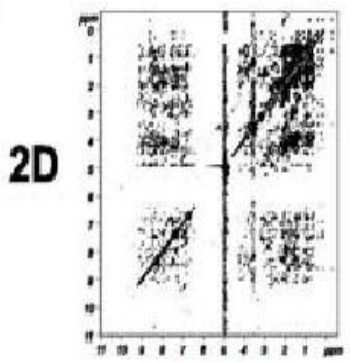
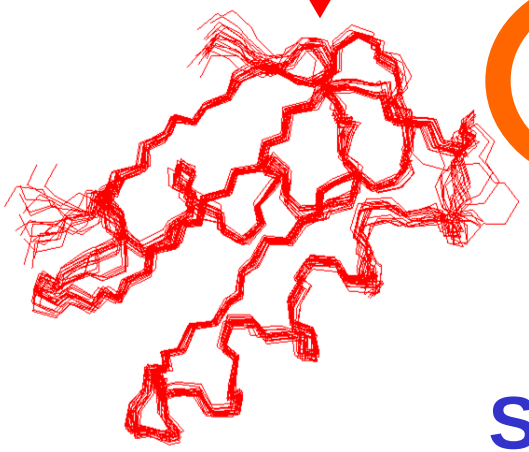
NMR Spectral Assignment and Structural Calculations

Lucia Banci

CERM – University of Florence

Structure determination through NMR





Protein Sample

↓

NMR spectroscopy

↓

Sequential resonance assignment

↓

Collection of conformational constraints

↓

3D structure calculations

↓

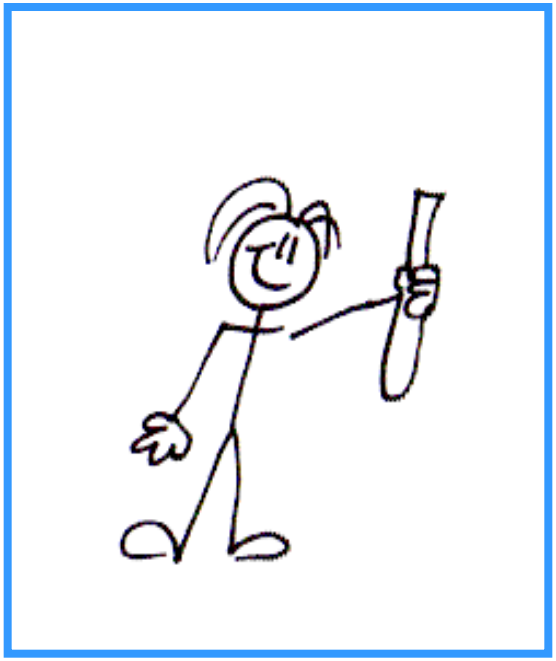
Structure refinement and Analysis

The protein in the NMR tube!

- Protein overexpression
- Purification
- $^{15}N/^{13}C$ labelling

< 25 KDa
About 240 AA → **^{13}C, ^{15}N labeling**

> 25 kDa
About 240 AA → **^{13}C, ^{15}N labeling**
+ ^{2}H labeling necessary!!

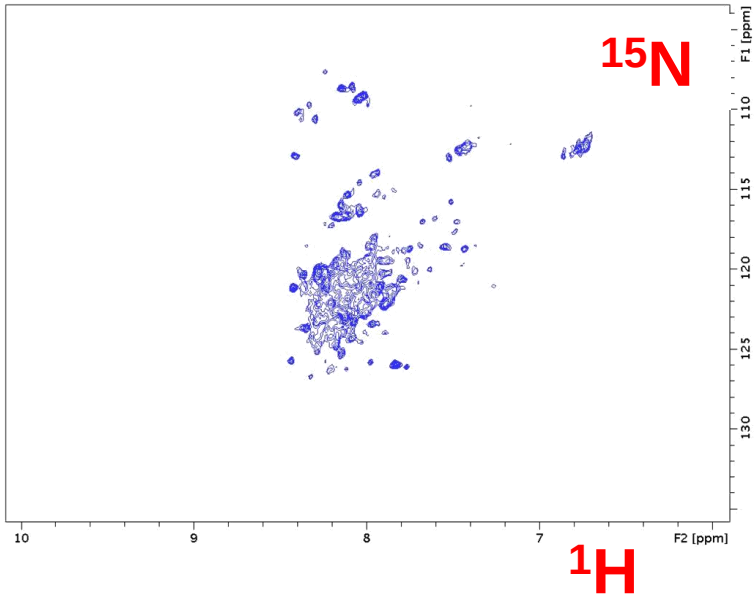


Which experiments should I run?

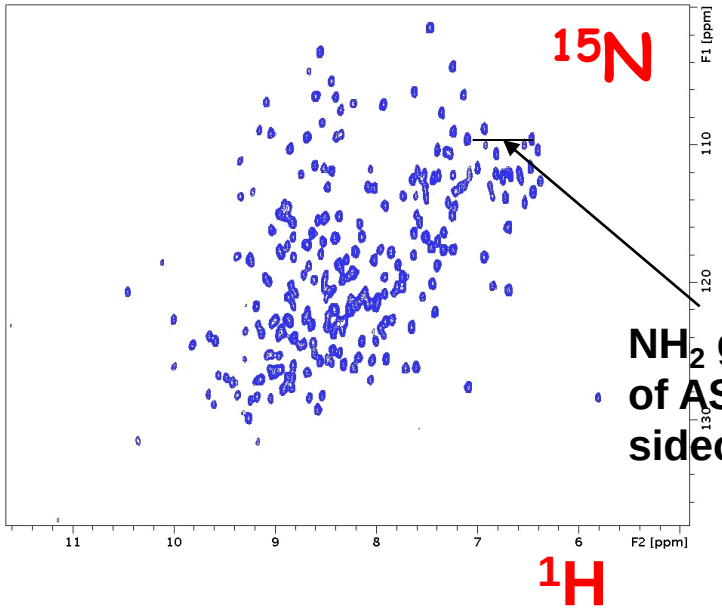
Is my sample OK for NMR?

^{1}H-^{15}N HSQC gives the protein fingerprint

unfolded



folded



Signals of unfolded proteins have little ^{1}H dispersion, that means the ^{1}H frequencies of all residues are very similar.

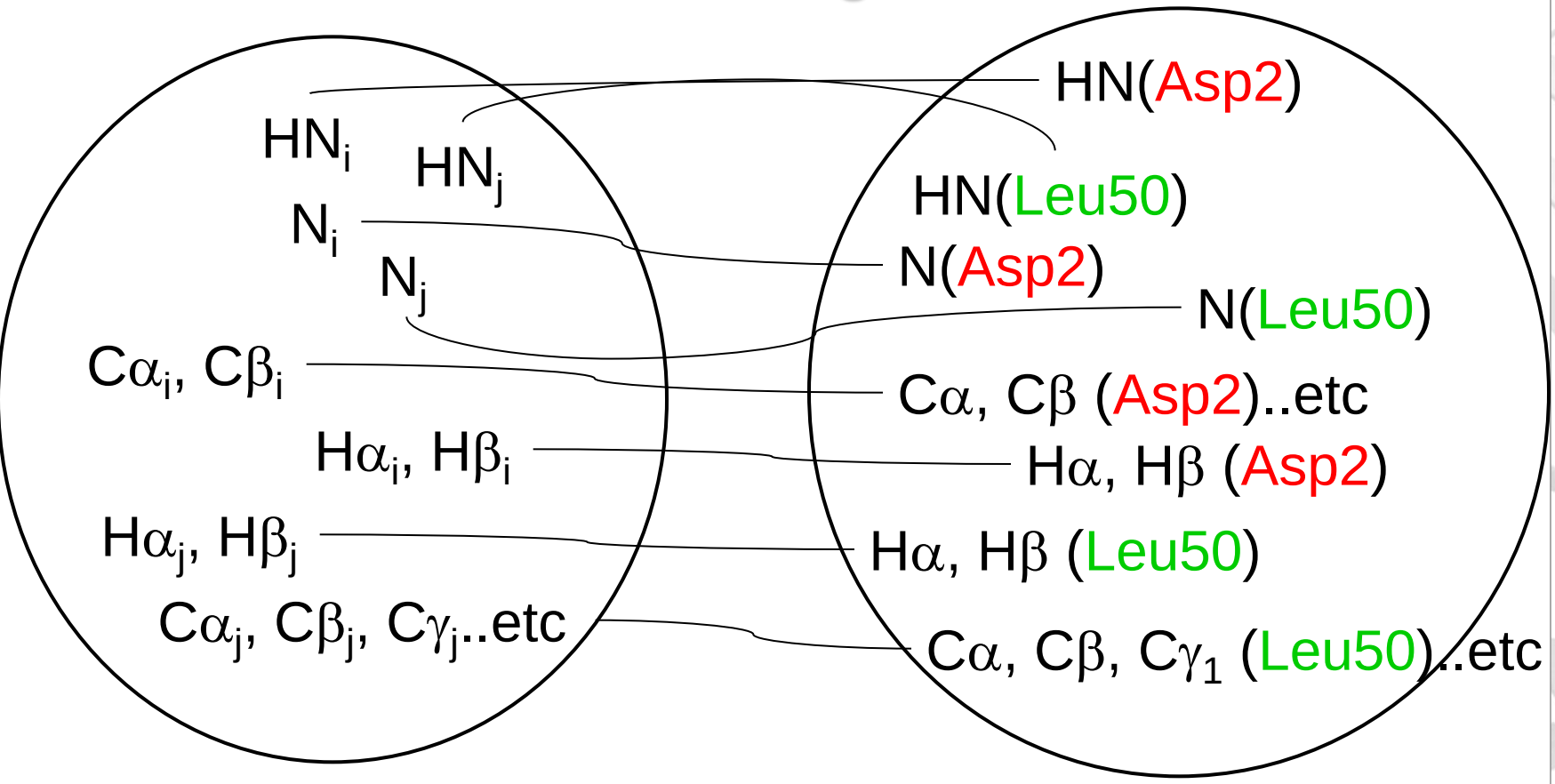
Folded proteins have larger dispersion

Can I see all the peaks I expect?

Count the peaks! → Backbone NH (excluding prolines!)

Making resonance assignment

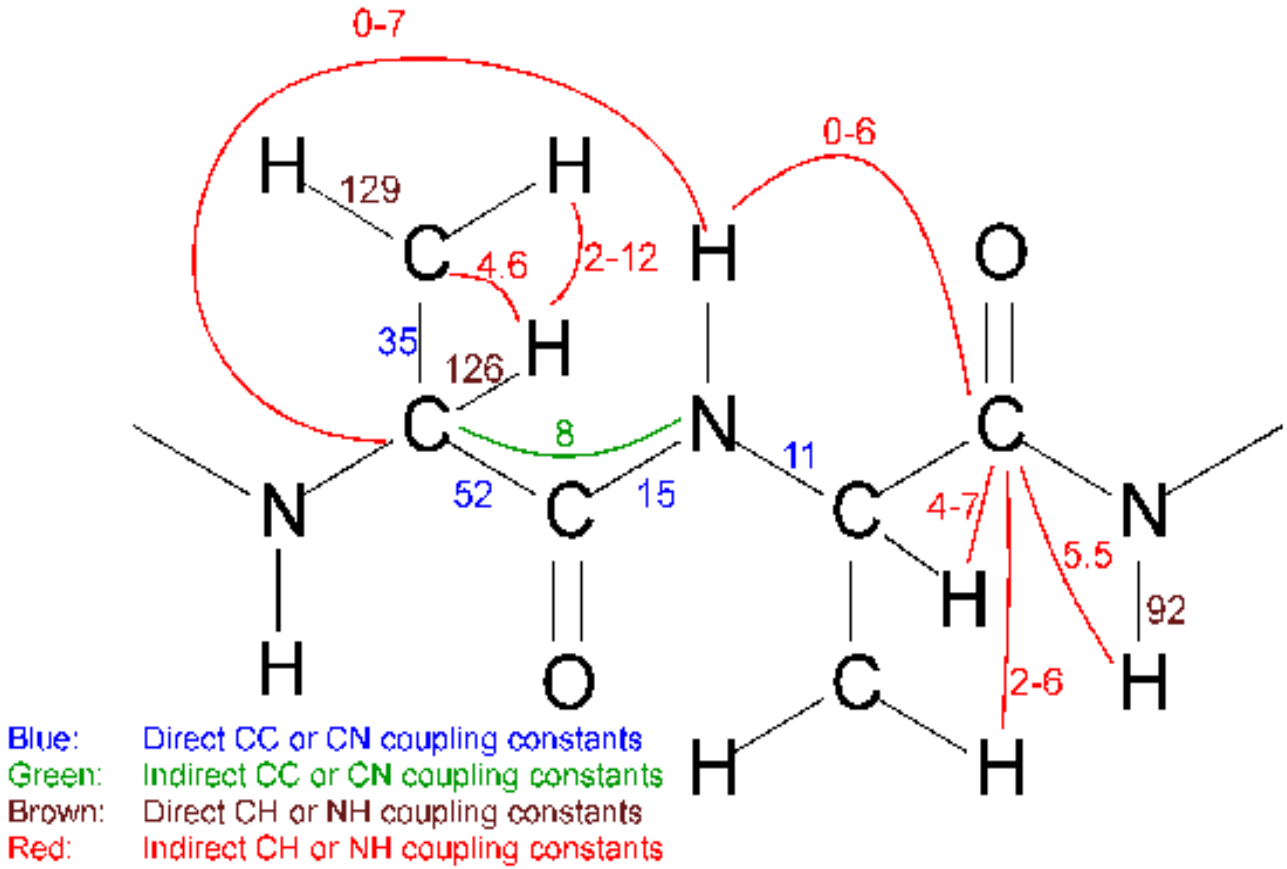
What does it mean to make sequence specific resonance assignment ?



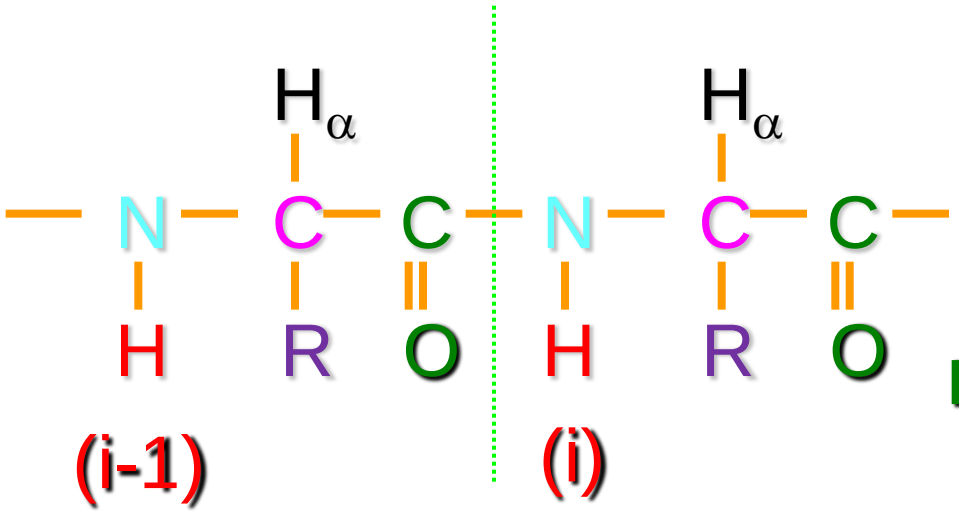
To associate each resonance frequency to each atom of the individual residues of the protein

Assignment Strategy

The strategy for assignment is based on scalar couplings



Triple resonance experiments have made assignment easy and fast



HNCO $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}CO_{(i-1)}$

HN(CO)CA $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}C_{\alpha\,(i-1)}$

HNCA { $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}C_{\alpha\,(i)}$; $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}C_{\alpha\,(i-1)}$

(HCA)CO(CA)NH { $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}CO_{(i)}$; $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}CO_{(i-1)}$

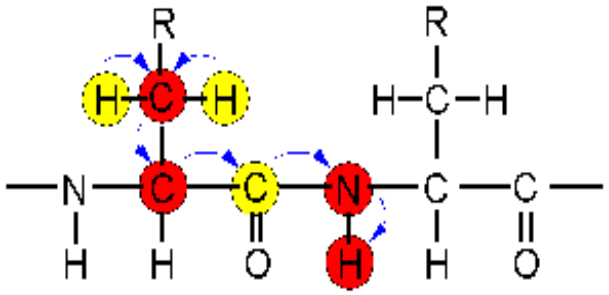
CBCA(CO)NH $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}C_{\alpha,\,\beta\,(i-1)}$

H(C)CH- TOCSY $^{1}H_{(i)}$-$^{13}C_{(i)}$-$R_{(i)}$

CBCANH { $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}C_{\alpha,\,\beta\,(i)}$; $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}C_{\alpha,\,\beta\,(i-1)}$

CC(CO)NH $^{1}H_{(i)}$-$^{15}N_{(i)}$-$^{13}C_{R\,(i-1)}$

Experiments for backbone assignment



CBCA(CO)NH

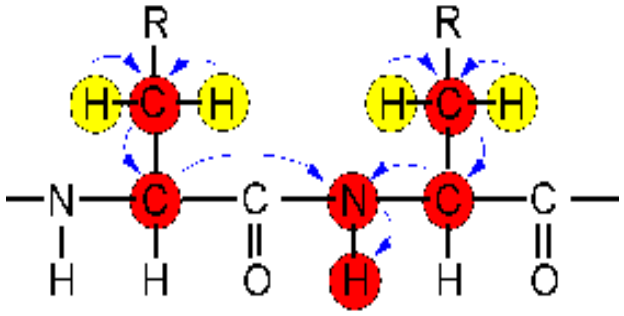
$^{1}H_{i}$-$^{15}N_{i}$-$^{13}C^{\alpha}_{i-1}$$^{13}C^{\beta}_{i-1}$

Res_{i-1} Res_{i}

CBCANH

$^{1}H_{i}$-$^{15}N_{i}$-$^{13}C^{\alpha}_{i}$$^{13}C^{\beta}_{i}$

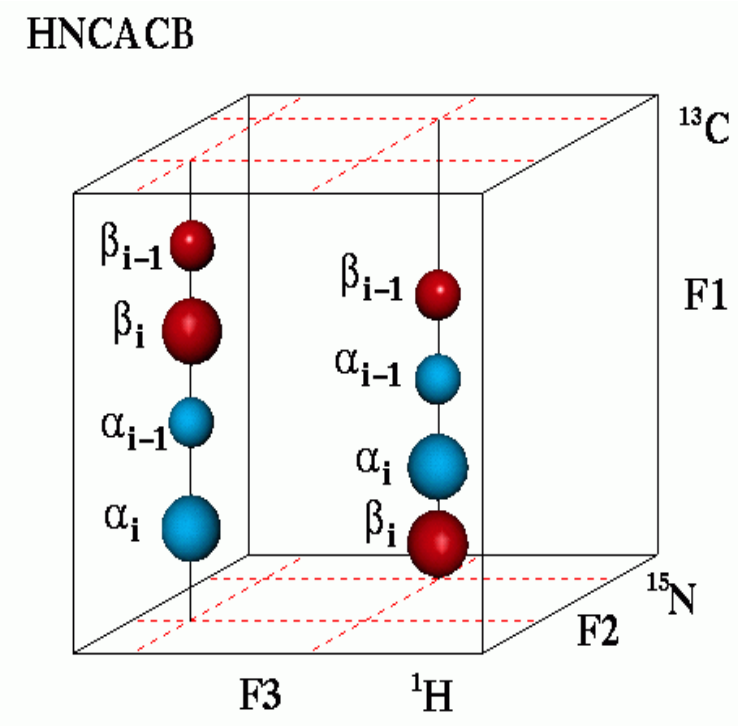
$^{1}H_{i}$-$^{15}N_{i}$-$^{13}C^{\alpha}_{i-1}$$^{13}C^{\beta}_{i-1}$

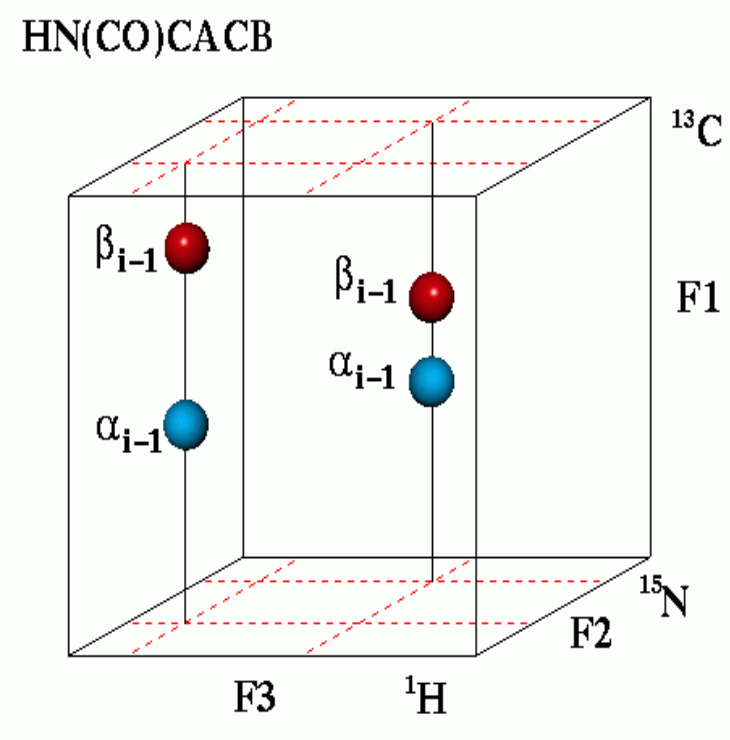


Res_{i-1} Res_{i}

CBCA(CO)NH and CBCANH correlate amide protons via Cα and Cβ resonances.

