Kennedy, M. A. *Proteins* **2009**, *75*, 147.
- (31) Mao, B.; Guan, R.; Montelione, G. T. *Structure* **2011**, *19*, 757.
- (32) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. *J. Mol. Biol.* **1997**, *268*, 209.
- (33) Simons, K. T.; Bonneau, R.; Ruczinski, I.; Baker, D. *Proteins* **1999**, *Suppl 3*, 171.
- (34) Misura, K. M.; Chivian, D.; Rohl, C. A.; Kim, D. E.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5361.
- (35) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364.
- (36) Qian, B.; Raman, S.; Das, R.; Bradley, P.; McCoy, A. J.; Read, R. J.; Baker, D. *Nature* **2007**, *450*, 259.
- (37) DiMaio, F.; Tyka, M. D.; Baker, M. L.; Chiu, W.; Baker, D. *J. Mol. Biol.* **2009**, *392*, 181.
- (38) DiMaio, F.; Terwilliger, T. C.; Read, R. J.; Wlodawer, A.; Oberdorfer, G.; Wagner, U.; Valkov, E.; Alon, A.; Fass, D.; Axelrod, H. L.; Das, D.; Vorobiev, S. M.; Iwai, H.; Pokkuluri, P. R.; Baker, D. *Nature* **2011**, *473*, 540.
- (39) Terwilliger, T. C.; Dimaio, F.; Read, R. J.; Baker, D.; Bunkoczi, G.; Adams, P. D.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Echols, N. *J. Struct. Funct. Genomics* **2012**, *13*, 81.
- (40) Adams, P. D.; Baker, D.; Brunger, A. T.; Das, R.; DiMaio, F.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C. *Annu. Rev. Biophys.* **2013**, *42*, 265.
- (41) Raman, S.; Huang, Y. J.; Mao, B.; Rossi, P.; Aramini, J. M.; Liu, G.; Montelione, G. T.; Baker, D. *J. Am. Chem. Soc.* **2010**, *132*, 202.
- (42) Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D. *Science* **2010**, *327*, 1014.
- (43) Lange, O. F.; Rossi, P.; Sgourakis, N. G.; Song, Y.; Lee, H. W.; Aramini, J. M.; Ertekin, A.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10873.
- (44) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res.* **2008**, *36*, D402.
- (45) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
- (46) Tejero, R.; Snyder, D.; Mao, B.; Aramini, J. M.; Montelione, G. T. *J. Biomol. NMR* **2013**, *56*, 337.
- (47) Chivian, D.; Kim, D. E.; Malmstrom, L.; Bradley, P.; Robertson, T.; Murphy, P.; Strauss, C. E.; Bonneau, R.; Rohl, C. A.; Baker, D. *Proteins* **2003**, *53* (Suppl 6), S24.
- (48) Kim, D. E.; Chivian, D.; Baker, D. *Nucleic Acids Res.* **2004**, *32*, W526.
- (49) Hyberts, S. G.; Goldberg, M. S.; Havel, T. F.; Wagner, G. *Protein Sci.* **1992**, *1*, 736.
- (50) Snyder, D. A.; Montelione, G. T. *Proteins* **2005**, *59*, 673.
- (51) Bradley, P.; Malmstrom, L.; Qian, B.; Schonbrun, J.; Chivian, D.; Kim, D. E.; Meiler, J.; Misura, K. M.; Baker, D. *Proteins* **2005**, *61*, 128.
- (52) Misura, K. M.; Baker, D. *Proteins* **2005**, *59*, 15.
- (53) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213.
- (54) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685.
- (55) Vernon, R.; Shen, Y.; Baker, D.; Lange, O. F. *J. Biomol. NMR* **2013**, *57*, 117.
- (56) Zemla, A. *Nucleic Acids Res.* **2003**, *31*, 3370.
- (57) Zhang, Y.; Skolnick, J. *Proteins* **2004**, *57*, 702.
- (58) DeLano, W. L. *The PyMOL Molecular Graphics System*, version 1.5.0.4; Schrödinger, LLC: Portland, OR, 2002.
- (59) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
- (60) Joosten, R. P.; te Beek, T. A.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W.; Schneider, R.; Sander, C.; Vriend, G. *Nucleic Acids Res.* **2011**, *39*, D411.
- (61) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
- (62) Saff, E. B.; Kuijlaars, A. B. J. *Math. Intell.* **1997**, *19*, 5.
- (63) Perrakis, A.; Harkiolaki, M.; Wilson, K. S.; Lamzin, V. S. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 1445.
- (64) Cohen, S. X.; Ben Jelloul, M.; Long, F.; Vagin, A.; Knipscheer, P.; Lebbink, J.; Sixma, T. K.; Lamzin, V. S.; Murshudov, G. N.; Perrakis, A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 49.
- (65) Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L. W.; Read, R. J.; Adams, P. D. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 61.
- (66) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240.
- (67) Snyder, D. A.; Bhattacharya, A.; Huang, Y. J.; Montelione, G. T. *Proteins* **2005**, *59*, 655.
- (68) Kossiakoff, A. A.; Randal, M.; Guenot, J.; Eigenbrot, C. *Proteins* **1992**, *14*, 65.
- (69) Andrec, M.; Snyder, D. A.; Zhou, Z.; Young, J.; Montelione, G. T.; Levy, R. M. *Proteins* **2007**, *69*, 449.
- (70) Montelione, G. T.; Nilges, M.; Bax, A.; Guntert, P.; Herrmann, T.; Richardson, J. S.; Schwieters, C. D.; Vranken, W. F.; Vuister, G. W.; Wishart, D. S.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L. *Structure* **2013**, *21*, 1563.
- (71) Lange, O. F.; Baker, D. *Proteins* **2012**, *80*, 884.
- (72) van der Schot, G.; Zhang, Z.; Vernon, R.; Shen, Y.; Vranken, W. F.; Baker, D.; Bonvin, A. M.; Lange, O. F. *J. Biomol. NMR* **2013**, *57*, 27.