PETER GÜNTERT, INSTITUT FÜR BIOPHYSIKALISCHE CHEMIE

NMR-basierte rechnergestützte Strukturbiologie

Rechnergestützte Methoden zur Untersuchung biomolekularer Systeme, insbesondere mit Hilfe der NMR Spektroskopie, sind der Fokus unserer Forschung. Die Beziehung zwischen Struktur, Dynamik und Funktion von biologischen Makromolekülen ist von fundamentaler Bedeutung für das Verständnis des Lebens auf molekularer Ebene und eine wichtige Grundlage der Arzneimittelforschung. Die dreidimensionale Struktur spielt dabei eine entscheidende Rolle, da ihre Kenntnis unverzichtbar ist, um die physikalischen, chemischen und biologischen Eigenschaften eines Proteins zu verstehen. Bis vor kurzem waren NMR Proteinstrukturbestimmungen sehr arbeitsintensive Vorhaben, die für jede neue Proteinstruktur einen erfahrenen Spektroskopiker während Monaten oder Jahren beschäftigten. Diese Situation hat sich mit der Einführung automatischer, rechnergestützter Systeme gewandelt. Wir erweitern die NMR Proteinstrukturanalyse auf bisher dafür nicht zugängliche Systeme wie Proteine mit einer Größe von mehr als 40 kDa, Membranproteine, und Proteine, die direkt in lebenden Zellen untersucht werden.

PROTEINSTRUKTURANALYSE

Dreidimensionale Strukturen von Proteinen in Lösung können auf der Grundlage von konformationellen Einschränkungen aus NMR Experimenten berechnet werden. Unser Programmpaket CYANA, das auf Simulated Annealing durch Moleküldynamiksimulation im Torsionswinkelraum und der automatischen Zuordnung von NOE Distanzeinschränkungen beruht, ist einer der meistgebrauchten Algorithmen für diese Aufgabe. Automatische Methoden zur NMR Proteinstrukturermittlung werden zunehmend akzeptiert und werden jetzt weitverbreitet eingesetzt, um Distanzeinschränkungen automatisch zuzuordnen und die dreidimensionalen Strukturen zu berechnen. Unser FLYA Algorithmus zur vollständig automatisierten NMR Proteinstrukturbestimmung kann die gesamte manuelle Spektrenauswertung ersetzen und beseitigt damit eine wesentliche Beschränkung der Effizienz der NMR Methode zur Proteinstrukturermittlung. Vollautomatische Strukturbestimmung von Proteinen in Lösung (FLYA) liefert, ausgehend von einer Serie mehrdimensionaler NMR-Spektren, ohne manuelles Eingreifen dreidimensionale Proteinstrukturen. Wie beim klassischen, manuellen Verfahren werden die Strukturen durch ein Netzwerk experimenteller NOE Distanzschranken bestimmt, ohne auf bereits bekannte Strukturen oder emprische Molekülmodellierung zurückzugreifen. Zusätzlich zur dreidimensionale Struktur des Proteins liefert FLYA Zuordnungen des Rückgrats und der Seitenketten sowie Kreuzsignalzuordnungen für alle Spektren.

STEREOARRAY ISOTOPENMARKIERUNG

Die NMR Spektroskopie kann dreidimensionale Strukturen von Proteinen in Lösung bestimmen. Dennoch begrenzte die Schwierigkeit, NMR-Spektren mit verbreiterten und überlappenden Resonanzlinen und niedrigem Signal-zu-Rauschen Verhältnis auszuwerten, ihr Poten-

PETER GÜNTERT, INSTITUTE OF BIOPHYSICAL CHEMISTRY

NMR-based Computational Structural Biology

Computation methods to study biomolecular systems, in particular by nuclear magnetic resonance (NMR), are the focus of our research. The relationship between structure, dynamics and function of biological macromolecules is of fundamental importance for understanding life at a molecular level, and a key element of rational drug design. The three-dimensional structure has a pivotal role, since its knowledge is essential to understand the physical, chemical, and biological properties of a protein. Until recently NMR protein structure determination was a laborious undertaking that occupied a trained spectroscopist for months or years for each new protein structure. This situation has changed by the introduction of automated, computational systems. We are extending NMR protein structure analysis to hitherto inaccessible systems, including proteins larger than 40 kDa, membrane proteins, and proteins studied directly inside living cells.