Experiments for backbone assignment

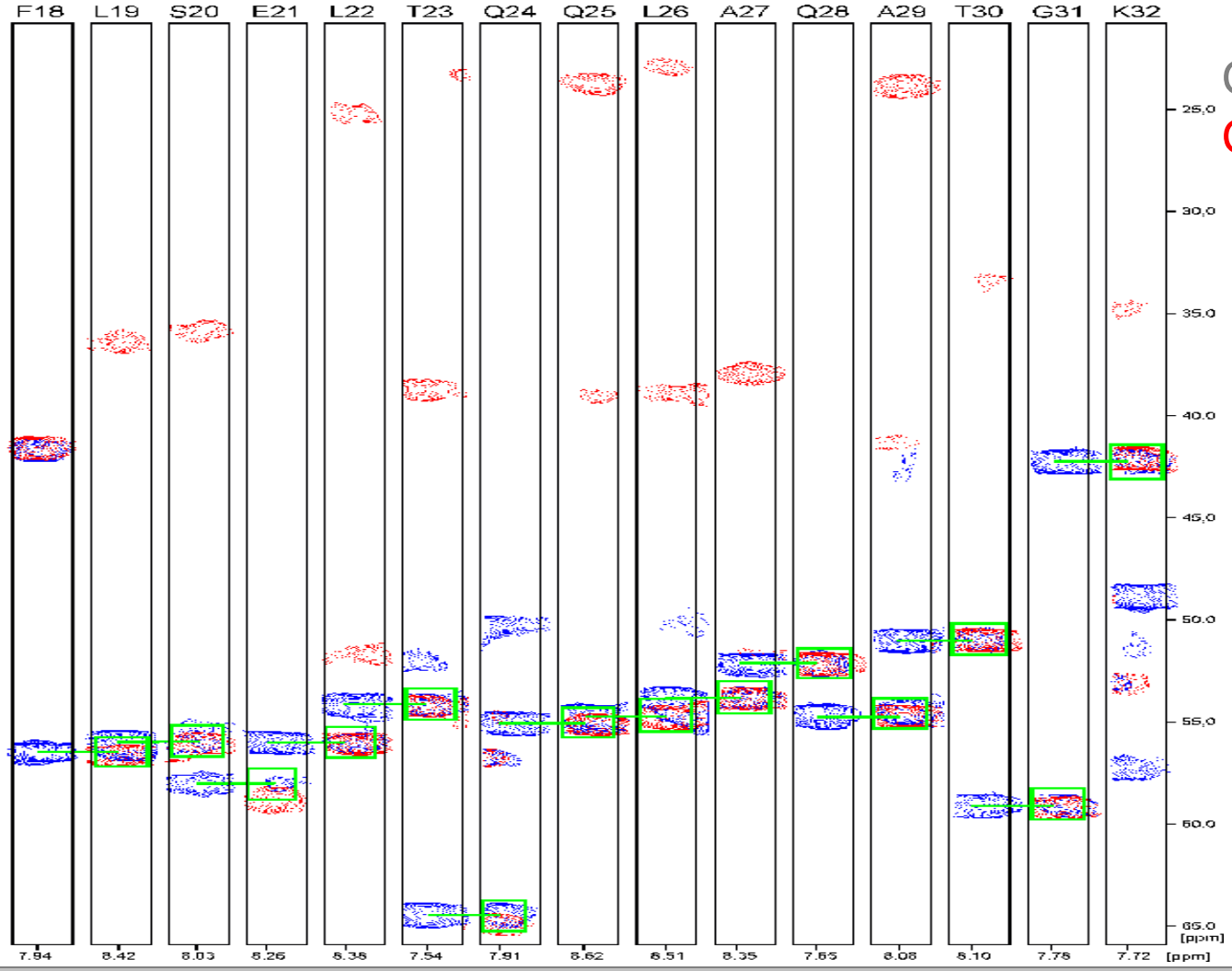




The chemical shifts of Cα and Cβ atoms can be used for a preliminary identification of the amino acid type.

Sequential Assignment

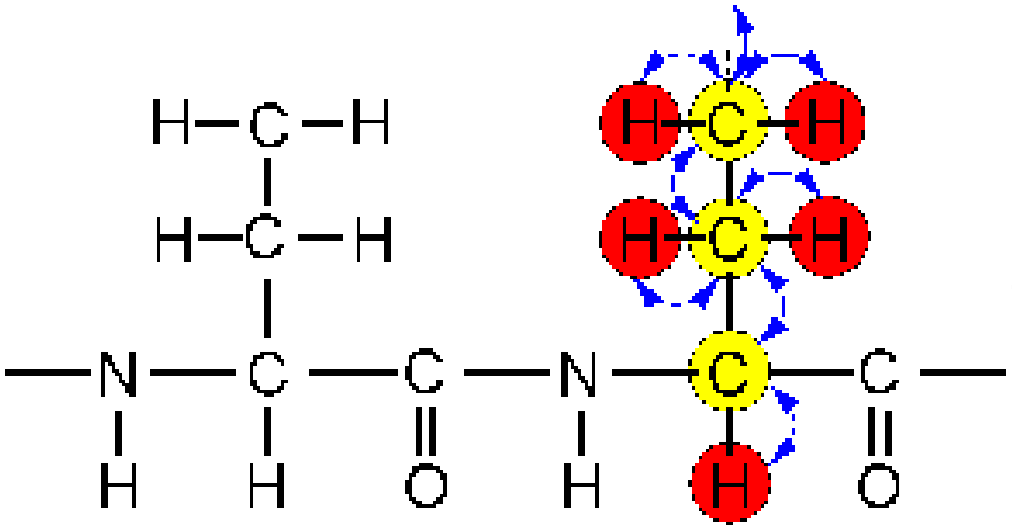
The 'domino pattern' is obtained during the sequential assignment with triple resonance spectra



CB CANH
CBCA(CO)NH

Green boxes indicate sequential connectivities from each amino acid to the preceeding one

Experiment for side-chain assignment

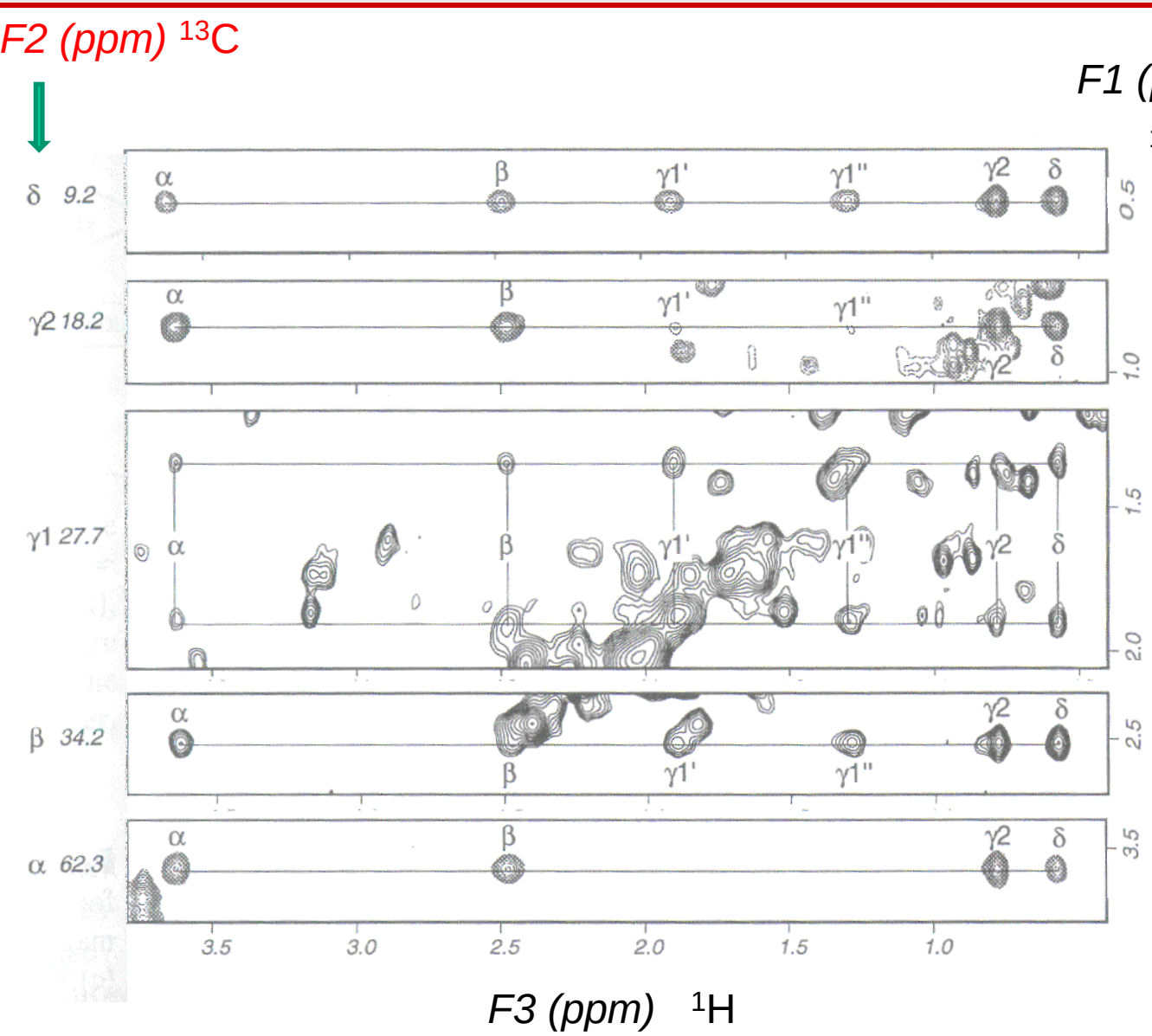


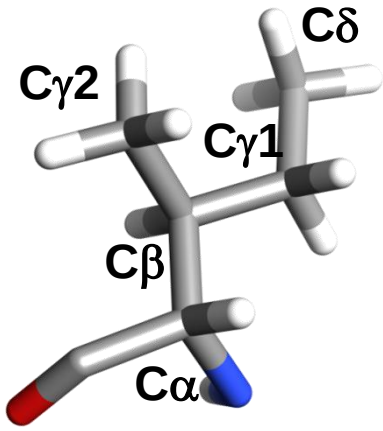
$^1H_i\alpha$, $^1H_i\beta$, $^1H_i\gamma1$.....

Res_{i-1} Res_i

In HCCH-TOCSY, magnetization coherence is transferred, through ^1J couplings, from a proton to its carbon atom, to the neighboring carbon atoms and finally to their protons.

hCCH-TOCSY experiment





Isoleucine

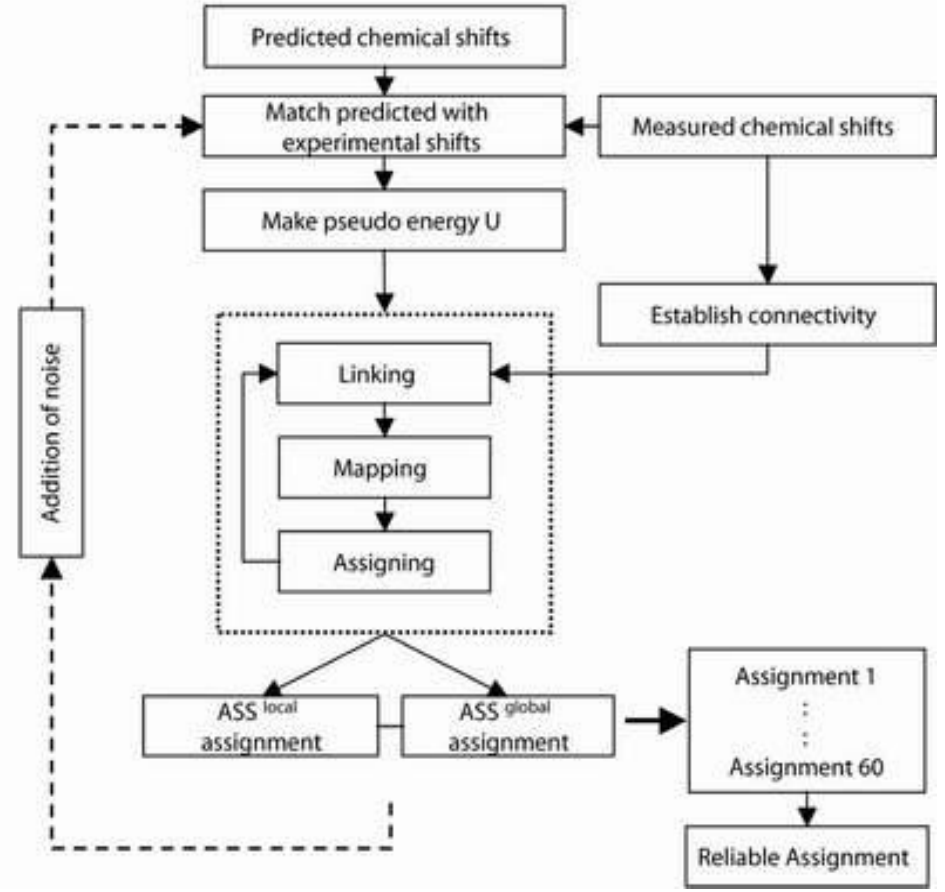
Automated assignment programs

MARS

For automated backbone assignment (NH, CO, Cα, Cβ) .
It requires manually pick-peaking of 3D spectra for backbone assignment, such as CBCAHN, CBCACOHN etc

Input:

- Primary sequence
- Spectral data, i.e chemical shifts of resonances grouped per residue and those of its preceding residue.
- Chemical shift tolerances
- Secondary structure prediction data (PSI-PRED)



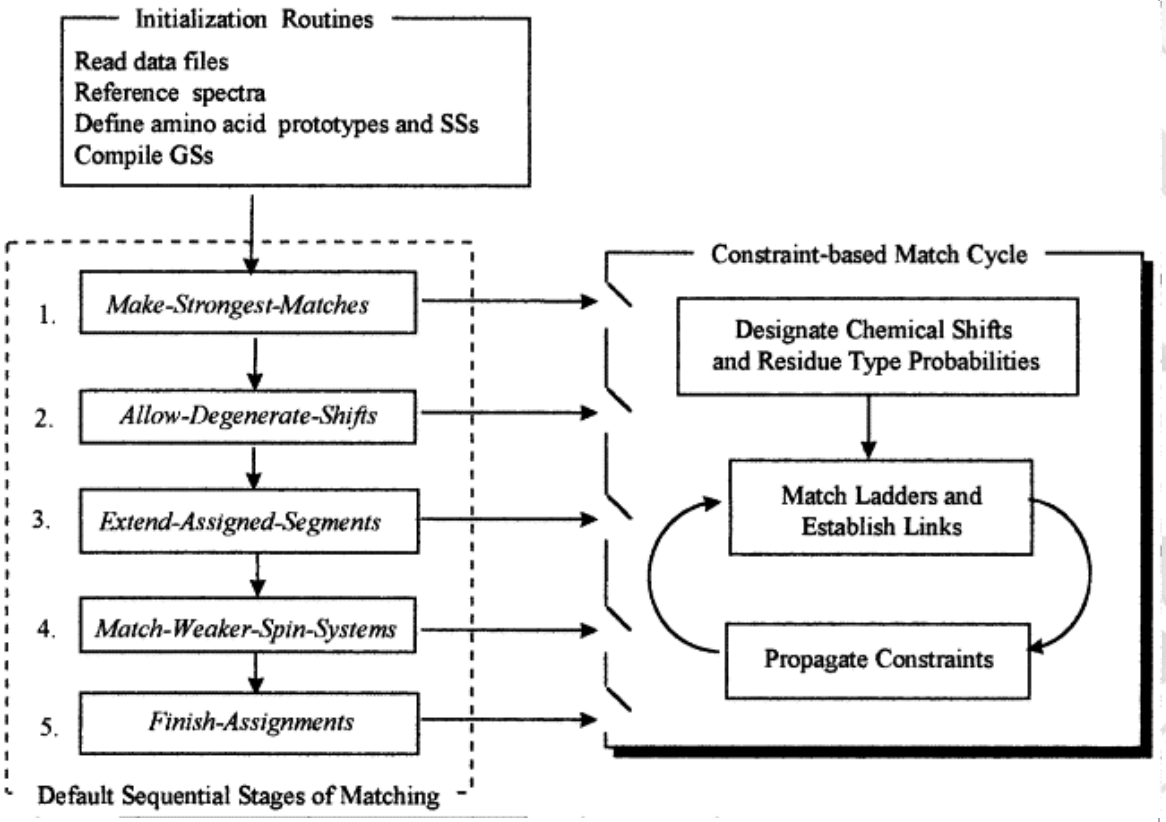
Automated assignment programs

AutoAssign

For automated backbone assignment (HN, NH, CO, Cα, Cβ, Hβ and Hα)
It requires manually pick-peaking of 3D spectra for backbone assignment, such as CBCAHN, CBCACOHN etc.

