

DISS. ETH NO. 23867

eNOE Method Development And Applications To Protein Allostery

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

DEAN STROTZ

MSc ETH in Biology, ETH Zurich

born on 17.09.1973

citizen of

Zurich, Switzerland

accepted on the recommendation of

Prof. Dr. Roland Riek, examiner

Prof. Dr. Gunnar Jeschke, co-examiner

PD. Dr. Beat Vögeli, co-examiner

2016

Abstract

Three-dimensional structure elucidation at atomic resolution provides key information to understanding biomolecular function. However, detailed understanding of enzyme activity, protein folding and signal transduction throughout biological molecules requires information of internal motion and dynamics in addition to the three-dimensional structure.

There are essentially two techniques to determine molecular structure at atomic resolution; X-ray and NMR. Mainly due to the relatively fast execution, X-ray is the preferred structure determination tool for many applications, if the molecule forms diffracting crystals. NMR however, is the only technique that can determine molecular structures and dynamics at atomic resolution in solution and at near physiological conditions. In the early 1980's the Wüthrich group at ETHZ was developing the conceptual framework to calculate protein structures in solution using NMR spectroscopy. Their approach to calculate structures required sequence-specific resonance assignment and was based on scalar couplings, hydrogen bond information obtained from amide exchange rates and short-range nuclear Overhauser enhancement (NOE) measurements. The structures were calculated using distance geometry computation.

20 years after the Wüthrich's Group seminal achievement and the work on NOE based structure elucidation by many others that followed, the Riek group was picking up the loose ends on the interpretation of NOE data in order to obtain more exact NOE (eNOE) data. The eNOE data is reflective not only of a single static structure but an occupied conformational space and dynamics. To reveal this information a novel ensemble based structure calculation protocol in combination with the analysis of NOE data was required and developed.

In the first part of this thesis (I. Methods) we present the current state of development in methods pertaining to the analysis of distance restraint information from NOE measurements and structure calculation. First we analyze the experimental accuracy of our eNOE data, dissecting in detail the experimental setup to measure the data and the errors contained therein. We then show an extension of our data set that allows analysis of NOE data in more crowded spectra resulting from large proteins with overlapping resonance signals. In order to extract the eNOE data from the large NOE data sets, a comprehensive software package was realized with eNORA2. The software package offers two complementary ways to analyze the NOE data and correct for spin

diffusion, the unwanted and indirect transfer of magnetization during NOE measurements. According to project and sample specific conditions a full-matrix (FM) or three-spin approach for spin diffusion correction may be used. The approach allows for various labeling schemes in deuterated samples enabling the eNOE analysis of large biomolecular systems. Structure calculation is presented in a novel and complementary protocol to the ensemble based structure calculation protocol previously established in the Riek group (minimal conformational space required to fulfill the data). In this novel approach we use replica exchange simulation and maximum entropy reweighing (maximum conformational space required to fulfill the data).

In the second part of this thesis (II. Applications) we illustrate the potential impact on the comprehensive elucidation of the action of biomolecules at atomic resolution using our high-resolution solution state NMR data. One of the major challenges presents itself in the accurate representation, interpretation and analysis of the data. We present the current state of our approach on the examples of Pin1-WW and Cyclophilin A. The Pin1-WW is a peptidyl-prolyl cis/trans isomerase of phospho-Ser/Thr-Pro motifs and key cell-cycle regulator playing a role in human diseases such as Alzheimer's disease, hepatitis C infection and cancer. In order to study information transduction in human Pin1-WW we use replica exchange simulations as basis for a detailed analysis. An extension of methods for the analysis in a statistical manner was required in order to determine atomistic thermodynamic models. Our analysis provides a basis for the understanding of the multi-functionality of the Pin1 protein. We then proceed to employ the ensemble based structure calculation protocol to analyze the allosteric coupling of the ligand-binding event at loop1 to the interface between the Pin1-WW domain and the Pin1-PPIase. Cyclophilin A is another peptidyl-prolyl cis/trans isomerase involved in signal transduction, protein folding, responding to inflammatory stimuli and playing a key role in several human diseases such as viral infections, neurodegeneration, cancer, HIV-1, hepatitis C and asthma. With our eNOE based ensemble structure calculation protocol and analysis of Cyclophilin A, we observe the presence of an open and closed state in the apo form of the protein. This finding infers conformational sampling of Cyclophilin A and a preorganization for catalysis. Furthermore we illustrate the allosteric coupling network connecting the ligand-binding loop to the active site.

Zusammenfassung

Strukturaufklärung steht nur am Anfang eines Verständnisses von molekularen Interaktionen in atomarer Auflösung. Um enzymatische Aktivität, Proteinfaltungsprozesse und Signaltransport durch Biomoleküle zu verstehen, benötigt es zusätzlich detaillierte Kenntnisse über interne Bewegung und Dynamik.

Prinzipiell gibt es zwei Techniken zur Aufklärung von molekularen Strukturen in atomarer Auflösung; Kristallographie und Kernspinresonanz (NMR). Hauptsächlich wegen der raschen Ausführung ist Kristallographie die bevorzugte Technik für viele Anwendungen, unter der Voraussetzung, dass das Molekül sich gut kristallisieren lässt. Kernspinresonanz ist allerdings die einzige Technik, mit welcher molekulare Struktur und Dynamik in molekularer Auflösung nahe an physiologischen Bedingungen ermittelt werden können. In den frühen 80er Jahren hat die Wüthrich Gruppe an der ETHZ die konzeptionellen Rahmenbedingungen geschaffen für die Berechnung von Proteinstrukturen in Lösung mittels Kernspinresonanz. Ihre Methode der Berechnung von Strukturen verlangte die kettenspezifische Resonanzbestimmung und stützte sich auf skalare Kopplungen, Informationen von Wasserstoffbrücken aus Amidaustauschraten und Kurzdistanzen aus Nuklearen Overhauser Effekt (NOE) Messungen. Die Strukturen wurden mittels Distanz-Matrix-Geometrie Berechnungen ermittelt.

