

Robust and highly accurate automatic NOESY assignment and structure determination with Rosetta

Zaiyong Zhang · Justin Porter · Konstantinos Tripsianes · Oliver F. Lange

Received: 15 February 2014 / Accepted: 19 April 2014 / Published online: 21 May 2014
© Springer Science+Business Media Dordrecht 2014

Abstract We have developed a novel and robust approach for automatic and unsupervised simultaneous nuclear Overhauser effect (NOE) assignment and structure determination within the CS-Rosetta framework. Starting from unassigned peak lists and chemical shift assignments, auto-NOE-Rosetta determines NOE cross-peak assignments and generates structural models. The approach tolerates incomplete and raw NOE peak lists as well as incomplete or partially incorrect chemical shift assignments, and its performance has been tested on 50 protein targets ranging from 50 to 200 residues in size. We find a significantly improved performance compared to established programs, particularly for larger proteins and for NOE data obtained on perdeuterated protein samples. X-ray crystallographic structures allowed comparison of Rosetta and conventional, PDB-deposited, NMR models in 20 of 50 test cases. The unsupervised autoNOE-Rosetta models were often of significantly higher accuracy than the corresponding expert-

supervised NMR models deposited in the PDB. We also tested the method with unrefined peak lists and found that performance was nearly as good as for refined peak lists. Finally, demonstrating our method's remarkable robustness against problematic input data, we provided correct models for an incorrect PDB-deposited NMR solution structure.

Keywords Nuclear magnetic resonance · Automatic data analysis · Structure determination

Significance statement

Along with X-ray crystallography, NMR structure determination is the only available high-resolution method for protein structure determination. NMR spectroscopy is conducted in aqueous solution and thus be the only route towards high-resolution 3D structures for proteins that cannot be crystallized. However, the analysis of NMR data is very time-consuming and can generally only be conducted by highly trained NMR experts, which require between weeks and months to obtain accurate and precise structures. Here, we provide a novel method to automate the analysis process and show that accurate structures can be obtained. Remarkably, the automatically generated structures are, in a majority of the cases, more accurate than the structures laboriously generated by NMR experts. Our method promises to significantly increase the attractiveness and viability of NMR structure determination.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9832-4) contains supplementary material, which is available to authorized users.

Z. Zhang · J. Porter · O. F. Lange (✉)
Department Chemie, Biomolecular NMR and Munich Center for Integrated Protein Science, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany
e-mail: oliver.lange@tum.de

K. Tripsianes
Central European Institute of Technology (CEITEC), Masaryk University, Kamenice 5, 62500 Brno, Czech Republic

O. F. Lange
Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany

Introduction

Structure determination by nuclear magnetic resonance (NMR) spectroscopy is largely driven by distance

information gathered through nuclear Overhauser effect spectroscopy (NOESY). To use such data as distance restraints, the NOESY crosspeaks in multidimensional spectra have to be assigned to individual atoms of the biomolecular system. NOESY crosspeak assignment and structure generation steps are usually performed in an integrated, iterative manner. This maximizes the number of conformational restraints, while guaranteeing self-consistency amongst distance restraints (Wüthrich 1986).

Many of the repetitive tasks in NMR structure determination have been successfully automated (Guerry and Herrmann 2011; Moseley and Montelione 1999; Baran et al. 2004; Guntert 2009). Two such crucial tasks in the chain of the data analysis are the assignment of NOE crosspeaks and the determination of accurate structural models. Popular programs that perform these two tasks include ARIA (Linge et al. 2003a), CYANA (Guntert 2009; Guntert et al. 1997), AutoStructure (Huang et al. 2006) and UNIO (Serrano et al. 2012) and have recently been tested with good results in a blind-testing challenge (Rosato et al. 2012). However, a limitation of these programs is that they have to be able to generate a sufficiently accurate model from the initial set of assignments. This usually limits the methods to small proteins with high quality spectra, complete and accurate chemical shift assignments, and well-refined peak lists. When conditions are suboptimal, a calculation either does not converge, or worse, converges to a precise but inaccurate fold (Guerry and Herrmann 2011). Accordingly, these programs are not usually used unsupervised, and must instead be applied in combination with manual assignment and possibly peak list refinement by a skilled NMR expert. Indeed, in our own work on larger proteins, a few manual assignments were required to bootstrap the automated analysis with CYANA (Lange et al. 2012).

Here, we aim to develop an NOE assignment and structure determination algorithm that can—unsupervised—produce results that are both reliable and accurate. This algorithm should take chemical shift assignments and unassigned NOE peak lists as input and produce, without further user interaction, refined models of protein structures in atomic resolution.

To achieve this goal, we combine Rosetta structure prediction with automatic NOE assignment. It has been demonstrated that Rosetta, which searches for the lowest energy conformation of the polypeptide chain using physically realistic force fields, requires only very sparse NMR data to guide its search to accurate structures (Lange et al. 2012; Raman et al. 2010). The question we ask here is whether the very noisy automatically assigned NOE restraints might be able to provide sufficient guidance for Rosetta to yield accurate initial models. These models would then allow iterative refinement of NOE assignments until accurate high-quality structures and self-consistent

assignments can be generated. Iteration of automatic NOE assignment with structural modeling is, however, also the basis of established algorithms. Thus, the crucial question to be explored in this study is not whether iteration between modeling and assignment is a successful strategy, but rather if a significant benefit is gained by using the improved, but computationally more demanding, ROSETTA structural modeling, and if we can solve the engineering challenge to render the ROSETTA structure calculation sufficiently robust against the very noisy automatically assigned NOE restraints of the initial assignment stage. In cases where established programs cannot find converged initial models, and thus fail, the new approach might converge and thus applicability is broadened to include more challenging cases. Additionally, the more accurate modeling provided by the ROSETTA energy function might render the method more robust against erroneous input data and yield more accurate final 3D models.

To couple NOE assignment with ROSETTA, we build on the previously developed iterative structural modeling algorithm, RASREC, and extend it to become an algorithm for automatic NOE assignment. This entails the implementation of a new ROSETTTA module for automatic NOE assignment as well as the development of a robust protocol to couple the iterative search for the near-native protein structures in RASREC with iterative NOE assignment. The assignment module employs among other techniques, network anchoring (Herrmann et al. 2002), ambiguous restraints (Nilges et al. 1997), covalent structure compliance (Herrmann et al. 2002; Huang et al. 2005), structure dependent and independent peak calibration, and restraint combination (Herrmann et al. 2002). In our final protocol, the calculation consists of multiple iterations of structural sampling guided by automatically assigned NOE restraints. In early iterations, cross-peak assignments compatible with preliminary models are reinforced, but incompatible assignments are not removed. In later iterations, incompatible cross-peak assignments are removed from the restraint list. Throughout the whole process, however, a pool of best fitting structures is maintained that is ranked by the initial NOE assignments. This is a major difference to existing programs and helps us to prevent convergence on inaccurate but self-consistent solutions. Implementation details of the new method will be described elsewhere (Lange 2013).

