

DISS. ETH NO. 23431

Protein structure determination and methods
development for Solid - State NMR

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
FRANCESCO RAVOTTI
Physics University of Turin

born on 13.05.1985

Citizen of
Italy

Accepted on the recommendation of

Prof. Dr. Beat H. Meier

Prof. Dr. Roland Riek

Dr. Anja Böckmann

2016

Abstract

Proteins are large macromolecules consisting in one or more chains of amino acids. Their role is of great importance in cells, accomplishing specific tasks which are coded in the genes. To understand the function of a protein, knowing its 3D structure is a key point. Solid - state NMR provides powerful techniques that can be applied to determine the structure of protein fibrils and protein complexes that cannot be solved with standards methods, as X - ray crystallography and solution - state NMR. In fact, solid - state NMR does not require the protein to form nice crystal for diffraction or high solubility of the sample.

Structure calculation based on NMR data combines chemical information from the sample (amino acid sequence, bond length and bond angles) with restraints and constraints derived from experimental data. The most commonly used classes of constraints and restraints are torsion angle restraints and distance restraints and constraints. These data can be obtained from the analysis of solid - state NMR spectra.

In the first part of the thesis we present two examples of applications of solid - state NMR methods for protein structure determination. In Chapter 1 we present the structure of the Amyloid β (1 - 42) fibrils. This protein belongs to the amyloid fibrils family, which play a role in different neurodegenerative diseases, including Alzheimer's disease. Firstly, we obtained the chemical shifts of the fibrils from 2D and 3D spectra. These spectra show a well formed fibrils, with ^{13}C line width in the order of 0.5 ppm. We identified the presence of 5 β - strands in the protein and we deduced the parallel in - register β sheets architecture. The stretch of residues 11 - 14 are not rigid and invisible in solid state NMR spectra, therefore we calculated the structure of the rigid part of the protein (residues 15 - 42).

In Chapter 2 we introduced a new hybrid approach that combines solid - state NMR methods with cryo - electron microscopy data. We described this method and applied it for the structure determination of the filaments composed by the PYRIN domain of the mouse ASC protein. The two techniques provide complementary information, leading to an overall comprehensive description extending beyond the power and resolution of cryo - EM or solid - state NMR alone. In the first step of this approach, we used the dihedral angle restraints obtained from solid - state NMR together with the structural information of the cryo - EM density map. This led to a well converged structure of the backbone. To increase the precision of this approach, we refined the initial structure

adding solid - state NMR distance restraints and distance constraints. These two final steps were important to obtain more information on the side chains position.

In the second part of the thesis, we proposed two new approaches that can be applied in solid - state NMR, efficient in the fast MAS regime. In this regime, the main factor that limits resolution in ^{13}C spectra is the presence of the homonuclear J - coupling. This interaction is challenging to remove during the detection of the NMR signal. In Chapter 3 we discuss the applications of a family of J - decoupling sequences, the Spin State Selective Excitation, to obtain high resolution ^{13}C spectra. We present here possible applications of these sequences to obtain highly resolved correlation spectra that can be used for the assignment of proteins backbone atoms.

In Chapter 4 we present a new approach to obtain highly resolved ^{15}N - ^{13}C intermolecular correlations. These correlations can be used in structure determination of proteins as source of information on the stacking of the monomers. The 3D TEDOR experiment we performed combined information on long range inter molecular ^{15}N - ^{13}C contacts with a intra residue ^{15}N - ^1H correlations. The strong restriction that the ^{15}N and ^1H have to belong to the same residue allowed to dramatically reduce the ambiguities of the inter molecular ^{15}N - ^{13}C constraints, which is usually a limiting factor in distance restraints experiment. We have used the so-obtained correlations in an automatic structure calculation of the HET-s fibrils.

Sommario

Le proteine sono grosse macromolecole, consistenti in una o più catene di amino acidi. Il loro ruolo è di fondamentale importanza nelle cellule, svolgendo compiti specifici che sono codificati nei geni. Per determinare la funzione di una proteina, la conoscenza della sua struttura in 3D è un punto fondamentale. RMN stato solido fornisce potenti tecniche che possono essere applicate per determinare la struttura di fibrille e complessi di proteine che non possono essere determinate con metodi standard, come cristallografia a raggi X e RMN stato liquido. Infatti, RMN stato solido non richiede che la proteina formi cristalli per la cristallografia e non richiede una sua alta solubilità.

