NMR (Nuclear Magnetic Resonance Spectroscopy)

- literature / background (already in Stine)
  - Thomas James chapter http://www.biophysics.org/img/James.T.pdf
  - Ferentz, A.E. and Wagner, G., Q. Rev. Biophys, 33, 29-65 (2000)

current standing

- ≈ 13 % of all current structures solved by NMR
- about 1/3 of smaller structures

Next 3 Weeks

- Background to NMR – chemistry
- Calculating structures
  - distance geometry
  - problems with structures

History

- younger field than X-ray
  - one Nobel prize in early 90's (Ernst – technical)
  - ½ Nobel prize 2002 (Wüthrich)
- first real protein structure about 1985 or 1986

NMR from our viewpoint

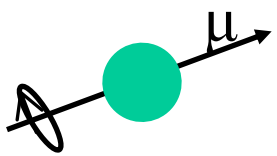
- a way to get structures
- can provide information on
  - dynamics, stability
  - interactions (other proteins, small molecules)
- we concentrate on structural aspects

Overview – how we get coordinates

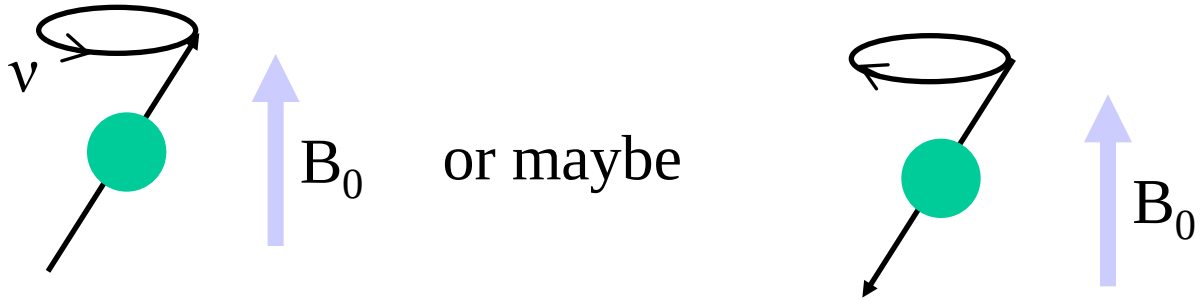
- protein in solution
- record spectra
- assign peaks to ^{1}H, ^{13}C, ^{15}N nuclei
- record some more spectra
  - distance information (mostly)
  - some internal angles
- reconstruct structure

Nuclei have spin

- have a charge and act like magnets
- put them in a field and they will align with it



- now apply a magnetic field
  - they "precess" around the field
  - two possible states

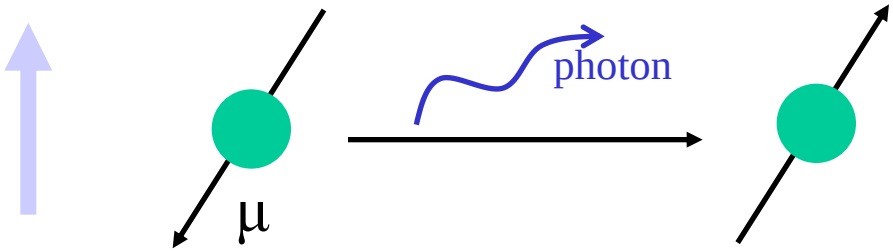


B_0 is applied field

ν speed of rotation (many MHz / 10^6 Hz)

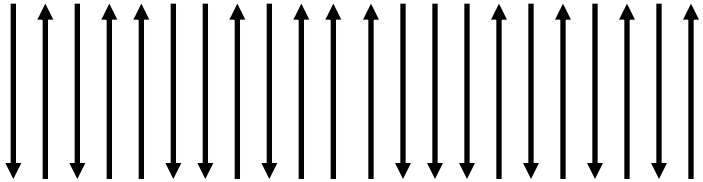
Do nuclei like fighting the field ?

- is a nucleus really happy facing the wrong way ?
- what if we push it the wrong way ?
  - wants to get to low energy state – emits a photon



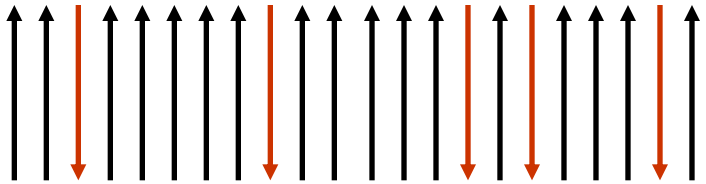
- energy difference very small

What NMR records



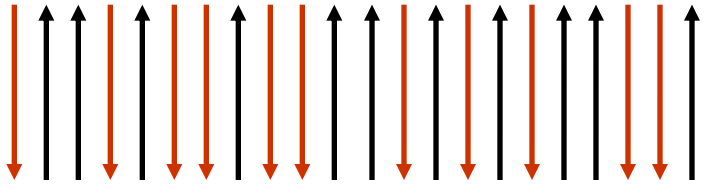
some nuclei not doing much

turn on a field



B_0 in an applied field, some align

put in energy

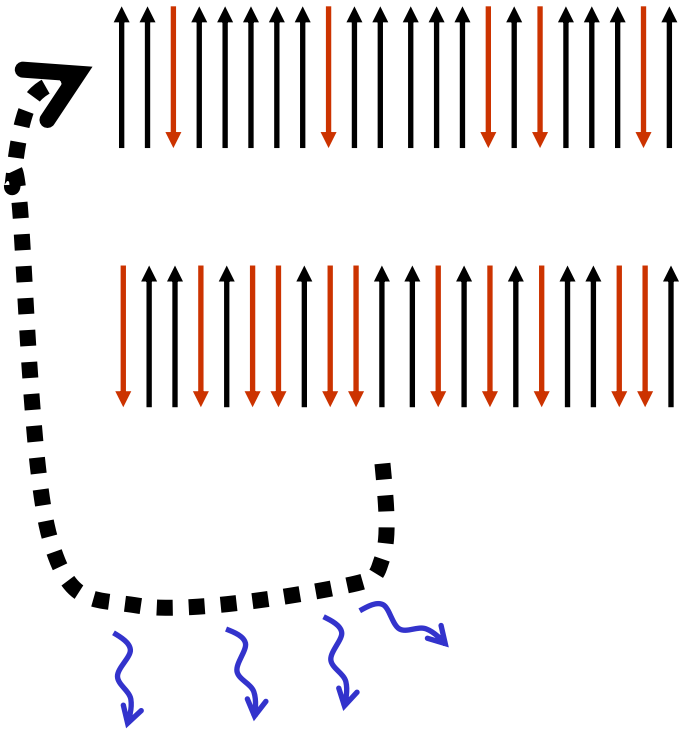


B_0

let them relax

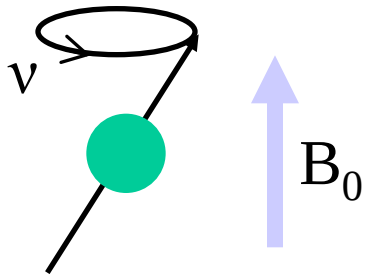
Still not really interesting spectrocopy

Is this useful ?

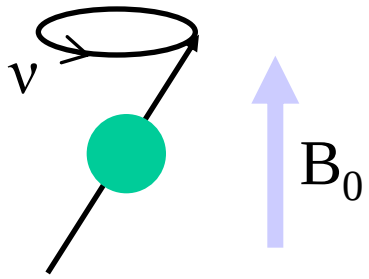


- record some photons (radio freq) no information (yet)
- what if the nuclei emit slightly different frequency energy ?

- what determines the frequency ?
  - energy difference
    - field strength

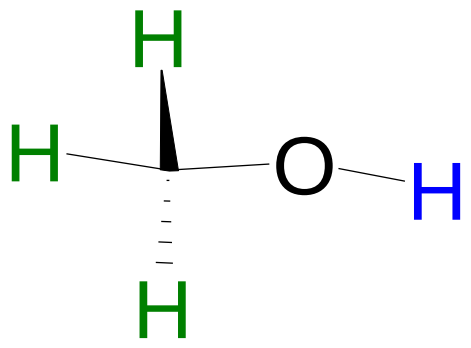


- B_0 applied field
- υ Larmor frequency
- γ magic number for nucleus (gyromagnetic ratio) purely empirical



$$\nu = \gamma B_0 / 2\pi$$

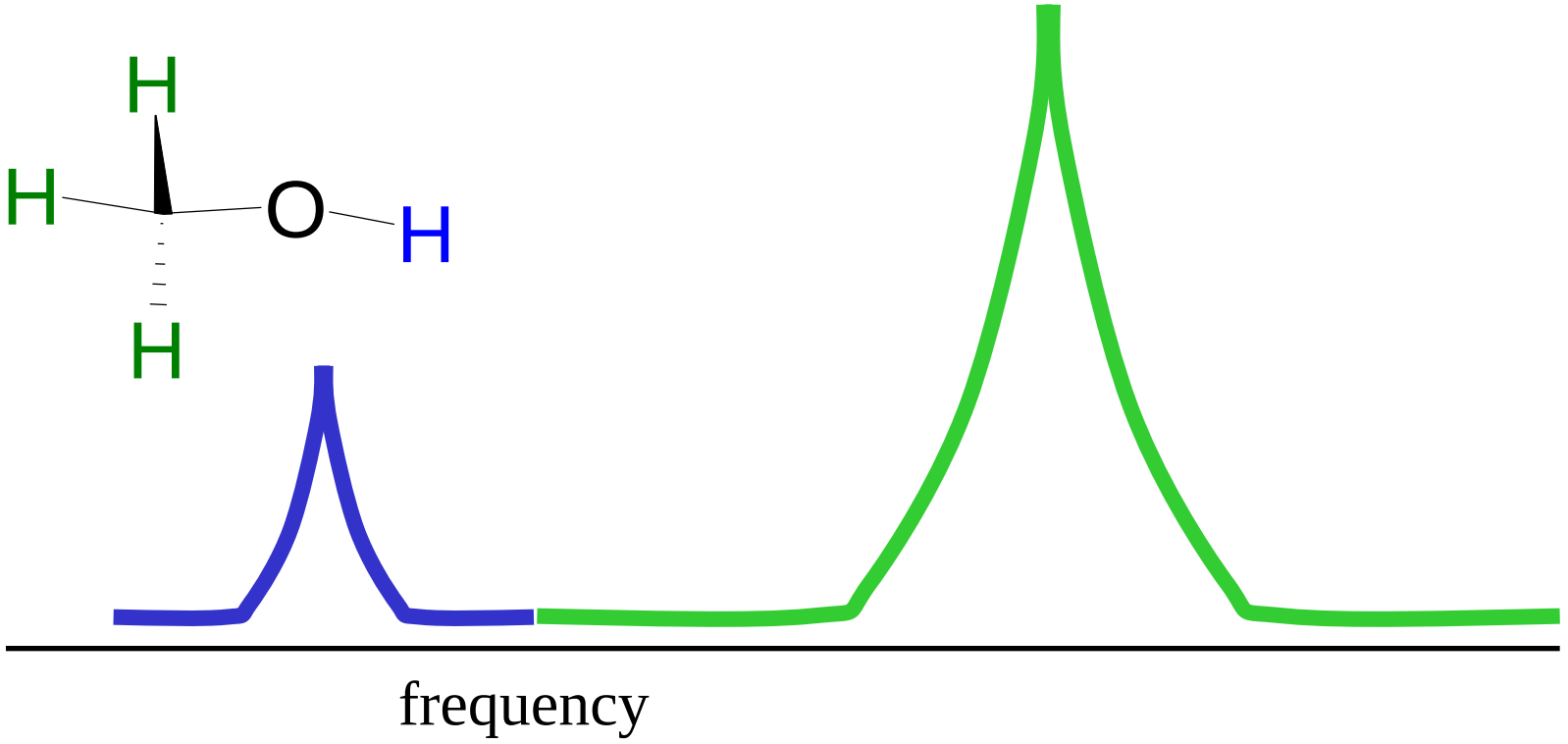
- What is the real field that a nucleus sees ?
- mixture of outside field and local environment



blue H is different to green H
so frequency should change

A possible toy spectrum

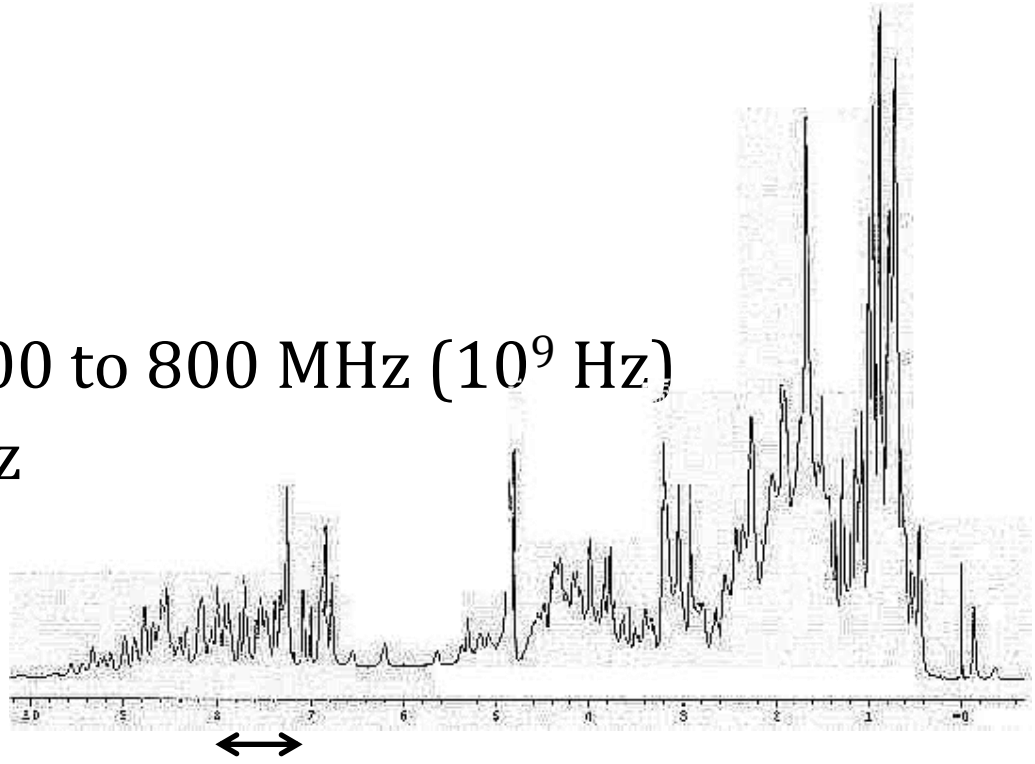
- different atoms / nuclei give different peaks



