

CASD-NMR 2: robust and accurate unsupervised analysis of raw NOESY spectra and protein structure determination with UNIO

Paul Guerry¹ · Viet Dung Duong¹ · Torsten Herrmann¹

Received: 17 February 2015 / Accepted: 18 April 2015 / Published online: 28 April 2015
© Springer Science+Business Media Dordrecht 2015

Abstract UNIO is a comprehensive software suite for protein NMR structure determination that enables full automation of all NMR data analysis steps involved—including signal identification in NMR spectra, sequence-specific backbone and side-chain resonance assignment, NOE assignment and structure calculation. Within the framework of the second round of the community-wide stringent blind NMR structure determination challenge (CASD-NMR 2), we participated in two categories of CASD-NMR 2, namely using either raw NMR spectra or unrefined NOE peak lists as input. A total of 15 resulting NMR structure bundles were submitted for 9 out of 10 blind protein targets. All submitted UNIO structures accurately coincided with the corresponding blind targets as documented by an average backbone root mean-square deviation to the reference proteins of only 1.2 Å. Also, the precision of the UNIO structure bundles was virtually identical to the ensemble of reference structures. By assessing the quality of all UNIO structures submitted to the two categories, we find throughout that only the UNIO–ATNOS/CANDID approach using raw NMR spectra consistently yielded structure bundles of high quality for direct deposition in the Protein Data Bank. In conclusion, the results obtained in CASD-NMR 2 are another vital proof for robust, accurate and unsupervised NMR data analysis by UNIO for real-world applications.

Keywords Nuclear magnetic resonance · Unsupervised data analysis · Protein structure determination · UNIO · ATNOS · CANDID · CASD-NMR

Introduction

Little more than a decade ago, protein NMR structure determination implied months, if not years of laborious, interactive work that required the expertise of a well-trained NMR analyst. Nowadays, owing to stunning advances in NMR experiments, instrumentation, and notably computational algorithms for NMR data analysis, the three-dimensional structure of a relatively propitious protein target may be determined in a few weeks. Despite this significant progress, the motivation remains high and ongoing to establish a general and robust protocol for NMR structure determination that is framed in man-hours, not man-weeks (Billeter et al. 2008; Williamson and Craven 2009). The ultimate aim of current research is to further promote NMR spectroscopy as a universal toolbox for the broader Structural Biology community (Wassenaar et al. 2012). So that it becomes possible even for newcomers to the NMR field to pursue protein structure determination with a minimum of training, by trustfully relying on accurate computational algorithms and protocols for unsupervised NMR data analysis and structure calculation.

A central and indispensable promoter for achieving this goal is the worldwide initiative ‘Critical assessment of automated structure determination of proteins from NMR data’, or CASD-NMR, that is a community-wide experiment for unsupervised protein structure determination based on NMR chemical shifts and/or NOESY data (Rosato et al. 2012, 2009). CASD-NMR offers an ideal platform for research groups both to objectively test their

✉ Torsten Herrmann
torsten.herrmann@ens-lyon.fr

¹ Institut des Sciences Analytiques, Centre de RMN à très Hauts Champs, Université de Lyon (UMR 5280 CNRS, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon 1), 5 rue de la Doua, 69100 Villeurbanne, France

methods and to initiate new developments. Notably, CASD-NMR provides an independent, impartial assessment of state-of-the-art computational approaches to the Structural Biology community and software users. The NMR software space is huge and continuously growing, populated by hundreds of approaches for automated data analysis proposed (Guerry and Herrmann 2011). Too easily, newcomers but also experienced researcher can get lost and deluded by hasty claims and conclusions. Hence, an independent and objective control authority such as provided by CASD-NMR is highly valuable for software developers and end-users alike (Rosato et al. 2009).