To investigate the performance of the new methodology, we carried out a benchmark on 50 NOE data sets obtained from 41 protein samples of 63–370 residues length. To test the impact of *difficult* inputs on the performance of auto-NOE-Rosetta, we have included unrefined and automatically picked peak lists, as well as sparse data sets obtained from perdeuterated ILV-methyl labelled protein samples. To avoid unwittingly cherry-picking targets that work especially well for our method, we chose three pre-existing

benchmark sets and used *all* monomeric proteins from each (Rosato et al. 2012; Lange et al. 2012; Mao et al. 2011).

In the following, we report on the results of the benchmark. First, we will contrast the performance of AutoNOE-Rosetta with CYANA. Subsequently, we compare the accuracy of the unsupervised method with the state of the art of expert guided NMR structure determination as reflected in PDB-deposited NMR models. This is followed by an analysis of structure validation metrics and NOE completeness scores. Finally, we stress test the method with non-ideal input data, such as raw or unrefined peak lists or incomplete and erroneous chemical shift assignments.

Results

Comparison with CYANA

We have defined a single set of parameters that is used to run all targets, including data preparation (e.g., automatic trimming of flexible tails), structure calculation and final model selection. Thus, results in similar quality as reported here should be achievable from application of the method to as yet unknown targets. We also provide a suite of scripts that allow the user to run the software in this unsupervised fashion. The entire benchmark set and the final models can be obtained from our website (www.csrosetta.org/benchmarks) and our results can be scrutinized by interested readers using our software and accompanying toolchain.

The benchmark comprises 50 NOE data sets derived from 40 different proteins ranging in size from 5.5 to 40 kDa. Input data are the sequence, chemical shift assignments and NOE peak lists (“Methods” section, SI Appendix, Table S1). In 20 cases, RDC data of the N–H bond vectors in one or more alignment medium was also included (SI Appendix, Table S1).

Multiple calculations are carried out with different weighting of the NOE data against the Rosetta Energy. One is selected from these based on a combination of final Rosetta Energy and the intrinsic precision of the resulting models (“Methods” section). Finally, to be accepted as a successful solution, the structures must fulfill two criteria: convergence and intrinsic NOE consistency (SI Appendix: Methods Section 2.2). AutoNOE-Rosetta was run successfully on 42 of 50 data sets, comprising 35 different proteins. Final models are shown in SI Appendix, Fig. S1 for all targets, and their accuracy is reported in SI Appendix, Table S2 as C_{α} -RMSD with respect to the reference structure. A number of targets have only been used after the autoNOE-algorithm was finalized, including all parameters, and the run selection protocol. These targets are DrR147D, MrR110B, OR8C, PFR13A, PsR293, SR384, SgR42, VpR247, and HmR11 and display similar performance as the other targets (SI Appendix, Table S2).

To provide a reference for the performance of autoNOE-Rosetta, we chose to run the popular program CYANA 3.0, which obtained the most accurate models in a recent community-wide blind structure determination challenge (CASD) (Rosato et al. 2012). In analogy to AutoNOE-Rosetta, we have defined an acceptance rule for CYANA. Based on suggestions of CYANA’s creator, Peter Güntert, we use a combination of convergence and CYANA’s target function (SI Appendix: Methods Section 2.1, Fig. S2).

CYANA was *successful* for 31 of 50 data sets according to its acceptance rule (“Methods” section). Thus, a significant improvement in both accuracy and radius of convergence for autoNOE-Rosetta is observed with respect to CYANA (Fig. 1). All structures that failed the automatic acceptance criteria in autoNOE-Rosetta also failed in CYANA, but eleven of the failing targets in CYANA were acceptable according to the criteria in autoNOE-Rosetta, and yielded accurate structures below 2.5 Å RMSD (Fig. 1). Furthermore, 10 of 17 inaccurate CYANA-structures (RMSDs > 2.5 Å) were determined accurately by autoNOE-Rosetta (RMSDs < 2.5 Å). Numerical values of the C_{α} -RMSD against the reference structures for CYANA and autoNOE-Rosetta can be found in (SI Appendix: Table S2).

Comparison with PDB-deposited NMR structures

The state-of-the-art in high-resolution NMR structure determination typically involves not just a single CYANA run, but performing several rounds of CYANA-based NOE assignment and refinement of the input peak lists (or even manual assignments, going through peak-by-peak), followed by simulated annealing in XPLOR or CNS (considered to have a better force field than CYANA), and finally a high-resolution refinement in explicit water (Linge et al. 2003b), where RDCs are used if present. To directly compare autoNOE-Rosetta to this more complex structure determination protocol, we included 20 protein targets in our benchmark for which both a conventionally determined solution NMR structure and an X-ray crystal structure are available. We further assume that the state-of-the-art in NMR structure calculation is well reflected in these 20 PDB-deposited NMR solution structures. Indeed, all these structures were deposited in the last decade, the program CNS is listed in all PDB headers (except 1xpv), and whenever the respective remark section is provided in the PDB header (12 of 20 cases), water refinement is mentioned explicitly.

In this study, we assume that the X-ray structure is an accurate representation of the dominant solution structure; accordingly, the RMSD of atomic coordinates between NMR and X-ray structure provides a measure for the accuracy of the NMR structure. This view is supported by the NMR data (SI Appendix: Table S3). Based on this criterion, autoNOE-Rosetta significantly outperforms conventional

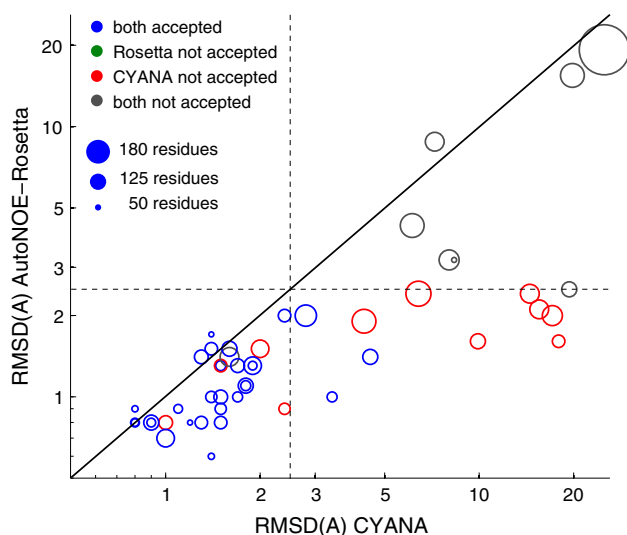


Fig. 1 Comparison of autoNOE-Rosetta with CYANA. Shown are the median C_{α} -RMSDs of final models with respect to their reference structure on logarithmic scale. The *diagonal line* indicates points of equal performance, points above the line correspond to targets for which CYANA yields lower RMSDs, and points below the line correspond to targets for which autoNOE-Rosetta yields lower RMSDs. The *dashed lines* mark 2.5 Å RMSD. The *size* of the proteins is proportional to the area of the symbol as indicated by the legend. The *color* indicates whether for CYANA, autoNOE-Rosetta or for both programs the final models are considered as success based on convergence and NOE consistency (SI Appendix, Method Section 2.0). RMSDs are capped at a maximum of 25 Å. Assignment statistics, convergence and accuracy of final models can be found in SI Appendix Table S7 and SI Appendix Table S8 for autoNOE-Rosetta and CYANA models, respectively. Comparing heavy-atom RMSDs instead of C_{α} -RMSDs yields a similar picture (SI Appendix, Fig. S6)

