J Biomol NMR (2008) 40:251–261
DOI 10.1007/s10858-008-9228-4

ARTICLE

The war of tools: how can NMR spectroscopists detect errors in their structures?

Edoardo Saccenti · Antonio Rosato

Received: 1 October 2007 / Accepted: 13 February 2008 / Published online: 5 March 2008

Abstract Protein structure determination by NMR methods has started in the mid-eighties and has been growing steadily since then. Ca. 14% of the protein structures deposited in the PDB have been solved by NMR. The evaluation of the quality of NMR structures however is still lacking a well-established practice. In this work, we examined various tools for the assessment of structural quality to ascertain the extent to which these tools could be applied to detect flaws in NMR structures. In particular, we investigated the variation in the scores assigned by these programs as a function of the deviation of the structures induced by errors in assignments or in the upper distance limits used. These perturbations did not distort radically the protein fold, but resulted in backbone RMS deviations up to 3 Å, which is in line with errors highlighted in the available literature. We found that it is quite difficult to discriminate the structures perturbed because of misassignments from the original ones, also because the spread in score over the conformers of the original bundle is relatively large. ϕ–ψ distributions and normality scores related to the backbone conformation and to the distribution of side-chain dihedral angles are the most sensitive indicators of flaws.

Keywords NMR · Quality assessment · Structure validation · Structure quality · Validation software

Abbreviations

NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
RMSD	Root mean standard deviation

Introduction

NMR structures are both an achievement by themselves and a complementary characterization of X-ray structures under many aspects (Billeter 1992; Wagner et al. 1992; Garbuzynskiy et al. 2005; Andrec et al. 2007). Indeed, when structure determination by NMR is coupled with, as it often happens, characterization of protein backbone dynamics, one gains information beyond the single snapshot of a static structure and actually obtains direct experimental insight into protein flexibility. A long-standing issue for protein structures solved by NMR (Clore et al. 1993; Hooft et al. 1996a; Doreleijers et al. 1998; Nabuurs et al. 2006; Brown and Ramaswamy 2007) or X-ray (Branden and Jones 1990; Kleywegt 2000; DePristo et al. 2004) is the evaluation of their quality. A set of recommendations for the presentation of NMR structures of proteins and nucleic acids, including some indications for structure validation, that dates back to 1998 (Markley et al. 1998) still constitutes the main contribution to the standardization of quality assessment in bio-NMR. In practice, although several measures of quality have been proposed

Electronic supplementary material The online version of this article (doi:10.1007/s10858-008-9228-4) contains supplementary material, which is available to authorized users.

E. Saccenti · A. Rosato (✉)
Magnetic Resonance Center, University of Florence,
Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy
e-mail: rosato@cerm.unifi.it

E. Saccenti
FiorGen Foundation, Via Sacconi 6, 50019 Sesto Fiorentino, Italy

A. Rosato
Department of Chemistry, University of Florence,
Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

also recently (Vriend and Sander 1993; Brünger et al. 1993; Hooft et al. 1996b; Lovell et al. 2003; Huang et al. 2005b), each laboratory is still using its own ensemble of quality checks and parameters. Even when different laboratories are using the same measures, they often compute them differently or assess these measures using subjective, non-reproducible criteria. Therefore, it is well possible that when analysed by different experts, the same structure will be assessed through different statistics, and thus assigned a different quality level. The fact that NMR structures often have a relatively low resolution can cause further difficulties in this respect.

For NMR structures, commonly used measures of accuracy are the size and number of residual restraint violations (Laskowski et al. 1996; Herrmann et al. 2002) and the statistics of the distribution of residues in the regions of the Ramachandran plot (Ramachandran et al. 1963; Morris et al. 1992; Hooft et al. 1997). A related issue is that of the measure of precision of an NMR structure, which is typically evaluated by the RMSD of the backbone atoms or of all heavy atoms within the bundle of conformers that is released in the PDB. However, this measure often leads to overestimating the precision of a structure. Residual violations and RMSD have been shown in the literature to be inadequate as proper indicators of structure quality (Spronk et al. 2003) In fact, solution structures can be refined to very good values for these parameters, and thus apparently show high accuracy and precision, while still having major flaws. A number of software tools have been developed over the years to assess the quality of protein 3-dimensional experimental structures (typically focusing only on X-ray structures) and/or of structures generated through molecular modelling approaches. These tools, or a combination thereof, can in principle be applicable also to the validation of NMR structures.

In this study we wanted to investigate how one can identify structural models derived from a partly incorrect analysis or assignment of NMR data using available software tools, by looking only at the resulting structure (which could be feature up to a 2–3 Å backbone RMSD from the real structure). We did not evaluate the agreement between the NOE data and the resulting structure as a quality measure because we aimed at analysing measures that are *independent* of the experimental data. Such measures are highly desirable because they can allow the assessment of a structure in a transparent, impartial manner and, in addition, they should be less easy to adjust. In addition, the importance and usefulness of carefully comparing the NOE data and the structure obtained from calculations have been already firmly established (Laskowski et al. 1996; Gronwald et al. 2002; Herrmann et al. 2002; Moseley et al. 2004). Situations of the kind addressed here can be incurred relatively often, e.g. when a few errors are introduced at the beginning of a structure calculation procedure and subsequent additional NOE assignments in practice tend to crystallize the structure around the initial error. We observe that relatively simple analyses based on the distribution of ϕ–ψ pairs in the structure and on *Z*-score values e.g. relative to the distribution of χ_1–χ_2 dihedral angles or to packing quality can be informative. It is likely that the analysis of these parameters from the early stages of structure calculation and throughout all phases of structure improvement until the final result would be useful to prevent deposition of wrong structures.

