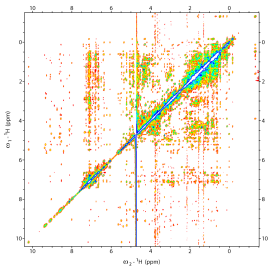
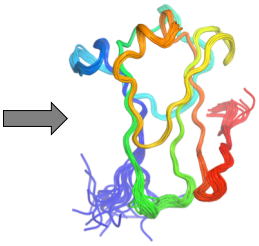
NOESY spectrum



structure



NOEs and ambiguity

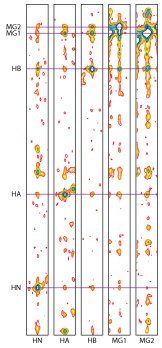
- ^{1}H-^{1}H NOESY 2D $^{1}H \leftrightarrow {}^{1}H$
- ^{15}N- or ^{13}C edited ^{1}H-^{1}H NOESY

3D $^{1}H \leftrightarrow {}^{1}H$ (^{15}N) $^{1}H \leftrightarrow {}^{1}H$ (^{13}C)

- ^{15}N- or ^{13}C edited ^{1}H-^{1}H NOESY

4D $^{1}H \leftrightarrow {}^{1}H$ (^{15}N / ^{13}C) $^{1}H \leftrightarrow {}^{1}H$ (^{13}C / ^{13}C) $^{1}H \leftrightarrow {}^{1}H$ (^{15}N / ^{15}N)

Working with 3D-NOESY



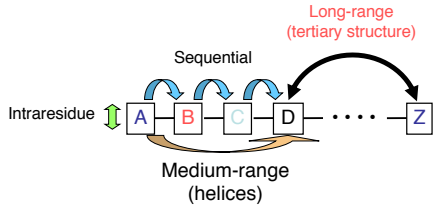
assign only intra-residue cross-peaks to generate accurate chemical shift lists

NB: do not assign long-range NOEs !!!
Let CYANA do the job

User considerations

- Completeness of chemical shift assignments should be higher than 90%
- Lack of aromatic chemical shifts is harmful to the outcome of a structure calculation because they give rise to a higher-than-average number of NOEs

4. ^{1}H-^{1}H Distances from NOEs



Challenge is to assign all peaks in NOESY spectra - semi-automated processes for NOE assignment using NOESY data and table of chemical shifts yet still significant amount of human analysis

Traditionally NOE Assignment is done manually

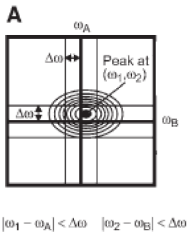
- User is biased against the data (erroneous assignments - rejected peaks)
- Time consuming (several months)

Automated NOE Assignment and de novo Structure Calculation

Distance restraints from not uniquely assigned NOEs:
→Ambiguous distance restraints
Robustness against erroneous assignments:
→Constraint combination / violation confinement
Reduction of assignment ambiguity prior to the structure calculation:
→Probabilistic network-anchored assignment

CANDID/CYANA

NOE assignment and ambiguous distance restraints



$|\omega_1 - \omega_A| < \Delta\omega \quad |\omega_2 - \omega_B| < \Delta\omega$

In general, several different ^{1}H chemical shifts ω_A, ω_B match the position of a NOESY peak within the experimental uncertainty $\Delta\omega$.

→Assignment ambiguity

Manual assignment is very cumbersome!

For peak lists obtained from 13C- or 15N-resolved 3D NOESY spectra, the ambiguity in one of the proton dimensions can usually be resolved by reference to the heteroatom

Ambiguous distance constrains

$$d_{\text{eff}} = \left(\sum_k d_k^{-6} \right)^{-1/6} \leq b$$

upper distance bound; distance for assignment possibility k; sum over all assignment possibilities

Constraint with multiple assignments

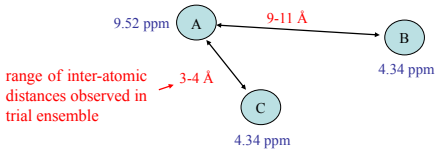
Allows delay of assignment choice until structures are better defined

If one assignment possibility leads to a sufficiently short distance, then the ambiguous distance restraint will be fulfilled.