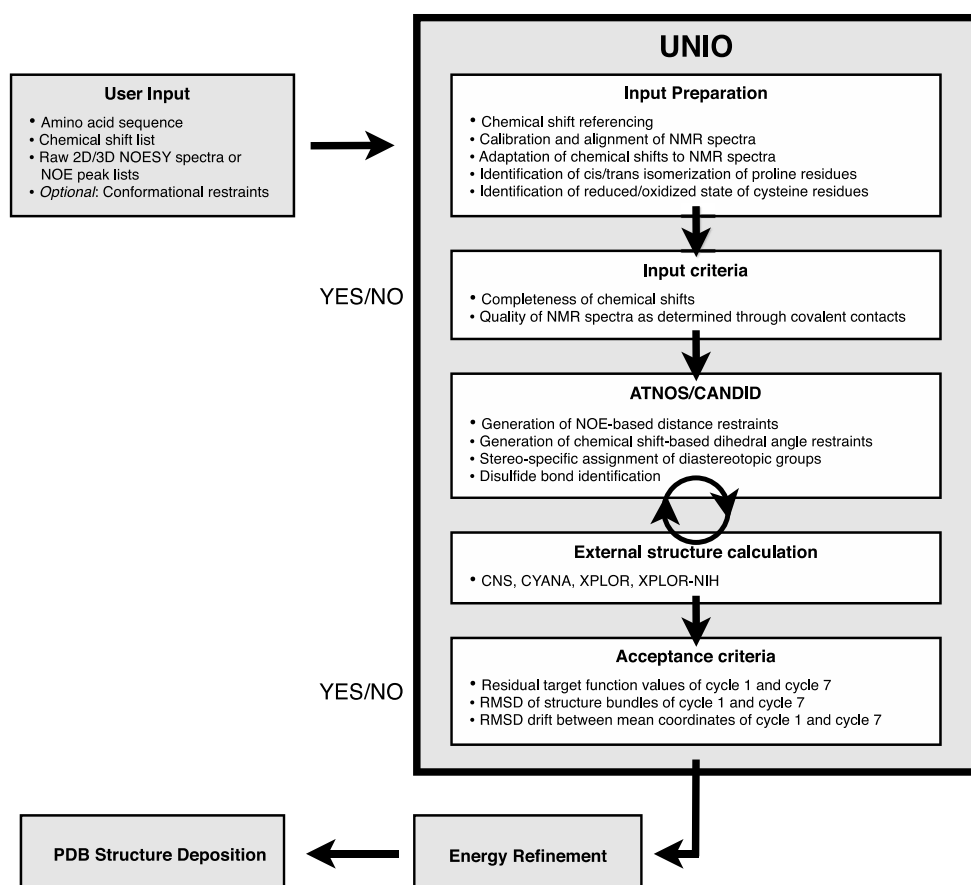
Here we report on the performance of the UNIO software for structure determination of blind protein targets released in the second round of the worldwide CASD-NMR experiment (CASD-NMR 2). UNIO is probably the most comprehensive NMR data analysis suite for protein NMR currently available to the Structural Biology community. More than a thousand of free-of-charge UNIO software licenses have been distributed to individual research groups and university computing centers worldwide. Since its first software release in 2008, UNIO has been used for hundreds of protein NMR structure determinations in the liquid- and solid-state state deposited in the Protein Data Bank. UNIO enables high to full automation of all data analysis steps involved—including signal identification in multi-dimensional NMR spectra, sequence-specific backbone and side-chain resonance assignment, NOE assignment and structure calculation (Guerry and Herrmann 2012). UNIO auxiliary algorithms for automated chemical shift referencing, automated alignment of NMR spectra and peak lists, automated adaptation of control parameters to input quality, automated RMSD evaluation for best structure superposition etc., are decisive UNIO components in order to guarantee proper daily laboratory operation. Also, thoroughly tested acceptance criteria for UNIO results enable the novice and experienced user to clearly distinguish between successful and doubtful structures.

Within the current scope of CASD-NMR 2, we exclusively focus on the final stage of the NMR structure determination process, namely the task of NOESY data analysis and structure calculation. We participated in two categories of CASD-NMR 2, using as input either raw NMR spectra or unrefined peak lists together with a list of nearly complete NMR chemical shifts. A third category of CASD-NMR 2 offering to start from manually refined peak lists appeared to us less attractive, so we didn't take part. This was so for two main reasons. First, UNIO had already flawlessly succeeded in determining all blind protein targets based on refined peak lists in the first round of CASD-NMR (Rosato et al. 2012). Secondly, the preparation of refined peak lists is quite cumbersome and usually requires multiple rounds of interactive, subjective NMR data

refinement. Hence, the use of refined NOE data contradicts in a certain sense the advanced spirit of CASD-NMR 2 compared to CASD-NMR 1 for assessing the robustness of unsupervised procedures confronted with imperfect and thus realistic input data. In UNIO, all major and auxiliary data analysis techniques were developed for proper performance with raw, imperfect NMR spectra, so to simultaneously guarantee efficiency, objectivity and operability even by relatively inexperienced users in daily practice.

