

## CASD-NMR: Critical Assessment of Automated Structure Determination by NMR

Rutgers University has made this article freely available. Please share how this access benefits you.

Your story matters. [\[https://rucore.libraries.rutgers.edu/rutgers-lib/47166/story/\]](https://rucore.libraries.rutgers.edu/rutgers-lib/47166/story/)

This work is an **ACCEPTED MANUSCRIPT (AM)**

This is the author's manuscript for a work that has been accepted for publication. Changes resulting from the publishing process, such as copyediting, final layout, and pagination, may not be reflected in this document. The publisher takes permanent responsibility for the work. Content and layout follow publisher's submission requirements.

Citation for this version and the definitive version are shown below.

**Citation to Publisher** Rosato, Antonio, Bagaria, Anurag, Baker, David, Bardiaux, Benjamin, Cavalli, Andrea, Doreleijers, Jurgen F., Giachetti, Andrea, Guerry, Paul, Güntert, Peter, Herrmann, Torsten, Huang, Yuanpeng J., Jonker, Hendrik R. A., Mao, Binchen, Malliavin, Thérèse E., Montelione, Gaetano T., Nilges, Michael, Raman, Srivatsan, van der Schot, Gijs, Vranken, Wim F., Vuister, Geerten W. & Bonvin, Alexandre M. J. J. (2009). CASD-NMR: Critical Assessment of Automated Structure Determination by NMR. *Nature Methods* 6, 625-626. <https://dx.doi.org/10.1038/nmeth0909-625>.

**Citation to this Version:** Rosato, Antonio, Bagaria, Anurag, Baker, David, Bardiaux, Benjamin, Cavalli, Andrea, Doreleijers, Jurgen F., Giachetti, Andrea, Guerry, Paul, Güntert, Peter, Herrmann, Torsten, Huang, Yuanpeng J., Jonker, Hendrik R. A., Mao, Binchen, Malliavin, Thérèse E., Montelione, Gaetano T., Nilges, Michael, Raman, Srivatsan, van der Schot, Gijs, Vranken, Wim F., Vuister, Geerten W. & Bonvin, Alexandre M. J. J. (2009). CASD-NMR: Critical Assessment of Automated Structure Determination by NMR. *Nature Methods* 6, 625-626. Retrieved from [doi:10.7282/T32Z17B4](https://doi.org/10.7282/T32Z17B4).

**Terms of Use:** Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

*Article begins on next page*

## CASD-NMR: Critical Assessment of Automated Structure Determination by NMR

Antonio Rosato<sup>1,2,\*</sup>, Anurag Bagaria<sup>3,4</sup>, David Baker<sup>5</sup>, Benjamin Bardiaux<sup>6</sup>, Andrea Cavalli<sup>7</sup>, Jurgen F. Doreleijers<sup>8</sup>, Andrea Giachetti<sup>1</sup>, Paul Guerry<sup>9</sup>, Peter Güntert<sup>3,4</sup>, Torsten Herrmann<sup>9</sup>, Yuanpeng J. Huang<sup>10</sup>, Hendrik R. A. Jonker<sup>4,11</sup>, Binchen Mao<sup>10</sup>, Thérèse E. Malliavin<sup>6</sup>, Gaetano T. Montelione<sup>10</sup>, Michael Nilges<sup>6</sup>, Srivatsan Raman<sup>5</sup>, Gijs van der Schot<sup>12</sup>, Wim F. Vranken<sup>13</sup>, Geerten W. Vuister<sup>8</sup>, Alexandre M.J.J. Bonvin<sup>12</sup>

<sup>1</sup> Magnetic Resonance Center, University of Florence, 50019 Sesto Fiorentino, Italy.

<sup>2</sup> Department of Chemistry, University of Florence, 50019 Sesto Fiorentino, Italy.

<sup>3</sup> Institute of Biophysical Chemistry and Frankfurt Institute for Advanced Studies, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany.

<sup>4</sup> Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany.

<sup>5</sup> Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.

<sup>6</sup> Unité de Bioinformatique Structurale, Institut Pasteur, CNRS URA 2185, Paris, France.

<sup>7</sup> Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, United Kingdom.

<sup>8</sup> Protein Biophysics, Institute of Molecules and Materials, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands.

<sup>9</sup> Centre de RMN à très Hauts Champs, Université de Lyon, FRE-3008 CNRS, ENS Lyon, UCB Lyon 1, 69100 Villeurbanne, France.

<sup>10</sup> Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA.

<sup>11</sup> Institute of Organic Chemistry and Chemical Biology, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany.

