NMR for the non-spectroscopist

What is NMR?

- NMR spectra arise from transitions made by atomic nuclei between different energy states.
- The nuclei of many isotopes carry magnetic dipoles, that take up different orientations in a magnetic field → in the magnet of an NMR spectrometer.
- According to quantum mechanics, transitions between states with certain energies are permitted.
- When we apply pulses of electromagnetic radiation at frequencies that precisely match those energy gaps, we are able to observe transitions that give rise to NMR signals.
- Nuclei in different environments resonate at different frequencies.
- 1D NMR spectrum: plot pf intensity vs resonance frequency.
- Raw frequencies (MHz) scale with the size of the NMR magnet. To correct for this, we report resonance frequencies as "chemical shifts" (ppm).
- Excited states are relatively long-lived (nano-millisecond range) → we can have multiple transfer steps in a single experiment
- Since we can measure the frequencies of each of the nuclei through which excitation is passed, we can obtain signals that correlate the frequencies of many nuclei → multidimensional spectra: each transfer can be visualized a san independent nuclear frequency dimension (axis), and signals occurring at the intersections of two or more frequencies indicate a correlation between the corresponsing nuclei.

Types of NMR spectra:

- 1D H-NMR: shows signals for each H atom.
- 2D N-HSQC: a signal for each covalently bonded H-N group. Each signal has an intensity and two chemical shifts (one for the H and one for the N). In essence, one peak for each residue in the protein → excellent high-resolution fingerprint of the protein.
  - One peak for every backbone amide proton.
  - A pair of peaks for each Trp
  - A pair of peaks the side chain amide groups in Asn and Gln.
- C-HSQC: a signal for each covalently-bonded H-C pair. Not as well resolved as N-HSQC.

How much sample do I need?

- 50 ug are enough, and can be recovered and used for other purposes after the experiment.
- We can work with natural abundances of the isotopes, but normally we will have to work on recombinant proteins that have been overexpressed in minimal media containing nutrients enriched in those isotopes.

Information that can be deduced from an NMR experiment:

- Is the protein folded? Unfolded → all residues are exposed to solvent → poor signal dispersion.
- Is the protein agreggated? Agreggation broadens signals Conformational exchange only broadens the signals of the nuclei whose environment has changed.
- Is the protein dynamic? We can know if the protein undergoes us-ms conformational exchange.
- How stable is the protein? Degradation is visible.

Other important parameters:

- Strength of the magnetic field: higher strength → better signal
- Molecular weight: over 50kDa it becomes difficult.
- Temperature: higher T → better sigal (except for proteins that are not stable over some T).
- Buffer: No firm guidelines.

- Membrane proteins: most important → the choice of solubilizing detergents

Advantages of NMR:
- We measure much more thana static image of a protein
- We can study proteins in their native solution

Triple resonance experiments:
- First task: find the CS of every atom in the molecule → resonance assignment (H, N, C)
- We gather structural restraints. Usually: proton-proton distances, dihedral angles and hydrogen bonds.
  - Dipolar interactions between protons (measured as a H, H nuclear Overhauser effect (NOE) in 2D NOESY experiments → interproton distance restraints. NOEs are the most important source of structural restraints. Only ones that can be used on their own.
  - Scalar couplings (between pairs of nuclei that are cloe in the covalent structure of the molecule) → backbone phi and psi angles.
  - Hydrogen bonds can also be inferred as structural restraints.

NMR structure validation

Some interesting notes:

- The more specific data included, the less general modeling is required.
- The very high precision with which atoms are described in silico can be misleading, as the uncertainties of the process are not immediately apparent.
- Validation can be knowledge-based or model vs data.
- The set of structures obtained in the end do NOT correspond to the conformers. The set will depend on the experimental data used and also on the modeling process.

Blind testing of routine, fully automated determination of protein structures from NMR data

Is it possible to generate NMR structures in a fully automated manner?
Why atuomated? Because it would be faster, easier and more reproducible.

They took 10 experimental data sets with unassigned nuclear Overhauser effect spectroscopy (NOESY) peak lists for various proteins of unknown structure. Their structures were computed using different fully automatic programs. The resulting structures were compared among them and also against manually solved reference structures (that were not available during the blind test).

Typical NMR protocol:
1) Resonance assignment: the chemical shifts (CS) observed in multidimensional spectra are assigned to their corresponding protein atoms.
2) NOESY assignment: thousands of dipolar-coupling effects are identified in multidimensional NOE spectra and assigned and converted into interatomic distance restraints.
3) Additional conformational restraints from RDCs, scalar couplings, CS data.
4) Structure generation: software programs are used generate a set of protein conformations (called a bundle of conformers) that should satisfy these restraints.
5) Structure refinement: the bundle of conformers is energetically refined through restrained MD simulations.

2-4 are performed iteratively in an integrated manner to maximize the number of conformational restraints obtained while guaranteeing self-consistency. We want them to converge to a self-consistent set of conformational restraints from which the final bundle of low pseudoenergy conformers is calculated.

Data: assigned CS lists and unassigned NOESY peak lists.

Three groups of automated methods:

1) Only NOESY data to obtain distance restraints for structure calculations: CYANA, UNIO, ASDP, ARIA.

2) CS data augmented by NOESY data:
- NOESY information is used to re-rank CS-based results: CS-DP-Rosetta
- Uses CS-generated structures to perform NOESY assignments and extract distance restraints: Chesire-YAPP.

3) Only CS data as experimental information: Chesire and CS-Rosetta.

They compared two of the structures with X-ray crystallography structures of close homologs.

How do we quantify convergence? Rmsd among the automatically generates structures.

How do we measure accuracy? Rmsd agains the reference structure. Also, Global Distance Test total score (GDT_TS) (thee two meaures are anticorrelated).

Results:
- Three of the NOESY-based programs automatically and consistently generated acceptable structures.
- CS-based methos augmented with NOESY data.
- Pure CS-methods had the poorest performance. Also, they are computationally more demanding.
- Geometric and stereochemical quality of the structures is not a good indicator of their accuracy, because it is mainly determined by the force field used in the refinement step. Even structures with a wrong fold can give good stereochemical results.
- DP score: quantifies the agreement between the structures and unassigned NOESY data.