NMR structure determination

of a peptide using the ARIA webportal

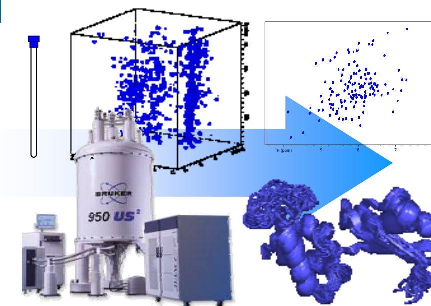
Henry Jonker

Center for Biomolecular Magnetic Resonance (BMRZ)
Institute for Organic Chemistry and Chemical Biology
Goethe University - Frankfurt am Main (Germany)

Structure calculation with ARIA (NMR Intensiv kurs 2016)

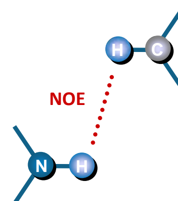
NMR observables that contain structural information:

- Atom distances
  - NOE (Nuclear Overhauser Effect)
- Secondary structure, interaction
  - Chemical Shift & Hydrogen Bond
- Torsion angles
  - ^{3}J Scalar Coupling
  - CCR (Cross Correlated Relaxation)
- Orientation, shape & long distance
  - RDC (Residual Dipolar Coupling)
  - PCS (Pseudo Contact Shift)
  - PRE (Paramagnetic Relaxation Enhancement)
  - Relaxation data (diffusion anisotropy)



Distances from NOEs

- NOEs are due to through-space dipole-dipole interactions
  - relate to atom-atom distance (up to ~5 Ångstrom)
  - commonly derived from 2D homonuclear ^{1}H^{1}H NOESY spectra and 2D, 3D or 4D heteronuclear (^{15}N and/or ^{13}C) filtered NOESY spectra
- $V_{NOE} = C / d^6$
  - V_{NOE} : NOE Volume (or intensity)
  - d : distance, C : calibration constant

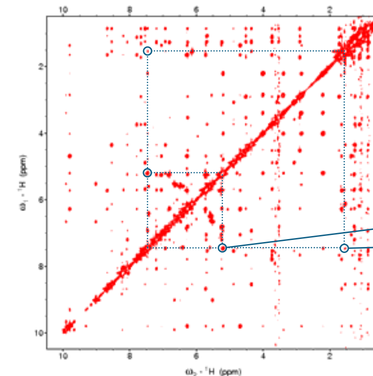


Calibration methods

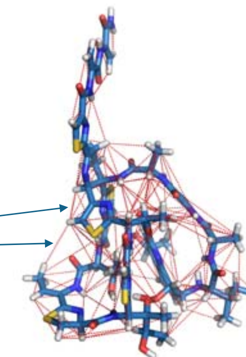
- divide into classes (old style: strong, medium or weak crosspeak)
- from known distances (intraresidual or secondary structures)
- from preliminary structures (if available)
- median cross peak V_{NOE} corresponds to a given distance

From NOE to structure

- NOE crosspeak » proton-proton distance



2D ^{1}H^{1}H-NOESY spectrum



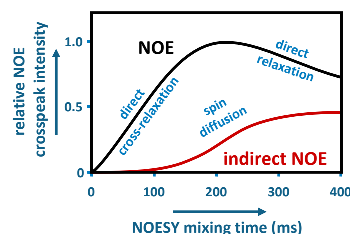
Structure with NOEs

❖ NOE Buildup curves

- Measure multiple NOESY spectra with different mixing times to correct for **spin-diffusion** and identify indirect NOEs at long mixing times
- A single NOESY spectrum with a sufficient **short mixing time** can be considered linear in good approximation
- Automatic callibration of NOESY for **longer mixing times** with relaxation matrix using spin-diffusion & exchange correction

- NOE buildup rate depends on the **mobility** and can thus differ per molecule, region or domain (fast for rigid parts with long τ_c and slow for flexible loops and termini)

- In complex, **transferred NOEs** can be observed for small (μM binder) ligands with fast off-rates (detect conformational information of the bound form on the easily observed signals of the free ligand)