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

Nilges et al., *J. Mol. Biol.* 269, 408–422 (1997)

Resolving ambiguity during structure calculation



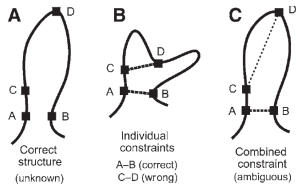
Due to resonance overlap between atoms B and C, an NOE crosspeak between 9.52 ppm and 4.34 ppm could be A to C or A to B - this restraint is ambiguous.

But if an ensemble generated with this ambiguous restraint shows that A is never close to B, then the restraint must be A to C.

80

Constrain combination

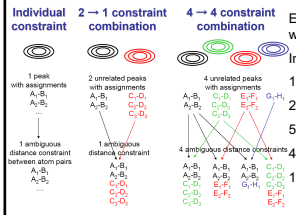
Problem: Peaks with wrong (long-range) assignments may severely distort the structure, especially in the first cycles, and may lead to convergence to a wrong structure.



Idea: From two long-range peaks each, combine the assignments into a single distance constraint for the first two cycles.

Result: occurrence of erroneous assignments is reduced at the expense of temporary loss of information

Effect of constrain combination



Example: 1000 long-range peaks, 10% of which would lead to erroneous constraints

Individual constraints:

1000 constraints, ≈1000 x 0.1 = 100 wrong (10%)

2->1 constraint combination:

500 constraints, ≈500 x 0.1^2 = 5 wrong (1%)

4->4 constraints combination:

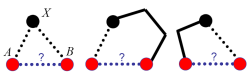
1000 constraints, ≈1000 x 0.1^2 = 10 wrong (1%)

The number of long-range constraints is halved by the 2->1 combination but stays constant on 4->4 pair-wise combination!!!

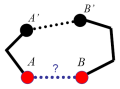
Network-anchoring

Distance d_{AB} restricted by covalent structure

Another NOE (e.g. a transposed peak) with the same assignment exists



A pair of NOEs (or an NOE and a covalently restricted distance) connecting atoms A and B through a third atom X exist

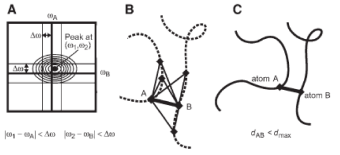


Atoms A and B are covalently close to two atoms A' and B' that are connected by a NOE

The generalized relative contribution is determined from chemical shift tolerance, cross-peak symmetry, covalent structure compatibility, and the convergence of network anchoring and three-dimensional structure compatibility of multidimensional experiments

Herrmann et al., *J. Mol. Biol.* 319, 209-227 (2002)

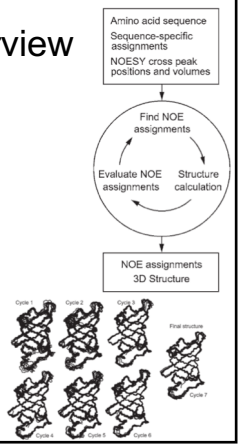
Conditions for valid assignment of a NOESY cross-peak



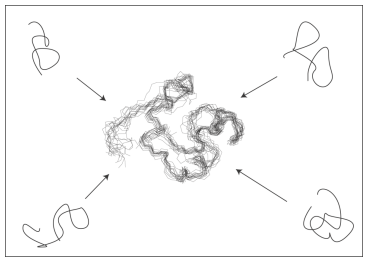
chemical shift agreement | network anchoring | spatial proximity in the structure

