NOESY-based strategy for assignments of backbone and side chain resonances of large proteins without deuteration

Yingqi Xu
National University of Singapore
Yu Zheng
National University of Singapore
Daiwen Yang
National University of Singapore

Method Article

Keywords: NMR, assignment, large protein, NOESY, protein structure, protein NMR

Posted Date: May 11th, 2006

DOI: https://doi.org/10.1038/nprot.2006.410

Abstract

Introduction

NMR spectroscopy has been widely used to determine high resolution structures of proteins and protein complexes in solution[1, 2]. However, it is is not suitable for larger proteins \(> 30 kDa) due to increased signal decay rates. Deuterium labeling of protein samples overcomes this problem[3], but deuteration makes sample preparation much harder and structure determination more difficult[4, 5]. Here we introduce a strategy that uses non-deuterated samples[6]. Backbone and side chain resonances of large proteins are assigned from 3D TROSY-HNCA[7], 3D MQ-CCH-TOCSY[8] and 4D ^{15}N,^{13}C-edited NOESY[9] spectra. The three spectra with reasonable quality can be obtained within 11 days using a uniformly ^{13}C,^{15}N-labeled sample at a protein concentration of 1 mM. Unlike the conventional strategy mainly relying on through-bond correlation experiments, the strategy used here mainly uses a through-space correlation experiment \(4D ^{15}N,^{13}C-edited NOESY) to obtain intra-residue and sequential correlations, which are separated from other inter-residue NOE correlations using the 3D HNCA and CCH-TOCSY experiments. Similar to the conventional strategy, sequential assignment is achieved from matching intra-residue and sequential correlations. Using this strategy, one can obtain nearly complete backbone assignments and >80% side chain assignments for monomeric proteins < 42 kDa and multimeric proteins < 65 kDa.

Reagents

• ^{13}C and ^{15}N labeled protein sample with ~1 mM monomer and 5-10% D2O; if a cryoprobe is available, lower concentration is needed

Equipment

• 800 MHz NMR spectrometer; for multimeric proteins with less complexity, a 500 MHz or higher field machine equipped with a cryoprobe can be used as an alternate • A personal computer \(Linux operating system) • NMRPipe and Sparky softwares; other softwares with similar functions can be used

Procedure

Recording data • TIMING ~11 days 1. Record 3D TROSY-HNCA, 3D MQ-CCH-TOCSY and 4D ^{13}C, ^{15}N-edited NOESY for resonance assignment. Use half-point delays for the evolution periods of C^{α} spins \(t_1) in the HNCA, ^{13}C spins \(bonded with the directly detected ^{1}H, t_2) in the MQ-CCH-TOCSY, and both ^{13}C and ^{15}N spins \(t_2 and t_3) in the 4D NOESY. In this way, the peaks aliased with odd times have opposite signs to those aliased with even times. Record a 2D ^{1}H-^{15}N TROSY-HSQC with enough spectral width in the ^{15}N dimension to facilitate the identification of peaks aliased in the ^{15}N dimension in 3D HNCA and 4D ^{13}C, ^{15}N-edited NOESY. **Processing data** • TIMING ~1 day 2. Process the data with NMRPipe. Use linear