The driving force for developments in UNIO is the motivation to provide a single computational framework for determining protein structures at atomic resolution without any or minor user intervention. To achieve this goal for unsupervised NOESY interpretation and structure calculation, numerous original concepts for NMR data analysis were developed for the UNIO-ATNOS/CANDID approach over now more than a decade, aiming at enabling advanced, unattended protein studies by solution and solid-state NMR alike (Herrmann et al. 2002a, 2002b; Knight et al. 2012; 2011; Manolikas et al. 2008). The UNIO-ATNOS/CANDID workflow is depicted in Fig. 1 and illustrates the major and auxiliary building blocks used for achieving unattended NOESY analysis. Many of these numerical techniques were adopted by other research groups and are nowadays employed by popular software programs (Lee et al. 2011; Rieping et al. 2007; Zhang et al. 2014). The key elements of UNIO-ATNOS (Herrmann et al. 2002b) for NOESY spectral analysis are local baseline correction and evaluation of local noise level amplitudes, determination of spectrum-specific threshold parameters, the use of spectral symmetry relations, chemical shift adaptation, and the incorporation of chemical shift information and intermediate protein structures into the process of NMR signal identification (peak picking). The key techniques of UNIO-CANDID (Herrmann et al. 2002a) for NOE assignment and structure calculation are network-anchored assignment, ambiguous distance restraints (ODonoghue et al. 1996), structure-guided calibration of NMR peak volumes, and distance restraint combination. The full UNIO-ATNOS/CANDID approach proceeds, as all commonly used NOE assignment algorithms, in iterative cycles, each consisting of exhaustive NOE signal identification and, in part, ambiguous NOE assignments followed by a structure calculation (Güntert 2003). In contrast to the predominant majority of other NOE assignment approaches that operate on invariant lists of peak positions and chemical shifts, the combined use of UNIO-ATNOS NOESY signal identification and UNIO-CANDID NOE assignment waives the common requirement for performing multiple rounds of calculations with gradually (manual) refined input data. This fundamental conceptual difference of the UNIO-ATNOS/CANDID procedure compared to other approaches leads to the

Fig. 1 Flowchart of unsupervised protein NMR structure determination using UNIO–ATNOS/CANDID



decisive advantage of significantly increased efficiency, objectivity and reproducibility of protein NMR structure determination.

In the following, we describe some recent developments and present the UNIO results for all 15 structure bundles submitted to the aforementioned two categories of CASD-NMR 2. We list input quality criteria for proper performance of UNIO–ATNOS/CANDID, and practical and reliable guidelines are given for judging the correctness of the resulting protein structures.

Materials and methods

UNIO protocol

For all blind protein targets, NOESY data analysis was performed with the modules ATNOS and/or CANDID incorporated into the software platform UNIO (Guerry and Herrmann 2012), using as input either raw NOESY spectra or unrefined peak lists together with a list of NMR chemical shifts and the amino acid sequence of the target protein (Fig. 1). For all target proteins, the experimental NMR input data consisted of a 3D ^{15}N -resolved (^1H , ^1H)-

NOESY and two 3D ^{13}C -resolved (^1H , ^1H)-NOESY with the carrier frequency in the aliphatic or aromatic region, respectively. The standard UNIO protocol was employed that consisted of seven cycles of concert NOESY peak identification, NOE assignment and structure calculation. Each cycle comprised automated NOESY peak picking with ATNOS (Herrmann et al. 2002b), use of the resulting lists of peak positions and intensities as input for automated CANDID NOE assignment (Herrmann et al. 2002a), and use of the final set of meaningful, non-redundant NOE distance restraints from CANDID as input for structure calculation by simulated annealing using a suitable external program (Fig. 1). At the outset of the spectral analysis, UNIO used highly permissive criteria to identify and assign a comprehensive set of peaks in the NOESY spectra or the unassigned peak lists provided. Only the knowledge of the covalent polypeptide structure and the chemical shifts were initially exploited to guide NOE cross peak identification and NOE assignment. In the second and subsequent UNIO–ATNOS/CANDID cycles, the intermediate protein three-dimensional structures were used as an additional guide for the interpretation of the NOESY spectra or unassigned input peak lists. Since the precision of the protein structure models normally improves with each subsequent cycle, the

