RPF: a quality assessment tool for protein NMR structures

Yuanpeng Janet Huang[1], Antonio Rosato[2], Gautam Singh[1] and Gaetano T. Montelione[1,3,*]

[1]Center for Advanced Biotechnology and Medicine, Northeast Structural Genomic Consortium, Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854, USA, [2]Magnetic Resonance Center and Department of Chemistry, University of Florence, 50019 Sesto Fiorentino, Italy and [3]Robert Wood Johnson Medical School, University of Medicine and Dentistry of NJ, 675 Hoes Lane, Piscataway, NJ 08854, USA

ABSTRACT

We describe the RPF web server, a quality assessment tool for protein NMR structures. The RPF server measures the 'goodness-of-fit' of the 3D structure with NMR chemical shift and unassigned NOESY data, and calculates a discrimination power (DP) score, which estimates the differences between the fits of the query structures and random coil structures to these experimental data. The DP-score is an accuracy predictor of the query structure. The RPF server also maps local structure quality measures onto the 3D structure using an online molecular viewer, and onto the NMR spectra, allowing refinement of the structure and/or NOESY peak list data. The RPF server is available at: http://nmr.cabm.rutgers.edu/rpf.

INTRODUCTION

Protein NMR spectroscopy provides infrastructure for research in biophysical chemistry. One of the challenges of protein structure determination by NMR is the lack of a broadly accepted 'R factor', comparing 3D structures with the raw, uninterpreted experimental data. Such R factors have been critical to the development of X-ray crystallography as a routine protein structure analysis method (1). Instead, NMR structures are generally validated against the derived experimental distance constraint lists, which are an interpreted and incomplete representation of the data in the NOESY and other NMR spectra (2).

RPF is an 'R-factor'-like protein structure validation tool, which assesses the completeness of experimental data and its agreement with the 3D structure (3). Because it is difficult to compare structures directly against raw experimental NMR spectral data, these analyses were performed with respect to minimally interpreted experimental data, i.e. NOESY spectra peak lists and resonance assignments. RPF also calculates a discriminating power (DP) score that estimates how well the query structure satisfies the data relative to a statistical random-coil structure. The DP-score ranges from 0 to 1 (3).

The RPF protein structure quality assessment program has been used by the Northeast Structural Genomics (NESG) Consortium of the NIGMS Protein Structure Initiative over the last several years. It is a core component of the Protein Structure Validation Server (PSVS) analysis (4), and has been applied in the assessment and/or refinement of more than 400 protein NMR structures. RPF has also been used as a key component of the recently developed CS-DP-Rosetta method (5), the GLM-RMSD accuracy prediction score (6) and the Critical Assessment of Automated Protein Structure Determination by NMR (CASD-NMR) project (7,8).

Some commonly used knowledge-based protein structure validation tools that assess the geometric and stereochemical quality of the structure include (i) Verify3D (9), (ii) ProsaII scores (10), which evaluate the global fold likelihood, (iii) PROCHECK scores (11), which assess the distribution of backbone and side-chain dihedral angles and the (iv) MolProbity clash score (12), which assesses the occurrence of high-energy interatomic contacts. We examined the correlations between these quality scores (including the DP score) and accuracy of the structures using 63 protein structure ensembles generated in the CASD-NMR2010 project (8). In this communication, we summarize data showing that, of these measures, only the DP-score has significant correlation with the accuracy of the protein structures.

CALCULATING RPF/DP SCORES

The algorithm to calculate RPF scores (i.e. Recall, Precision, *F*-measure) and the DP-score are described elsewhere (3). Briefly, Recall measures the percentage of input NOESY peaks that can be explained by the input query

---

structure(s) with a distance cut-off $\leq 5\,\text{Å}$. Precision measures the percentage of ^1H–^1H distances $\leq 5\,\text{Å}$ calculated from the query structure that are observed in the NOESY data. *F*-measure combines the Recall and Precision scores, and estimates how well the input NMR structure ensemble fits with the input NMR data. DP score is a normalized score of *F*-measure, which estimates the significance of the F-measure score for the query structure relative to what would be obtained for a random-coil structure fit to the same experimental data.

