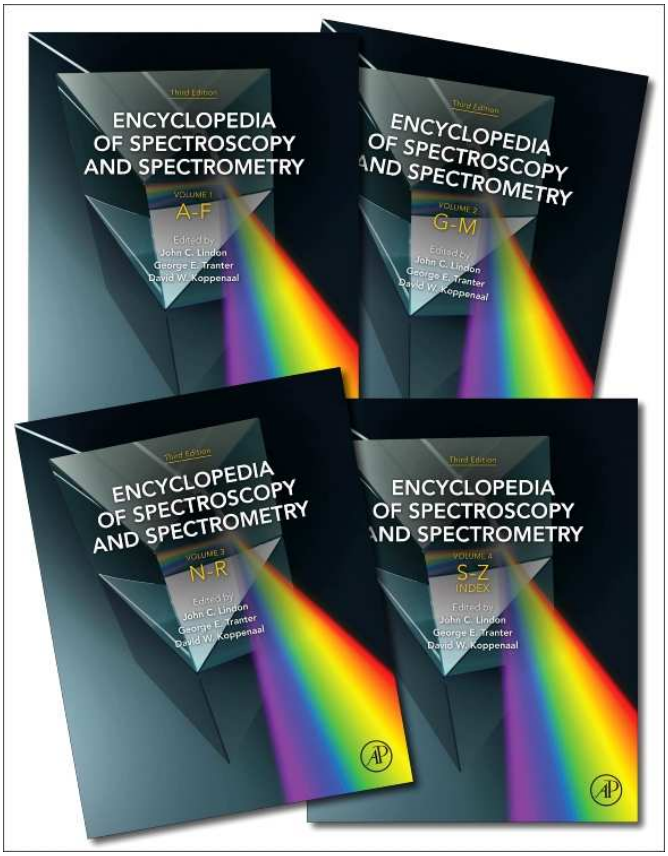

Marius Clore G. (2017) Structure Determination of Large Macromolecular Complexes Using NMR. In: Lindon, J.C., Tranter, G.E., and Koppenaal, D.W. (eds.) The Encyclopedia of Spectroscopy and Spectrometry, 3rd edition vol. 4, pp. 316-318. Oxford: Academic Press.

Structure Determination of Large Macromolecular Complexes Using NMR

G Marius Clore, National Institutes of Health, Bethesda, MD, United States

Abbreviations

EI	Enzyme I of the PTS
EIC	The C-terminal domain of enzyme I
EIN	The N-terminal domain of enzyme I
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser enhancement
PTS	Phosphotransferase system
RDC	Residual dipolar coupling
SAXS	Small angle X-ray scattering
WAXS	Wide angle X-ray scattering

Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a powerful solution technique that permits one to obtain structural information on proteins and their complexes at atomic resolution. Many reviews have been written on conventional NMR structure determination, the mainstay of which relies on a large number of short (<6 Å) interproton distance restraints derived from nuclear Overhauser enhancement (NOE) measurements. However, conventional strategies of NMR structure determination fall short when dealing with large systems that tumble slowly in solution, resulting in broad lines that preclude complete resonance assignments. Here, we summarize some recent developments that show how judicious application of hybrid methodology combining NMR with other biophysical techniques such as solution X-ray scattering, together with prior knowledge of the structures of individual domains or proteins, can significantly extend the molecular weight range and complexity of structural problems that can be addressed by NMR.

Brief Background on Modern Conventional Structure Determination

The general strategy used to solve NMR structures of proteins by conventional means relies on obtaining near-complete backbone and side-chain ^{1}H, ^{15}N, and ^{13}C assignments using three-dimensional (3D) double- and triple-resonance heteronuclear correlation experiments to connect nuclear spins via through-bond scalar couplings. With assignments in hand, NOEs between protons close (<6 Å) together in space can then be assigned using a variety of heteronuclear-separated 3D and 4D experiments, thereby minimizing NOE assignment ambiguities. Once a set of solidly assigned NOEs are available, initial atomic structures can be calculated using methods such as simulated annealing and subsequently refined in an iterative manner using the structures to resolve ambiguous NOE assignments. The interproton distance restraints derived from NOE data can be supplemented by torsion angle restraints derived from backbone ^{1}H, ^{15}N, and $^{13}C\alpha/\beta$ chemical shift data and heteronuclear and homonuclear three-bond J couplings. Further improvements in accuracy can be obtained by making use of residual dipolar couplings (RDCs) measured in weakly aligned media (such as dilute bicelles or phages) to provide bond vector orientations relative to an external alignment tensor.