criteria for accepting NMR cross peaks and NOE assignments were successively tightened during the iterations. In each UNIO–ATNOS/CANDID cycle, the output consisted of an updated list of assigned NOE cross peaks for each input spectrum and a final set of meaningful upper limit distance restraints which constituted the input for the torsion angle dynamics algorithm of CYANA for three-dimensional (3D) structure calculation (Güntert et al. 1997). In addition, torsion angle restraints for the backbone dihedral angles ϕ and ψ derived from all backbone chemical shifts were automatically generated by UNIO (Herrmann et al., to be published) and added to the input for each cycle of structure calculation. During the first six UNIO–ATNOS/CANDID cycles, ambiguous distance restraints (ODonoghue et al. 1996) were used. For the final structure calculation in cycle 7, only distance restraints were retained by UNIO that could be unambiguously assigned based on the protein three-dimensional structure from cycle 6. Residual dipolar coupling data were used where available. The computation time for all 15 structures bundles submitted to CASD-NMR 2 was in the range of only 1.0–2.5 h on a single 2.4 GHz Intel processor and was spent predominantly with CYANA structure calculation (approximately 80–90 % of the total CPU time).

The 20 conformers with the lowest residual CYANA target function values obtained from cycle 7 were energy-refined in a water shell with the program OPALp (Koradi et al. 2000; Luginbuhl et al. 1996) using the AMBER force field (Ponder and Case 2003).

Input criteria for proper performance of UNIO–ATNOS/CANDID

For accessing proper performance and enabling structure validation, the following two input criteria should be fulfilled: (1) the input chemical shift list should contain more than 90 % of the non-labile and backbone amide ^1H chemical shifts. If 3D heteronuclear-resolved NOESY are used, more than 90 % of the ^{15}N and/or ^{13}C chemical shifts must be available. (2) UNIO–ATNOS should validate NOE signals for at least 85 % of all pairwise combinations of protons for which sequence-specific NMR assignments are available, and which have covalent structure-imposed upper distance limits shorter than 5 Å.

Acceptance criteria for successful UNIO–ATNOS/CANDID calculations

The following three acceptance criteria have to be met for validation of the resulting structure: (1) The average final target function value from the first UNIO–ATNOS/CANDID cycle should be below 250 \AA^2 , and the corresponding value for the last UNIO–ATNOS/CANDID cycle should be

below 10 \AA^2 , with more than 80 % of all picked NOESY cross peaks assigned and less than 20 % of the peaks with exclusively long-range assignments eliminated by the filtering step applied in UNIO–CANDID. (2) The average backbone RMSD to the mean coordinates for the structured parts of the polypeptide chain should be below 3 Å for the bundle of conformers used to represent the protein structure from the first UNIO–ATNOS/CANDID cycle. (3) The RMSD drift between the mean atom coordinates after the first and the last UNIO–ATNOS/CANDID cycles calculated for the backbone heavy atoms of the structured part of the polypeptide chain should be smaller than 3 Å.

Results

Standard protocol for UNIO–ATNOS/CANDID and recent developments

All 15 protein structures submitted to the two CASD-NMR categories (Table 1) were calculated using the standard UNIO protocol (see “Materials and methods”) that is part of the UNIO distribution and employs a single set of control parameters used for NOESY signal identification, NOE assignment and structure calculation. At the outset of a UNIO run, all control parameters are initialized to their default values and are thus accessible to any software user (Guerry and Herrmann 2012). Over more than a decade, these default values have continuously been optimized by applications to hundreds of protein projects and thanks to valuable user feedback. Nowadays, the standard UNIO protocol offers a balanced, general robustness for protein studies and is well capable to cope with different input NMR data sets and data quality, as also documented by the UNIO results in CASD-NMR 2 (see below). A non-exhaustive search of PDB-deposited solution and solid-state NMR structures revealed that so far the UNIO–ATNOS/CANDID approach led to hundreds of PDB depositions of proteins and protein complexes, belonging to various topology classes and with a molecular weight up to 28 kDa.