Material and methods

Protein test systems

We used different proteins as test systems to evaluate the performance of the validation tools tested: bovine Calbindin D_{9K}, a vitamin D-dependent calcium-binding protein, the second PDZ domain of the human neuronal adaptor X11α and the apo form of the A69P mutant of the sixth soluble domain of the Menkes protein (MNK6 hereafter). Ubiquitin was also used for a partial test. All deposited models have been retrieved from the RCSB Protein Data Bank (PDB) (Berman et al. 2000).

Calbindin D_{9K} (UniProt code P02633) is a 75 aminoacid calcium-binding protein, whose structure consists of 4 helices and 3 loops. A total of 1675 meaningful NOE data, 37 dihedral angles and 1,097 pseudocontact shifts have been used for structure calculations (Bertini et al. 2001). The PDB entry 1KQV has been used as reference after restrained energy minimization (which had not been performed on the original entry).

The second PDZ domain of the human neuronal adaptor X11α (UniProt code Q02410, PDB entry 1Y7N, Duquesne et al. 2005) is a 90 residues protein and its fold consists of 2 helices and 6 beta sheets. This is a high quality NMR structure determined by a group deeply involved in the development of structure validation and refinement methodologies. The original experimental data (1,725 meaningful NOE's and 95 dihedral angle restraints) have been retrieved from the BioMagResBank (Doreleijers et al. 2003) (mrblock_id 51978, bmrb_id 6113).

The structure of MNK6 (UniProt code Q04656) has been recalculated using 1956 meaningful NOE data and 82 dihedrals angles. MNK6 is a 75 residues protein with a $\beta\alpha\beta\beta\alpha\beta$ fold (PDB entry 1YJR, Banci et al. 2005).

The NMR (PDB entry 1D3Z, Cornilescu et al. 1999) and X-Ray structures (PDB entry 1UBI, Ramage et al. 1994) of human Ubiquitin (UniProt code P62988) have been also analysed. Ubiquitin is a 76 residues protein containing 3 helices and four β-sheets.

Structure calculations

For structure calculations we used the package PARAMAGNETIC CYANA 2.1 (Barbieri et al. 2004) an implementation of the CYANA program (Güntert et al. 1997) that allows the use of pseudocontact shifts in structure calculation. Families of 300 structures were annealed in 10,000 steps starting from randomly generated conformers using the available constraints. The structures obtained were refined through restrained energy minimization with the program Amber 8 (Case et al. 2004).

To mimic inaccurate NMR experimental data and generate flawed structures still having a plausible fold, the original experimental distance constraints have been randomly perturbed by adding and/or subtracting up to 75% of their original values. These data have been subsequently used to perform structure calculations as mentioned above. For Calbindin D_{9K}, we additionally generated different families of structures with wrong NOE assignments. In the first case we swapped the NOE constraints for the δ protons of residues Lys7 and Lys41 (Error A). In the second case, in addition to the previous misassignments, we interchanged all NOE values of the δ protons Lys16 and Lys25 (Error B). In the third case we swapped all the assignments of residues Glu51 and Glu64 (Error C). Qualitatively, the seriousness of the mistake(s) increases going from Error A to C. Similarly, for the sixth domain of the Menkes protein we switched all the NOE assignments of residues Cys18 and Cys35. The importance of this misassignment is analogous to that of Error C of Calbindin D_{9K}. Some RMSD data and violation statistics for the recalculated families are summarized in Supplementary Table S1.

Description of the validation tools tested

The following programs were used: FRST, HARMONY, HOPPscore, MolProbity, PROCHECK, ProSA (ProSA-Web), TAP, WHAT IF (WHAT CHECK).

FRST (Tosatto 2005) is a statistical scoring function implemented as a weighted linear combination of four different potentials: pairwise, solvation, RAPDF (Samudrala and Mould 1998), and hydrogen bond potentials. A combined energy is given by a weighted sum of the four potentials with the weights optimized on a representative set.

HARMONY (Pugalenthi et al. 2006) is a structure validation method based on the compatibility between the sequence and the structure of a protein considering the local environment of the residues. Structural descriptors such as backbone conformation, solvent accessibility and hydrogen bonding are used to characterise the structural environment of each residue position. Propensity and Substitution values are used together to predict the probability of the occurrence of an amino acid at each position in the sequence on the basis of the local structural environment. Proteins with misfolded regions score low. The reverse sequence is used as a control: regions in which the reverse sequence has a better score are likely to be affected by errors.

HOPPscore (Sims and Kim 2006): Structures are evaluated by comparing multiple φ–ψ pairs of short fragments of structure to a reference database of multiple φ–ψ angles from high resolution X-ray crystal structures. The overall quality of a model is given in the form of a plot showing a logarithmic fit of HOPPscore values for fragments spanning from 1 to 10 φ–ψ pairs. The degree to which all the various fragments lengths fit the reference database is then evaluated. The best fit is in the form $[m\cdot\ln(x) + b]$. Because of the fact that the steeper the descent the lower the quality of the model, the slope m of the fit curves is a measure of the overall structure quality. It indicates the degree to which all of the various fragment lengths fit the selected reference database. The parameter b is the score of the model to a single 2-dimensional φ–ψ Ramachandran plot. In general poor structures have slope $m < -1.00$. HOPPscore also gives a per-fragment and per-residue validation. Each fragment is assigned to one of four categories according to the frequency of that conformation relative to the average frequency within a conformational reference database. The four categories are: favoured (F), allowed (A), unfavoured (U) and disallowed (D) corresponding to the scores +2, +1, +0.5 and −4. The overall score is calculated by averaging the contributions of each fragment. Based on these criteria, also a per-fragment and a per-residue validation is given showing the fragment frequency and the score for each residue according to the different fragment length. Default values (Resolution 1.8 Å and 12° grid size) have been used.

