SEVENTH FRAMEWORK PROGRAMME

Research Infrastructures

# Deliverable 3.5.1 & Deliverable 3.5.2

## CASD-NMR evaluation workshop


| | |
|---|---|
| Grant Agreement number: | RI-261572 |
| Project acronym: | WeNMR |
| Project title: | WeNMR, A worldwide e-Infrastructure for NMR and structural biology |
| Deliverable type and status | Report – Public |
| Due Date | July 2012 (M21; D3.5.1) and October 2013 (M35; D3.5.2) |
| Delivery Date | August 2013 (M34) |
| Project coordinator name, title: | Prof. Alexandre M.J.J. Bonvin |
| Organization: | Utrecht University, The Netherlands |
| Tel: | +31 30 2533859 |
| Fax: | +31 30 2537623 |
| E-mail: | a.m.j.j.bonvin@uu.nl |
| Project website address: | www.wenmr.eu |

# Table of Contents

# Section 1: Summary of Deliverable

## Background

CASD-NMR is a rolling community-wide experiment involving developers of software tools / protocols for the automated calculation of protein structures from NMR data. The goal of CASD-NMR is to help advance the relevant methodology in order to reach the level of quality and reliability required for direct structure deposition in the PDB. CASD-NMR also produces extensive data sets that will be useful to develop better methods for NMR structure validation. CASD-NMR collects and makes available to the participants NMR data sets that can be successfully used for protein structure determination. All data sets in CASD-NMR lead to a satisfactory structure through the traditional manually curated procedures. They are released as blind data sets, i.e., the corresponding protein structure is not publicly available at the time of release. The participants are given 8 weeks to generate a structure using fully automated methods as if they would directly deposit them into the PDB. The coordinates are stored in a central database at the Magnetic Resonance Center in Florence. The results will be analyzed through various validation tools and the corresponding outputs will be made publicly available.

On a regular basis, organizers and participants meet in CASD-NMR workshops to discuss the results. Originally, it was planned to organize two of such workshops during the lifetime of WeNMR, but due to a relatively low number of submitted structures, it was decided to combine both meetings (see also previous WeNMR reports). The CASD-NMR workshop 2013 was held at Rutgers University, Piscataway on June 1st, 2013.

More information on CASD-NMR and how to join is also available on the WeNMR website: http://www.wenmr.eu/wenmr/casd-nmr

## Goal

This document serves as the report for both Deliverable 3.5.1 and 3.5.2 and describes the preliminary outcomes of the 2012-2013 CASD-NMR rounds, which were also discussed at the CASD-NMR workshop that was held at Rutgers University, Piscataway on June 1st, 2013.

# Section 2: Preliminary results of CASD-NMR round 2013

## Overview

The CASD-NMR experiment (‘‘Critical Assessment of Automated Structure Determination of Proteins by NMR”) has gone through a new round of blind testing of structure determination programs, and is in the process of analysing the results. Blind testing, working on a data set for which the structure is not yet released, is the most realistic test of an automatic structure generation program. Bias and overfitting are excluded, so that programs can be evaluated by comparing their results with each other, and with the manually determined structure submitted to the Protein Data Bank.

## Changes

There have been a number of improvements since the last CASD-NMR round, which was held in 2010 and described in Rosato *et al.* (http://dx.doi.org/10.1016/j.str.2012.01.002).

The current round featured thirteen different calculation protocols (versus eight in the last round) from eleven different groups. Programs could compete using input from either chemical shifts only, raw spectra, unrefined peak lists, or refined (filtered) peak lists, all with a known chemical shift assignment as a starting point. With several programs participating in two categories, there were a total of seventeen data series encompassing at least half the targets, plus additional submissions from the programs I-TASSER and BE-metadynamics. A total of 167 calculations were submitted.

| Category | Protocols |
|---|---|
| Chemical shift only | BE-metadynamics , Cheshire, CS-Rosetta, CS-HM-Rosetta |
| Raw Spectra | Ponderosa, UNIO |
| Peak Lists | ARIA, ASDP-CNS, ASDP-Rosetta , Autonoe-Rosetta-α, CYANA, I-TASSER , UNIO , YAPP |

The table shows the different calculation protocols and the categories they are participating in. All protocols in the ’Peak Lists’ category have submitted calculations for raw peak lists and some software also submitted calculations for the refined peak lists. I-TASSER and BE-metadynamics only participated for the last two, respectively one, targets.

## Preliminary Results

Representatives of the participating calculation and validation groups met on the 1st of June at Rutgers University, Piscataway to discuss the available results. The analysis of the results is still in progress, but at the meeting the participants agreed on a number of preliminary results:

- Compared to CASD-NMR 2010, the results obtained from refined lists only are generally better in terms of agreement with the target structure and the lack of clearly wrong structures.
- Methods that use chemical shifts and database fragments for the initial structure generation (Cheshire and the various derivatives of Rosetta) seem no less reliable than methods using restrained dynamics. Methods starting from raw spectra likewise seem competitive relative to peak-list-based methods
- The structure quality was improved relative to CASD-NMR 2010. Structure quality parameters tended to be heavily influenced by the nature of the refinement protocol, notably the difference between fragment-based and dynamics-based generation methods, and the presence or absence of a final solvated dynamics refinement step.
- Calculations using either refined or unrefined peak lists produced good quality structures, with relatively limited advantage in using refined peak lists. In a number of cases calculations with refined peak lists converged where calculations with unrefined lists did not. Calculations starting from raw spectra showed a level of convergence and quality similar to or even better than calculations using unrefined lists.