Latest developments of UNIO–ATNOS/CANDID mainly aimed at minor improvements of the overall procedure. Previously, information about the cis/trans isomerization of proline and the reduced or oxidized state of cysteine residues needed to be provided as additional user input. The current UNIO version automatically checks the correctness of the user input for these two amino acids and modifies the input, if necessary (Fadel et al. 2005). Other more significant improvements concerned automatic generation of torsion angle restraints for the backbone dihedral angles ϕ and ψ derived from all backbone NMR chemical shifts; automatic referencing of all backbone NMR

Table 1 Blind protein targets of CASD-NMR Round 2

Blind target	PDB ID	Length	Topology	State	Sample concentration (mM)
HR2876C	2M5O	97	α - β	Monomer	1.0
HR8254A	2M2E	73	α	Monomer	0.5
YR313A	2LTL	119	α - β	Monomer	1.0
HR2876B	2LTM	107	α - β	Monomer	0.9
OR135	2LN3	83	α - β	Monomer	1.0
OR36	2LCI	134	α - β	Monomer	0.7
HR5460A	2LAH	160	α	Monomer	1.2
HR6430A	2LA6	99	α - β	Monomer	0.8
HR6470A	2L9R	69	α	Monomer	0.7

chemical shifts; automatic stereospecific assignment of methylene or isopropyl groups in concert with NOE assignment; automatic determination of residue ranges for optimal superposition of NMR structure bundles (Fig. 1). All these new routines in UNIO will be described in detail elsewhere (Herrmann et al., to be published). The majority of blind targets were determined with this latest UNIO version.

UNIO results for 9 blind protein targets

We participated in two categories of CASD-NMR 2, namely using either raw NMR spectra or unrefined NOE peak lists as input. A total of 15 resulting NMR structure bundles were submitted for 9 out of 10 blind protein targets. All submitted UNIO structures closely coincided with the corresponding blind targets as documented by an average backbone RMSD to the reference (RMSD bias) of only 1.2 Å. Numerical values of the backbone RMSD bias of UNIO structures to the blind targets calculated for all backbone heavy atoms of well-defined polypeptide regions in the reference structures are listed in Table 2. UNIO yielded highly accurate structures (RMSD bias <1.5 Å) for 13 of 15 data sets submitted (Table 1). Notably, the large majority of these 13 structures (8 out of 13 UNIO structures) showed even a remarkably small RMSD bias <1.0 Å. One data set resulted in an accurate structure with a slightly increased RMSD bias of 2.07 Å. A single UNIO structure (blind target YR313A) showed a distorted local conformation, but only for a short polypeptide segment of seven residues. Importantly, this region of the UNIO structure remained largely undefined and had not converged into a wrong, precisely defined local fold. By exclusion of this structural region, the value of the RMSD bias for the blind target YR313A drops to 1.48 Å (see Table 2) and becomes so in perfect agreement with the excellent results obtained for all other data sets. Correspondently, the RMSD of the structure bundle for YR313A decreases from 1.26 to 0.97 Å, proving that this seven

residues comprising polypeptide segment in the UNIO structure was largely disordered as stated above.

All UNIO and corresponding reference structures are also closely similar in terms of precision of the atomic coordinates as shown by the RMSD values of the structure bundles in Table 2. In summary, the nearly perfect agreement of the unsupervised UNIO structures with the blind targets in terms of both accuracy and precision of atom positions is evident also by visual inspection of the 3D models. Structural deviations can almost exclusively be seen for surface loop regions of the different protein targets. All over, unsupervised UNIO data analysis is well capable to yield resulting NMR structures with a quality both in terms of accuracy and precision at least equal to the structures obtained by the tedious interactive approach, but notably using only a fraction of the manually invested man-power, and moreover providing fully objective NMR data analysis as guaranteed by direct operation on the raw NMR spectra.

Quality of UNIO structures

A NMR structure can be defined as the ensemble of conformers that simultaneously fulfill all experimentally derived conformational restraints. In this context, an important criterion for evaluating the quality of NMR structures can be assessed by detailed restraint analysis (Montelione et al. 2013). The residual target function, used as hybrid energy potential during simulated annealing, strongly penalizes restraint violations and is thus a reliable measure for the consistency between conformational restraints and calculated NMR conformers. The target function values of all UNIO structures submitted to the two CASD-NMR categories are summarized in Table 2. Although the precision and accuracy of UNIO structures for both classes of input data are quite comparable, we detected throughout that the target function values for structures calculated from invariant unrefined peak lists are always higher than for structures determined from raw NMR spectra (Table 2). Apparently, the iterative UNIO-

