Biomolecular structure and dynamics by NMR

Proteins of up to 200 amino acids, nucleic acids up to 50 nucleotides can be studied. The necessary concentrations are at least 0.1 mM for simple tests, 0.5 mM for more complex studies. Much depends on the molecular weight, folding etc. For solving 3D structures, isotope labeling with ^{15}N and ^{13}C is always needed for proteins and for oligonucleotides larger that approximately 30 nucleotides.

Specialized methods, some of them developed in our laboratory, for charactering inherently disordered proteins (IDPs) are available. We can measure spectra with up to 5 dimensions to assign the protein backbone and sidechain resonances.

Evaluation of the folding state of proteins and nucleic acids

The folding state of proteins can be easily determined from the dispersion of the chemical shifts in the ^{1}H-^{15}N HSQC spectrum. The self-processing module of FrpC protein is disordered in the free state (Figure 1b) and it folds into a structure upon binding to calcium (Figure 1a).

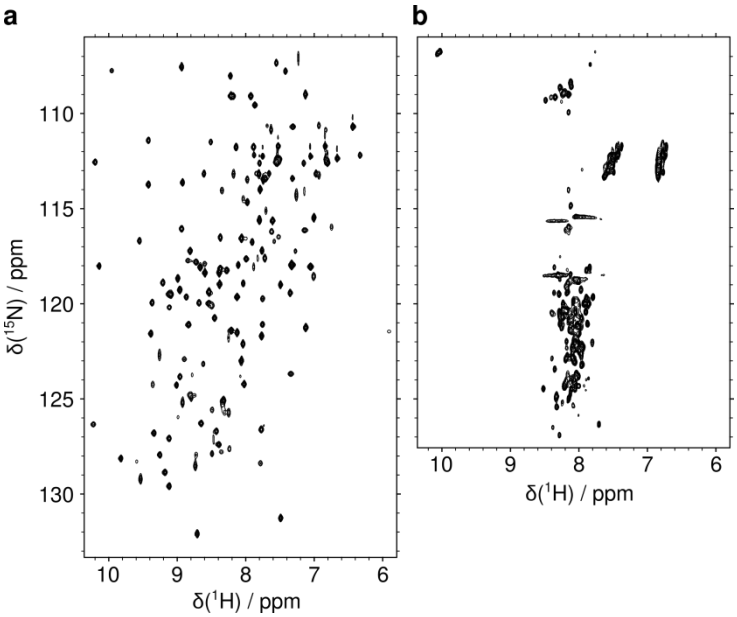

Fig. 1. Self-processing module of FrpC protein,

a) Ca bound, b) Ca free

NMR data courtesy Lukáš Žídek, protein sample courtesy Ladislav Bumba

The folding state of nucleic acids is best evaluated from the imino region of proton 1D spectrum. The spectrum of a well folded oligomer shows one peak for every imino proton bound in a base pair (Figure 2d). Higher number of signals in the spectrum indicates multiple structural forms (Figure 2a,b). A complex equilibrium of multiple forms give rise to broad unresolved peaks (Figure 2c).

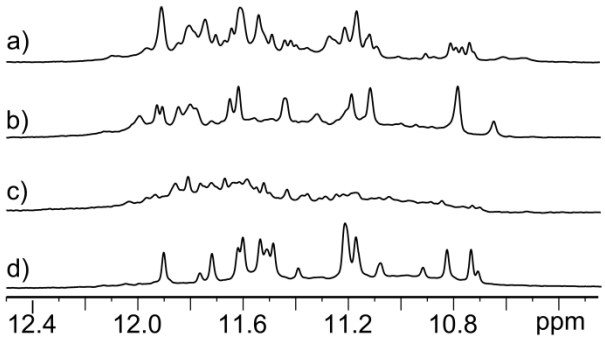

Fig. 2. Imino region of proton 1D spectra of oligonucleotide d[AGGGTTAGGGTTAGGGTTAGGG], a) wild type, b) with abasic site at position 7, c) with abasic site at position 13, and d) abasic site at position 19. From Babinský et al.(2014), Nucl. Acids Res. 42, 14031-14041.

