



Doctoral Thesis

The new software RADAR for automation of three-dimensional protein structure determination by NMR in solution

Author(s):

Bettendorff, Pascal

Publication Date:

2007

Permanent Link:

<https://doi.org/10.3929/ethz-a-005392847> →

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. 16917

The new software **RADAR** for automation of three-dimensional protein structure determination by **NMR** in solution

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doktor der Wissenschaften

presented by
PASCAL BETTENDORFF

Dipl. Phys. ETH
born June 30th 1975
citizen of Luxembourg

accepted on the recommendation of
Prof. Dr. K. Wüthrich, examiner
Prof. Dr. W. F. van Gunsteren, co-examiner
Dr. T. Herrmann, co-examiner

2007

Abstract

Nuclear magnetic resonance (NMR) spectroscopy has become an important tool for the elucidation of three-dimensional protein structures in solution as well as for the study of the dynamic behaviour of proteins. Typical applications are presented in chapter 2 of the present work. The workflow from the acquisition of experimental NMR spectra to a three-dimensional structure is complex and time-consuming, thus automated procedures are an important component for fast, reliable and unbiased structure determinations. The new algorithm RADAR, which is presented in chapter 3 of the present work combines the functionalities of the two previously developed algorithms ATNOS, for automated spectral signal identification and CANDID, for automated NOE (nuclear Overhauser effect) cross peak assignment. The input required for RADAR consists of the sequence of the amino acids of the protein, the sequence-specific resonance assignments and one or several NOESY (NOE spectroscopy) spectra. The RADAR algorithm is used in an iterative procedure in conjunction with a restrained molecular dynamics (rMD) algorithm, where after an initial RADAR run of NOE signal identification and NOE cross peak assignment a preliminary structure is calculated by rMD, which is subsequently used to guide the signal identification and cross peak assignment. Instead of a sequential use of ATNOS/CANDID, RADAR establishes a direct feedback between the cross peak assignments and signal identification routines, before an initial structure is calculated. As for any automated iterative approach, the first cycle of the iterative procedure is of critical importance for the correctness of the final result since the intermediate structures are used to guide the further analysis. The iterative structure determination procedure using RADAR is more robust than ATNOS/CANDID, due to the improved performance in the first cycle, as

evidenced by testing on seven protein datasets, which resulted in a lower r.m.s. deviation of the coordinates of the rMD conformers, a lower r.m.s. deviation between the initial and final bundles of rMD conformers and a lower r.m.s. deviation bias to the rMD conformers of the reference structure. A detailed analysis shows that RADAR collects 10–20% more distance restraints than ATNOS/CANDID as input for the rMD algorithm, while the residual target function value is comparable. Thus, the additional distance restraints are of high quality and fit well into the overall set of distance restraints. Furthermore, combining the functionalities of ATNOS/CANDID into a single algorithm, allows extended data analyses. The characteristic patterns generated by regular secondary structure elements allow direct identification of α -helices and β -strands, down to the topology of the β -sheet, from the input spectra prior to the structure calculation. Finally, RADAR introduces two algorithms for automated structure validation: the first establishes a measure of the compatibility between the input spectra and the final three-dimensional structure, the second validates the input chemical shifts against the structure by performing chemical shift calculations. Test calculations with correct and distorted conformers established that both methodologies can correctly identify local and global distortions and produces thus a meaningful measure for the compatibility of experimental input and three-dimensional structure.

Zusammenfassung

Kernspinresonanz-Spektroskopie hat sich zu einem wichtigen Werkzeug für die Aufklärung der dreidimensionalen Struktur von Proteinen in Lösung und der Studie dynamischer Phänomene in Proteinen entwickelt. Kapitel 2 der vorliegenden Arbeit stellt einige typische Anwendungen dar. Der Arbeitsablauf von den experimentellen Daten zur dreidimensionalen Struktur ist komplex und zeitaufwendig, so dass automatisierte Verfahren ein wichtiges Element für die schnelle, zuverlässige und neutrale Strukturermittlung sind. Der in Kapitel 3 der vorliegenden Arbeit vorgestellte neue Algorithmus RADAR kombiniert die Funktionalitäten von zwei bestehenden Algorithmen, ATNOS, für die automatisierte Signalidentifikation in Spektren und CANDID, für die automatisierte Zuordnung von NOE (nuklearer Overhauser Effekt) Kreuzsignalen. Die für RADAR erforderlichen Eingangsdaten sind ein oder mehrere NOE-Spektren, die Reihenfolge der Aminosäuren des Polypeptidrückgrats des Proteins und die sequenzspezifischen Resonanzzuordnungen von mehr als 80% der Atome des Proteins. Der RADAR Algorithmus wird in einem iterativen Verfahren in Verbindung mit einer Moleküldynamik-Simulation verwendet. Nach dem ersten RADAR Durchlauf wird eine initiale Struktur berechnet, welche anschliessend die Identifikation und Zuweisung der NOE Kreuzsignale zum Endresultat steuert. Anstelle der starren Abfolge von ATNOS/CANDID erlaubt RADAR eine Wechselwirkung zwischen Signalidentifikation und -zuweisungen ermöglicht, noch bevor die Ausgangsstruktur errechnet wird. Wie in vielen iterativen Verfahren, ist der erste Zyklus des iterativen Verfahrens von höchster Bedeutung für die Korrektheit des Endresultats, da die Zwischenstrukturen benutzt werden, um die weitere Analyse zu steuern. Die iterative Strukturermittlungsprozedur mit RADAR ist robuster als diejenige mit ATNOS/CANDID wegen der verbesserten Leistung im ersten