While improvements in spectrometer technology (eg, the advent of cryoprobe technology that increases the signal-to-noise ratio three- to fourfold, and higher field magnets that increase spectral resolution, thereby reducing spectral overlap) have reduced the measurement time to some extent, collecting all the data necessary to solve a protein NMR structure at high accuracy using the conventional approach may still require many months. Likewise, improvements in spectral analysis software and structure calculation algorithms have permitted the introduction of some degree of automation, but extensive human intervention is still necessary to fully and reliably interpret the data in all but the simplest of cases.

Approaches Designed to Speed Up the Structure Determination of Protein Complexes

The structure of a 40 kDa complex between the N-terminal domain of enzyme I (EIN) and the histidine-containing phosphocarrier protein HPr, the first complex in the bacterial phosphotransferase system (PTS), was published in 1999 using a conventional approach based on NOE and RDC data. The time required to acquire all the relevant NMR data was ~3500 h. It was soon realized, however, that this process could be sped up by many orders of magnitude since the structures of the two proteins were already known, and it was evident from both the small chemical shift perturbations upon complex formation and the RDC data that the structures of the two proteins were unchanged within coordinate errors upon complexation. Thus, all that was really required to determine the structure of the complex was a few intermolecular NOEs measured exclusively from 3D heteronuclear-separated and heteronuclear-filtered experiments to provide translational and orientational information, coupled with backbone amide RDC measurements to provide orientational information. With this information in hand, one can then use combined rigid-body/torsion angle simulated annealing driven by the intermolecular NOE data and RDCs to dock the two proteins accurately, treating the protein backbones as rigid bodies but giving the interfacial side chains torsional degrees of freedom. This approach was subsequently used to determine the other eight cytoplasmic complexes of the four sugar branches of the PTS. Further refinements in calculational methodology can even permit

Encyclopedia of Spectroscopy and Spectrometry, Third Edition, (2017), vol. 4, pp. 316-318

accurate docking based on RDCs and a set of highly ambiguous distance restraints derived from chemical shift mapping.

The rigid-body docking method does not require the assumption that the complete backbones of the constituent proteins remain identical in the complex and the free states. If portions of the protein backbone undergo conformational rearrangements upon complexation, intramolecular NOE data can be focused on those specific regions of the backbone that are then given torsional degrees of freedom during the course of simulated annealing.

Hybrid Approaches to Large Multidomain Complexes

Conventional NMR strategies are generally applicable to systems below about 50 kDa. For larger systems, broad lines owing to slow tumbling generally preclude complete assignments. As a consequence, only sparse structural information can be obtained, necessitating the use of hybrid technologies. An example is provided by the 128 kDa enzyme I dimer, as well as its 146 kDa complex with HPr.

