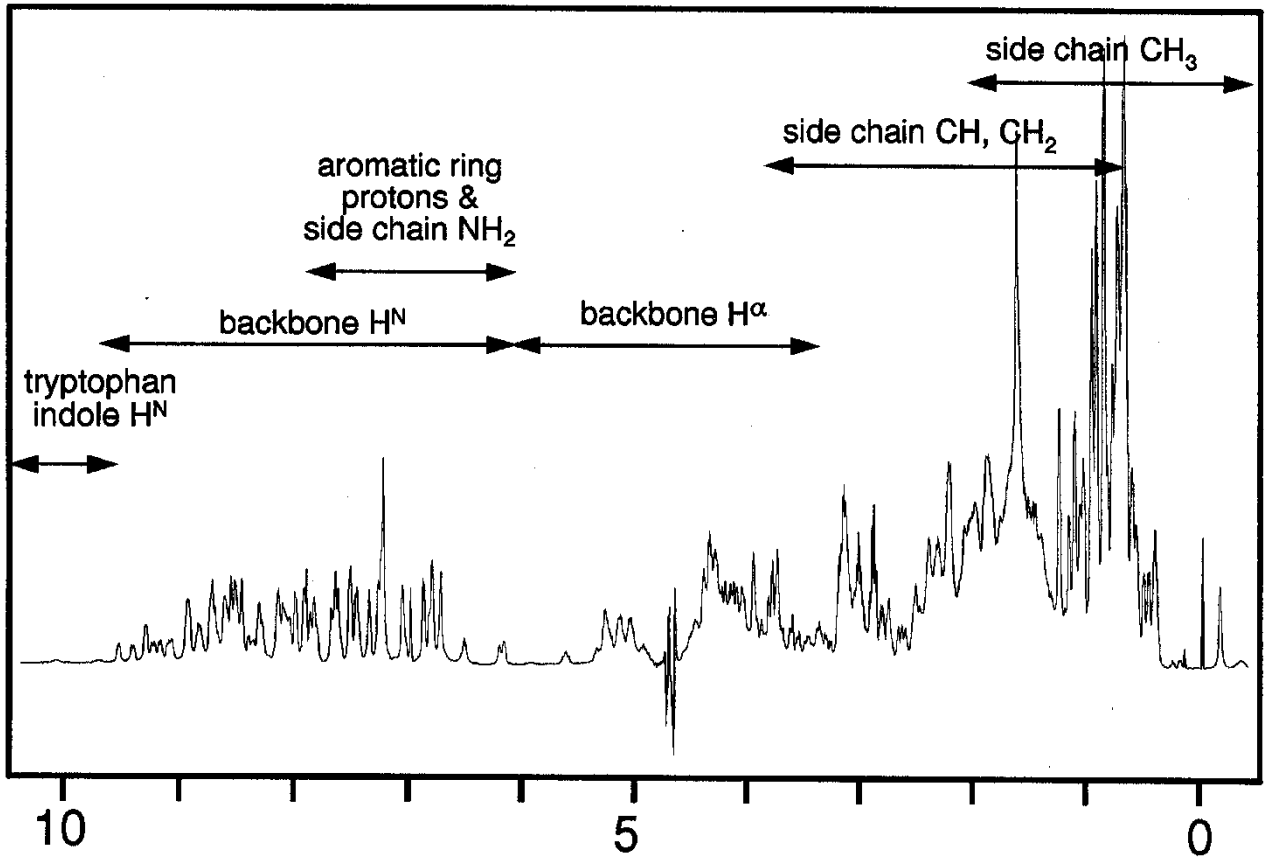
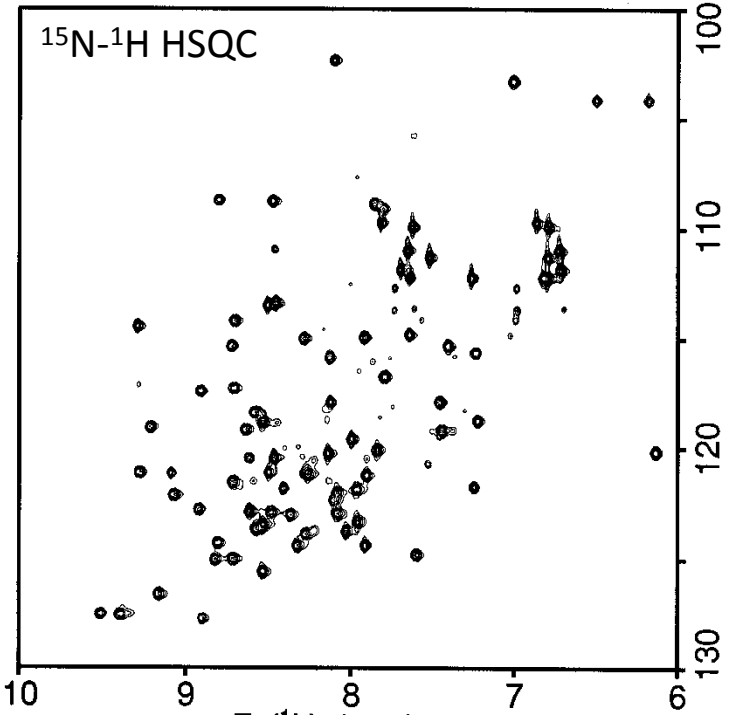
NMR Assay of Purity and Folding

Don't Need Resonance Assignments or Labeling

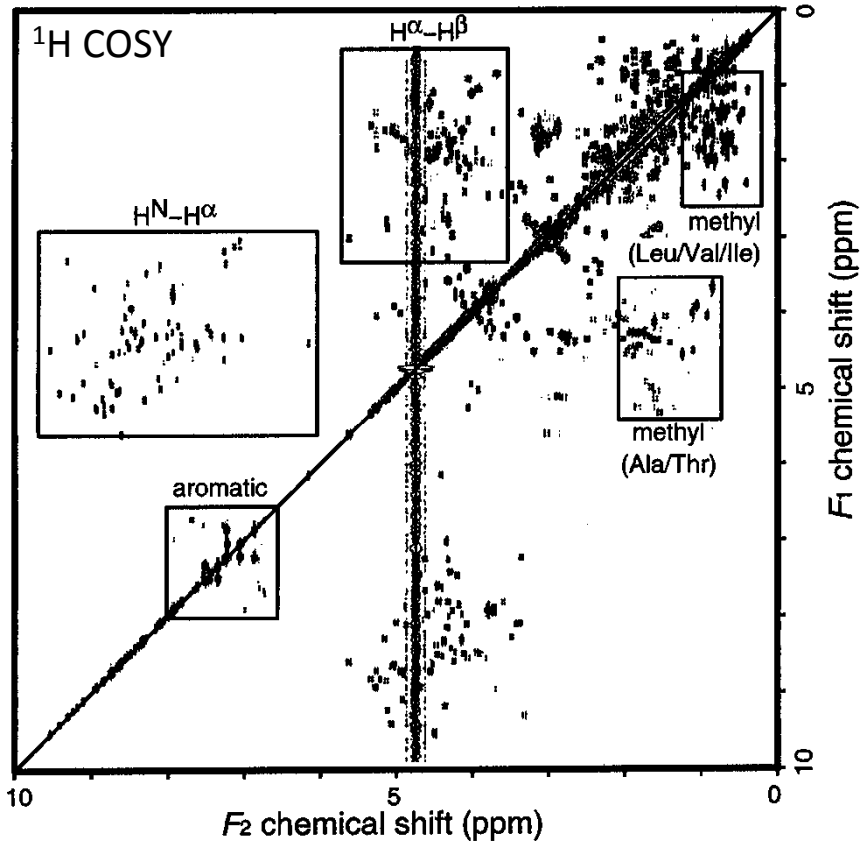


- **1D requires only 10-50 μM protein concentration**

2D Provides A More Detailed Assay



^{13}C HSQC also!

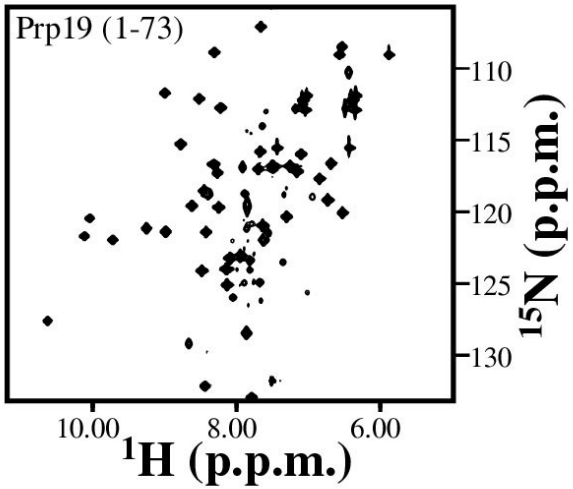


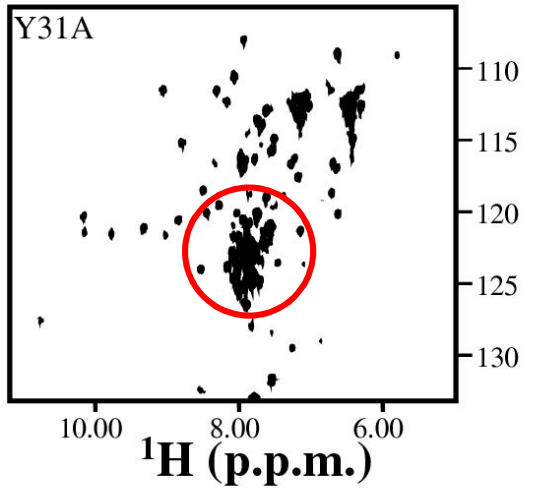
- **Analyze tertiary structure, check sequence**

