



Doctoral Thesis

Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment

Author(s):

Keller, Rochus Leonhard Josef

Publication Date:

2005

Permanent Link:

<https://doi.org/10.3929/ethz-a-005068942> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 15947

OPTIMIZING THE PROCESS OF
NUCLEAR MAGNETIC RESONANCE SPECTRUM ANALYSIS AND
COMPUTER AIDED RESONANCE ASSIGNMENT

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of
Doctor of Natural Sciences

presented by
ROCHUS LEONHARD JOSEF KELLER

Diplomierter Elektroingenieur ETH

born 03.28.1966

citizen of Endingen (AG)

supervised by
Prof. Dr. Kurt Wüthrich

accepted on the recommendation of

Prof. Dr. Konstantin Pervushin, examiner
Prof. Dr. Beat H. Meier, co-examiner
Prof. Dr. Rudolf Glockshuber, co-examiner
Prof. Dr. Joachim Buhmann, co-examiner
Prof. Dr. Iain Campbell, co-examiner
Dr. habil. Peter Güntert, co-examiner

2005

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is an important method which allows to determine the three-dimensional structure of proteins and other biological macromolecules. The process is based on the concept of sequence-specific resonance assignment, where cross-peaks between sequentially neighboring amino acids are observed in multidimensional NMR spectra and linked to fragments. The fragments can be uniquely mapped on the sequence, if they become sufficiently long, which allows the atoms of the residue types to be assigned to their corresponding chemical shifts. To assign a protein with 100 or more residues, several thousand signals have to be identified and analyzed. This task is very complex and usually takes several month of manual work. Up to now the available tools left most of the complex information management to the operator, who had to take care of plausibility and consistency by himself.

One of the major achievements of this dissertation is a formal and sufficiently complete information model. This model is able to describe and capture all information coming up during the analysis and resonance assignment of NMR spectra and to ensure consistency. The new software package CARA is a comprehensive implementation of this information model. It follows a semi-automatic approach and causes a significant increase of process efficiency and a decrease of error probability. In contrast to previous solutions like XEASY, whose information management is primarily based on peaklists, CARA makes use of a central repository able to manage abstract and semantically interlinked information objects. The availability of this repository allows CARA to dynamically calculate the needed projections (e.g. the cross-peaks expected in a concrete spectrum) by means of incremental inference algorithms. Further concepts have been developed to simulate the magnetization transfer pathways of NMR experiments, to integrate cross-peaks and to back-calculate and efficiently store NMR spectra.

Zusammenfassung

Die Kernresonanzspektroskopie (NMR) ist eine wichtige Methode zur Bestimmung der dreidimensionalen Struktur von biologischen Makromolekülen im allgemeinen und Proteinen im speziellen. Der Prozess basiert auf dem Konzept der sequenzspezifischen Zuordnung von Kernspinresonanzen. Aminosäuren, die in der Sequenz benachbart sind, manifestieren sich in mehrdimensionalen NMR-Spektren in der Form sog. *Cross-Peaks*, die anhand übereinstimmender chemischer Verschiebungen zu Fragmenten verbunden werden können. Wenn diese eine hinreichende Länge erreichen, ist deren eindeutige Abbildung auf die Sequenz und somit die Zuordnung der beobachteten Signale zu den einzelnen Atomen der Aminosäuren möglich. Die Zuordnung einer Sequenz von hundert oder mehr Aminosäuren erfordert die Analyse und Identifikation von tausenden von Signalen. Diese Aufgabe ist sehr komplex und erfordert oft einen mehrmonatigen Einsatz, verbunden mit viel Handarbeit. Bisherige Werkzeuge überliessen den Grossteil der komplexen Informationsverwaltung der Verantwortung des Operateurs, der sich selber um die Plausibilität und Konsistenz der Daten kümmern musste.

Im Rahmen der vorliegenden Dissertation wurde ein formales, hinreichend vollständiges Informationsmodell entwickelt, das in der Lage ist, sämtliche während der Analyse und Resonanzzuordnung von NMR-Spektren anfallenden Informationen aufzunehmen und deren Konsistenz zu gewährleisten. Das neue Anwendungsprogramm *CARA* ist eine vollständige Implementation dieses Informationsmodells und ermöglicht dank eines semi-automatischen Ansatzes eine erhebliche Steigerung der Prozesseffizienz und Reduktion des Fehlerrisikos. Im Gegensatz zu bisherigen Anwendungen wie XEASY, deren Informationsverwaltung primär auf Peaklisten basierte, verwendet *CARA* ein zentrales Repository, das alle Informationen in abstrakter und semantisch vernetzter Form verwaltet. Dies erlaubt die dynamische Berechnung der benötigten Projektionen (z.B. *Cross-Peaks* zu einem konkreten Spektrum) mittels inkrementeller Inferenzmechanismen. Weitere Algorithmen ermöglichen die Simulation von Magnetisierungspfaden, sowie die Integration von *Cross-Peaks* verbunden mit deren Rückrechnung auf NMR-Spektren.