Input:

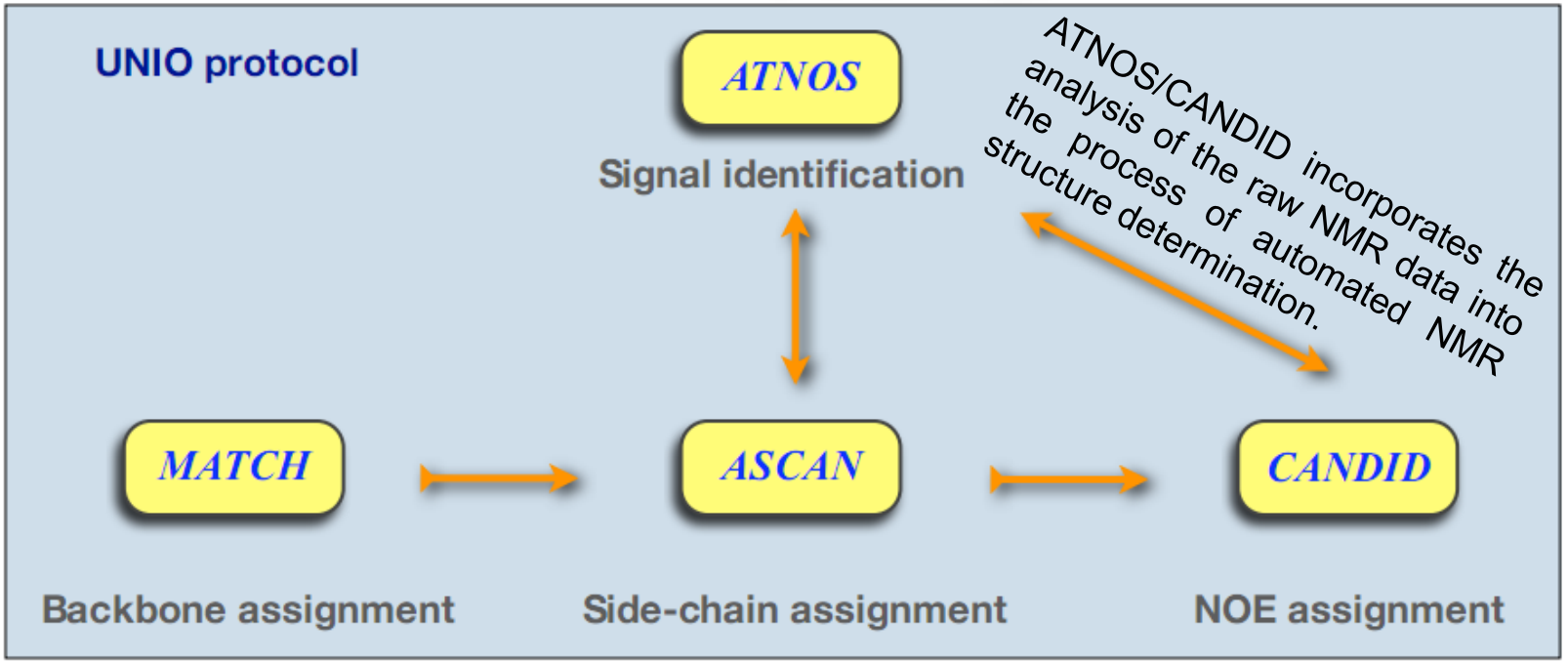
- peak list table of triple resonance spectra
- primary sequence



Automated assignment programs

UNIO

NMR data analysis interconnects the MATCH algorithm for backbone assignment, the ASCAN algorithm for side-chain assignment directly on NMR spectra

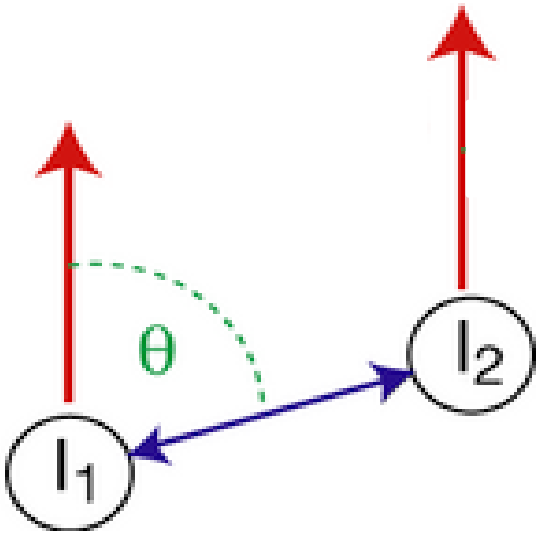


Conformational restraints

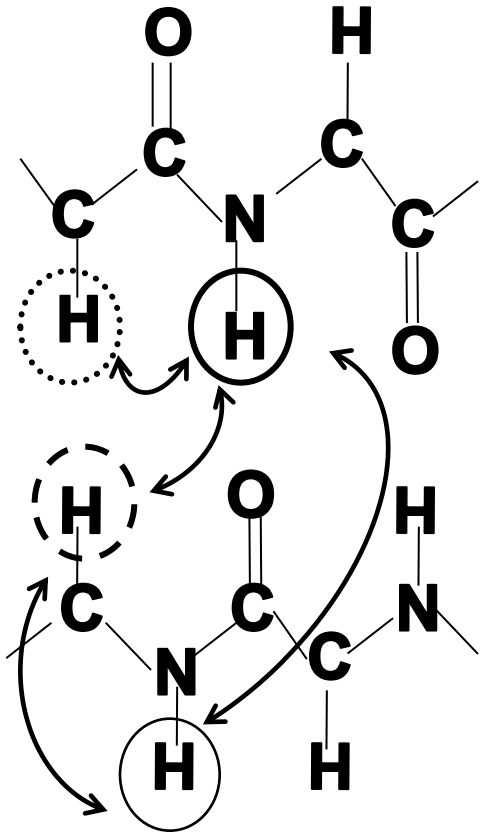
NMR experimental data		Structural restraints
NOEs	→	Proton-proton distances
Coupling constants	→	Torsion angles
Chemical shifts	→	Torsion angles
H -bonds	→	Proton-proton distances
RDCs	→	Bond orientations
Relaxation times	→	Metal-nucleus distances
PCSs	→	{ Metal-nucleus distances Orientation in the metal χ frame
Contact shifts	→	Torsion angles

Distance constraints

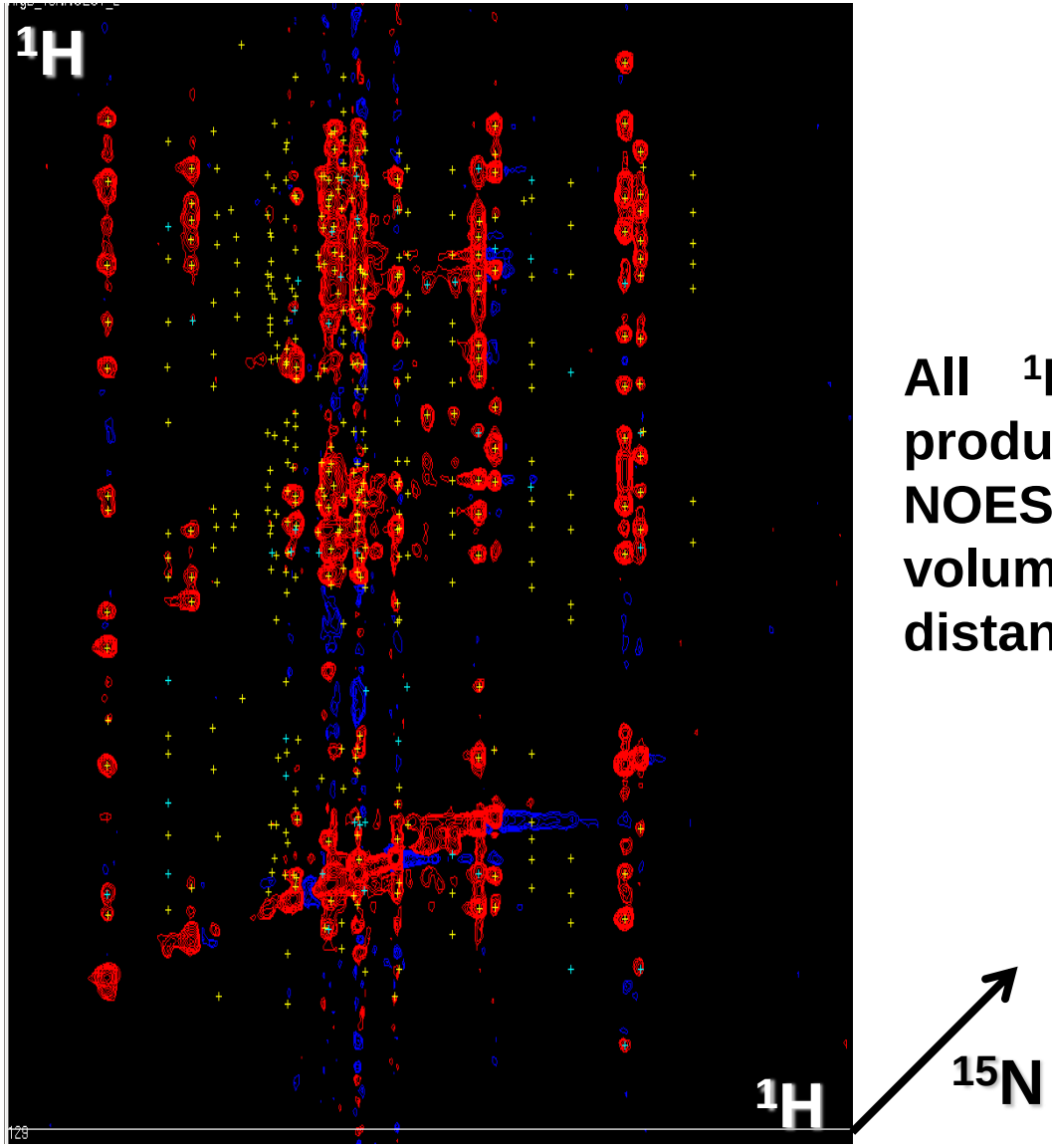
NOESY volumes are proportional to the inverse of the sixth power of the interproton distance (upon vector reorientational averaging)



$$\eta_{IJ} \propto \frac{\tau_c}{r_{IJ}^6}$$



The NOESY experiment:



All ^{1}H within 5-6 Å can produce a cross-peak in NOESY spectra whose volume provides ^{1}H-^{1}H distance restraints

How are the distance constraints obtained from NOEs intensities?

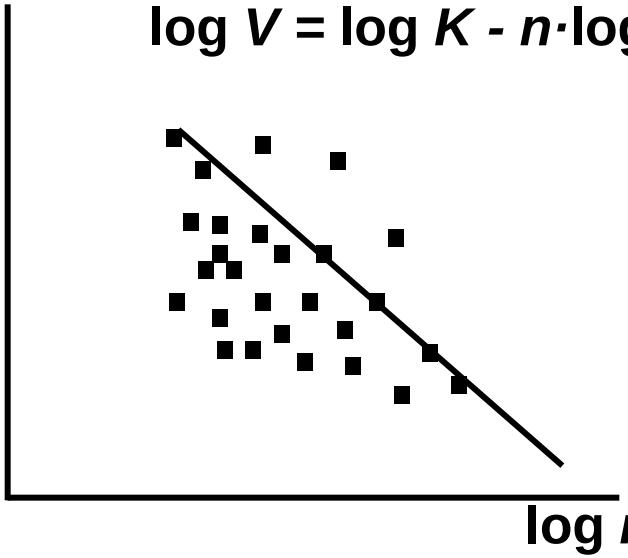
The NOESY cross-peak intensities (V) are converted into upper distance limits (r) through the relation:

$$V = \frac{K}{r^{n}}$$

where K is a constant and *n* can vary from 4 to 6.

***K* constant is initially determined from NOE's between protons at fixed distance**

$$\log V = \log K - n \cdot \log r$$



Classes of constraints

1. Backbone	$V = A/d^6$
2. Sidechain	$V = B/d^4$
3. Methyl	$V = C/d^4$

Wuthrich, K. (1986) "NMR of Proteins and Nucleic Acids"

How are the distance constraints obtained from NOEs intensities?

The NOESY cross-peak intensities are converted into upper distance limits

Classes of restraints		Distance ranges
1. Very Weak	0 – 20%	1.8–6.0 Å
2. Weak	20 – 50%	1.8–5.0 Å
3. Medium	50 – 80%	1.8–3.3 Å
4. Strong	80 –100%	1.8–2.7 Å

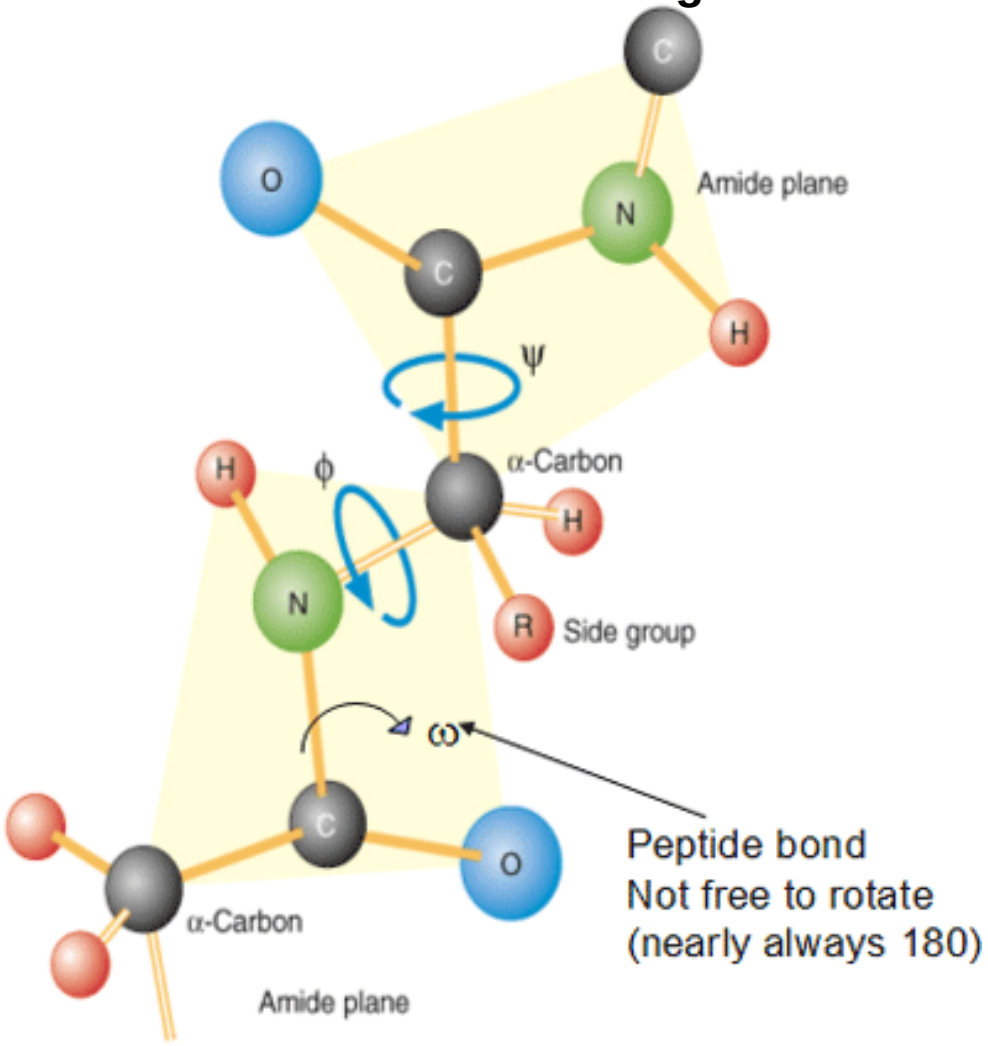
0.5 Å are added to the upper bound of distances involving methyl groups in order to correct for the larger than expected intensity of methyl crosspeaks

Xplor-NIH Calibration of NOEs

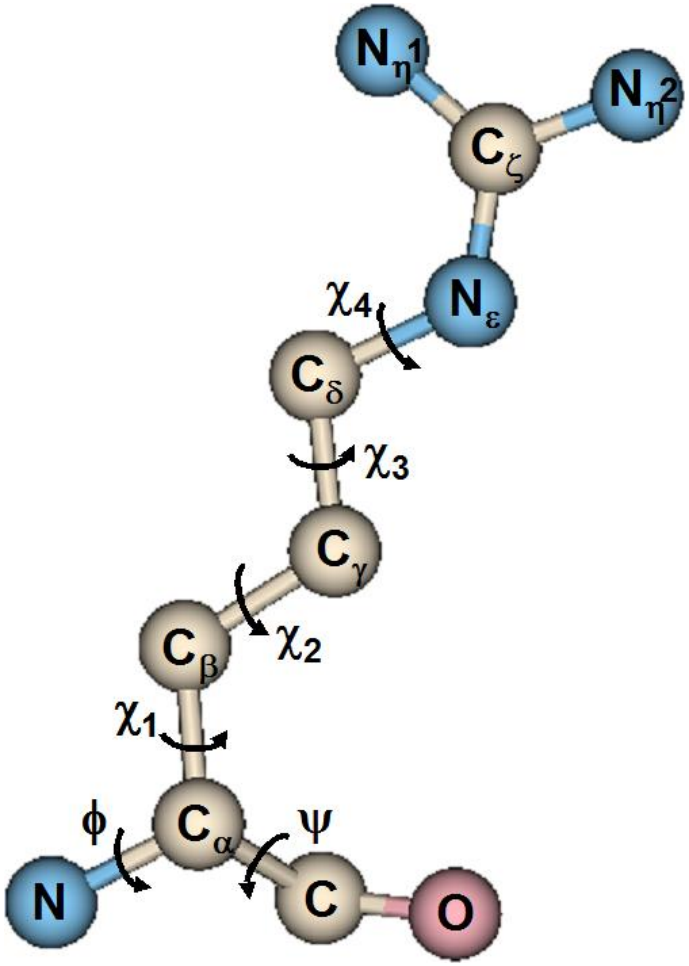
J. J. Kuszewski, R. A. Thottungal, G. M. Clore, Charles D. Schwieters *J Biol NMR* 2008