The *F*-measure provides an assessment of the overall fit between a query model structure and the experimental data. Low *F* scores indicate that the query structure does not fit well with the data. A high-quality NMR structure is expected to both (i) fit well to the NMR data (i.e. high *F*-measure score) and (ii) have enough long-range contacts to distinguish it from a freely rotating chain model (i.e. high DP scores). High *F* scores and low DP scores indicate that the NMR data does not have enough long-range information to distinguish the structures from a 'random coil' structure.

Calculating the Precision score requires identifying all ^1H–^1H distances $\leq 5\,\text{Å}$ from the query structure. Identifying all ^1H–^1H distances $\leq 5\,\text{Å}$ is a typical 3D range-searching problem in computational geometry (13). In the most recent version of RPF, we have implemented the *k-D* tree algorithm (14) to speed up the ^1H–^1H distance calculation time. Using the *k-D* tree, a set of n ^1H can be preprocessed in $O(n \log n)$ time into a data structure of $O(n)$ size so that any 3-D range query can be answered in $O(n^{1/3}+k)$ time, where k is the number of answers reported (14). Without using the *k-D* tree algorithm, the query time will be $O(n^2)$ time, which becomes prohibitively time consuming for larger-size proteins or for studies involving the assessment of hundreds of decoys (5).

THE RPF WEB SERVER

The input files required for RPF are: (i) the atomic coordinate files in PDB format, (ii) chemical shift data in BMRB format and (iii) NOESY peak lists in Sparky or Xeasy format. Examples of input data are provided on the home page of the web server.

The RPF server reports the quality scores for each individual conformer in the NMR ensemble, and for the structure ensemble as a whole (i.e. using the mid-range ^1H–^1H distances of the ensemble). We observe that the RPF scores for the ensemble as a whole are generally higher (i.e. fit the NOESY peak list data better) than scores for the individual conformers. Accordingly, the ensemble is a better representation of the uncertainty and/or the dynamic conformational averaging effects that underlie the NOESY data.

Precision Violations (i.e. false-positive interactions) are short ^1H–^1H distances in the query structures that are not supported by NOESY peak list data. The first result page displays the distribution of the Precision Violations on the query structure using the java viewer Jmol (15). The color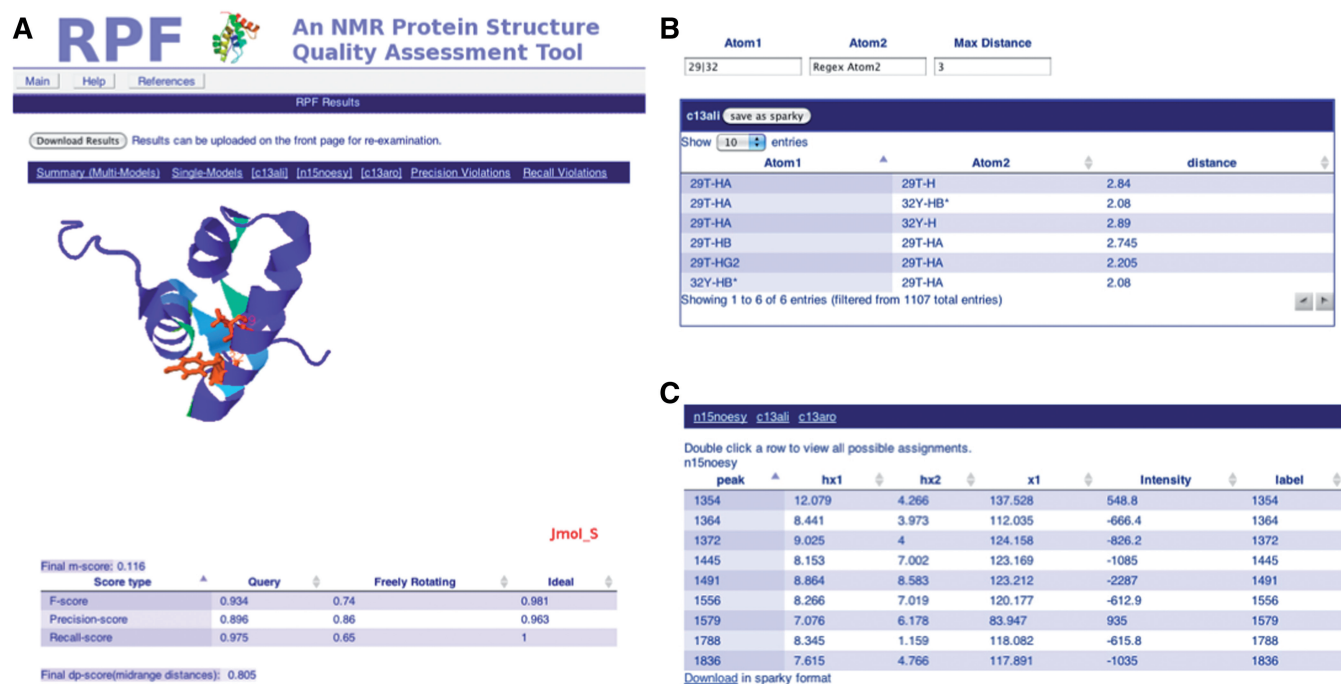 is coded based on a heat index where red represents residues with extensive (many and/or large) Precision Violations, and blue represents residues with few or no Precision Violations (Figure 1A) (3). In Figure 1A, as an example, residues 29 and 32 are colored red, indicating that some of the very short ^1H–^1H distances observed in the query structure is not supported by the combined NOESY peak list and chemical shift data. Such Precision Violations generally arise from either inaccurate local structure or inaccurate resonance assignments, or the effects of NMR resonance broadening due to intermediate timescale conformational exchange (3).

