ARTICLE

Resolution-by-proxy: a simple measure for assessing and comparing the overall quality of NMR protein structures

Mark Berjanskii · Jianjun Zhou · Yongjie Liang · Guohui Lin · David S. Wishart

Abstract In protein X-ray crystallography, resolution is often used as a good indicator of structural quality. Diffraction resolution of protein crystals correlates well with the number of X-ray observables that are used in structure generation and, therefore, with protein coordinate errors. In protein NMR, there is no parameter identical to X-ray resolution. Instead, resolution is often used as a synonym of NMR model quality. Resolution of NMR structures is often deduced from ensemble precision, torsion angle normality and number of distance restraints per residue. The lack of common techniques to assess the resolution of X-ray and NMR structures complicates the comparison of structures solved by these two methods. This problem is sometimes approached by calculating "equivalent resolution" from structure quality metrics. However, existing protocols do not offer a comprehensive assessment of protein structure as they calculate equivalent resolution from a relatively small number (<5) of protein parameters. Here, we report a development of a protocol that calculates equivalent resolution from 25 measurable protein features. This new method offers better performance (correlation coefficient of 0.92, mean absolute error of 0.28 Å) than existing predictors of equivalent resolution. Because the method uses coordinate data as a proxy for X-ray diffraction data, we call this measure "Resolution-by-Proxy" or ResProx. We demonstrate that ResProx can be used to identify under-restrained, poorly refined or inaccurate NMR structures, and can discover structural defects that the other equivalent resolution methods cannot detect. The ResProx web server is available at http://www.resprox.ca.

M. Berjanskii · J. Zhou · Y. Liang · G. Lin · D. S. Wishart (✉)
Department of Computing Science, University of Alberta, Edmonton, AB, Canada
e-mail: david.wishart@ualberta.ca

D. S. Wishart
Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

D. S. Wishart
National Research Council, National Institute for Nanotechnology (NINT), Edmonton, AB T6G 2E8, Canada

Introduction

The assessment of protein structure quality continues to be an important area for the "consumers", "brokers" and "producers" of protein structures. Structural biologists (the "producers") want improved measures for assessing structure quality to help with the structure determination process. Database curators (the "brokers") want improved tools to validate new entries and to identify potentially fraudulent or erroneous structures. Bioinformaticians, protein chemists, and drug designers (the "consumers") want improved methods to rapidly select high quality structures for the modeling of novel protein structures or novel ligand complexes. One of the challenges in structure assessment and validation is the inconsistency in metrics and protocols that are used to evaluate structures determined by X-ray crystallography and NMR. In X-ray crystallography, most investigators assess structure quality in terms of resolution, R_{free} factor, reflection completeness, B-factors, and stereo-

chemical normality parameters (Wlodawer et al. 2008). X-ray resolution can serve as a reasonable first approximation of protein structure quality when a very precise or exhaustive protein assessment is not required. Resolution is formally designed as the minimal distance between structural features that can still provide measurable X-ray diffraction and, as a result, can be distinguished from each other on electron density maps (Wlodawer et al. 2008). A protein with a resolution above 2.7 Å is considered to be low-resolution structure, while proteins with a resolution between 2.7 and 1.8 Å are classified as medium resolution structures, and those below 1.8 Å resolution are typically classified as high-resolution structures (Minor 2007; Wlodawer et al. 2008).

In protein NMR, structural quality is typically reported in terms of numbers of restraint violations, numbers of experimental restraints, ensemble precision and stereochemical normality parameters (Spronk et al. 2004). NMR parameters similar to the X-ray R_{free} factor do exist (Cornilescu et al. 1998; Clore and Garrett 1999; Gronwald et al. 2000; Huang et al. 2005) but they either have not been widely adopted or are not always reported in NMR publications. The number of NMR experimental restraints is not directly comparable to the number of X-ray reflections due to their different meanings and their different utilization in structure generation. NMR structures do not have a parameter identical to X-ray resolution. However, the term "resolution" is still widely used in protein NMR as a synonym of structural quality. Resolution range (high-, medium-, or low-resolution) is commonly assigned to NMR ensembles based on parameters, such as ensemble precision, the quality of the Ramachandran plot and the number of NOEs per residues (Kwan et al. 2011).

The differences in resolution definitions, criteria for evaluating quality of experimental data and goodness of structure-to-experiment fit make it difficult to compare protein structures determined by X-ray crystallography and NMR spectroscopy. The current situation makes it particularly challenging for structure "consumers" to make reasonable decisions how to properly select NMR structures for analysis or comparison in their research projects.

It is possible to do a partial evaluation of the quality of NMR and X-ray structures in a consistent manner through other methods that depend upon atomic coordinates rather than experiment-specific data (such as diffraction data or NOE data). Commonly assessed parameters include the normality of torsion angles, the presence of atom clashes, the normality of hydrogen bonding, violations of bond lengths and bond angles (Laskowski et al. 1993, 1996; Hooft et al. 1996; Willard et al. 2003; Davis et al. 2007; Berjanskii et al. 2010). Other techniques that assess cavities (Sheffler and Baker 2009), residue-specific packing

volumes (Richards 1977), packing efficiency (Seeliger and de Groot 2007), threading energies (Eisenberg et al. 1997; Wiederstein and Sippl 2007), atomic volumes (Pontius et al. 1996), as well as asparagine and glutamine flips (Vriend 1990; Word et al. 1999) have also been shown to be quite effective in assessing protein structure quality at a global level. However, the problem with these methods is that each approach has different criteria or non-obvious thresholds for calculating an "overall grade" for a structure or distinguishing good structures from bad.

To address this issue of inconsistency, methods that use a more universally understood criterion (i.e. X-ray resolution) have been developed over the past few decades. In particular, protocols to calculate "equivalent resolution" from coordinate data have been incorporated into several structure validation programs, such as Procheck-NMR (Laskowski et al. 1996), MolProbity (Chen et al. 2010), and RosettaHoles2 (Sheffler and Baker 2010). Unfortunately, all of the existing methods rely on a very small number of protein structure quality measures to predict resolution (4, 3, and 1 measures, respectively) and, therefore, cannot provide a comprehensive assessment of protein structure quality. Here, we show that it is possible to develop a computer algorithm that accurately predicts equivalent resolution from a much more complete set of 25 protein features. We also show that this new method is significantly more accurate than other equivalent resolution techniques and that it reproduces the expected behaviour of a true resolution function (i.e. sensitivity to the quantity of experimental data, sensitivity to coordinate errors, correlation with orthogonal measures of structure quality, etc.) better than existing techniques. Because the method uses coordinate data as a proxy for X-ray diffraction data, we have decided to call this measure "Resolution-by-Proxy" or ResProx. A more detailed description of the method follows.

Materials and methods

A brief summary of the ResProx protocol

ResProx consists of three main elements: (1) a machine learning predictor (SVR predictor); (2) a Z-score based predictor (Z-Mean predictor); and (3) a decision making module (DecisionMaker). This algorithmic structure is illustrated in Fig. S1. Both predictors are run sequentially for a query PDB file and the decision making algorithm uses the two estimated resolution values and selected protein quality scores to decide which resolution value should be returned to the user. We describe the training, implementation and optimization of these three components below.

Training and testing sets

Key to the development and testing of ResProx was the construction of a suitable training and testing set of X-ray structures with a wide span of known X-ray resolutions. This collection of structures was divided into 0.25 Å resolution bins and the initial training set was built so that, at least, 100 structures could be placed in each bin, spanning a range between 1.0 and 3.75 Å (Fig. S2). No minimum requirements were applied to the resolution bins outside this range. A total of 2,927 structures that were largely free of severe structural defects, as assessed by several measures of structure quality (Berjanskii et al. 2010), were selected from the PDB. These structures spanned an X-ray resolution range between 0.75 and 4.75 Å. A testing set of 500 proteins was randomly selected from this initial protein pool of 2,927 proteins, leaving 2,427 proteins for the training set. To have the required number of structures in each resolution bin, redundant proteins were allowed to be present in both the training and testing datasets. However, care was taken to ensure that no particular protein system dominated the training and testing sets. More specifically, proteins that had 2, 3, 4, and 5 or more similar structures with sequence identities above 98 % did not exceed 35, 2.5, 1 and 2 % of protein set, respectively. The PDB IDs of the proteins that were included in the training and testing sets are available from the ResProx website (http://www.resprox.ca).