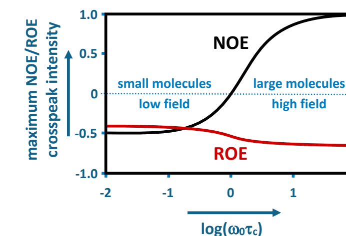


❖ Rotating-frame Overhauser Effect spectroscopY (ROESY)

- Measurement of NOEs using a spin-lock in the rotating-frame to overcome some undesirable features of NOE:
  - NOE is around zero for medium sized molecules (1-3 kDa)
  - Chemical exchange contribution can not be determined for large molecules since it has the same sign as the NOE
- ROE has the same distance corrrelation (d^6) as the NOE

- The **Exchange peaks** have opposite sign and can be identified

- The **COSY and TOCSY artifacts** have opposite sign as well (avoid by reducing the spin-lock power)
- Needs **more scans** per increment for same S/N ratio as NOESY

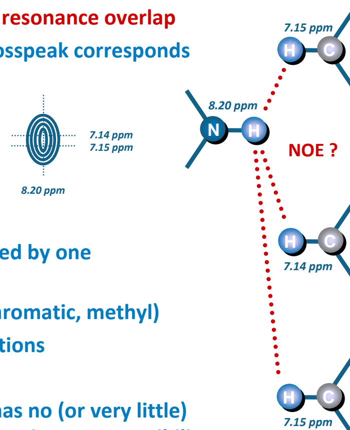


❖ Ambiguous Assignment of NOEs

- Frequently multiple assignments possible due to **resonance overlap**
- Volume and distance of an ambiguous NOESY crosspeak corresponds to the **sum of individual contributions**:

$$V_{NOE} = \Sigma_a V_{NOE,a}$$

$$d = (\Sigma_a d_a^{-6})^{-1/6}$$



- Ambiguous distance restraint (**ADR**) can be fullfiled by one or a combination of assignment possibilities
- ADR principle also used for **equivalent protons** (aromatic, methyl)
- Reduced **ambiguity cut-off** over the various iterations

- The presence of a wrong assignment possibility has no (or very little) influence on the structure, as long as the correct assignment possibility is present.

Nilges et al. JMB (1997)

❖ NOE distance restraints (example: SPARKY & CNS/ARIA)

- Peak and resonance list from SPARKY

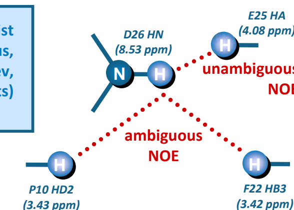
```
E25HA-D26HN      4.078      8.525      12101860
   ?-D26HN       3.429      8.528       3032037
```

peak list (ppm values & volume/intensity)

```
P10  HD2   1H     3,433  0.003   3
F22  HB3   1H     3.423  0.007   3
E25   HA   1H     4.078  0.004   3
D26   HN   1H     8.526  0.002   9
```

resonance list (group, atom, nucleus, chemical shift, st.dev, nr. of assignments)

Typical tolerance: 0.02-0.04 ppm (1H)



- ARIA final distance restraints: (un)ambig.tbl

```
ASSI ( resid 26 and name HN ) ( resid 25 and name HA )
2.250  0.630  0.630  peak 23 spectrum 1 weight 1.1 volume ...
```

unambiguous restraint (distance, -, +, peak info)

```
ASSI ( resid 26 and name HN ) ( resid 22 and name HB3 )
3.180  1.270  1.270  peak 24 spectrum 1 weight 1.0 volume ...
OR   ( resid 26 and name HN ) ( resid 10 and name HD2 )
```

ambiguous restraint
the associated .list file shows 58% F22HB3, 42% P38HD3

❖ Empirical Forcefield

- General chemical knowledge about the composition of the macromolecule
  - residue definitions (atom composition connectivity, chirality and planarity)
  - linkage of residues in order to generate the sequence
  - atoms types (masses, charges)
  - bond lengths and angles
  - van der Waals and electrostatic interactions
- There are several forcefield implementations available for **proteins**, **RNA** and **DNA**
  - gromos, gromacs, allhdg, amber, charmm, opls, prolsq
- Parametrisation for **other molecules** (such as modified residues and ligands)
  - have to be generated by hand or can be obtained via PRODRG and/or HIC-UP

