

New computational methods for automated cross-peak
assignment of solution- and solid-state NMR spectra of
proteins.

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie und Pharmazie
der Freien Universität Berlin

vorgelegt von

Michele Fossi
aus Florenz

Juli 2005

1. Gutachter: Prof. Dr. Hartmut Oschkinat

2. Gutachter: Prof. Dr. Udo Heinemann

Disputation am 23/11/2005

Acknowledgements

This thesis is dedicated to my father and to my mother, for their constant support and their moving attempts to understand and show interest for the behaviour of proteins immersed in magnetic fields.

I thank my family, Nonna Eda, Nonno Sandro, Nonna Lidia, Nonna Gemma, my cousin Sharon, my aunt Rita, Maria Teresa for always having believed in me and having encouraged me throughout my studies.

I thank Hartmut, my supervisor, who gave me the possibility to work on this interesting project and to experience two incredible cities, Berlin and Paris.

I thank Michael Nilges, my supervisor at the Pasteur Institute in Paris, for his kindness and for having introduced me into the world of bioinformatics. I thank Jens Linge for having guided me step by step, with patience, into learning Python. Michael Habeck, Wolfgang Rieping, for having always answered my daily questions. Nathalie, for having offered me so many interesting french novels.

I thank all my colleagues at the FMP, Linda Ball, Barth Van Rossum, Federica Castellani, Urs Wiedemann, Anne Diehl, Peter Schmieder, Dietmar Leitner, Dirk Labudde, Nikolaj Schröder and Henrik Holtmann for the interesting scientific conversations that have inspired me. In particular, Linda, Barth, I have learnt a lot from you.

My friends in Berlin and Paris deserve my gratitude for all the good time we had together in these years.

Abbreviations

2D, 3D, two-, three-dimensional;
Å, Ångström (1 Å = 0.1 nm);
ARIA, Ambiguous Restraints for Iterative Assignment;
ADR, Ambiguous Distance Restraint;
ArgR, Arginine Repressor N-terminal domain;
CNS, Crystallography and NMR system;
CSA, chemical shift anisotropy;
 γ , gyromagnetic ratio
 Δ , vector of chemical shift tolerances;
 Δ_{\max} , values of Δ for which the number of accepted peaks is maximal;
 δ^{pro1} , chemical shift tolerance for the indirect proton dimension;
 δ^{pro2} , chemical shift tolerance for the direct proton dimension;
EVH1, Ena/VASP Homology 1 domain;
HRDC, helicase and RNaseD C-terminal domain;
L, lower limit for distance restraints;
MAS, magic-angle-spinning;
MD, Molecular Dynamics
 $n(C_j)$, number of assignment options for the peak C_j ;
 n_{\max} , maximum number of assignment possibilities allowed per peak;
 n_{av} , average number of assignment possibilities per peak in the first iteration;
 ${}^mN_{\text{tot}}$, total number of entries in the merged list;
NMR, Nuclear Magnetic Resonance;
NOE, Nuclear Overhauser Effect;
PDB, Protein Database;
ppm, part per million;
rmsd, root mean square deviation;
SA, Simulated Annealing;
SH3, Src Homology domain 3;
 T_1 , T_2 , longitudinal and transverse relaxation rates.
U, upper limit for distance restraints;
VASP, Vasodilator Stimulated Phosphoprotein.