Enzyme I comprises two domains, an N-terminal domain (EIN) containing the active site histidine and a C-terminal dimerization domain (EIC) (Fig. 1). The EIN domain is itself divided into two subdomains, a helical subdomain (EIN^{α}) and a mixed helix/sheet subdomain ($EIN^{\alpha/\beta}$). While the line widths for the EIN domain are reasonable, those for the EIC domain are quite severely broadened. As a result, although the vast majority of resonances for the backbone amide groups of EIN in the context of full-length dimeric EI can be transferred from those obtained for the isolated EIN domain, only a few assignments could be obtained for the EIC domain. The strategy used to determine the structure of free EI made use of RDCs measured in phage to provide domain orientations and small and wide angle X-ray scattering (SAXS/WAXS) to provide shape and size information. First, while the RDCs measured for EIN^{α} and $EIN^{\alpha/\beta}$ subdomains in full-length EI agreed well with the corresponding coordinates from both isolated EIN and the closed structure of a phosphoryl transfer intermediate of EI, only the orientation of EIN^{α} and $EIN^{\alpha/\beta}$ seen in isolated EIN was consistent with the RDC data. Thus, one can immediately conclude that the EIN domain undergoes a large (90°) rigid-body conformational rearrangement of the α and α/β subdomains between the free state and the phosphoryl transfer intermediate. Second, the structure of the closed phosphoryl transfer intermediate was not consistent with the SAXS/WAXS data ($\chi^2 \sim 128$), implying an additional large rigid-body rotation of the $EIN^{\alpha/\beta}$ subdomain relative to the EIC dimer. The structure of EI was then solved by simulated annealing driven by the RDC and SAXS/WAXS data in which EIN and EIC were treated as rigid bodies, and the linker connecting EIN to EIC was given Cartesian degrees of freedom. The result is an open structure that satisfies both the RDC and SAXS/WAXS data within experimental error (RDC *R*-factors comparable with those for the individual subdomains and SAXS/WAXS $\chi^2 \sim 1$) and correctly predicts the WAXS data for scattering vector *q*-values above 0.4/Å not included in the simulated annealing calculations. Exactly the same strategy could be used to solve the EI–HPr complex as the RDC data indicated that the orientation of HPr bound to the EIN domain of EI was the same as that in the complex of HPr with isolated EIN. These studies demonstrated that the transition from the closed phosphoryl transfer intermediate to the free (and HPr-complexed) states of EI involves two large rigid-body rearrangements comprising a ~90° reorientation of EIN^{α} relative to $EIN^{\alpha/\beta}$ and a ~70° reorientation of $EIN^{\alpha/\beta}$ relative to EIC. These results make perfect physical sense. The closed structure is required to allow in-line phosphoryl transfer from phosphoenolpyruvate bound to EIC to His189 of $EIN^{\alpha/\beta}$. However, in the closed structure, the distance between His189 in $EIN^{\alpha/\beta}$ and the active site His15 of HPr is far too large to allow for subsequent phosphoryl transfer to HPr that can only occur in the open state.

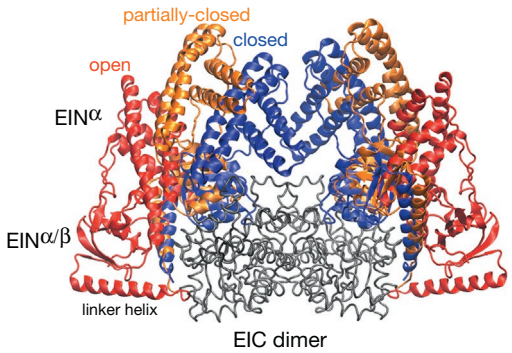

Fig. 1 Structure of the open (*red*), partially closed (*orange*), and closed (*blue*) forms of *E. coli* enzyme I of the bacterial phosphotransferase system. The EIC dimerization domain is the same in all structures and shown as a gray tube. The open (PDB ID 2KX9), partially closed (PDB ID 2N5T), and closed (PDB ID 2HWG) conformations of the EIN domain are shown as red, orange, and blue ribbons, respectively. The closed state corresponds to a phosphoryl transfer intermediate in which in-line phosphoryl transfer between phosphoenolpyruvate bound to EIC and the active site His189 in the $EIN^{\alpha/\beta}$ subdomain can occur. The open state allows in-line phosphoryl transfer from His189 on the $EIN^{\alpha/\beta}$ subdomain to HPr bound to the EIN^{α} subdomain. The partially closed state represents an intermediate along the open-to-closed transition that is occupied at around 50% in a complex of the EI(H189A) mutant with phosphoenolpyruvate. Adapted from Schwieters, C. D.; Suh, J. Y.; Grishaev, A.; Ghirlando, R.; Takayama, Y.; Clore, G. M. *J. Am. Chem. Soc.* **2010**, *132***See Further Reading for more detailed information, Schwieters et al, 2010 and";, 13026–13045; Venditti, V.; Schwieters, C. D.; Grishaev, A.; Clore, G. M. Proc. Natl. Acad. Sci. USA **2015,** *112*(37), 11565–11570.