- a more interesting spectrum …

chemical shift / real spectrum

- some protein
  - 100's ^{1}H
- Scales ?
  - all peaks resonating 100 to 800 MHz (10^{9} Hz)
  - whole spectrum 10^{4} Hz



100 Hz

Important nuclei (spin ½)

nucleus	sensitivity	notes
^{1}H	1	cheap and natural
^{13}C	1.6×10^{-2}	expensive, but only 1% of natural abundance
^{15}N	10^{-3}	bit less expensive, 0.4 % natural abundance
^{31}P	7×10^{-2}	fun for DNA and other PO_4 chemistry

- but the natural isotopes are ^{12}C and ^{14}N
  - (usually) these isotopes require labelling
- other nuclei ? ...

A simple spectrum

- an example protein (ubiquitin)
  - lots of peaks, but not useless
- could already
  - look at ligand binding
  - pk_a of residues
- no real structural information yet

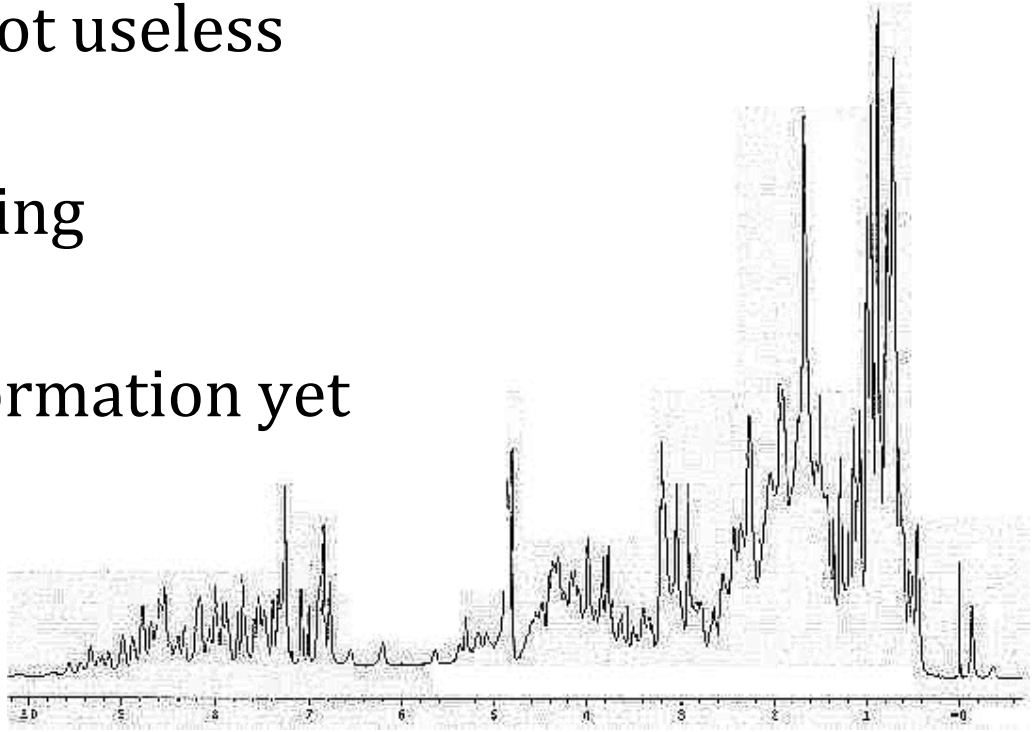
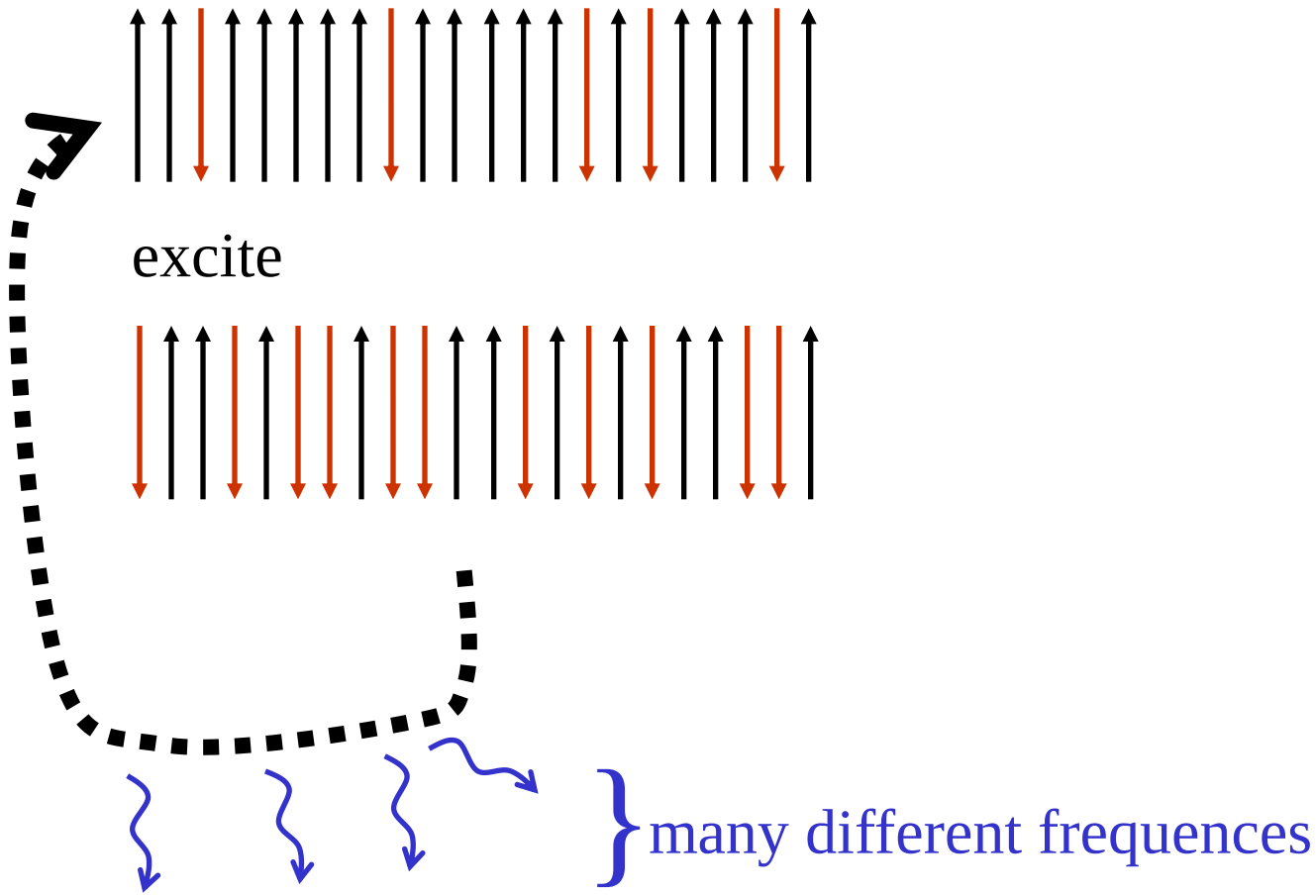


diagram from http://bouman.chem.georgetown.edu/nmr/protein.htm

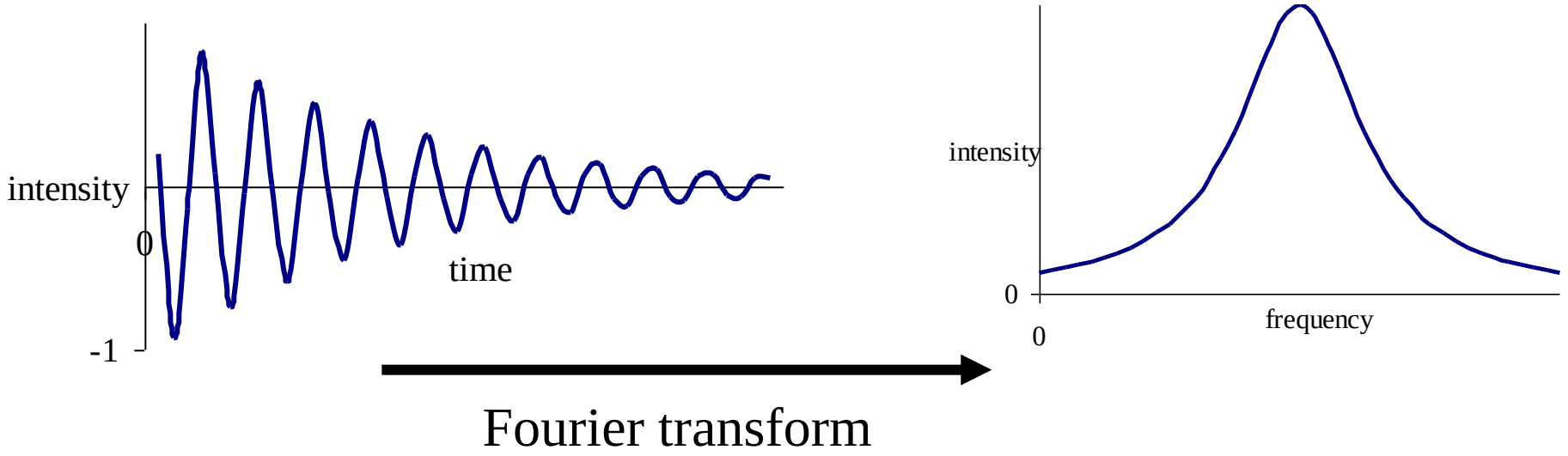
Recording a spectrum



sort out frequencies with Fourier transform

Raw data and Fourier transforms

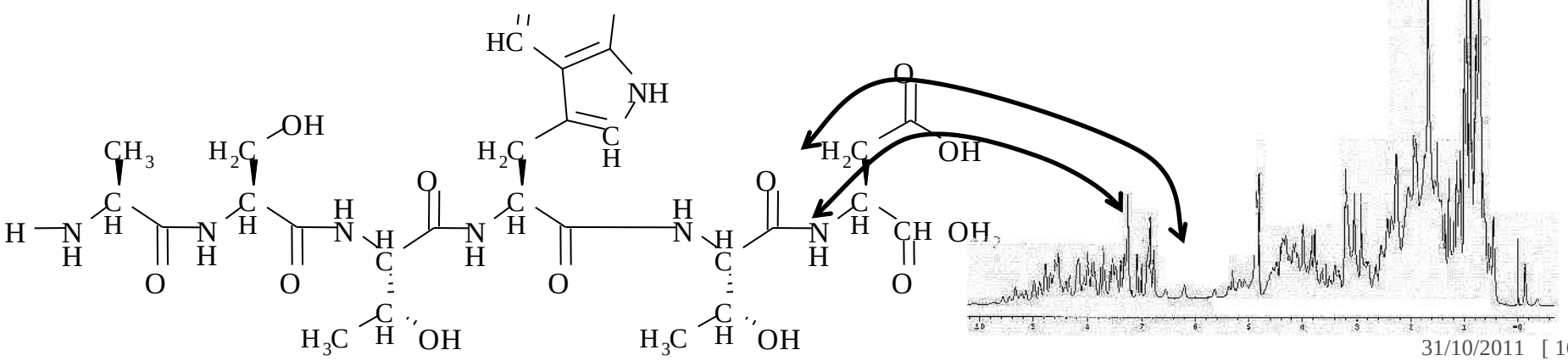
- raw data will be simple periodic functions + decay