Dihedral angles

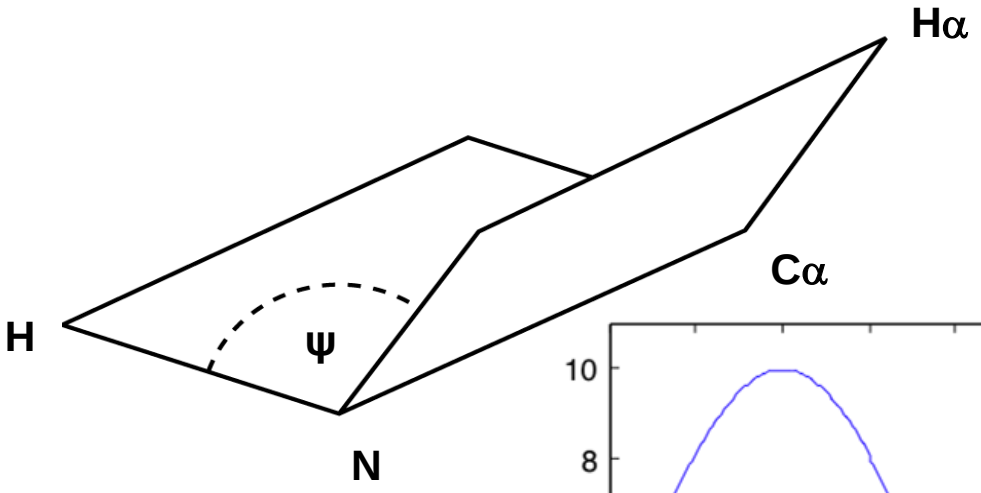
Backbone dihedral angles



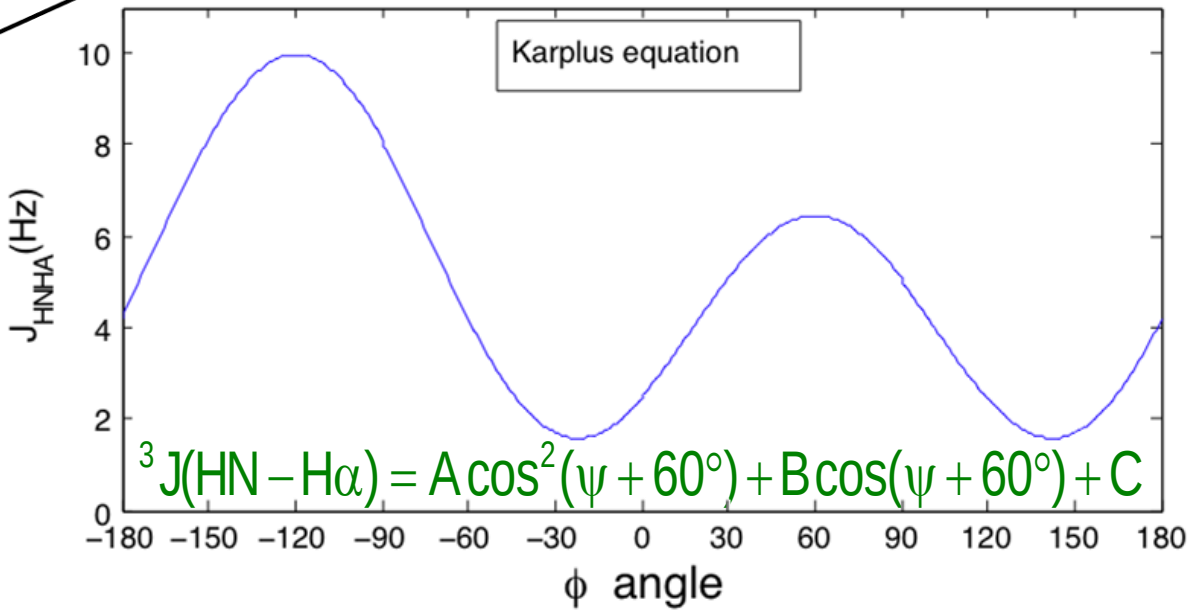
Sidechains dihedral angles



Dihedral angle restraints



^{3}J coupling constants are related to dihedral angles through the Karplus equation



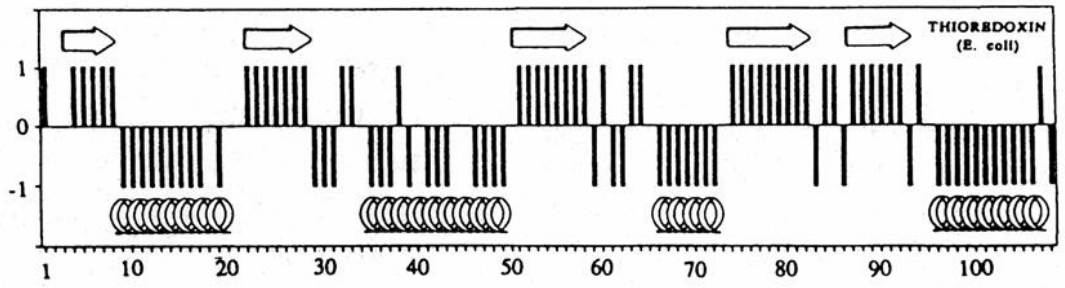
$J_{HNH\alpha}$ > 8Hz
$J_{HNH\alpha}$ < 4.5Hz
4.5Hz < $J_{HNH\alpha}$ < 8Hz

– 155° < ϕ=ψ+120° < – 85°
– 70° < ϕ=ψ+120° < – 30°
ϕ,ψ values depend on $J_{HNC'}$

Chemical Shift Restraints

As chemical shifts depend on the nucleus environment, they contain structural information. Correlations between chemical shifts of Cα, Cβ,CO, Hα and secondary structures have been identified.

Chemical Shift Index:



CSI's are assigned as:

Carbon chemical shift difference with respect to reference random coil values:

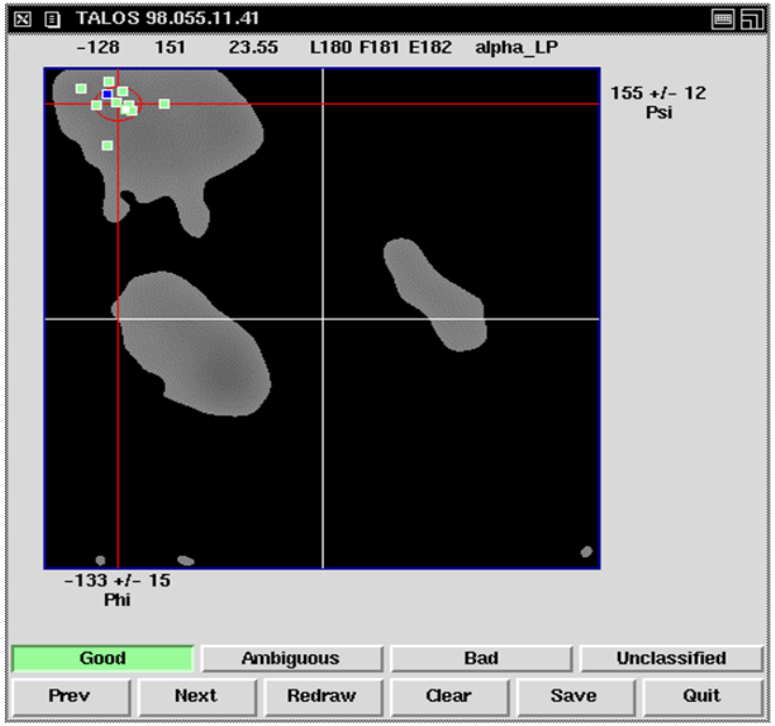
-0.7 ppm < $\Delta\delta$ < 0.7 ppm	0
$\Delta\delta$ < - 0.7 ppm	-1
$\Delta\delta$ > +0.7 ppm	+1

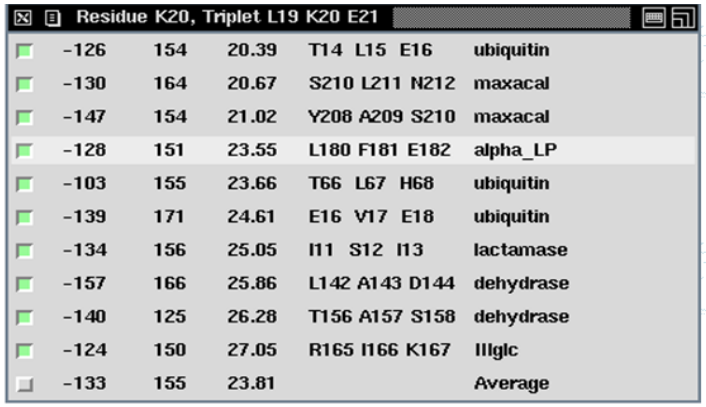
Any "dense" grouping of four or more "-1's", uninterrupted by "1's" is assigned as a helix, while any "dense" grouping of three or more "1's", uninterrupted by "-1's", is assigned as a sheet. Other regions are assigned as "coil".

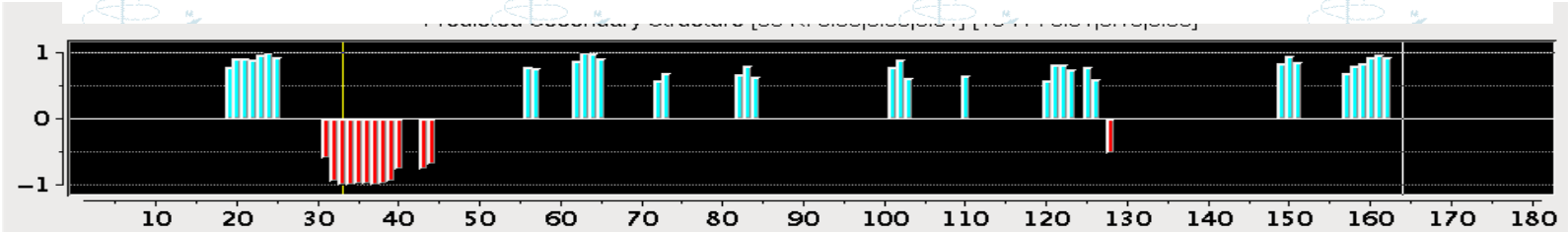
A "dense" grouping means at least 70% nonzero CSI's.

Chemical Shift Restraints

TALOS+ uses $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, $^{1}H\alpha$ and ^{15}N chemical shifts together with sequence information/chemical shift databases to predict values for backbone dihedral angles φ and ψ







Shen, Delaglio, Cornilescu, Bax *J. Biomol NMR*, 2009

H-bonds as Structural restraints

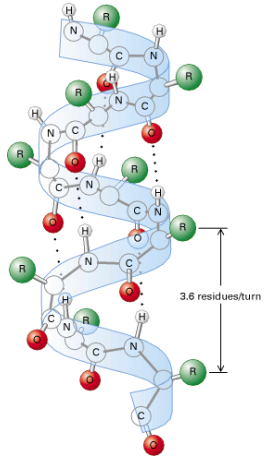
Experimental Determination of H-Bonds:

- **HNCO** — direct method
- **H/D exchange** — indirect method

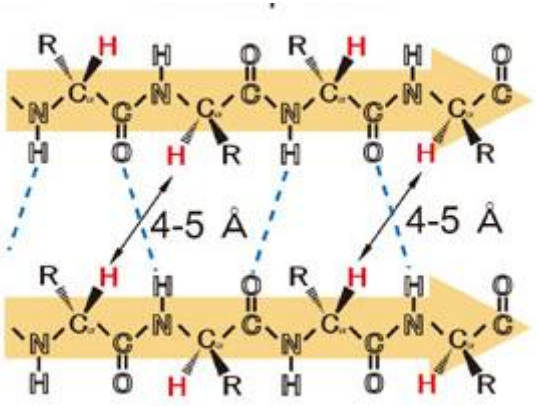
Distance and angle restraints

- **Upper distance limit**
- **Lower distance limit**

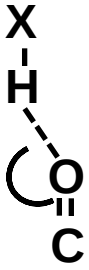
α-Helix



β–Sheet

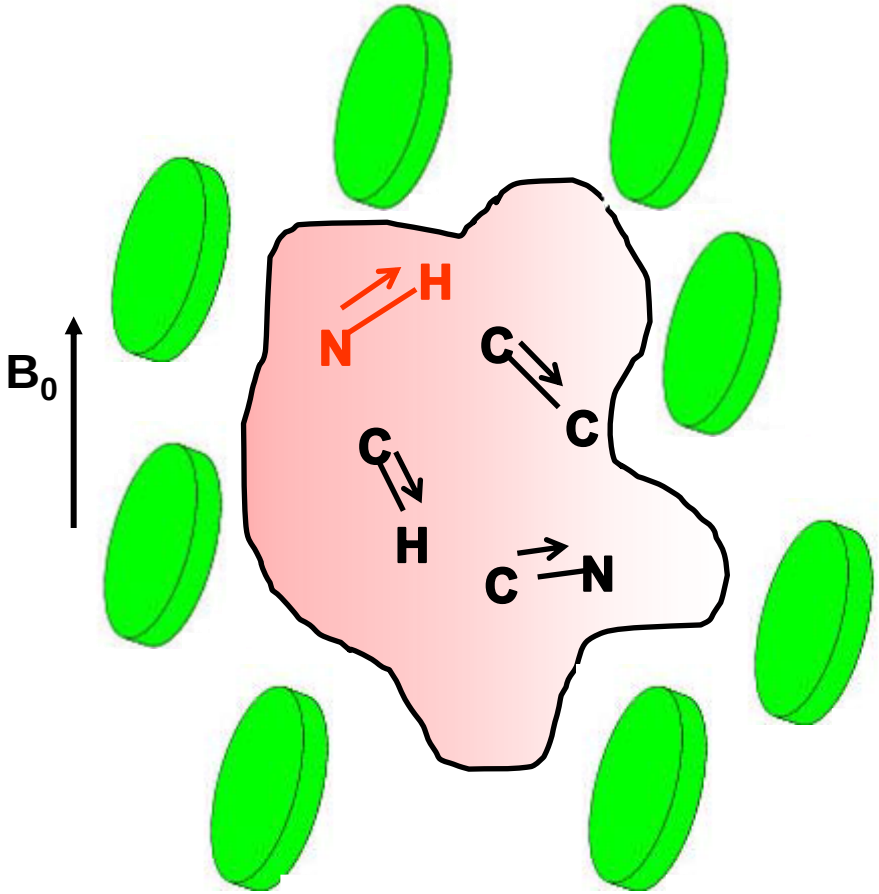


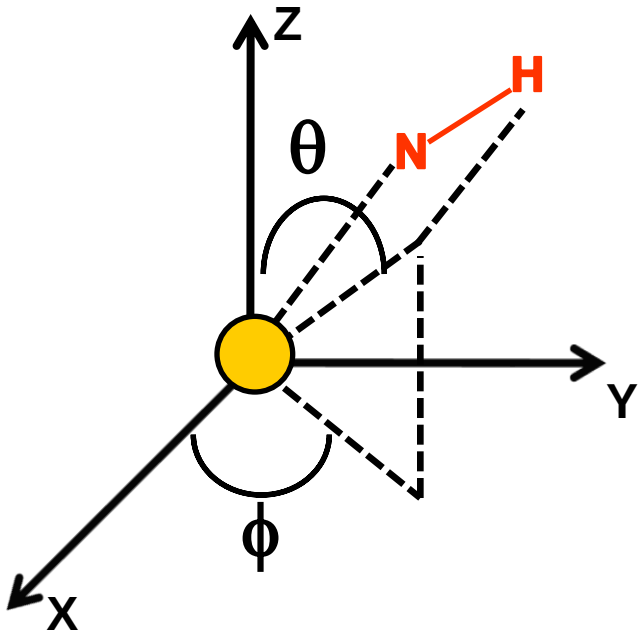
140° > N-H···O > 180°



X-H···O=C ~160°

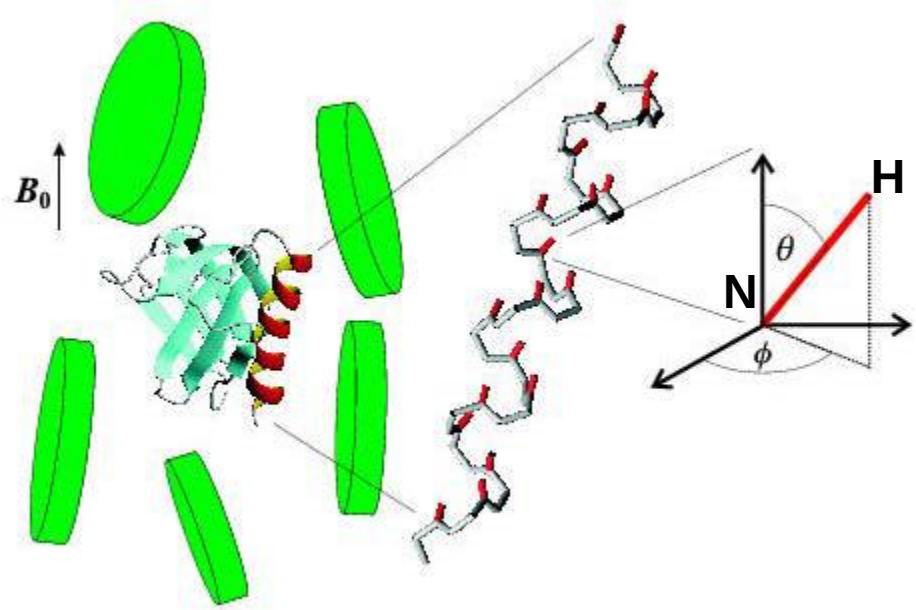
Residual dipolar couplings





RDCs provide information on the orientation of (in principle each) bond-vector with respect to the molecular frame and its alignment in the magnetic field

Residual dipolar couplings



$$RDC_{(IS)_i} \propto \Delta\chi \; f(\theta_i, \varphi_i)$$

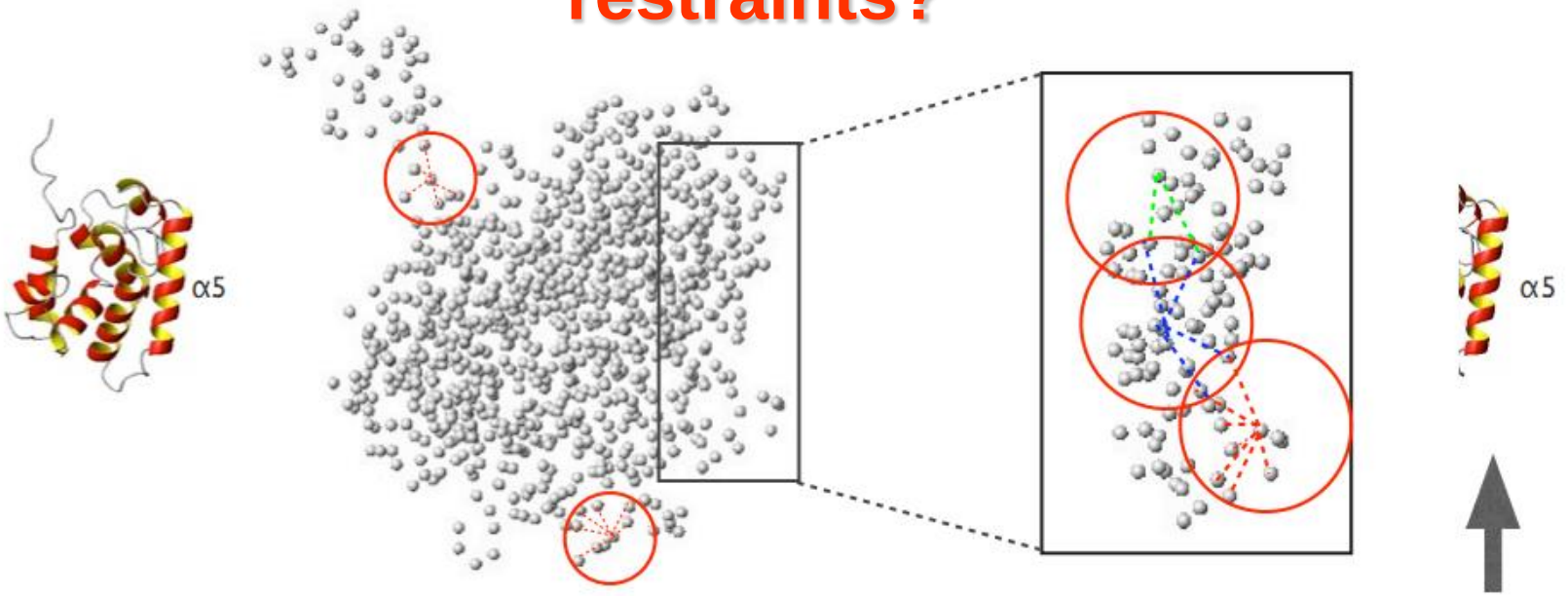
where χ is the molecular alignment tensor with respect to the magnetic field and θ_i, φ_i are the angles between the bond vector and the tensor axes

Proteins dissolved in liquid , orienting medium
Some media (e.g. bicelles, filamentous phage, cellulose crystallites) induce to the solute , some orientational order in a magnetic field
A small "residual dipolar coupling" results

Relative orientation of secondary structural elements can also be determined

General Consideration

How complete are the NMR Structural restraints?



