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, Wolfang 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, Ångstrøm (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;

$^{m}N_{tot}$, 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