The 'Precision Violations' report summarizes all ^1H–^1H distances $\leq 5\,\text{Å}$ in the query structures that are not supported by the NOE peak list data. It is possible to use regular expressions to filter the list of Precision Violations. Figure 1B illustrates an example using a regular expression search for Precision Violations involving residues 29 and 32 with max distance of $3.0\,\text{Å}$. The 'Recall Violations' page (i.e. false-negative interactions) (Figure 1C) reports the resonance frequencies of peaks in the NOESY spectrum that, considering all possible assignments of the NOESY peak, are not consistent with the 3D query structure(s). These 'Precision Violations' and 'Recall Violations' are local quality score measures, which can be overlooked when looking at global RPF and DP scores. Precision Violations and Recall Violations provided in these reports can be mapped back to the 3D structure and NMR spectrum, respectively, providing guidance for further peak list and/or 3D structure refinement. This validation process is used extensively by NESG consortium NMR scientists in the final stages of protein structure refinement.

The RPF web server provides a web-service for large-scale NMR structure quality assessments. The user can also save the RPF results locally and review them again on the website by uploading the file to the RPF web server. Sample data, including input files and results, are provided on the home page of the RPF web server. The Help Page of the RPF web server includes information on how to interpret the RPF results. Sample codes to access the RPF web-service are also provided on the Help Page.

CORRELATION OF DP SCORE WITH THE ACCURACY OF PROTEIN NMR STRUCTURES

The Critical Assessment of Automated Protein Structure Determination by NMR (CASD-NMR) (8) is an international NMR community project in which refined NOESY peak lists and resonance assignment lists are distributed while the manually refined NMR structure in held in confidence. Participants then carry out fully automated NMR structure analysis with these 'blind' data, which are subsequently compared with the manually refined NMR structure and/or X-ray crystal structures. In the 2010 session of CASD-NMR (CASD-NMR2010), participants submitted 63 NMR protein structure ensembles for 10 monomeric proteins, ranging in size from 60 to 150 amino acid residues. (8). For each of these structure ensembles, we calculated the backbone RMSD to the