CYANA overview

- input data
  protein sequence, chemical shift lists, NOESY peaks, other constarints (RDC's, angles, hydrogen or disulphide bonds)
- initial assignments
  one or several assignments are defined based on chemical shift lists
- rank of initial assignments
  filtering criteria include presence of symmetry-related cross-peaks, agreement of chemical shifts and peak position, self consistency with the entire NOE network
- calibrate distance constraints
  from the NOESY peak volumes or intensities upper distance bounds are derived for the corresponding, ambiguous or unambiguous distance restraints
- eliminate spurious NOESY cross-peaks
- constraints combination
  unrelated long-range distance constraints are combined into new virtual distance constraints
- structure calculation



NMR structure calculation



molecular dynamics

We use a force field, or equations that describe the energy of the system as a function of **<xyz>** coordinates.

In general, it is a sum of different energy ***terms***:

$$E_{total} = E_{vdW} + E_{bs} + E_{ab} + E_{torsion} + E_{elctrostatics} + \ldots$$

NOE data

As long NOEs relate our experimental data with the **<xyz>** coordinates, we include them at the end of the energy function.

Strong NOE	**1.8 - 2.7 Å**
Medium NOE	**1.8 - 3.3 Å**
Weak NOE	**1.8 - 5.0 Å**

The potential energy function related to these ranges looks like this:

$$E_{NOE} = K_{NOE} * (r_{calc} - r_{max})^2 \quad \text{if } r_{calc} > r_{max}$$
$$E_{NOE} = 0 \quad \text{if } r_{max} > r_{calc} > r_{min}$$
$$E_{NOE} = K_{NOE} * (r_{min} - r_{calc})^2 \quad \text{if } r_{calc} < r_{min}$$

It is a flat-bottomed quadratic function. The further away the distance calculated by the computer (r_{calc}) is from the range, the higher the penalty.

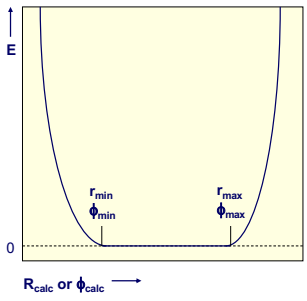
Torsion angles

Similarly, we include torsions as a range constraint:

$$E_J = K_J * (\phi_{calc} - \phi_{max})^2 \quad \text{if } \phi_{calc} > \phi_{max}$$
$$E_J = 0 \quad \text{if } \phi_{max} > \phi_{calc} > \phi_{min}$$
$$E_J = K_J * (\phi_{min} - \phi_{calc})^2 \quad \text{if } \phi_{calc} < \phi_{min}$$

Or any other type of contraints (RDC, PRE, PCS, chemical shifts, etc)

Penalty function



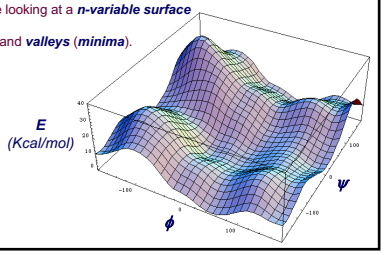
Energy minimization

functions in the potential energy expression for the molecule, represent bonded interactions (bonds, angles, and torsions), and non-bonded interactions (vdW, electrostatic, NMR constraints).

to get the structural model we must be able to ***minimize*** the energy of the system, which means to find a low energy (or the lowest energy) conformer or group of conformers.

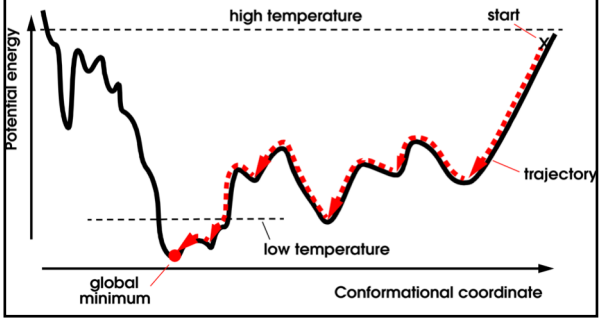
Nearly impossible, because we are looking at a ***n-variable surface***

We have energy ***peaks*** (***maxima***) and ***valleys*** (***minima***).

