



## ARIA: automated NOE assignment and NMR structure calculation

Jens P. Linge, Michael Habeck, Wolfgang Rieping and Michael Nilges

Unité de Bio-Informatique Structurale, Institut Pasteur, 25-28 rue du docteur Roux, F-75015 Paris, France

Received on July 25, 2002; revised on September 5, 2002; accepted on September 9, 2002

### ABSTRACT

**Motivation:** In the light of several ongoing structural genomics projects, faster and more reliable methods for structure calculation from NMR data are in great demand. The major bottleneck in the determination of solution NMR structures is the assignment of NOE peaks (nuclear Overhauser effect). Due to the high complexity of the assignment problem, most NOEs cannot be directly converted into unambiguous inter-proton distance restraints.

**Results:** We present version 1.2 of our program ARIA (Ambiguous Restraints for Iterative Assignment) for automated assignment of NOE data and NMR structure calculation. We summarize recent progress in correcting for spin diffusion with a relaxation matrix approach, representing non-bonded interactions in the force field and refining final structures in explicit solvent. We also discuss book-keeping, data exchange with spectra assignment programs and deposition of the analysed experimental data to the databases.

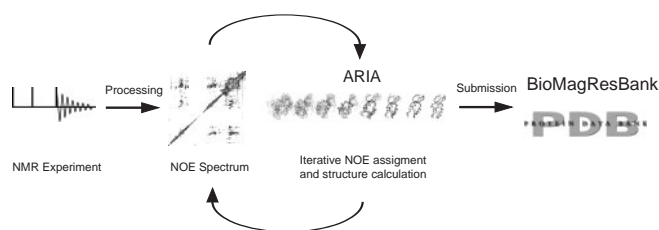
**Availability:** ARIA 1.2 is available from: <http://www.pasteur.fr/recherche/unites/Binfs/aria/>

**Supplementary information:** XML DTDs (for chemical shifts and NOE crosspeaks), Python scripts for the conversion of various NMR data formats and the results of example calculations using data from the *S. cerevisiae* HRDC domain are available from: <http://www.pasteur.fr/recherche/unites/Binfs/aria/>

**Contact:** [linge@pasteur.fr](mailto:linge@pasteur.fr)

### INTRODUCTION

The most time-consuming step in the analysis of NMR data and structure calculation is the assignment of NOE peaks. Manual NOE assignment is tedious due to the large number of possible assignments for each peak, overlap and artefacts in the spectra. Several programs for automated NOE assignment have been developed: CLOUDS (Grishaev and Llinás, 2002), CANDID (Herrmann *et al.*, 2002), NOAH (Mumenthaler *et al.*,



**Fig. 1.** Typical data flow in an NMR structure determination project using ARIA.

1997), AUTOSTRUCTURE (Montelione *et al.*, 2000) and ARIA (Nilges *et al.*, 1997; Linge *et al.*, 2001). ARIA automatically assigns NOEs in an iterative manner using the concept of ambiguous distance restraints (for an overview, see Figure 1). Currently, more than 50 structures deposited in the PDB cite ARIA as refinement method. Version 1.2 of ARIA has several new features which we summarize in this article.

### SPIN DIFFUSION CORRECTION

To tackle the problem of indirect magnetization pathways, we implemented a CPU efficient spin diffusion correction. It relies on a new algorithm to calculate NOE intensities from preliminary structures (Linge, 2001). By comparing theoretical and measured NOE intensities, ARIA estimates target distances for the next iteration of structure calculation; the calculated NOE intensities serve as a simple correction factor that allows an easy integration of the algorithm in the calibration process. The user only has to specify spectrometer frequency, rotation correlation time and NOE mixing time. The spin diffusion correction permits tighter error bounds for the distance restraints. This helps to distinct between data and noise in the analysis of violated distance restraints

## PARALLHDG 5.3 FORCE FIELD

We further improved the PARALLHDG force field (Linge and Nilges, 1999); most notably, we modified the side-chain dihedral angles. Test calculations on Interleukin-4, Crambin and Ubiquitin demonstrated that the force field dihedral angle term in version 5.3 improves the  $\chi_1$ - $\chi_2$  distribution without altering other quality indices like the Ramachandran map or the packing quality (Linge et al., 2003).