Base pairing patterns of nucleic acids

Proton NMR spectrum is an excellent indicator of the base pairing pattern in nucleic acids. The imino hydrogens involved in Watson-Crick base pairs give rise to signals between 12 and 14 ppm (a). Signals between 10 and 12 ppm are characteristic of Hoogsteen base pairs (b). Absence of signals beyond 10 ppm suggests that no stable base pairs are formed.

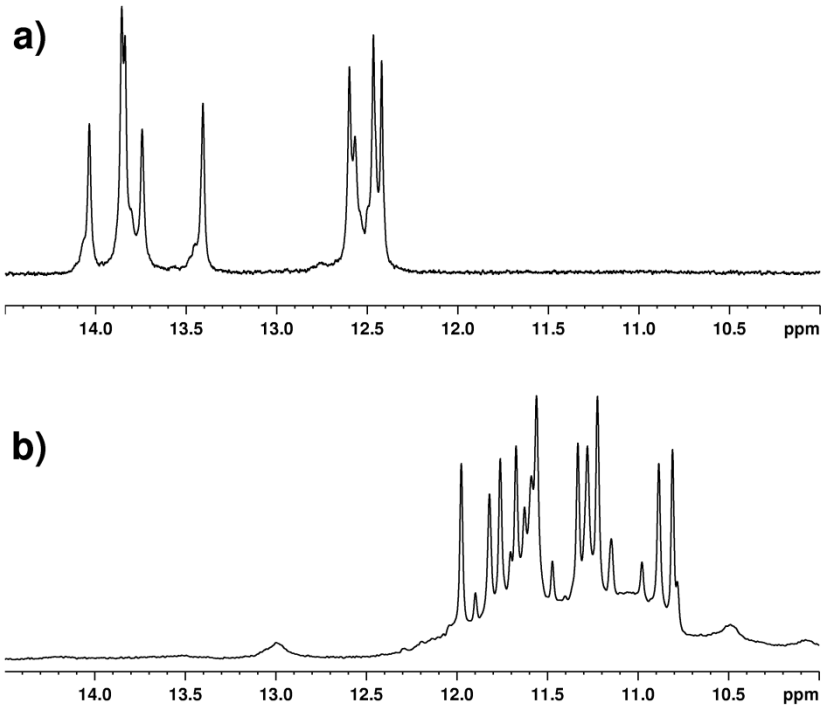

Fig. 3. Imino regions of proton NMR spectra of oligonucleotides d(TCTTGTGTTCT*AGAACACAAGA) forming duplex (a) and d[AGGGTTAGGGTTAGGGTTAGGG] (b) forming G-quadruplex. The signals between 12 and 13 ppm arise from GC base pairs, signal between 13 and 14 ppm from AT base pairs, and signal between 10 and 12 ppm from GG Hoogsteen base pairs.

Protein and nucleic acid interactions

NMR spectroscopy provides several methods to study interactions between molecules. Examples are given below