Figure 1. RPF output. (**A**) The distribution of the Precision Violations (a.k.a. false-positive interactions) mapped on the query structure based on a heat index. Red represents residues with strong Precision Violations and blue represents residues with few or no Precision Violations. In this example, residues 29 and 32 are colored red, indicating that several very short distances based on the input structure do not have corresponding NOE data in the NOESY peak list and/or one or more of the corresponding resonances are mis-assigned in the chemical shift list. (**B**) The 'Precision Violations' page reports all distances ≤5.0 Å calculated from the query structures that are not supported by the NOESY data. In this example, there are six Precision Violations involving residues 29 or 32 with max distance of 3.0 Å. (**C**) The 'Recall Violations' page reports the input NOESY peaks that are not supported by the query structures within the average distance of 5.0 Å.

corresponding manually refined structure in the PDB. We refer to this measure of accuracy (assuming that the manually refined reference structures are correct) as the *RMSD bias*. We also computed the global distance test total score (GDT_TS) (16), a structural similarity measure that does not require residue ranges to be pre-defined with RMSD calculations and is independent of protein size. The GDT_TS score has been developed as a local–global alignment method for structure comparison, and has been extensively used for assessing the accuracy of protein structure predictions in CASP assessments (17). High structural similarity corresponds to low RMSD and high GDT_TS values.

The DP score from the RPF program, along with five additional geometric and stereochemical quality scores, were calculated for each of the submitted 63 CASD-NMR2010 structure ensembles using the Protein Structure Validation Server (PSVS) (4). The five knowledge-based structure quality scores assessed included PROCHECK-Φ/ψ score (11), the PROCHECK-All dihedral score (11), the Molprobity clash score (12), the Verify3D fold score (9) and the ProsaII fold score (10). Using these 63 protein structure ensembles, a significant correlation is observed between the DP-score and the structure accuracy (Figure 2). However, no significant correlation is observed between any of the other five knowledge-based validation scores and the RMSD bias (Table 1).

We define a structure as 'accurate' when the condition (i) backbone RMSD ≤ 2.0 Å or (ii) GDT_TS ≥ 80 is met.

Table 2 summarizes the confusion matrix and metrics for accuracy prediction on the basis of the DP score. Very few false-positive and false-negative errors are found for the CASD-NMR2010 structures (Table 2) (8). The range of DP-scores for the manually refined reference structures is 0.79–0.90, except for one (AR3426A, 0.64; in this case, the NOESY data are unusually weak; many expected NOEs with very close distances have rather weak intensities or are missing from the spectra). A DP-score cut-off ≥0.7 allowed the identification of acceptable accurate CASD-NMR2010 structures with a reliability of 95% (Table 2), based on the available NOESY peak lists. All structures with an RMSD to the reference >3.0 Å or a GDT_TS score <60% had DP-scores lower than 0.6, except for a single instance. Based on these data, we conclude that a protein NMR structure will usually satisfy our definition of 'accurate' when its ensemble DP-score is ≥0.7.

DISCUSSION

RPF versus NOE restraint violation scores

NOE restraint violation statistics measure the fitness of structure coordinates with the NOE-derived distance restraints. Several protein structure validation servers compute both restraint violation statistics and DP score [e.g. the PSVS server (4)]. A high quality structure tends to have high DP score and also low NOE restraint violations.
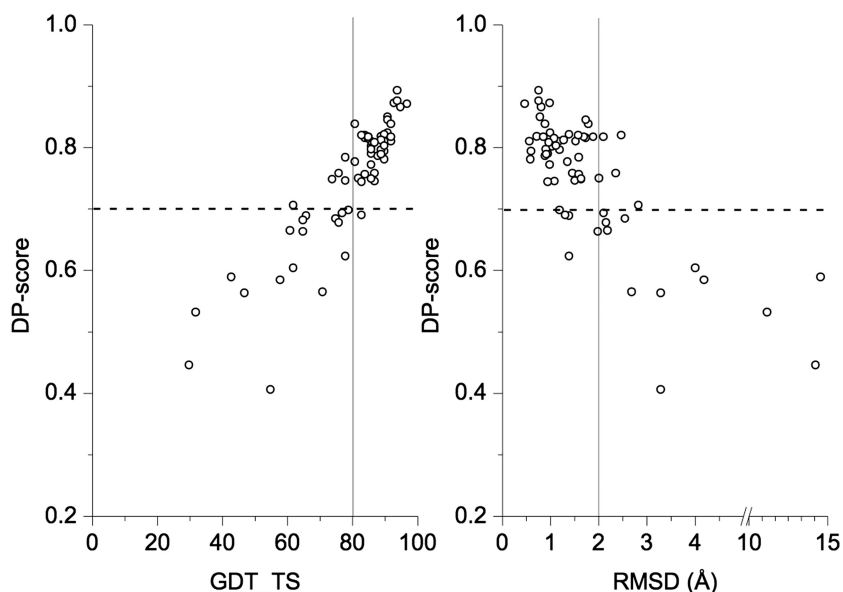
Figure 2. Correlation between accuracy measures (backbone RMSD to the reference structure and GDT_TS score) and the DP-score. The various thresholds mentioned in the text are highlighted by the continuous (RMSD ≤ 2 Å; GDT_TS ≥ 80) and dashed (DP-score ≥ 0.7) lines. These results demonstrate the discriminating power of the DP score in distinguishing accurate from less accurate protein NMR models.