prediction to improve the resolution of all indirectly detected dimensions. Covert the processed spectra into ucsf format for further analysis with Sparky. **Picking peaks and grouping peaks into clusters** • TIMING ~1-2 days 3. Use peak picking functions in Sparky to pick HNCA peaks automatically. Group them into clusters according to \(H^N, N) chemical shifts. Usually one cluster contains two peaks: one from the intraresidue C^α, another from the C^α of the preceding residue. However, some clusters contain only one peak due to degenerate C^α chemical shifts, while some clusters may contain more than two peaks due to the overlap of amide chemical shifts. Identify peaks from NH_2 groups according to their paired patterns and unusual "C^α" chemical shifts. These peaks can be excluded during backbone assignment. Check the number of clusters identified and the expected number and ensure that there is not a large difference. 4. Pick peaks in 4D ^{13}C,^{15}N-edited NOESY manually or automatically. Group NOE peaks into their corresponding clusters according to \(H^N, N) chemical shifts of individual clusters. Due to lower resolution of the 4D NOESY, some peaks cannot be grouped into only one cluster unambiguously. These peaks will remain ungrouped. 5. Correct chemical shifts of the aliased peaks in 3D TROSY-HNCA and 4D NOESY with Sparky. 6. Pick peaks in CCH-TOCSY. **Identifying and classifying spin-systems** • TIMING ~5-10 days 7. Each cluster is characterized by a pair of \(H^N, N) chemical shifts. Typically, one cluster contains two HNCA peaks and a number of HC-NH NOE peaks. First, initialize a spin-system from one HNCA peak in a given cluster. Second, among the NOE peaks in this cluster, identify the peak that matches the HNCA peak in ^{13}C chemical shift within a certain tolerance \(e.g., ~0.3 ppm). If only one peak is identified, this peak is considered as a sequential or intra-residue $C^\alpha H^\alpha$-NH NOE[10]. Note that there are two α protons for glycine residues and thus two peaks with identical C^α shifts may be observed. Third, display the TOCSY slice defined by the \(H, C) chemical shifts of each NOE peak in this cluster and check if the CH spin-pair of the NOE peak correlates with the C^α spin of this initial spin-system. If yes, the NOE peak is regarded as a sequential or intraresidue side chain NOE in this spin-system. The ^{13}C spin of the CH spin-pair should also correlate with $C^\alpha H^\alpha$ of this spin-system via TOCSY. In some cases, this correlation cannot be detected because it is too weak or close to the intense water signal. For spin-systems of long side chains, a few more CH spin-pairs may be added into the spin-systems according to their TOCSY correlations with side chain CH spin-pairs that are already identified. 8. In cases where only one HNCA peak is picked for one cluster due to degenerate C^α chemical shifts and two NOE peaks match this HNCA peak, build two spin-systems using the two NOE peaks. The NOE peaks belonging to these two spin-systems can be separated because side chain ^{13}C spins correlate with two non-degenerate $C^\alpha H^\alpha$ spin-pairs in CCH-TOCSY, provided that ^{13}C-$C^\alpha H^\alpha$ correlations are detectable. 9. For each spin-system, check whether the CH spin-pairs correlate with each other in CCH-TOCSY and whether the correlations can be explained by one of the 20 amino acids in terms of number of peaks and chemical shifts. 10. Classify a spin-system by residue type according to its C^α chemical shift and NOE peaks. Typically it is possible to unambiguously recognize most glycine, alanine, threonine, valine, isoleucine, and leucine, which account for ~42% amino acid residues in a protein. Glycine can be identified according to its small C^α chemical shift, and sometimes two H^α peaks. Alanine can be identified based on its strong intraresidue $C^\beta H^\beta$-NH NOE peak, strong C^α-$C^\beta H^\beta$ TOCSY peak, and characteristic C^α and C^β chemical shifts. For

threonine, valine, isoleucine and leucine, classification can be easily done based on their characteristic side chain chemical shifts. In particular, the methyl groups of these residues always give rise to well resolved TOCSY correlations with other ^{13}C spins. Some serine, lysine, arginine, and proline can be distinguished, provided that the $C^{\beta}H^{\beta}$ spin-pair for serine, $C^{\beta}H^{\beta}$, $C^{\gamma}H^{\gamma}$ and $C^{\delta}H^{\delta}$ spin-pairs for lysine, arginine and proline can be identified. Classify other residues into three groups: \(Asp, Asn, Phe, Tyr), \(Gln, Glu, Met) and \(Cys, Trp, His) according to both ^{1}H and ^{13}C chemical shifts. If a spin-system contains too few peaks to be classified, leave it without classification. Usually classification of about ~35-50% spin-systems is sufficient for obtaining backbone assignment. **Matching spin-systems to form fragments and Mapping fragments** • TIMING ~1-3 days 11. Link clusters to build fragments through matching spin-systems. Two spin-systems match each other when they have i). identical C^{α} chemical shifts, ii). the largest number of matched NOE peaks, and iii). no inconsistent \(C, H) chemical shifts. 12. Compare spin-system types of a connectivity fragment with the protein sequence. If only one segment in the sequence matches the fragment in residue type at every amino acid position, map this fragment onto the segment of the sequence. Note that it is unknown which end of the fragment is the N-terminus since intraresidue and sequential spin-systems are not distinguished. The intraresidue HNCA peak is usually stronger than the sequential one. If a mapping is inconsistent with this, it may contain errors and need to check more carefully. 13. Use common NOEs shared by two clusters to resolve ambiguities in sequential connectivities for unassigned spin-systems[6]. If two spin-systems have a true sequential relationship, no mismatch of intra-residue and sequential correlations should exist. On the basis of this fact, some ambiguities can be removed too. 14. Repeat steps 11-13 and assign the unassigned fragments and clusters in an iterative way. 15. Note that some initial errors may be introduced during the processes of identifying, classifying and matching spin-systems and mapping fragments to the sequence. Most errors can be recognized during the process of assignment and will not lead to wrong final assignments. Those lead to wrong initial assignments should be recognized when some other clusters can't be assigned[6]. **Side chain assignments** • TIMING ~2-7 days 16. Most side chain chemical shifts are ready to be extracted according to the identified spin-systems and backbone assignments. Assign some more side chain resonances by comparing two ^{13}C,^{15}N-edited NOESY planes defined by two adjacent amides, and by exploring the MQ-CCH-TOCSY more extensively[10]. 17. Assign NH_2 chemical shifts according to i) C^{β}-NH_2 and C^{γ}-NH_2 correlations in the HNCA spectrum for asparagine and glutamine respectively, ii) $H^{\beta}C^{\beta}$-NH_2 and $H^{\gamma}C^{\gamma}$-NH_2 /$H^{\beta}C^{\beta}$-NH_2 correlations in the 4D ^{13}C,^{15}N-edited NOESY for asparagine and glutamine respectively. 18. Assign $C^{\delta}H^{\delta}$ groups of aromatic residues according to their NOE correlations with intraresidue and sequential amides[11]. Other aromatic resonances can be assigned later with a 4D ^{13}C,^{13}C-edited NOESY \(7 days for data collection) or a 3D $^{15}N/^{13}C$-simutaneous NOESY \(3 days for data collection). **NOE assignment and structure determination** • TIMING ~2-6 weeks 19. With a relatively complete chemical shift list, assign other NOE peaks in the 4D ^{13}C,^{15}N-edited NOESY and ^{13}C,^{13}C-edited NOESY. Follow the conventional procedure for NOE assignment and structure calculation. The chemical shifts of a few more side chain atoms may be assigned during this process.

