



## **Deliverable 1.2**

### **Test data sets for software development and testing**

INFRA-2007-1.2.2 - Deployment of eInfrastructures for scientific communities

**Grant agreement for:**      **Combination of Collaborative projects &  
Coordination and support actions**

Proposal/Contract no.: 213010 – e-nmr

Project full title: Deploying and unifying the NMR e-Infrastructure in System Biology

Project coordinator: Prof. Dr. Harald Schwalbe

Project website: <http://www.enmr.eu/>

**Date due: 31-01-2009**

**Date released: 31-01-2009**

All datasets are released in reference to the [1<sup>st</sup> eNMR workshop for the assessment of software tools](#) which will be held in Florence, Italy on May 4-6, 2009.

- Datasets 1-4 constitute an ensemble useful to evaluate the programs' capability to i. handle prolines in cis-configuration and ii. Highlight possibly subtle structural changes that occur e.g. upon binding of a metal cofactor
- Dataset 6 comprises an extensive amount of experimental data for a relatively small protein, which should enable a high precision of structure calculations
- Dataset 8 is characterized by the presence of two domains that are structurally very similar but essentially independent, with no contact with one another. This poses a challenge for the programs to discriminate properly the signals from each domain.
- Dataset 9 is relative to a homodimeric protein, posing the challenge to discriminate intra- and inter-subunit NOEs. A specific experiment addresses in part this matter.
- The other datasets regard examples of relatively "standard" protein structure determinations

As a whole, the above group of datasets constitute an extensive variety of possible cases, both for the characteristics of the systems under investigation, while remaining in the realm of an individual protein, and for the extent of the experimental data available. This initial basis of data will make it possible to assess the capability of the protocols to:

1. capture subtle structural changes in a series of related system
2. handle cis prolines
3. handle dimeric systems
4. handle multi-domain systems
5. handle systems non completely folded

#	Name	2D <sup>1</sup> H- <sup>1</sup> H NOESY	3D <sup>13</sup> C-separated NOESY	3D <sup>15</sup> N-separated NOESY	Length	Relevant properties	Comments
1	<a href="#">apoTTHA_cis</a>	Yes	Yes	Yes	120	Pro14 in cis configuration	
2	<a href="#">apoTTHA_trans</a>	Yes	Yes	Yes	120	Pro14 in trans configuration	
3	<a href="#">cuTTHA_cis</a>	Yes	Yes	Yes	120	Pro14 in cis configuration	Copper-bound form
4	<a href="#">cuTTHA_trans</a>	Yes	Yes	Yes	120	Pro14 in trans configuration	Copper-bound form
5	<a href="#">Mia40</a>	Yes	Yes	Yes	146		Only 45-109 folded
6	<a href="#">Parvulostat</a>	Four spectra: three in water with $\tau_m$ 50 ms, 100 ms and 150 ms; one in D <sub>2</sub> O with $\tau_m$ 325 ms	Yes, folded in 13C dimension	Yes	68		
7	<a href="#">TTSco</a>	Yes	Yes	Yes	172		
8	<a href="#">Win34</a>		Two spectra, of which one folded in 13C dimension	Yes	202		Two 70-amino acid domains separated by a long flexible linker
9	<a href="#">BASP</a>	Two spectra, of which one is a 13C/15N filtered NOESY for the intersubunit NOEs	Yes	Yes	54	Homodimer	
10	<a href="#">TAF3</a>	Yes	Yes	Yes	75		Bound to two zinc ions