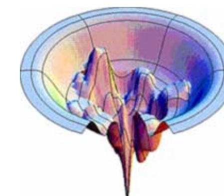
GOETHE UNIVERSITÄT FRANKFURT AM MAIN — Center for Biomolecular Magnetic Resonance — Structure calculation with ARIA — Henry Jonker Biomolecular NMR

❖ NMR structure calculations

- Energy potentials (for the force field parameters and NMR data)
- Structure calculation programs try to "fold" a molecule into a 3D structure that agrees with the experimental **NMR restraints** and the **forcefield constraints**
- Minimization of the differences, manifested as **energy violations**, drive the molecule towards its conformation
- The target **energy landscape** is the sum of squares (or similar) of the violations and has many **local minima**

❖ Target energy function

$E_{total} = \Sigma E_{bonds} + \Sigma E_{angles} + \Sigma E_{distances} + \Sigma E_{rdcs} + \Sigma E_{...}$

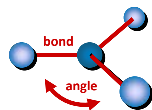


- E : energy from forcefield / restraint violation

GOETHE UNIVERSITÄT FRANKFURT AM MAIN — Center for Biomolecular Magnetic Resonance — Structure calculation with ARIA — Henry Jonker Biomolecular NMR

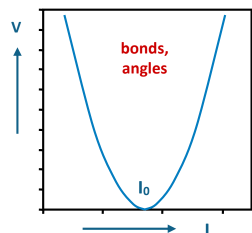
❖ Harmonic Potential

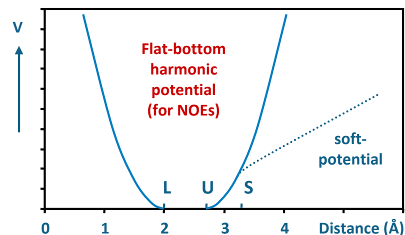
- For forcefield: eg. bonds, angles
- $V = c \times (l - l_0)^2$



❖ Flat-bottom Harmonic Potential

- For distance restraints: eg. NOEs, Hydrogen bonds

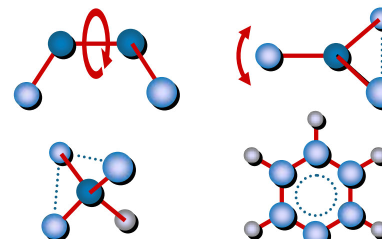


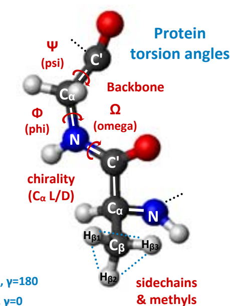


GOETHE UNIVERSITÄT FRANKFURT AM MAIN — Center for Biomolecular Magnetic Resonance — Structure calculation with ARIA — Henry Jonker Biomolecular NMR

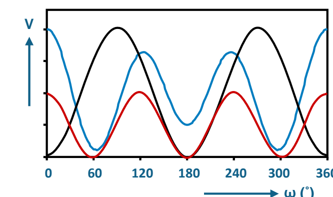
❖ Torsion angles

- Dihedral / Improper Angle:
  - contraint in forcefield and restraint by NMR data
  - planarity (peptide bond, ring, base, ...)
  - chirality, stereochemistry, methyl stagger





- Usually expressed as cosine series expansions
  - Can have multiple minima which do not have to be equal when using multiple terms
  - $V(\omega) = \Sigma_n \frac{1}{2} V_n \times [1 + \cos(n\omega - \gamma)]$