<sup>12</sup> Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, 3584 CH Utrecht, the Netherlands.

<sup>13</sup> European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, United Kingdom.

Correspondence to:

Dr. Antonio Rosato, [Rosato@cerm.unifi.it](mailto:Rosato@cerm.unifi.it), Tel.: +39 055 4574267 Fax.: +39 055 4574253

We report the completion of the first comparison of automated NMR protein structure calculation methods and announce its continuation in the form of an ongoing, community-wide experiment: CASD-NMR (Critical Assessment of Automated Structure Determination of Proteins by NMR). CASD-NMR is open for any laboratory to participate and/or to submit targets.

NMR spectroscopy is the only technique for the determination of the solution structure of biological macromolecules. This typically requires both the assignment of resonances and a labor-intensive analysis of multidimensional NOESY spectra, where peaks are matched to assigned resonances. Software tools for the full automation of the NOESY assignment and the structure calculation steps have the potential to boost the efficiency, reproducibility and reliability of NMR structures. Within the e-NMR project ([www.e-nmr.eu](http://www.e-nmr.eu)), which is funded by the European Commission (Project number 213010), we are developing an approach to assess whether such automated methods can indeed produce structures that closely match those manually refined using the same experimental data (the “reference structures”). The concept closely resembles that of other community-wide experiments, such as CASP, the Critical Assessment of Techniques for Protein Structure Prediction<sup>1</sup>, and CAPRI, the Critical Assessment of Prediction of Interactions<sup>2</sup>. At variance with both CASP and CAPRI, CASD-NMR is entirely based on experimental data, presenting special issues in assembling, organizing, and distributing these data among participants.

We provided seven research teams in the field with ten experimental data sets for various protein systems of known structure and two sets for protein structures not yet publicly available (“blind tests”), courtesy of the NorthEast Structural Genomics consortium (NESG). We then met in Florence, Italy on May 4-6, 2009 to analyze the structures generated (**Fig. 1**), by comparison to the reference structures and by using software tools for structure validation. This first experiment indicated that while most submissions had correct overall folds, on certain targets some programs failed to calculate accurate packing and length of secondary structure elements. The root mean square deviations (RMSDs) of the backbone coordinates from the manually-solved structures were typically in the 1-2 Å range, but reached values as high as 9 Å in some cases.

**The future** – The complete automation of protein solution structure determination from assigned chemical shift lists and unassigned NOESY peak lists may soon reach the point where unsupervised results can be directly deposited to the PDB. It is therefore meaningful and timely<sup>3</sup> to implement CASD-NMR as a community-wide rolling experiment. **We invite software developers to test their fully automated protocols on blind data sets and produce structures as if they would directly deposit them to the PDB.** An assessment meeting is planned for mid-2010. During CASD-NMR, we will regularly release blind test data sets, provided by the whole NMR community, for proteins whose solution structure will be kept on hold by the PDB for at least eight weeks. **We therefore invite any NMR group that is about to deposit a structure in the PDB to contribute a blind test case to CASD-NMR.** The NESG consortium of the NIH Protein Structure Initiative (PSI) will provide one data set per month. The information in a blind data set will include the protein sequence, chemical shift assignments, and unassigned integrated NOESY peak lists. Data providers may also include additional biochemical information and raw spectral data. These experimental data will be available from a central database of the e-NMR project and through the PSI Knowledge Base (<http://kb.psi-structuralgenomics.org/>), also after the release of the reference PDB structure. This will allow participants to join CASD-NMR after its beginning. For more information, including file format requirements and details on the submission procedures, please visit <http://www.e-nmr.eu/CASD-NMR>.

The CASD-NMR participants will have eight weeks to automatically generate structures and deposit in the aforementioned central database their coordinates and the conformational restraints used in calculations. Manual intervention on the data other than re-calibration of chemical shifts is

forbidden. For all structures, the CASD-NMR web site will provide access to the coordinates, their comparison to the reference structure and their various validation scores. These data will foster the development of better algorithms and validation tools, and the adoption of state-of-the-art protocols by the wider bio-NMR community.

We look forward to a fascinating experiment!

1. Moult, J. *et al. Proteins*, **23**, ii-v (1995).
2. Janin, J. *et al. Proteins*, **52**, 2-9 (2003).
3. Editorial, *Nature Methods*, **5**, 659 (2008).

**Figure 1. Performance of various automated structure calculation methods.** The results of fully automated calculations by various programs for one of the blind test data sets of the 2009 Florence workshop are compared to the reference structure (bottom right) determined by Aramini et al. (PDB ID 2kif).