Contents

Chapter 1	General Introduction	10
1.1	Nuclear Magnetic Resonance (NMR)	10
1.1.1	The nuclear spin Hamiltonian	10
1.1.2	Magnetisation and pulses: 1D and multidimensional NMR spectra	15
1.2	Solution NMR	16
1.2.1	Protein structure determination <i>via</i> solution NMR	17
1.2.1.1	Collection of structural restraints	17
1.2.1.2	Structure calculation <i>via</i> Simulated Annealing	20
1.3	Solid-state NMR	24
1.3.1	Protein structure determination <i>via</i> solid-state NMR	28
1.3.1.1	Collection of structural restraints	28
1.3.1.2	Structure calculations of the α -spectrin SH3 domain	31
1.4	ARIA	32
1.4.1	Unambiguous results from ambiguous data	32
1.4.2	Description of the software	34
1.4.2.1	Peak annotation	35
1.4.2.1.1	Δ and n_{\max} : two important parameters that strongly influence the peak annotation	38
1.4.2.1.2	The parameter p : removal of assignment options during the iterations	40
1.4.2.2	Distance restraint evaluation	42
1.4.2.3	Merging of the peak lists	45
1.4.2.4	A description of the SA protocol used in ARIA	45
1.4.2.5	Recognition of artefacts and of noise-peaks	47
1.5	Overview of the thesis	48
Chapter 2	Material and Methods	50
2.1	ARIA calculations	50
2.1.1	Protein datasets used	50
2.1.2	Pre-calculation analysis by Cesta.py	50
2.1.3	Structure calculations	51
2.2	SOLARIA calculations	52
2.2.1	Protein dataset used and MAS NMR spectroscopy	52
2.2.2	Structure calculations	53
Chapter 3	Influence of chemical shift tolerances on NMR structure calculations using ARIA protocols for assigning NOE data.	54
3.1	Introduction	54
3.2	Cesta.py: a pre-calculation analysis of the influence of Δ and n_{\max} on peak annotation	55
3.2.1	A description of Cesta.py	55

3.2.2	The output of Cesta.py: four diagnostic functions	56
3.2.2.1	$N_{\text{noassig}}^{\text{rej}}(\Delta)$: the number of rejected peaks due to a lack of assignment options as a function of Δ	56
3.2.2.2	$N_{\text{n}_{\text{max}}}^{\text{rej}}(\Delta, \text{n}_{\text{max}})$: the number of rejected peaks due to an excess of assignment options as a function of Δ	58
3.2.2.3	$N^{\text{acc}}(\Delta, \text{n}_{\text{max}})$: the number of accepted peaks as a function of Δ and n_{max}	58
3.2.2.4	$n_{\text{av}}(\Delta, \text{n}_{\text{max}})$: the average number of assignments per peak n_{av} as a function of Δ and n_{max}	60
3.3	Results and discussion	61
3.3.1	Analysis of the NOE assignment in the first iteration: evaluation of $N_{\text{noassig}}^{\text{rej}}(\Delta)$, $N_{\text{n}_{\text{max}}}^{\text{rej}}(\Delta, \text{n}_{\text{max}})$, $N^{\text{acc}}(\Delta, \text{n}_{\text{max}})$ and $n_{\text{av}}(\Delta, \text{n}_{\text{max}})$ for five different protein datasets by means of Cesta.py	61
3.3.2	Influence of Δ and n_{max} on the quality of the structures	65
3.3.2.1	Calculations with $\text{n}_{\text{max}}=5$	66
3.3.2.2	Calculations with $\text{n}_{\text{max}}=200$	69
3.3.2.3	Calculations with $\text{n}_{\text{max}}=20$	70
3.3.3	Calculations of the PB1 domain	71
3.3.4	A strategy for choosing most suitable values for Δ and n_{max}	71
3.4	Conclusions	73
Chapter 4	Quantitative study of the effects of chemical shift tolerances and rates of SA cooling on structure calculation from automatically assigned NOE data.	76
4.1	Introduction	76
4.2	Results and discussion	77
4.2.1	Effect of SA cooling rate on the accuracy of structures obtained with increased Δ	77
4.2.2	Slower cooling <i>versus</i> increased number of structure calculations	84
4.2.3	Time costs of the method	87
4.3	Conclusions	92
Chapter 5	SOLARIA: a protocol for automated cross-peak assignment and structure calculation for solid-state NMR.	94
5.1	Introduction	94
5.2	Computational aspects	95
5.2.1	Different correlations observed in distance-dependent solid-state NMR spectra and NOESY-type solution NMR spectra	96
5.2.2	Carbon-labelling pattern	96
5.2.3	Definition of a labelling and power-adapted weighting factor w_{ij}	97
5.2.4	Fixed boundaries for distance restraints	98
5.2.5	Inter-molecular peaks	98
5.3	Results and discussion	99
5.4	Conclusions and perspectives	105

Chapter 6	Summary	108
6.1	Summary of the work and future perspectives	108
6.2	Zusammenfassung	111
	Bibliography	116