supervised NMR structure determination (Fig. 2a and SI Appendix, Table S4). For 10 of 21 targets, accuracy is significantly improved, and only for 2 of 21 it is decreased (CcR55, partially converged; ER690 unconverged). Moreover, if we restrict the analysis to the 19 converged targets, accuracy never deteriorates more than 33 %, whereas it improves for 7 targets significantly beyond 33 %. This is in stark contrast to the performance of established automatic assignment programs. Only 3 of the smallest targets of the benchmark set (<80 residues) yield sufficiently accurate results in CYANA to compete with PDB deposited NMR structures. For the other 18 of 21 targets, the structures obtained unsupervised with CYANA are >25 % worse in accuracy than PDB-deposited NMR structures (Fig. 2b). Of these 18 with deteriorated accuracy, 13 yield a tight structural bundle and 10 are acceptable according to the success criteria introduced above (SI Appendix: Tables S2–S8).

Analysis of structural quality

In addition to a high accuracy, we generally would like to obtain 3D models of proteins with a high structural quality.

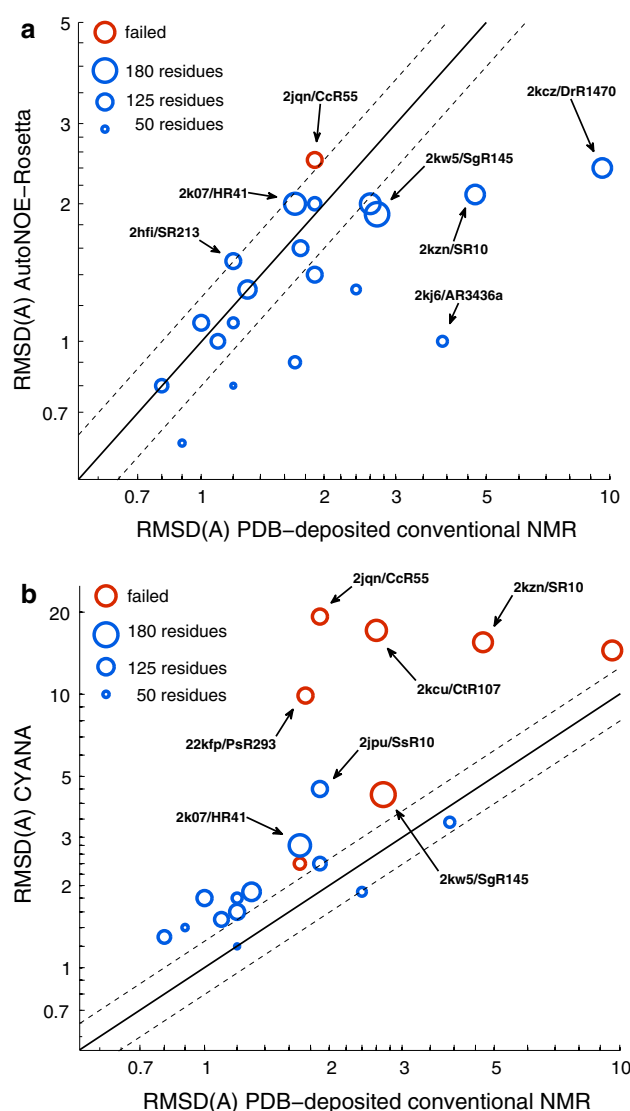


Fig. 2 Comparison of unsupervised automatic NOE models with expert-analyzed NMR solutions structures. The C_{α} -RMSDs of PDB deposited NMR models is plotted against final models obtained with **a** autoNOE-Rosetta and **b** CYANA. For AR3436a no X-ray structure is available as reference, but a new manually refined NMR solution structure, which supersedes 2kj6 (Fig. 4 and “Results” section). The *solid diagonal line* indicates points of equal performance, points above the line correspond to targets where PDB-deposited NMR structures have higher accuracy, and points below the line correspond to targets with higher accuracy of the autoNOE-Rosetta models. *Dashed lines* mark $\pm 25\%$ accuracy. The *size* of the proteins is proportional to the area of the *symbol* as indicated by the legend. AutoNOE-Rosetta or CYANA runs that are not converged (<90 % of residues converged) are shown in *red*. Comparing heavy-atom RMSDs instead of C_{α} -RMSDs yields a similar picture (SI Appendix, Fig. S7)

This quality is generally assessed by structural validation packages through various metrics, such as packing quality, Ramachandran consistency, and Janin-plots. NMR solution structures based on NOE distance restraints are prone to show deficits (Doreleijers et al. 2011), whereas un-restrained CS-Rosetta models were previously reported to show high