Effect of Mutations

NMR assays for proper folding/stability

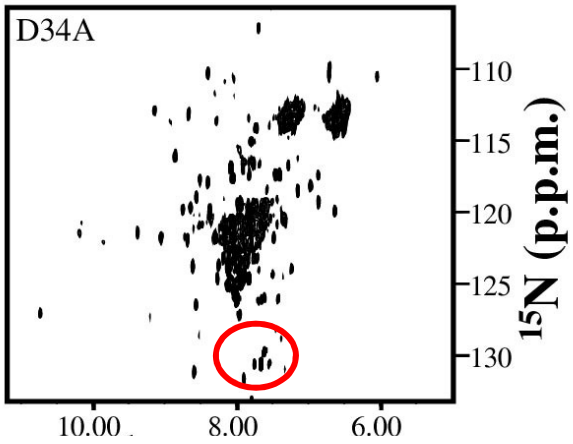
Wild-type

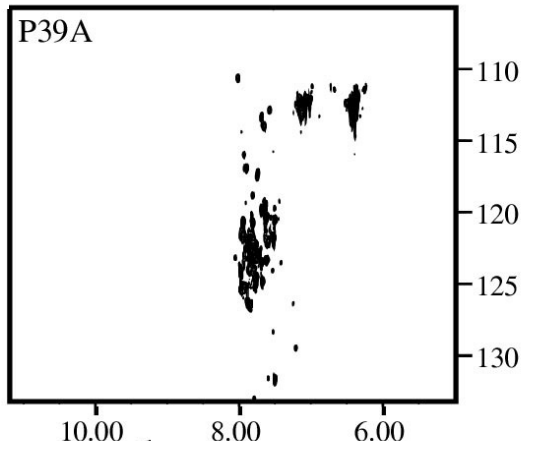




Partially destabilized

Structural heterogeneity

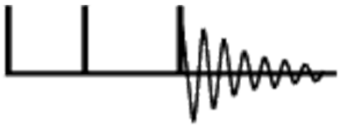


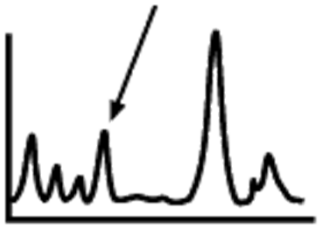


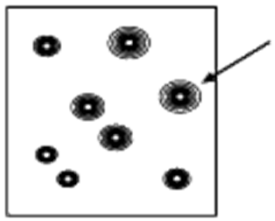
Unfolded

Ohi et al., NSB (2003)

NMR structure determination steps

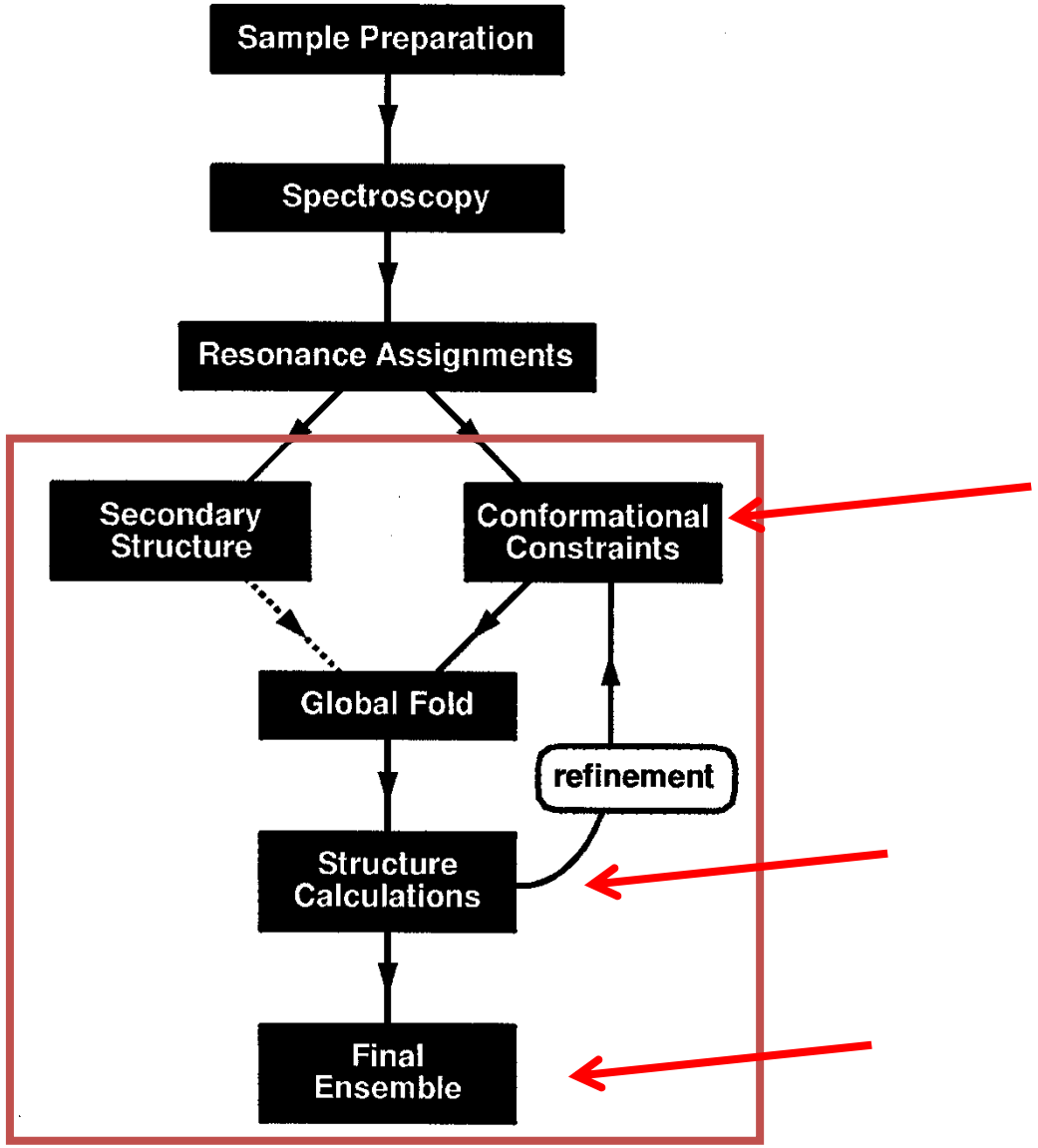






- NMR experiment
- Resonance assignment (connect the spin systems with short-range NOEs)
- Structural restraints
  - Distances (from NOEs)
  - torsion angles (from J coupling)
- Structure calculations
  - Conformation of polypeptide that satisfies all distance restraints
- Structure validation (cross-check your data)

NMR Structure Determination

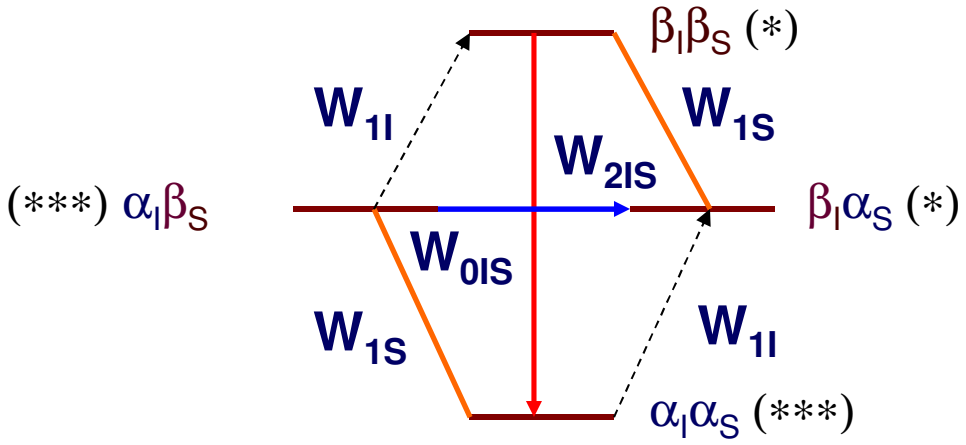


NMR Experimental Observables Providing Structural Information

- Distances from dipolar couplings (NOEs)
- Orientations of inter-nuclear vectors from residual dipolar coupling (RDCs)
- Backbone and side chain dihedral angles from scalar couplings (J)
- Backbone (ϕ,ψ) angles from chemical shifts (Chemical Shift Index- CSI, TALOS)
- Hydrogen bonds: NH exchange + NOES, J

Connections through space

• We had seen at the very beginning that if we saturate a proton in the sample, it will relax by either ***zero-*** or ***double-quantum*** processes, giving energy (enhancing) the signals of protons dipolarly coupled to it (protons close by…). This was the ***nuclear Overhauser effect*** (***NOE***):

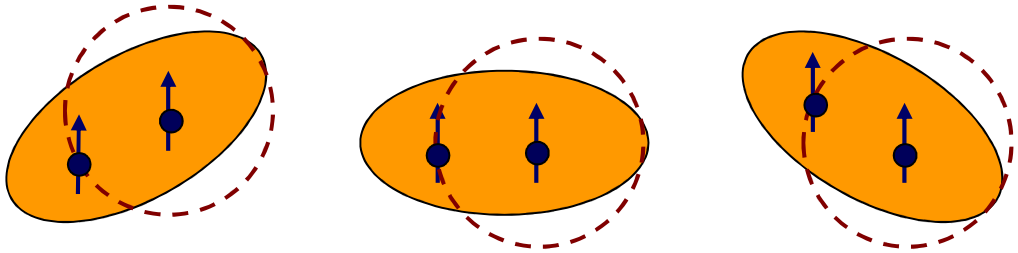


• We had seen that relaxation by either W_{2IS} or W_{0IS} will occur depending on the size of the molecule, actually its rate of tumbling or its reorientation time (i.e., its correlation time, τ_c).

