

DSM Science & Technology Awards 2006

Name	Sander Nabuurs
University	Radboud University Nijmegen (NL)
Department	Center for Molecular and Biomolecular Informatics
PhD Supervisor	Prof. Dr. G. Vriend

Description of PhD work of:
Sander Nabuurs



CMBI
Toernooiveld 1
6525 ED Nijmegen
Tel: +31 24 3653397
Fax: +31 24 3652977
E-mail: vriend@cmbi.ru.nl

Sander Nabuurs started his Ph.D. in the field of structural bioinformatics after completing his undergraduate studies in Biophysical Chemistry with honours. The primary goal of his Ph.D. project was the development of novel validation methodology aimed at evaluating, and where possible, improving the structural quality of biomolecular NMR structures. To illustrate the importance of the having accurate knowledge on the quality of biomolecular structures, a brief introduction to the scientific field in which Nabuurs has been active will be given.

The genetic inventories of numerous organisms are being revealed by genome sequencing efforts at an ever increasing rate. However, the resulting genomic information reveals comparatively little on how the individual parts fulfil their function in the machinery of life (See BioNieuws of 17 Feb 2006). This is a direct consequence of the fact that the genes are not the actual actors within a cell. In order to perform their encoded function, genes are transcribed and translated into proteins, and these in turn mediate most of the essential structures and functions of cells.

Much of our understanding of protein function at the atomic level originates from studying three-dimensional protein structures. There are a few techniques to elucidate protein structures and two of those are producing structures in large quantities: single-crystal X-ray diffraction and nuclear magnetic resonance (NMR) spectroscopy. For the proteins that form suitable crystals, X-ray crystallography represents a mature and rapid approach. Soluble proteins that do not crystallize readily can only be studied by solution state NMR spectroscopy.

X-ray crystallographers, aided by largely automated procedures, sometimes can solve a protein structure within hours of data collection. NMR spectroscopists have to go through a more laborious process of resonance assignment and structure calculation, and this can take months. Nevertheless, due to the development of more advanced spectrometers, more sophisticated experiments, and automated assignment and structure calculation procedures, NMR protein structure determination methods have advanced to the point where the structure of small- to medium-sized proteins (up to roughly 30 kDa) can now be determined in a routine manner. This has led to the integration of NMR spectroscopy as a structure determination tool in many of the currently ongoing structural genomics projects.

One of the foremost goals of these structural genomics projects is to determine a basic set of protein structures, which includes at least one member of each of the many different protein fold classes. This set of protein folds should provide the basis for the prediction of the three-dimensional structure of most of the remaining proteins using homology modelling techniques. However, for such an approach to be successful it is of the utmost importance that the protein structures present in the

basic set are accurate and of high structural quality. Also, for further interpretation and use of these structures in follow-up studies it is essential to have detailed knowledge of structural quality. Therefore, it is important that the structures be extensively validated, using both the experimental data and structural knowledge obtained from a reference set of high quality (X-ray) structures. The high rate at which biomolecular structures are being determined within structural genomics projects and especially the ever increasing amount of automation within these projects renders proper validation of the resulting structures even more important. It is within this context that the Ph.D. work of Sander Nabuurs, with its special focus on the validation of biomolecular structures determined using NMR spectroscopy, should be placed. Validation of NMR structures is typically aimed at two aspects: how well do the structures agree with the experimental NMR data and how do the structures compare to statistics derived from a reference database of high quality protein structures. Both aspects are extensively addressed in Nabuurs' thesis.

Within one-and-a-half year after he embarked on his project, he published a novel method to analyze the information content of NMR data in the *Journal of the American Chemical Society*. This method, named QUEEN, is based on a description of the structures in distance space and concepts taken from information theory. It allows for an objective description of the amount of information contained in complete datasets as well as individual restraints. The method was tested on several experimental datasets, and it was shown that QUEEN can be used to successfully identify the crucial restraints in a structure determination project. As a result of this, it gained quite some attention within the field, and the method is currently being used in many research groups within the international NMR community.

Subsequently, Nabuurs applied the information measures implemented in QUEEN to investigate the relation between the information contained in experimental datasets and the quality of resulting structure ensembles. His results show, for the first time, that there is a direct relation between data information content and structural quality. This knowledge was used to derive and publish a new per-residue quality parameter, which provides direct insight into the extent to which structural quality is governed by the experimental input data.

In addition to the quality of the data, the energy parameters used in the final refinement step contribute significantly to the quality of biomolecular NMR structures. With the development and publication of the DRESS database, Sander Nabuurs provides a clear example of this finding. In this database, a set of 100 NMR derived protein structures was re-refined using restrained molecular dynamics in explicit solvent. Validation of the structure ensembles, using approaches developed in the Nijmegen structural bioinformatics group, demonstrates that both the geometric and overall quality of the NMR structures in DRESS is significantly improved compared to the original ensembles. Obviously, these improved structures will prove a better starting point for structure-based follow-up research.. As a result of this, the DRESS database has been incorporated in several international databases

at both the BioMagResBank (BMRB, Wisconsin, USA) and the European Bioinformatics Institute (EBI, Hinxton, UK).

It has become increasingly clear in recent years that the precision of deposited NMR ensembles often overestimates their accuracy. In close collaboration with his co-workers Nabuurs reported on a method that yields a more realistic estimate of the uncertainty in the atomic coordinates by maximizing the structural variance within an ensemble of structures, while maintaining accordance with the experimentally derived data. The results indicate that the structural variance of most NMR structure ensembles can be significantly increased without compromising geometric quality or the fit to the experimental NMR data, indicating that the ensembles deposited in the international databases generally provide a too rose picture of their quality.

Finally, Nabuurs recently presented an application of the methodology, tools, and knowledge described in his thesis to a practical example in the journal *PloS Computational Biology*. In search of a suitable template to build a homology model for the protein Dynein Light Chain 2A, two NMR structures of this protein were obtained from the Protein Data Bank, both originating from different structural genomics efforts. The folds of the two structures are remarkably different, despite their high sequence identity (96%). In this paper a detailed analysis of both structure ensembles is presented, which allowed him to identify one of the two ensembles as incorrect. Subsequently, the analysis of a large set of structures solved as part of structural genomics efforts shows that this erroneous structure is unfortunately not an isolated incident. In the conclusion, Nabuurs offers clear suggestions on how the methods and tools described in his thesis could be applied to prevent such serious errors from occurring in the future.

The clear relevance of Nabuurs' work for both NMR spectroscopists as the end-users of biomolecular structure models has resulted in quite some interest from not only the international scientific community, but also the Dutch press, culminating in articles in several life sciences magazines as also the Dutch nationwide paper *De Volkskrant*.

In addition to the work directly related to this Ph.D. project Nabuurs has been very active in several collaborations with experimental groups. Together with scientists from diverse backgrounds (human genetics, animal physiology, cell biology etc.) he applied his knowledge of protein structure in different homology modelling projects. Nearly all these projects have lead to publications in renowned international journals.

On the 9th of February 2006 Sander Nabuurs defended this thesis "On the quality of NMR structures – methodology and tools for NMR data and structure validation" in order to obtain his Ph.D. degree. This title was awarded with the honours 'cum laude'.