SVR predictor

The development of the ResProx machine learning predictor started from a feature selection step to identify a set of protein structure quality parameters that could be used for resolution prediction. Using a small number of highly-relevant features in machine learning usually produces a more robust predictive model than using large numbers of low quality or low relevance features (Mitchell 1997). On the other hand, a protein feature set for evaluating protein structure quality should be relatively comprehensive in order to effectively detect a variety of structural abnormalities that can occur during X-ray or NMR model generation. The feature selection used for ResProx was primarily based on absolute Pearson correlation coefficients between the observed X-ray resolution and the structure quality scores from a variety of in-house and externally developed programs, such as VADAR (Willard et al. 2003), MolProbity (Davis et al. 2007), GeNMR (Berjanskii et al. 2009), PROSESS (Berjanskii et al. 2010), and RosettaHoles (Sheffler and Baker 2009, 2010). Scores with higher absolute correlation coefficients were generally given preference. In addition, care was taken to ensure that major categories of protein quality parameters (e.g.

covalent bond quality, protein packing quality, torsion angle quality, etc.) had at least one representative in the final feature set, even if no feature in these categories correlated particularly well with resolution. While the presence of these low-correlating features could possibly decrease the performance of the machine learning model, their scores were included to make the model more sensitive to rarely occurring defects in protein structures. In total, 25 features were selected (Table S1). Figure S3 illustrates the relationship between several of these high- and low-correlating protein structure quality parameters and X-ray resolution.

This set of 25 features was first used to construct a support vector regression (SVR) model for prediction of the corresponding X-ray resolution, using version 3.6.2 of WEKA (Hall et al. 2009). An improved version of the sequential minimal optimization (SMO) algorithm (She-vade et al. 2000) was chosen to solve the regression problem via this support vector machine method. The SMO algorithm was run with the default options for the epsilon insensitive loss function, stopping criteria, round-off error and algorithm variant. A radial basis function (RBF) was selected as the SVR kernel. The optimal value of 0.35 for the gamma parameter that describes the width of the RBF kernel was found through a systematic grid search. The model was trained on the aforementioned training set of 2,427 proteins. During the SVR model optimization, it was found that developing an accurate SVR predictor require additional post-processing that included applying minimal and maximal limits to the score values, and converting some scores to their logarithmic form as described in Table S1. When tested on the 500-protein testing set, the SVR model demonstrated the same outstanding agreement with experimentally measured resolution as it did with the 2,427 protein training set (see "Results and discussion").

Z-Mean predictor

Initial tests of the SVR model on NMR and X-ray structures revealed that the algorithm excelled at predicting the experimentally observed resolution for published X-ray structures, but sometimes produced unreasonable resolution estimates for deliberately misfolded or under-determined NMR models. For example, the SVR model could assign an equivalent resolution value of 2.3 Å for a structure with as many as 10 NOE violations. This result was not totally unexpected because certain structural defects observed in poor NMR models, such as packing defects, unrealistic bond lengths, and distorted bond angles were not commonly seen in the SVR training set of X-ray structures. It is conceivable that the SVR model failed for some NMR structures because the machine learning method did not have proper training data to learn how to

relate NMR-specific defects to X-ray resolution. To calculate equivalent resolution for proteins with a larger array of structural defects, we developed a second algorithm, called Z-Mean. Rather than using machine learning methods, the Z-Mean approach uses a simpler regression model. In particular, it empirically selects features that correlate well with the measured X-ray resolution of well-determined structures and extrapolates this fit to the problem proteins. Z-mean was optimized and tested on the same X-ray structure training and testing sets that were used for the development of the SVR model. In addition, Z-Mean, as well as the SVR model and the final ResProx protocol were tested on 920 NMR models that were deliberately under-constrained or contained high numbers of NOE constraint violations.

To develop Z-Mean, a subset of 15 structure quality parameters (Table S1) was empirically selected from the SVR model scores to reduce feature redundancy and remove scores with exceptionally poor correlations to experimental resolution. For example, only two parameters were selected among 5 different metrics of protein hydrogen bonding. A similar approach was applied to various normality measures of side-chain orientations and backbone torsion angles. This extra feature selection step was necessary to achieve comparable contributions from each category of protein quality into resolution predictions while preserving the accuracy of the predictions. A subset of 235 high-resolution X-ray structures (resolution below or equal to 1.0 Å) was selected from the original ResProx protein set to find the means and standard deviations of the selected quality parameters for calculation of their Z-scores.

The structure quality parameters were divided into two groups: structural measures and normality parameters, which are derived by comparing the structural measures of a query protein with their counterparts in high-quality structures. An example of a structural measure is the hydrogen bond energy. Both excessive hydrogen bonding due to incorrect $HN_i–CO_{i-2}$ hydrogen bonding (Xia et al. 2002) and insufficient hydrogen bonding can be a good indicator of poor model quality. Therefore, both positive and negative Z-scores of hydrogen bond energy were taken into account in the calculation of the absolute Z-score. For normality parameters, such as the percentage of residues in the disallowed regions of the Ramachandran plot, only the sign (either positive or negative) indicates an increase or decrease in protein quality. The signs of the Z-scores that were used in predicting protein resolution are shown in Table S1. Z-scores of the opposite sign are evidence of exceptional structural quality that may arise from either heroic research efforts in structure determination or by specifically refining the model using similar normality parameters such as the Rama energy term in the XPLOR-NIH program (Schwieters et al. 2003). Since the origin of

such "too-good-to-be-true" Z-scores cannot be determined from protein coordinates, we decided not to penalize protein models for having them and used zero as a default Z-score value.

Individual Z-scores were combined into a single measure of overall quality by calculating a simple average. No weighting coefficients were assigned to quality parameters in the calculations of the average Z-score in order to make this ResProx step equally sensitive to all selected types of protein quality characteristics. The observed X-ray resolution for the training data set was plotted against the average Z-score and the linear part of the plot (average Z-score from 0 to 1.2) was fit by a first-order polynomial equation (Fig. S4) using QtiPlot 0.9.8.4 (Vasilief 2011). The following formula for converting mean Z-scores to resolution values (Å) was obtained via curve fitting:

$$Z\text{-Mean resolution} = 1.52 \times \bar{Z} + 0.76 \qquad (1)$$

where $\bar{Z}$ is the average Z score.

DecisionMaker

Since two equivalent resolution values are available from these two ResProx predictors, a decision needs to be made regarding which prediction should be returned to the user. ResProx's decision making module uses the predicted resolution values and GeNMR (Berjanskii et al. 2009) scores to determine if a model has structural problems that are rare in published structures and, therefore, may not be suitable for the SVR model. More specifically, a protein model is considered to be poor if its total GeNMR knowledge-based score, excluding the radius of gyration score, is above 10 (Fig. S5). Also, if the Z-Mean algorithm reports a predicted resolution that is significantly worse than the one predicted by the SVR model (i.e. the Z-Mean equivalent resolution is larger than the SVR model resolution by more than an empirical threshold of 1.5 Å), a preference is given to the Z-Mean value, which is designed for handling low-quality structures. A special rule is also applied to unfolded structures or small extended peptides, whose structures cannot be confidently determined by NMR or X-ray methods and are beyond the range of applicability of both ResProx predictors. Such models are identified by extremely weak or non-existent hydrogen bonding. In these cases, if the GeNMR hydrogen bond energy score is above 1.5, the protein model is believed to be unstructured and is assigned an upper resolution limit of 10 Å. Analysis of the ResProx training set (see below) suggests that these empirical rules create an effective switch between applications of the SVR model for good quality models and Z-Mean or an upper resolution limit for poorly determined or incorrectly determined structures.