Assignment

- Peaks correspond to ^{1}H / ^{13}C from specific atoms
- structural information given by peaks
  - more later
- Assignment – which peaks correspond to which atoms

- Assumption
  - we know exactly which atoms are present
    - sequence of protein known
    - know the structure of each amino acid

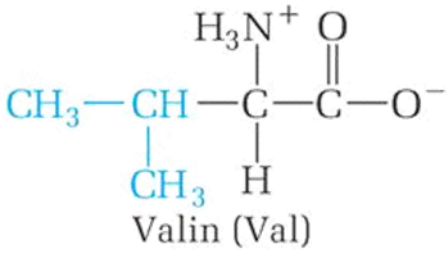


Assignment

- Which peaks correspond to which atoms ?
  - location / chemical shift

  - with which atoms are you bonded ?
  - which atoms must you be near in space

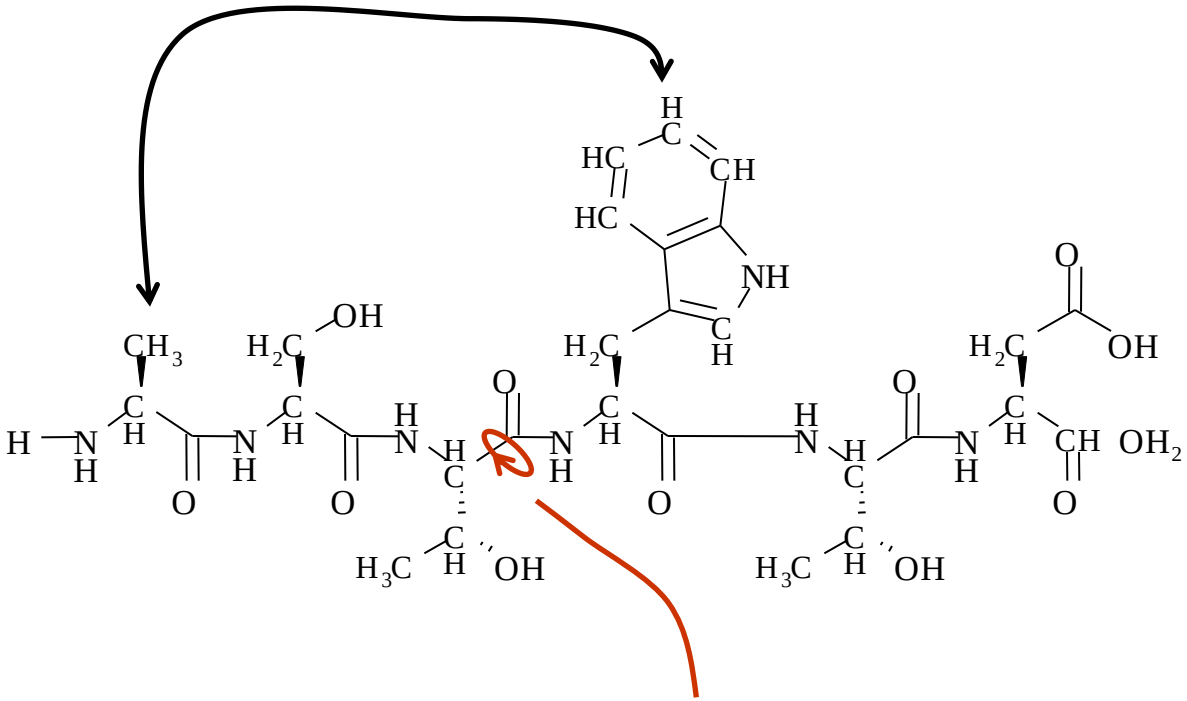
- Example
  - find a methyl peak
    - connected to a methyne
      - connected to a methyl & another methyne
      - …
- mostly automated



Valin (Val)

To calculate structures ?

1. distance information



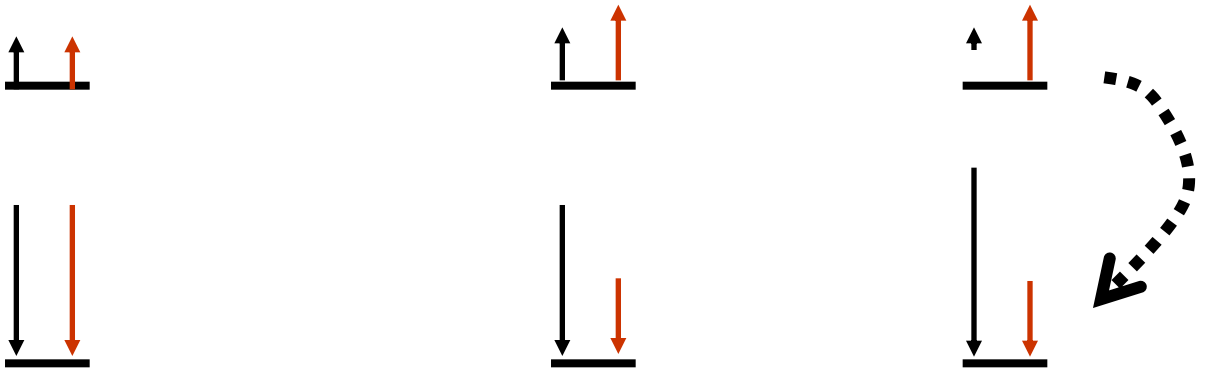
2. dihedral / torsion angle information

Distance information / the NOE

- most important
  - an effect which depends on how close in space nuclei are
  - NOE $\propto r^{-6}$
  - usually only up to about 5 - 6 Å
- story
  - two spins' dipoles interact
  - saturating one spin affects populations of other spin
- who wants an explanation ?
  - cross relaxation phenomenon

• red relaxing (jumping to lower energy) affects black
• can one create this situation ?

Cross relaxation and the NOE



equilibrium

saturate red spins

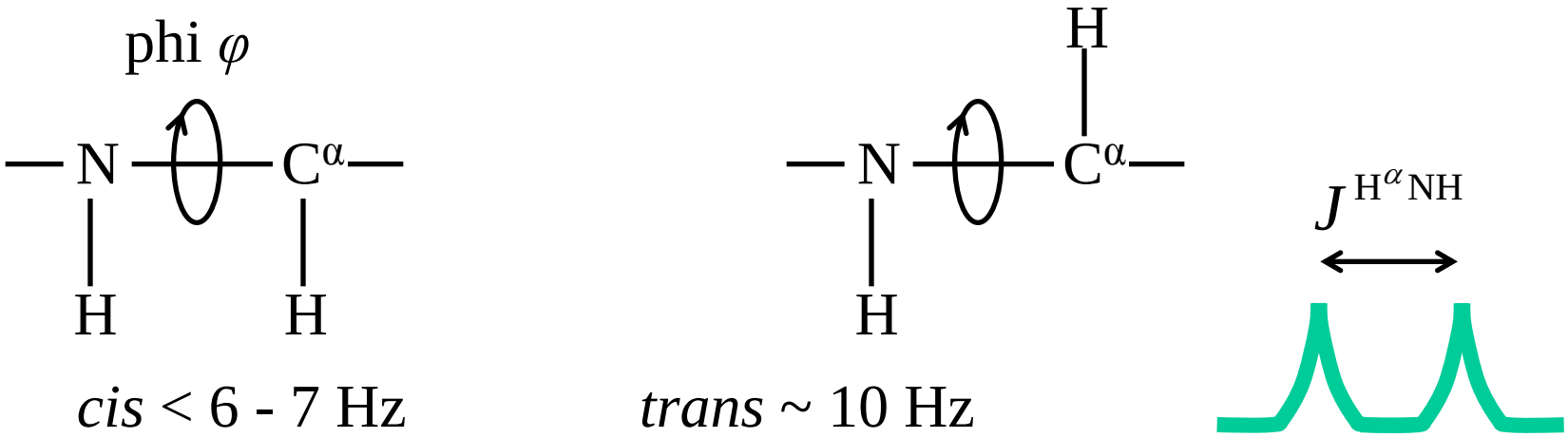
both relax together

- now, the population difference is bigger than normal
  - bigger signal
- record a normal spectrum
  - red is not there
  - black is "enhanced"
- via another mechanism
  - population difference can become smaller
- only happens if nuclei are very close in space

Other structural information

- NOE – information about short (< 5 or 6 Å) distances
- there is more – angles
  - mainly *J* coupling

Amide NH to H^{α} coupling

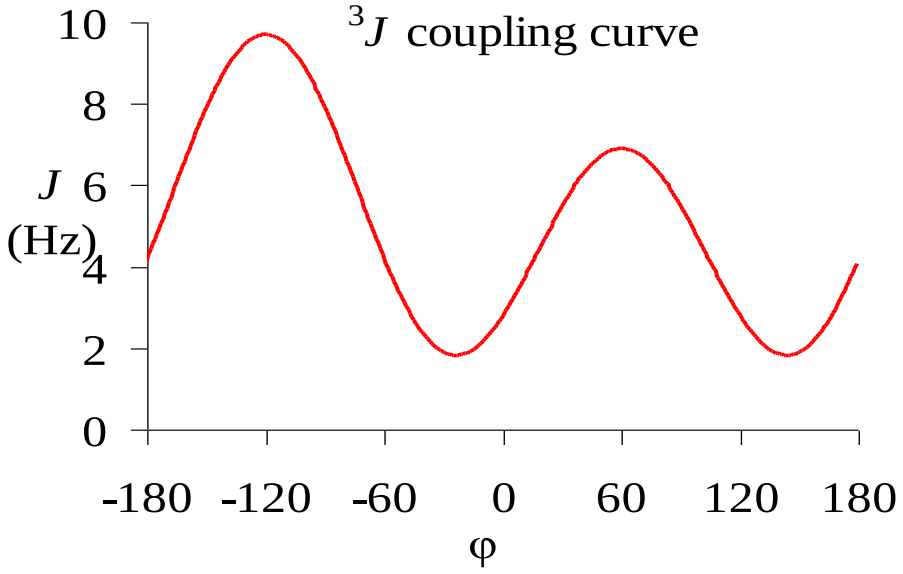


$^3J_{HN\alpha}$ coupling

- formalised as

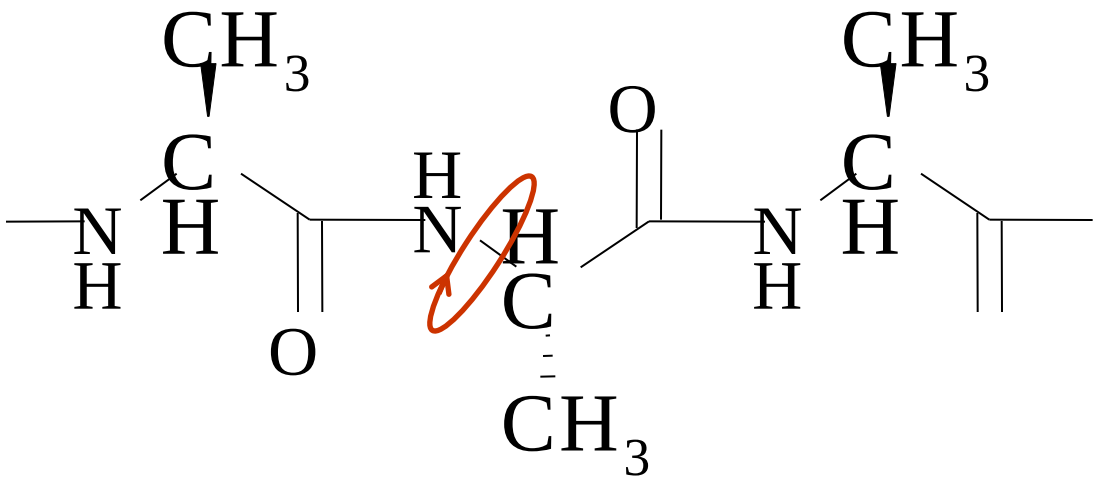
$$^3J_{HN^\alpha} = 6.4\cos^2\theta - 1.4\,cos\theta + 1.9$$