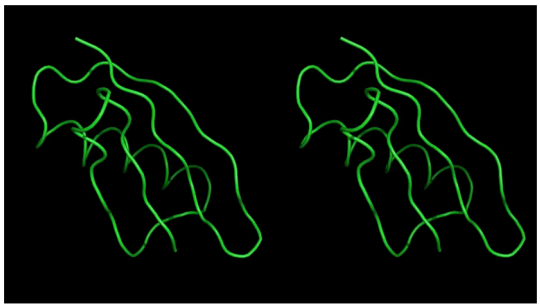


Simulated annealing

Provide energy to the system (rise the 'temperature') and see how it evolves with time. Temperature usually translates into kinetic energy, which allows the peptide to surmount energy barriers.



Restrained Molecular Dynamics structure determination of protein 1GB1 from NMR



Judge your structure CANDID criteria

- Average CYANA target function value of cycle 1 below 250 Å^2
- Average final CYANA target function value below 10 Å^2
- Less than 20% unassigned NOEs

  good data sets can reach 95% of input peaks assigned

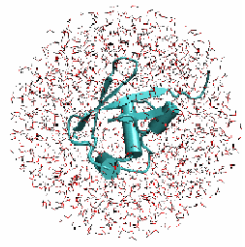
  always check the unassigned peaks !!!
- Less than 20% discarded long-range NOEs

  not straightforward to assess due to chemical shift ambiguity
- RMSD value in cycle 1 below 3 Å
- RMSD between the mean structures of the first and last cycle below 3 Å

Water refinement

Improving the Quality of NMR Structures

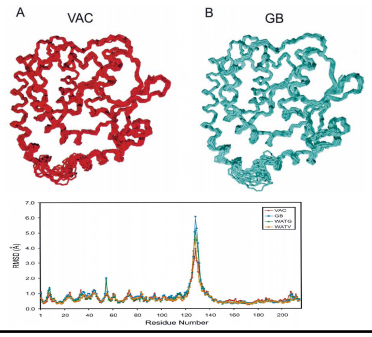
- Water Refinement
  - protein structures generally calculated in vacuum.
  - water has a significant effect on protein structures
    - explicit solvent model
      - MD simulation in box of water
      - box > 10 Å, keep solvent from edge
      - 1000 to 10,000s water molecule
      - Computationally expensive



Water refinement

compare structures in vacuum to water

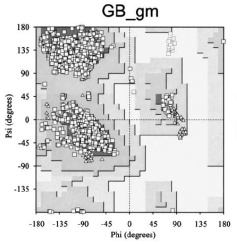
- no visible difference



Water refinement

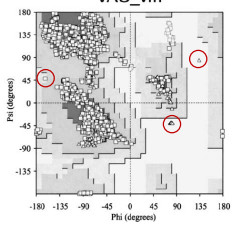
subtle, but significant improvements
- compare structures in vacuum to water
  - improves NH to CO hydrogen bonds
  - improves ϕ and ψ angle distributions

GB_gm



Residues in most favorable region 90.7
Residues in additional allowed region 8.4

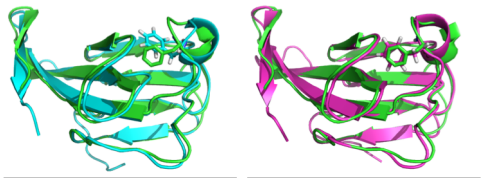
VAC_vm



Residues in most favorable region 86.6
Residues in additional allowed region 12.1

to keep or not to keep manual assignments

[do not keep]

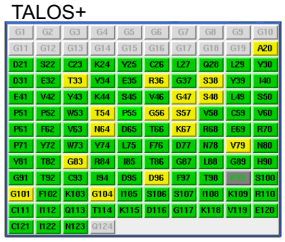


Phe102 HZ has been assigned based on unique NOE cross-peaks
CYANA consistently rejected these NOEs
Xray structure confirmed our suspicions
We fixed only 3 NOEs from Phe102 HZ
Introducing Val/Leu stereospecific assignments resolved the problem