Hydrogen cloud of the protein

Successive distance measurements in an alpha helix.

100 of distance measurements with NMR.

NMR mainly determines short range structural restraints but provides a complete network over the entire molecule

Algorithms for 3D structure calculations

- Simulated annealing/MD in cartesian coordinates

  XPLOR-NIH

- Simulated annealing/MD in torsion angle space

  XPLOR-NIH and CYANA

Basic concepts on 3D solution structure calculations

- The various types of NMR parameters provide conformational restraints to be used in structure calculation

- Calculation of the 3D structure is performed as a minimization problem of a target or penalty function

- The target/penalty function measures the deviation of the restraints in a calculated conformation with respect to the experimental ones

Basic concepts for 3D solution structure calculations

- NMR data alone would not be sufficient to determine the position of all atoms in a biological macromolecule (protein)

- The experimental data are supplemented with information on the covalent structure of the protein (bond lengths, bond angles, planar groups...) and the atomic radii (i.e. each atom pair cannot be closer than the sum of their atomic radii)

Hybrid energy function

- A hybrid energy function is defined, that incorporates *a priori* information and NMR structural restraints as potential and pseudopotential energy terms, respectively

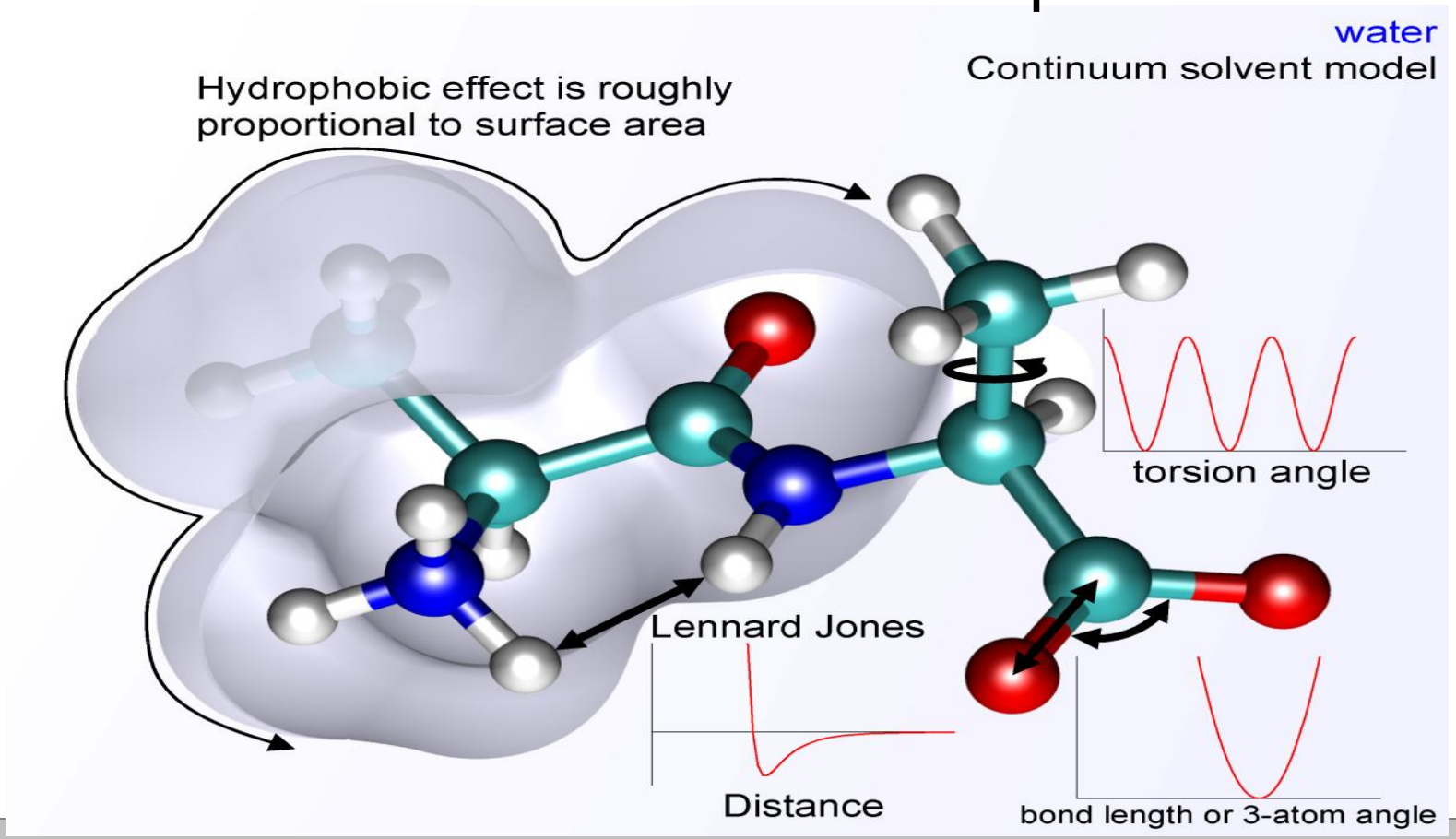
$$E_{hybrid} = \sum_{bonds} k_b (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2$$

$$+ \sum_{dihedrals} k_\phi (1 + \cos(n\phi + \delta)) + \sum_{\substack{nonbonded \\ pairs}} k_{nb} (r - r_0)^2$$

$$+ \sum_{\substack{distance \\ restraints}} k_d (d - d_0)^2 + \sum_{\substack{torsional \\ restraints}} k_\psi (\psi - \psi_0)^2 + \ldots$$

Potential energy terms: example

- Simplified description of the forces in the system
- Potential energy differs from zero if the conformation deviates from the equilibrium one

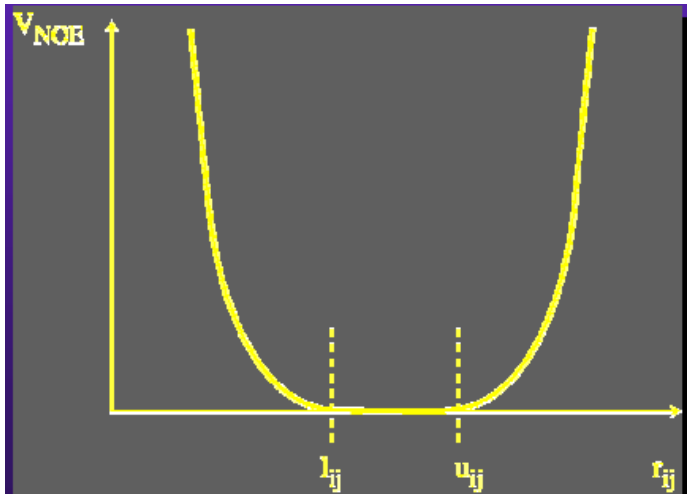


Pseudopotential energy terms: an example

- The atom pair distance r_{ij} (derived from NOE) is restrained between an upper (u_{ij}) and a lower (l_{ij}) limit as:

$$
\begin{aligned}
V_{NOE} &= k(r_{ij} - u_{ij})^2 && \text{if} \quad r_{ij} > u_{ij} \\
&= 0 && \text{if} \quad l_{ij} \le r_{ij} < u_{ij} \\
&= k(l_{ij} - r_{ij})^2 && \text{if} \quad r_{ij} < l_{ij}
\end{aligned}
$$

- The shape of the energy term looks like (if l_{ij} is not available, the sum of the atomic radii is used):



Pseudopotential energy terms

- Several other types of NMR-derived restraints can be used (provided that they are implemented in the program!)
- As an example, residual dipolar couplings (rdc's) provide information on the orientation of bond vectors (e.g. N-H, C-H) relative to the molecular magnetic susceptibility tensor, as:

$$rdc = -\frac{1}{4\pi}\frac{B_0^2}{15kT}\frac{\gamma_I \gamma_S h}{4\pi^2 r_{IS}^3}\left[\Delta\chi_{ax}\left(3\cos^2\theta - 1\right) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\phi\right]$$

- These restraints contribute to the hybrid energy function with terms such as:

$$\mathbf{E_{RDC}} = \sum_{\mathbf{rdc's}} \mathbf{w_{rdc}}\left[\mathbf{max}\left(\left|\mathbf{rdc_i^{exp}} - \mathbf{rdc_i^{calc}}\right| - \mathbf{tol_i}, \mathbf{0}\right)\right]^2$$

How the algorithms work:

Molecular Dynamics (MD)

- MD was developed with the aim of simulating the time evolution of a molecular system

- MD calculations numerically solve the equation of motion to obtain a trajectory for the molecular system

- In Cartesian coordinates, the Newton's equation of motion is:

$$m_a \frac{\mathrm{d}^2 \mathbf{r}_a}{\mathrm{d}t^2} = \frac{\partial}{\partial \mathbf{r}_a} U\left[t \middle| \sigma\left(\mathbf{r}_1, \ldots, \mathbf{r}_N\right)\right],$$

How the algorithms work:

Molecular Dynamics (MD)

- In structure calculations, the purpose of MD is quite different

- MD simply provides a means to search the conformation space of the protein for structures that match the restraints

- This corresponds to take the hybrid energy function as the potential energy of the system and to minimize it

How the algorithms work:

Why does MD minimize the energy?

- A distinctive feature of MD simulation, when compared to the straightforward minimization of a target function, is the presence of kinetic energy that allows to cross barriers of the potential surface

- The potential energy landscape of a protein is indeed very complex and studded with many local minima where a conformation can become trapped

How the algorithms work:

Simulated annealing (SA)

- MD is combined with simulated annealing protocols
- The kinetic energy (provided in terms of temperature) defines the maximal height of energy barrier that can be overcome in a MD simulation
- In protein structure calculations, the temperature is varied along the MD simulation so as to sample a wide conformational space of the protein and to optimize the ability of finding the minimum of the hybrid energy function

How the algorithms work:

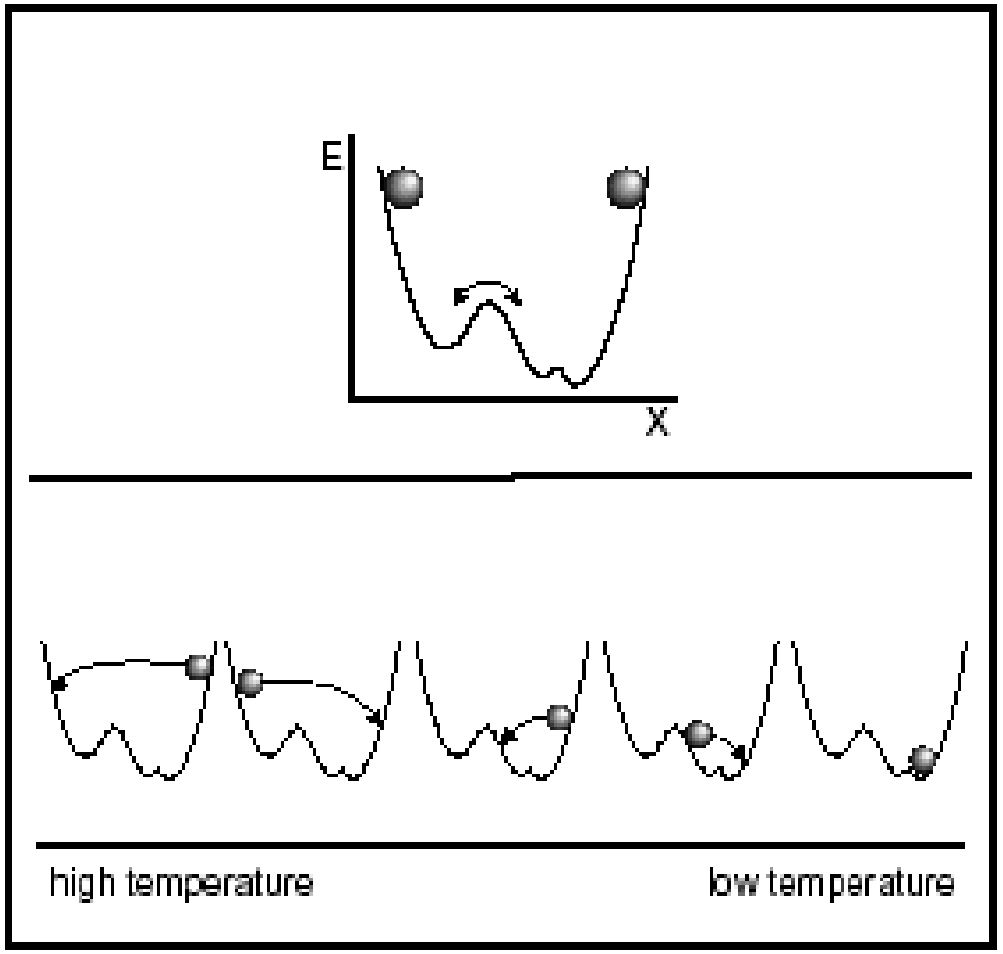
Simulated annealing (SA)

- SA mimics the annealing process through which a molecules attains its minimum energy configuration by its slow cooling after having sampled a broad conformation range at high temperatures
- It is a general optimization method used to search for the minimum of very complex functions
- Elaborated SA protocols have been devised to optimize the exploration of protein conformational space (e.g., several stages of heating and cooling, switching on/off atom-atom repulsion, etc.)

How the algorithms work:

Example of SA protocol

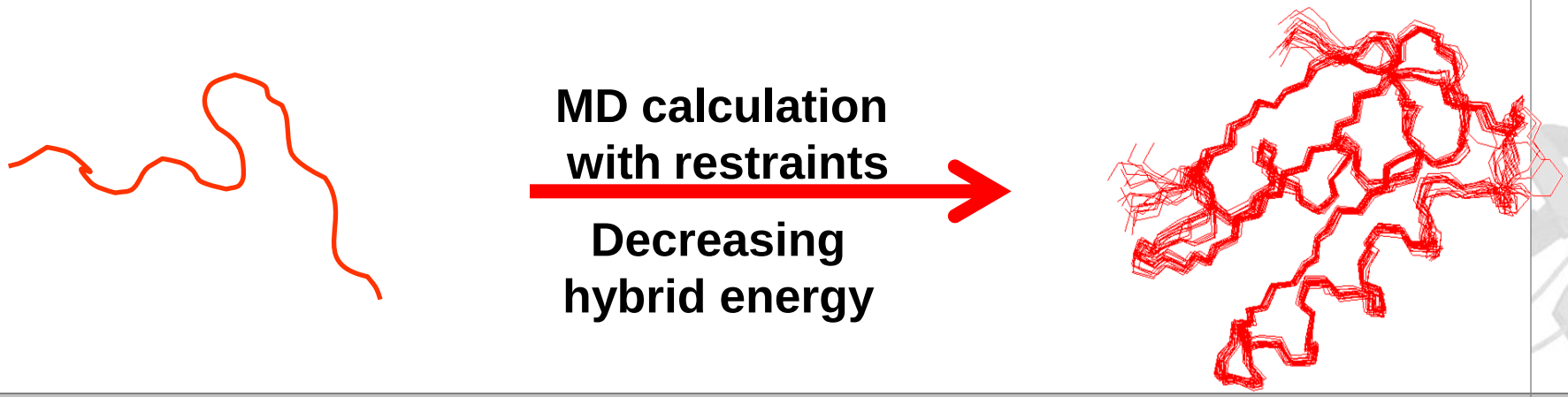
- A starting random structure is heated to very high temperature
- During many cooling steps the starting structure evolves towards (i.e., folds into) the energetically favorable final structure under the influence of the force field derived from the restraints



How the algorithms work:

Molecular Dynamics (MD)

- In a nutshell:
  - a random coil conformation is generated
  - an MD trajectory is calculated using the hybrid energy function as the potential energy
  - the end point of the trajectory is (close to) the minimum of the hybrid energy function

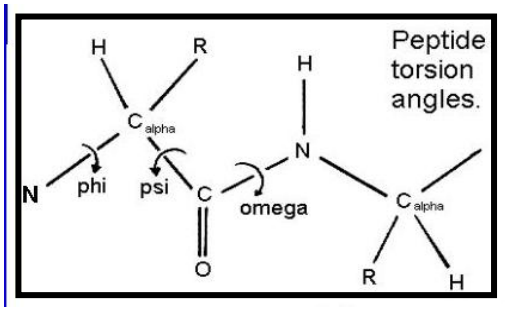


TAD versus MD in Cartesian space

- TAD (Torsion Angle Dynamics) is MD in torsion angle space
- The equations of motion (Lagrange equations) are solved in a system with N torsion angles as the only degrees of freedom

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} = 0$$

L = Ekin – Epot
q = generalized coordinates



- About 10 times less degrees of freedom than in MD in Cartesian space
- Fixed bond lengths and bond angles:
  - no high-frequency motions
  - longer integration time-steps, higher annealing temperatures