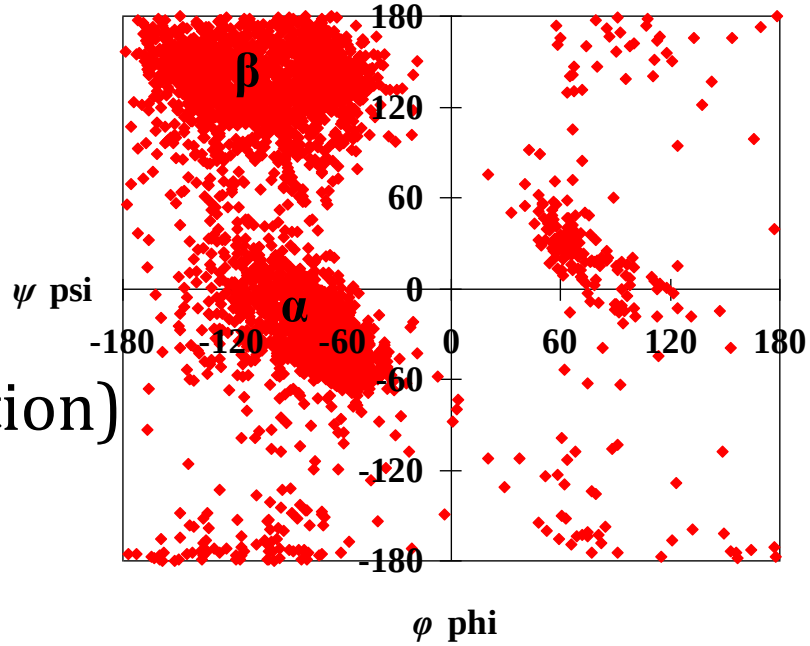
Problems later



from Pardi, A, Billeter, M and Wüthrich, K, J. Mol. Biol. 180, 741-751 (1984)

Amide NH to H^{α} coupling



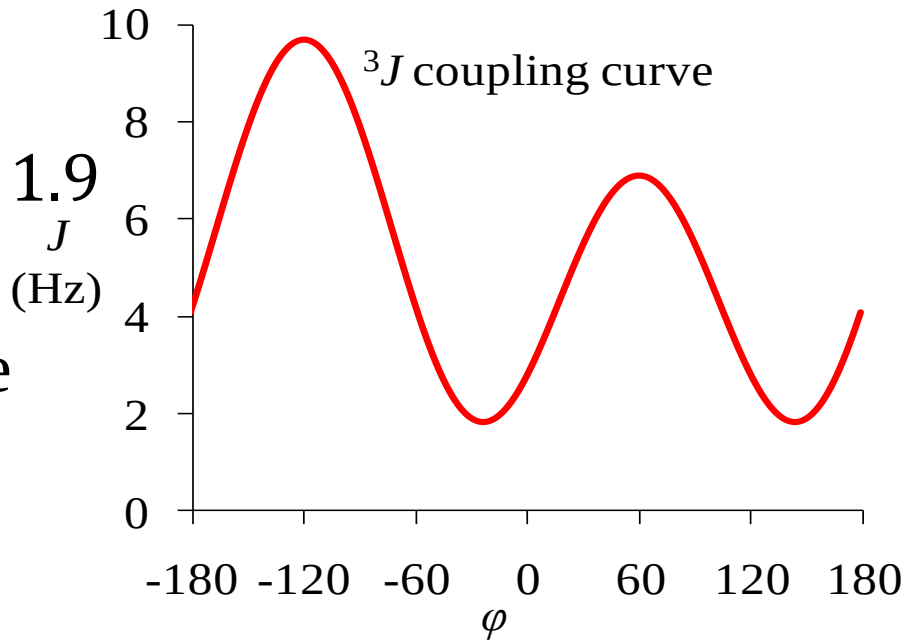


- can help distinguish α from β
- not always seen (exchange / motion)
- NH not always present
- other angles ?
  - other vicinal protons
    - C^{α} to C^{β}

Problems with *J*-coupling

1. we have a formula

$$^{3}J_{HN^{\alpha}} = 6.4\cos^{2}\theta - 1.4\,cos\theta + 1.9$$



- most of the time, there is more than one solution
- only use very big *J* values

2. dynamics

more serious than they appear ! look around -90°

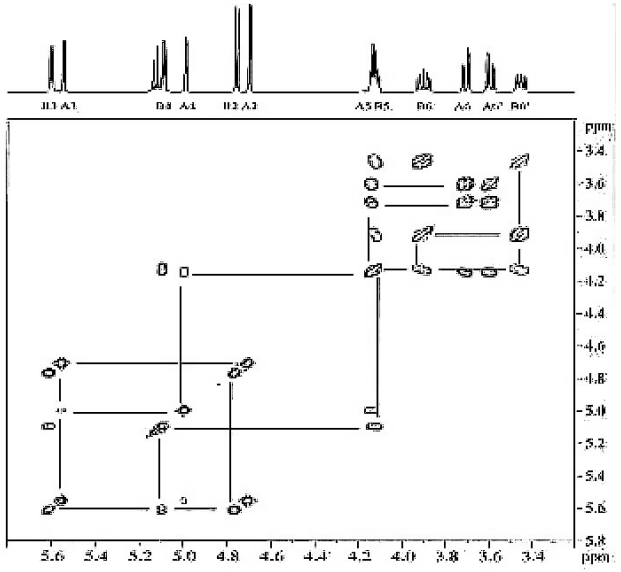
Practical NMR

- We have some basic methods
- Real NMR
  - more techniques
    - identifying specific kinds of atom
    - spreading peaks out

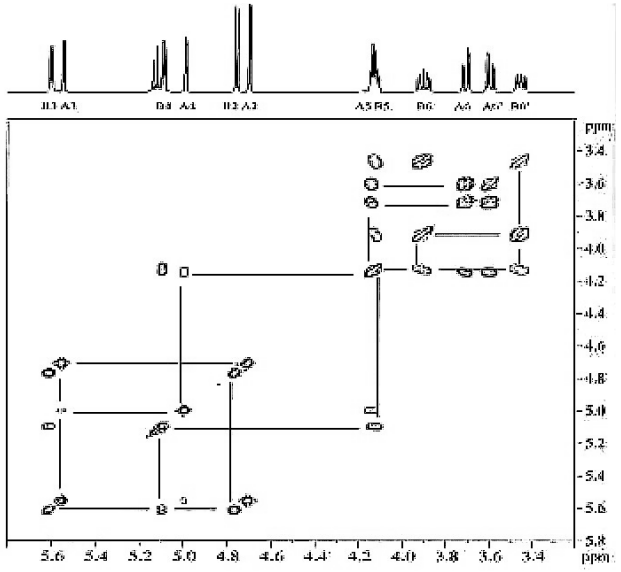
- Briefly mention the most important…
  - 2D NMR

2D NMR

- two reasons
  1. spread spectrum out
     - resolve peaks / remove overlap
  2. add information



2D spectra information



What do the off-diagonal peaks mean ?

- depends on spectrum

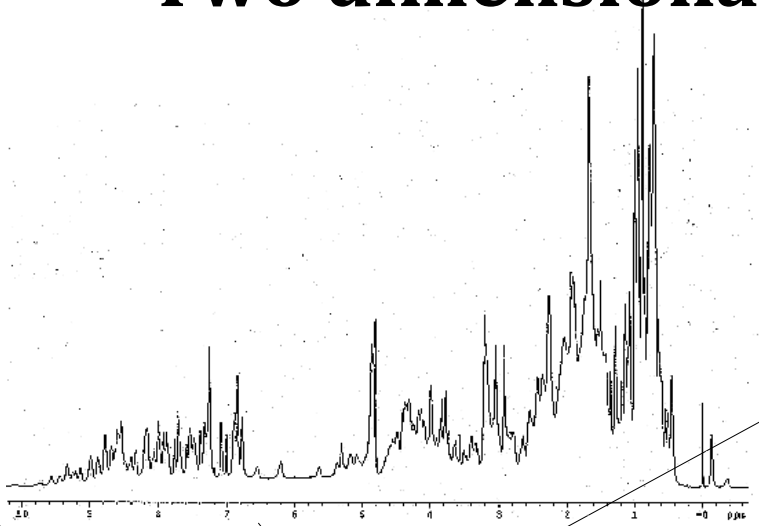
Example 1

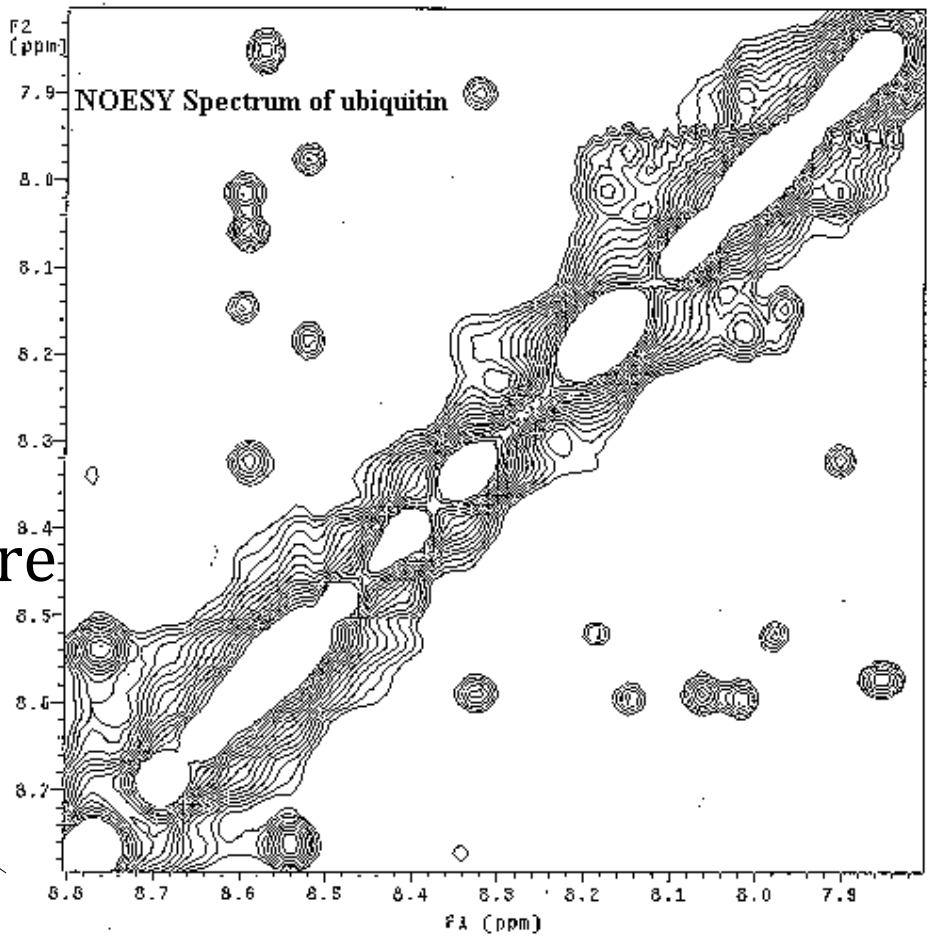
- COSY (correlated spectroscopy)
- peaks indicate J-coupling
- look at spectrum and quickly see which peaks are connected

Example 2

- NOESY (NOE …)
- peaks indicate NOE
- corresponding nuclei close in space

Two dimensional NOE spectra example





2 D NOE spectrum

- NOESY

- what determines if peaks are present ?

diagram from http://bouman.chem.georgetown.edu/nmr/protein.htm

Information summary

phenomenon	assignments	structure
chemical shift	important	not used
spin-spin (*J*) coupling	important	torsion angles
NOE / distances	important	main information

- more spectroscopy
  - filtering according to chemistry, atom types
  - *n*-dimensional methods
- structural information
  - labels for broadening / shifting peaks
  - orientation of bonds to reference ..