Results and discussion

Performance of ResProx on X-ray structures

As noted earlier, ResProx consists of two resolution predictors, an SVR predictor and Z-Mean predictor. The first is designed to handle properly folded or properly determined protein structures, the second is developed to handle unusual, partially refined or misfolded structures. The performance of the SVR predictor was evaluated by calculating absolute errors of prediction and correlation coefficients achieved by the machine learning model for both the independent test set of 500 proteins and a fivefold cross validation assessment of the 2,427 training proteins. If a machine learning algorithm has not been over-trained, the performance for the fivefold cross validation should match closely (~1–2 %) with the performance on the independent test set. As seen on Fig. 1, this is indeed the case for the SVR predictor. The absolute error of prediction and the correlation coefficient between ResProx's estimated resolution and the experimentally observed resolution are 0.28 and 0.92 Å, respectively, for both training and testing sets. This result indicates that the SVR approach is robust and that the regression model has not been over-trained.

The correlation coefficients between the Z-Mean prediction and the observed X-ray resolution are 0.85 and 0.86 for the training and testing sets, respectively (Fig. 1). Detailed analysis of ResProx's training/testing protein set indicates that the combined ResProx method uses the SVR model in 99.3 % cases when evaluating published or PDB-deposited X-ray structures. ResProx resorts to the Z-Mean predictor or assignment of the maximal resolution for just 0.2 and 0.5 % of published X-ray structures, respectively. ResProx uses one of these two alternative methods almost exclusively when poor quality structures are encountered.

Overall, these results show that ResProx's estimate of equivalent resolution strongly correlates with the experimentally observed X-ray resolution. Since diffraction resolution is closely tied to the number of reflections in a diffraction pattern, this also implies that ResProx is sensitive to the amount of X-ray experimental data that is used in structure generation.

Comparative evaluation of ResProx

Resprox's performance was compared with other existing equivalent resolution methods, namely, Procheck-NMR's equivalent resolution (Laskowski et al. 1996), MolProbity's score (Chen et al. 2010), and the RosettaHoles2 S_{RESL} score (Sheffler and Baker 2010). Procheck-NMR calculates four equivalent resolution values from the following quality parameters: Ramachandran plot quality, main-chain
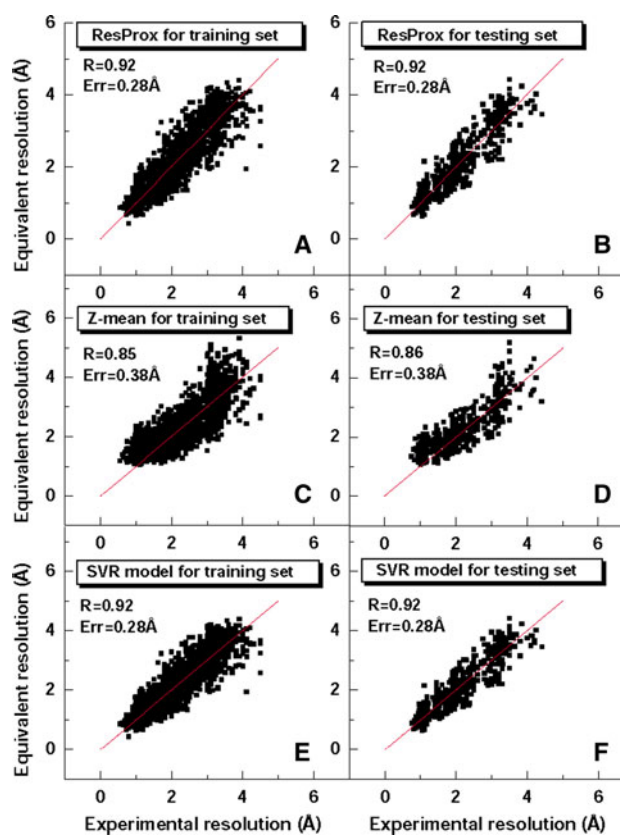

Fig. 1 Correlation between ResProx equivalent resolution and X-ray experimental resolution for the ResProx training and testing sets. **a** Final ResProx values for the ResProx training set. **b** Final ResProx values for the ResProx testing set. **c** Z-Mean equivalent resolution for the ResProx training set. **d** Z-Mean equivalent resolution for the ResProx testing set. **e** SVR predictions for the ResProx training set. **f** SVR predictions for the ResProx testing set. R and Err parameters indicate the Pearson correlation coefficient and the absolute mean error of resolution prediction, respectively. The *red lines* correspond to a perfect correlation ($y = x$)

hydrogen bond energies, χ_1 pooled standard deviation, and the standard deviation of χ_2 trans angle (Laskowski et al. 1996). By averaging these individual equivalent resolution values, one can determine an overall or mean equivalent resolution for a given protein structure. The MolProbity score is a log-weighted combination of MolProbity's clashscore, the percentage of residues in the "not-favored" region of Ramachandran space and the percentage of bad side-chain rotamers (Chen et al. 2010). The MolProbity score was originally optimized to predict crystallographic resolution from a database of high-resolution X-ray structures. RosettaHoles2 uses information about protein packing to predict X-ray resolution via a SVR model (Sheffler and Baker 2010). For the ResProx testing set, the correlation coefficients between the observed resolution and the predicted "equivalent resolution" values of Procheck-NMR, MolProbity, RosettaHoles2 S_{RESL} score and ResProx were 0.78, 0.86, 0.87, and 0.92, respectively (Figs. 1,
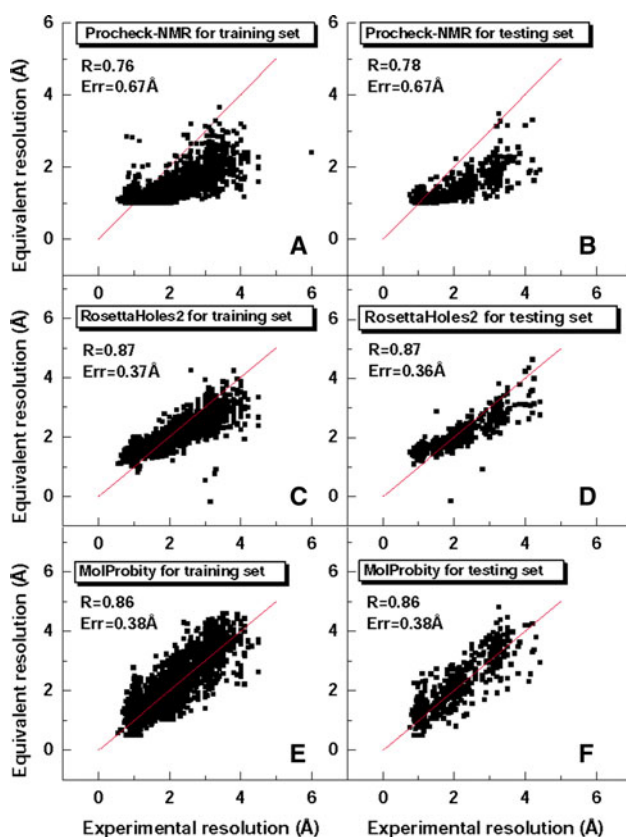
Fig. 2 Correlation between equivalent resolution and X-ray experimental resolution as calculated by Procheck-NMR, MolProbity, and RosettaHoles2. **a** Procheck-NMR equivalent resolution for the ResProx training set. **b** Procheck-NMR equivalent resolution for the ResProx testing set. **c** RosettaHoles2 S_{RESL} equivalent resolution for the ResProx training set. **d** RosettaHoles2 S_{RESL} for the ResProx testing set. **e** MolProbity score for the ResProx training set. **f** MolProbity score for the ResProx testing set. R and Err parameters indicate the Pearson correlation coefficient and the absolute mean error of resolution prediction, respectively. The *red lines* correspond to a perfect correlation ($y = x$)

2). ResProx outperforms MolProbity and RosettaHoles2 by about 7 % and Procheck-NMR by 15 %. ResProx also has a significantly smaller mean absolute error (0.28 Å) than Procheck-NMR (0.67 Å), MolProbity (0.38 Å), and RosettaHoles2 (0.36 Å).

Resolution predictions by the Z-Mean method are less accurate than those obtained by the SVR model or the combined ResProx method (Fig. 1). This result is expected because the high sensitivity of the Z-Mean method to rare protein defects requires a uniform contribution of all scores, including those that do not correlate well with X-ray resolution (see Methods). However, even with these limitations, Z-Mean's error is still half of that seen for Procheck-NMR and is similar to the prediction errors seen for MolProbity and RosettaHoles2 (Figs. 1, 2).