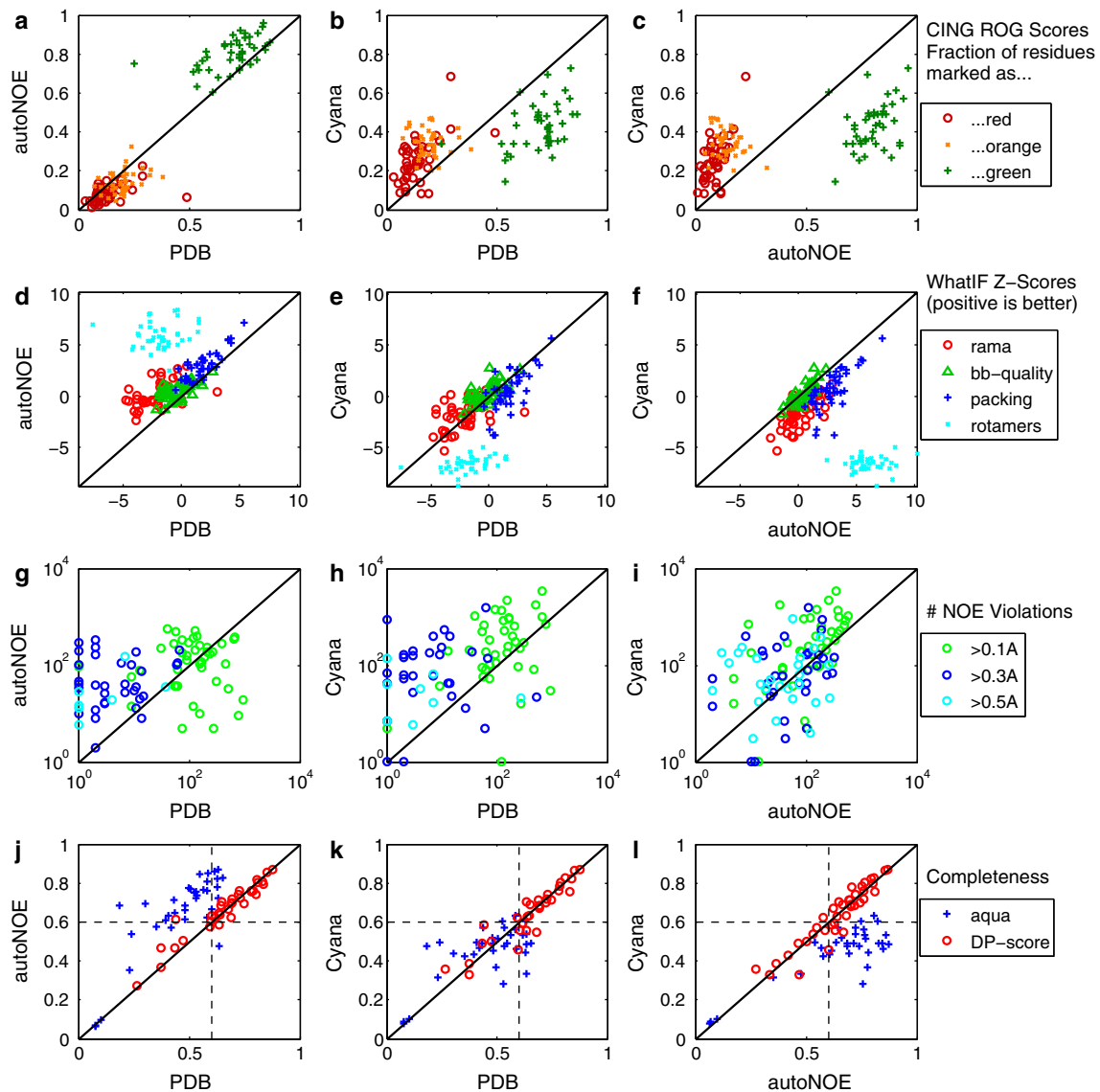


Fig. 3 Validation metrics for autoNOE-Rosetta, CYANA and PDB-deposited NMR models. Metrics computed for autoNOE-Rosetta and CYANA-models are compared to metrics computed on PDB-models, in panel-columns 1 and 2, respectively. Metrics between autoNOE and CYANA are compared directly in panel-column 3. **a–c** Fraction of residues annotated as *red*, *orange* and *green* by the iCING server's ROG score(legend). Less *red* and *orange* and more *green* residues is better. **d–f** WhatIF Z-scores for Ramachandran plot appearance,

backbone-quality, packing and chi-1/chi-2 rotamer normality (legend). Higher Z-scores are better. **g–i** The number of NOE restraints violated by structural models. Structural models of CYANA and autoNOE are analyzed together with the restraints produced by the respective algorithms. PDB-deposited models are analyzed with respect to the NOE restraints uploaded with the structures. **j–l** Completeness scores computed with AQUA (Doreleijers et al. 1999) and AutoStruct-DP (Huang et al. 2005). Higher numbers are better

structural quality but significantly lower accuracy than NOE-driven structure calculations (Rosato et al. 2012). We were curious to see whether autoNOE-Rosetta both, preserves the high structural quality, despite being subjected to a large number of automatically assigned NOE restraints, and yields more accurate structures than CS-Rosetta. To assess the structural quality of AutoNOE-Rosetta models, we used the online validation server iCING (Doreleijers et al. 2012), which performs WhatIF (Vriend 1990), PROCHECK (Laszkowski et al. 1996) and its own structural analysis.

The iCING-ROG score summarizes and integrates different validation measures into a single score and annotates individual residues as *green*, *orange* and *red* to convey an increasing level of alertness for unphysical local structure (Doreleijers et al. 2012). AutoNOE-Rosetta models produce generally less red and orange residues than PDB NMR-models or CYANA models (Fig. 3a–c). WhatIF compares local structure of the protein against common structural knowledge derived from high-resolution X-ray structures (Vriend 1990). Figure 3d–f shows the WhatIF

structure Z-scores on *Ramachandran plot appearance*, *backbone quality*, *1st generation packing quality*, and *chi-1/chi-2 rotamer quality*. AutoNOE-Rosetta models generally are of higher quality than PDB NMR models or CYANA models.

Analysis of restraint violations

Another popular criterion for judging NMR structure quality is a low count of restraint violations by the final models. Figure 3g–i shows how often the final models violate the NOE-derived restraints by >0.1 , >0.3 and >0.5 Å. Generally, NMR restraint-sets deposited with their corresponding PDB structures have less violations above >0.3 or >0.5 Å than those obtained with CYANA or autoNOE-Rosetta, but CYANA and AutoNOE-Rosetta yield similar results. We found that, for AutoNOE-Rosetta ensembles, many of the violations occurred at side-chains that adopted multiple conformations. In these cases, each conformation would actually be consistent with a subset of the violated NOE restraints involving this side-chain, and it would be plausible that dynamic averaging causes the assigned NOE cross-peaks. Since it is well possible that dynamic averaging might be the reason for some of the observed violations, as well as the fact that programs could trivially remove any violated restraint from the restraint-list, it is questionable whether the count of restraint violations is actually a valuable criterion for NMR structure validation. Indeed, we see no particular correlation between this measure and accuracy of the final models (C_{α} -RMSD) regardless whether they were downloaded from the PDB or generated with CYANA or autoNOE-Rosetta (SI Appendix: Fig. S3).