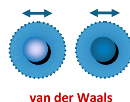


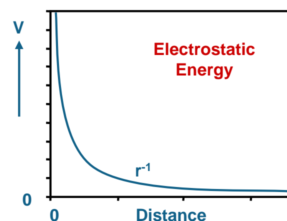
Electrostatic interactions

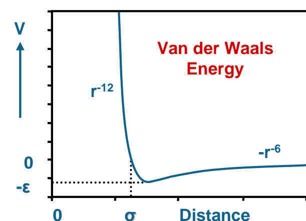
- Coulombs Law: $V_{elec} = \Sigma_i \Sigma_j\, q_i q_j / 4\pi\varepsilon_0 r_{ij}$
  ε = dielectric constant, q = partial point charge

Van der Waals interactions

- Attractive long-range and repulsive short-range forces
- Lennard-Jones: $V_{L\text{-}J} = 4\varepsilon\, [\, (\sigma/r)^{12} - (\sigma/r)^{6}\,]$
  ε = well depth, σ = collision diameter



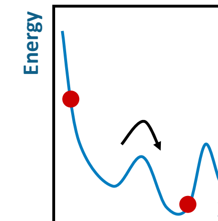


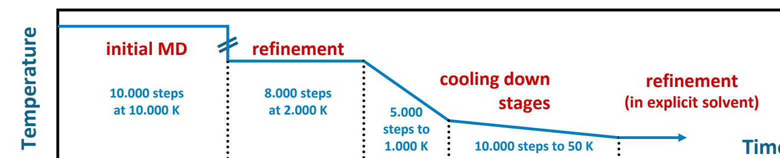


GOETHE UNIVERSITÄT FRANKFURT AM MAIN · Center for Biomolecular Magnetic Resonance · Structure calculation with ARIA · Henry Jonker Biomolecular NMR

Molecular Dynamics (MD)

- Direction of motion depends on forces (restraints and forcefield) and momentum
  - Cartesian Angle Dynamics (CAD, coordinate space)
    (timestep for flexible bonds and vibrations: 2-5 fs)
  - Torsion Angle Dynamics (TAD, angular space)
    (about 10 times faster - less degree of freedom)
- MD can overcome local energy barriers
  - temperature relates to kinetic energy and thus velocity
  - scaling and use of different energy terms
    (increase of restraint energy terms during cool-down)
- Use Simulated Annealing (SA) protocol

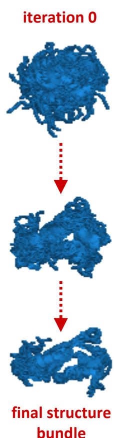




GOETHE UNIVERSITÄT FRANKFURT AM MAIN · Center for Biomolecular Magnetic Resonance · Structure calculation with ARIA · Henry Jonker Biomolecular NMR

Typical NMR structure calculation

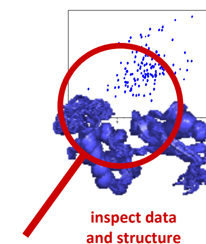
- NMR data:
  - (un)assigned NOESY crosspeaks, resonance assignments, RDCs, dihedral angles, J-couplings, H-bonds, planarity, PRE distances and relaxation data
- Initial structure
  - generate the molecular topology and an extended structure, model or random bundle using the sequence and forcefield
- Iterative structure calculation
  - assign, analyse and callibrate NOEs based on the energetically best structures of the previous iteration
- Energy minimization or the final structures
  - in ARIA, refinement in explicit solvent (water)
- Analyse and validate the final structures
  - repeat the calculation with corrected dataset



GOETHE UNIVERSITÄT FRANKFURT AM MAIN · Center for Biomolecular Magnetic Resonance · Structure calculation with ARIA · Henry Jonker Biomolecular NMR

Structure Validation

- Does the 3D structure agree well with all the experimental data
  - check violations and new assignments, backcalculate RDCs, ...
- Cross-validate the structure
  - does unused data (eg. some % NOEs, RDCs, ...) fit well?
- Is the structure physically correct?
  - geometry, bonds, angles, planarity, ... (check energies)
- Validation parameters
  - RMSD with reference structure (when available...)
  - Local-Global alignment method, Global Distance Test Score
  - TM-score, Z-score, Ramachandran plot quality, G-factor
- Validation software
  - Cing (Common Interface for NMR structure Generation)
  - PSVS (Protein Structure Software suite)
  - Procheck, Whatcheck, Whatif, Aqua, Prosa, Molprobity, ...
- Use another method (Aria, Cyana, Xplor, Unio, ...) to confirm