Correlation functions and spectral density

• We had mentioned before that the pathways for the system had to release energy to the lattice depended on the frequencies of different processes the system can undergo.

• In solution, this means rotation of the molecule (τ_c). The spins stay aligned with the external $\mathbf{B_o}$, while the molecule turns, and this generates magnetic fields (fluctuating dipoles) at the frequency of the rotation that allow spins to release energy:



• We need a way of analyzing the way a molecule tumbles in solution. We define the ***correlation function*** of a system as the average of the molecular orientation at a certain time (**t**), and a little while (**t** + τ) after that:

$$g(\tau) = \overline{R(t) * R(t + \tau)}$$

• It basically (cor)relates the orientation of the molecule at two different times. **g(0)** = **1**, and **g(τ)** decays exponentially as a function of τ / τ_c, being τ_c the correlation time of the molecule.

• Since the probability of a transition depends on the different frequencies that the system has (the ***spectral density***), the **W** terms are proportional the **J(ω)**.

• Also, since we need two magnetic dipoles to have dipolar coupling, the NOE depends on the strength of the two dipoles involved. The strength of a dipole is proportional to r_{IS}^{-3}, and the **Ws** will depend on r_{IS}^{-6}:

$$W_{0IS} \propto \gamma_I^2 \gamma_S^2 r_{IS}^{-6} \tau_c / [\, 1 + (\omega_I - \omega_S)^2 \tau_c^2]$$

$$W_{2IS} \propto \gamma_I^2 \gamma_S^2 r_{IS}^{-6} \tau_c / [\, 1 + (\omega_I + \omega_S)^2 \tau_c^2]$$

$$W_{1S} \propto \gamma_I^2 \gamma_S^2 r_{IS}^{-6} \tau_c / [\, 1 + \omega_S^2 \tau_c^2]$$

$$W_{1I} \propto \gamma_I^2 \gamma_S^2 r_{IS}^{-6} \tau_c / [\, 1 + \omega_I^2 \tau_c^2]$$

$$\eta \propto \frac{\tau_c}{r^6}$$

• The relationship is to the inverse sixth power of r_{IS}, which means that the NOE decays very fast as we pull the two nuclei away from each other.

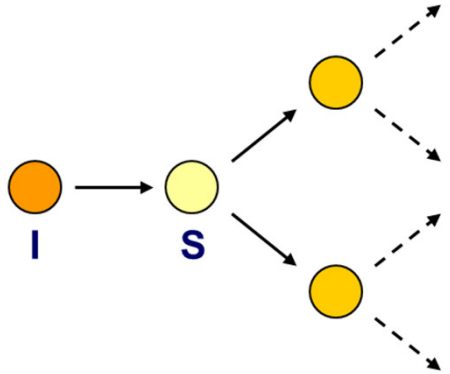
• For protons, this means that we can see things which are at most 5 to 6 Å apart in the molecule (under ideal conditions…).

Transient NOE

• One of the problems of steady-state NOE is that we are continuously giving power to the system (saturation). This works well for small molecules, because **W2** processes (double-quantum) are dominant and we have few protons.

• However, as the size and τ_c increase, other processes are more important (normal single-quantum spin-spin relaxation and zero-quantum transitions).

• Additionally, there are more protons in the surroundings of a larger molecule, and we have to start considering a process called ***spin diffusion***:

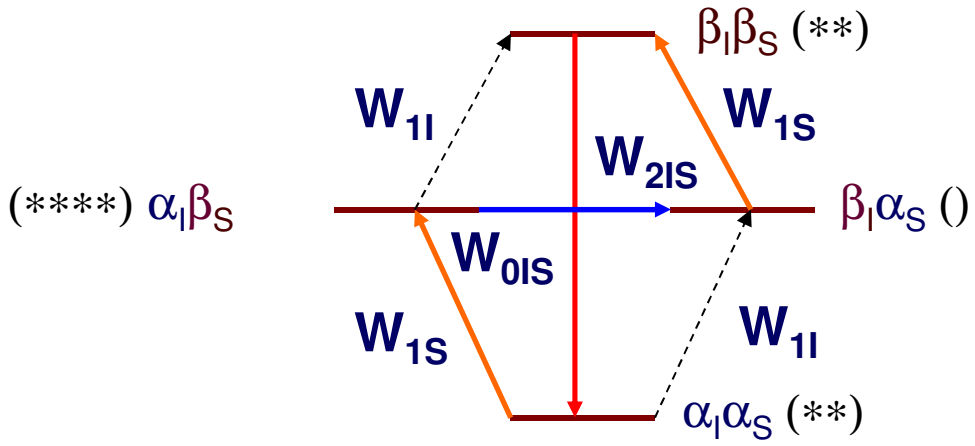


• Basically, the energy transferred from **I** to **S** then diffuses to other nuclei in the molecule. We can see an enhancement of a certain proton even if it is really far away from the center we are irradiating, which would give us ambiguous results.

• Therefore, we need to control the amount of time we saturate the spins in the system. The longer we irradiate, the more spin diffusion we get…

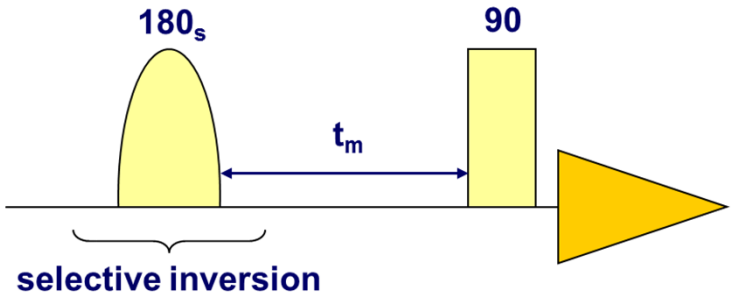
Transient NOE (continued)

• There are also some technical difficulties if we try to do this by selective saturation. Since what we need is to see how a system returns to equilibrium through cross-relaxation, we can selectively invert one transition and then see how the NOE grows with time:



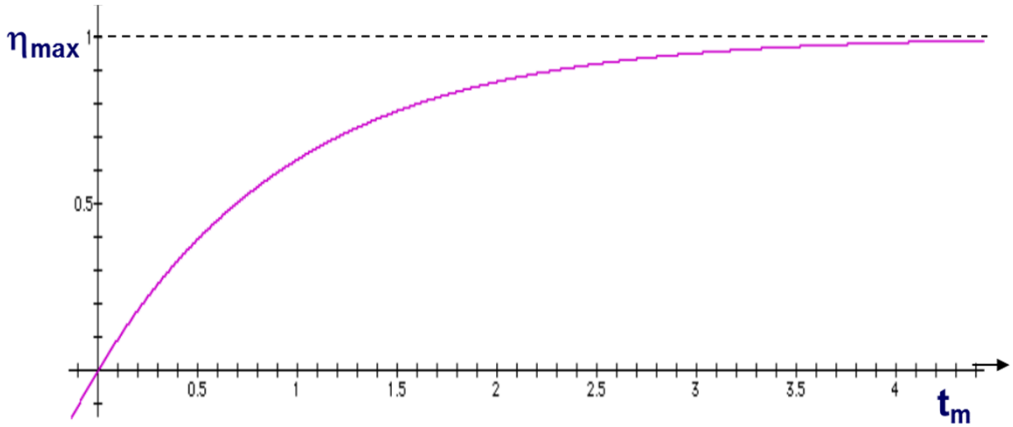
• A pulse sequence to do this could be the following:

• The last **π / 2** pulse is usually called a ***read*** pulse, and its only job is to convert whatever magnetization is on **z** after t_m into **<xy>** magnetization (detectable). All the equations are the same, but the NOE will also depend on the mixing time, t_m.

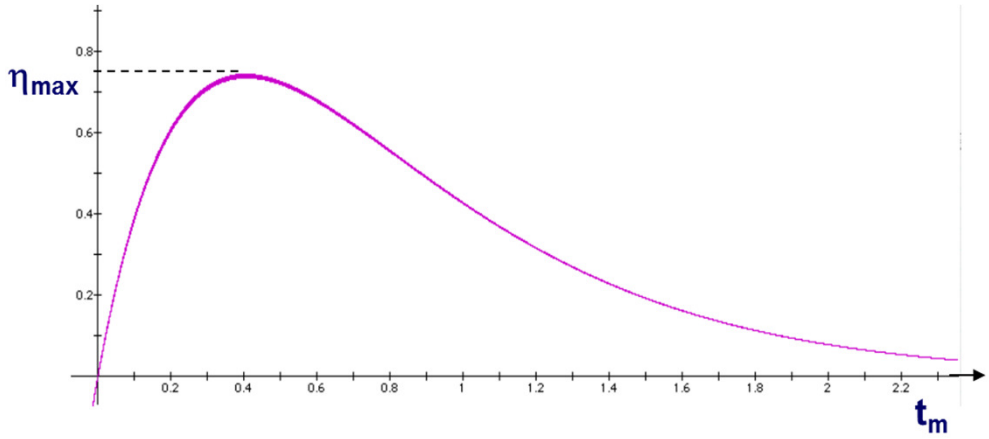


Transient NOE (…)

• If we do it for different t_m values, we get ***NOE build-up*** curves, which in the case of two isolated protons and ideal conditions are exponentials that grow until they reach η_{max}.