Analysis of the error patterns revealed that Procheck-NMR systematically overestimates resolution (i.e. the

predicted resolution values are smaller than experimental ones) by 0.61 Å on average (Fig. 2a, b). On the other hand, RosettaHoles2 S_{RESL} score predicts a higher equivalent resolution value than the experimental one by 0.28 Å (on average) for the resolution range from 0.5 to 2 Å. For X-ray structures with a resolution worse than 2 Å, RosettaHoles2 overestimates the resolution by 0.35 Å, on average (Fig. 2c, d). Systematic differences between equivalent resolution and experimental resolution for the full resolution range are 0.09, 0.05, 0.16, and 0.05 Å for MolProbity, SVR model, Z-Mean, and ResProx, respectively (Figs. 1, 2e, f).

Sensitivity to completeness of experimental NMR restraints

Different values of crystallographic resolution relate to different number of observables (reflections) in X-ray diffraction patterns. The higher the resolution, the larger the number of reflections that one can expect to measure and use in protein structure generation (Wlodawer et al. 2008). Among many parameters that are reported for NMR models, the number of NMR experimental restraints (NOE-, disulfide-, and hydrogen bond distance restraints, torsion angle restraints, J-couplings, etc.) is the closest experimental match to the number of reflections. If ResProx is to have the equivalent utility for NMR structures as it does with X-ray structures, it is important to demonstrate that ResProx is also sensitive to the number of NMR restraints. To conduct this test, several sets of distance restraints with different percentages of randomly removed entries (0, 20, 40, 60, 70, 75, 80, 85, 90, 93, and 95 %) were created for the NMR model of ubiquitin (PDB ID: 1D3Z). For each level of distance restraint completeness, 5 sets of randomly reduced restraints were created. 200 ubiquitin models, per restraint set, were calculated using a standard simulated annealing protocol included with the XPLOR-NIH distribution (Schwieters et al. 2003). Ensembles of 20 models were selected based on those exhibiting the lowest overall XPLOR energy. Average ResProx values were calculated for each ensemble and plotted against the percentage of missing distance restraints (Fig. 3a).

As can be seen from Fig. 3a, the ResProx values for ubiquitin demonstrate an exponential-like relationship with the amount of experimental information (i.e. the number of distance restraints). In other words, ResProx's estimated resolution appears to be sensitive to the amount of experimental information, as one would expect from an equivalent of diffraction resolution that is related to the number of X-ray reflections. This observation is consistent with the observed relationship between NOE completeness and other measures of equivalent resolution that are calculated by Procheck-NMR (Fig. 3b), MolProbity (Fig. 3c), and RosettaHoles2 (Fig. 3d). However, it is important to note

Fig. 3 Correlation between completeness of experimental information (distance restraints) and equivalent resolution of ubiquitin. **a** ResProx score. **b** Procheck-NMR equivalent resolution. **c** MolProbity score. **d** RosettaHoles2 S_{RESL}. Different completeness of experimental data was achieved by random removal of different portions of distance restraints from the total restraint set (i.e. 0, 20, 40, 60, 70, 75, 80, 85, 90, 93, and 95 % of removed restraints; five removal attempts per completeness level). Distance restraints consisted of NOE-based distance restraints and hydrogen bond distance restraints of the ubiquitin NMR ensemble 1D3Z
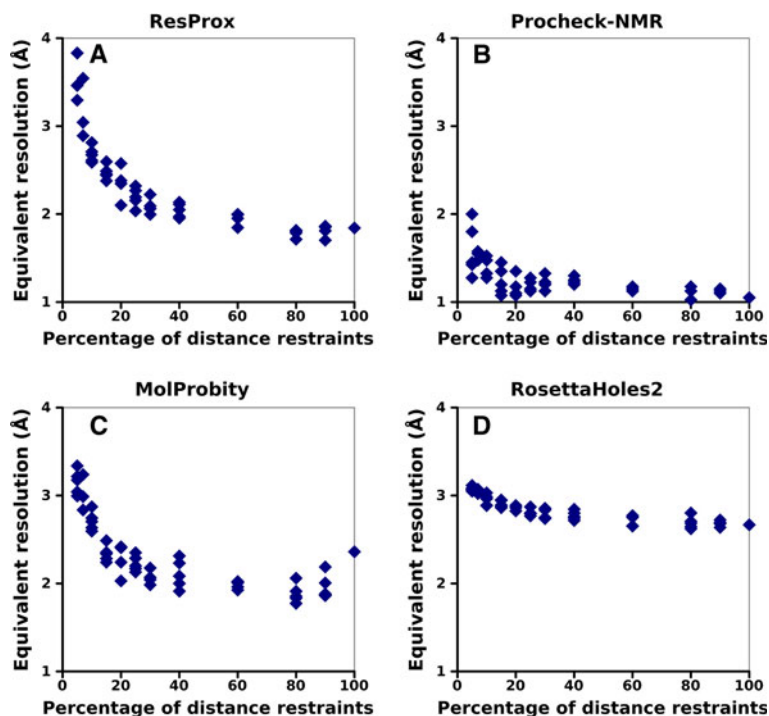
that, unlike ResProx, both Procheck-NMR and MolProbity can generate very similar scores for both poorly restrained ubiquitin models (20 % of distance restraints) and well-defined ones (100 % of distance restraints). In other words, these two methods demonstrate limited sensitivity to the completeness of the experimental restraints. In addition, Procheck-NMR's equivalent resolution overestimates model quality by assigning an equivalent resolution below 2 Å to all models, even those that are calculated with just 5 % of the distance restraints. Indeed, Procheck-NMR gives the majority of models a resolution below 1.5 Å. It is also notable that the RosettaHoles2 S_{RESL} score changes within a small very range as a function of restraint completeness and seems to underestimate quality of good models (assigning a resolution of 2.6 Å for models from the 100 % restraint set) and overestimate quality of bad models (predicting a resolution of 3.1 Å for models from the 5 % restraint set).

In a practical sense, our results suggest that high Res-Prox values (i.e. >2.5 Å) could help a researcher detect experimentally underdetermined NMR structures. As will be discussed later, ResProx is sensitive not only to a lack of experimental information, but also to other problems in NMR structure determination process.

Sensitivity to orthogonal measures of NMR structure quality

The precision of NMR ensembles is not explicitly optimized by structure generation programs and, therefore, is

considered to be a valuable parameter to evaluate the quality of an NMR structures. Other non-optimized, orthogonal measures of protein structure quality include the correlation between observed and predicted backbone chemical shifts (Neal et al. 2003; Wishart 2011; Han et al. 2011), the agreement between a protein model and its NOE spectra, as reported by NOE R-factors (Gronwald et al. 2000), RPF scores (Huang et al. 2005) and RDC Q or R-factors (Cornilescu et al. 1998; Clore and Garrett 1999). Using the same set of ubiquitin structures described earlier (obtained with randomly excluded distance restraints), we calculated the ensemble precision and absolute mean difference between experimentally observed backbone proton chemical shifts and those shifts predicted from protein models by SHIFTX2 (Han et al. 2011). These parameters were then plotted against the corresponding ResProx values (Figs. 4a, 5a, respectively). As seen in these figures, Res-Prox demonstrates an almost linear dependence on ensemble precision and model-to-chemical shift agreement, with a rank-order Spearman correlation coefficient of 0.95 for both parameters. A similar trend is observed for equivalent resolution values determined by Procheck-NMR (Figs. 4b, 5b), MolProbity (Figs. 4c, 5c) and Rosetta-Holes2 (Figs. 4d, 5d), albeit with lower correlation coefficients for MolProbity and Procheck-NMR. As mentioned earlier, Procheck-NMR overestimates quality of this model set, assigning an equivalent resolution of 2 Å to ensembles with a precision as bad as 4 Å. RosettaHoles2 does not appear to be very sensitive to ensemble precision, as it predicts an equivalent resolution of 2.6 Å to highly precise