CYANA and Xplor-NIH

	Cyana	Xplor-NIH
Covalent structure	Fixed	Restrained by potential energy terms
MD in Cartesian coordinates	No	Yes
MD in Torsion Angle Space (TAD)	Yes	Yes
SA protocol	Yes	Yes
Structure refinement (in explicit water)	No	Yes

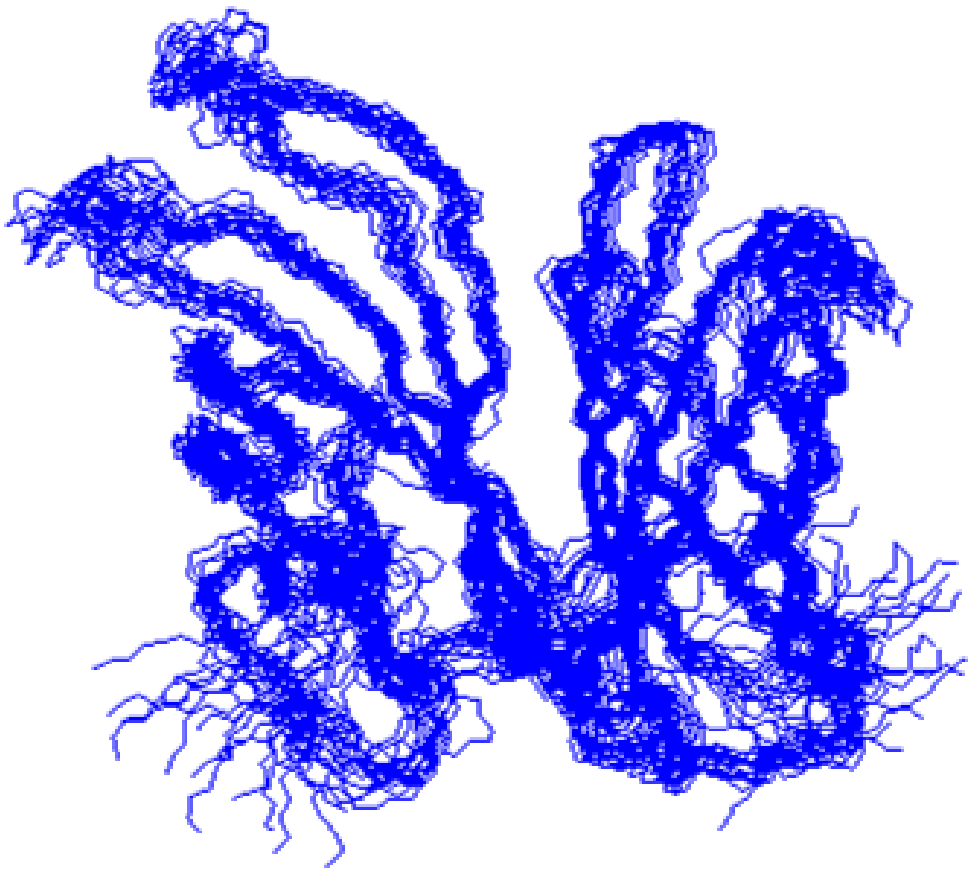
NMR structure determination & GRID

http://wenmr.eu/wenmr/nmr-services

Not just one time

- NMR structure calculations are always performed by computing, using the same restraints and algorithm, several different conformers, each starting from different initial random coil conformations
- In general, some of the conformers will be good solutions (i.e. exhibit small restraint violations) whereas others might be trapped in local minima
- The usual representation of an NMR structure is thus a bundle of conformers, each of which being an equally good fit to the data
- Conformational uncertainty may be correlated to true flexibility of the molecule

Bundles of conformers



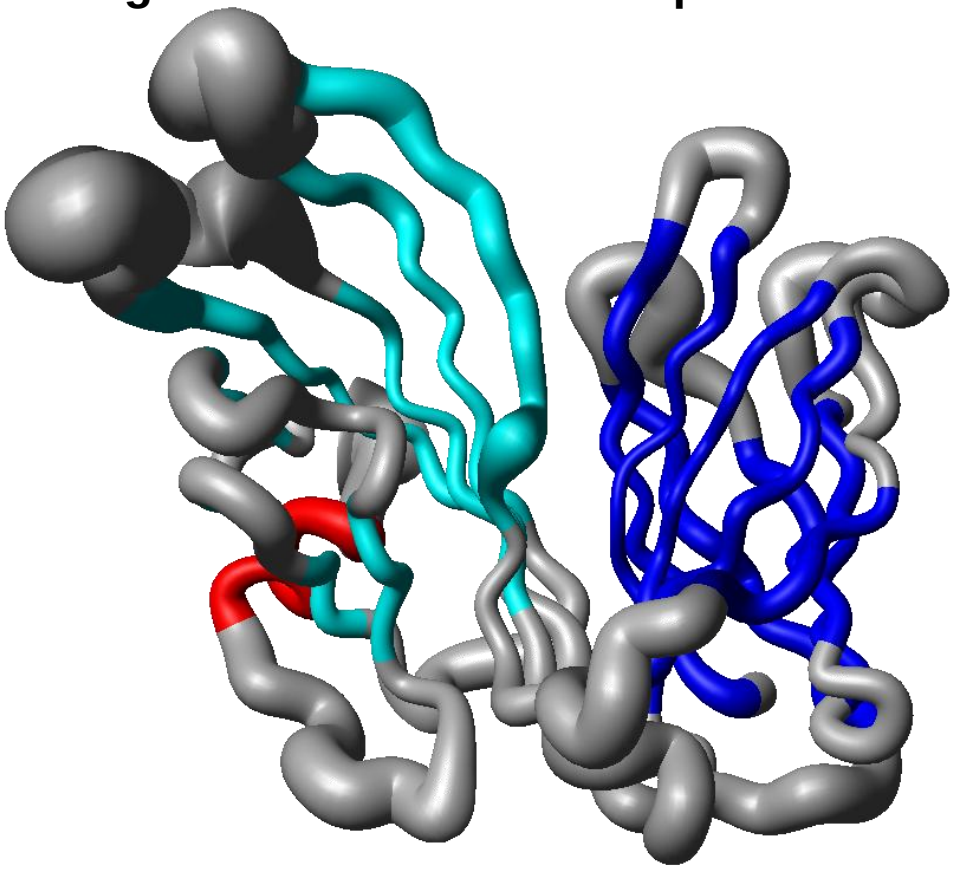
- ***2987 meaningful NOE***
- ***158*** *dihedral* ψ *and* ***158*** *dihedral* ϕ *angle constraints*
- ***RMSD*** *to the mean structure is* ***1.25 ± 0.23 Å*** *for the backbone and* ***1.75 ± 0.14 Å*** *for all heavy atoms*

The NMR solution structure of a protein is hence represented by a bundle of equivalent conformers.

Cantini, F., Veggi, D., Dragonetti, S., Savino, S., Scarselli, M., Romagnoli, G., Pizza, M., Banci, L., and Rappuoli, R. (2009) *J. Biol. Chem. 284*, 9022-9026.

Bundles of conformers

The backbone of a protein structure can be displayed as a cylindrical "sausage" of variable radius, which represents the global displacements among the conformers of the protein family:



- ***2987 meaningful NOE***
- ***158*** *dihedral* ψ *and* ***158*** *dihedral* ϕ *angle constraints*
- ***RMSD*** *to the mean structure is* ***1.25 ± 0.23 Å*** *for the backbone and* ***1.75 ± 0.14 Å*** *for all heavy atoms*

Cantini, F., Veggi, D., Dragonetti, S., Savino, S., Scarselli, M., Romagnoli, G., Pizza, M., Banci, L., and Rappuoli, R. (2009) *J. Biol. Chem. 284*, 9022-9026.

Structure refinement

(Restrained) Energy Minimization (EM) and MD on the bundle of conformers

- EM: the conformation with the local energy minimum is obtained
- MD: the conformational space is sampled through internal motions which depend on the potential generated by the atoms in the molecule
- (R)EM/(R)MD: in addition to the classical force field, the structural restraints are applied as pseudopotential
- Performed in vacuum and in explicit solvent (water)

Structure refinement

- With CYANA an external MD program is needed (e.g., AMBER). Xplor-NIH can also perform
- AMBER force field:

$$E = \sum K_r (r - r_0)^2 + \sum K_\theta (\theta - \theta_0)^2 + \sum_n \sum \frac{V_n}{2} [\cos(\eta_n \phi - \gamma_n)] +$$

$$\sum_{i<j} \varepsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i<j} \frac{q_i q_j}{r_{ij}}$$

Analysis of the results

- *How many conformers should be used to represent the solution structure?*

Around 10% of calculated structures. It should be a number that is a reasonable compromise between statistics significance and data size with respect to their manageability in graphics and analysis programs.

- *How should they be selected from the ensemble of conformers?*

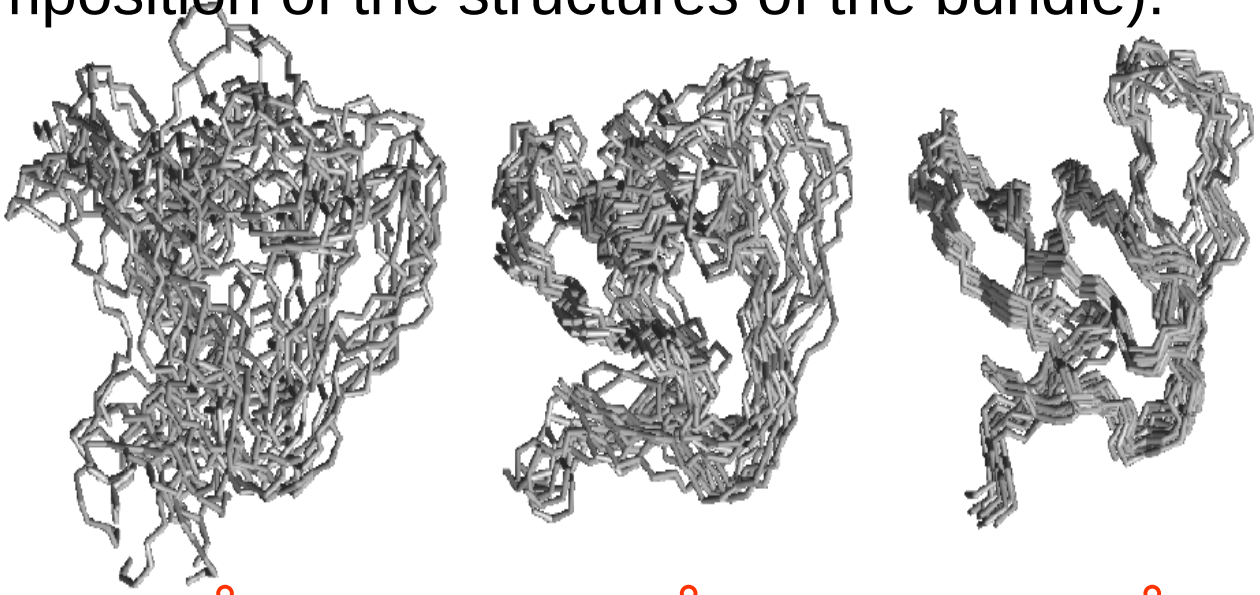
The conformers with the lowest target/penalty function, i.e. with the best agreement with the experimental structural restraints are selected

→ **Accuracy of the Structure**

RMSD

For two sets of n atoms, RMSD is defined as the normalized sum of the root mean square deviations of the position of a given atom with that of the same atom in the second set (after superimposition of the structures of the bundle):

Precision of the structure



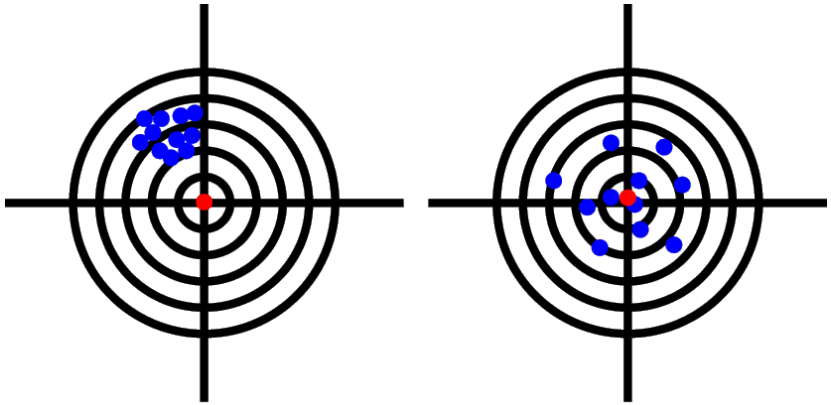
RMSD: 4.2 Å 1.9 Å 1.1 Å

$$RMSD = \sqrt{\frac{\sum (r_{ai} - r_{bi})^2}{n}}$$

- two identical structures will have an rmsd of 0Å
- larger is the rmsd and more dissimilar are the structures

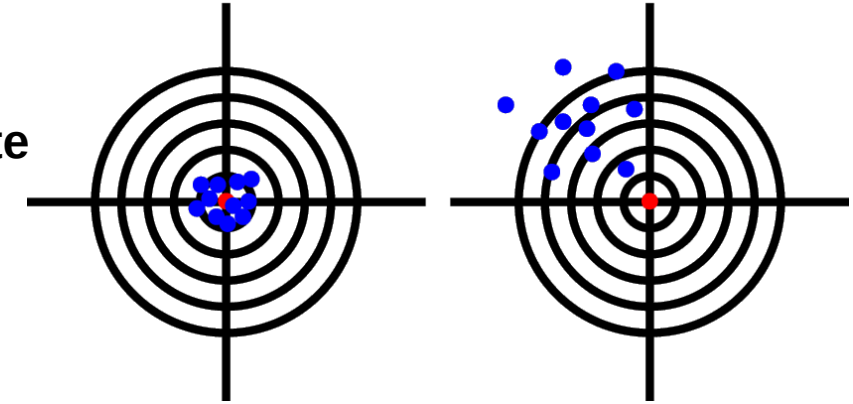
Precision versus Accuracy

Precise,
not accurate



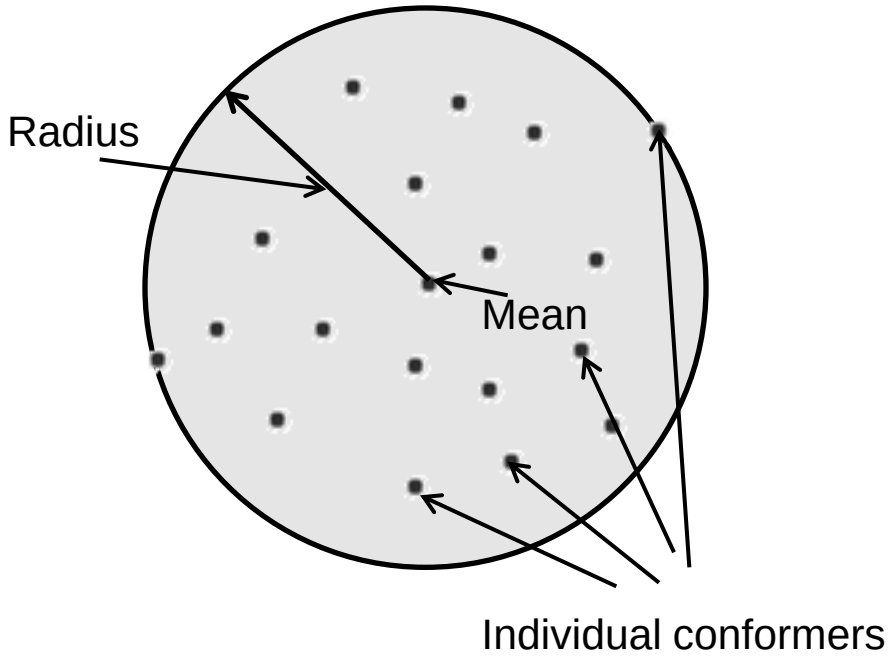
Accurate,
not precise

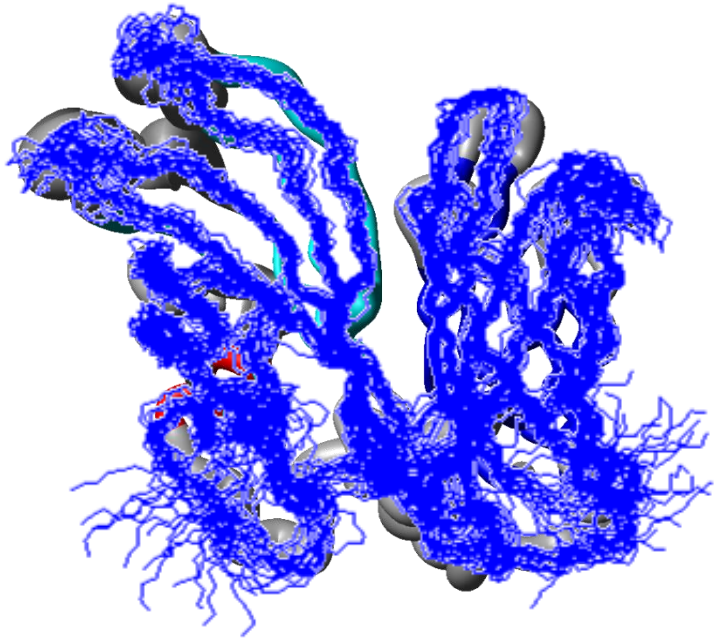
Precise
and accurate



Not accurate
and not precise

RMSD = precision





When RMSD values are used to measure the spread among the N conformers in a structure bundle, the most convenient value is the`RMSD radius‘, defined as the average of the m pairwise RMSD values between the individual conformers and their mean structure.