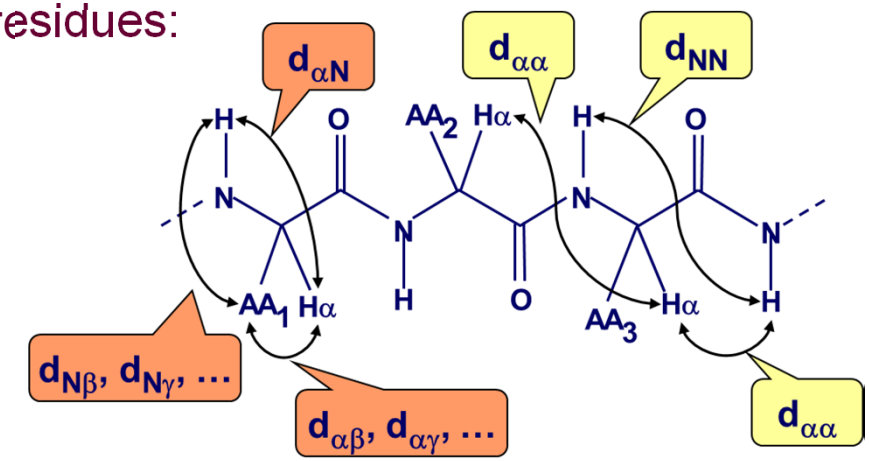


• If we also take into account T_1 and T_2 relaxation, the NOE grows and then falls to zero (all the magnetization returns to the **z** axis…):

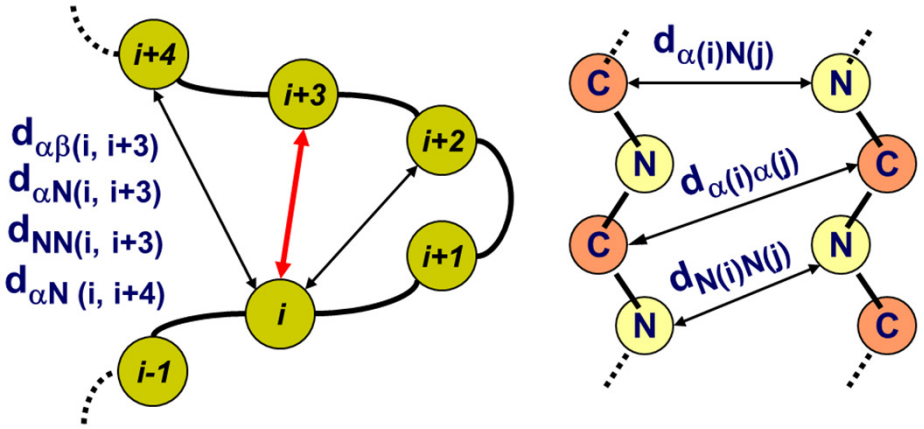


Characteristic NOE patterns.

• The easiest to identify are ***interesidue*** and ***sequential*** NOE, cross-peaks, which are NOEs among protons of the same residue and from a residue to protons of the ***(i + 1)*** and ***(i - 1)*** residues:



• Apart from those, regular secondary structure will have regular NOE patterns. For ***α-helices*** and ***β-sheets*** we have:



What the NOEs does and doesn't mean

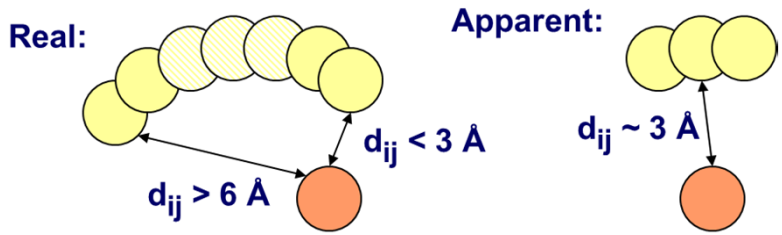
• So now we have everything: All spin systems identified, all their sequential, medium, and long range NOEs assigned, and their intensities measured.

• At this point (and very likely before this point also), we will have several conflicting cases in which we see a particular NOE but we don't see others we think should be there.

• The reason is because the NOE not only depends on the distance between two protons, but also on the dynamics between them (that means, how much one moves relative to the other). This is particularly important in peptides, because we have lots of side chain and backbone mobility.

• The most important 'law' from all this is that not seeing an NOE cross-peak ***does not mean*** that the protons are at a distance larger than 5 Å.

• Also, an NOE can arise from an average of populations of the peptide. We see something as medium (1.8 to 3.3 Å), when it is actually a mix of strong (1.8 - 2.7 Å) and no NOE:



Couplings and dihedral angles

• The previous slides showed us how to use NMR to obtain some of the structural parameters required to determine 3D structures of macromolecules in solution.

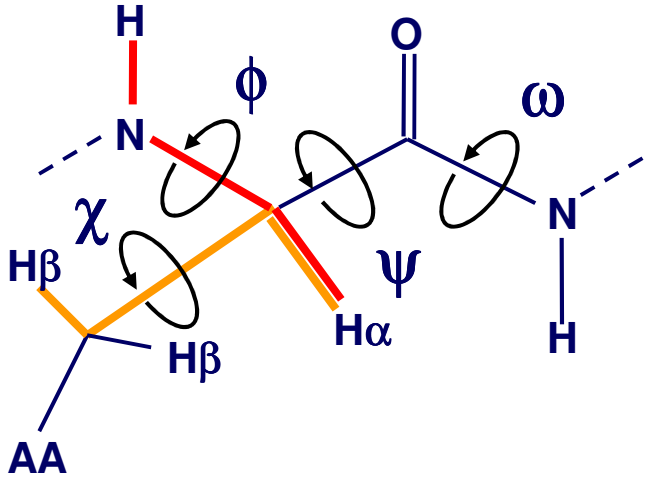
• NOEs let us find out approximate distances between protons. They can tell us a lot when we find one that report on things that are far away in the sequence being close in space.

• However, we cannot say anything about torsions around rotatable bonds from NOEs alone. What we can use in these cases are the ^3J coupling constants present in the peptide spin system (also true for sugars, DNA, RNA). We can use homonuclear or heternonuclear Js, but we'll concentrate on the former (^3J).

• These are $^3J_{N\alpha}$, which reports on the conformation of the peptide backbone, and $^3J_{\alpha\beta}$ which is related to the side chain conformation:

$^3J_{N\alpha} \longrightarrow \phi$

$^3J_{\alpha\beta} \longrightarrow \chi$



Couplings and dihedral angles (continued)

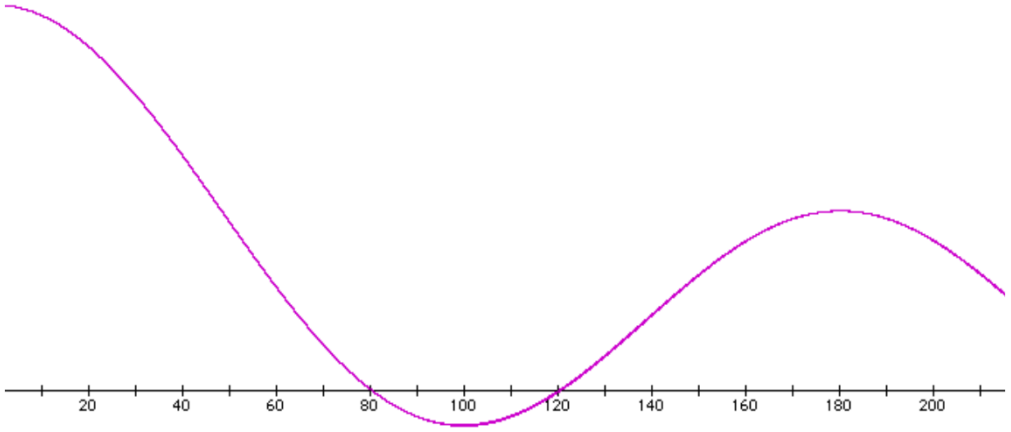
• The ^3J coupling constants are related to the dihedral angles by the ***Karplus equation***, which is an empirical relationship obtained from rigid molecules for which the crystal structure is known (derived originally for small organic molecules).

• The equation is a sum of cosines, and depending on the type of topology (**H-N-C-H** or **H-C-C-H**) we have different parameters:

$$^3J_{N\alpha} = 9.4\cos^2(\phi - 60) - 1.1\cos(\phi - 60) + 0.4$$

$$^3J_{\alpha\beta} = 9.5\cos^2(\psi - 60) - 1.6\cos(\psi - 60) + 1.8$$

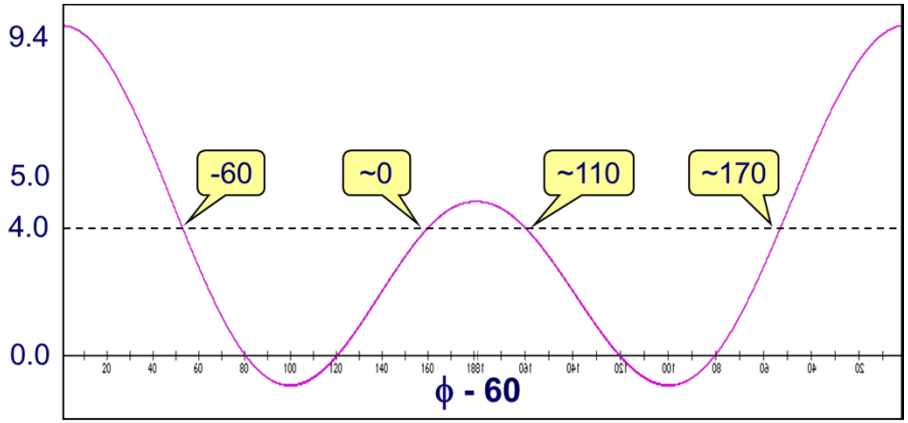
• Graphically:



Couplings and dihedral angles (…)

• How do we measure the ^3J values? When there are few amino acids, directly from the 1D. We can also measure them from HOMO2DJ spectra (remember what it did?), and from COSY-type spectra with high resolution (MQF-COSY and E-COSY).