MolProbity (Davis et al. 2007) is a web server offering quality validation for proteins structures. The output of a default analysis consists in two tables showing data for all-atoms contact and protein geometry (Ramachandran and rotamers outliers). This data are summarized in two overall scores: the clashscore and the MolProbity score. The clashscore is defined as the number of overlaps >0.4 Å. C^{β} are defined as the modelled C^{β} position to the ideal position calculated from the coordinates. Deviations >0.25 Å are likely to indicate an incompatibility between the sidechain and the mainchain.

PROCHECK (Laskowski et al. 1993, 1996) assesses the global structural quality through the G-factor, a measure of the fit to the most frequently occupied regions in the distribution of the dihedral angles. The G-factor provides a measure of how normal or how unusual a given stereochemical property is. It is computed for torsion angles (φ–ψ, χ_1–χ_2 combinations, χ_1 torsion angle for those

residues that do not have a χ_2, combined χ_3 and χ_4 torsion angles, ω torsion angles) for covalent geometry (main-chain bond lengths and main-chain bond angles) The *G*-factor is a log-odds score based on the observed distributions of these stereochemical parameters. Given a residue, the *G*-factor indicates the probability of a certain property to correspond to the given conformation. On the basis of the *G*-factor each rear side is assigned to a Core (*C*) Allowed (*A*) Generously Allowed (*G*) or Disallowed (*D*) class.

ProSA (and ProSA-web) (Sippl 1993; Wiederstein and Sippl 2007) uses knowledge based potentials of mean forces to evaluate model accuracy using C^{α} and/or C^{β} potentials. The program outputs a *Z*-score indicating the overall quality of the structure and a plot showing its local quality. The *Z*-score is normally negative and becomes more negative the larger the protein.

The TAP score (Tosatto and Battistutta 2007) is a criterion based on torsion angles propensities normalized against the global minimum and maximum. The output of TAP consists in a TAP score deriving from normalized torsion angle potential and indicating the degree of nativeness of a protein model. Native structures should have a TAP score close to 1 while lower values indicate some sort of incompatibility between sequence and structure (Tosatto and Battistutta 2007). TAP assigns the queried structure to an experimental quality class (MEdium, HIgh quality, Very High quality) inferring this from the analysis of a large data set of X-ray structures from Protein Data Bank. TAP score confidence estimate is given by two other parameters describing accuracy (that is the probability that the actual TAP score corresponds to a structure in the given quality class) and coverage (that is the fraction of structures in a given quality class that have, at least, the current TAP score).

WHAT IF (WHAT CHECK) (Hooft et al. 1996a) is a tool combining a number of different checks on file syntax and stereochemical and geometric properties of the structure. Global stereochemical quality parameters (Ramachandran plot appearance, 2nd generation packing quality, χ_1–χ_2 rotamer normality, Backbone conformation and others) are also checked and scored in the form of *Z*-scores.

All these validation tools are available trough web servers. Web server addresses are shown in Table 1.

Other software tools used

Original experimental data have been converted from the CNS format to the Cyana format using the CcpNmr FormatConverter software (Wanker et al. 2005). Figures have been prepared using the program MOLMOL (Koradi et al. 1996). Consumers for data analysis have been written in Python (van Rossum and Drake 2001). Energy minimizations have been carried out with the program Amber 8 (Case et al. 2004).

Results and discussion

Validation tools can be generally divided two broad groups, the first including tools that rely on the evaluation of interaction preferences or profiles (i.e. on the empirical evaluation of the energetics of folding), the second group including programs that evaluate geometrical and stereochemical properties. Among the methods analysed, ProSA-web, TAP, FSRT and HARMONY give an unique parameter indicating the overall quality of a structural model, while the other tools produce more extensive outputs, listing several different parameters usually concerning conformational and geometrical aspects. Energy-evaluating programs can also estimate the local quality of a structure, by calculating the energy on a per-residue basis. The availability of an unique parameter to globally describe the quality of a structure can be appealing especially to occasional users not expert in structural biology, who, for instance, are just looking for a good template for homology modelling.

We first of all checked whether the flawed families that we generated had a significantly worse overall energy of

Table 1 Programs analysed in the present study

Program	Web site	Reference
FRST	http://protein.cribi.unipd.it/frst/	Tosatto (2005)
HARMONY	http://caps.ncbs.res.in/harmony/	Pougalenthi et al. (2006)
HOPscore	http://hoppscore.lbl.gov/run.html	Sims and Kim (2006)
MolProbity	http://molprobity.biochem.duke.edu/	Davis et al. (2007)
PROCHECK	On-line version available through http://www-nmr.cabm.rutgers.edu/PSVS/	Laskowski et al. (1993, 1996)
ProsSA	http://prosa.services.came.sbg.ac.at/prosa.php	Wiederstein and Sippl (2007)
TAPscore	http://protein.cribi.unipd.it/tap/	Tosatto and Battistutta (2007)
WHAT CHECK	http://swift.cmbi.kun.nl/WIWWWI/	Hooft et al. (1996)

folding. This is actually not necessarily the case, because most of our perturbations in the data still maintain the overall fold of the protein, whereas individual secondary structure elements are distorted or wrongly packed one against the other (Figs. 1 and 2). These distortions are of the kind that can go through a normal NMR structure determination procedure. For the evaluation of the overall energy, we first used ProSA, which has a very convenient web interface. Table 2 reports ProSA *Z*-scores for the families analysed. All the unperturbed families scored in the same range of protein structures of the same size solved by NMR (as shown in the output display). In the case of both PDZ domain and Calbindin D_{9k} we could observe a trend of increasing *Z*-score values with increasing structural perturbation. The structures with 50% and 75% perturbation of the distance constraints indeed stood out as being affected by potential errors (Table 2). For the Calbindin D_{9k} and MNK6 families calculated with wrong assignments, there was not a consistent discrimination between the original and the flawed structure.