Table 1. Pearson's correlation coefficient between various accuracy and quality scores for the same data shown in Figure 2

	DP-score	Verify3D	ProsaII	PROCHECK (phi–psi)	PROCHECK (all)	MolProbity clash score
RMSD	−0.659	−0.139	−0.156	0.108	0.257	0.065
GDT_TS	0.887	0.283	0.260	−0.065	−0.246	−0.085

Table 2. Confusion matrix and metrics for accuracy prediction on the basis of the DP-score

	Success	
	Positive	Negative
DP-score prediction		
Positive	44 (TP[a])	2 (FP[b])
Negative	4 (FN[c])	13 (TN[d])
Metrics		
Sensitivity [TP/(TP + FN)]		0.917
Specificity [TN/(TN + FP)]		0.867
Precision[e] [TP/(TP + FP)]		0.957

[a]True positives (TP) are accurate structures (i.e. RMSD ≤ 2.0 Å or GDT_TS ≥ 80) that are correctly predicted to be accurate on the basis of their DP-score higher than the threshold (i.e. 0.7).
[b]False positives (FP) are inaccurate structures that are erroneously predicted to be accurate on the basis of their DP-score higher than the threshold.
[c]False negatives (FN) are accurate structures that are erroneously predicted to be inaccurate on the basis of their DP-score lower than the threshold.
[d]True negatives (TN) are inaccurate structures that are correctly predicted to be inaccurate on the basis of their DP-score lower than the threshold.
[e]The precision (i.e. the ratio of true positives among all positive predictions) becomes 1.00 at a DP-cut-off of 0.76.

However, it is possible that an incorrect structure with low DP score can also have low restraint violations; e.g. the NOE restraints may have been incorrectly assigned or otherwise incorrectly derived from the NOESY data.

Limitations for analyzing larger size proteins and homodimeric proteins

For larger size proteins (e.g. >200 amino acids), it is often necessary to use perdeuterated samples for structure determination. The RPF program can handle validation of protein structures using data from such perdeuterated protein samples, by excluding the deuterated atoms from the chemical shift assignment table. The computed RPF score provides useful measures of how good the data fits with the structure. However, the correlation between the RPF/DP scores and the structure accuracy is not as high as with fully protonated proteins, because data from perdeuterated proteins is much sparser. In particular, close H–H contacts, which may be critical to distinguish the correct from incorrect fold, are less extensive in the perdeuterated data set, making the DP score less sensitive to the structure accuracy. We suspect that an accurate structure will require a higher DP score cut-off using data from a perdeuterated protein compared to using

full-protonated protein data. Additional test data sets are needed to assess the best way to use the DP score for data obtained on perdeuterated proteins.

The RPF program can also analyze homodimeric proteins. This requires the user to first combine the two identical chains into a single chain with a different residue index. The RPF/DP score may be less sensitive for highly degenerate homodimeric proteins if many intermolecular NOEs, which may be critical to define the correct intermolecular packing, are degenerate with intramolecular NOEs. We also suspect that an accurate highly degenerate homodimeric protein structure will require a higher DP score cut-off than a protein structure with less degenerate resonance frequencies.