• The biggest problem of the Karplus equation is that it is ambiguous - If we are dealing with a $^3J_{N\alpha}$ coupling smaller than 4 Hz, and we look it up in the graph, we can have at least 4 possible ϕ angles:



• In these cases there are two things we can do. One is just to try figuring out the structure from NOE correlations alone andthen use the couplings to confirm what we get from NOEs.This is fine, but we are sort of dumping information to the can.

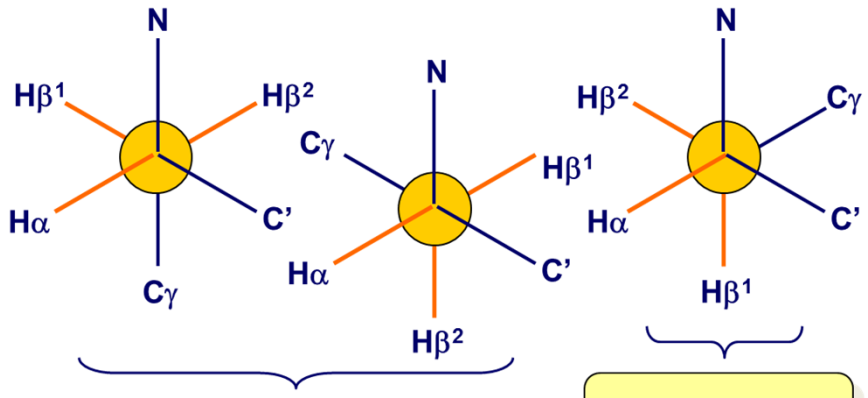
Couplings and dihedral angles (…)

• Another thing commonly done in proteins is to use only those angles that are more common from X-ray structures. In the case of ϕ, these are the negative values (in this case the -60 and 170). Also, we use ranges of angles:

$^3J_{N\alpha} < 5$ Hz $\longrightarrow$ $-80 < \phi < -40$

$^3J_{N\alpha} > 8$ Hz $\longrightarrow$ $-160 < \phi < -80$

• For side chains we have the same situation, but in this case we have to select among three possible conformations (like in ethane…). Since we usually have two $^3J_{\alpha\beta}$ values (there are 2 β protons), we can select the appropriate conformer:



Use of chemical shifts

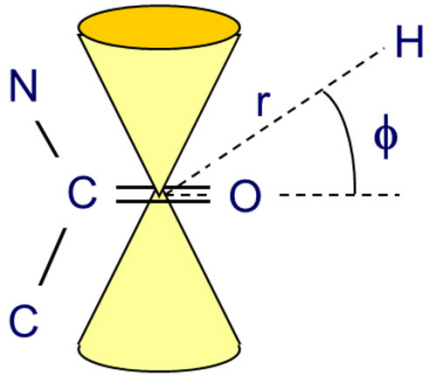
• What about chemical shifts? After all, we have chemical shifts because we have different conformations for different amino acids in the peptide.

• However, nobody really cared about them until recently. The main problem is that, as opposed to couplings, rules or parameters for chemical shifts can only be used in regular structures.

• Since nobody looked at proteins by NMR until the mid ‘80s, there were no good parametrizations or good reference data.

• The idea is that we can assign a random coil chemical shift value to all the protons in an amino acid. Any deviation from it, or ***secondary shift***, arises from different effects:

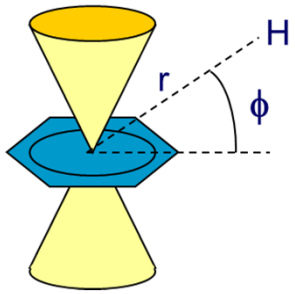
a) Peptide group anisotropy. The local magnetic field of the peptide group (CO-NH) will make protons lying above or to the side be shifted up- or down-field.



$$\sigma_{pga} = C_{CO} * r^{-3} * [\, 1 - 3 * \cos(\phi)^2 \,]$$

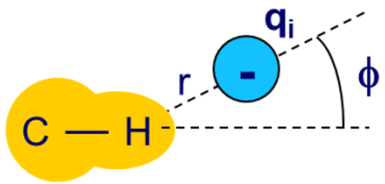
Use of chemical shifts (continued)

b) Ring current effects. The local magnetic field created by the e^- current of aromatic rings will cause protons lying above or to its the side be shifted up- or down-field. This example is archetypal and you'll find it in every organic chemistry book.

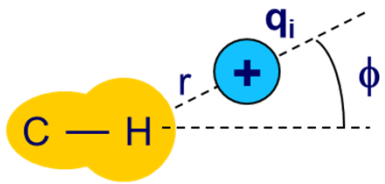


$$\sigma_{rc} = C_{ring} * r^{-3} * [\, 1 - 3 * \cos(\phi)^2 \,]$$

c) Polarization of C-H bonds by polar/charged groups. The electron cloud of the σ bond goes back or forth the C-H bond depending of the presence of groups of different polarity aligned with them:



Upfield shift



Downfield shift

$$\sigma_{elec} = C * r^{-2} * q_i * \cos(\phi)$$

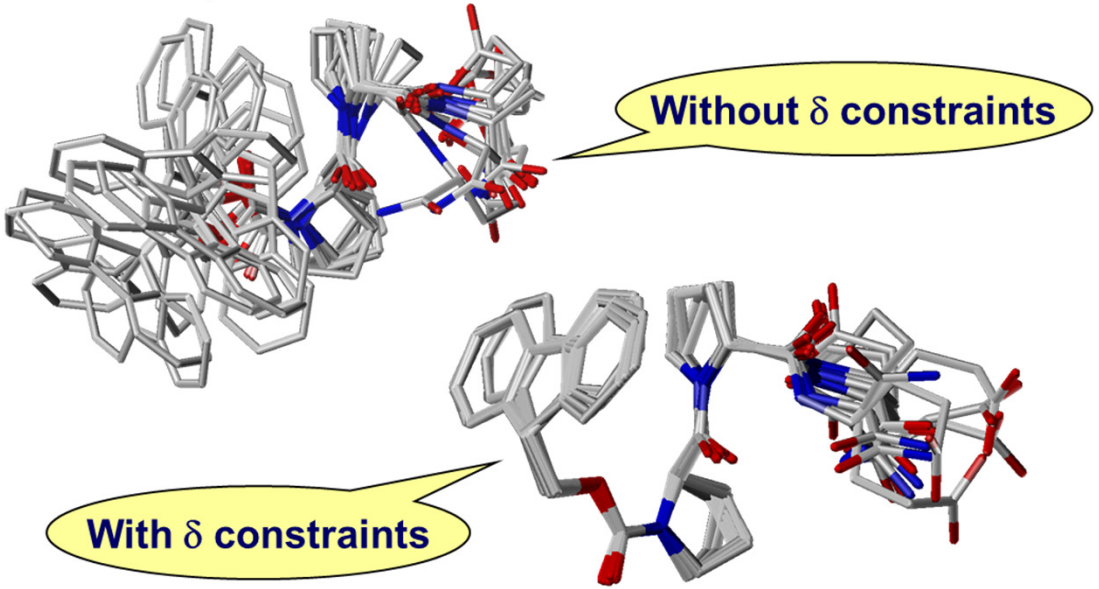
Use of chemical shifts (...)

• So, since we have equations for each effect, we can calculate it to a certain degree of accuracy in the computer. If we know both the random coil and the experimental value we can tell the MM program to make the calculated mach the observed values or else put an energy penalty:

$$E_{\sigma} = K_{\sigma} * [(\delta_{obs} - \delta_{random}) - (\sigma_{pga} + \sigma_{rc} + \sigma_{elec})]^2$$

$\delta_{obs} - \delta_{random}$ is the ***secondary shift***

• This works great in some cases. The following case had no NOEs, but a lot of secondary shifts...



Brief introduction to molecular modeling

• Now we have all (almost all…) the information pertaining structure that we could milk from our sample: NOE tables with all the different intensities and angle ranges from ^{3}J coupling constants.

• We will try to see how these parameters are employed to obtain the 'picture' of the molecule in solution.

• As opposed to X-ray, in which we actually 'see' the electron density from atoms in the molecule and can be considered as a 'direct' method, with NMR we only get indirect information on some atoms of the molecule (mainly ^{1}Hs…).

• Therefore, we will have to rely on some form of theoretical model to represent the structure of the peptide. Usually this means a computer-generated ***molecular model***.

• A molecular model can have different degrees of complexity:

  • ***ab initio*** - We actually look at the atomic/molecular orbitals and try to solve the Schröedinger equation. No parameters. Hugely computer intensive (10 - 50 atoms).

  • ***Semiempirical*** - We use some ***parameters*** to describe the molecular orbitals (50 - 500 atoms).

  • ***Molecular mechanics*** - We use a simple ***parametrized mass-and-spring*** type model (everything else…).

Introduction to molecular modeling (continued)

• We are dealing with peptides here (thousands of atoms), so we obviously use a molecular mechanics (**MM**) approach.

• The center of **MM** is the ***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***:

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

• Each term depends in a way or another in the geometry of the system. For example, E_{bs}, the ***bond stretching*** energy of the system is:

$$E_{bs} = \sum_i K_{bsi} * (r_i - r_{oi})^2$$

• The different constants (K_{bs}, r_o, etc., etc.) are called the ***parameters*** of the force field, and are obtained either from experimental data (X-ray, microwave data) or higher level computations (*ab initio* or semiempirical).