## REFINEMENT IN EXPLICIT SOLVENT

A short molecular dynamics trajectory in a thin layer of explicit solvent (water or DMSO) refines the final structure ensemble. The employed force field is consistent with the one used for structure calculation and validation. Thus, there are no systematic differences between the force field and the validation software that could bias the validation results. We have shown that few solvent molecules in a thin layer around the protein are sufficient to account for most important effects of water, notably polar and non-polar interactions on the protein surface (Linge et al., 2003). The refinement is CPU efficient (9 min for a 75-residue protein on an Athlon 1.5 GHz processor) and is fully integrated into ARIA.

## DATA FORMATS, INTERACTION WITH OTHER SOFTWARE PACKAGES, SUBMISSION TO THE DATABASES

Users can start ARIA calculations with peak and chemical shift lists in the formats of ANSIG, NMRView or XEASY. They can include predicted dihedral angles from the programs TALOS and CSI to improve the convergence in an early stage of a structure determination project.

After the last iteration, ARIA exports the assigned NOE peak tables in several formats, including NMRView, XEASY, MOLMOL, BioMagResBank STAR files and XML files. Thus, it is straightforward to compare the results of an ARIA analysis with the original spectra and to validate the automated assignments. To improve book-keeping for protons swapped during the floating chirality assignment, ARIA generates a list of all the swapped protons for each structure.

With the flood of new structures, NMR software must allow easy deposition of all relevant information to the databases (BioMagResBank, PDB). Therefore, we integrated a Python module (written by Jurgen Doreleijers, BMRB) that enables us to write BioMagResBank compliant STAR files. The users can send their final data

(unambiguous and ambiguous NOE assignments, chemical shifts) directly to the BMRB without any further modifications. This saves a lot of time for both structural biologists and database curators.

We are working on a full implementation of the CCPN data-model (Fogh et al., 2002) to further facilitate data exchange. The CCPN project will provide more elaborate means to exchange data between different programs and for the submission to the databases.

## OUTLOOK

Future developments will include statistical means to validate the quality of the assignments and to study internal dynamics (similar to the *B*-factor in X-ray diffraction), full support of the CCPN data-model and XML DTDs for all experimental NMR data.

## REFERENCES

- Fogh,R., Ionides,J., Ulrich,E., Boucher,W., Vranken,W., Linge,J.P., Habeck,M., Rieping,W., Bhat,T.N., Westbrook,J. et al. (2002) The CCPN project: an interim report on a data model for the NMR community. *Nat. Struct. Biol.*, **9**, 416–418.
- Grishaev,A., Llinás,M. (2002) CLOUDS, a protocol for deriving a molecular proton density via NMR. *PNAS*, **99**, 6707–6712.
- Herrmann,T., Güntert,P., Wüthrich,K. (2002) Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *J. Mol. Biol.*, **319**, 209–227.
- Linge,J.P., Williams,M.A., Spronk,C.A.E.M., Bonvin,A.M.J.J., Nilges,M. (2003) Refinement of protein structures in explicit solvent. *Proteins Struct. Funct. Genet.*, in press.
- Linge,J.P., O'Donoghue,S.I., Nilges,M. (2001) Automated Assignment of Ambiguous Nuclear Overhauser Effects with ARIA. *Meth. Enzymol.*, **339**, 71–90.
- Linge,J.P. (2001) *New Methods for Automated NOE Assignment and NMR Structure Calculation*. Book on Demand Verlag, Norderstedt, Germany, ISBN: 383111482X
- Linge,J.P., Nilges,M. (1999) Influence of non-bonded parameters on the quality of NMR structures: a new force-field for NMR structure determination. *J. Biomol. NMR*, **13**, 51–59.
- Montelione,G.T., Zheng,D., Huang,Y.J., Gunsalus,K.C., Szyperski,T. (2000) Protein NMR spectroscopy in structural genomics. *Nat. Struct. Biol.*, **7**, S982–985.
- Mumenthaler,C., Güntert,P., Braun,W., Wüthrich,K. (1997) Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *J. Biomol. NMR*, **10**, 351–362.
- Nilges,M., Macias,M.J., O'Donoghue,S.I., Oschkinat,H. (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the Pleckstrin homology domain from  $\beta$ -Spectrin. *J. Mol. Biol.*, **269**, 408–422.