Table 2 UNIO results for CASD-NMR round 2 targets

Blind target	Length	RMSD range	RMSD PDB (Å) ^a	Raw NMR spectra			Unrefined peak lists		
				TF (Å ²) ^b	RMSD (Å)	RMSD Bias (Å) ^c	TF (Å ²) ^b	RMSD (Å)	RMSD Bias (Å) ^c
HR2876C	97	17–92	0.38	4.89	0.44	1.20	10.88	0.36	1.02
HR8254A	73	6–60	0.51	–	–	–	1.37	0.70	1.30
YR313A	119	17–111	0.68	5.74	1.26	2.98	–	–	–
		17–37, 45–111		5.74	0.97	1.48	–	–	–
HR2876B	107	12–105	0.44	15.41	0.41	0.87	30.85	0.32	0.89
OR135	83	4–74	0.45	1.37	0.39	0.81	2.35	0.43	0.87
OR36	134	1–128	0.66	5.33	0.92	1.37	19.23	0.60	1.34
HR5460A	160	13–158	0.53	5.80	1.09	2.07	–	–	–
HR6430A	99	13–99	0.37	5.27	0.52	0.95	6.74	0.52	0.68
HR6470A	69	11–58	0.33	0.71	0.47	1.12	0.60	0.42	0.63
Average	104.6	–	0.48	5.56	0.65	1.23	10.29	0.48	0.96

^a Values for the RMSD of the reference NMR structure bundle deposited in the protein data bank are given. The RMSD is the average of the RMSD values between the individual conformers in the bundle and their mean coordinates for the backbone atoms N, C α and C' of residues in the well-defined regions of the protein

^b The residual UNIO target function value is the average for the bundles of conformers representing the NMR structure. The target function values before energy minimization are given

^c The RMSD bias is calculated as the RMSD between the mean coordinates of the UNIO and the reference structure bundles

ATNOS/CANDID approach yielded a more self-consistent analysis of NOESY data, mainly by avoiding over-restraining of local protein structure caused by data overfitting. This is an attractive observation and the key strength of UNIO–ATNOS/CANDID compared to UNIO–CANDID, and so also to all other popular approaches. In practice, programs operating on invariant input peak lists are typically executed not only once, but require several rounds of automatic NOE assignment and manual refinement of the input peak lists in order to achieve high quality NMR structures. Automatically generated unrefined peak list might facilitate the start of a structure determination project, but can hardly remove entirely the subsequent need for tedious, interactive editing of NOE peak lists. Hence, by assessing the quality of all UNIO structures submitted to the two categories, we find that only the UNIO–ATNOS/CANDID approach using raw NMR spectra consistently yielded structure bundles of sufficient quality for direct deposition in the Protein Data Bank.

Input and acceptance criteria for successful UNIO calculations

It is central for the operation of unsupervised approaches to define minimal input quality requirements for guaranteeing proper software performance and reliable criteria for discriminating between successful and failed runs. Here, we represent the opinion that it is better to give rather too strict than too loose guidelines to the UNIO end-users. A few

runs erroneously flag as failed are acceptable, while acceptance criteria become completely useless if they are not able to detect all wrong structures.

The UNIO input requirements (see materials and methods) for proper software performance are consciously set too strict. High quality of the NOESY spectra and accurate calibration of the input chemical shifts to the NOESY spectra are imposed. From our experience, UNIO also yields correct structures with input data that does not meet—in reasonable limits—these entry requirements. However, a low percentage of validated NOE cross peaks typically results when the signal-to-noise ratio is too poor for automated spectral analysis, or the input chemical shifts are not well-calibrated to the NOESY spectra. In this situation, the input data need to be critically reevaluated before attempting a new automated NOESY interpretation. In particular the adaptation of the chemical shifts to the NOESY spectra needs to be improved.

The three UNIO acceptance criteria emphasize the crucial importance of getting the correct protein fold already after the first iteration cycle. For reliable automated NOESY analysis, the initial 3D structure obtained should be reasonably compatible with the input data and show a defined fold of the protein. Structural changes between the first and subsequent UNIO–ATNOS/CANDID cycles should only occur within the conformation space determined by the initial bundle of conformers obtained after UNIO–ATNOS/CANDID cycle 1. As for the input requirements, UNIO runs that show slight and reasonable

deviations from these output criteria can usually still be considered as success.