Fig. 4 Correlation between
equivalent resolution and the
ensemble precision of ubiquitin.
a ResProx score. **b** Procheck-
NMR equivalent resolution.
c MolProbity score.
d RosettaHoles2 S_{RESL}
Ensemble precision was
assessed by calculating the
backbone RMSD of ubiquitin
NMR ensembles with MolMol
(Koradi et al. 1996). The
Spearman rank-order
correlation coefficient is 0.95,
0.69, 0.84, and 0.90 for
ResProx, Procheck-NMR,
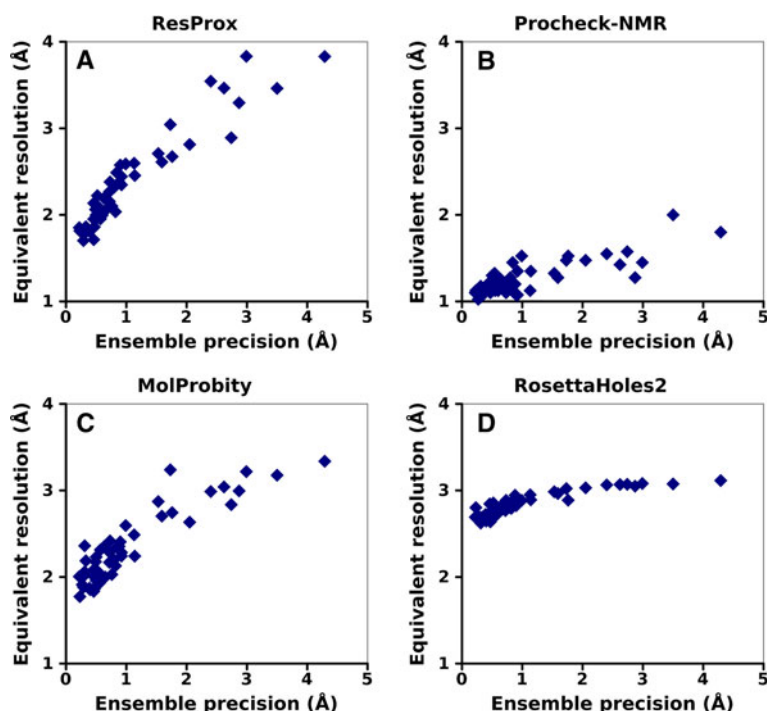MolProbity, and RosettaHoles2,
respectively

Fig. 5 Correlation of
equivalent resolution with
backbone proton chemical
shifts. **a** ResProx score.
b Procheck-NMR equivalent
resolution. **c** MolProbity score.
d RosettaHoles2 S_{RESL} The
agreement between ubiquitin
models and backbone proton
chemical shifts was assessed by
predicting the chemical shifts
from different NMR models
with ShiftX2 (Han et al. 2011)
and calculating the mean
absolute difference between
predicted and experimentally
measured chemical shifts. The
Spearman rank-order
correlation coefficient is 0.95,
0.73, 0.85, and 0.95 for
ResProx, Procheck-NMR,
MolProbity, and RosettaHoles2,
respectively
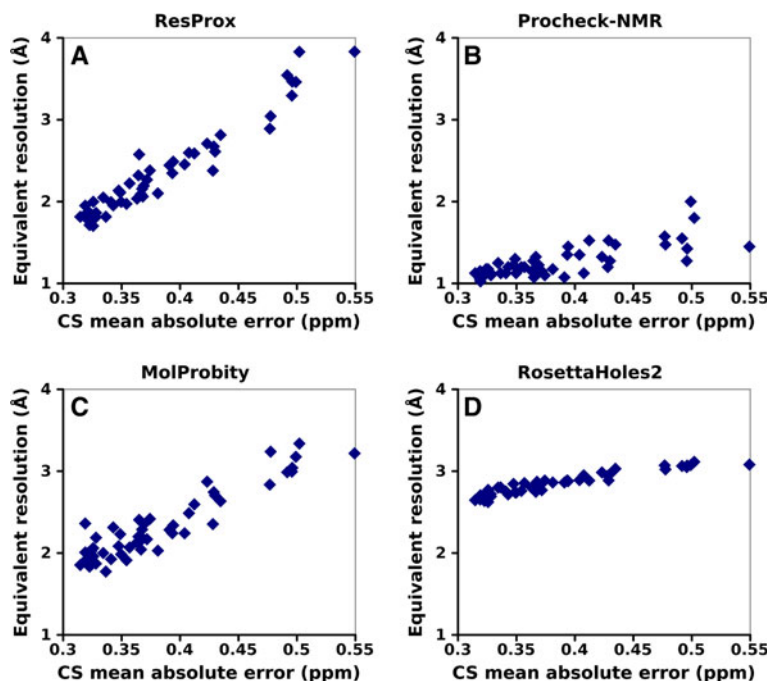
models (secondary structure backbone RMSD of 0.22 Å)
and 3.1 Å to imprecise ensembles (RMSD of 4.29 Å).
These results suggest that ResProx correlates quite well to
a number of orthogonal structure quality measures, such as
ensemble precision and model-to-chemical shift agreement. Furthermore, ResProx performs better than most
existing "equivalent resolution" methods in detecting
under-restrained NMR structures.

**Sensitivity to NMR restraint violations and model
accuracy**

Since the magnitude of the coordinate errors in X-ray
models correlates with crystallographic resolution, it is
reasonable to expect that ResProx should also demonstrate
similar sensitivity to coordinate errors in NMR structures.
Therefore, we investigated if ResProx was capable of

Fig. 6 Correlation between equivalent resolution of ubiquitin and the number of distance violations. **a** ResProx score. **b** Procheck-NMR equivalent resolution. **c** MolProbity score. **d** RosettaHoles2 S_{RESL}
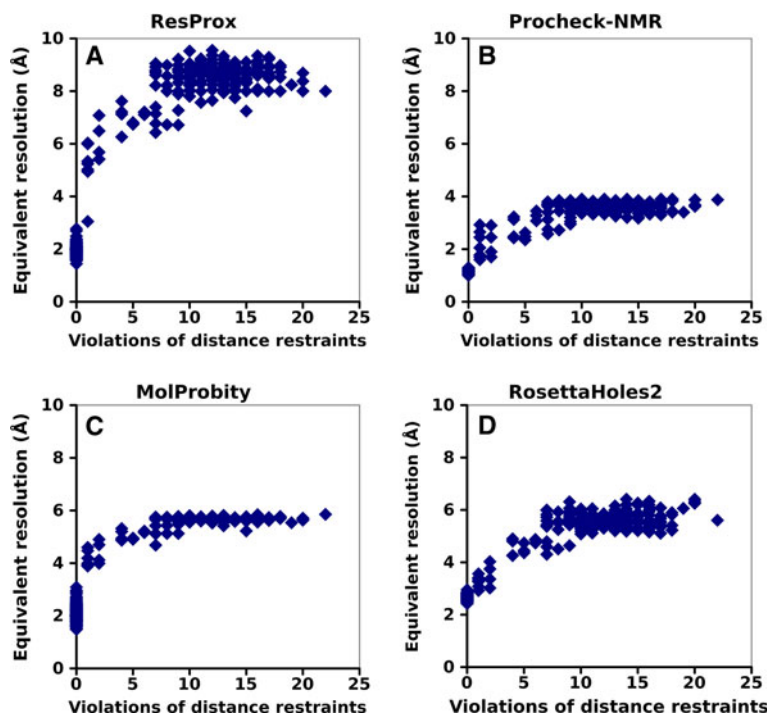