Structures from NMR data

To come

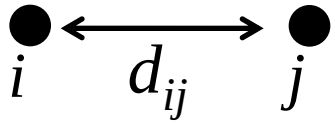
- Distances in 2 and 3 D
- Distance geometry
  - 2 approaches
- Restrained molecular dynamics (MD)

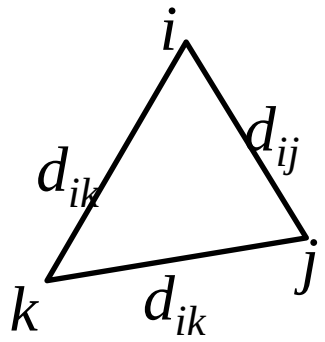
Available information

- distances
  - short (5 to 6 Å)
  - incomplete
- some dihedral / torsion angles
- does this define a structure ?
  - strictly no
  - with chemical information ?
    - still not

Determining distances (ideal)

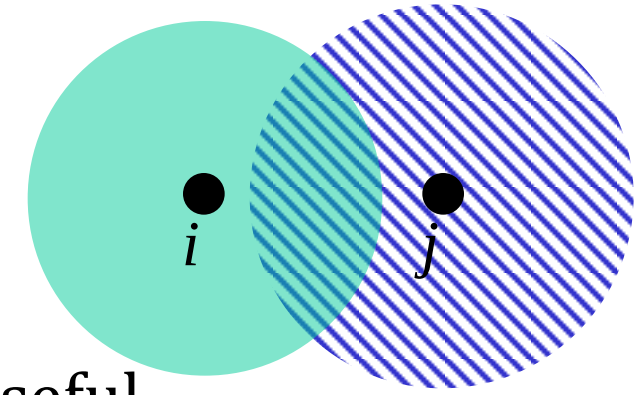
- 2 points 1 distance
- 3 points 3 distances…
  - think of $3N_{atom}$ distances
  - remember $N_{atom} \approx 10$ or $20\ N_{res}$

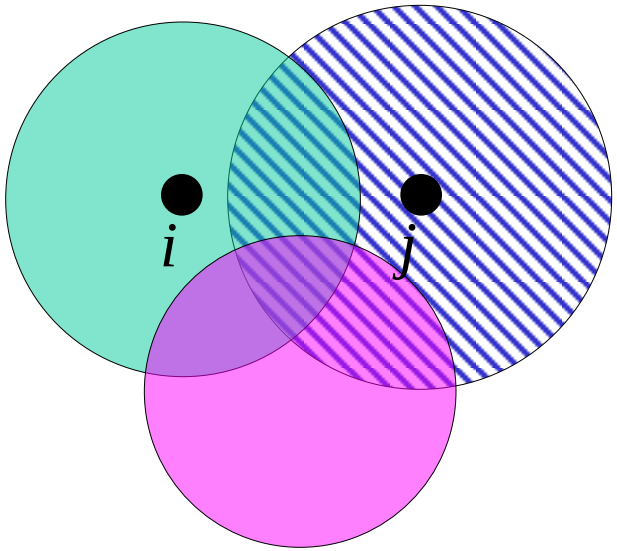




Underdetermined distances

- think in terms of triangles …
  - $d_{ik} < 6$ Å, $d_{jk} < 6$ Å
  - where is k ?
- a few more distances…
  - more and more distances are useful

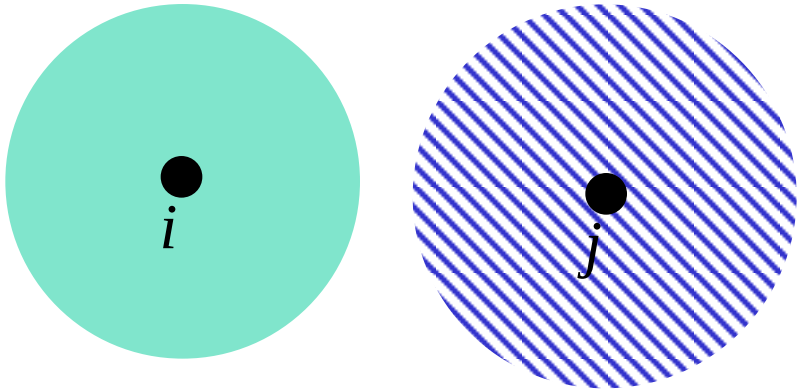




Impossible distances

No overlap ?

- experimental error
- nowhere for *k* to go



Real data

For N residue protein, maybe 5 N_{res} or 10 N_{res}

- want more like $3N_{atom}$ (30 – 60 N_{res}) distances if perfect
  - needs much more data…
    - lots of chemical data

Mission

- gather all experimental data
- mix in chemical data
- make all distance information as tight as possible
- put an upper bound on the distance between every pair of points
- put a lower bound on every distance (less important)

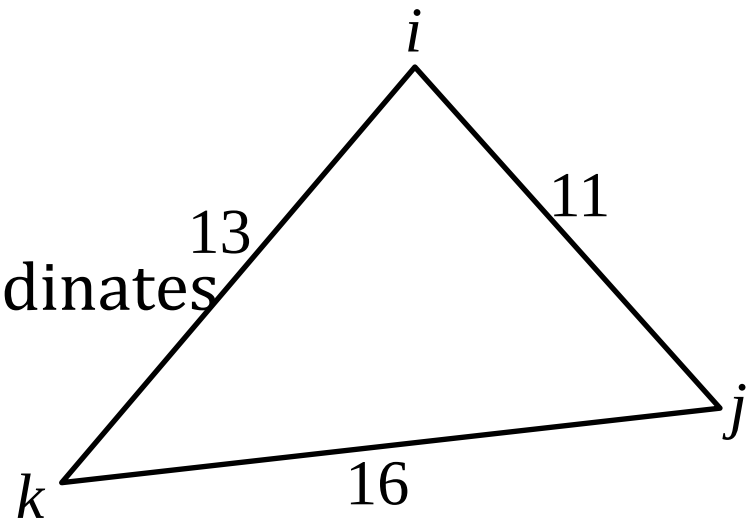
- somehow generate coordinates

- start with toys and triangles

Structures from distance information

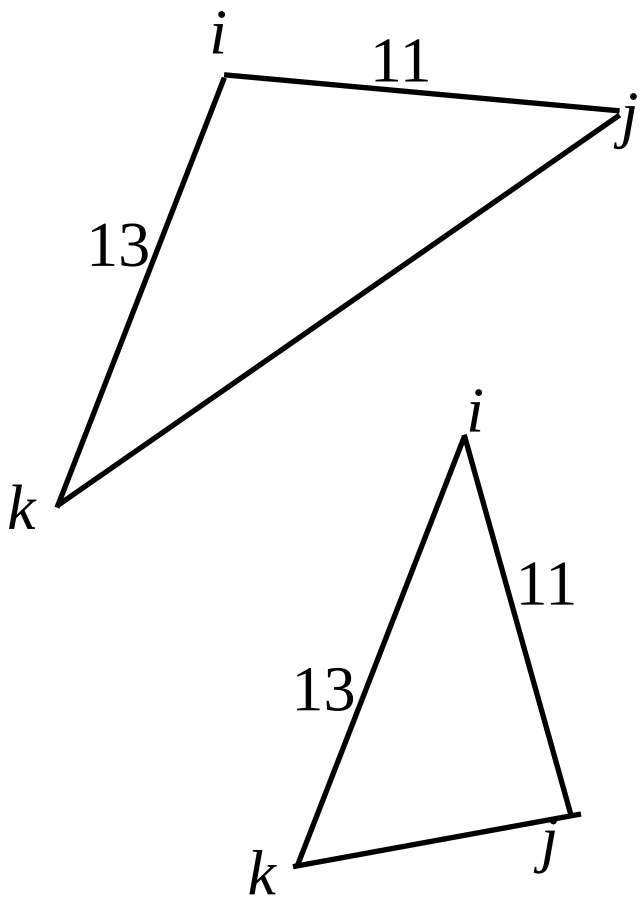
Start in two dimensions..

- ein freundliches Dreieck
  - d_{ij}=11 d_{ik}=13 d_{jk}=16
- fix *i*, put *j* on x-axis and make coordinates
- solve analytically



Underdetermined data

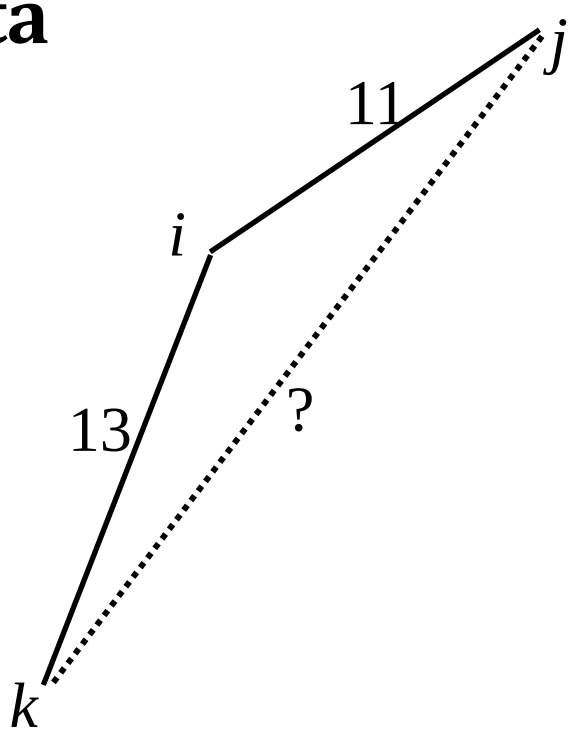
- d_{ij}=11 d_{ik}=13 d_{jk}=12 – 20
  - more like NMR data
- unique solution ?
  - no



Impossible data

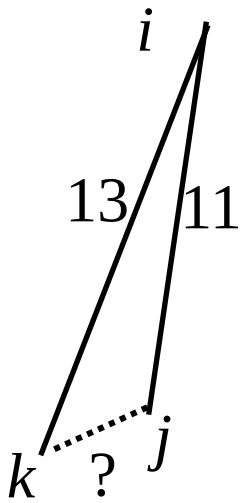
- distance too big

  d_{ij}=11 d_{ik}=13 d_{jk}=25



- distance too small
  - d_{ij}=11 d_{ik}=13 d_{jk}=1

- no 3D structure



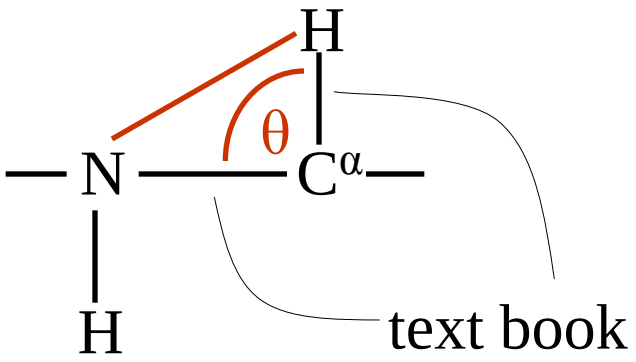
Gathering data

- add in chemistry
- use to get more
  - mix chemistry + measurements
- what comes easily from chemistry ?

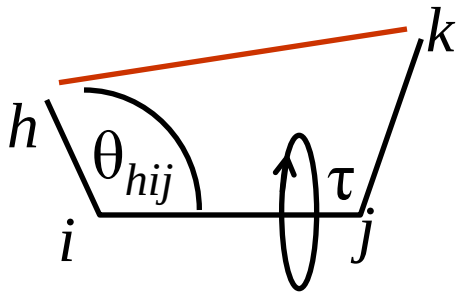
Gather as much data as possible

Simple, geometric information

- bonds – standard
- angles – standard
- simple distances from bond angles
- dihedral / torsion angles



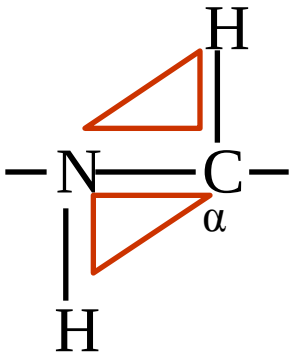
$$d_{hk}^2 = \left(d_{ij} - d_{hi}\cos\theta_{hij} - d_{jk}\cos\theta_{ijk}\right)^2 + \left(d_{hi}\sin\theta_{hij} - d_{jk}\sin\theta_{ijk}\cos\tau_{hijk}\right)^2 + \left(d_{jk}\sin\tau_{hijk}\right)^2$$

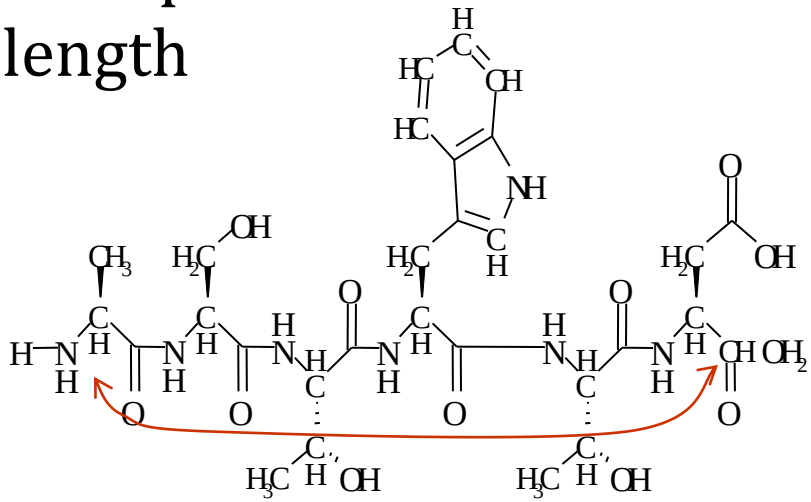


- set τ =0
  - minimum
- $\tau = \pi$
  - maximum

How to get more distance information

- impose some distance limits generally
- intuitively
  - stretch out a protein and there is a limit to length





??

- can we formalise this ?