PROTEIN STRUCTURE ANALYSIS

Three-dimensional structures of proteins in solution can be calculated on the basis of conformational restraints derived from NMR measurements. Our CYANA program package, based on simulated annealing by molecular dynamics simulation in torsion angle space and the automated assignment of NOE distance restraints, is one of the most widely used algorithms for this purpose. Automated methods for protein structure determination by NMR have increasingly gained acceptance and are now widely used for the automated assignment of distance restraints and the calculation of three-dimensional structures. Our FLYA algorithm for the fully automated NMR structure determination of proteins is suitable to substitute all manual spectra analysis and thus overcomes a major efficiency limitation of the NMR method for protein structure determination. Fully automated structure determination of proteins in solution (FLYA) yields, without human intervention, three-dimensional protein structures starting from a set of multidimensional NMR spectra. As in the classical manual approach, structures are determined by a set of experimental NOE distance restraints without reference to already existing structures or empirical molecular modeling information. In addition to the three-dimensional structure of the protein, FLYA yields backbone and side-chain chemical shift assignments, and cross peak assignments for all spectra.

STEREO-ARRAY ISOTOPE LABELING

NMR spectroscopy can determine the three-dimensional structure of proteins in solution. Nevertheless, its potential has been limited by the difficulty of interpreting NMR spectra in the presence of broadened and overlapped resonance lines and low signal-to-noise ratios. Stereo-array isotope labelling (SAIL) can overcome many of these problems by applying a complete stereo- and regiospecific pattern of stable isotopes, which is optimal with regard to the quality and information content of the resulting NMR spectra. SAIL utilizes exclusively chemically and enzymatically synthesized amino acids for cell-free protein expression

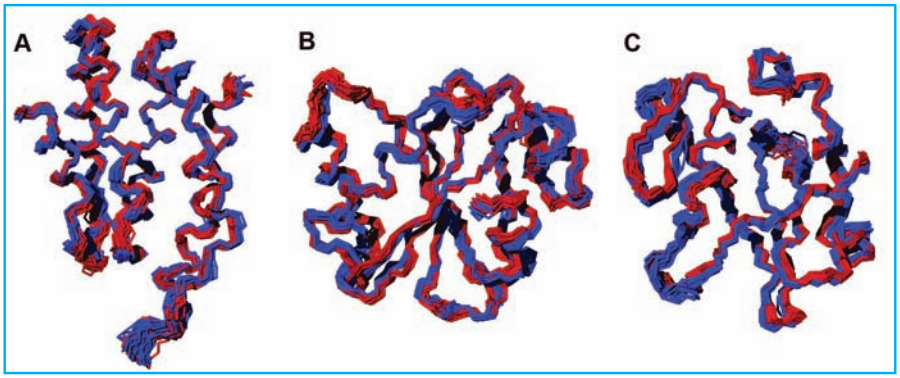

Abb. 1: Vollautomatische NMR Proteinstrukturbestimmung: Proteinstrukturen, die durch vollautomatische NMR Proteinstrukturermittlung mit Hilfe des FLYA Algorithmus' erhalten wurden (blau), sind fast identisch mit den entsprechenden NMR Strukturen, die mit dem herkömmlichen Verfahren bestimmt wurden (rot). (A) ENTH Domäne At3g16270(9–135) von Arabidopsis thaliana. (B) Rhodanase-homologe Domäne At4g01050(175–295) von Arabidopsis thaliana. (C) Src-homologe Domäne 2 (SH2) des menschlichen Katzensarkom Onkogens Fes.

Fig. 1: Fully automated NMR protein structure determination: Protein structures obtained by fully automated structure determination with the FLYA algorithm (blue) are virtually identical to the corresponding NMR structures determined by conventional methods (red). (a) ENTH domain At3g16270(9–135) from Arabidopsis thaliana. (b) Rhodanese homology domain At4g01050(175–295) from Arabidopsis thaliana. (c) Src homology domain 2 (SH2) from the human feline sarcoma oncogene Fes.