Fig. 7 Correlation between the equivalent resolution of ubiquitin and model accuracy. **a** ResProx resolution **b** Procheck-NMR equivalent resolution. **c** MolProbity score. **d** RosettaHoles2 S_{RESL} Model accuracy was measured by calculating the backbone RMSD of ubiquitin models with respect to the ubiquitin X-ray structure 1UBQ. NMR models of ubiquitin with different distance restraint violations were analyzed (see text for details)
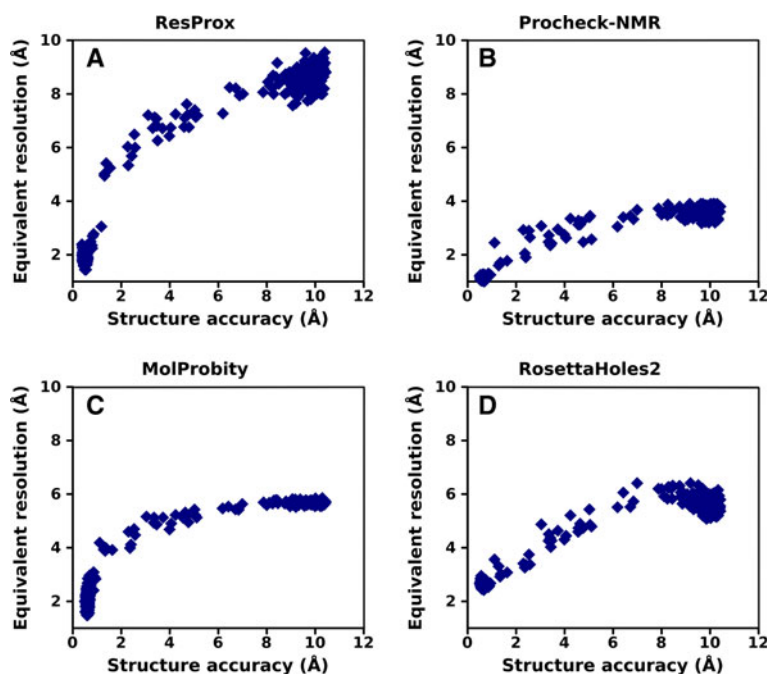
detecting the degree of structural inaccuracy that corresponds to typical violations seen for a high-quality NMR restraint set. To do so, a model of ubiquitin was generated from published NOE and hydrogen bond distance restraints of a high-resolution NMR model of ubiquitin (PDB ID: 1D3Z). A set of NMR models of ubiquitin with different numbers of distance restraint violations, ranging from 0 to

20, was prepared for this test by varying the length of simulated annealing step and, as a result, the convergence of the XPLOR-NIH structure determination protocol. ResProx values were calculated for individual structures and plotted against the number of NOE distance restraint violations (Fig. 6a) and model accuracy, as measured by backbone RMSD with respect to the X-ray ubiquitin

structure 1UBQ (Fig. 7a). As can be seen from Fig. 6a, ResProx is sensitive to the agreement with experimental data, increasing from a mean predicted resolution of 2–5 Å after only 1 NOE violation or, in terms of model accuracy, an RMSD of 1.2 Å. ResProx values reach a plateau of ∼8.5 Å after 7 NOE violations. However, when model accuracy is analyzed (Fig. 7a), it is clear that models with the same number of NOE violations can have different coordinate errors and that ResProx displays a sensitivity to model accuracy within a range from 1 to 8 Å. Similar dependencies are also observed for equivalent resolution predicted by Procheck-NMR (Figs. 6b, 7b), MolProbity (Figs. 6c, 7c), and RosettaHoles2 (Figs. 6d, 7d), albeit with a smaller range of resolution estimates. Within the group of models with no NOE violations, the ResProx resolution prediction still varies substantially (from 1.4 to 2.8 Å), confirming a well-known notion that the absence of restraint violations by a model does not always imply high structure quality (Nabuurs et al. 2006) and that other parameters (e.g. XPLOR energy terms) should be included to select the final NMR ensemble. While there is no strong correlation between model accuracy and ResProx values within this group of near-native models (data not shown), it is important to note that the no-violations model with the worst ResProx resolution (2.8 Å) also had the worst backbone accuracy (0.86 Å).

ResProx's ability to distinguish restraint-compliant or accurate NMR models of ubiquitin from those with just a few restraint violations or those with coordinate errors as little as 1.3 Å (as measured by backbone RMSD) suggests that ResProx could be used to detect inaccurate NMR structures and that it could also be used as an additional quality metric for selecting final NMR ensembles. These results also show that ResProx exhibits a correlation with structure accuracy in a manner consistent with the correlation seen between crystallographic resolution and coordinate error.

To further explore ResProx's ability to detect coordinate inaccuracies in NMR models, we compared ResProx values for several pairs of obsolete and current PDB entries that had significant structural differences. Typically if a PDB entry is updated with a significantly altered structure, this is a strong indication of significant problems with the original (obsolete) coordinates. For all tested cases, ResProx showed a significant improvement in the predicted resolution (i.e. the NMR model quality) after the PDB entries were updated (Table 1). This result is in agreement with improvements in equivalent resolution predicted by other methods, but as seen in Table 1, the average difference in equivalent resolution between the obsolete model and the corrected model is greater and more obvious for ResProx (1.99 Å) than for Procheck-NMR (0.61 Å), MolProbity (0.95 Å) or RosettaHoles2 (0.52 Å). It is

reasonable to assume that had ResProx been used prior to the coordinate deposition of these obsolete structures, the researchers might have been able to avoid the problems associated with publishing and/or depositing these incorrect structures.

This survey of obsolete NMR structures also revealed an example of ResProx's ability to detect structural flaws that go unnoticed by other equivalent resolution methods. More specifically, ResProx clearly identified models with a misplaced residue Glu105 in the obsolete NMR ensemble of the *E. coli* heme chaperone CcmE (PDB ID: 1LIZ; Fig. S6). For example, ResProx predicts an equivalent resolution of 4.7 Å for a "broken" model 3 of 1LIZ (Fig. S6B), while assigning an equivalent resolution of 2.4 Å to a properly built, "intact" model 1 (Fig. S6A). All the other methods fail to discriminate between the "broken" and "intact" models of 1LIZ and predict identical or almost identical equivalent resolution for these models. More detailed analysis reveals that the "bump score" from the GeNMR knowledge-based potential (Table S1), which is uniquely used in the ResProx method, plays a critical role in detecting the structural defect in this specific case (bump Z scores are −0.1 and 16.9, for "intact" model #1 and "broken" model #3 of 1LIZ, respectively). Similar to the GeNMR bump score, other quality scores that are unique to the ResProx method should allow ResProx to detect a larger variety of structural problems than what other equivalent resolution methods could identify.

Sensitivity to refinement

In a typical NMR structure generation protocol, model refinement occurs during the last few low-temperature steps of simulated annealing and during the final minimization. More sophisticated refinement schemes have been shown to produce a significant improvement of NMR structure quality, in comparison with the standard refinement techniques (Linge and Nilges 1999; Xia et al. 2002; Linge et al. 2003; Ramelot et al. 2009). For instance, refinement in explicit water may reduce the number of atomic clashes, unrealistic hydrogen bonds and unsatisfied hydrogen bond donors/acceptors. Water refinement can also improve the normality of backbone and side-chain torsion angles, and decrease the over-constraining of bond lengths, bond angles, omega torsion angles, and side-chain planarity (Linge and Nilges 1999; Xia et al. 2002; Linge et al. 2003; Nabuurs et al. 2004). Interestingly, when water refinement was applied to well-restrained NMR models, no significant changes in structure accuracy were observed (Spronk et al. 2002). To assess ResProx's ability to detect improvements in protein structure quality in the absence of changes in global accuracy, we tested the program on several NMR models before and after water refinement. All

Table 1 Equivalent resolution of ResProx, Procheck-NMR, MolProbity, and RosettaHoles2 for obsolete and current PDB entries of NMR structures

Protein	Version	PDB	ResProx (Å)	Procheck (Å)	MolProbity (Å)	RosettaHoles2 (Å)
AbrB N-terminal domain	Obsolete	1EKT	5.14	3.20	4.73	3.58
	Current	1Z0R	2.68	1.95	3.76	2.62
Ets-1	Obsolete	1ETC	6.29	3.00	5.03	3.63
	Current	1R36	2.77	1.78	3.53	2.74
CcmE	Obsolete	1LIZ	4.91	2.42	3.24	2.20
	Current	1SR3	2.22	2.40	2.94	2.14
Domain IV from the YbbR	Obsolete	2KPS	3.22	2.05	2.90	2.66
	Current	2L3U	2.86	1.75	2.78	2.62
SH3 of phospholipase C-gamma	Obsolete	1HSP	5.80	2.90	4.40	2.94
	Current	2HSP	4.78	3.13	4.16	2.85
MRF-2 DNA-binding domain	Obsolete	1BMY	5.05	3.13	4.62	3.47
	Current	1IG6	1.80	1.60	1.88	2.34
E. coli thioredoxin	Obsolete	1TRX	2.03	1.50	2.07	2.25
	Current	1XOB	1.41	1.35	1.27	1.8

protein models were obtained from the DRESS database (Nabuurs et al. 2004). In every case, ResProx was able to detect improvements of model quality due to water refinement. In particular, the decrease in ResProx's resolution values correlated well with the improvement (i.e. increase) in the Z-scores (a structure quality measure) given in the DRESS database (Table 2). The sensitivity of ResProx scores to variations in protein model quality due to very local adjustments (e.g. after water refinement) and changes in global structure (e.g. due to NOE violations or under-restraining) suggests that it could be used as an initial step in the detection of NMR models with a large range of structural problems. If required, the exact structural defects in those structures could be further identified by analysing restraint violations, restraint quality, and individual global or per-residue quality scores.