Troubleshooting

Some amide groups have degenerate \(H^N, N) chemical shifts, which lead to some clusters containing more than one residues. These clusters can be identified since they contain more than two HNCA peaks or more than one HNCO peak. They are treated as normal clusters during spin-system identification. However, when linking clusters through matching spin-systems, we place these clusters at the ends of fragments. In this way, each spin-system in such a cluster is linked and mapped independently. When the number of degenerate amides is so large that many fragments are too short to be mapped onto the sequence, backbone assignment may be difficult to be obtained. For very large proteins, the CCH-TOCSY may give weak signals for many correlations, especially for those involved in H^{α} and H^{β} spins, and the 4D ^{13}C,^{15}N edited-NOESY may also lack many intraresidue and sequential NOE peaks. This leads to spin-systems with very few NOE peaks, which are difficult to be classified and matched unambiguously. In such cases, one may rely on C^{α} chemical shifts and common NOE peaks between two clusters, although there will be more ambiguity in spin-system matching and mapping.

Anticipated Results

For monomeric proteins smaller than $_{42\ kDa\ or\ multimeric\ proteins\ smaller\ than}$ 65 kDa, the 3D and 4D spectra used in our method can be obtained with reasonable quality on an 800 MHz NMR spectrometer equipped with a normal probe using a sample of ^{13}C,^{15}N-labeled protein at a monomer concentration of $_{1\ mM.\ The\ number\ of\ the\ clusters\ obtained\ from\ the\ spectra\ should\ be\ very\ close\ to\ the\ expected\ one.\ Most\ spin-systems\ generated\ from\ the\ spectra\ should\ contain\ at\ least\ one\ HC-NH\ NOE\ peak.}$ 60% spin-systems can be classified by residue types. ~50% spin-systems can be unambiguously matched to their corresponding partners. Nearly complete backbone assignment can be obtained, provided that all amides give rise to ^{1}H-^{15}N HSQC correlations. On the basis of the backbone assignment, >80% side chain resonances can be assigned from the 3D MQ-CCH-TOCSY and 4D ^{13}C,^{15}N-edited NOESY spectra. With these assignments, NOE assignments should be possible and a high resolution structure can be determined.

References

1. Wuthrich, K. _NMR of Proteins and Nucleic Acids_, \(Wiley, New York, 1986). 2. Cavanagh, J., Palmer, A. G., Fairbrother, W. & Skelton, N. _Protein NMR Spectroscopy: Principles and Practice_ \(Academic, San Diego, 1996). 3. LeMaster, D.M. & Richards, F.M. NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteriation. _Biochemistry_ **27**, 142-150 \(1988). 4. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. _Nature Protocols_ **1**, 749-754 \(2006). 5. Kainosho, K. et al. Optimal isotope labelling for NMR protein structure determinations. _Nature_ **440**, 52-57 \(2006). 6. Xu, Y.Q., Zheng, Y., Fan, J.S. & Yang, D.W. A novel strategy for solution structure determination of large proteins without deuteration. _Nature Methods_ \(in press). 7. Yang, D.W. & Kay, L.E. TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. _J. Am. Chem. Soc._ **121**,

2571-2575 \(1999). 8. Yang, D.W., Zheng, Y., Liu, D. & Wyss, D.F. Sequence-specific assignments of methyl groups in high-molecular weight proteins. _J. Am. Chem. Soc._ **126**, 3710-3711 \(2004). 9. Kay, L.E., Clore, G.M., Bax, A. & Gronenborn, A. M. Four-dimensional heteronuclear triple-resonance NMR spectroscopy of Interleukin-1β in solution. _Science_, **249**, 411-414 \(1990). 10. Xu, Y.Q., Lin, Z., Ho, C. & Yang, D.W. A general strategy for the assignment of aliphatic side-chain resonances of uniformly 13C,15N-labeled large proteins. _J. Am. Chem. Soc._ **127**, 11920-11921 \(2005). 11. Lin, Z., Xu, Y.Q., Yang, S. & Yang, D.W. Sequence-specific assignment of aromatic resonances of uniformly 13C,15N-labeled proteins by using 13C- and 15N-edited NOESY spectra. _Angew. Chem. Int. Ed._ **45**, 1960-1963 \(2006).