zial. Die Stereoarray Isotopenmarkierung (SAIL) überwindet viele dieser Probleme durch die Anwendung eines vollständigen, stereo- und regiospezifischen Musters stabiler Isotopen, das bezüglich der Qualität und des Informationsgehalts der Spektren optimal ist. SAIL verwendet ausschließlich chemisch und enzymatisch synthetisierte Aminosäuren für die zellfreie Proteinherstellung, so dass in allen Methylengruppen ein ^{1}H stereoselektiv durch ^{2}H ersetzt ist, in allen einzelnen Methylgruppen zwei ^{1}H gegen ^{2}H ausgetauscht sind und in den prochiralen Methylgruppen von Leucin und Valin stereospezifisch eine Methylgruppe -$^{12}C(^{2}H)_3$ und die andere -$^{13}C^{1}H(^{2}H)_2$ markiert ist. In sechsatomigen aromatischen Ringen wechseln ^{12}C-^{2}H und ^{13}C-^{1}H Gruppen miteinander ab. SAIL erreicht ein 4–7-fach verbessertes Signal-zu-Rauschen Verhältnis, schärfere Resonanzlinien und eine um 40–60% reduzierte Anzahl von Signalen, ohne essentielle Information über das Rückgrat und die Seitenketten aller Aminosäuretypen zu verlieren. Daraus ergeben sich eine Verringerung des Überlapps in den Spektren, genauere Frequenzbestimmungen, vollständige stereospezifische Zuordnungen und die Messung längerer ^{1}H–^{1}H Distanzen, die es ermöglichen, qualititativ hochstehende Strukturen von Proteinen, die größer als 30 kDa sind, zu bestimmen.

such that in all methylene groups, one ^{1}H is stereo-selectively replaced by ^{2}H, in all single methyl groups, two ^{1}H are replaced by ^{2}H, and in the prochiral methyl groups of Leu and Val, one methyl is stereo-selectively -$^{12}C(^{2}H)_3$ and the other is -$^{13}C^{1}H(^{2}H)_2$. In six-membered aromatic rings, the ^{12}C-^{2}H und ^{13}C-^{1}H moieties alternate with each other. SAIL achieves a 4–7-fold increase in signal-to-noise, sharper resonance lines, and a 40–60% reduction in the number of signals without sacrificing essential information about the backbones and the side chains of all amino acid residue types. The resulting reduction of spectral overlap, the higher accuracy of the frequency determination, the complete stereospecific assignment, and the measurement of longer ^{1}H–^{1}H distances makes it possible to determine high-quality solution structures of proteins larger than 30 kDa.

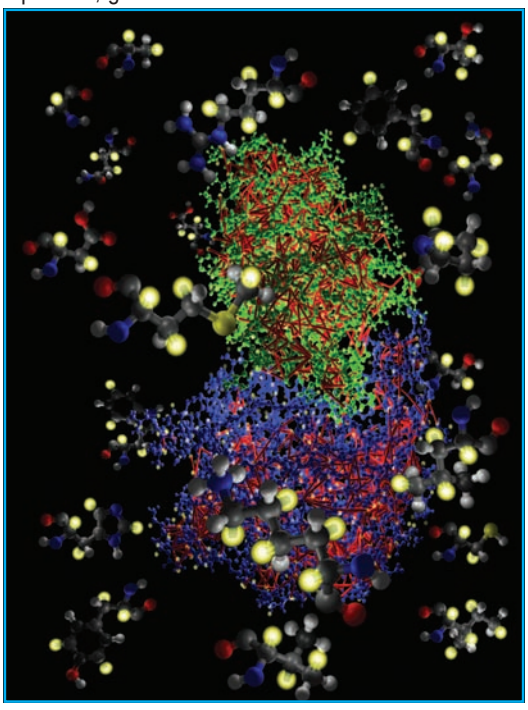

Abb. 2: Stereoarray Isotopenmarkierung (SAIL): Die 20 Standardaminosäuren werden so markiert, dass jede CH_n Gruppe höchstens einen NMR-sichtbaren ^{1}H Kern trägt, während die anderen durch NMR-unsichtbares ^{2}H ersetzt sind. Die verbleibenden ^{1}H Kerne, in der Figur als Lichtquellen dargestellt, liefern Daten, die die NMR Strukturbestimmung von Proteinen ermöglichen, die etwa doppelt so groß sind als bei herkömmlichen NMR Methoden. Die in der Mitte der Figure gezeigte Struktur des 42 kDa Maltodextrin bindenden Proteins MBP wurde in Zusammenarbeit mit dem Labor von Masatsune Kainosho an der Tokyo Metropolitan University in Japan mit Hilfe der SAIL Isotopenmarkierung und des Strukturberechnungsprogramms CYANA gelöst.