• Depending on the problem we will need different parameter sets that include (or not) certain interactions and are therefore more or less accurate.

Inclusion of NMR data

• The really good thing about MM force fields is that if we have a function that relates our experimental data with the **<xyz>** coordinates, we can basically lump it at the end of the energy function.

• This is exactly what we do with NMR data. For NOEs, we had said before that we cannot use accurate distances. We use ranges, and we don't constraint the lower bound, because a weak NOE may be a long distance or just fast relaxation:

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

• Now, the potential energy function related to these ranges will look 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. We call them ***NOE constraints***.

Inclusion of NMR data (continued)

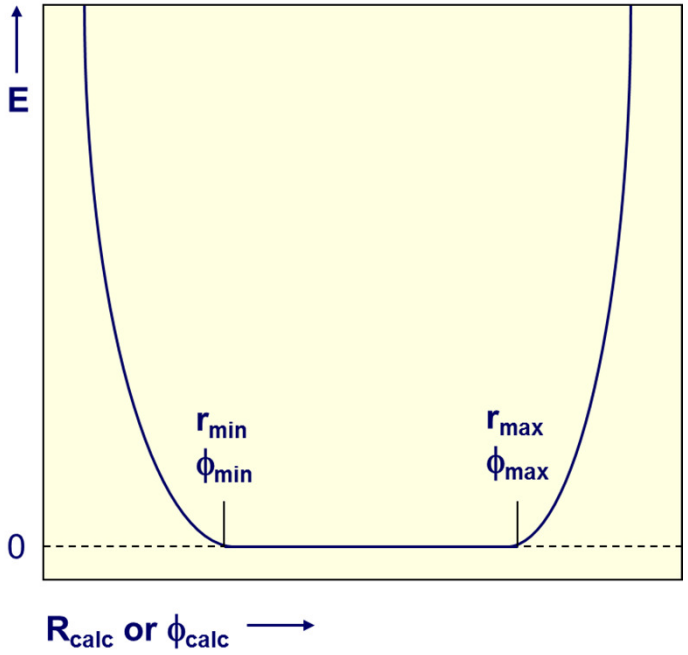
- Similarly, we can include torsions as a range constraint:

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

$$E_J = 0 \quad \text{if } \phi_{max} > \phi_{calc} > \phi_{min}$$

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

- Graphically, these penalty functions look like this:

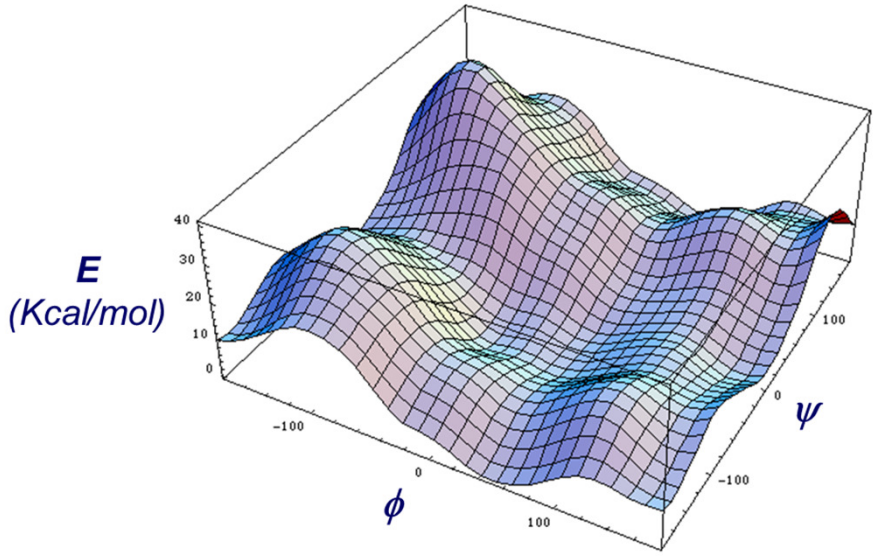


Structure optimization

• Now we have all the functions in the potential energy expression for the molecule, those that represent bonded interactions (bonds, angles, and torsions), and non-bonded interactions (vdW, electrostatic, NMR constraints).

• In order to obtain a decent model of a peptide 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.

• In a function with so many variables this is nearly impossible, because we are looking at a ***n-variable surface*** (each thing we try to optimize). For the two torsions in a disaccharide:



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

Structure optimization (continued)

• Minimizing the function means going down the ***energy (hyper)surface*** of the molecule. To do so we need to compute the derivatives WRT **<xyz>** (variables) for all atoms:

$$\frac{\partial E_{total}}{\partial xyz} > 0 \longrightarrow E_{total} \uparrow \qquad \frac{\partial E_{total}}{\partial xyz} < 0 \longrightarrow E_{total} \downarrow$$

• This allows us to figure out which way is 'down' for each variable so we can go that way.

• Now, minimization only goes downhill. We may have many local minima of the energy surface, and if we only minimize it can get trapped in one of these. This is bound to happen in a protein, which has hundreds of degrees of freedom (the number of rotatable bonds…).

• In these cases we have to use some other method to get to the lowest minima. A common way of doing this is ***molecular dynamics*** (**MD**).

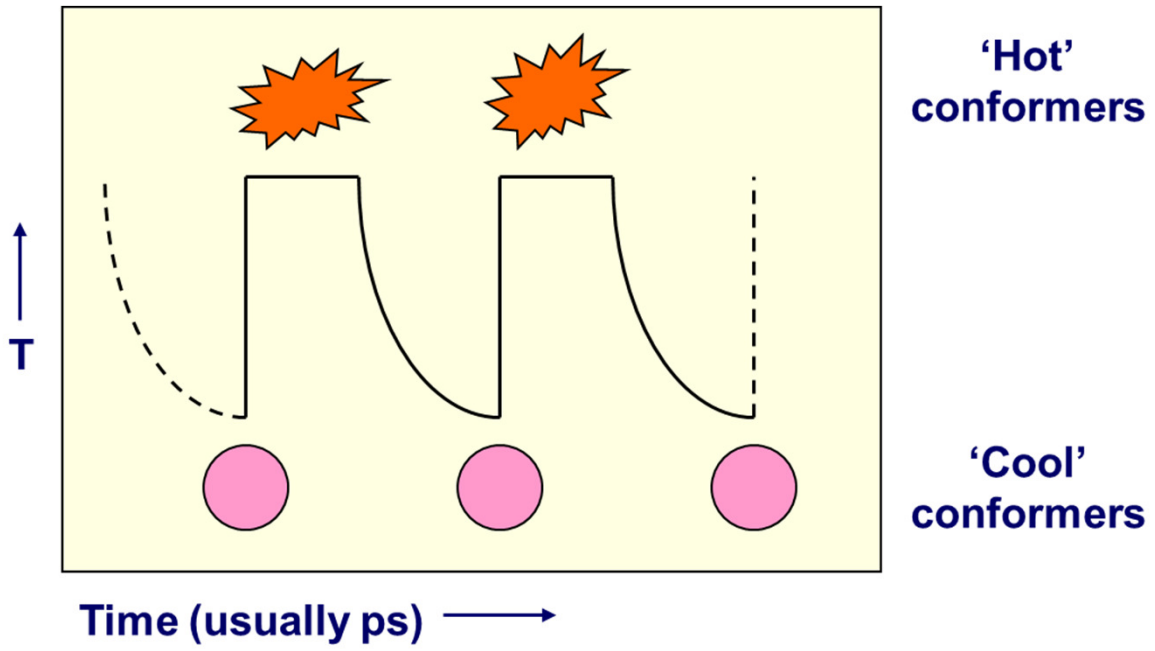
• Since we have a the energy function we can give energy to the system (usually we rise the 'temperature') and see how it evolves with time. Temperature usually translates into kinetic energy, which allows the peptide to surmount energy barriers.

Molecular dynamics and simulated annealing

• In MD we usually 'heat' the system to a physically reasonable temperature around 300 K. The amount of energy per mol at this temperature is ~ $\mathbf{k_BT}$, were $\mathbf{k_B}$ is the Boltzmann constant. If you do the math, this is ~ 2 Kcal/mol.

• This may be enough for certain barriers, but not for others, and we are bound to have this 'other' barriers. In these cases we need to use a more drastic searching method, called ***simulated annealing*** (called that way because it simulates the annealing of glass or metals).

• We heat the system to an obscene temperature (1000 K), and then we allow it to cool slowly. This will hopefully let the system fall into preferred conformations:



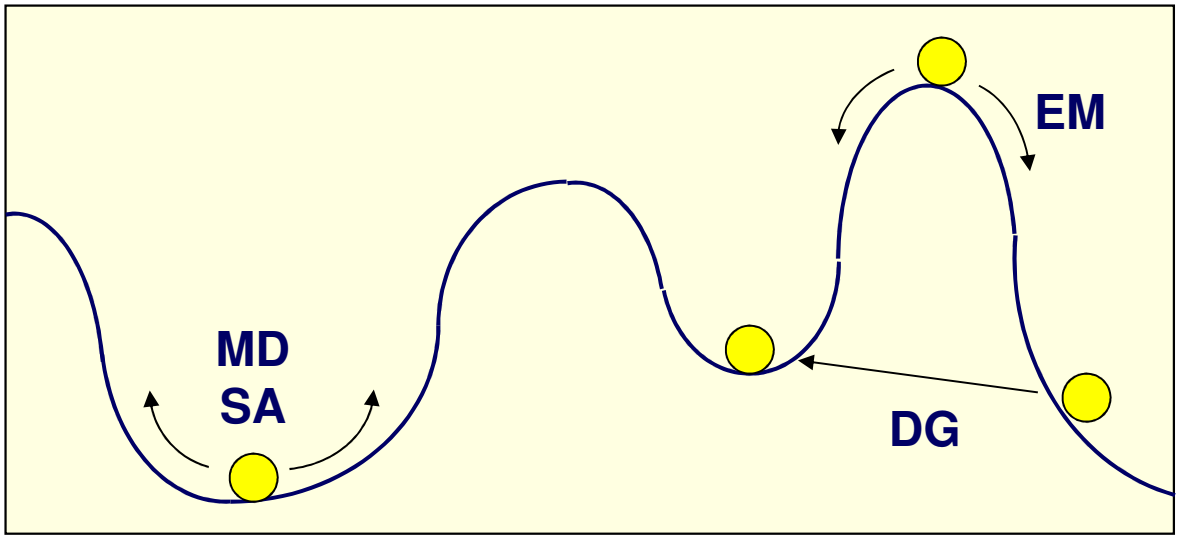
Distance geometry

• Another method commonly used and completely different to **MD** and **SA** is distance geometry (**DG**). We'll try to describe what we get, not so much how it works in detail.