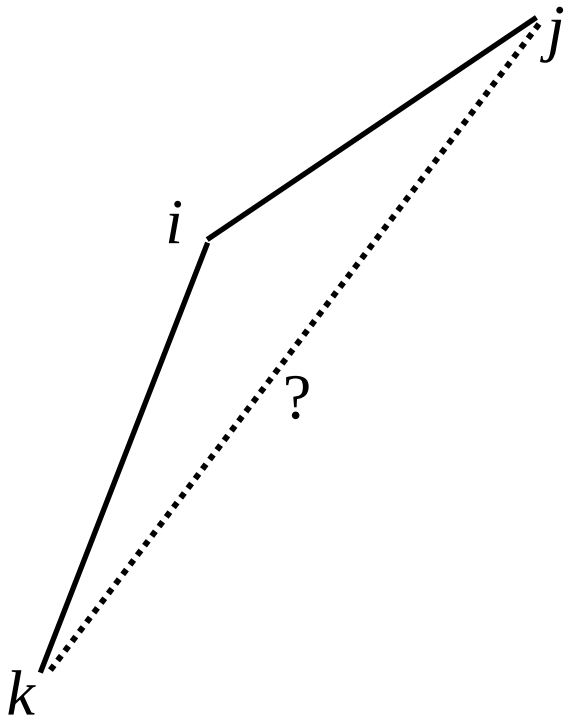
More general / triangle inequality

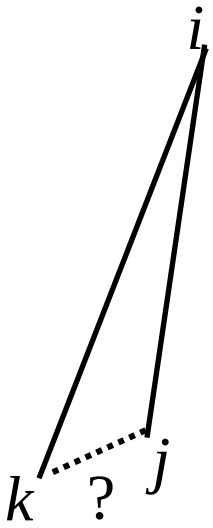
What limits can be worked out ?

- upper bound

  $d_{jk} \leq d_{ij} + d_{ik}$

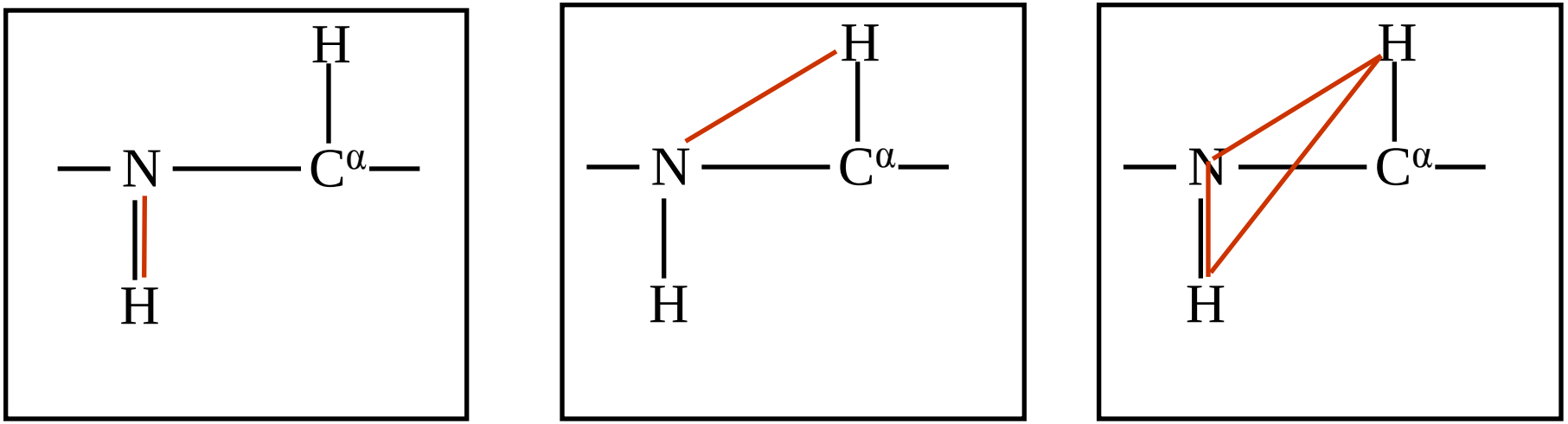
- lower bound
  - $d_{jk} \geq |d_{ij} - d_{ik}|$



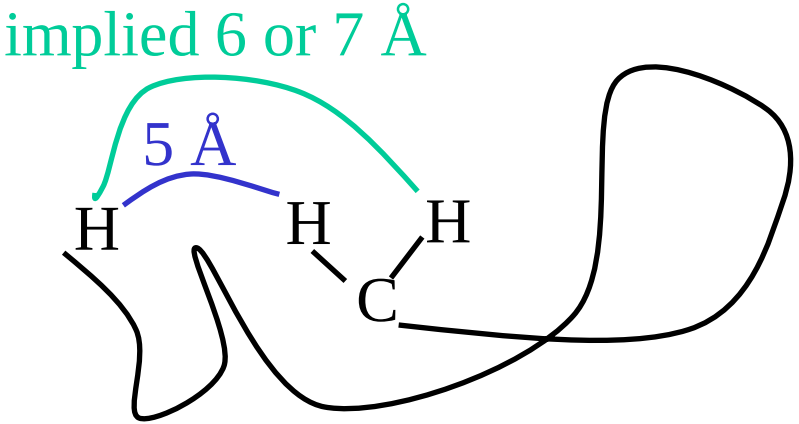


Where to use triangle inequality

- we could avoid some ugly trigonometry

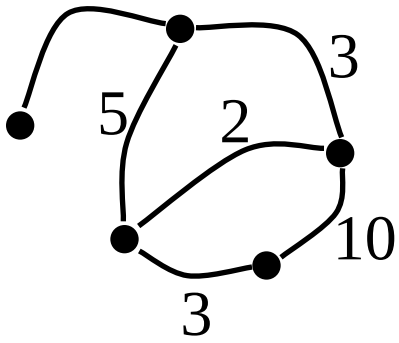


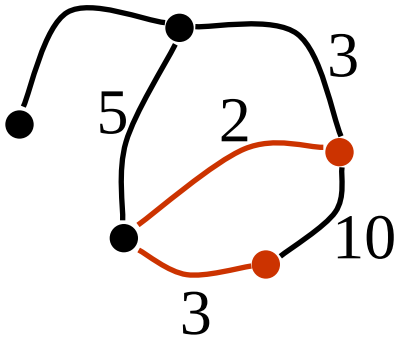
- more general



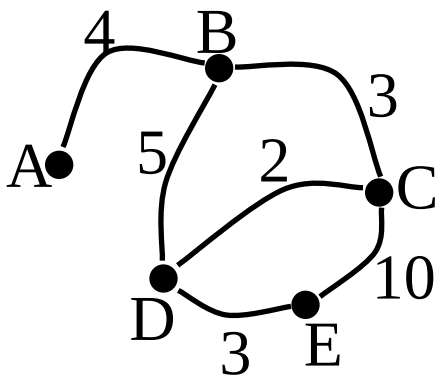
Most general triangle bound inequality

- triangle bound should be satisfied by any three points
- chemists
  - triangle bound smoothing
- informatik
  - all points shortest path problem





All points shortest path (Floyd)

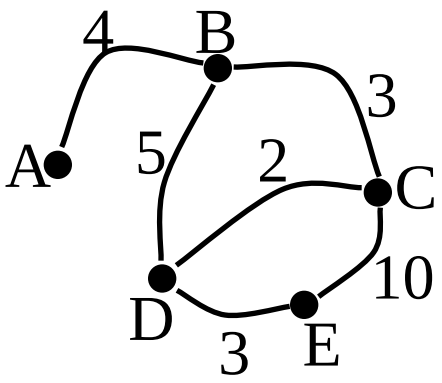


	A	B	C	D	E
A		4			
B			3	5	
C				2	10
D					3
E					

	A	B	C	D	E
A		4	max	max	max
B			3	5	max
C				2	10
D					3
E					

Bound smoothing / Floyd

	A	B	C	D	E
A		4	max	max	max
B			3	5	max
C				2	10
D					3
E					




```
for k = 0; k < n_last; k++)
        for (i = 0; i < n_last; i++)
                for (j = 0; j < n_last; j++)
                        if ij > ik + jk
                                ij := ik + jk
```

Running time

- $O(n^3)$

	A	B	C	D	E
A		4	7	9	12
B			3	5	8
C				2	5
D					3
E					

Distance matrix so far

- we can build a distance matrix of upper limits
  - consistent with all bonds and angles and other information
- can do the same for lower bounds
  - every pair of atoms
    - invent some lower bound (atomic radii)

Does this define a structure ?

- almost certainly not
  - still no way to get to a 3D model

From distances to coordinates

How would you build coordinates from distances

- stepwise ?
  - error prone, errors add
- history
  - early 80's
  - methods which are tolerant of errors
    - metric matrix method

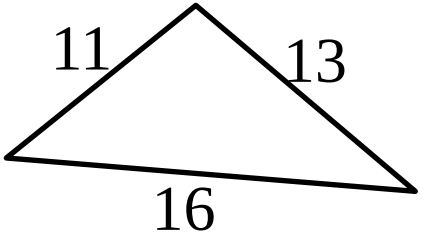
Metric matrix method

- get best upper bounds
- get best lower bounds
  - guess distances between
    → trial distance matrix
  - convert to centre of mass matrix (metric matrix)
  - magic conversion to coordinates
    - if metric matrix has three positive eigenvalues
      - error free coordinates
- real coordinates
  - lots of errors
  - initial coordinates not healthy
  - refine

Chirality

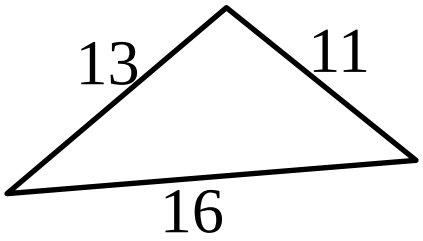
2D version

- can *not* be rotated on to each other
- can not be distinguished by distances



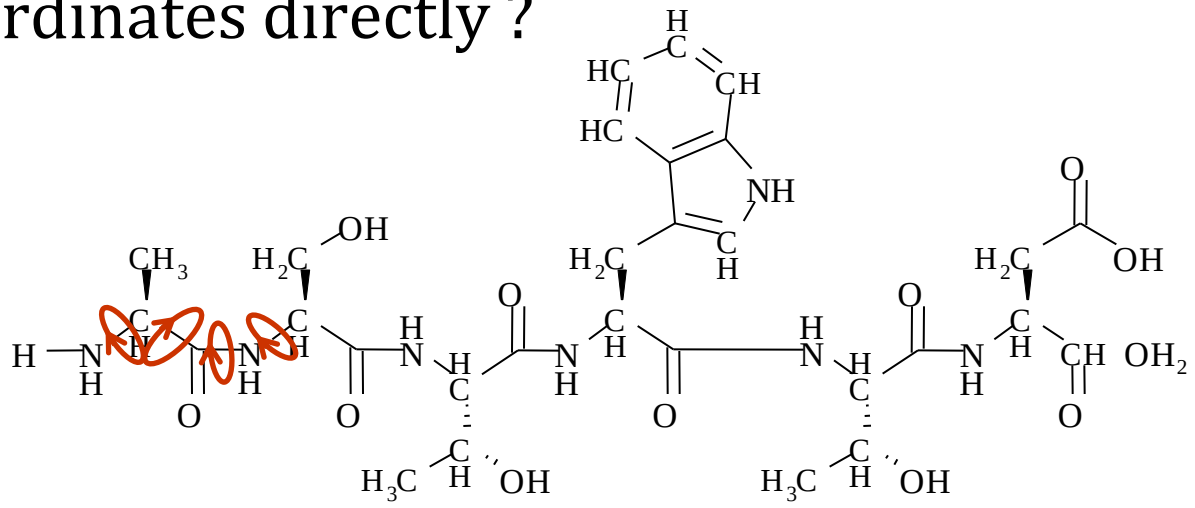
3D

- chirality is random
- problem ? no
  - flip all coordinates and check
- local chirality
  - mixture of good and bad
  - difficult to fix



Other distance geometry

Can we adjust coordinates directly ?

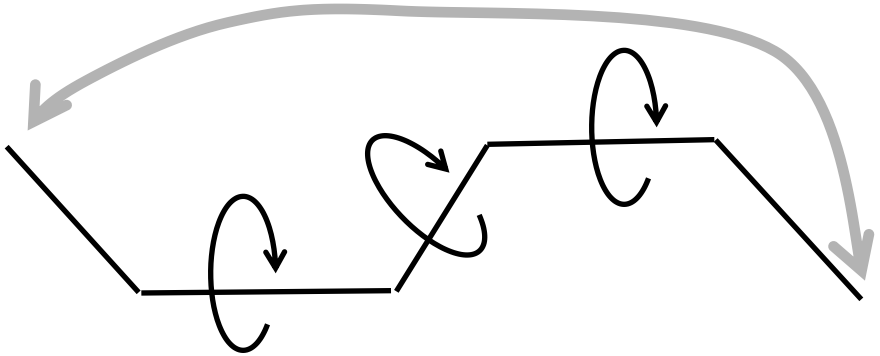


Can we work with angles ?