Analysis of NOE completeness

Since the autoNOE-Rosetta structures fit more accurately to X-ray structural models, a possible concern might be that Rosetta modeling is biased towards X-ray crystallographic artifacts rather than solution state structure. To verify that this is not the case, we show that autoNOE-Rosetta models yield a better or equivalent interpretation of the NMR data in comparison to conventional NMR solution structures, as quantified by the AQUA completeness (Doreleijers et al. 1999) and the AutoStruct DP score (Huang et al. 2005). AQUA reads the models and restraint list and checks how many of the proton–proton contacts in the model are actually observed as assigned NOEs. The more modern DP score uses chemical shift assignments and unassigned peak lists as input, and is thus independent of the specific restraint list. AQUA's completeness score is systematically better for AutoNOE-Rosetta than for PDB-NMR or CYANA models (Fig. 3j–l, black circles). For most targets the DP scores are comparable between the

different methods (Fig. 3j–l, blue crosses). However, for some PDB NMR structures with low DP-scores (<0.6), autoNOE-Rosetta was able to yield significant improvements. Overall these quality measures show that autoNOE-Rosetta models yield an interpretation of the NMR data that is as good or better than that of the PDB-deposited NMR models.

Performance under non-ideal input

Next, we were interested how autoNOE-Rosetta behaves when provided with problematic data. We tested both automatic (raw) and refined peak lists for 8 targets from round II of the blind, community-wide NMR structure determination challenge (CASD) (Rosato et al. 2009). In addition to the 8 *raw* data sets, we use 7 *unrefined* data sets from previous work (Lange et al. 2012) and one from CASD round I. For these *unrefined* data sets, peaks have been picked manually and chemical shift assignments have been validated, but the peak lists and chemical shift assignments have not yet undergone iterative refinement using structural models. Of the *unrefined* data sets, 6 stem from ILV-methyl labeled perdeuterated protein samples. Restraints obtained from such ILV-samples are inherently sparse, rendering structure calculation more challenging due to a lower restraint density. Moreover, the sparser NOE networks render the automatic validation of NOE cross-peak assignments via network anchoring less effective.

The availability of 9 targets with both *raw/unrefined* and *refined* data allows us to investigate the robustness of autoNOE-Rosetta. AutoNOE-Rosetta turns out to be remarkably robust; for 7 of the 9 *raw/unrefined* peak lists differences in accuracy are insignificant (<0.3 Å). In only two cases, StT322 and HR5460, was the accuracy significantly decreased. The automatic acceptance criteria successfully identified both these raw data sets as having produced untrustworthy results. Interestingly, autoNOE-Rosetta tends to select a lower weight for the NOE-based pseudo-energy contribution for *raw* peak lists compared to *refined* peak lists (Table 1), which is consistent with the presumed lower quality of the data.

Another type of challenging input is given by the 8 *unrefined* ILV data sets. AutoNOE-Rosetta succeeded on four of these data sets and yielded a partially converged structure for another (HmR11). CYANA, however, did not succeed on any of these 8 data sets. Is the deciding factor, which makes these data sets so challenging, the sparseness of the ILV data, the quality of the data sets (*unrefined* vs. *refined*), or the increased molecular weight (ILV data sets have a molecular weight between 15 and 21 kDa)? One can mostly exclude the increased molecular weight, as the driving factor for these failures, since both autoNOE and CYANA were significantly more successful on the *refined*

Table 1 Impact of raw peak lists

Target	Reference	Size	Residue ranges			Weight ratio ^a	C_{α} -RMSD (Å) to reference structure			
			of NMR sample	Used in Rosetta	RMSD analysis		Raw peak list		Refined peak list	
							CYANA	autoNOE-Rosetta	CYANA	autoNOE-Rosetta
StT322	2loj	38	1–63	26–63	26–63	0.04	8.3	3.2	1.4	1.7
HR6470	219r	48	1–69	11–58	11–58	1.00	0.8	0.9	0.8	0.8
OR135	2ln3	69	1–79	5–73	5–73	0.50	0.9	0.8	1.1	0.9
AR3436a	tbd ^c	80	1–97	14–93	14–93	0.40	3.4	1.0	1.7	1.0
HR6430	2la6	89	1–99	11–99	11–99	1.00	1.4	1.0	1.5	0.9
HR2876	2ltm	95	1–107	13–107	13–107	0.04	Not converged	1.6	1.4	1.5
YR313	2ltl	102	1–119	18–119	18–40, 46–115	0.10	1.6	1.4	1.7	1.3
OR36	2lci	128	1–134	1–128	1–128	0.50	2.0	1.5	n/a ^b	1.2
HR5460	2lah	150	1–160	11–160	19–160	0.02	Not converged	3.2	n/a ^b	1.8

^a Ratio of NOE-pseudo energy weights selected automatically by autoNOE-Rosetta for *raw* versus *refined* data sets

^b A segmentation fault in CYANA 3.0 prohibited us from finishing the structure calculation

^c The reference structure for AR3436a is a new NMR structure (PDB accession code: TBD) that results from a (manual) re-evaluation of the original NMR spectra

data sets in the same size range. Furthermore, we showed above that the influence of data quality (*raw* vs. *refined*) on autoNOE-Rosetta is low for small, double-labeled data sets. Thus, the lower success rate is likely a result of the sparseness of the ILV data.

