John von Neumann Institute for Computing

Automatic Sequential NOESY Assignment and NMR Structure Improvement by X-Ray

K. Brunner, W. Gronwald, A. Fischer, J. Trenner,
K.-P. Neidig, H. R. Kalbitzer

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 173-176, 2008.

http://www.fz-juelich.de/nic-series/volume40

Automatic Sequential NOESY Assignment and NMR Structure Improvement by X-Ray

K. Brunner, W. Gronwald, A. Fischer, J. Trenner, K.-P. Neidig, and H. R. Kalbitzer

Institute of Biophysics and Physical Biochemistry, University of Regensburg,
93040 Regensburg, Germany
E-mail: konrad.brunner@biologie.uni-regensburg.de

and Bruker Biospin, 76287 Rheinstetten, Germany

We are developing AUREMOL[1] (`www.auremol.de`), which goal is the reliable and automatic structure determination of biological macro molecules such as proteins from NMR data. For a fully automatic sequential NOESY assignment the tool ASSIGN[2] has been developed. The required input consists of a homologous structure for a NOESY spectrum simulation and the experimental NOESY spectrum. ASSIGN fits the simulated NOE signals to the experimental spectrum. The fit quality given by a probability depends on the line shapes and volumes of the signals. The assignment is varied by moving or swapping spin system assignments using a Monte Carlo approach. A threshold accepting algorithm (TA[3]) is employed to find the maximum of accordance.

1 Introduction

For a fully automated sequential NOESY assignment the tool ASSIGN has been developed. The assignment is driven by the comparison of experimental spectra of a protein and simulated spectra. The simulated spectra are derived from a preliminary structure model. ASSIGN is part of the AUREMOL NMR software suite.

2 Method

The basic idea is to use preliminary structural information together with the NOESY peaks to drive the assignment process. Therefore ASSIGN expects additionally to the NOESY spectra a preliminary structure model of the protein to be solved. Such a model can be provided for example by homology modelling[4]. A start assignment can be provided as an optional input. The first step is the recording and the processing of a NOESY spectrum. In this spectrum the signals are identified and the corresponding chemical shifts are stored in a slot list. The second step is to simulate a NOESY spectrum for the structure. Each expected coupling signal is simulated with a proper line shape and volume. The shifts for the signals are not calculated. Instead of that shifts are taken from the slot list and randomly assigned to the simulated signals. If a start assignment is provided as input the shifts are assigned according to the start assignment. In the third step the resulting simulated spectrum is compared with the experimental one with respect to line shapes and signal volumes. The degree of accordance is expressed as a probability. In the following a quenching protocol is applied to the simulation procedure to improve the agreement of the spectra. A random perturbation swaps the shifts of two simulated signals and the probability of accordance is recalculated. If the new parameters lead to an improved agreement with the experimental data, they are accepted, otherwise declined. This method is repeated until the agreement between experiment and simulation cannot be further improved for a defined number of iterations. As a result a sequential shift chemical assignment is obtained that can explain the experimental spectra with the final probability of accordance.

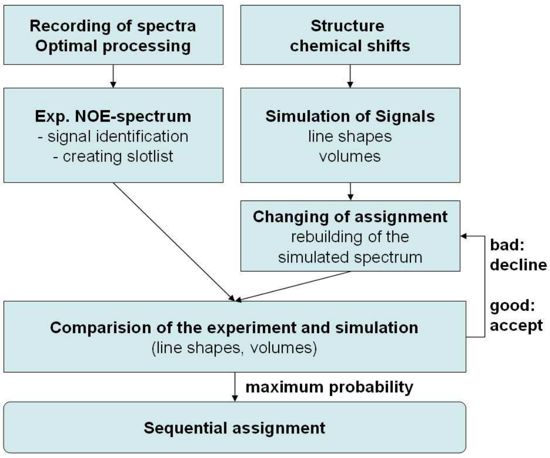

Figure 1. Scheme of the ASSIGN algorithm.

The agreement with the experimental data is expressed as probability and calculated on each test region where an experimental signal is found. Shapes are compared with the help of the cosinus criterion, volumes are compared directly by summing up the intensities of the testing regions. With the help of frequency distributions of solutions with random and partial correct assignments probabilities of line shape and volumes are derived. For the test region p the Bayesian probability of shape and volume (PSV) is given as

$$PSV_p = \frac{P_p^{S,ok} P_p^{V,ok}}{P_p^{S,ok} P_p^{V,ok} + P_p^{S,rnd} P_p^{V,rnd}} \tag{1}$$

$P^{S,ok}$ and $P^{V,ok}$ are the probabilities of the partial correct solution and $P^{S,rnd}$ and $P^{V,rnd}$ are the probabilities of the random solution. The sum over all (N_{ex}) test regions ESV (Energy Shape Volume)

$$ESV = \sum_{p=1}^{N_{ex}} |\ln(PSV_p)| \tag{2}$$

is optimized by the TA-algorithm[3].

3 Results

In order to evaluate the method, a set of test cases is considered based on the structure of HPr *S. aureus* (H15A) that has already been solved by NMR. In case of an ideal data set the simulated spectrum is calculated from the known structure and is also used as an experimental spectrum. Here 450 of 455 shifts (99.3 %) were found correct without any partial start assignment. In case of a real experimental spectrum of HPr *S. aureus* (H15A) from up to 20 % correct start assignments over 90 % to 99 % correct assignments were found (Fig. 2).