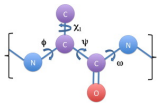
TALOS predictions and their effect on NMR structures

TALOS

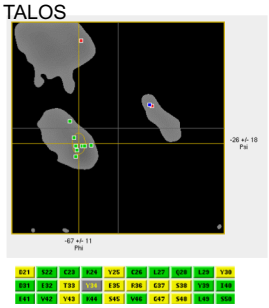
TALOS+



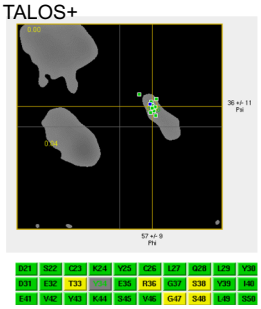
TALOS vs Xray (1)



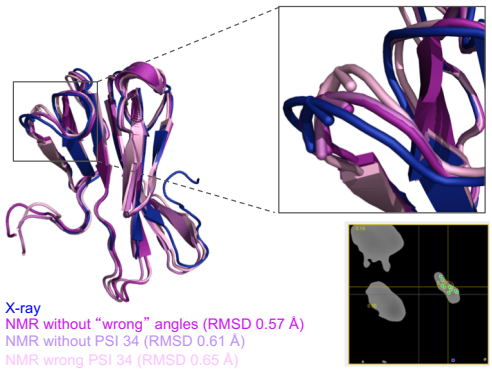
TALOS



TALOS+



TALOS in structure calculations



X-ray
NMR without "wrong" angles (RMSD 0.57 Å)
NMR without PSI 34 (RMSD 0.61 Å)
NMR wrong PSI 34 (RMSD 0.65 Å)

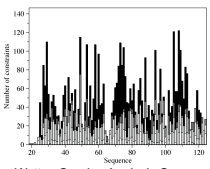
what is a good NMR structure IPSE (106aa)

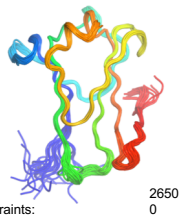
Input spectra
hNH
hCH
hCH2
cNH_
CcH_
hH_noN_
hH

selected peaks: 9345
assigned: 8865 (94.8%)
unassigned: 480

sequential:	499
intra-residual:	651
medium-range:	286
long-range:	1214
total:	2650

ramachandran	
core:	91.1%
allowed:	8.9%
generous:	0.0%
disallowed:	0.0%





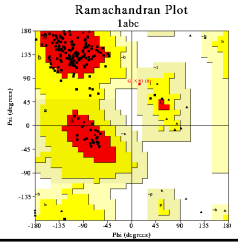
Wattos Surplus Analysis Summary

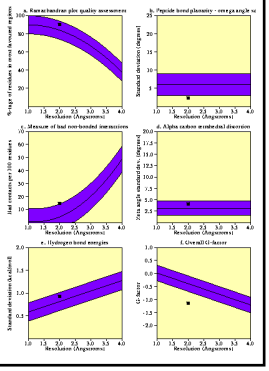
Found number of to do constraints:	2650
Found number of exceptional constraints:	0
Found number of constraints to be double with others:	17
Found number of impossible constraints:	0
Found number of fixed constraints:	2
Found number of redundant constraints:	1
Found number of non-redundant constraints:	2630
Found number of constraints to be surplus (E+C+D+I+F+R):	20

Overall NOE completeness is 68.10 percent

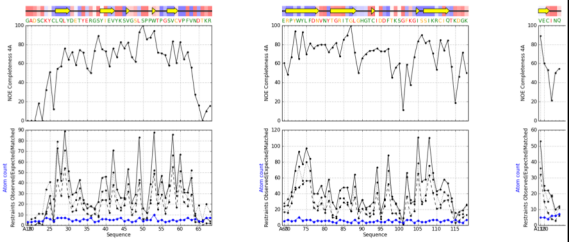
PROCHECK analysis

ϕ, ψ, χ_1, χ_2 distribution
Comparison of main chain and side-chain parameters to standard values