• Basically, we randomize the **<xyz>** coordinates of the atoms in the peptide, putting a low and high bounds beyond which the atoms cannot go. These include normal bonds and NMR constraints.

• This is called ***embedding*** the structure to the bound matrix. Then we optimize this matrix by triangle inequalities by ***smoothing*** it. We get really shuffled and lousy looking molecules. Usually they have to be refined, either by **MD** followed by minimization or by straight minimization.

• What the different methods do in the energy surface can be represented graphically:

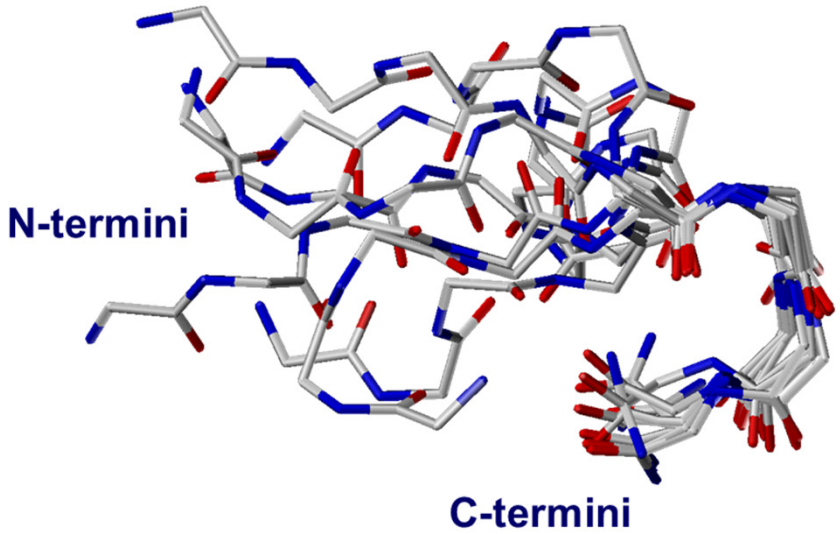


Presentation of results

• The idea behind all this was to sample the conformational space available to the protein/peptide under the effects of the NOE constraints.

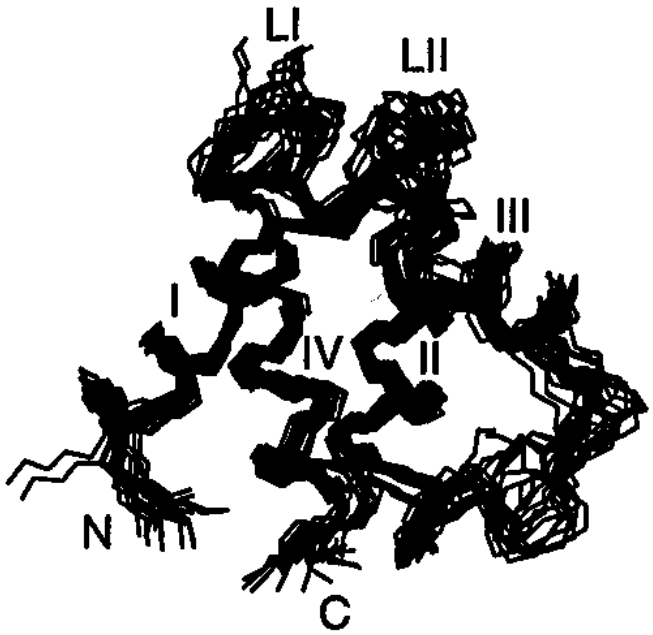
• The several low energy structures we obtain by these methods which have no big violations of these constraints are said to be in agreement with the NMR data.

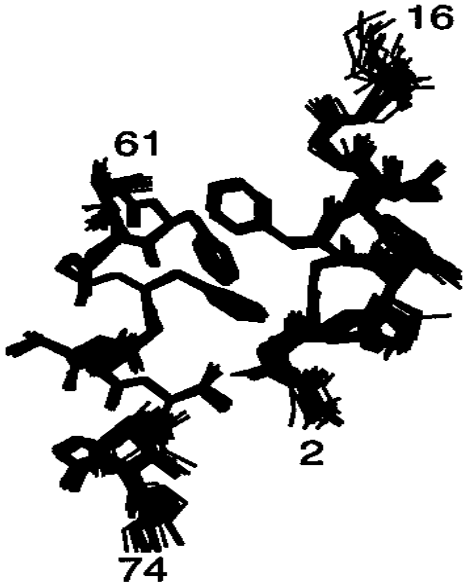
• Since there is no way we can discard any of this structures, we normally draw a low energy set of them superimposed along the most fixed parts of the molecule:



• In this one we are just showing the peptide backbone atoms. Although this is not a sought for thing, the floppiness of certain regions is an indication of the lack of NOE constrains, which reflects the real flexibility of the molecule in solution.

Characteristics of Structures Determined in Solution by NMR

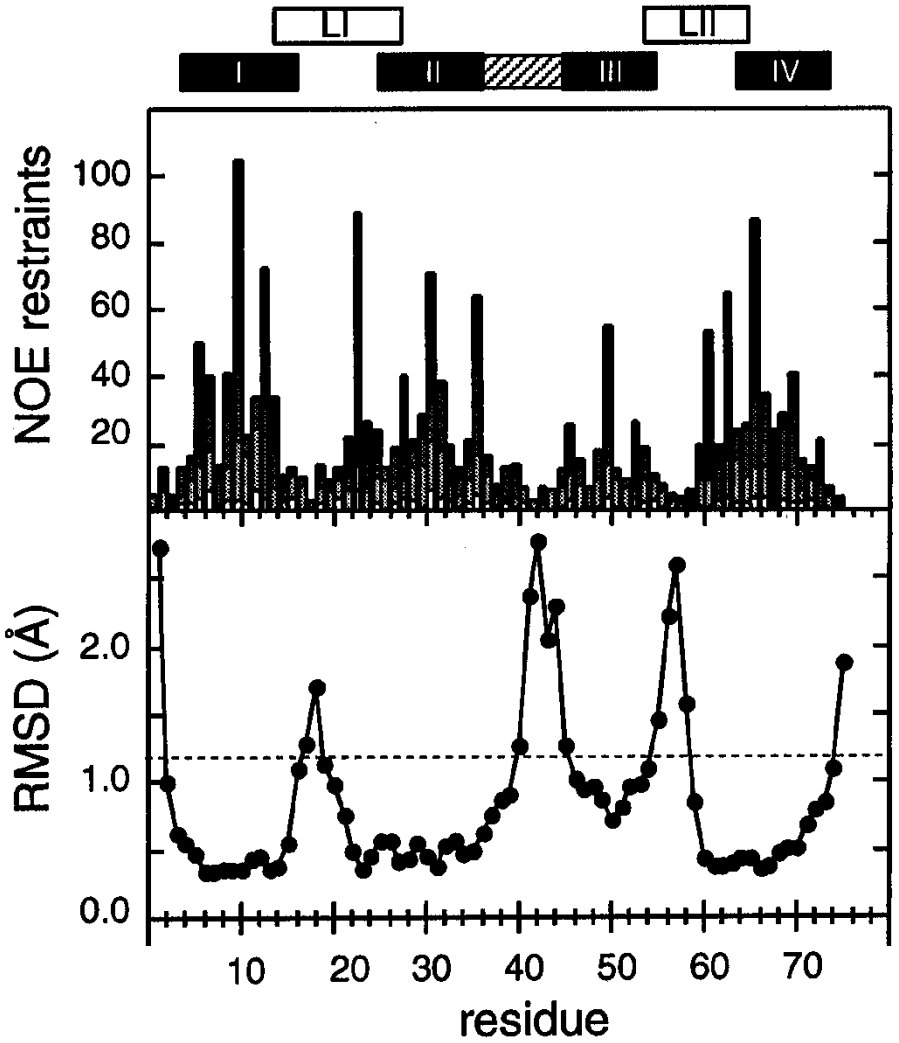




- Secondary structures well defined, loops variable
- Interiors well defined, surfaces more variable
- RMSD provides measure of variability/precision (but not accuracy!)