20 Jahre nach den zukunftsweisenden Errungenschaften der Wüthrich Gruppe und den Beiträgen von vielen anderen Wissenschaftlern zur Weiterentwicklung von NOE basierten Strukturberechnungen hat die Riek Gruppe die losen Enden der Interpretation von NOE Daten neu aufgegriffen mit dem Ziel einer exakteren Auswertung der Daten. Es hat sich gezeigt, dass diese Daten nicht nur repräsentativ für ein einziges und statisches Ensemble von Strukturen sind, sondern Einsicht in den belegten konformellen Raum und die Dynamik vermitteln. Damit ist ein neues Protokoll für die Analyse von exakten NOE-Daten (eNOE) und die Multi-Zustands-Ensembles basierte Strukturberechnung geschaffen worden.

Im ersten Teil dieser Doktorarbeit (I. Methods) präsentieren wir den gegenwärtigen Stand der Entwicklungen in Methoden zur Analyse der Distanzinformation, die in NOE Daten enthalten sind, aber auch Entwicklungen in Strukturberechnungsmethoden. Zuerst analysieren wir die experimentelle Genauigkeit unserer eNOE Daten, indem wir den experimentellen Aufbau unserer Experimente und die darin enthaltenen Fehler im Detail analysieren. Danach zeigen wir eine Möglichkeit zur Vergrößerung des Datensatzes, welche die Analyse der Daten von grossen Proteinen mit vielen überlappenden

Resonanzsignalen erlaubt. Wir realisierten die umfassende Software eNORA2, um die Extraktion von eNOE Daten aus so grossen Datensätzen, wie sie in NOE Experimenten anfallen, zu ermöglichen. Die Software offeriert zwei sich ergänzende Wege für die Analyse der Daten und die Korrektur von Spindiffusion, den unerwünschten, indirekten Übertrag von Magnetisierung während der Messung von NOE Experimenten. Aufgrund von projekt- und probenspezifischen Kriterien kann entweder die Vollmatrix- oder die Dreispinkorrektur für Spindiffusion gewählt werden. Mit der Dreispinkorrektur für Spindiffusion, ermöglichen wir die Berücksichtigung von verschiedenen Markierungsmöglichkeiten in deuterierten Proben und bereiten damit einen Weg zur Analyse von grossen biomolekularen Systemen mittels eNOE Daten. Komplementär zur bereits etablierten, auf Multi-Zustands-Ensembles basierten Strukturberechnung (Ermittlung des minimalen konformellen Raumes der die Daten reproduziert), präsentieren wir auch ein neues Strukturberechnungsprotokoll wozu wir Replika-Austausch Simulationen und Maximum-Entropie-Gewichtung (Ermittlung des maximalen konformellen Raumes der die Daten zu reproduziert) verwenden.

Im zweiten Teil der Doktorarbeit (II. Applications) illustrieren wir den potentiellen Einfluss unserer hochaufgelösten Strukturdaten auf das Verständnis der Funktion von Biomolekülen in atomarer Auflösung. Eine grosse Aufgabe stellt sich dabei in der korrekten Darstellung, Interpretation und Analyse der Daten. Wir präsentieren unser gegenwärtiges Verständnis der Aussagekraft der eNOE Daten anhand der Pin1-WW Domäne und Cyclophilin A. Die Pin1-WW ist eine Peptidyl-Prolyl-cis/trans-Isomerase für phosphorylierte Serin/Threonin-Prolin-Motive. Diese Domäne ist ein Schlüsselprotein in der Regulierung des Zell-Zyklus und spielt eine signifikante Rolle in menschlichen Krankheiten wie Alzheimer, Hepatitis-C-Infektionen und Krebs. Um die Übertragung von Information in der menschlichen Pin1-WW Domäne zu studieren, verwenden wir Replika-Austausch Simulationen als Basis für unsere Analyse. Um atomare, thermodynamische Modelle entwickeln zu können, verwenden wir eine statistische Analyse. Unsere Analyse bereitet damit eine Basis zum Verständnis der Multifunktionalität des Pin1-Enzyms. Wir benutzen das auf Multi-Zustands-Ensembles basierte Strukturberechnungsprotokoll, um die Bindung des Liganden an die Schleife (Loop 1) und die allosterische Kopplung dieses Events zur Interaktionsoberfläche zwischen der Pin1-WW Domäne und der Pin1-PPIase zu analysieren. Cyclophilin A ist eine weitere Peptidyl-Prolyl-cis/trans-Isomerase, welche ein Bestandteil von Signalübertragungswegen und Proteinfaltungsmechanismen ist, auf Entzündungen regiert und eine wichtige Rolle spielt in verschiedenen menschlichen Krankheiten wie

viralen Infektionen, Neurodegeneration, Krebs, HIV-1, Hepatitis-C und Asthma. Mit unseren eNOE Daten und dem auf Multi-Zustands-Ensembles basierten Strukturberechnungsprotokoll analysieren wir Cyclophilin A und können eine offene und geschlossene Konformation in der Apoform feststellen. Diese Feststellung suggeriert eine konformelle Abtastung von Cyclophilin A und eine Vororganisation für die Katalyse. Im weiteren illustrieren wir mit unserer Analyse die allosterische Kopplung zwischen der Bindungsstelle für den Liganden in der Schleife (Loop 1) und der enzymatisch aktiven Stelle des Enzyms.