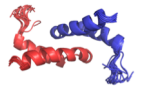


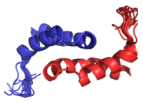


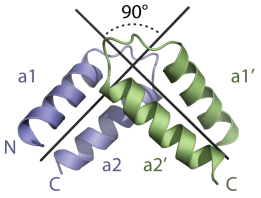
Wattos analysis



examples
Qua1 symmetric dimer



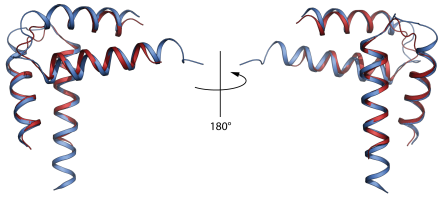




Full NOE set + RDCs

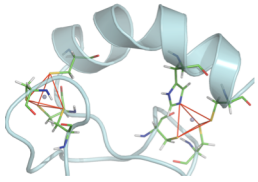
- two 13C edited noesy spectra as input
- no filter NOESY experiments

crystals coming to your rescue



bb rmsd 1.4 Å
dimer vs dimer

examples
MYND “the sinful structure”
45 aa !!!



- Structure calculation without the zinc atoms
- Identification of the zinc coordination residues from the fold
- Repeat calculation with the zinc atoms fixed

Defining tetrahedra

CYANA uses only distances

S - ZN: 2.3 **S - S: 3.65**

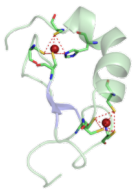
N - ZN: 2.0 **S - N: 3.35**

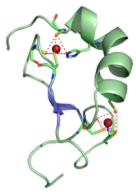
CNS uses both distances and angles definitions with possibility of using different weights

S – ZN: 2.3 **S – ZN – S : 109.5**

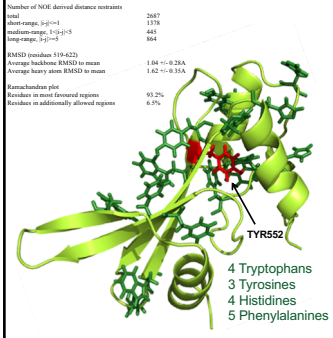
N – ZN: 2.0 **N – ZN – S : 120**

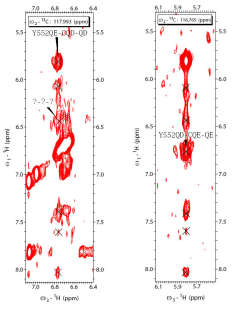
In both cases one needs to give the Zn chelating residues





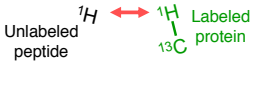
examples
aromats troubleshooting

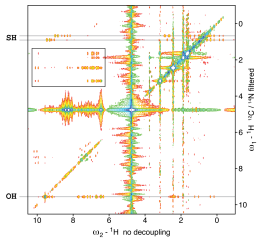




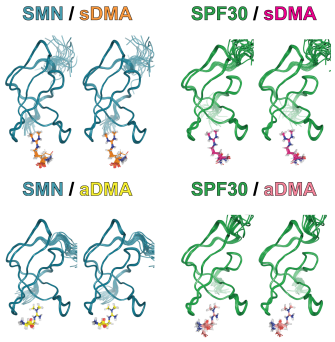
Intermolecular NOEs

Filtered/Edited NOE: based on selection of NOEs from two molecules with unique labeling patterns





Protein-ligand structures



Summary

- CYANA will determine the correct fold
- you should take care for the input data
- you should take care for the local geometry
- understand how CNS works to refine your structure

In general to determine an NMR structure is (not) straightforward