1. Chemical shift perturbation

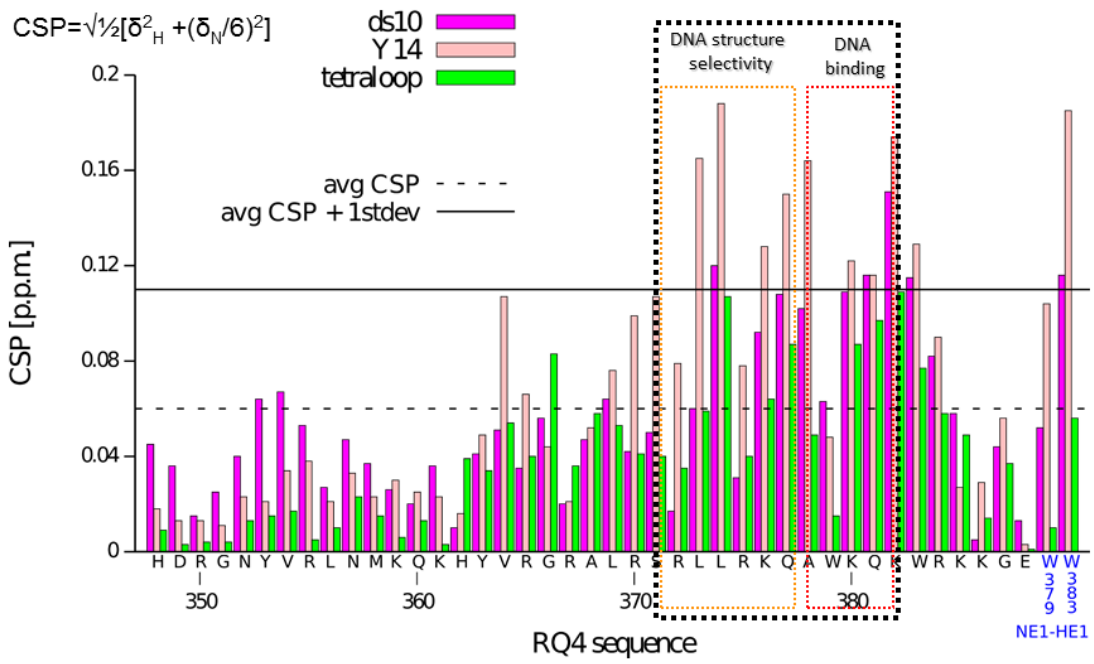

Fig. 4. Example of a chemical shift pertrubation study of binding 3 DNA constructs to RecQ4 DNA helicase protein. Picture by Anna Papageorgiou, CEITEC.

Example of using CSP to study binding of 3 DNA constructs to the protein called RecQ4 DNA helicase is given in Figure 4. Ligand binding changes the chemical shifts (y-axis) of the protein for the residues (X-axis) where it binds. The CSP thus reveals the binding site (residues 371 to 382, in dotted boxes) and ranks the binding. The construct Y14 shows the highest affinity (tallest bars, beige), followed by the double stranded decamer (ds10, purple), with tetraloop least tightly bound (green).