Using the program TAP, we observed a partial overlap of the scores between the correct and the wrong MNK6 families (Table 3). Averaging the score over the families differentiated the good from the bad structures. In the case of the PDZ domain and of the perturbed Calbindin D_{9k} structures, the TAP scores for the original structure were in the range of scores of structures of high experimental quality and there was no overlap with the perturbed

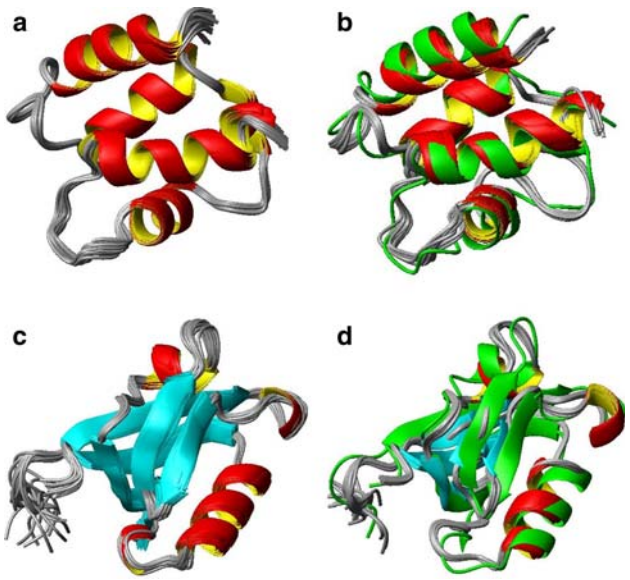

Fig. 1 The deposited reference structures of (**a**) Calbindin and (**c**) the PDZ domain (PDB entry 1KQV and 1Y7N, respectively) are shown together with the corresponding 75% perturbed structure families (panels **b** and **d**, respectively). In panels **b** and **d**, the first model of the reference family is shown in green superimposed to the perturbed bundle

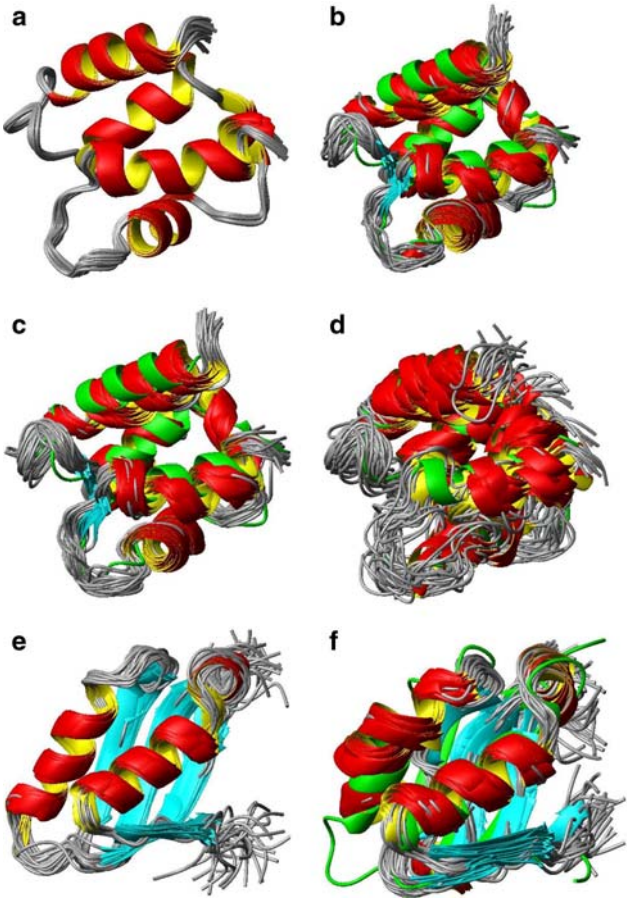

Fig. 2 The three flawed families calculated for Calbindin (Error A, B and C) are shown in panels **b**, **c** and **d**. The reference 1KQV structure is shown in (**a**). The flawed family calculated for MNK6 is shown in (**f**) whereas the reference family 1YJR is shown in panel **e**. In panels **b**–**d** and **f**, the first model of the reference family is shown in green superimposed to the flawed bundle

models. For misassigned Calbindin D_{9k}, the TAP score failed to differentiate the good family from any of the wrong ones, but again averaging could help. The torsion angle potential on which TAP is based has been also implemented in the FRST package (Tosatto 2005), which provides a number of different energetic terms (Supplementary Table S2). However, the analysis of all these parameters was not obvious also due to the significant intra-family fluctuations. These resulted in a considerable overlap of the ranges spanned by the various structures, preventing the identification of bad models.

PROCHECK was one of the first methods implemented to assess structural quality and soon the φ–ψ plot itself became a tools for crystallographers to standardize the stereochemistry quality of a structure. As reported in Table 4, the φ–ψ plot correlates well with structure quality. In spite of its simplicity, PROCHECK seemed able to discriminate the flawed structures from the original ones relatively well: only a couple of misassigned Calbindin

Table 2 ProSA *Z*-scores

Structure	ProSA *Z*-score
PDZ (1Y7N)	−4.66/−3.82
PDZ perturbed 25%	−4.35/−3.51
PDZ perturbed 50%	−2.22/−1.15
PDZ perturbed 75%	−0.89/+1.36
Calb D_{9k} (1KQV)	−6.78/−5.96
Calb D_{9k} perturbed 25%	−4.08/−2.69
Calb D_{9k} perturbed 50%	−2.81/−1.81
Calb D_{9k} perturbed 75%	−2.85/−1.6
Calb D_{9k} Error A	−6.55/−4.11
Calb D_{9k} Error B	−5.96/−4.67
Calb D_{9k} Error C	−5.47/−3.46
Menkes 6 (1YJR)	−4.66/−3.72
Menkes 6 Error	−4.78/−3.13