Validation criteria

Protein Structures are assessed with respect to:

- **Back-calculation of the experimental restraints**
- **Local geometry:**
  - **Bond lengths, bond angles, chirality, omega angles, side chain planarity**
- **Overall quality:**
  - **Ramachandran plot, rotameric states, packing quality, backbone conformation**
- **Others:**
  - **Inter-atomic bumps, buried hydrogen-bonds, electrostatics**

Validation of the NMR Structures

The most common programs used to evaluate the quality of the structures are

- **WHATIF (swift.cmbi.ru.nl)**
- **QUEEN**
- **CiNG http://nmr.cmbi.ru.nl/icing (WHATIF and PROCHECK-NMR)**
- **PSVS (http://psvs-1_4-dev.nesg.org/) (PROCHECK-NMR, MolProbity, Verify3D, Prosa II)**

Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. (1993) *J.Magn.Reson.Ser.B* 101, 333-337

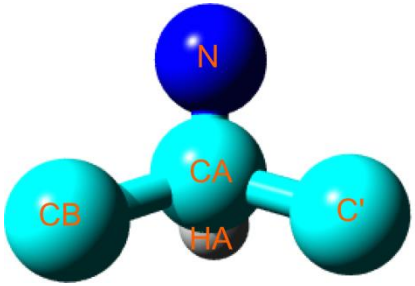
Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) *J.Biomol.NMR* 4, 845-858

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* 33, 5984

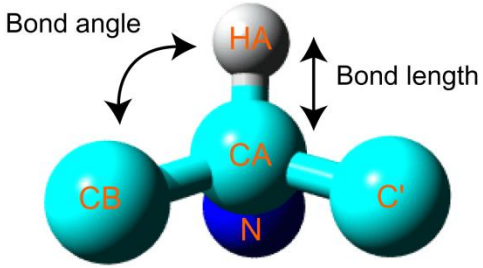
Battacharya, A., Tejero, R., and Montelione, G. T. (2007) *Proteins* 66, 778-795

Structural Parameters

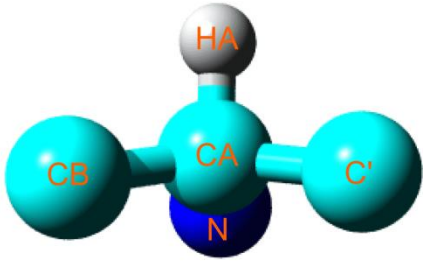
Bonded geometry



D-amino acid

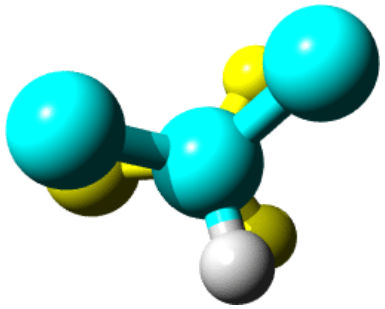


L-amino acid

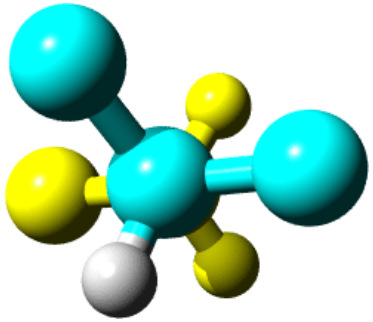


Distorted Cα-chirality

Rotameric states



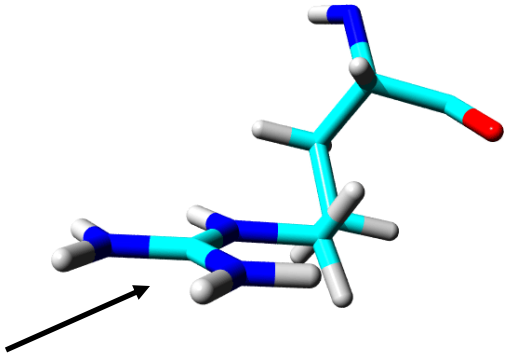
Eclipsed



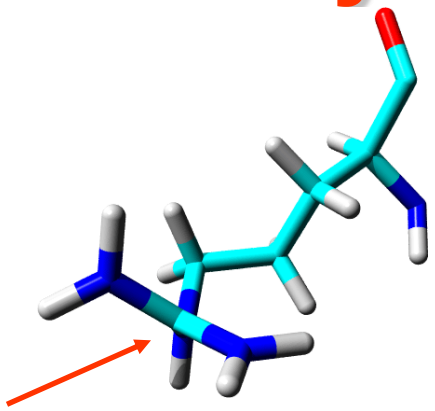
Staggered

Structural Parameters

Side chain planarity

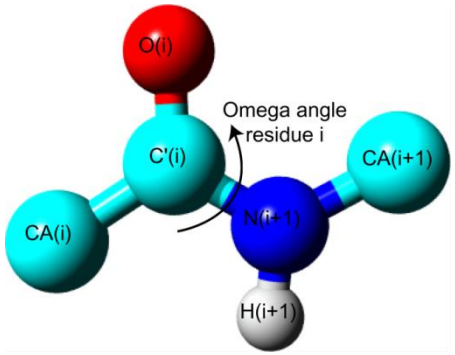


Planar ARG side-chain (Good)

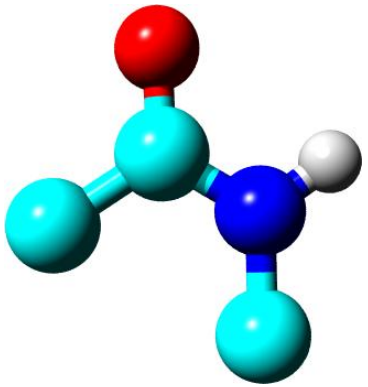


Non-planar ARG side-chain (Bad)

Omega angles



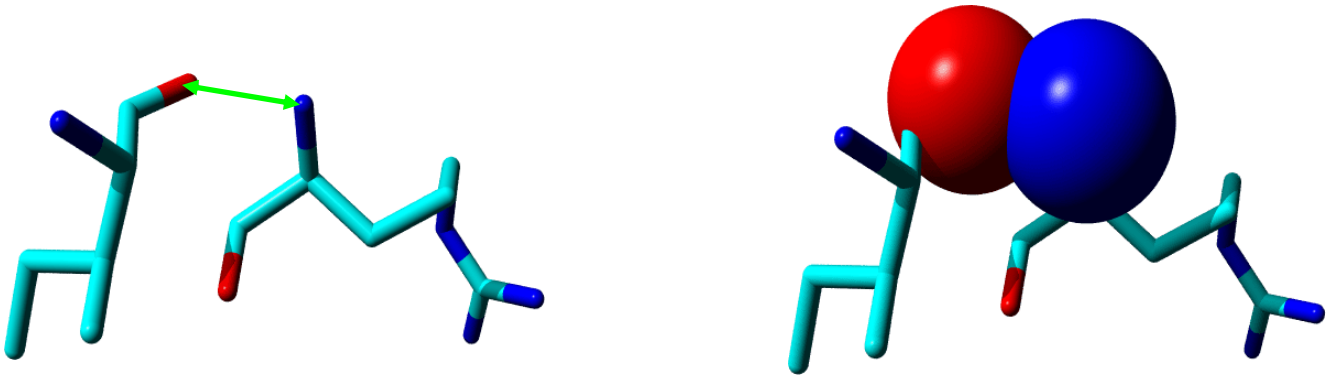
Trans-conformation (omega=180°)



Cis-conformation (omega=0°)

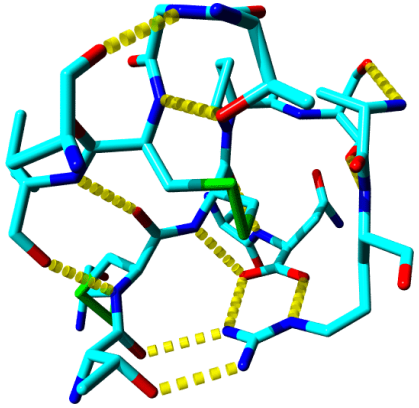
Structural Parameters

Inter-atomic bumps



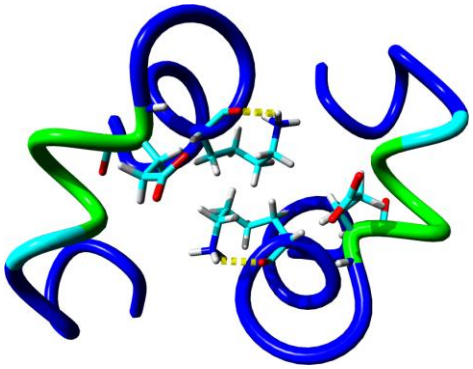
Overlap of two backbone atoms

Internal hydrogen bonding

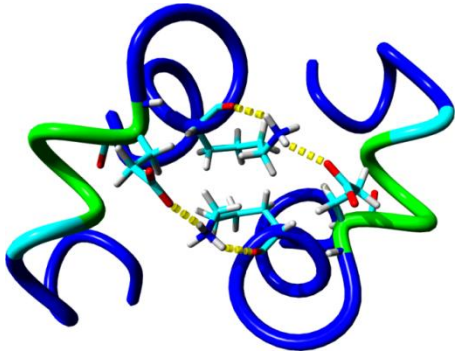


Structural Parameters

Electrostatics

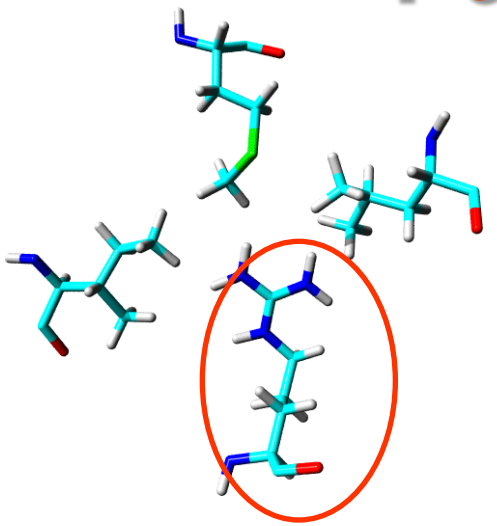


"Bad" electrostatics

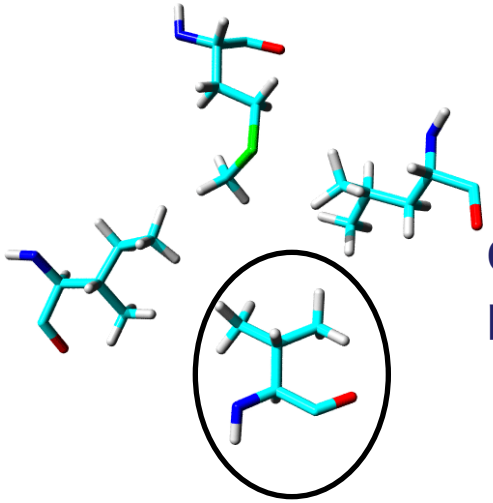


After energy minimization including electrostatics

Packing quality



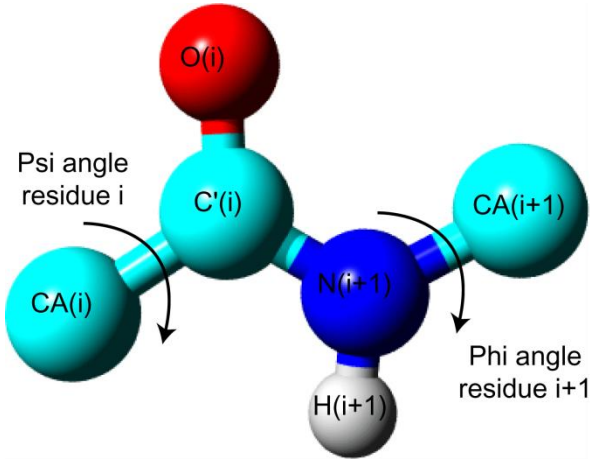
Bad packing



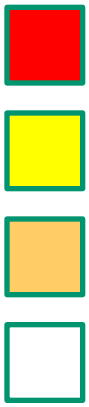
Good packing

Structural Parameters

Ramachandran Plot

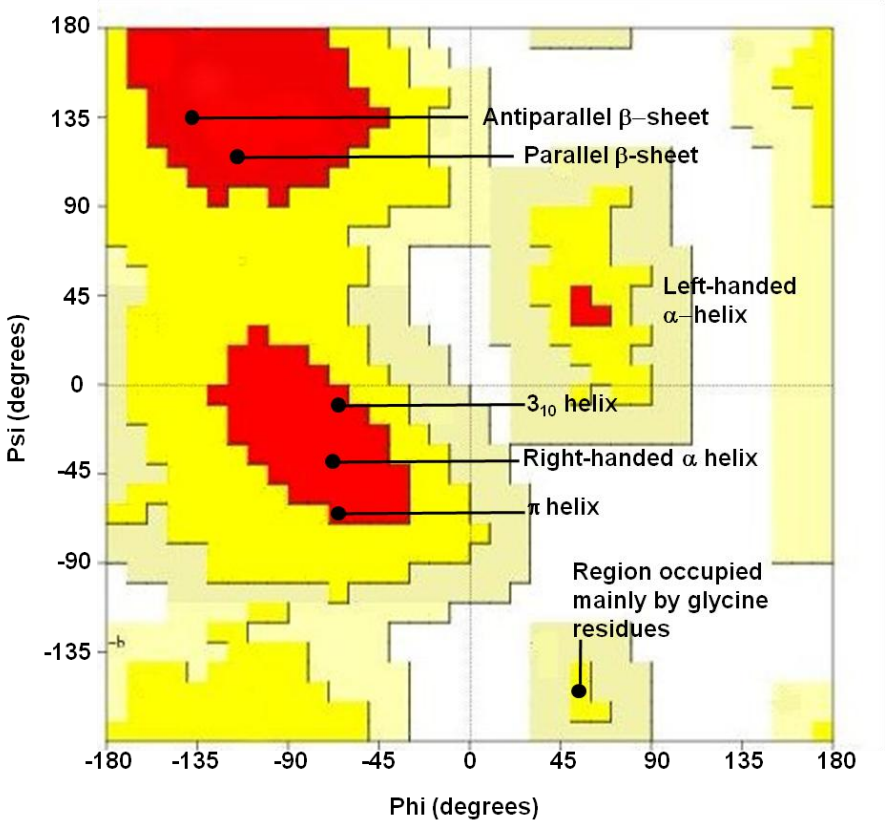


Phi and Psi angles



Generously allowed

Disallowed

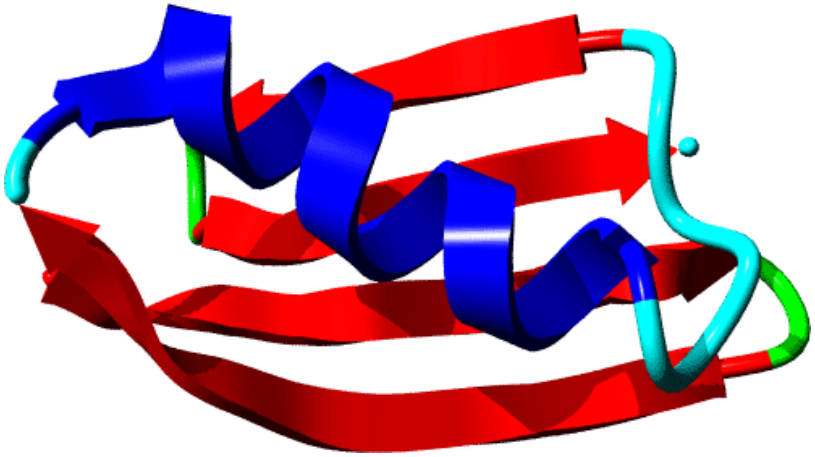


Ramachandran plot

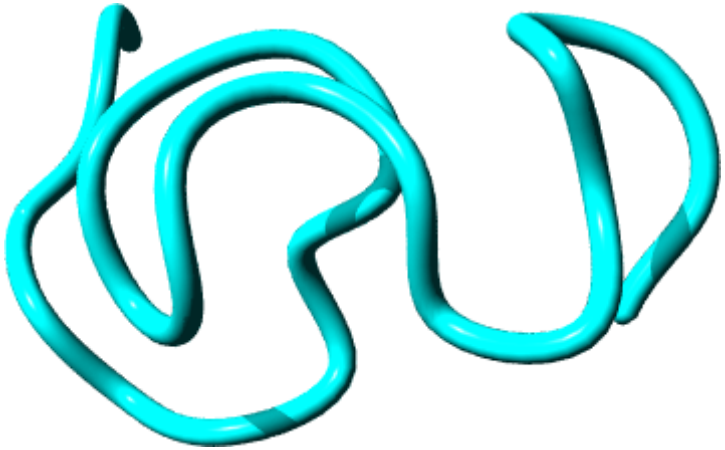
Ideally, over 90% of the residues should be in the "core" regions

Structural Parameters

Backbone Conformation (still in agreement with Ramachandran plot)



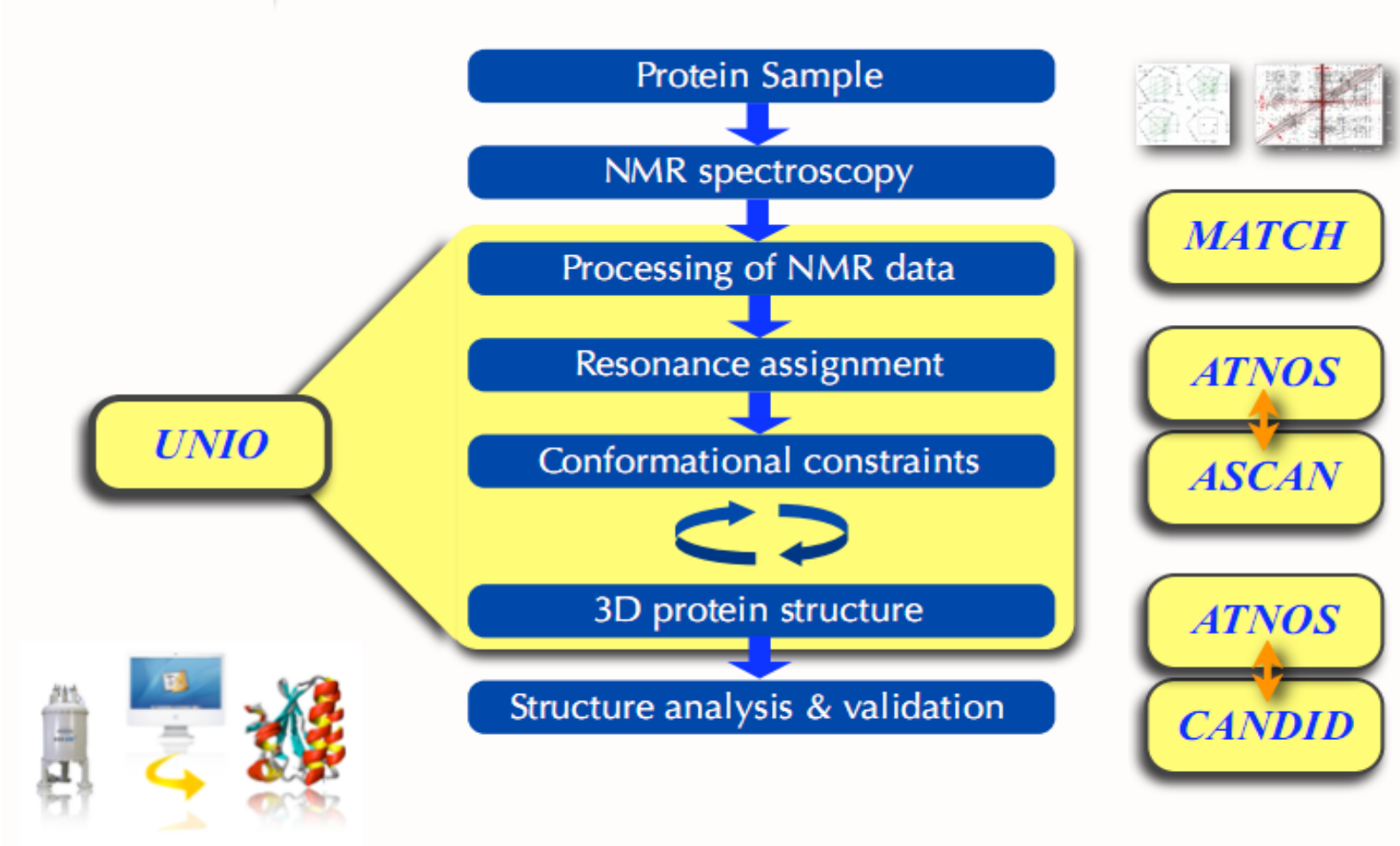
Very normal



Warning!!
Deviates from the already reported conformations

Automated Structure determination

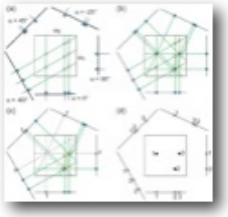
UNIO – Computational suite for fully/highly Automated NMR protein structure determination