2. Saturation Transfer Difference

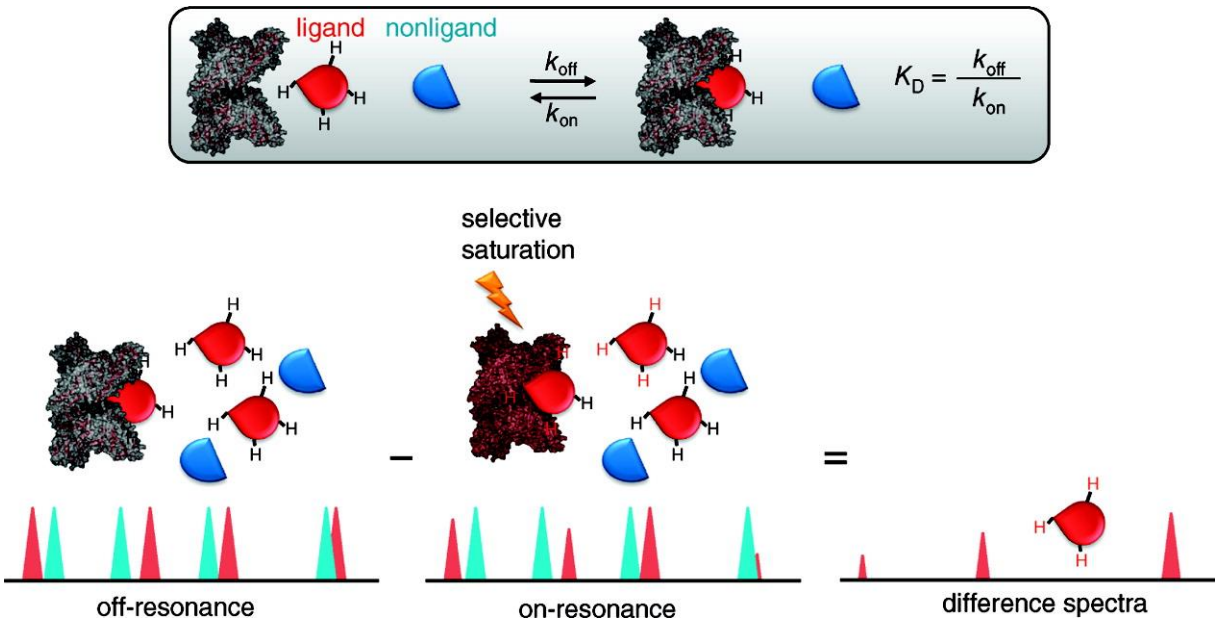

Fig. 5. Scheme of the STD-NMR experiment. The exchange between free and bound ligand allows intermolecular transfer of magnetization from the receptor to the bound small molecule. Illustration from Aldino Viegas; João Manso; Franklin L. Nobrega; Eurico J. Cabrita; J. Chem. Educ. 2011, 88, 990-994.

The principle of STD is illustrated in Figure 5. The saturation of the macromolecule is transferred to the bound molecule, while molecules that do not bind remain unchanged. The difference between spectra with- and without saturation reveals the bound molecule.

3. X-filtered NOESY

The NOESY spectrum filtered by heteronuclei (usually ^{13}C and ^{15}N) can selectively detect interactions between isotopically labeled protein and unlabeled ligand.

3D structures of proteins and nucleic acids

The solution of a complete 3D structure by NMR spectroscopy consists of several steps

1. Measuring a series of 2D and 3D spectra and assignment of the NMR signals
2. Getting structural restraints (interatomic distances, dihedral angles, bond vector directions) from NOESY, COSY and HSQC type spectra
3. Running molecular dynamics simulation using the restraints obtained in the previous step.

The core facility has the equipment and expertise to measure the spectra by all current state-of-the-art methods. However, the interpretation of spectra in terms of a 3D structure is laborious and requires moths of full time work by a qualified person. The CF does not have personnel to provide this kind of service. This can only be done in individual cases on a basis of collaboration with one of CEITEC's structural biology research groups.

If a detailed 3D structure in the form of a PDB file is not necessary, all three steps outlined above may not be necessary and the procedure is simpler. In the example below, the goal was to determine the topologies of abasic forms of oligonucleotide d[AGGGTTAGGGTTAGGGTTAGGG] with missing adenines at positions 7 (ap7) and 19 (ap19). From the NOE connectivities (Figure 6), it was possible to determine that the ap7 and ap19 oligonucleotides adopt the so called hybrid-1 and hybrid-2 forms of 3+1 quadruplex topologies, respectively.

