CrossMark

EDITORIAL

Automated protein structure determination by NMR

Antonio Rosato[1] · Martin Billeter[2]

The current approaches to protein structure determination from NMR data adopted in many laboratories all around the world crucially depend on the determination of a large number of upper distance limits for proton–proton pairs. These distance restraints are typically generated through a labor-intensive manual analysis of several NMR NOESY spectra, involving many cycles of NOESY peak assignment and structure generation. Some regions of the structure may prove more difficult than others to characterize, sometimes forcing researchers to spend a significant amount of time just to evaluate all possible peaks for just a handful of amino acids. This entails taking decisions on individual ambiguous peaks that are typically driven by the spectroscopist's experience rather than by objective reasons. To reduce this burden and improve reproducibility, significant efforts have been and are currently being made to fully automate the entire process of structure determination, from analysis of the spectra to structure calculations. The goal of the present special issue of the Journal of Biomolecular NMR is to provide an up-to-date overview of the available methods for automated assignment of NOESY spectra and structure generation, including comprehensive tests on the reliability and success of these approaches, and comparisons among them.

Critical assessment of automated structure determination by NMR (CASD-NMR) is an international collaboration for the evaluation and comparison of different automated software tools for NMR determination of protein structures using the same experimental data sets. The concept of the CASD-NMR initiative was designed in 2008–2009, as part of the activities of the e-NMR/WeNMR electronic infrastructure (Wassenaar et al. 2012), funded by the European Commission through grants 213010 and 261572. CASD-NMR aimed to contribute to advance the methodological development in the field, with the overarching goal to foster the adoption of automated methods throughout the biomolecular NMR community by demonstrating their reliability and the high quality of their results.

As a first proof-of-concept of CASD-NMR, various research groups historically involved in the development of automated structure determination methods engaged in the recalculation of a few NMR structures already deposited in the PDB by some of the participating partners of WeNMR. This experience was successful and spurred enthusiasm among the participants. Therefore, we joined forces with the NorthEast Structural Genomics Consortium (NESG) to implement CASD-NMR in the form that produced the results recapitulated by the contributions included in the present special issue of the Journal of Biomolecular NMR. In short, the key points of the initiative have been:

- The use of blind targets, i.e. of proteins for which a structure was manually solved by NESG researchers but not made publicly available (this design is intentionally very similar to that of CASP);
- The use of real experimental NOESY data, acquired at NESG centers and distributed to all participants via the WeNMR website;
- The involvement of an external team with extensive expertise in validation of NMR structures to analyze the results generated by the participants for each target.

✉ Antonio Rosato
  rosato@cerm.unifi.it

[1] Magnetic Resonance Center (CERM), University of Florence, Florence, Italy

[2] Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden

In the first round of CASD-NMR (see later), the available data consisted only of chemical shift assignments and unassigned, manually curated ("refined") NOESY peak lists. Manual intervention was limited to the removal of noise peaks. In the present second round of CASD-NMR, the focus was on the use of automatically picked peak lists without any manual intervention ("unrefined"). Four weeks after the release of the unrefined lists, refined lists were provided for each target for a second, separate calculation. In addition, raw spectral data were available to allow the participating software to use their own peak picking algorithms. CASD-NMR is in fact unique among conceptually similar community efforts because it uses real experimental data before any related publication is available in the literature. Packaging such data in a format that could be easily distributed to and readily exploited by the participants entailed a dedicated effort by data "providers" at NESG. This aspect of the CASD-NMR design ensured that all automated software used exactly the same input data, consisting of chemical shift assignments, unassigned peak lists and raw spectral data. The very same data were also used to generate the corresponding structure through conventional manual methods. This had a two-fold implication: (1) participants were guaranteed that each dataset they received was indeed suitable for high quality structure determination; (2) for each target, the comparative assessment of results was not affected by possible differences in the quality of input data.

The CASD-NMR manifesto was published in Nature Methods in 2009 (Rosato et al. 2009). In the first round of CASD-NMR, seven different teams provided structures for 10 targets. The corresponding results demonstrated that automated methods could use a manually curated peak list to consistently generate a structure with a fold very close (backbone RMSD <2 Å) to the reference manual structure (Rosato et al. 2012). A second round of CASD-NMR was initiated shortly after, which involved a larger number of software developers (eventually as many as 11, although some joined CASD-NMR only by the end of the second round), including also teams from the neighboring field of structural bioinformatics. It is thus fair to say that a first achievement of CASD-NMR has been to boost the visibility of automated NMR methods for protein structure determination also beyond the inner circle of biomolecular NMR scientists. Furthermore, CASD-NMR fuelled significant improvement of such methods, by providing an opportunity for researchers in the field to truly appreciate

the weaknesses and strengths of their protocols with respect to other approaches, as well as to identify common bottlenecks across the field. The purpose of the present Special Issue is to allow CASD-NMR participants not only to show their results for the new 10 targets of the second round, but also to recapitulate the evolution of their software within and beyond the whole of CASD-NMR. In addition to the contributions specifically dedicated to the achievement of each software, which altogether provide an extensive view of where automation in NMR structure determination stands, two articles focus respectively on the performance achieved for the targets included in the second round of CASD-NMR and on the validation of the structures submitted by the CASD-NMR participants. The experimental NMR data for all 20 targets in the two rounds of CASD-NMR are publicly available (see Rosato et al. 2015 for details).

References

Rosato A, Bagaria A, Baker D, Bardiaux B, Cavalli A, Doreleijers 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. doi:10.1038/nmeth0909-625

Rosato A, Aramini JM, Arrowsmith C, Bagaria A, Baker D, Cavalli A, Doreleijers JF, Eletsky A, Giachetti A, Guerry P, Gutmanas A, Güntert P, He Y, Herrmann T, Huang YJ, Jaravine V, Jonker HRA, Kennedy MA, Lange OF, Liu G, Malliavin TE, Mani R, Mao B, 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 Y, Bonvin AMJJ (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 20:227–236. doi:10.1016/j.str.2012.01.002

Rosato A, Vranken W, Fogh RH, Ragan TJ, Tejero R, Pederson K, Lee H-W, Prestegard JH, Yee A, Wu B, Lemak A, Houliston S, Arrowsmith CH, Kennedy M, Acton TB, Xiao R, Liu G, Montelione GT, Vuister GW (2015) J Biomol NMR. doi:10.1007/s10858-015-9953-4

Wassenaar TA, Dijk M, Loureiro-Ferreira N, Schot G, Vries SJ, Schmitz C, 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, Doreleijers JF, Vriend G, Vuister GW, Franke D, Kikhney A, Svergun DI, Fogh RH, Ionides J, Laue ED, Spronk CAEM, Jurkša 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