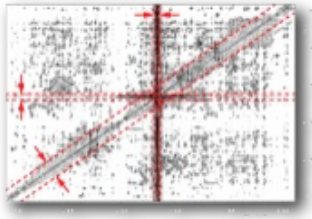


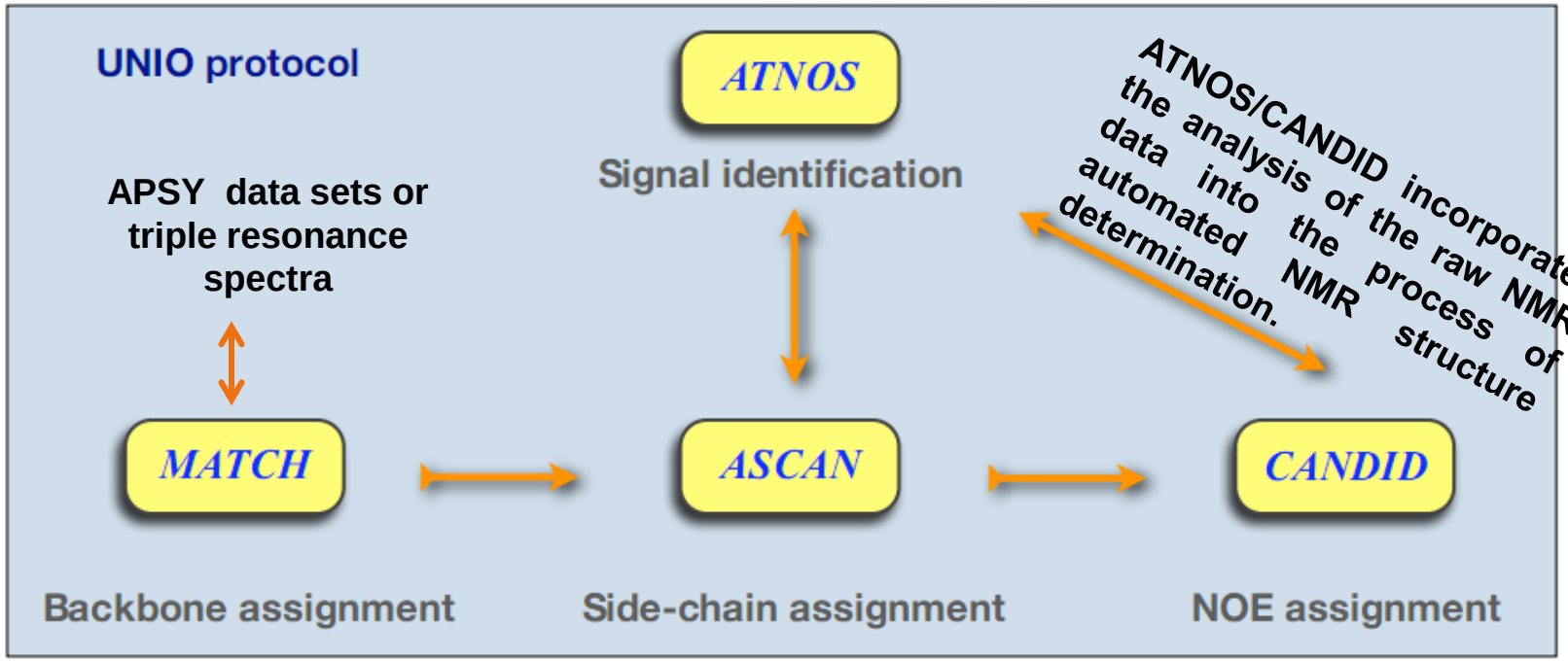
- UNIO provides accurate and automated 3D protein structure determination.
- UNIO enables protein NMR structure determination within one week including the collection of NMR experiments.

[1] Herrmann, T., Güntert, P., Wüthrich, K. (2002). J. Biomol. NMR 24
[2] Herrmann, T., Güntert, P., Wüthrich, K. (2002). J. Mol. Biol. 319 [4] Volk, J., Herrmann, T., Wüthrich, K. (2008). J. Biomol. NMR 41.
[3] Fiorito, F., Damberger, F.F., Herrmann, T., Wüthrich, K. (2008). J. Biomol. NMR 42.

UNIO for protein structure determination







UNIO protocol operates directly on the NMR spectra.

UNIO standard protocol

Amino acid sequence of the protein

MATCH backbone assignment

Input : 4D and 5D APSY spectra or triple resonance spectra

Output :backbone chemical shifts

ATNOS/ASCAN side chain assignment

Input : 3D NOESY spectra

Output :side-chain chemical shifts

ATNOS/CANDID NOE assignment

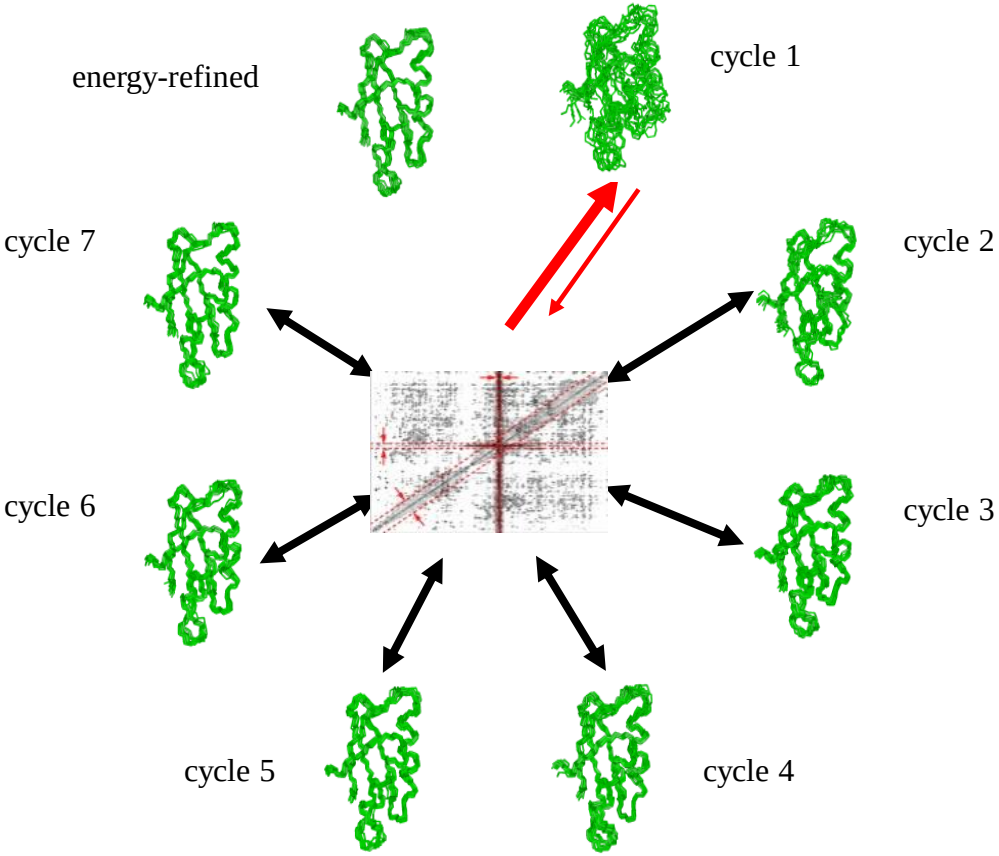
Input : 3D NOESY spectra

Output :assigned 3D NOESY peak lists and 3D protein structure with external program (XPLOR, CYANA, CNS etc)

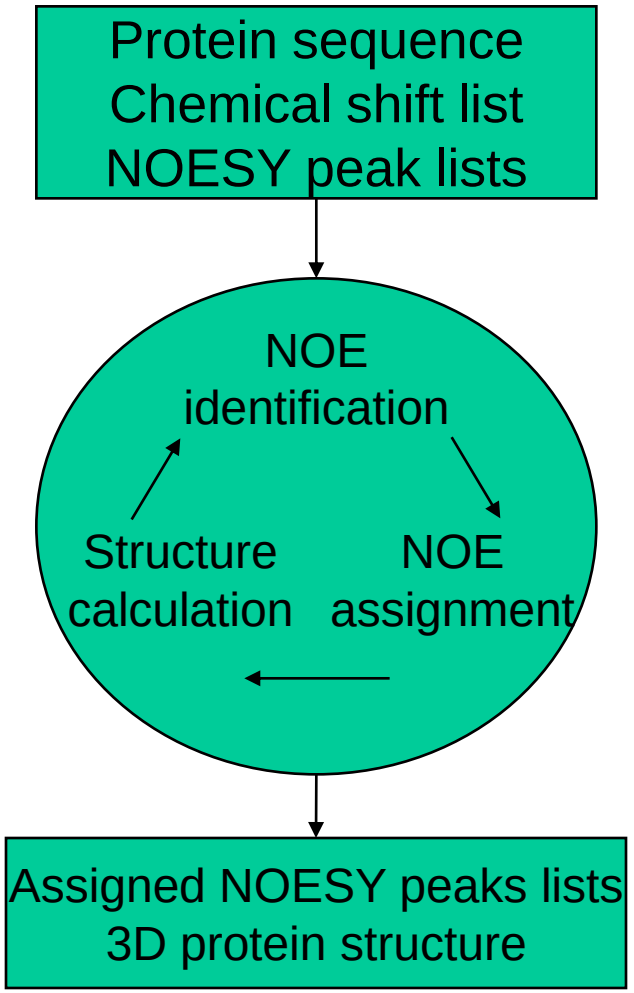
Automated NMR structure determination

•Automated NOESY spectral analysis using ATNOS-CANDID/CYANA (external program)

- Iterative process
  - all but the first cycle use the intermediate structures from the preceding cycle



- Correctness of cycle 1 is crucial for reliablity of automated approach



T. Herrmann and K. Wüthrich

Ambiguous distance constraints

- A NOESY cross peak with a single initial assignment (n=1) gives rise to a conventional upper distance constraint.
- A NOESY cross peak with initial multiple assignments (n>1) gives rise to an ambiguous distance constraint.

$$d_{eff} \equiv (\sum d_k^{-6})^{-1/6} \leq b$$

b : upper distance bound

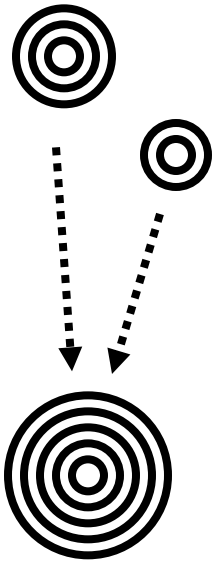
d_k: distance for assignment possibility *k*

Sums run over all assignment possibilities

Nilges et al., 1997, J. Mol. Biol. **269**, 408-422

Characteristics of ambiguous distance constraints

Isolated spin approximation: $NOE \sim d^{-6}$



Peak 1: $NOE_1 \sim d_1^{-6}$

Peak 2: $NOE_2 \sim d_2^{-6}$

$NOE_1 + NOE_2 \sim d_1^{-6} + d_2^{-6}$

$NOE_{12} \sim d_{eff}^{-6}$

$d_{eff} = (d_1^{-6} + d_2^{-6})^{-1/6}$

.

Output criteria

The correctness of resulting 3D protein structure

- **Residual CYANA target function value:**

  $TF^{cycle1} < 200Å^2$, $TF^{cycle7} < 2Å^2$

- **Root mean square deviation (RMSD) value:**

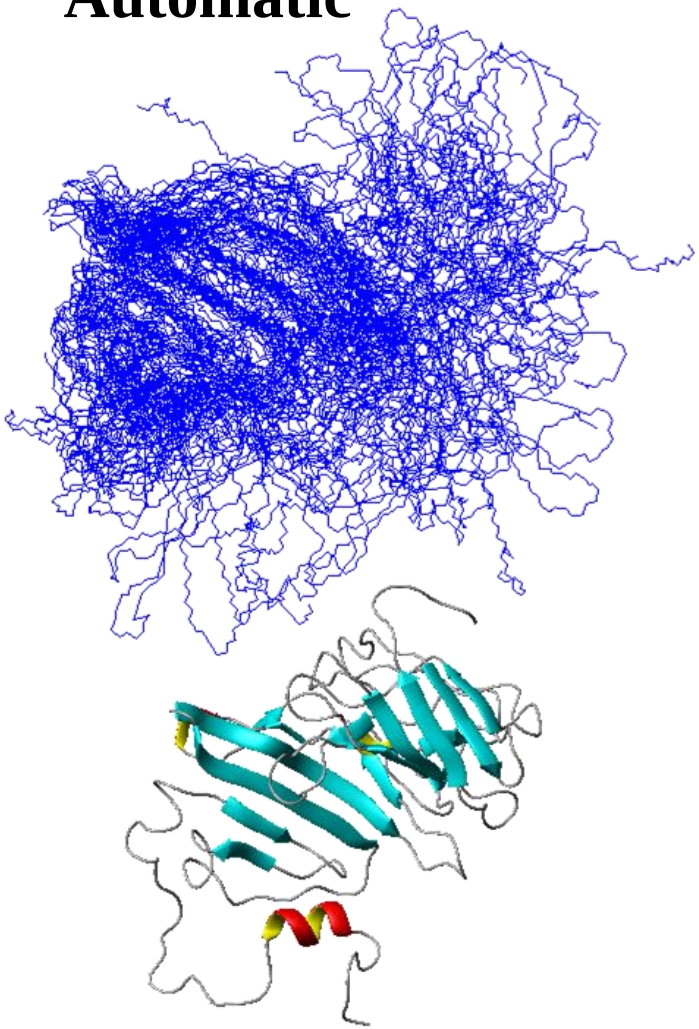
  $RMSD^{cycle1} < 3Å$

- **Evolution of $RMSD^{drift}$ value:**

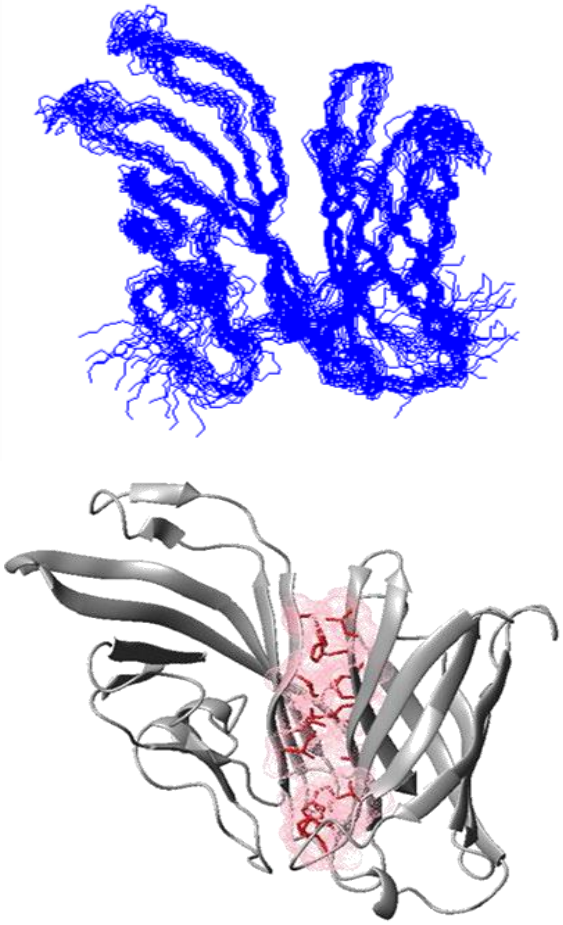
The RMSD value between the mean coordinates of the k-th and the last cycle should be in the order of the RMSD value of the k-th cycle.

Does it always work ??

Automatic



Manual



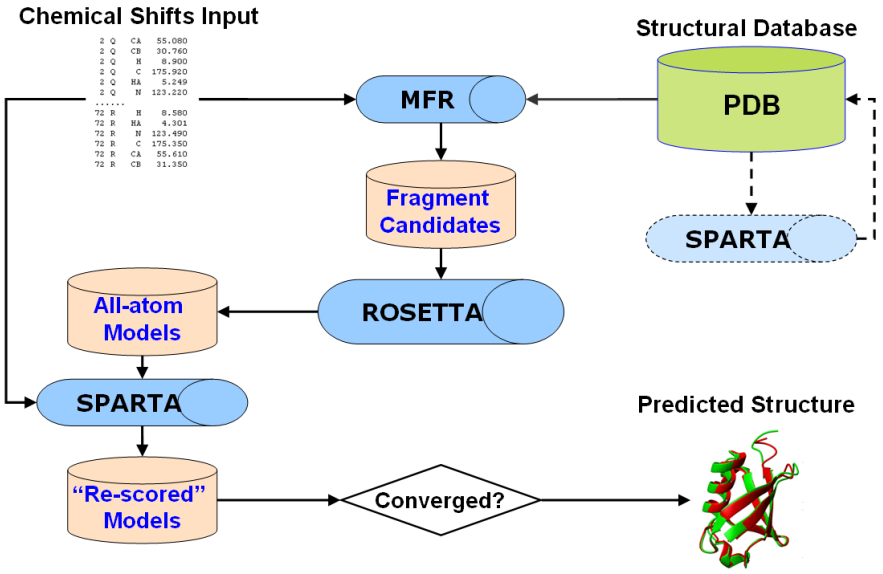
Cantini, F., Veggi, D., Dragonetti, S., Savino, S., Scarselli, M., Romagnoli, G., Pizza, M., Banci, L., and Rappuoli, R. (2009) *J. Biol. Chem. 284*, 9022-9026.

Chemical Shift-based structure calculations

CS ROSETTA generates 3D models of proteins, using only the $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, ^{15}N, $^{1}H\alpha$ and ^{1}HN NMR chemical shifts as input

CS-ROSETTA involves two separate stages:

1. Polypeptide fragments are selected from a protein structural database, based on the combined use of $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, ^{15}N, $^{1}H\alpha$, and ^{1}HN chemical shifts and the amino acid sequence pattern.
2. These fragments are used for generate a structural model, using the standard ROSETTA Monte Carlo assembly and relaxation methods



Shen, Lange, Delaglio, Bax *et al. PNAS 2008*

Thank you