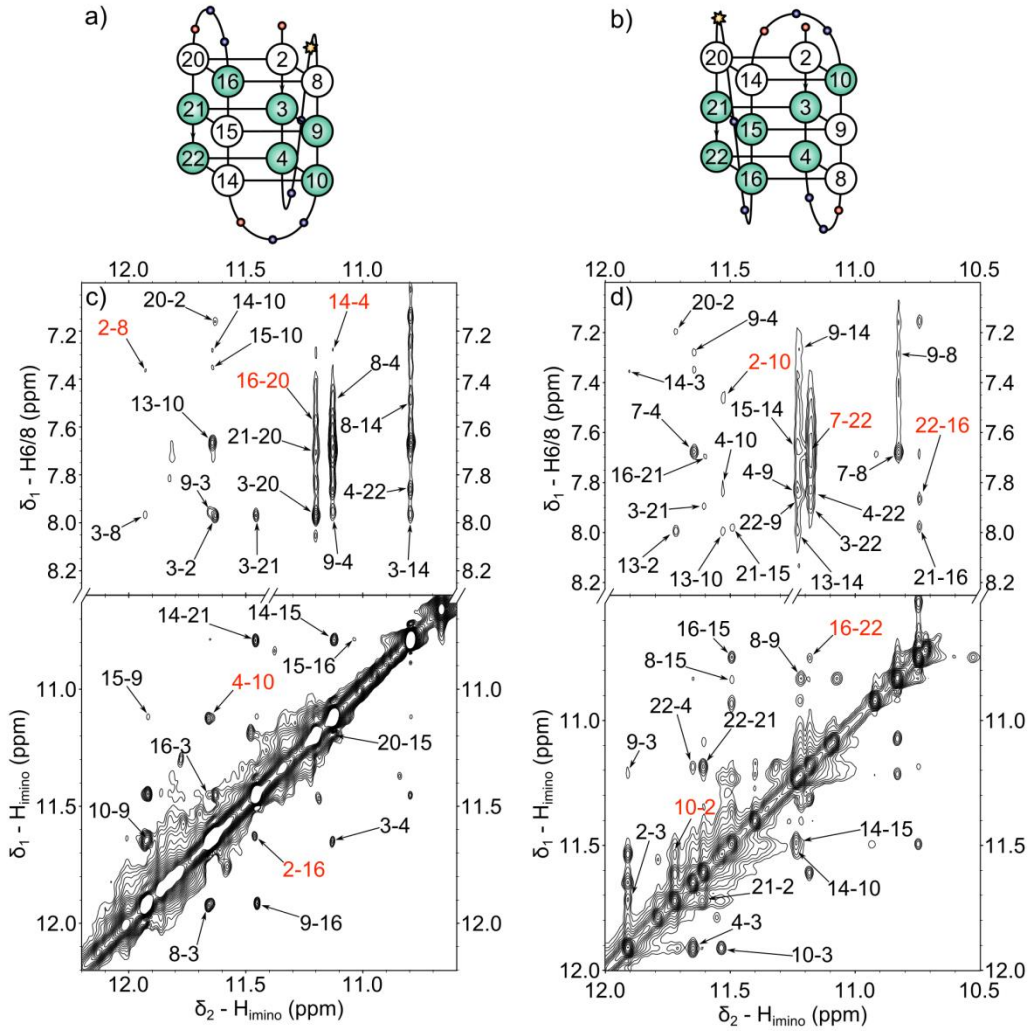

Fig. 6. Schematic representations of (a) ap7 (hybrid-1) and (b) ap19 (hybrid-2) quadruplex topologies. Guanines in syn- and anti-conformation are denoted as white and green rectangles, respectively. Adenines are shown as red and thymines in blue. Large yellow circles denote the positions of AP sites. Parts of 150 ms NOESY spectra of ap7 (c) and ap19 (d) showing Himino–Himino (lower part) and Himino–H8 (upper part) connectivities between bases. The labels denote the positions of the two interacting bases in the sequence. The crosspeaks specific to the particular form of (3 + 1) quadruplex are labeled in red. From Babinský et al.(2014), Nucl. Acids Res. 42, 14031-14041.

Biomolecular dynamics studied by NMR relaxation

NMR relaxation rates are influences by molecular motions, both overall and internal. Therefore, the dynamics of biomolecules can be studied based on measured relaxation rates. Most often, the protein backbone dynamics is characterized by relaxation properties of amide nitrogen nuclei. NMR provides two time windows to characterize either fast motions (time scale of nanoseconds and faster) or motions on intermediate time scale of milliseconds to microseconds. The information obtained is not detailed enough to characterize the motions completely. Rather, it provides information about the amplitudes and times of motions in the former case and the relative populations of two or more states and exchange rates between them in the latter case.

Examples of dynamics studies of a small 9.4 kDa protein (domain 1.1 of the σ^A factor from *Bacillus subtilis*) are shown below. The results were kindly provided by Lukáš Žídek's research group.

In the fast dynamics, the amplitudes of motions are characterized by the *generalized order parameter* S^2 whose values vary between 0 (for a perfectly flexible residue, all orientations of the interatomic vector N-H would be represented equally) and 1 (completely rigid residue, the interatomic vector N-H would be pointing always in the same direction). The times scales of the motion are characterized by correlation times for the internal motions τ_e. The contributions of all other motions are interpreted as chemical exchange expressed by the rate constant R_{ex}.