For each family of structures the minimum and the maximum *Z*-score values are shown

Table 3 TAP scores

Structure	TAP score
1Y7N	0.7576/0.7946
PDZ perturbed 25%	0.6811/0.7181
PDZ perturbed 50%	0.6287/0.6362
PDZ perturbed 75%	0.6299/0.6457
Calb D_{9k} (1KQV)	0.7198/0.7480
Calb D_{9k} perturbed 25%	0.6247/0.6371
Calb D_{9k} perturbed 50%	0.6157/0.6409
Calb D_{9k} perturbed 75%	0.6159/0.6246
Calb D_{9k} Error A	0.6937/0.7359
Calb D_{9k} Error B	0.6858/0.7451
Calb D_{9k} Error C	0.6656/0.7171
Menkes 6 (1YJR)	0.6737/0.7211
Menkes 6 Error	0.6401/0.6937

For each family of structures the minimum and the maximum TAP scores are shown. The TAP score derives from a normalized torsion angle potential and indicates the degree of nativeness of a protein model. Native structures should have a TAP score close to 1

D_{9K} families scored similarly to the reference structure (Table 4).

Other strategies based on the fit of most occupied φ–ψ regions have been proposed, such as HOPPscore (Sims and Kim 2006) which is based on a new evaluation method for higher order backbone torsion angle maps. Out of the HOPPscore output, Table 5 reports only the *m* and *b* fit values and the HOPPscore for four φ–ψ pairs. According to the authors, models with a score of the four φ–ψ pair lower than 0.50 should be carefully re-examined. The first result worth of notice is that all our test structures scored below zero and thus should considered of poor quality. As authors do not mention explicitly NMR structure validation but generally refer to HOPPscore as a "method for scoring the model quality of experimental and theoretical protein structures", we decided to compare the NMR and a X-ray structure of human ubiquitin. The RMSD between the first conformer of the NMR family and the X-Ray structure is 0.52 Å. HOPPscore values indicated in both cases good models with high local quality: both structures have an HOPPscore >0.5 even for 8 φ–ψ pairs. In the case of the PDZ domain, the HOPPscore system failed to separate the flawed families from the original one based on the overall quality score, but could discriminate them on the basis of the four φ–ψ pair score as well as on the basis of the *b* parameter (which essentially corresponds to a measure of the Ramachandran plot quality). Indeed, some differentiation was already apparent from the PROCHECK data of Table 4. Also in the case of original vs. incorrect structures of Calbindin D_{9k} and MNK6, the four φ–ψ pair is better for the original structure. However, the *b* parameter here was less informative, even though the PROCHECK evaluation of the Ramachandran plot appearances was quite selective (e.g. for MNK6, the percentage of residues in core regions dropped from 77% to 60%, and the RMSD between the two structures was as large as 3.5 Å).

We then checked the possible use of the program MolProbity, focusing on its unique features such as the evaluation of all-atoms contacts and C^{β} deviations. The use of both MolProbity score (Table 6) and the clashscore could identify the structures resulting from perturbed NOE data, but were not sufficient to consistently highlight the structures obtained after swapping some assignments. A per-residue analysis of clashes, rotamer and/or Ramachandran outliers was able to identify as problematic the residues whose NOE assignments were switched in the case of MNK6.

WHAT IF currently is the most used, if not the most complete, tool for assessing the quality of a structure and checking errors. Together with PROCHECK it is a *de facto* standard for structure and structural model validation. As the output of the complete WHAT IF check is quite verbose and given in the form of an extended report, we have summarized only a few selected WHAT CHECK *Z*-scores (Table 7). The values obtained for the correct structures fell in the range of values commonly found for NMR deposited in the PDB (Nabuurs et al. 2004; Nederveen et al. 2005) even if many of them were flagged as worrisome by WHAT CHECK. Notably, for the original Calbindin D_{9k} structure, the Ramachandran plot as evaluated by PROCHECK was good (Table 4). Instead, WHAT CHECK flagged it as problematic (even though still well in the range of NMR structures in the PDB). This discrepancy could be due to the way in which the Ramachandran plot appearance *Z*-score is defined or to the reference database,

Table 4 Results of the PROCHECK analysis

Structure	Core (%)	Allowed (%)	Generously allowed (%)	Disallowed (%)
PDZ (1Y7N)	90.4	9.1	0.4	0.1
PDZ perturbed 25%	68.1	27.5	4.3	0
PDZ perturbed 50%	20.3	45.5	20.3	15.9
PDZ perturbed 75%	30.4	21.7	29.3	18.8
Calb D_{9k} (1KQV)	89.2	10.8	0.0	0.0
Calb D_{9k} perturbed 25%	41.5	35.4	15.4	7.7
Calb D_{9k} perturbed 50%	12.3	43.1	20.0	24.6
Calb D_{9k} perturbed 75%	20.0	33.8	32.3	13.8
Calb D_{9k} Error A	87.7	10.8	1.5	0.0
Calb D_{9k} Error B	81.5	16.9	1.5	0.0
Calb D_{9k} Error C	60.8	30.5	7.2	1.5
Menkes 6 (1YJR)	77.9	18.3	2.0	1.7
Menkes 6 Error	60.0	36.9	0.0	3.1

The percentages of non-Gly, non-Pro residues of all conformers in the bundle falling in the various regions of the Raamachandran plot are given

or both. This observation highlights the opportunity of using a combination of more measures to assess the quality of a structure, even if the measures are relatively similar. Indeed, it is always possible to bias a structure calculation, e.g. by tuning the force field used in molecular dynamics, to obtain good scores on a given desired aspect. However, the present data suggest that there is little correlation between different quality measures therefore limiting the possibility of fudging calculations to obtain artificially structures with high scores. For all flawed structures, at least one of the WHAT CHECK parameters showed a significant deterioration with respect to the original structures. Even a relatively minor perturbation, such as in the case of Calbindin D_{9k} Error A, caused a large drop in the quality of backbone conformation and some worsening of the other parameters. However for some of the flawed structures the WHAT CHECK parameters still remained in the range of published NMR structures (Nabuurs et al. 2004; Nederveen et al. 2005).