Il calcolo di strutture basato su dati da RMN combina informazioni chimiche del campione (sequenza di amino acidi, lunghezze dei legami e angoli dei legami) con vincoli ottenute da dati sperimentali. Le classi più comuni di vincoli sono su angoli di torsione e vincoli di distanza. Questi dati possono essere ottenuti dall'analisi degli spettri di RMN stato solido.

Nella prima parte della tesi, presentiamo due esempi di applicazioni di metodi da RMN stato solido per la determinazione di strutture di proteine. Nel Capitolo 1 presentiamo la struttura delle fibrille dell'amiloide β (1 - 42). Questa proteina risiede nella famiglia di fibrille amiloidi che giocano un ruolo nelle malattie neurodegenerative, come la sindrome di Alzheimer. In primo luogo abbiamo ottenuto gli spostamenti chimici (chemical shifts) delle fibrille da spettri in 2D e in 3D. Questi spettri presentano fibrille ben formate con una linea spettrale di 0.5 ppm. Abbiamo identificato la presenza di 5 fogli β nella proteina e abbiamo dedotto la struttura parallela in β -registrata di fogli β . I residui tra 11 - 14 non sono rigidi e sono invisibili negli spettri di RMN stato solido, dunque abbiamo calcolato la struttura della parte rigida della proteina (residui 15 - 42).

Nel Capitolo 2 abbiamo introdotto un nuovo approccio ibrido che combina metodi di RMN stato solido con dati di crio - microscopia elettronica. Abbiamo descritto questo metodo e applicato per la determinazione di struttura dei filamenti composti dal dominio PYRIN della proteina ASC del topo. Le due tecniche forniscono informazioni complementari, portando a una descrizione di insieme che va oltre il potere di risoluzione della singola crio - microscopia elettronica o di RMN stato solido. Nel primo passo di questo approccio, abbiamo usato i vincoli per gli angoli diedri ottenuti da RMN stato solido con le informazioni strutturali della densità elettronica della crio - microscopia elettronica. Questo ha portato a una struttura convergente della catena dorsale (backbone). Per

aumentare la precisione di questo approccio, abbiamo rifinito la struttura iniziale aggiungendo vincoli di distanza da RMN stato solido. Questi due passi finali sono importanti per ottenere più informazioni sulla posizione delle catene laterali.

Nella seconda parte della tesi, abbiamo proposto due nuovi approcci che possono essere applicati nella RMN stato solido, efficienti nel regime di MAS veloce. In questo regime, il fattore principale che limita la risoluzione negli spettri di ^{13}C è la presenza del accoppiamento J omonucleare. Questa interazione è complessa da rimuovere durante la detezione del segnale di RMN. Nel Capitolo 3 discutiamo le applicazioni di una famiglia di sequenze per il disaccoppiamento J durante l'acquisizione in spettri di ^{13}C , le sequenze di eccitazione selettiva di stati di spin. Presentiamo qui possibili applicazioni di queste sequenze per ottenere spettri di correlazione ad alta risoluzione che possono essere usati per l'assegnazione di atomi del bacbone delle proteine.

Nel Capitolo 4 presentiamo un nuovo approccio per ottenere correlazioni intermolecolari tra ^{15}N - ^{13}C ad alta risoluzione. Queste correlazioni possono essere usate per il calcolo di strutture di proteine come una sorgente di informazioni sulla simmetria dei monomeri. L'esperimento in 3D TEDOR che abbiamo misurato combina informazioni su contatti a lunga distanza fra differenti molecole di ^{15}N - ^{13}C con correlazioni nello stesso residuo di ^{15}N - ^1H . La stringente restrizione che il ^{15}N e il ^1H debbano appartenere allo stesso residuo ha permesso di ridurre drasticamente le ambiguità dei vincoli intermolecolari tra ^{15}N - ^{13}C , che solitamente rappresenta un fattore limitante negli esperimenti di vincoli di distanza. Abbiamo utilizzato le correlazioni così ottenute in un calcolo di struttura automatico delle fibrille del HET-s.