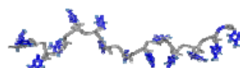
inspect data and structure

❖ Peptide structure calculation using the ARIA webportal

- The webportal uses an adapted version of ARIA (based on 1.2) and automates various steps for the setup and the structure calculation
  - the webportal creates the full setup and submits the structure calculation to one of the computational nodes. The completed full run can be downloaded as a gzipped tar file (.tgz)

❖ Data needed

- Protein Sequence file
  - protein/peptide sequence (eg. "xxx.seq" file)
- Chemical Shift Assignment
  - resonance list (eg. "xxx.shifts" file)
- NOESY or ROESY peak list
  - assigned/unassigned peak list (eg. "xxx.list" file)



- Sequence file made by yourself, resonance and peak files exported by Sparky
- Carefully check and understand the input files before submitting!

❖ Sequence file

- manually create a "xxx.seq" file

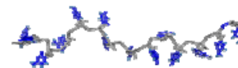
```
MET HIS ASN THR TRP GLY VAL CYS SER PRO LEU

```

at the end: enter / blanc line (counts for all input files)

- create a simple text file (not a word document) with the sequence of your peptide
- use 3-letter amino acid codes separated by a space (do not use tabs)
- first residue is residue number 1

- used with the selected forcefield to generate the initial extended template PDB file



❖ Export the resonance list from Sparky

- open your project
- delete unused resonances (dr, Peak - Resonances - Delete unused resonances)
- show the resonance list (rl, Peak - Resonances - Resonance list)
- press update after changes

- check the Group and Atom names
  - right nomenclature ?!
  - wrong and unconventional names cause errors or will not be used

- check the error of the resonances (SDev)
  - large error indicates that something has been misassigned
  - selecting a line will show the involved peaks
  - correct the wrong assignments ....

- save the resonance list (xxx.shifts)

Resonance List

Molecule / Condition:

Group	Atom	Nuc	Shift	SDev	Assignments
L28	HD1#	1H	0.060	0.003	2
L28	HD2#	1H	0.122	0.001	3
L28	HG	1H	1.168	0.000	2
L28	HN	1H	7.881	0.002	9
L28	N	15N	114.187	0.040	9
G29	CA	13C	47.154	0.053	5
G29	HA1	1H	3.907	0.002	3
G29	HA2	1H	3.995	0.003	3
G29	HN	1H	8.693	0.005	18
G29	N	15N	109.915	0.012	15
Q30	CA	13C	57.483	0.044	2
Q30	CB	13C	28.244	0.041	5
Q30	CG	13C	32.975	0.039	5
Q30	HA	1H	4.030	0.001	4
Q30	HB1	1H	1.672	0.002	3
Q30	HB2	1H	1.907	0.002	3
Q30	HE21	1H	6.724	0.002	4
Q30	HE22	1H	7.310	0.002	4

Update | Save... | Load... | Close | Help

❖ Export the NOESY or ROESY peak list from Sparky

- open your project
- select your NOESY or ROESY spectrum and show the peak list (lt, Peak - Peak list)

- under options select to show " Data height "
  - remove noise / negative peaks ... ?
- set assignment format to " %A1-%A2 "
- save the peak list (xxx.list)

Spectrum Peaks NOESY

Assignment	w1	w2	Data Height
?-?	2.501	7.944	22160696
?-?	3.818	8.063	2169447
A2HA-K3HN	3.985	8.342	3984323
K3HA-K3HN	4.257	8.346	1603697
K4HN-K3HN	8.809	8.345	605837
V5HA-V5HN	3.543	8.658	4741525
A6HN-A7HN	9.564	7.798	21789304
R41HN-F42HN	8.032	8.761	11556816
D63HN-E62HN	7.750	8.781	7462892
P73HD1-T72HN	3.458	8.091	931151
P73HD2-T72HN	3.844	8.096	1223515

Search for:

Update | Save... | Options... | Close | Help

Options Spectrum Peaks NOE..