Regarding the CASD-NMR targets, 12 out of 15 resulting structures passed all input quality and acceptance criteria (see “Materials and methods”). For three data sets, always the same of the three UNIO output criteria was violated, namely an increased value of the final target function was detected. As discussed above, these runs were indeed a bit problematic in terms of restraint violations. Since all other input and output criteria were fulfilled, it was obvious to us prior to the public release of the reference structure that the resulting UNIO structures had nonetheless converged into an overall correct protein fold. In summary, the UNIO guidelines provide an effective and informative multi-pass filter that is able to reliably label successful calculations, to detect slightly problematic runs, and to clearly spot erroneous results.

Discussions and conclusions

We presented the results of the unsupervised UNIO procedure for the second round of the community-wide blind NMR structure determination challenge (CASD-NMR 2). All 15 submitted UNIO structures for the two classes of input data showed excellent agreement to the reference structures in terms of both precision and accuracy of the atomic coordinates. The performance of UNIO for the nine blind targets clearly demonstrated that our unsupervised procedure was well able to cope with various NMR data quality and protein topologies. We found that the UNIO-ATNOS/CANDID approach that enables direct feedback between NMR spectra, NOE assignment and protein structure, was superior to the alternatively used UNIO-CANDID approach that exclusively operated on invariant peak lists.

The findings presented here are in line with those of the UNIO-ATNOS/CANDID approach used within the J-UNIO protocol (Dutta et al. 2015; Serrano et al. 2012) for extensive automation of NMR structure determination that also includes unsupervised algorithms for the preceding data analysis steps for obtaining sequence-specific resonance backbone and side-chain assignment using UNIO-MATCH (Volk et al. 2008) and UNIO-ATNOS/ASCAN (Fiorito et al. 2008), respectively. Successful and routine application of the J-UNIO protocol to more than 50 de novo protein targets within the Joint Center for Structural Genomics (JCSG) through the NIH Protein Structure Initiative (PSI) showed that unsupervised NOESY analysis is well feasible and routinely used for studies of protein and protein complexes with more than 200 residues in size (Jaudzems et al. 2015). In conclusion, the results obtained in CASD-NMR 2 are another vital proof for robust,

accurate and unsupervised NMR data analysis by UNIO for real-world applications.

References

- Billeter M, Wagner G, Wüthrich K (2008) Solution NMR structure determination of proteins revisited. *J Biomol NMR* 42:155–158
- Dutta SK, Serrano P, Proudfoot A, Geralt M, Pedrini B, Herrmann T, Wüthrich K (2015) APSY-NMR for protein backbone assignment in high-throughput structural biology. *J Biomol NMR* 61:47–53. doi:10.1007/s10858-014-9881-8
- Fadel V, Bettendorff P, Herrmann T, de Azevedo WF Jr, Oliveira EB, Yamane T, Wüthrich K (2005) Automated NMR structure determination and disulfide bond identification of the myotoxin crostamine from *Crotalus durissus terrificus*. *Toxicol* 46:759–767. doi:10.1016/j.toxicol.2005.07.018
- Fiorito F, Herrmann T, Damberger FF, Wüthrich K (2008) Automated amino acid side-chain NMR assignment of proteins using C-13- and N-15-resolved 3D [H-1, H-1]-NOESY. *J Biomol NMR* 42:23–33
- Guerry P, Herrmann T (2011) Advances in automated NMR protein structure determination. *Q Rev Biophys* 44:257–309
- Guerry P, Herrmann T (2012) Comprehensive automation for NMR structure determination of proteins. *Methods Mol Biol* 831:429–451. doi:10.1007/978-1-61779-480-3_22
- Güntert P (2003) Automated NMR protein structure calculation. *Prog Nucl Magn Spectrosc* 43:105–125
- 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 (2002a) 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
- Herrmann T, Güntert P, Wüthrich K (2002b) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 24:171–189
- Jaudzems K, Pedrini B, Geralt M, Serrano P, Wüthrich K (2015) J-UNIO protocol used for NMR structure determination of the 206-residue protein NP_346487.1 from streptococcus pneumoniae TIGR4. *J Biomol NMR* 61:65–72. doi:10.1007/s10858-014-9886-3
- Knight MJ, Webber AL, Pell AJ, Guerry P, Barbet-Massin E, Bertini I, Felli IC, Gonnelli L, Pierattelli R, Emsley L, Lesage A, Herrmann T, Pintacuda G (2011) Fast resonance assignment and fold determination of human superoxide dismutase by high-resolution proton-detected solid-state mas nmr spectroscopy. *Angew Chem Int Edit* 50:11697–11701
- Knight MJ, Pell AJ, Bertini I, Felli IC, Gonnelli L, Pierattelli R, Herrmann T, Emsley L, Pintacuda G (2012) Structure and backbone dynamics of a microcrystalline metalloprotein by solid-state NMR. *Proc Natl Acad Sci USA* 109:11095–11100
- Koradi R, Billeter M, Güntert P (2000) Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput Phys Commun* 124:139–147
- Lee W, Kim JH, Westler WM, Markley JL (2011) PONDEROSA, an automated 3D-NOESY peak picking program, enables automated protein structure determination. *Bioinformatics* 27:1727–1728
- Luginbuhl P, Güntert P, Billeter M, Wüthrich K (1996) The new program OPAL for molecular dynamics simulations and energy