Fig. 2: Stereo-array isotope labeling (SAIL): The 20 standard amino acids are labeled such that each CH_n group carries at most a single NMR-visible ^{1}H nucleus, the others being replaced by NMR-invisible ^{2}H. The remaining ^{1}H nuclei, shown as lights in the Figure, provide data that allows the NMR structure determination of proteins about twice as large as by conventional NMR approaches. The structure of the 42 kDa maltodextrin-binding protein MBP that is shown in the center of the Figure was solved in collaboration with the laboratory of Prof. Masatsune Kainosho at Tokyo Metropolitan University, Japan, using SAIL in conjunction with the structure calculation program CYANA.

PROTEINSTRUKTURBESTIMMUNG IN LEBENDEN ZELLEN DURCH IN-CELL NMR SPEKTROSKOPIE

In lebenden Zellen arbeiten Proteine in einer sehr gedrängten Umgebung, in der sie spezifisch mit anderen Proteinen, Nukleinsäuren, Kofaktoren und Liganden wechselwirken. Verfahren zur Bestimmung der dreidimensionalen Struktur gereinigter Proteine in Einkristallen oder Lösung sind weitverbreitet und haben sehr wertvolle Beiträge zum Verständnis biologischer Prozesse geleistet. Es ist allerdings schwierig, die zelluläre Umgebung in vitro nachzubilden. In vivo Beobachtungen der dreidimensionalen Strukturen, Dynamik und Wechselwirkungen von Proteinen sind notwendig um die strukturelle Basis ihrer Funktion im Innern von Zellen zu verstehen. Proteine „an der Arbeit" in einer lebenden Umgebung zu untersuchen is deshalb eines der großen Ziele der Molekularbiologie. Jüngste Entwicklungen der NMR Geräte und Methoden haben es ermöglicht, hochaufgelöste, hetereonukleare, mehrdimensionale NMR Spektren von Makromolekülen in lebenden Zellen zu messen (in-cell NMR). Verschiedene intrazelluläre Abläufe wie Konformationsänderungen, Dynamik und Bindungsereignisse wurden mit dieser Methode untesucht. Dennoch haben die geringe Signalstärke und die kurze Lebensdauer der Proben bis vor kurzem die Messung von genügend struktureller Information für die Bestimmung von Proteinstrukturen durch in-cell NMR verunmöglicht. Kürzlich konnten wir jedoch die erste dreidimensionale Struktur eines Proteins vorstellen, die ausschließlich auf Grund von Daten berechnet wurde, die in lebenden Zellen erhalten wurden. Der in-cell NMR Ansatz kann somit präzise, hochaufgelöste Strukturen von Proteinen in lebender Umgebung liefern.

Abb. 3: In lebenden Zellen bestimmte Proteinstruktur: Die erste dreidimensionale Proteinstruktur, die ausschließlich auf Grund von Informationen berechnet wurde, die aus lebenden Zellen stamen, wurde mit Hilfe der in-cell NMR Spektroskopie für das vermutlich Schwermetall bindende Protein TTHA1718 von Thermus thermophilus HB8 bestimmt, das in E. coli Zellen überexprimiert wurde. Die begrenzte Lebensdauer der Zellen im NMR Probenröhrchen stellt eine große Hürde für die in-cell Strukturbestimmung dar. Gewöhnliche NMR Experimente benötigen 1–2 Tage für die Datenaufnahme, was für lebende Zellen zu lange ist. Die Messdauer konnte auf 2–3 Stunden verkürzt werden, indem für jedes Experiment eine neue Probe vorbereitet und in den indirekten Spektraldimensionen ein nichtlineares Abtastschema in Kombination mit Entropie maximierender Datenprozessierung eingesetzt wurde.

PROTEIN STRUCTURE DETERMINATION IN LIVING CELLS BY IN-CELL NMR SPECTROSCOPY

Proteins in living cells work in an extremely crowded environment where they interact specifically with other proteins, nucleic acids, co-factors and ligands. Methods for the three-dimensional structure determination of purified proteins in single crystals or in solution are widely used and have made very valuable contributions to understanding many biological processes. However, replicating the cellular environment in vitro is difficult. In vivo observations of three-dimensional structures, dynamics and interactions of proteins are required for fully understanding the structural basis of their functions inside cells. Investigating proteins "at work" in a living environment at atomic resolution is thus a major goal of molecular biology. Recent developments in NMR hardware and methodology have enabled the measurement of high-resolution heteronuclear multi-dimensional NMR spectra of macromolecules in living cells (in-cell NMR). Various intracellular events such as conformational changes, dynamics and binding events have been investigated by this method. However, the low sensitivity and short life time of the samples have so far prevented the acquisition of sufficient structural information to determine protein structures by in-cell NMR. Recently we presented the first three-dimensional protein structure calculated exclusively on the basis of information obtained in living cells. The in-cell NMR approach can thus provide accurate high-resolution structures of proteins in living environments.