CONCLUSIONS

The RPF scores measure the fitness of NOESY peak list and resonance assignment data with NMR structure models. RPF scores, particularly the DP score, have a strong correlation with structure accuracy. Although other structure quality assessment tools [e.g. PROCHECK-all (11) and Molprobity (12)] do not correlate well with the structure accuracy based on the CASD-NMR2010 data, these knowledge-based assessments are none the less, very important tools for protein structure validation. Such knowledge-based methods compare observed conformational distributions and packing interactions with values observed in nature and/or expected on first principles. In general, an accurate NMR-derived protein structure should score well in all of these different and complementary views of structure quality (3).

High RPF scores and high PROCHECK and Molprobity scores indicate that a structure both fits the data well and has good stereochemical qualities. This is a goal of the structure determination process. High RPF scores and slightly lower PROCHECK and Molprobity scores indicate that a structure fits the data well, but that the data may not be sufficient to correctly define local conformations. In this case, additional data and/ or refinement may be required. However, good PROCHECK, Molprobity and other knowledge-based scores may be obtained for inaccurate structures which do not in fact fit well to the NMR data (8). Provided that the quality of input NOESY data is high, such structures would have poor RPF scores, and particularly DP score ≤0.6. The RPF server provides an effective and convention tool for evaluating and validating protein structures derived from NOESY data.

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Stout,G.H. and Jensen,L.H. (1968) *X-ray Structure Determination: A Practical Guide*. The Macmillan Company, London.
2. Doreleijers,J.F., Raves,M.L., Rullmann,T. and Kaptein,R. (1999) Completeness of NOEs in protein structure: a statistical analysis of NMR. *J. Biomol. NMR*, **14**, 123–132.
3. Huang,Y.J., Powers,R. and Montelione,G.T. (2005) Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.*, **127**, 1665–1674.
4. Bhattacharya,A., Tejero,R. and Montelione,G.T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins*, **66**, 778–795.
5. Raman,S., Huang,Y.J., Mao,B., Rossi,P., Aramini,J.M., Liu,G., Montelione,G.T. and Baker,D. (2010) Accurate automated protein NMR structure determination using unassigned NOESY data. *J. Am. Chem. Soc.*, **132**, 202–207.
6. Bagaria,A., Jaravine,V., Huang,Y.J., Montelione,G.T. and Guntert,P. (2012) Protein structure validation by generalized linear model root-mean-square deviation prediction. *Protein Sci.*, **21**, 229–238.
7. Rosato,A., Bagaria,A., Baker,D., Bardiaux,B., Cavalli,A., Doreleijers,J.F., Giachetti,A., Guerry,P., Guntert,P., Herrmann,T. *et al.* (2009) CASD-NMR: critical assessment of automated structure determination by NMR. *Nat. Methods*, **6**, 625–626.
8. Rosato,A., Aramini,J., Arrowsmith,C., Bagaria,A., Baker,D., Cavalli,A., Doreleijers,J.F., Eletsky,A., Giachetti,A., Guerry,P. *et al.* (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. *Structure*, **20**, 227–236.
9. Eisenberg,D., Luthy,R. and Bowie,J.U. (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.*, **277**, 396–404.
10. Sippl,M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins*, **17**, 355–362.
11. Laskowski,R.A., Rullmann,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
12. Davis,I.W., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.*, **32**, W615–W619.
13. Preparata,F.P. and Shamos,M.I. (1988) *Computational Geometry: An Introduction*, 2nd edn. Springer, New York.
14. Bentley,J.L. (1975) Multidimensional binary search trees used for associative searching. *Commun. ACM*, **18**, 509–517.
15. Jmol: an open-source Java viewer for chemical structures in 3D. Available online at http://www.jmol.org.
16. Zemla,A. (2003) LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Res.*, **31**, 3370–3374.
17. Clarke,N.D., Ezkurdia,I., Kopp,J., Read,R.J., Schwede,T. and Tress,M. (2007) Domain definition and target classification for CASP7. *Proteins*, **69(Suppl. 8)**, 10–18.