A completely different approach is used by the validation tool HARMONY that assesses the compatibility between the sequence and the structure of a protein considering the local environment. As also mentioned for the program ProSA, the kind of perturbations presently introduced,

Table 5 Outputs of HOPPscore

Structure	HOPPscore overall quality score (m slope)		HOPPscore b parameter 2D φ–ψ plot score		HOPPscore for 4 φ–ψ pairs	
	Min/max	Average	Min/max	Average	Min/max	Average
PDZ (1Y7N)	−1.9951/−1.6866	−1.8420	2.0022/2.3166	2.1183	−0.75/0.06	−0.31
PDZ perturbed 25%	−2.3542/−2.0771	−2.1960	1.1974/1.7644	1.4197	−2.37/−1.70	−2.05
PDZ perturbed 50%	−1.7429/−1.4726	−1.5917	−1.2302/−0.6750	−0.9879	−4.00/−3.76	−3.86
PDZ perturbed 75%	−1.6293/−1.3600	−1.4906	−1.4399/−0.8577	−1.1582	−4.00/−3.65	−3.85
Calb D_{9k} (1KQV)	−1.5597/−1.329	−1.4245	1.5941/2.1610	1.8472	−0.54/0.14	−0.11
Calb D_{9k} perturbed 25%	−1.8034/−1.5866	−1.7101	−0.7793/−0.2571	−0.5160	−3.87/−3.24	−3.60
Calb D_{9k} perturbed 50%	−1.5450/−1.2653	−1.3835	−1.6267/1.0696	−1.4001	−4.00/−3.85	−3.87
Calb D_{9k} perturbed 75%	−1.4507/−1.2030	−1.3285	−1.7327/1.2599	−1.4882	−4.00/+3.64	−3.90
Calb D_{9k} Error A	−1.9504/−1.4770	−1.7203	1.5761/2.1211	1.8612	−2.97/−0.06	−0.67
Calb D_{9k} Error B	−2.2115/−1.5919	−1.8234	1.5576/2.2756	1.8366	−1.191/−0.08	−0.88
Calb D_{9k} Error C	−2.2546/−1.6343	−1.8819	0.6820/1.4447	1.1444	−2.52/−1.12	−1.87
Menkes 6 (1YJR)	−1.7002/−1.3442	−1.5088	1.0840/1.6326	1.3466	−1.38/+0.01	−0.85
Menkes 6 Error	−2.1935/−1.7179	−1.9603	0.5659/1.3437	0.9472	−3.24/−1.36	−2.06
Ubiquitin 1UBI (X-Ray)	−0.9625	–	2.2716	–	+1.40	–
Ubiquitin 1D3Z (NMR)	−0.9420/−0.7295	−0.8674	2.1364/2.3696	2.2809	1.26/1.46	1.36

For each family of structures the minimum, the maximum and the average HOPPscores are shown

Table 6 Outputs of MolProbity

System	Clashscore	Rotamer outliers (%)	Ramachandran outliers (%)	C^{β} deviations	MolProbity score
PDZ (1Y7N)	29.9	7.2	0	1	2.7
PDZ perturbed 25%	47.9	20.3	4.9	0	3.9
PDZ perturbed 50%	224.6	50.7	20.8	0	5.1
PDZ perturbed 75%	476.0	46.0	39.0	0	5.4
Calb D_{9k} (1KQV)	7.7	13.2	1.4	0	2.7
Calb D_{9k} perturbed 25%	43.0	44.0	24.6	0	4.4
Calb D_{9k} perturbed 50%	123.2	51.5	41.0	0	4.9
Calb D_{9k} perturbed 75%	297.8	48.5	42.5	0	5.3
Calb D_{9k} Error A	10.1	20.6	0	0	3.1
Calb D_{9k} Error B	7.4	17.6	0	0	2.4
Calb D_{9k} Error C	23.9	14.7	13.7	7	3.6
Menkes 6 (1YJR)	10.3	12.3	4.1	0	3.0
Menkes 6 Error	35.5	21.9	5.5	6	3.8

The table reports for the first model of each family of structures all-atoms data (Clashscore) and structure geometry data (Rotamer outliers, Ramachandran outliers, C^{β} Deviations). Structure geometry data are summarized by the MolProbity score

Table 7 Outputs of WHAT CHECK

Structure	Ramachandran Plot appearance	2nd Gener packing quality	$\chi_1-\chi_2$ rotamer normality	Backbone conformation
PDZ (1Y7N)	−1.954	0.904	−0.437	−1.249
Recalc	−2.088	−0.733	−2.327	−0.174
PDZ perturbed 25%	−4.515	−1.307	−3.988	−4.588
PDZ perturbed 50%	−7.006	−4.460	−4.623	−16.942
PDZ perturbed 75%	−6.864	−5.673	−5.133	−17.615
Calb D_{9k} (1KQV)	−4.191	−0.756	−1.263	−0.578
Calb D_{9k} perturbed 25%	−7.268	−6.174	−4.716	−20.509
Calb D_{9k} perturbed 50%	−7.938	−6.005	−4.907	−18.079
Calb D_{9k} perturbed 75%	−7.573	−4.728	−4.372	−20.034
Calb D_{9k} Error A	−4.347	−2.330	−2.876	−4.190
Calb D_{9k} Error B	−4.456	−3.076	−3.076	−4.117
Calb D_{9k} Error C	−5.617	−4.393	−3.383	−8.039
Menkes 6 (1YJR)	−4.148	−2.591	−2.726	−2.498
Menkes 6 Error	−5.210	−3.409	−3.652	−7.415