Fig. 3: Protein structure determined in living cells: The first three-dimensional protein structure calculated exclusively on the basis of information obtained in living cells was solved by in-cell NMR for the putative heavy metal-binding protein TTHA1718 from Thermus thermophilus HB8 overexpressed in E. coli cells. A major hurdle for determining in-cell NMR structures is the limited lifetime of the cells inside the NMR sample tube. Standard NMR experiments usually require 1–2 days of data collection, which is an unacceptably long time for live cells. This time could be shortened to 2–3 hours by preparing a fresh sample for each experiment and by applying a nonlinear sampling scheme in combination with maximum entropy processing for the indirectly acquired dimensions.

PARALLELES UND GRID-BASIERTES HOCHLEISTUNGSRECHNEN IN DER STRUKTURBIOLOGIE

Die verfügbare Rechenleistung begrenzt viele Answendungen in der Strukturbiologie. Die bestmögliche Ausnutzung der vorhandenen vielfältigen Rechnerarchitekturen, Compiler und Betriebssysteme ist eine anspruchsvolle Aufgabe, für die die maximale Leistung mit Anforderungen an die Portabilität und den Unterhalt der Software in Einklang gebracht werden muss. Softwareentwicklungsstrategien, die optimale Effizienz der wissenschaftlichen Algorithmen mit voller Einhaltung von Standards erreichen, sind insbesondere für rechenaufwändige vollautomatische Proteinstrukturberechnungen wichtig. Ein Web-basierter externer Zugang zu NMR Strukturbestimmungsalgorithmen ist, attraktiv, um die Nutzergemeinschaft zu stärken und deren Sichtbarkeit zu erhöhen, wie für viele Bioinformatikwerkzeuge, die im Allgemeinen nicht durch ihre oft (technisch) unerfahrenen Benutzer lokal installiert werden, sondern auf professionell und konsistent eingerichteten dezentralen Servern angeboten werden.

HIGH-PERFORMANCE PARALLEL AND GRID COMPUTING IN STRUCTURAL BIOLOGY

Computation power is limiting many applications in structural biology. Making the best possible use of the available variety of hardware architectures, compilers and operating systems makes it non-trivial to reconcile maximal efficiency with code portability and maintainability. Software engineering strategies to achieve optimal performance of scientific codes while fully adhering to standards are important, in particular for computation-intensive fully automated protein structure calculations. A web-based service for external access to NMR structure calculation algorithms will be attractive to strengthen the user base and to generate additional visibility in a similar way as for many bioinformatics tools, which are in general not installed locally by their often technically inexperienced users but run in a consistent way on professionally maintained, decentralized servers.

LITERATUR / REFERENCES

[1] Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, A. M. & Güntert, P. (2006). Optimal isotope labelling for NMR protein structure determinations. Nature 440, 52–57.
[2] Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Wälchli, M., Smith, B. O., Shirakawa, M., Güntert, P. & Ito, Y. (2009). Protein structure determination in living cells by in-cell NMR spectroscopy. Nature 458, 102-105.
[3] Koglin, A., Löhr, F., Bernhard, F., Rogov, V. V., Frueh, D. P., Strieter, E. R., Mofid, M. R., Güntert, P., Wagner, G., Walsh, C. T., Marahiel, M. A. & Dötsch, V. (2008). Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. Nature 454, 907-911.
[4] Güntert, P. (2009). Automated structure determination from NMR spectra. Eur. Biophys. J. 38, 129-143.
[5] López-Méndez, B., & Güntert, P. (2006). Automated protein structure determination from NMR spectra. J. Am. Chem. Soc. 128, 13112–13122.

KONTAKT / CONTACT:

Prof. Dr. Peter Güntert

Fachbereich Biochemie, Chemie und Pharmazie
Institut für Biophysikalische Chemie
Max-von-Laue-Str. 9
D-60438 Frankfurt am Main

Telefon: ++49 (0)69 798-29621
Fax: ++49 (0)69 798-29643
Email: guentert@em.uni-frankfurt.de
http://www.bpc.uni-frankfurt.de/guentert/