Kordel et al., JMB (1993)

Restraints and Uncertainty



- **Large # of restraints = low values of RMSD**

Kordel et al., JMB (1993)

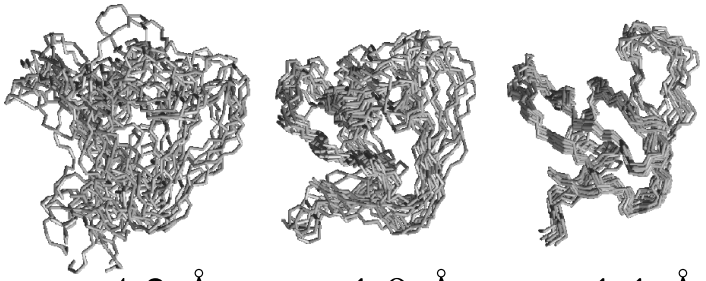
Validation criteria for protein structures

- 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

Quality of structures

- Precision

$$RMSD = \sqrt{\frac{1}{n}\sum \left| r_i - Rav_i \right|^2}$$



4.2 Å 1.9 Å 1.1 Å

Too low RMSD values are meaningless in solution at room temperature

- Accuracy

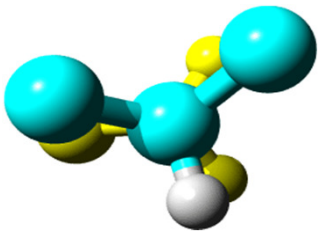
$$E_{pot} = \sum U_i \left(d_i - d_i^0\right)^2 + \sum W_i \left(1 + \cos\left(\vartheta_i - \vartheta_i^0\right)\right)^2 + VdW$$

+ other constraint contributions

- Procheck statistics expected for a good quality structure:

< 10 bad contacts per 100 residues
Average hydrogen bond energy in the range of 2.5-4.0 kJmol^{-1}
Overall G-factor > –0.5

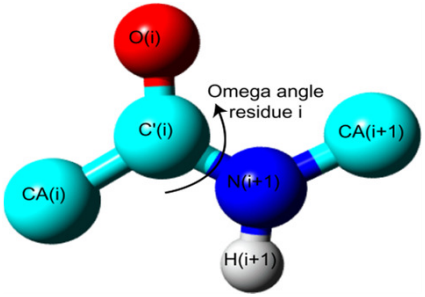
Rotameric states



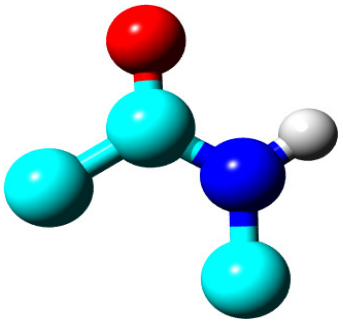
Eclipsed

Staggered

Omega angles

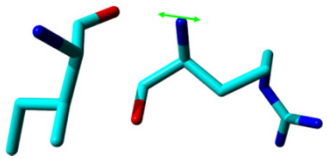


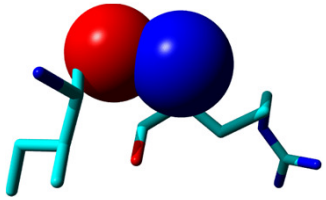
Trans-conformation
(omega=180°)



Cis-conformation
(omega=0°)

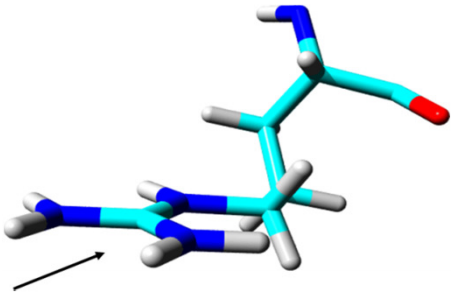
Inter-atomic bumps



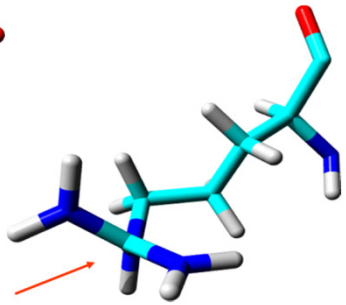


Overlap of two backbone atoms

Side-chain planarity

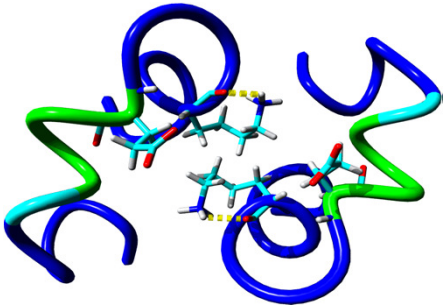


Planar ARG side-chain
(Good)

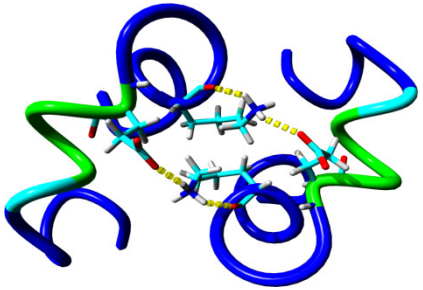


Non-planar ARG side-chain
(Bad)

Electrostatics

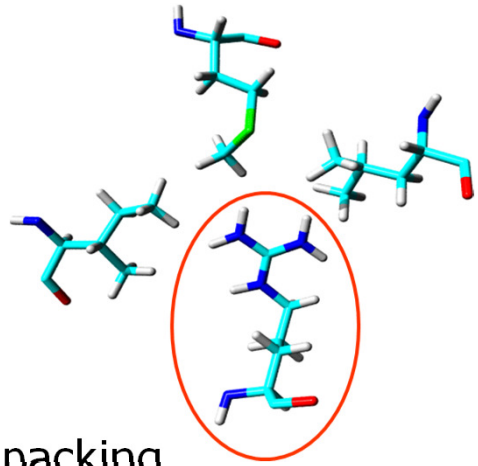


"Bad" electrostatics

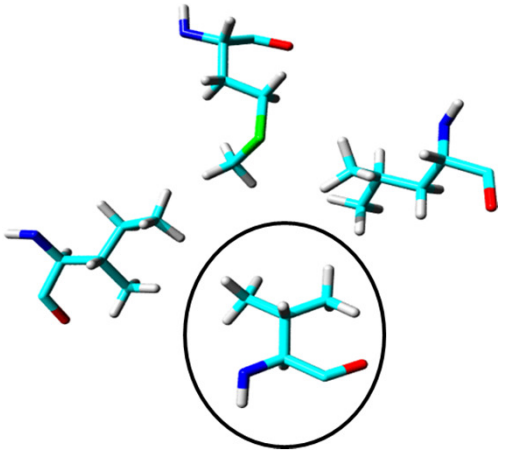


After energy minimization
including electrostatics

Packing quality

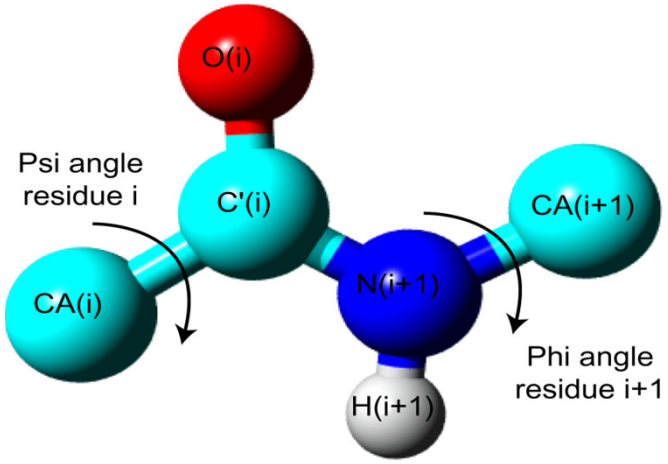


Bad packing

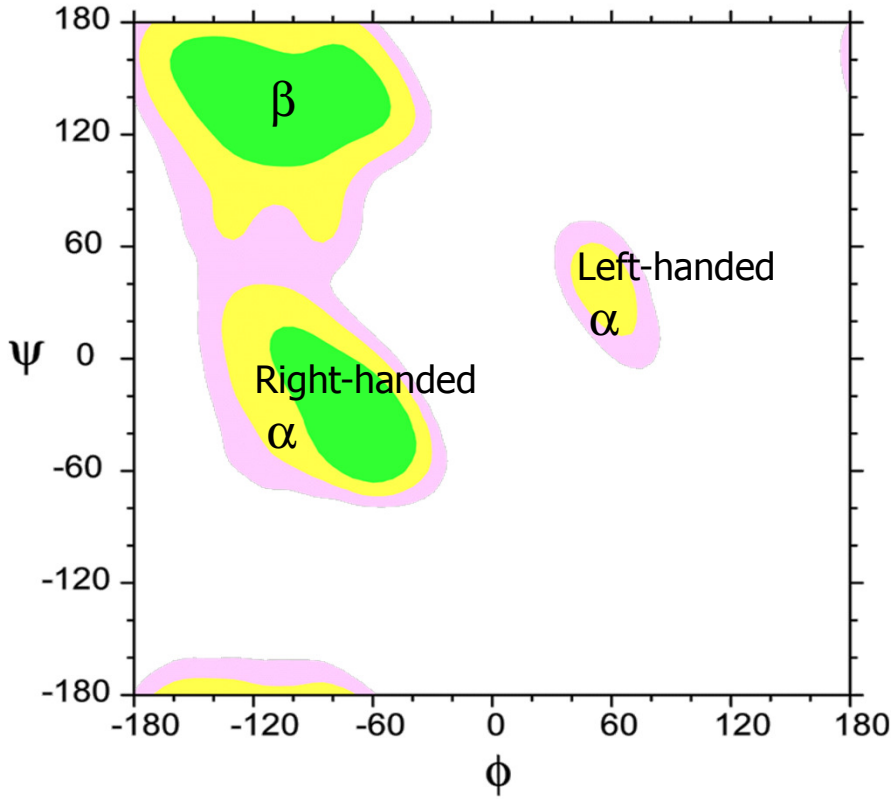


Good packing

Ramachandran Plot



Phi and Psi angles



Ramachandran plot

Ideally, one would hope to have over 90% of the residues in these "core" regions