For each family of structure *Z*-scores for the Ramachandran plot appeareance, the 2nd generation packing quality, the $\chi_1-\chi_2$ rotamer normality and the backbone conformation are shown. *Z*-scores values <-3 are usually flagged as *poor*, *Z*-scores values <-4 are flagged as *bad.* Positive values are better then average, negative values worse than average

though representative of common errors in normal NMR practice, affect the protein fold to a modest extent. Therefore, the application of HARMONY to detect essentially local flaws in NMR structures is really not within the scope of the program. Indeed, there is not a consistent trend in the program evaluation of quality (not shown).

Conclusions

Many excellent research teams have been contributing over the last years fine ideas and tools for assessment of the quality or validation of NMR structures (Spronk et al. 2002; Bertini et al. 2003; Huang et al. 2005a, b; Nabuurs et al. 2006; Bhattacharya et al. 2007). These teams are still working on this topic (Joosten and Vriend 2007), and will be producing more results in the future. In this work we evaluated how one can apply these tools to detect possible flaws in NMR structures, flaws that are of the kind that can result at the end of structure calculation protocols (e.g. swapped assignment of side chain NOEs). It is to be noted that the kind of errors introduced may be considered relatively gross (see supplementary table S1). Nevertheless, because they lead to significant distortions of the structures while maintaining several features intact, such as individual secondary structure elements, they are representative of

perhaps more subtle errors that can arise e.g. due to misinterpretation of experimental data (Nabuurs et al. 2006). These errors may become severely misleading when one attempts to use the structure to interpret biochemical/biological evidence. The spread of the scores in Tables 2–7 and the backbone RMSD within each bundle (supplementary table S1) do not correlate, corroborating the notion that the precision of a structure does not impact on its quality.

A first conclusion is that for typical NMR structures, the ranges of the scores output by several of the validation tools comprise both good and flawed structures without consistent differentiation. A significant contribution to this feature is also due to the fact that there is a sizable intra-family variation of the scores, regardless of the intra-family RMSD values. Because of this variability and also because often there is not a clear functional form for the dependence of the scores on e.g. protein size, it is difficult to set threshold levels that discriminate "good" from "bad" structures. Consequently, at a first analysis significantly perturbed structures can bypass several validation tools. This conclusion is particularly valid in cases where there have been one or a few misassignments or modest errors in integration of the NOEs. Grossly incorrect structures could instead be detected easily by nearly all the tools analysed here. However, problems of the first kind are much more common in the practice of the determination of the solution structure of proteins.

Operatively, it appears that the distribution of ϕ–ψ pairs, either in the "simple" Ramachandran plot (taking into account only the percentage of residues in the core region) or in the more sophisticated analysis performed by HOPPscore, constitutes one the best indicators of structure quality, provided that backbone dihedral angle restraints are applied loosely as done here. WHAT CHECK *Z*-scores for χ_1–χ_2 rotamer normality or for backbone conformation and 2nd generation packing are also quite informative. One has to keep in mind that some scores can be significantly affected by the force field applied in molecular dynamics calculations/refinement. As an example of this, note the significant worsening of the score for χ_1–χ_2 rotamer distribution for the PDZ domain simply upon recalculating the structure with a simple protocol involving CYANA plus AMBER minimization (Table 7). The χ_1–χ_2 rotamer distribution was also the most sensitive to the application of a standardized refinement protocol on a database of 500 NMR structures (Nederveen et al. 2005). State-of-the-art force fields and molecular dynamics protocols should thus be used. The final validation of a structure should nevertheless always be based on the combination of a panel of tools such as those mentioned in this work. When possible, quality parameters should also be inspected on a per-residue basis (Nabuurs et al. 2005), to help in the identification of problematic regions.

Various programs can give indications of the presence of problems, but their performance is not always consistent, and some flawed structures scored as well as the correct one. Averaging over all conformers in a bundle proved useful with TAP (Table 3). Therefore, it may be relatively easy to evaluate the relative quality of two different bundles, but, in the absence of an absolute threshold, this of course does not solve the problem of whether the structure is correct or not. The definition of a threshold for structural quality parameters does not appear feasible at present.

The large spread of scores observed within a bundle of conformers and, correspondingly, within the whole ensemble of the NMR structures deposited in the PDB, can be due, at least in part, to the fact that normal calculation protocols do not explicitly take into account structure quality. These protocols are typically iterative and comprise various steps of structure analysis/error correction/addition of new NOEs, in which the only goal is to minimize the RMSD of the bundle and the residual violations of upper distance limits. Care should be taken to evaluate quality parameters also during the iterative calculation process and not just at the end, in order to detect potential problems as early as possible. It is likely that this approach would result in tighter and better ranges of quality scores for deposited NMR structures.

Acknowledgements This work was stimulated by the activities of the Coordination Action "NMR-Life" (funded by the European Commission, project no. 18758). We thank Prof. Ivano Bertini for many useful discussions. We thank the Ente Cassa di Risparmio di Firenze, the EC (project no. 213010), and MIUR (project PRIN 2005) for financial support. E.S. is the recipient of a fellowship from the FiorGen Foundation.