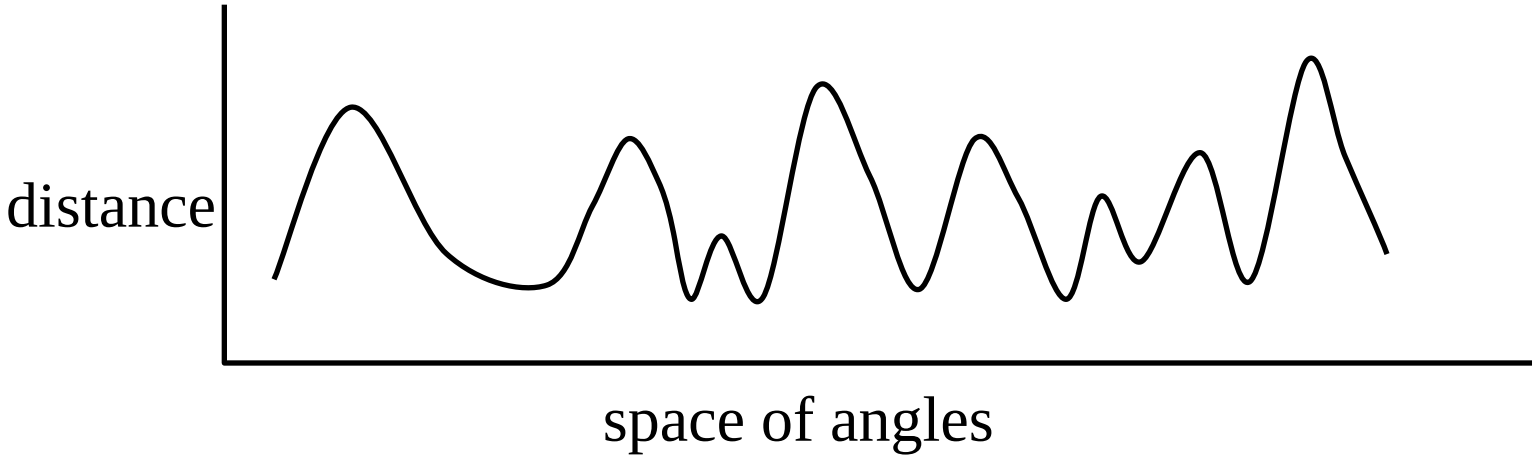
- many fewer angles than atoms

Distances and angles

- each distance may depend on many included angles

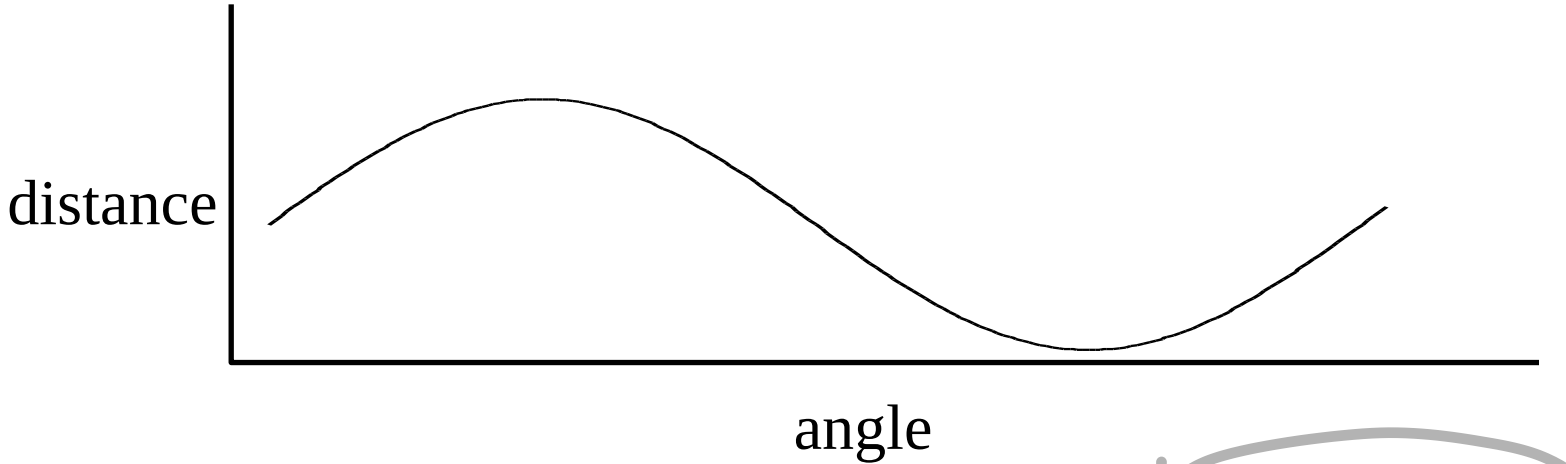


- high dimensional space of angles…

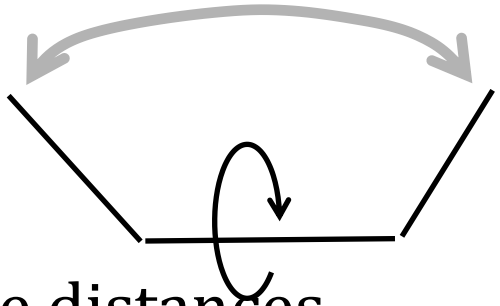


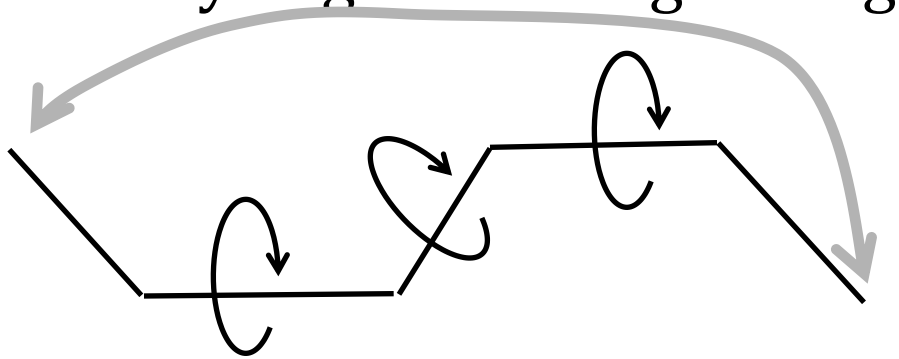
Distances and angles

- moving one angle affects one distance simply



- one angle is very important for a
  - short range distance
- just one of many angles for longer range distances





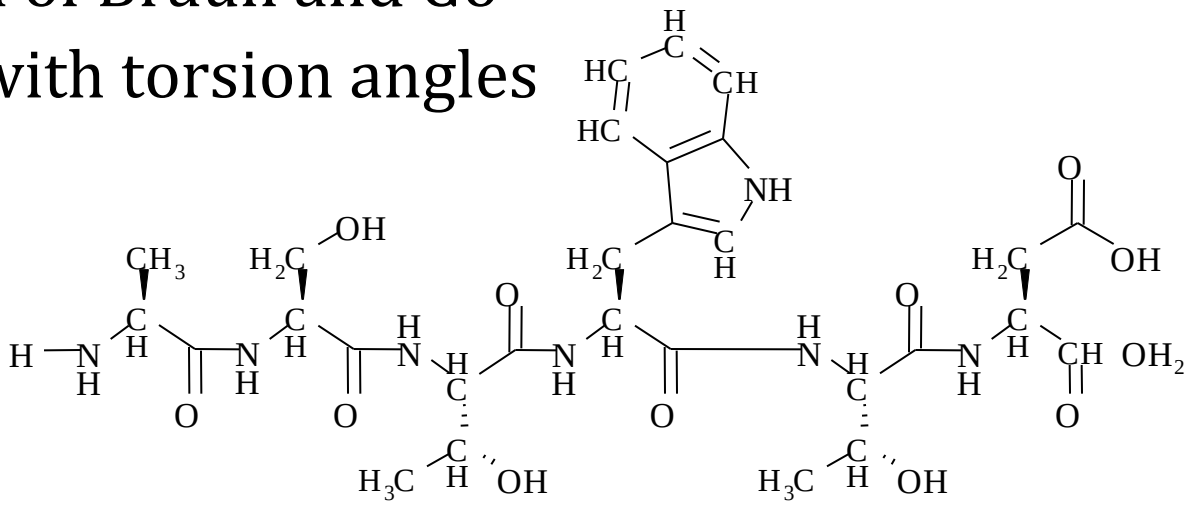
Optimisation Strategy

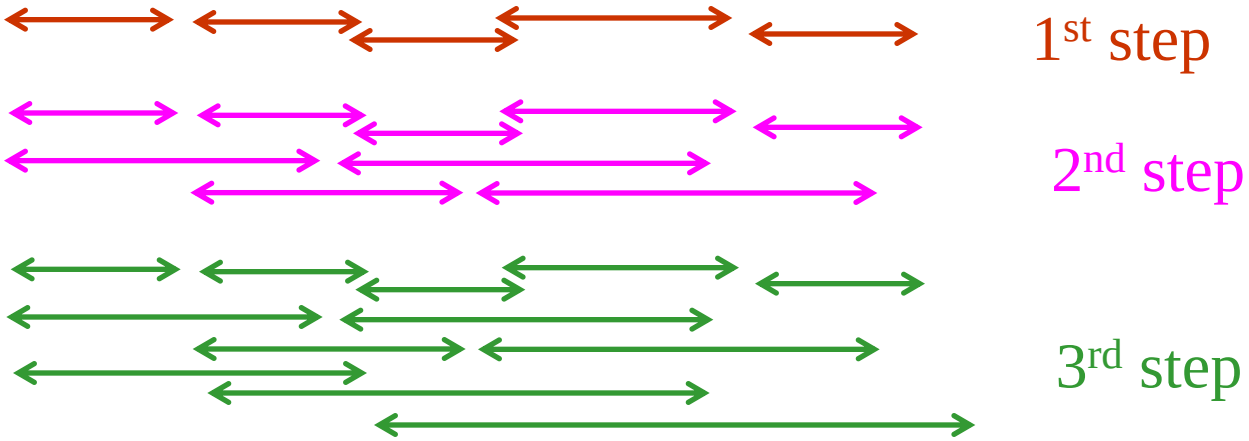
- start
  - concentrate on distances with few angles in between
  - shorter distances become correct
- add in more distances
  - re-optimise
- add in more distances
  - …

Variable target function

approach of Braun and Gō

- work with torsion angles





Stepwise variable target function method

- Collect experimental data

distance in sequence	residue 1	atom 1	residue 2	atom 2	distance in space (Å)
1	5	H^{α}	6	H^{N}	4.0
0	8	H^{α}	8	H^{γ}	4.4
80	2	H^{α}	82	H^{N}	4.5
2	3	H^{α}	5	H^{γ}	5.0
1	7	H^{β}	8	H^{γ}	3.8
0	3	H^{α}	3	H^{N}	5.0

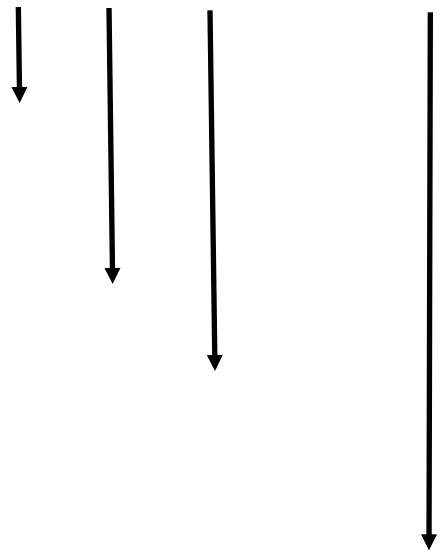
- Sort according to distance in sequence