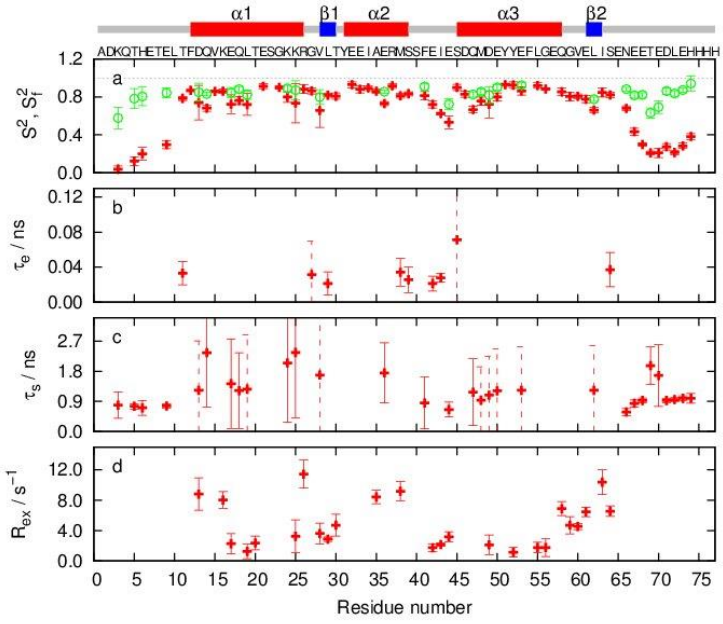

Fig. 7: Example of model-free analysis of a small 9.4 kDa protein. Results of model-free analysis using R_1,

R_2, and NOE relaxation rates measured at 600 MHz, 850 MHz, and 950MHz fields. Sequence and secondary structure elements are shown above the graph. (a) Generalised order parameter S^2 (indicated as crosses) and the faster motion model-free generalised order parameter S^{2f} (circles). (b) effective correlation time τ_e (crosses) and (c) slow timescale correlation time τ_s (circles). (d) Chemical exchange rates Rex.

Relaxation_dispersion

The protein dynamics on intermediate time scale of micro- to nanosecons is studied based on the relaxation rates measured under spin-lock conditions. The data are interpreted in the terms of Exchange rates, populations of the conformation states, and chemical shift differences between the states.

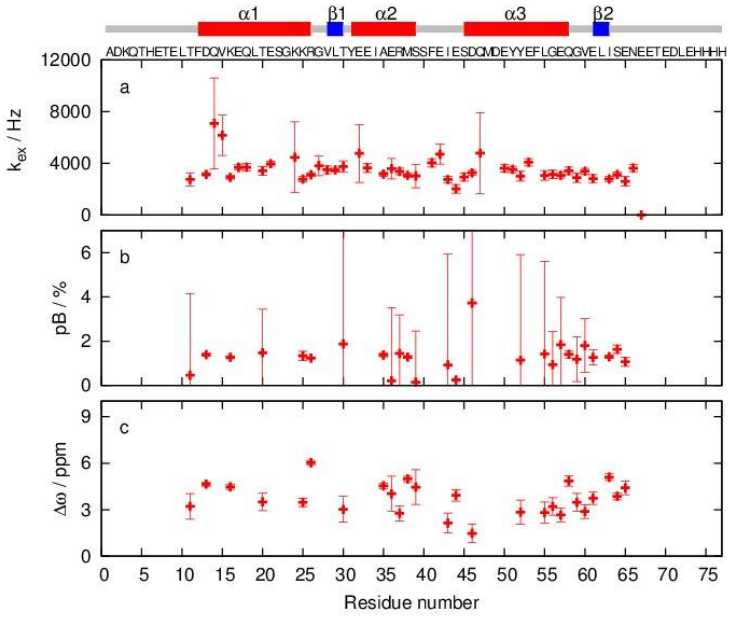

Fig. 8: Relaxation dispersion analysis of the intermediate time-scale motions of a small 9.4 kDa protein. CPMG relaxation dispersion experiments were measured at 600 MHz, 850 MHz, and 950MHz fields. Sequence and secondary structure elements are shown above the graph, (a) exchange rate for a two-state model, (b) population of a minor conformational state, (c) chemical shift difference between the two states.

Note.
NMR relaxation studies of protein dynamics require a sample isotopically enriched by ^{15}N and the protein structure (PDB file) must be available to interpret the data. The measurements require several weeks of experimental time and the data analysis is laborious. The core facility can obtain the data but has no capacity to interpret the results. The assistance in data analysis is available only in individual cases in collaboration with one of the structural biology research groups.