Typical equivalent resolution of NMR structures

Having shown that ResProx fulfills the requirements for a robust measure of structure resolution (excellent agreement with X-ray resolution, sensitivity to number of experimental restraints, sensitivity to NMR restraint violations and coordinate errors, sensitivity to orthogonal measures of NMR structure accuracy, and sensitivity to refinement), we decided to apply ResProx to calculate typical equivalent resolution for a large sample of randomly-selected NMR structures. For this study, a total of 500 randomly selected X-ray and 500 randomly selected NMR structures were assessed. Our results clearly show that NMR structures consistently exhibit lower ResProx values than X-ray structures, with the average NMR ResProx value being 3.2 ± 1.24 Å and the average X-ray ResProx value being 2.1 ± 0.40 Å. In other words, the equivalent atomic

Table 2 Improvements in the quality of water refined models: comparison between ResProx values and DRESS Z-scores

Protein	PDB	Refined	DRESS Z-score	ResProx (Å)
Intestinal fatty acid-binding protein	1A57	−	−4.46	5.50
		+	−2.72	2.56
Designed protein G core variant	1FD6	−	−1.4	2.49
		+	0.33	1.42
Rho GDP-dissociation inhibitor	1AJW	−	−2.79	2.90
		+	−1.24	2.03
Nudix enzyme hydrolase	1F3Y	−	−2.27	3.19
		+	−1.31	2.29
MTH1175	1EO1	−	−3.31	3.69
		+	−1.63	2.42

resolution of a randomly chosen NMR structure is about 1.1 Å worse than a randomly chosen X-ray structure. A bar graph displaying the distribution of NMR ResProx values along with the actual resolution distribution of 500 randomly selected X-ray structures is shown in Fig. S6. The low equivalent resolution of typical NMR structures should not be too surprising. Previous attempts to estimate the equivalent resolution of NMR structures have given similar values. In particular, a study that compared the level of agreement between observed and predicted backbone chemical shifts of X-ray and NMR structures estimated the apparent resolution of NMR structures as 3.0–3.5 Å (Neal et al. 2003). Likewise, a comparison of various Z-scores for X-ray and NMR structures led Spronk et al. (Spronk et al. 2004) to conclude that the average NMR structure has an equivalent resolution of 4 Å. Beyond these estimates of apparent or equivalent resolution for NMR structures, there is a significant body of evidence supporting the idea that

NMR structures are generally of lower quality than X-ray structures. For instance, an investigation of proton density distributions revealed that many high-precision NMR structures have an over-packed interior, whereas lower-precision structures display a mix of under- and over-packed regions (Ban et al. 2006). These distortions in protein packing could be behind another observation that NMR structures are significantly less stable in MD simulations and have higher internal energies than crystallographic models (Fan and Mark 2003). More recently, Andrec et al. (Andrec et al. 2007) assessed the structures of 35 proteins that were solved by both NMR and X-ray crystallography using Procheck (Laskowski et al. 1993), Verify3D (Eisenberg et al. 1997), ProSaII (Sippl 1993), and MolProbity (Davis et al. 2007) and came to the conclusion that NMR structures were consistently of lower quality than X-ray structures.

NMR structures need not necessarily be doomed to be perpetual basement-dwellers in the structure quality pennant race. In fact, with ResProx, we found a number of NMR structures (Fig. S6) that exhibited an equivalent resolution equal to or better than an average X-ray structure (ResProx value < 2.1 Å). As noted earlier, the equivalent resolution of an NMR structure can be improved by adding more experimental observables (Fig. 3) and/or a priori information, such as protein solvation effects (Table 2). The ubiquitin structure (PDB ID: 1D3Z) determined via multiple RDCs along with extensive NOEs and torsion angle restraints (Cornilescu et al. 1998) can serve as a useful example of a high-quality NMR structure (mean ResProx value of 1.8 Å). So far, the smallest equivalent resolution or ResProx score we have observed to date is 0.65 Å, corresponding to a model in a ubiquitin NMR ensemble (PDB ID: 1XQQ) that underwent a unique, simultaneous refinement of its structure and dynamics by NOEs and experimental order parameters in explicit water with the CHARMM 22 force-field (Lindorff-Larsen et al. 2005). These data suggest that the capacity to determine very high resolution NMR structures is already in hand and that the use of functions like ResProx (or other measures of equivalent of resolution) will not only help the NMR community achieve higher standards in protein structure determination but also improve perceptions about the quality of NMR structures in the scientific community.

ResProx limitations

While we firmly believe that ResProx is a powerful method of evaluating protein structure quality, its range of utility is not infinite. As with any equivalent resolution method, ResProx inherits the shortcomings of crystallographic resolution, which ResProx attempts to emulate. In particular, it is important to remember that crystallographers use resolution only as an approximate or simplified measure of protein quality. In protein crystallography, there are several other parameters that should be considered when evaluating protein structure quality (Wlodawer et al. 2008). For instance, the likelihood that a particular structure can be solved to a certain level of coordinate accuracy will depend not only on its experimental resolution but also other measures of X-ray data quality, such as data completeness, accuracy of the averaged reflection intensities (R-merge), reflection signal-to-noise ratios, and the average ratio of reflection intensities to their estimated errors. However, even high-quality X-ray data cannot guarantee that mistakes will be not made by the researcher(s) who solves the structure. These mistakes can be detected by evaluating the agreement between the model and experiment with R-factor and R_{free} parameters. They can also be identified by assessing the quality of the stereochemical parameters (e.g. bond lengths, bond angles, and peptide bond planarity) and the normality of its structural features, such as torsion angles and hydrogen bonding.

Similar to crystallographic resolution, ResProx scores should be treated as a first approximation for assessing protein structure quality. Furthermore, the following limitations need be considered. First, it must be recognized that ResProx evaluates the global quality of protein structures and, like any method for global protein structure assessment, is not a substitute for local or per-residue analysis. For instance, a single broken bond or a single incorrect torsion angle in otherwise perfect structure will likely be undetected by ResProx. Methods for per-residue analysis of protein structure, such as Procheck (Laskowski et al. 1993), WhatCheck (Hooft et al. 1996), MolProbity (Davis et al. 2007), VADAR (Willard et al. 2003) or PROSESS (Berjanskii et al. 2010) should always be used when a comprehensive analysis of structural quality is required. Second, the aforementioned sensitivity of ResProx to under-restrained models and NOE violations should not mislead researchers into using ResProx as a substitute for a detailed analysis of the quality of experimental data or the agreement between an NMR model and its restraints. More specialized measures, such as RPF scores (Huang et al. 2005), NOE R-factors (Gronwald et al. 2000), U-scores (Nabuurs et al. 2005), RDC Q-factors (Cornilescu et al. 1998), RDC R-factors (Clore and Garrett 1999) and traditional restraint violations should be used for these purposes. Third, ResProx may overestimate the quality of misfolded idealized models, such as models that can be generated by the Rosetta family of methods such as Rosetta-NMR (Bowers et al. 2000), CS-Rosetta (Shen et al. 2009), and CS-DP Rosetta (Raman et al. 2010). Indeed, ResProx is primarily designed to handle near-native protein structures. Even though ResProx is able to detect inaccurate NMR structures in our tests, the method was not

developed to serve as an ab initio protein folding potential or decoy detector and, therefore, may overestimate the resolution of misfolded idealized models with well-formed secondary structure elements. Fourth, ResProx should not be routinely used to estimate the resolution of unfolded or intrinsically disordered proteins and small extended peptides. Unfolded protein models are often used as the starting point in NMR structure generation. Consequently, ResProx detects and treats unfolded or disordered structures as "starting" structures and assigns them the upper limit of equivalent resolution (i.e. 10 Å). Finally, ResProx doesn't perform particularly well with protein complexes containing ligands such as DNA, RNA, and larger drugs. This is due to certain limitations in ResProx's sub-programs. Binding to DNA or RNA may induce a significant change in a protein molecule. As a consequence, removal of DNA or RNA from a protein complex often leaves the protein structure in an unrealistic conformation and therefore reduces the accuracy of ResProx's resolution estimate by about 10–15 %. Although NMR structures of protein-DNA and protein-RNA complexes are not yet very abundant in the PDB, efforts to improve ResProx's performance for such complexes are underway in our group.