## Future plans

The current CASD-NMR round will be continued through a few additional targets in order to improve the counting statistics.

Given the good performance of programs working on unrefined peak lists, future CASD-NMR rounds will no longer provide refined peak lists for calculation.

The next CASD-NMR round will release assignment and NOESY NMR spectra ahead of the assigned chemical shift list. This will allow programs to compete also in unsupervised automatic assignment, either as a stand-alone activity or as part of a structure generation pipeline.

## Annex 1: Programme of the CASD-NMR 2013 meeting

The following mail with information on the venue and programme of the meeting was distributed among all participants:

---

The CASD-NMR-2013 meeting will be held on the Saturday, June 1st and be hosted by Guy Montelione, of Rutgers University in Piscataway, NY, USA

The venue location is: Center for Advanced Biotechnology and Medicine Rutgers University Busch Campus, 679 Hoes Lane, Piscataway New Jersey 08854

Directions: http://www3.cabm.rutgers.edu/about/visitor_info.php

The hotel for meeting participants is the Hyatt Hotel, New Brunswick
http://newbrunswick.hyatt.com/en/hotel/our-hotel.html

The program is as follows:

| | |
|---|---|
| 9.00 - 9.55: | Arrival |
| 9.55- 10.00: | Welcome |
| 10.00-10.15: | Overview of the targets and experimental data handling (Antonio Rosato and Wim Vranken) |
| 10.15-10.40: | Montelione Lab (J.Huang) |
| 10.40-11.05: | Nilges/Malliavin Lab |
| 11.05-11.15: | Break |
| 11.15-11.40: | Guentert Lab |
| 11.40-12.05: | Torsten Lab |
| 12.05-12.30: | Baker Lab |
| 12.30-13.30: | Lunch |
| 13.30-13.55: | Markley Lab |
| 13.55-14.20: | Vendrusculo Lab |
| 14.20-14.45: | Zong Lab |
| 14.45-15.10: | Lange Lab |
| 15.10-15.35: | Utrecht Lab |
| 15.35-16.00: | Break |
| 16.00-16.30: | Evaluation of results/Discussion (Piscataway/Florence) |
| 16.30-17.00: | Evaluation of results/Discussion (Leicester) |
| 17.00-18.45: | Discussion/Conclusions/Wrap up |
| 18.00: | Departure |
| 18.30: | Joint dinner (optional) |

All lab representatives are requested to present their procedures, methods, successes, failures, etc.

At the end, we hope to answer a number of questions (among which):

- What changes, modifications, improvements were implemented during this round in the protocols; what were the reasons for this?
- Given that we now use unfiltered peak lists, do the automated methods still yield reliable results?
- Are the automated methods robust towards errors? Where it is not successful - what issues limit the success?
- Are there advantages for also using the experimental data?
- Are the resulting structures more bias-free compared to procedures using manually edited input?
- Do the automated methods extract more information from the available data?

Discussion topics:

- what did we learn from CASD-NMR2
- what will we need to do further in order to write a paper
- what format problems were encountered in the assesssment - can we make stricter rules
- (e.g. must have H atoms)
- what are open issues for the next cycle of CASD-NMR -- will we go to sparse restraint data sets? RDC? etc

Best regards,

Geerten Vuister
also on behalf of Guy Montelione, Alexandre Bonvin, Antonio Rosato

---

## Annex 2: Participants of the CASD-NMR 2013 meeting

| Name | City | E-mail address |
|---|---|---|
| **Alexandre Bonvin** | Utrecht | a.m.j.j.bonvin@uu.nl |
| **Antonio Rosato** | Florence | rosato@cerm.unifi.it |
| **Guy Montelione** | Piscataway | guy@cabm.rutgers.edu |
| **Janet Huang** | Piscataway | yphuang@cabm.rutgers.edu |
| **Roberto Tejero** | Piscataway/Spain | Roberto.Tejero@uv.es |
| **Geerten Vuister** | Leicester | gv29@le.ac.uk |
| **Rasmus Fogh** | Leicester/Cambridge | r.h.fogh@bioc.cam.ac.uk |
| **Wim Vranken** | Brussels | wvranken@vub.ac.be |
| **Richard Jang** | Michigan | rjang@umich.edu |
| **Micheal Nilges** | Paris | michael.nilges@pasteur.fr |
| **Benjamin Bardiaux** | Nilges Lab, Paris | benjamin.bardiaux@pasteur.fr |
| **Paul Guerry** | Lyon | Paul.Guerry@gmail.com |
| **Andra Cavalli** | Cambridge | amc82@cam.ac.uk |
| **Oliver Lange** | Munich | oliver.lange@tum.de |
| **Zaiyong Zhang** | Munich | - |
| **Woonghee Lee** | Wisconsin | whlee@nmrfam.wisc.edu |
| **Lei Shi** | Seattle | shileiustc@gmail.com |
| **Binchen Mao** | Seattle | binchen@cabm.rutgers.edu |
| **Peter Guentert** | Frankfurt | guentert@em.uni-frankfurt.de |
| **Pieter Hendrickx** | EBI | Pieter@ebi.ac.uk |