☑ Assignment	☑ Data height
☐ User label	☐ Signal / Noise
☑ Frequency (ppm)	☐ Linewidth
☐ Frequency (Hz)	☐ Freq of resonance axis
☐ Resonance deviation	☐ Dev of resonance axis
☐ Volume	☐ Spectrum Name
☐ Volume error	☐ Assignment distance
☐ Transpose volume	☐ Mardigras format
☐ Fit residual	☐ Dyana format
☐ Fit height	☐ Note

Sort by: Resonance name

Sort axis: 1

☐ Pair crossdiagonal peaks?

Assignment Format: %A1-%A2

Ok | Apply | Close | Help

WeNMR ARIA Webportal

- from the WeNMR website (www.wenmr.eu) go to: NMR - Structure calculation - ARIA
- or directly visit: http://enmr.chemie.uni-frankfurt.de/portal/aria.html

Running ARIA

- Settings
  - protocol for protein
  - 50 structures per iteration
  - 100 structures in last cycle
  - 10 final structures
  - enter your 'lucky' random number
- Input files
  - sequence: xxx.seq
  - no distance restraints, no hydrogen bonds, no dihedral angles, no planarity restraints

« select protocol for protein

« sequence file

Running ARIA

- NOESY or ROESY peak list
  - sparky format
  - enable swapping (option selected)
  - peak list 1: xxx.list
  - resonances: xxx.shifts
  - proton 1: column 1
  - tolerance 0.02 ppm
  - hetero 1: absent
  - proton 2: column 2
  - tolerance 0.02 ppm
  - hetero 2: absent

« sparky format

« NOESY or ROESY

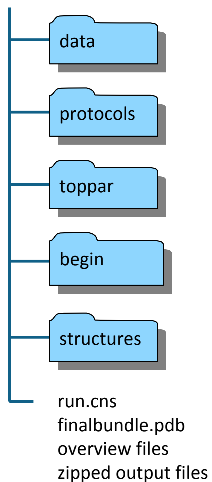
- use all peaks and keep assigned peaks
- do not reset bound for systematic violations (option not selected)
- do not use spin diffusion and exchange correction (options not selected)

Running ARIA

- Enter user name (testffm) and password (ffmtest) and press submit
- Depending on the complexity and computer usage, it wil take about an hour
- Note down the job ID code !!

❖ ARIA directory setup (as retrieved from the webportal)

- data
  - sequence, h-bonds, planarity, distances, dihedrals, diffusion, j-couplings, RDCs, s-bonds and converted NOESY data
- protocols and toppar
  - all ARIA protocols and forcefield files
- begin
  - molecular topology and a template extended PDB file which is used as a initial start for the structure calculation
- structures
  - iterations for the structure calculation
  - last iteration inlcudes a water refinement and (violation) analysis
- run.cns
  - settings file for the ARIA run
- finalbundle.pdb
  - final bundle of structures (from structures/it8/water/analysis)
- overview files
  - various info and new restraints from NOESYs



run.cns
finalbundle.pdb
overview files
zipped output files

❖ Retrieve and analyse the results

- Check the std.err and std.out file for errors
- Retrieve the results and unpack the aria.tgz file

- Check the template extended PDB file (molmol, pymol) and the input data
  - /begin/aria_template.pdb and /data (eg. sequence and NOESY or ROESY)
- Check for NOE violations
  - /structures/it8/water/analysis/ana_noe_viol_unambig.lis
  - /structures/it8/water/analysis/ana_noe_viol_ambig.lis
- Check for new assignments
  - overview.newunambig (and overview.newambig)
- Check the ramachandran score
  - /structures/it8/water/analysis/procheck_comp/ps-files.zip - aria_run1_01.ps
- Inspect the final bundle of structures (using molmol/pymol)
  - finalbundle.pdb, converged? backbone RMSD?
- Try to make a judgement of the succes of the run