Conclusion

Single value, experimentally-derived parameters, such as atomic resolution or number of NOEs per residue, still dominate the way we think about and assess protein structures determined by crystallographers or NMR spectroscopists. Like grades on a test, they are easy to remember, easy to understand and easy to compare. However, experimental measures don't necessarily catch experimentalist's errors. Over the past few years a number of efforts have been made to develop "derived" single-value parameters that assess proteins more on their coordinate quality than their experimental quality. These include, but are not limited to, predictors of equivalent resolution that are built into structure validation programs, such as Procheck-NMR (Laskowski et al. 1996), MolProbity (Davis et al. 2007), and Rosetta-Holes2 (Sheffler and Baker 2009, 2010). Unfortunately, these methods take into account only limited number of protein features (<5) and cannot provide a very comprehensive assessment of protein quality. To address this issue, we have developed a new measure of equivalent resolution—called Resolution-by-Proxy or ResProx—that is based on 25 protein quality metrics. Extensive assessments show that ResProx fulfills all the requirements for a robust measure of structure resolution including excellent agreement with observed X-ray data (R = 0.92), sensitivity to experimental numbers of restraints, sensitivity to NMR restraint violations and coordinate error, sensitivity to orthogonal measures of

NMR structure accuracy, and sensitivity to refinement. Applying this method to 500 randomly selected NMR structures, we found that the average equivalent resolution for NMR structures is about 3.2 Å (about 50 % worse than the average for X-ray structures). Furthermore, we were able to demonstrate that ResProx could be used to identify under-restrained, poorly refined or incorrectly modeled NMR structures and detect protein defects that other equivalent resolution methods could not. We believe that ResProx now offers a robust route for the scientific community to directly compare the quality of NMR structures to the quality of X-ray structures using the same common and universally understood measure—atomic resolution. The ResProx web server and a VMware virtual image of the ResProx program are available at http://www.resprox.ca.

References

Andrec M, Snyder DA, Zhou Z, Young J, Montelione GT, Levy RM (2007) A large data set comparison of protein structures determined by crystallography and NMR: statistical test for structural differences and the effect of crystal packing. Proteins 69(3):449–465

Ban YE, Rudolph J, Zhou P, Edelsbrunner H (2006) Evaluating the quality of NMR structures by local density of protons. Proteins 62(4):852–864

Berjanskii M, Tang P, Liang J, Cruz JA, Zhou J, Zhou Y, Bassett E, MacDonell C, Lu P, Lin G, Wishart DS (2009) GeNMR: a web server for rapid NMR-based protein structure determination. Nucleic Acids Res 37(Web Server issue):W670–W677

Berjanskii M, Liang Y, Zhou J, Tang P, Stothard P, Zhou Y, Cruz J, MacDonell C, Lin G, Lu P, Wishart DS (2010) PROSESS: a protein structure evaluation suite and server. Nucleic Acids Res 38(Web Server issue):W633–W640

Bowers PM, Strauss CE, Baker D (2000) De novo protein structure determination using sparse NMR data. J Biomol NMR 18(4):311–318

Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66(Pt 1):12–21

Clore GM, Garrett DS (1999) R-factor, Free R, and complete cross-validation for dipolar coupling refinement of NMR structures. J Am Chem Soc 121(39):9008–9012

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120(27):6836–6837

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB III, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35(Web Server issue):W375–W383

Eisenberg D, Luthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404

Fan H, Mark AE (2003) Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: a molecular dynamics simulation study. Proteins 53(1):111–120

Gronwald W, Kirchhofer R, Gorler A, Kremer W, Ganslmeier B, Neidig KP, Kalbitzer HR (2000) RFAC, a program for automated NMR R-factor estimation. J Biomol NMR 17(2):137–151

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. SIGKDD Explor Newsl 11(1):10–18

Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50(1):43–57

Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381(6580):272

Huang YJ, Powers R, Montelione GT (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(6):1665–1674

Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14(1):51–55, 29–32

Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. FEBS J 278(5):687–703

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26(2):283–291

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8(4):477–486

Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. Nature 433(7022):128–132

Linge JP, Nilges M (1999) Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. J Biomol NMR 13(1):51–59

Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins 50(3):496–506

Minor DL Jr (2007) The neurobiologist's guide to structural biology: a primer on why macromolecular structure matters and how to evaluate structural data. Neuron 54(4):511–533

Mitchell TM (1997) Machine learning. McGraw-Hill, New York

Nabuurs SB, Nederveen AJ, Vranken W, Doreleijers JF, Bonvin AM, Vuister GW, Vriend G, Spronk CA (2004) DRESS: a database of REfined solution NMR structures. Proteins 55(3):483–486

Nabuurs SB, Krieger E, Spronk CA, Nederveen AJ, Vriend G, Vuister GW (2005) Definition of a new information-based per-residue quality parameter. J Biomol NMR 33(2):123–134

Nabuurs SB, Spronk CA, Vuister GW, Vriend G (2006) Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. PLoS Comput Biol 2(2):e9

Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein 1H, 13C and 15 N chemical shifts. J Biomol NMR 26(3):215–240

Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol 264(1):121–136

Raman S, Huang YJ, Mao B, Rossi P, Aramini JM, Liu G, Montelione GT, Baker D (2010) Accurate automated protein NMR structure determination using unassigned NOESY data. J Am Chem Soc 132(1):202–207

Ramelot TA, Raman S, Kuzin AP, Xiao R, Ma LC, Acton TB, Hunt JF, Montelione GT, Baker D, Kennedy MA (2009) Improving NMR protein structure quality by Rosetta refinement: a molecular replacement study. Proteins 75(1):147–167

Richards FM (1977) Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng 6:151–176

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160(1):65–73

Seeliger D, de Groot BL (2007) Atomic contacts in protein structures. A detailed analysis of atomic radii, packing, and overlaps. Proteins 68(3):595–601

Sheffler W, Baker D (2009) RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. Protein Sci 18(1):229–239

Sheffler W, Baker D (2010) RosettaHoles2: a volumetric packing measure for protein structure refinement and validation. Protein Sci 19(10):1991–1995

Shen Y, Vernon R, Baker D, Bax A (2009) De novo protein structure generation from incomplete chemical shift assignments. J Biomol NMR 43(2):63–78

Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KK (2000) Improvements to the SMO algorithm for SVM regression. IEEE Trans Neural Netw 11(5):1188–1193

Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17(4):355–362

Spronk CA, Linge JP, Hilbers CW, Vuister GW (2002) Improving the quality of protein structures derived by NMR spectroscopy. J Biomol NMR 22(3):281–289

Spronk CA, Nabuurs SB, Krieger E, Vriend G, Vuister GW (2004) Validation of protein structures derived by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 45(3):315–337

Vasilief I (2011) QtiPlot—data analysis and scientific visualisation. http://soft.proindependent.com/qtiplot.html, 0.9.8.4 edn

Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8(1):52–56

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35(Web Server issue):W407–W410

Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS (2003) VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res 31(13):3316–3319

Wishart DS (2011) Interpreting protein chemical shift data. Prog Nucl Magn Reson Spectrosc 58(1):62–87

Wlodawer A, Minor W, Dauter Z, Jaskolski M (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J 275(1):1–21

Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 285(4):1735–1747

Xia B, Tsui V, Case DA, Dyson HJ, Wright PE (2002) Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. J Biomol NMR 22(4):317–331