Zyklus, was mittels verschiedenener Proteindatensätze überprüft wurde, die allesamt eine niedrigere Streuung der Koordinaten der durch die Moleküldynamik-Simulation erzeugten Konformere, eine niedrigere Abweichung zwischen den initialen und endgültigen Konformeren und eine niedrigere Abweichung von der Bezugsstruktur ergaben. Eine ausführliche Analyse zeigt, dass im ersten Zyklus RADAR 10-20% mehr Distanzeinschränkungen für die Moleküldynamik-Simulation als ATNOS/CANDID sammelt, wobei der Restwert der Zielfunktion vergleichbar ist. Die zusätzlichen Distanzeinschränkungen sind somit von vergleichbarer Qualität wie der komplette Satz von Distanzeinschränkungen. Durch die Kombination der Funktionalitäten von ATNOS/CANDID in einen einzigen Algorithmus wurden ferner erweiterte Datenanalysen ermöglicht. Die charakteristischen Muster, die durch regelmässige Sekundärstrukturelemente erzeugt werden, erlauben die direkte Identifikation von α -Helizes und von β -Strängen inklusive der Topologie der β -Faltblätter. Schliesslich stellt RADAR zwei neue Algorithmen für automatisierte Strukturvalidation vor: der erste bietet ein Mass für die Kompatibilität zwischen den Eingangsspektren und der abschliessenden dreidimensionalen Struktur, der zweite erlaubt den Vergleich der chemischen Verschiebungen mit der Struktur mittels Berechnungen der chemischen Verschiebungen. Testrechnungen mit korrekten und verzerrten Konformeren belegten, dass diese Algorithmen lokale und globale Verzerrungen richtig identifizieren können.

Résumé

La spectroscopie par résonance magnétique nucléaire (RMN) est devenue un outil important pour l'élucidation des structures tridimensionnelles de protéines en solution aussi bien que pour l'étude du comportement dynamique des protéines. Des applications sont présentées dans le deuxième chapitre de cette thèse. Le chemin qui mène de l'acquisition des données expérimentales à la structure tridimensionnelle est complexe et long, ainsi les procédures automatisées sont un composant important pour des déterminations de structures rapides, fiables et impartiales. Le nouveau algorithme RADAR, présenté dans le troisième chapitre de cette thèse, combine les fonctionnalités des deux algorithmes développés précédemment ATNOS, pour l'identification automatisée de signaux et CANDID, pour l'attribution automatisée des signaux de NOE (effet nucléaire d'Overhauser). RADAR exige au départ un ou plusieurs spectres de NOESY (spectroscopie de NOE), l'ordre des acides aminés de l'épine dorsale du polypeptide de la protéine et les attributions spécifiques à la séquence de plus de 80% des atomes de la protéine. RADAR est utilisé dans un procédé itératif en coopération avec un algorithme de simulation de dynamique moléculaire. Après une analyse initiale par RADAR d'identification de signaux de NOE et de l'attribution de ces signaux, une structure préliminaire est calculée par simulation de dynamique moléculaire. La structure préliminaire est ensuite utilisée pour guider l'identification et l'attribution des signaux. Au lieu de l'usage séquentiel de ATNOS/CANDID, RADAR permet une interaction entre les attributions et les identifications des signaux avant qu'une première structure ne soit calculée. Comme beaucoup de procédures itératives, le premier cycle du procédé itératif est d'une importance critique pour l'exactitude du résultat final puisque les structures intermédiaires sont employées pour guider l'analyse

supplémentaire. Le procédé itératif de détermination de structure avec RADAR est plus robuste qu'avec ATNOS/CANDID, dû à la performance améliorée dans le premier cycle, comme montré dans des essais sur sept protéines, où ont pu être constatées une déviation plus basse des coordonnées des conformères, une déviation plus basse entre les conformères initiales et finales et une déviation plus basse de la structure de référence. Une analyse détaillée montre que RADAR rassemble 10 à 20 % de plus de contraintes de distance que ATNOS/CANDID pour l'algorithme de simulation de dynamique moléculaire, alors que la valeur résiduelle de la fonction de cible est comparable. Ainsi, les contraintes additionnelles de distance sont d'une qualité comparable à l'ensemble global de contraintes de distance. En outre, la combinaison des fonctionnalités de ATNOS/CANDID dans un seul algorithme permet des analyses de données plus profondes. Les groupes caractéristiques de signaux produits par les éléments réguliers de structure secondaire permettent l'identification directe de hélices α et des brins β , jusqu'à la topologie des feuilletts β . Finalement, RADAR présente deux algorithmes pour la validation automatisée de structures : le premier établit une mesure de la compatibilité entre les spectres d'entrée et la structure tridimensionnelle finale, le deuxième vérifie les déplacements chimiques d'entrée et la structure tridimensionnelle finale par des calculs des déplacements chimiques. Des calculs avec des conformères correctes et tordues ont établi que ces algorithmes peuvent correctement identifier des distorsions locales et globales.