Detecting problematic or unsuccessful calculations

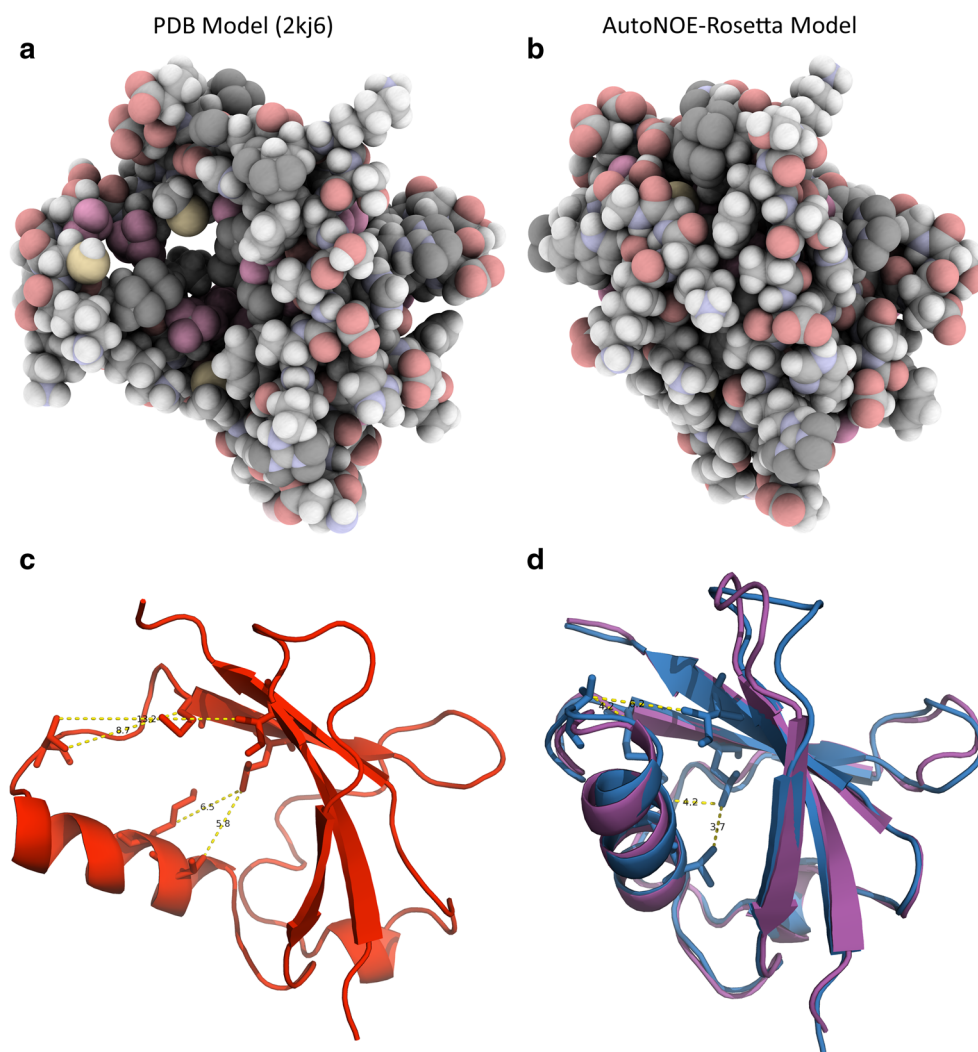
To run autoNOE-Rosetta or CYANA unsupervised, it is important to have clear criteria to flag problematic runs. This filter mechanism has to catch most, if not all, problematic results. In other words, the filter should produce little or no false positives. Some false negatives, on the other hand, are not as worrisome, as human experts can inspect a few such calculations. Here, we have introduced clear definitions for such a filter rule based on convergence of structures and NOE self-consistency for CYANA and autoNOE-Rosetta (SI Appendix: Methods Section 2). Of the eight declined calculations performed with autoNOE-Rosetta, four failed both criteria, and four (two each) failed only one of the criteria. The data sets that only failed the consistency criterion, are YR313(*raw*) and StT322(*raw*). While YR313 yielded accurate structures (1.4 Å) in autoNOE-Rosetta, StT322 did not (C_{α} -RMSD 3.2 Å; SI Appendix Fig. S4). For CcR55, HR5460(*raw*), and HmR11(*unrefined*) only 88, 79 and 75 % of residues converged, respectively, failing the criterion of 90 % convergence by only a small margin. In these cases, the converged part of the structure is reasonably accurate (C_{α} -RMSDs of 1.3, 2.0, and 3.1 Å, respectively) and would provide an advanced starting point for further iterative and structure based refinement of the data set (SI Appendix, Fig. S4).

Hence, as intended, the filter has been successful in producing no false positives and only very few false negatives.

Detailed discussion of target AR3436a

The data presented here shows that autoNOE-Rosetta yields accurate results even when the peak lists are not well refined. In the following we discuss a fortuitous discovery that demonstrates that autoNOE-Rosetta is not only robust against problematic peak lists, but also shows remarkable accuracy in the face of incomplete or erroneous side-chain chemical shift assignments. During our work on the here-presented benchmark we were initially puzzled by one outlier. For this outlier, AR3436a, autoNOE-Rosetta yielded structures that were 3.8 Å away from the PDB-deposited NMR solution structure (2kj6). The AR3436a data set stems from the CASD set, and was originally posed as a blind challenge to the community. The results of this competition seemed fairly standard except the CS-Rosetta models were identified as an outlier (Rosato et al. 2012): all NOE driven programs produced structures close to the PDB-deposited structure (1.4–2.2 Å) and with acceptable, albeit slightly borderline, validation scores. However, a closer inspection of the NMR solution models (2kj6) reveals that the main helix is at an angle causing the hydrophobic core of the protein to be exposed (Fig. 4a–c). In the autoNOE-Rosetta models, in contrast, the helix is well packed against the core (Fig. 4b–d), which is more consistent with our understanding of the physical chemistry of hydrophobic protein cores. Moreover, the CS-Rosetta based submissions to the blind structure determination challenge also packed the helix against the core (with

Fig. 4 Structure determination of AR3436A from incomplete and erroneous input data. Shown are two models of AR3436A in space-fill (a, b) and cartoon visualization (c, d) to highlight the differences in packing of the hydrophobic core between the PDB-deposited NMR solution structure (a, c) and the structure obtained with AutoNOE-Rosetta from the same input data (b, d). Due to the incomplete and erroneous chemical shift assignments AutoNOE-Rosetta can only assign a few NOE-crosspeaks (yellow lines) that support the packing of the helix, nevertheless, these are sufficient to yield well packed structures. **d** The PDB-deposited models violate these NOE crosspeaks, demonstrating that the respective assignments were discarded because they didn't fit initial models



RMSD $> 4 \text{ \AA}$ to the reference NMR structure), but did not converge to a high-precision structural bundle.

These observations prompted us to investigate whether the better-packed structure obtained with AutoNOE-Rosetta might actually be better supported by the raw NMR data as well. Indeed, a careful analysis of the raw input data conducted together with members of the laboratory that authored the original data set revealed a number of problems [manuscript in preparation]. Although the backbone assignment was nearly complete and correct, the side-chain chemical shifts were incomplete and had miss-assignments. Additionally, the NOESY data were under-picked as indicated by the unbalanced Recall-Precision scores of the PSVS analysis, such that many potentially well resolved peaks were not contained in the original peak list. These issues hindered the structure calculations of NOE-driven programs, but had no influence on the CS-Rosetta calculations. After correcting these issues with the input data, the structures obtained with conventional methods matched with the autoNOE-Rosetta models obtained with either the

original data (1.0 \AA) or the new data (1.0 \AA). This shows that autoNOE-Rosetta is not only reliable with unrefined (raw) peak lists but also with raw (i.e., incomplete and erroneous) sidechain chemical shift assignments. We are now in the process of systematically investigating the influence of such raw chemical shift assignments on automatic NOE assignment methods and our preliminary results support the anecdotal case reported here. The advantage of this robustness of AutoNOE-Rosetta for the full NMR pipeline is obvious. Assignment of side-chain chemical shifts is often a major bottleneck to progress in an NMR structure determination project. Automatic methods, such as FLYA (Schmidt and Guntert 2012), might take the burden of manual assignment, but cannot be relied on to always yield the highest quality of resonance assignments. However, paired with autoNOE-Rosetta, which is more fault-tolerant than other methods, an accurate structure might still be generated either as final result, or as a starting point for further refinement of the chemical shift assignments.