- refinements of biological macromolecules. *J Biomol NMR* 8:136–146
- Manolikas T, Herrmann T, Meier BH (2008) Protein structure determination from C-13 spin-diffusion solid-state NMR spectroscopy. *J Am Chem Soc* 130:3959–3966
- Montelione GT, Nilges M, Bax A, Güntert P, Herrmann T, Richardson JS, Schwieters CD, Vranken WF, Vuister GW, Wishart DS, Berman HM, Kleywegt GJ, Markley JL (2013) Recommendations of the wwPDB NMR validation task force. *Structure* 21:1563–1570
- ODonoghue SI, King GF, Nilges M (1996) Calculation of symmetric multimer structures from NMR data using a priori knowledge of the monomer structure, co-monomer restraints, and interface mapping: the case of leucine zippers. *J Biomol NMR* 8:193–206
- Ponder JW, Case DA (2003) Force fields for protein simulations. *Adv Protein Chem* 66:27–85
- Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23:381–382
- Rosato A, Bagaria A, Baker D, Bardiaux B, Cavalli A, Dorelejers JF, Giachetti A, Guerry P, Güntert P, Herrmann T, Huang YJ, Jonker HRA, Mao B, Malliavin TE, Montelione GT, Nilges M, Raman S, van der Schot G, Vranken WF, Vuister GW, Bonvin AMJJ (2009) CASD-NMR: critical assessment of automated structure determination by NMR. *Nat Methods* 6:625–626
- Rosato A, Aramini JM, Arrowsmith C, Bagaria A, Baker D, Cavalli A, Dorelejers JF, Eletsky A, Giachetti A, Guerry P, Gutmanas A, Güntert P, He YF, Herrmann T, Huang YPJ, Jaravine V, Jonker HRA, Kennedy MA, Lange OF, Liu GH, Malliavin TE, Mani R, Mao BC, Montelione GT, Nilges M, Rossi P, van der Schot G, Schwalbe H, Szyperski TA, Vendruscolo M, Vernon R, Vranken WF, de Vries S, Vuister GW, Wu B, Yang YH, Bonvin AMJJ (2012) Blind testing of routine fully automated determination of protein structures from NMR data. *Structure* 20:227–236
- Serrano P, Pedrini B, Mohanty B, Geralt M, Herrmann T, Wüthrich K (2012) The J-UNIO protocol for automated protein structure determination by NMR in solution. *J Biomol NMR* 53:341–354
- Volk J, Herrmann T, Wüthrich K (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J Biomol NMR* 41:127–138
- Wassenaar TA, van Dijk M, Loureiro-Ferreira N, van der Schot G, de Vries SJ, Schmitz C, van der Zwan J, Boelens R, Giachetti A, Ferella L, Rosato A, Bertini I, Herrmann T, Jonker HRA, Bagaria A, Jaravine V, Güntert P, Schwalbe H, Vranken WF, Dorelejers JF, Vriend G, Vuister GW, Franke D, Kikhney A, Svergun DI, Fogh RH, Ionides J, Laue ED, Spronk C, Jurksa S, Verlato M, Badoer S, Dal Pra S, Mazzucato M, Frizziero E, Bonvin AMJJ (2012) WeNMR: Structural biology on the grid *J Grid Comput* 10:743–767 doi: [10.1007/s10723-012-9246-z](https://doi.org/10.1007/s10723-012-9246-z)
- Williamson MP, Craven CJ (2009) Automated protein structure calculation from NMR data. *J Biomol NMR* 43:131–143
- Zhang ZY, Porter J, Tripsianes K, Lange OF (2014) Robust and highly accurate automatic NOESY assignment and structure determination with Rosetta. *J Biomol NMR* 59:135–145