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 69:449–465

Banci L, Bertini I Cantini F, Migliardi M, Rosato A, Wang S (2005) An atomic-level investigation of the disease-causing A629P mutant of the Menkes protein, ATP7A. J Mol Biol 352:409–417

Barbieri R, Luchinat C, Parigi G (2004) Backbone-only protein solution structures with a combination of classical and paramagnetism-based constraints: a method that can be scaled to large molecules. ChemPhysChem 5(6):797–806

Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. Proteins 66:778–795

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucl Acids Res 28:235–242

Bertini I, Cavallaro G, Luchinat C, Poli I (2003) A use of Ramachandran potentials in protein solution structure determinations. J Biomol NMR 4:355–366

Bertini I, Donaire A, Jimenez B, Luchinat C, Parigi G, Piccioli M, Poggi L (2001) Paramagnetism-based versus classical constraints: an analysis of the solution structure of Ca Ln calbindin D9k. J Biomol NMR 21:85–98

Billeter M (1992) Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals. Q Rev Biophys 25:325–377

Branden CI, Jones TA (1990) Between objectivity and subjectivity. Nature 343:687–689

Brown EN, Ramaswamy S (2007) Quality of protein crystal structures. Acta Crystallogr D Biol Crystallogr 63:941–950

Brünger AT, Clore GM, Gronenborn AM, Saffrich R, Nilges M (1993) Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. Science 261:328–331

Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2004) AMBER 8. University of California, San Francisco

Clore GM, Robien MA, Gronenborn AM (1993) Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. J Mol Biol 231:82–102

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1999) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120:6836–6837

Davis IW, Murray LW, Richardson JS, Richardson DC (2007) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res 32(Web Server issue):W615–W619

DePristo MA, de Bakker PI, Blundell TL (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. Structure 12:831–838

Doreleijers JF, Rullmann JA, Kaptein R (1998) Quality assessment of NMR structures: a statistical survey. J Mol Biol 281:149–164

Doreleijers JF, Mading S, Maziuk D, Sojourner K, Yin L, Zhu J, Markley JL, Ulrich EL (2003) BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. J Biomol NMR 26:139–146

Duquesne AE, de Ruijter M, Brouwer J, Drijfhout JW, Nabuurs SB, Spronk CAEM, Vuister GW, Ubbink M, Canters GW (2005) Solution structure of the second PDZ domain of the neuronal adaptor X11alpha and its interaction with the C-terminal peptide of the human copper chaperone for superoxide dismutase. J Biomol NMR 32:209–218

Garbuzynskiy SO, Melnik BS, Lobanov MY, Finkelstein AV, Galzitskaya OV (2005) Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray- and NMR-resolved protein structures? Proteins 60:139–147

Gronwald W, Moussa S, Elsner R, Jung A, Ganslmeier B, Trenner J, Kremer W, Neidig KP, Kalbitzer HR (2002) Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). J Biomol NMR 23:271–287

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Bio 273:283–298

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 24:209–227

Huang YJ, Moseley HN, Baran MC, Arrowsmith C, Powers R, Tejero R, Szyperski T, Montelione G (2005a) An integrated platform for automated analysis of protein NMR structures. Methods Enzymol 394:111–141

Huang YJ, Powers R, Montelione GT (2005b) 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

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

Hooft RW, Sander C, Vriend G (1996b) Verification of protein structures: side-chain planarity. J Appl Cryst 29:714–716

Hooft RWW, Sander C, Vriend G (1997) Objectively judging the quality of a protein structure from a Ramachandran plot. Comput Appl Biosci 13:425–430

Joosten RP, Vriend G (2007) PDB improvement starts with data deposition. Science 317:195–196

Kleywegt GJ (2000) Validation of proteins crystal structures. Acta Crystallogr D 56:249–265

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

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

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 8:477–486

Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Cα geometry: φ, ψ, and C^{β} deviation. Proteins 50:437–450

Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. J Biomol NMR 12:1–23

Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structures coordinates. Proteins 12:345:364

Moseley HN, Sahota G, Montelione GT (2004) Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR 28:341–355

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 33:123–134

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 55:483–486

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. doi: 10.1371/journal.pcbi.0020009

Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Güntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AM (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins 59:662–672

Pugalenthi G, Shameer K, Srinvasan N, Sowdhamini R (2006) HARMONY: a server for the assessment of protein structures. Nucl Acids Res 34(Web Server issue):W231–W234

Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configuration. J Mol Biol 7:95–99

Ramage R, Green J, Muir TW, Ogunjobi OM, Love S, Shaw K (1994) Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. Biochem J 299:151–158

Samudrala R, Moult J (1998) An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol 275:895–916

Sims GE, Kim S-H (2006) A method for evaluating the structural quality of protein models by using higher-order φ–ψ pairs scoring. Proc Natl Acad Sci USA 103(12):4428–4432

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

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

Spronk CA, Nabuurs SB, Bonvin AM, Krieger E, Vuister GW, Vriend G (2003) The precision of NMR structure ensembles revisited. J Biomol NMR 25:225–234

Tosatto SCE (2005) The Victor/FRST function for model quality estimation. J Comput Biol 12:1326–1327

Tosatto SCE, Battistutta R (2007) TAP score: torsion angle propensity normalization applied to local protein structure evaluation. BMC Bioinformatics 8:155–168

van Rossum G, Drake FL (eds) (2001) Python reference manual. PythonLabs, Virginia. Available via www.python.org

Vriend G, Sander C (1993) Quality controls of protein models: directional atomic contact analysis. J Appl Cryst 26:47–60

Wagner G, Hyberts SG, Havel TF (1992) NMR structure determination in solution: a critique and comparison with X-ray crystallography. Annu Rev Biophys Biomol Struct 21:167–198

Wanker WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas ML, Eldon JL, Ulrich MJI, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

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