Discussion

We developed a new method for automatic NOE assignment and NMR structure determination, which we tested on a benchmark of 50 data sets including 20 for which X-ray crystallographic reference structures were available. A final convergence and NOE consistency filter accurately discriminates between successful and failed runs, and all 42 runs that pass this filter yield an accuracy better than 2.5 Å C_{α} -RMSD. Thus, we successfully combined the most important traits of CS-Rosetta with those of NOE-driven structure determination. The new algorithm is robust against missing or erroneous data as CS-Rosetta, but in the end exploits the full NOESY data to achieve the optimal precision and accuracy in final structures. In particular the lack of precision is problematic for CS-Rosetta, even if NOE-based filtering is applied (CS-DP-Rosetta (Raman et al. 2010)), as shown by the community wide assessment of structure determination (CASD) (Rosato et al. 2012).

The usefulness of an automatic NOESY assignment algorithms hinges on its ability to handle a wide variety of data. In fact, the quality of NOESY peak lists can vary dramatically as a function of the quality of the raw data, the method of picking peaks, and the level of peak list refinement. With 50 data sets from 41 different proteins, we are confident that our benchmark covers a realistic range of NMR data quality. To enhance the variety in the benchmark, we also included data sets at different stages of refinement (termed *raw*, *unrefined*, and *refined*). And in spite of this wide variety of input data quality autoNOE-Rosetta yields accurate results with striking consistency, which demonstrates a remarkable robustness of the method against challenging input data. Thus, autoNOE-Rosetta is a significant advance in *fully automatic* analysis of NMR data.

We were able to compare autoNOE-Rosetta ensembles with PDB-deposited NMR ensembles which reflect the state-of-the-art in NMR structure determination including final refinement in explicit water. Remarkably, the autoNOE-Rosetta results are either very close in accuracy (within 25 %) or significantly better (Fig. 2) than the PDB-deposited models. The most significant improvements were from 9.6 to 2.3 Å for the double-labelled sample, DrR1470, and from 4.7 to 2.1 Å for the triple-labelled, ILV-protonated sample, SR10 for which our calculations started from an *unrefined* data set.

AutoNOE-Rosetta ensembles' high accuracy—both relative and absolute—is especially remarkable considering that we are comparing an automated, unsupervised method with expert driven iterative and structure based refinement, as it is reflected in PDB deposited structures. For experts in NMR data analysis the method will provide better starting points for refining challenging data sets. For

non-experts it will allow a safe and straightforward application of NMR structure determination to routine cases. We are confident that our method provides a significant progress towards unsupervised automatic NMR structure determination, which is likely to broaden the applicability of NMR for structure determination in academic and non-academic labs.

Methods

Benchmark

The 50 data sets comprising target sequence, assigned chemical shifts, and unassigned peak lists were obtained from three published sources (SI Appendix: Table S1): (1) all data sets available by December 2012 at the community wide assessment of NMR structure determination (CASD) (Rosato et al. 2012; Rosato et al. 2009) (currently hosted at <http://www.wenmr.eu/wenmr/casd-nmr-data-sets>), (2) all monomer data sets from a recent molecular replacement (MR) benchmark (Mao et al. 2011) (http://psvs-1_4-dev-nesg.org/MR/dataset.html) (SI Appendix: Fig. S5), (3) all targets from our previous work (Lange et al. 2012).

Peak lists from the first prediction period of CASD (CASDI) are refined. For targets from the second prediction period of CASD (CASDII), both, *refined* and *raw* (automatically picked) peak lists are available. For MR targets, the status of the peak lists is unknown but assumed *refined*, and for ILV-targets the peak lists and chemical shift files are *unrefined*, that is chemical shift assignments have been verified and peaks have been picked by a human expert (Lange et al. 2012), but the data sets have not undergone iterative refinement using structural models.

To analyze the accuracy of final structures, we computed the C_{α} -RMSD on all residues that are structured in the reference. Tails that were not well defined (flexible) in the reference structure are excluded from RMSD computation as specified in Table S1. For 11 reference structures, also internal loop-regions were not well defined and had to be excluded from RMSD calculations. Detailed justifications for these exclusions are given in Table S9. For a given method, autoNOE-Rosetta or CYANA, the ten final models are superimposed with the reference structures to compute C_{α} -RMSDs and heavy-atom RMSDs.

AutoNOE-Rosetta

AutoNOE-Rosetta structure calculations were run with parameters as detailed here (Lange 2013). Fragments were picked by the Rosetta3 fragment picker (Vernon et al. 2013) using the provided chemical shift data. Homologous proteins using an e-value cutoff of 0.05 (sequence identity

>20 %) were excluded from fragment picking. Tolerances for NOESY cross-peak assignment were set for all targets to 0.3, 0.3, 0.03 and 0.04 for ^{13}C , ^{15}N , direct ^1H , and indirect ^1H dimension, respectively. Residual Dipolar Coupling data were used where available (SI Appendix, Table S1).

For data sets with *unrefined* or *refined* peak lists, NOE-restraint strengths of 5, 10, 25 and 50, respectively are chosen, and for targets with *raw* peak lists restraint strengths of 1, 2, 5, 10, 25, and 50. For each restraint weight 3 independent runs were carried out with different random seeds. The 10 lowest energy structures yield the final ensemble of a given run.

To identify the optimal run the resulting ensembles were ranked as follows: The converged residues are identified as those with a C_α -RMS fluctuation of less than 2 Å, as reported previously (Lange et al. 2012). The average pairwise RMSD is computed on converged regions (SI Appendix, Methods), and an effective precision (EP) is computed from pairwise RMSD and fraction of converged residues. For each run with constraint weight w_{cst} a cumulative score $S = E - 12 \log w_{\text{cst}} + 5\text{EP}$ is computed, where E denotes the median Rosetta all-atom energy of the ensemble. If in any of the runs more than 2,000 peaks with initial assignments are removed, because final models violate them (column ‘violated’ in Table S11), only E is considered for selection of runs, otherwise the final run is selected using S .

The models of the top-ranking ensemble are further relaxed against the automatically assigned NOEs including intra-residue and sequential NOEs using a tenfold increased NOE-restraint weight. If this procedure reduces the number of NOE violations to less than 40 % of the violations counted in the ensemble of un-relaxed models, the relaxed models are accepted as final models, otherwise the un-relaxed models are kept as final models. This was the case for data sets HR2876(raw), YR313(raw), and CTR107. For all other data sets the relaxed models are kept as final models. This refinement step generally reduces NOE-violations without significantly affecting backbone RMSD to the reference structure.

We established two criteria for *successful* calculations: (1) reasonable NOE consistency (target-function <500) and (2) convergence (SI Appendix: Methods 2.2). For the convergence criterion the number of well defined residues has to reach 90 % or more of the total number of residues with random coil index (RCI) derived S (Guerry and Herrmann 2011) order parameter (Berjanskii and Wishart 2005) larger than 0.7 (“Methods” section and SI Appendix, Table S2d).