This same strategy was then applied with a twist to study the complex of the His189A mutant of EI complexed to phosphoenolpyruvate (PEP). In this instance, both the RDC and SAXS/WAXS data indicated that the structure was predominantly closed, but no single structure could be found that was capable of satisfying both the RDC and SAXS data simultaneously. Thus, the EI(H189A)–PEP complex had to exist as an ensemble of states in solution. Two extensions were made over the previous study on free EI described in the preceding text. First, the RDC data were measured in a neutral bicelle medium where alignment is purely steric, and hence, the alignment tensor can be directly calculated from the molecular shape determined from the coordinates. Second, an ensemble approach was used in which the EIN domain was represented by two states in rapid exchange with one another in solution. The result was the discovery of a partially closed state or intermediate in

dynamic equilibrium with the closed state in an approximately 50:50 mixture (Fig. 1). A similar calculational and experimental approach was used to determine the conformational space sampled by the N-terminal domain of HIV-1 capsid relative to its C-terminal dimerization domain, but in this instance, the N-terminal domain samples a large region of conformational space that necessitates the use of a larger ensemble size.

Concluding Remarks

Solving structures of large macromolecular complexes in solution, as well as structures of multidomain proteins in which domain orientations are not unique, can be accomplished by the combined use of NMR and solution X-ray scattering data. Because domains of known structure can be treated as rigid bodies, such problems can be tackled reliably using relatively sparse experimental data without the necessity of resorting to conventional full-blown NMR structure determination that is both timeconsuming and increasingly difficult with increasing size and molecular complexity of the systems under study.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

See also: Chemical Exchange Effects in NMR; Chemical Shift and Relaxation Reagents in NMR; CIDNP Applications; Electron Paramagnetic Resonance of Membrane Proteins; NMR Applications, ^{15}N; NMR Methods, ^{13}C; NMR Parameter Survey, ^{13}C; NMR Spectroscopy, ^{14}N and ^{15}N; NMR Spectroscopy of Nanoparticles; Nuclear Overhauser Effect; Peptides and Proteins Studied Using Mass Spectrometry; Proteins Studied by NMR; Residual Dipolar Couplings in Small-Molecule NMR.

Further Reading

Cavanagh J, Fairbrother WJ, Palmer AG, and Skelton NJ (2007) *Protein NMR Spectroscopy: Principles and Practice*, 2nd ed. San Diego: Elsevier Academic Press.
Clore GM (2000) *Proc. Natl. Acad. Sci. USA* 97: 9021–9025.
Clore GM and Gronenborn AM (1989) *Crit. Rev. Biochem. Mol. Biol.* 24: 479–564.
Clore GM and Gronenborn AM (1991) *Science* 252: 1390–1399.
Clore GM and Gronenborn AM (1998) *Trends Biotechnol.* 16: 22–34.
Clore GM and Schwieters CD (2003) *J. Am. Chem. Soc.* 125: 2902–2912.
Clore GM and Venditti V (2013) *Trends Biochem. Sci.* 38: 515–530.
Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM, and Clore GM (1999) *Nat. Struct. Biol.* 6: 166–173.
Herrmann T, Guntert P, and Wuthrich K (2002) *J. Biomol. NMR* 24: 171–189.
Kuszewski J, Schwieters CD, Garrett DS, Byrd RA, Tjandra N, and Clore GM (2004) *J. Am. Chem. Soc.* 126: 6258–6273.
Linge JP, Habeck M, Rieping W, and Nilges M (2003) *Bioinformatics* 19: 315–316.
Schwieters CD and Clore GM (2014) *Prog. Nucl. Magn. Reson. Spectrosc.* 80: 1–11.
Schwieters CD, Kuszewski JJ, and Clore GM (2006) *Prog. Nucl. Magn. Reson. Spectrosc.* 48: 47–62.
Schwieters CD, Suh JY, Grishaev A, Ghirlando R, Takayama Y, and Clore GM (2010) *J. Am. Chem. Soc.* 132: 13026–13045.
Venditti V, Schwieters CD, Grishaev A, and Clore GM (2015) *Proc. Natl. Acad. Sci. USA* 112(37): 11565–11570.
Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*. New York: John Wiley & Sons.