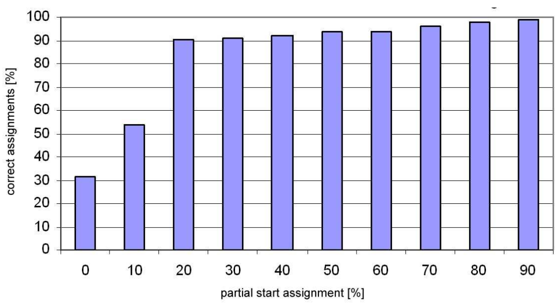

Figure 2. Correct assignments against partial start assignment.

In a third test case only easily obtainable NMR data such as chemical shifts of the backbone atoms H^N and H^α is used. In HPr *S. aureus* (H15A) these are 36.5 % of the whole assignment. With this start assignment 500 structures were calculated. The 10 best in respect of energy were taken for further analysis (Fig. 3B). Then ASSIGN was used to assign the missing shifts of the side chain atoms. 85.2 % of the shifts were correctly found. The structure bundle calculated from these data is shown in Fig. 3C. A clear improvement of the structure can be seen easily and the bundle is very similar to the original bundle (Fig. 3A). Also the AUREMOL NMR R-factor[5] (0.589 $\rightarrow$ 0.328), the RMSD (0.151 nm $\rightarrow$ 0.015 nm) and the Ramachandran[6] values are improved.

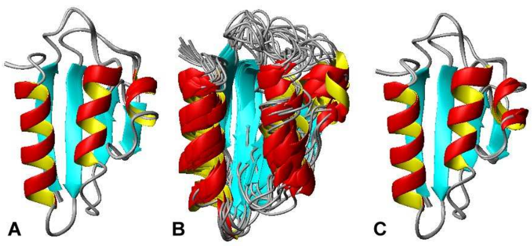

Figure 3. (A) Original structure, (B) structure calculated from bb-atoms, (C) ASSIGN improved structure of HPr *S. aureus* (H15A).

In the last test case the sequential assignment of the mutant HPr *S. aureus* (H15A) should be found with help of the solved structure of HPr *S. aureus* (wt). In this example 79.8 % of the assignments were already given (Fig. 4B). ASSIGN finds 96.7 % of the correct assignments, which leads to the structure seen in Fig 4C. Again quality values such as AUREMOL R-factor (0.384 → 0.356), RMSD (0.046 nm → 0.027 nm) and the Ramachandran are improved.

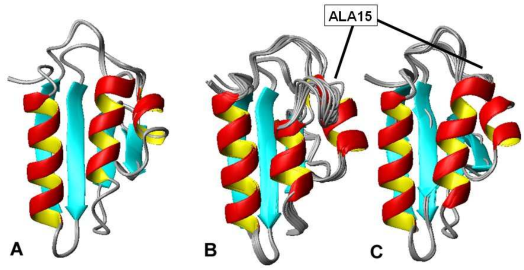

Figure 4. (A) Original structure, (B) structure calculated from the sequential assignment from the wt, (C) ASSIGN improved structure of HPr *S. aureus* (H15A).

For NMR structure improvement by X-ray data please see the ISIC algorithm[7].

Acknowledgments

Financial support from the BMBF, the Deutsche Forschungsgemeinschaft (DFG), and the European Union is gratefully acknowledged.

References

1. Gronwald, W. and Kalbitzer, H. R., *Automated Structure Determination of Proteins by NMR Spectroscopy*, Progr. NMR Spectr. **44**, 33–96, 2004.
2. Brunner, K., *Modellierung, Strukturverbesserung und sequentielle Zuordnung als vollautomatische Module für die automatisierte Proteinstrukturbestimmung im Softwareprojekt AUREMOL*, (Dissertation, Universität Regensburg, Germany, 2006).
3. Dueck,G. & Scheuer,T., *Threshold accepting: A general purpose algorithm appearing superior to simulated annealing*, J. Comput. Phys. **90**, 161–175, 1990.
4. Möglich,A., Weinfurtner,D., Gronwald,W., Maurer,T., & Kalbitzer,H.R., *PERMOL: Restraint-Based Protein Homology Modeling Using DYANA or CNS*, Bioinformatics **21**, 2110–2111, 2005.
5. Gronwald W., Kirchhofer R., Gorler A., Kremer W., Ganslmeier B., Neidig K.P., Kalbitzer H.R., *RFAC, a program for automated NMR R-factor estimation*, J Biomol NMR. **Jun;17(2)**, 137-51, 2000.
6. Laskowski,R.A., MacArthur,M.W., & Thornton,J.M., *Validation of protein models derived from experiment*, Curr. Opin. Struct. Biol. **8**, 631–639, 1998.
7. Brunner K., Gronwald W., Trenner J.M., Neidig K.P., Kalbitzer H.R., *A General Method for the Unbiased Improvement of Solution NMR Structures by the Use of Related X-Ray Data, the AUREMOL-ISIC Algorithm.*, BMC Struct Biol. **Jun 26;6(1):14**, 2006.