A few NOE data sets were recorded with reduced sweep width leading to *peak folding*. AutoNOE-Rosetta unfolds such frequencies on the fly, if the sweep-window is noted

in the header of the respective peak list. For CYANA calculations we manually *unfolded* by replicating peaks with integer multiples of the sweep width subtracted or added to the respective frequencies. This applies to four peak lists of two proteins of our benchmark and the corresponding sweep-width parameters are given in (SI Appendix, Table S6).

AutoNOE-Rosetta is parallelized for the MPI framework and runs were either carried out on our in-house cluster or on JUROPA at the Juelich Supercomputer center using 184 or 192 parallel processes, respectively.

Instructions to run autoNOE-Rosetta including command-lines can be found in the Manual or Tutorial sections of our website (www.csrosetta.org) and in SI Appendix: Methods.

Cyana structure calculations

Cyana 3.0 calculations were carried out to provide readers with a familiar reference for each target. TALOS+ restraints were generated from the chemical shift data, and 100 initial, and 20 final models were generated using 20,000 steps of torsion angle dynamics. RMSDs were computed from the 20 final models using the same residues and reference structure as for autoNOE-Rosetta models (SI Appendix: Methods for example script). All TALOS+ predicted phi and psi angles with prediction class ‘Good’ are used. Two schemes to derive torsion restraints from TALOS+ predictions were tested. ACO_TIGHT restraints were generated by computing the lower- and upper bound as $\phi \pm \Delta\phi$, where ϕ denotes the TALOS+ predicted torsion angle in degree, and $\Delta\phi$ the TALOS+ estimated standard deviation. For ACO_LOOSE, we obtained bounds as $\phi \pm 2\max(\min(\Delta\phi, 35), 10)$. ACO_TIGHT is the recommended protocol at the NMR facility of the Center for Advanced Biotechnology and Medicine (CABM) as described here (http://www.nmr2.buffalo.edu/enter/NMRWiki/images/2/2e/Talos2dyana_taloseerrors.txt). ACO_LOOSE is the protocol that derives from applying the talos2dyana.com executable packaged with the TALOS+ software. A comparison of both protocols shows that ACO_TIGHT yields better accuracy over all targets (SI Appendix, Fig. S2a). Thus, ACO_TIGHT is used in all further CYANA calculations.

Where RDC data was available, CYANA runs were carried out both, with and without RDC data. A weight of 0.02 was used for the RDC restraint, and 0.2 as cutoff for RDC violation output. For each alignment medium 5 additional pseudo-residues of type LL5 and 1 of type ORI are attached at the end of the protein sequence. Alignment tensor parameters, D_{zz} and R , are estimated using the macro FindTensor.cya which employs the histogram method (Clare et al. 1998). This protocol was obtained from <http://>

www.nmr2.buffalo.edu/nesc/wiki/CYANA. RMSDs of CYANA calculations with RDCs were generally higher than CYANA calculations without RDCs (SI Appendix, Fig. S2b), whereas RDC data leads to improved results for autoNOE-Rosetta (SI Appendix, Fig. S2c). Thus, CYANA calculations *without* RDCs are compared to autoNOE-Rosetta *with* RDCs throughout the study.

Acknowledgments We are indebted to the work of the Northeast Structural Genomics Consortium, which has made available NMR data sets for such a high number of proteins including those with known X-ray structure. Only the extensive high-quality work of the NESG and its open attitude towards sharing data enabled this study in the breadth and thoroughness now presented. Furthermore, we have to thank Peter Güntert and Paolo Rossi for many helpful discussions, as well as Michael Sattler for carefully reading the manuscript. This work was supported by DFG grant LA 1817/3-1.

References

- Baran MC, Huang YJ, Moseley HNB, Montelione GT (2004) Automated analysis of protein NMR assignments and structures. *Chem Rev* 104:3541–3556
- Berjanskii M, Wishart D (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* 127:14970–14971
- Clore GM, Gronenborn AM, Bax A (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson* 133:216–221
- Doreleijers JF, Raves ML, Rullmann T, Kaptein R (1999) Completeness of NOEs in protein structures: a statistical analysis of NMR data. *J Biomol NMR* 14:123–132
- Doreleijers JF et al (2011) NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. *Nucleic Acids Res* 40:D519–D524
- Doreleijers JF et al (2012) CING: an integrated residue-based structure validation program suite. *J Biomol NMR* 54:267–283
- Guerry P, Herrmann T (2011) Advances in automated NMR protein structure determination. *Quart Rev Biophys* 44:1–53
- Güntert P (2009) Automated structure determination from NMR spectra. *Eur Biophys J* 38:129–143
- Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 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* 319:209–227
- Huang Y, Powers R, Montelione G (2005) Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J Am Chem Soc* 127:1665–1674
- Huang YJ, Tejero R, Powers R (2006) A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins Struct* 62:587–603
- Lange OF (2013) Implementation of automatic NOE assignment in Rosetta. *J Biomol NMR*. doi:10.1007/s10858-014-9833-3
- Lange OF et al (2012) Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc Natl Acad Sci USA* 109:10873–10878
- Laskowski RA et al (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486
- Linge JP, Habeck M, Rieping W, Nilges M (2003a) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 19:315–316
- Linge JP et al (2003b) Refinement of protein structures in explicit solvent. *Proteins* 50:496–506
- Mao B, Guan R, Montelione GT (2011) Improved technologies now routinely provide protein NMR structures useful for molecular replacement. *Structure* 19:757–766
- Moseley HN, Montelione GT (1999) Automated analysis of NMR assignments and structures for proteins. *Curr Opin Struct Biol* 9:635–642
- Nilges M, Macias MJ, Odonoghue SI, Oschkinat H (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol* 269:408–422
- Raman S et al (2010) Accurate Automated Protein NMR Structure Determination Using Unassigned NOESY Data. *J Am Chem Soc* 132:202–207
- Rosato A et al (2009) CASD-NMR: critical assessment of automated structure determination by NMR. *Nat Methods* 6:625–626
- Rosato A et al (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. *Structure* 20:227–236
- Schmidt E, Güntert P (2012) A new algorithm for reliable and general NMR resonance assignment. *J Am Chem Soc* 134:12817–12829
- Serrano P et al (2012) The J-UNIO protocol for automated protein structure determination by NMR in solution. *J Biomol NMR* 53:341–354
- Vernon R, Shen Y, Baker D, Lange OF (2013) Improved chemical shift based fragment selection for CS-Rosetta using Rosetta3 fragment picker. *J Biomol NMR* 57:117–127
- Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8:52–56
- Wüthrich K (1986) NMR of proteins and nucleic acids. The George Fisher Baker non-resident lectureship in chemistry at Cornell University. Wiley, p 292