Stepwise variable target function method

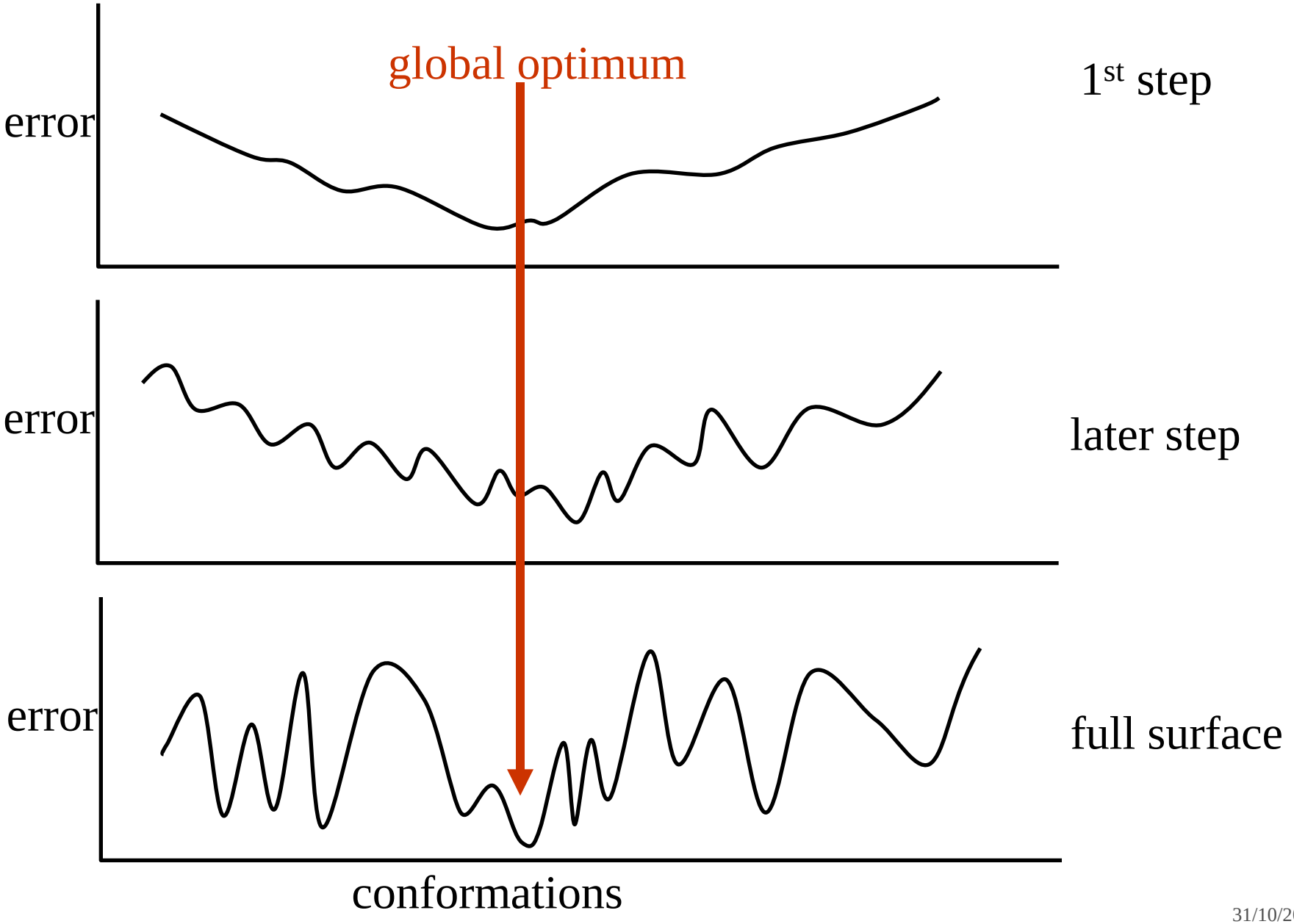
distance in sequence	residue 1	atom 1	residue 2	atom 2	distance in space (Å)
0	8	H^{α}	8	H^{γ}	4.4
0	3	H^{α}	3	H^{N}	5.0
1	5	H^{α}	6	H^{N}	4.0
1	7	H^{β}	8	H^{γ}	3.8
2	3	H^{α}	5	H^{γ}	5.0
…					
80	2	H^{α}	82	H^{N}	4.5
…	…				

Stepwise variable target function method

distance in sequence	residue 1	atom 1	residue 2	atom 2	distance in space (Å)
0	8	H^{α}	8	H^{γ}	4.4
0	3	H^{α}	3	H^{N}	5.0
1	5	H^{α}	6	H^{N}	4.0
1	7	H^{β}	8	H^{γ}	3.8
2	3	H^{α}	5	H^{γ}	5.0
…					
80	2	H^{α}	82	H^{N}	4.5
…	…				



Hope..


global optimum
error
1st step
error
later step
error
full surface
conformations

Variable target function vs metric matrix

- metric matrix vs variable target function
  - proponents of both
- variable target function probably more popular
  - no problems with chirality

Real implementations of distance geometry

- not small programs
- what kind of input would they like ?
  - list of protein sequence
  - set of distances
- most of code
  - libraries of standard amino acids
  - code to do geometry and work with standard geometries
- other information
  - angle restraints
    - convert to distances for metric matrix
    - natural for variable target function

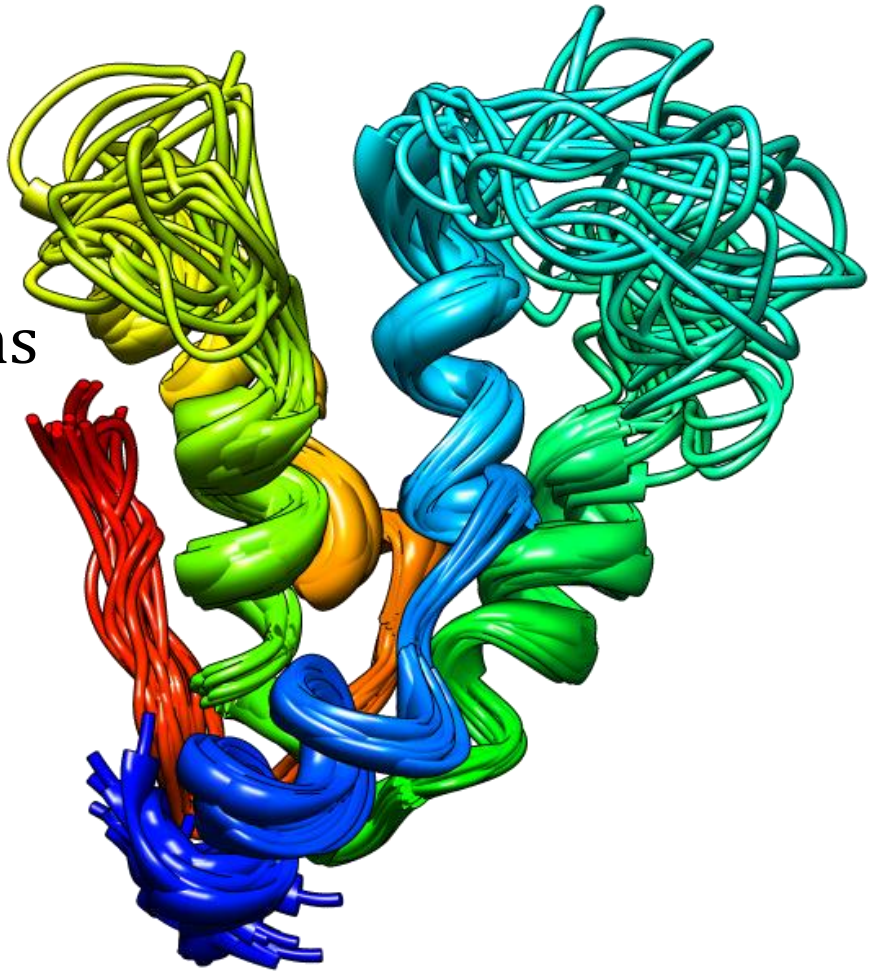
Output from programs

Structure impossible ?

- program dies or
- best possible solution

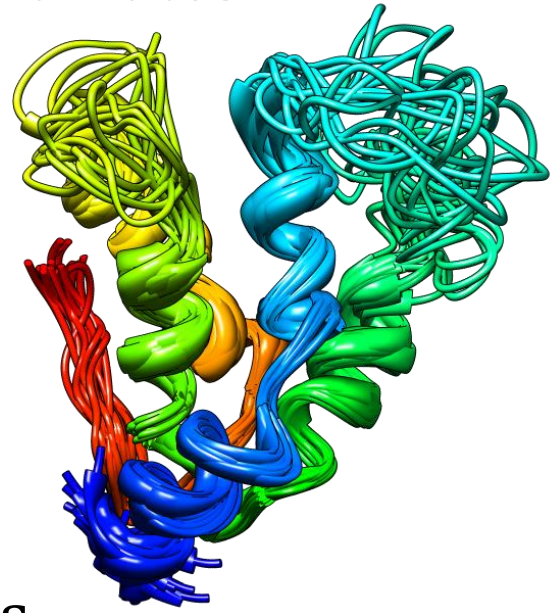
Structure not determined ?

- set of possible conformations (10 to 100)
- example 1sm7.pdb



Lots of models in a PDB file

- big difference compared to most x-ray coordinates
- typical
  - ends (C- and N-termini) badly defined
  - loops poorly defined
- spectroscopists say this reflects mobility
- problems with many models
  - difficult to work with
  - arbitrary which to select for calculations
  - averaging usually not a good idea
- Is this the absolute truth ? No.
  - number of models arbitrary
  - different methods (programs /details) give different results



Are we finished with making coordinates ?

- structures may not be well defined
- can they be improved ? probably
  - restrained molecular dynamics (more next semester)
- normal MD $E_{phys}(\vec{r}) = bonds + angles + electrostatics$ …

- restrained MD $E_{total}(\vec{r}) = E_{phys}(\vec{r}) + E_{restr}(\vec{r})$

- and… $E_{restr} = \sum_i k_i \left(r_i^{struct} - r_i^{measured}\right)^2$

- where *i* refers to the distance restraint

Mission - to minimise E_{total}

- result ?
- structures
  - agree with restraints + low energy

What else can one do with NMR ?

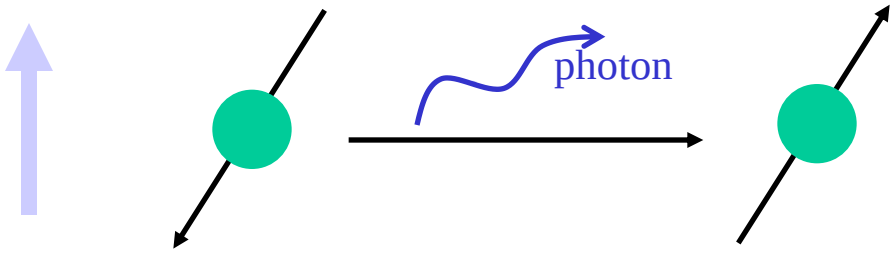
- NMR sensitive to dynamics

Timescales

- for phenomena where peaks are separated by Hz
  - timescale are Hz
- fast chemistry (small molecules drift in and out)
  - completely averaged
- very special to NMR
  - relaxation and dynamics...

What makes a nucleus relax ?

What makes a nucleus relax ?

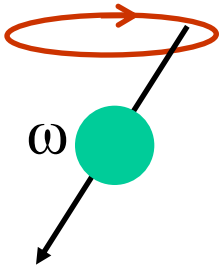


Is this really spontaneous ?

- no (think of metastable state)

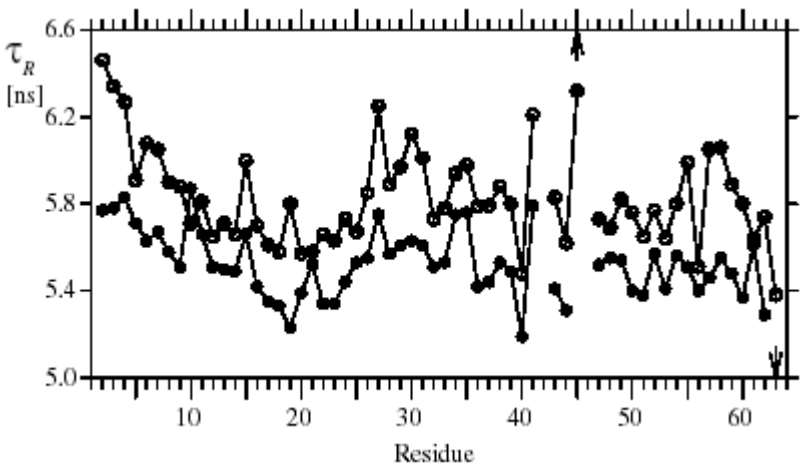
What will make it relax ?

- movement
  - overall
  - local
- Certain frequencies most important
- low frequency
- ω
- 2ω



NMR Relaxation

- different phenomena
  - NOE, T_1, T_2, …
  - different sensitivities to low frequency, ω, 2ω
    - plus, we have ω ^{13}C, ^{15}N, ^{1}H
- different sites in molecule have different motions
- define τ_c characteristic time of motion
- example, 64 residue protein
  - overall τ_c 5 ns
  - individual residues..
- do we see this in PDB files ?
  - no



Korzhnev, DM, Billeter, M., arseniev, AS, Orekhov, VY (2001), Prog. NMR. Spectr. 38, 197-266

NMR last words (almost)

- NMR good for
  - dynamics
  - deuterium exchange
  - screening / binding / ligands
- timescales

very slow	separate peaks	very different conformations
	NOE disappears poorly determined structure	
	broad peaks	solvent exchange sidechains turning
fast	averaged sharp peaks	fast side chain rotation (methyls) ligands on / off

Generating Structures Summary

- Information from NMR
  - is not complete
  - may be conflicting - methods must handle these problems
- Metric Matrix method
  - use distance information directly
  - convert ^3J (angle information) to distances
  - add chemical information (bonds, angles)
- Variable target function
  - angles and bonds are fixed - will generate good chemical geometry
  - attempts to solve an optimisation problem with a smoothing procedure (remove local minima and gradually add them)