NMR (Nuclear Magnetic Resonance Spectroscopy)

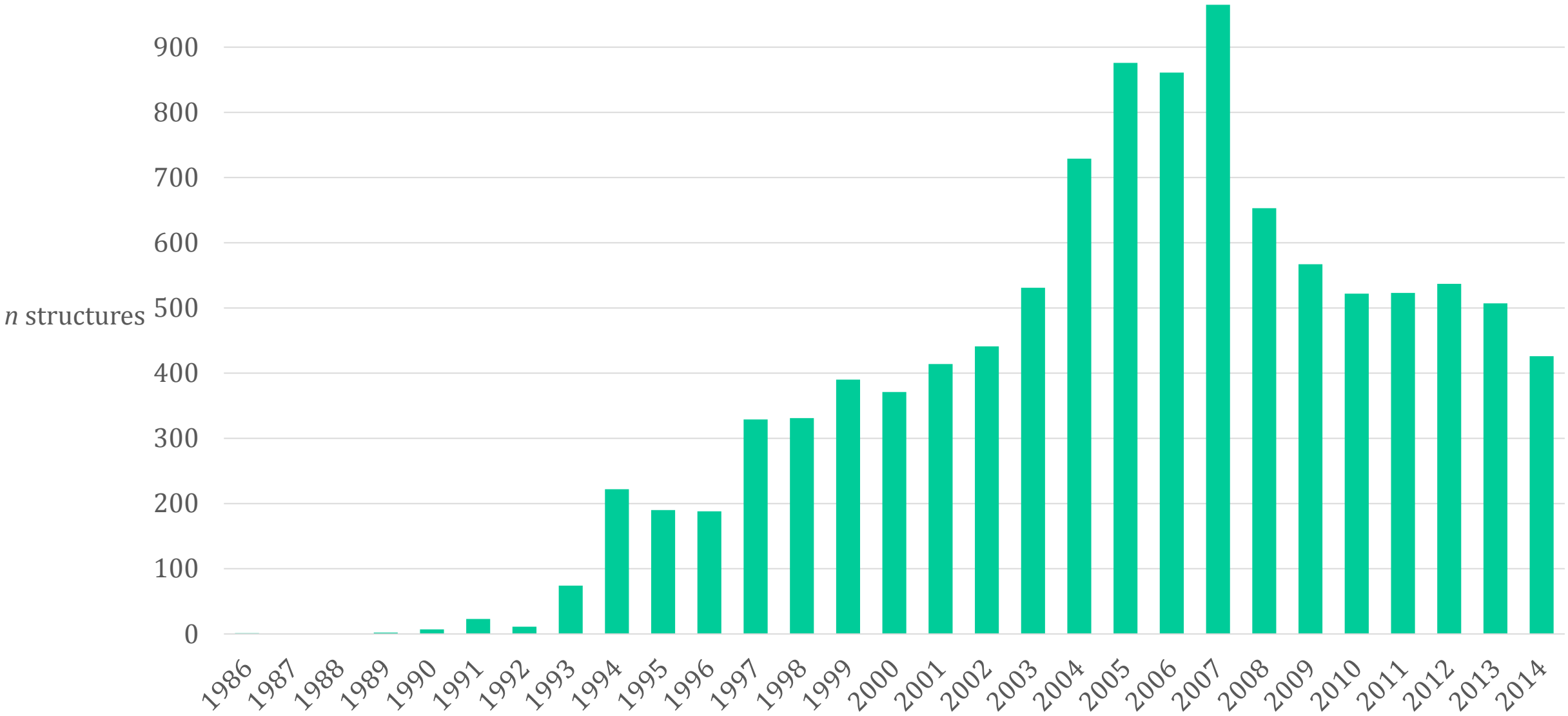
Literature / background (already in Stine)

- Ferentz, A.E. and Wagner, G., Q. Rev. Biophys, 33, 29-65 (2000) – in Stine

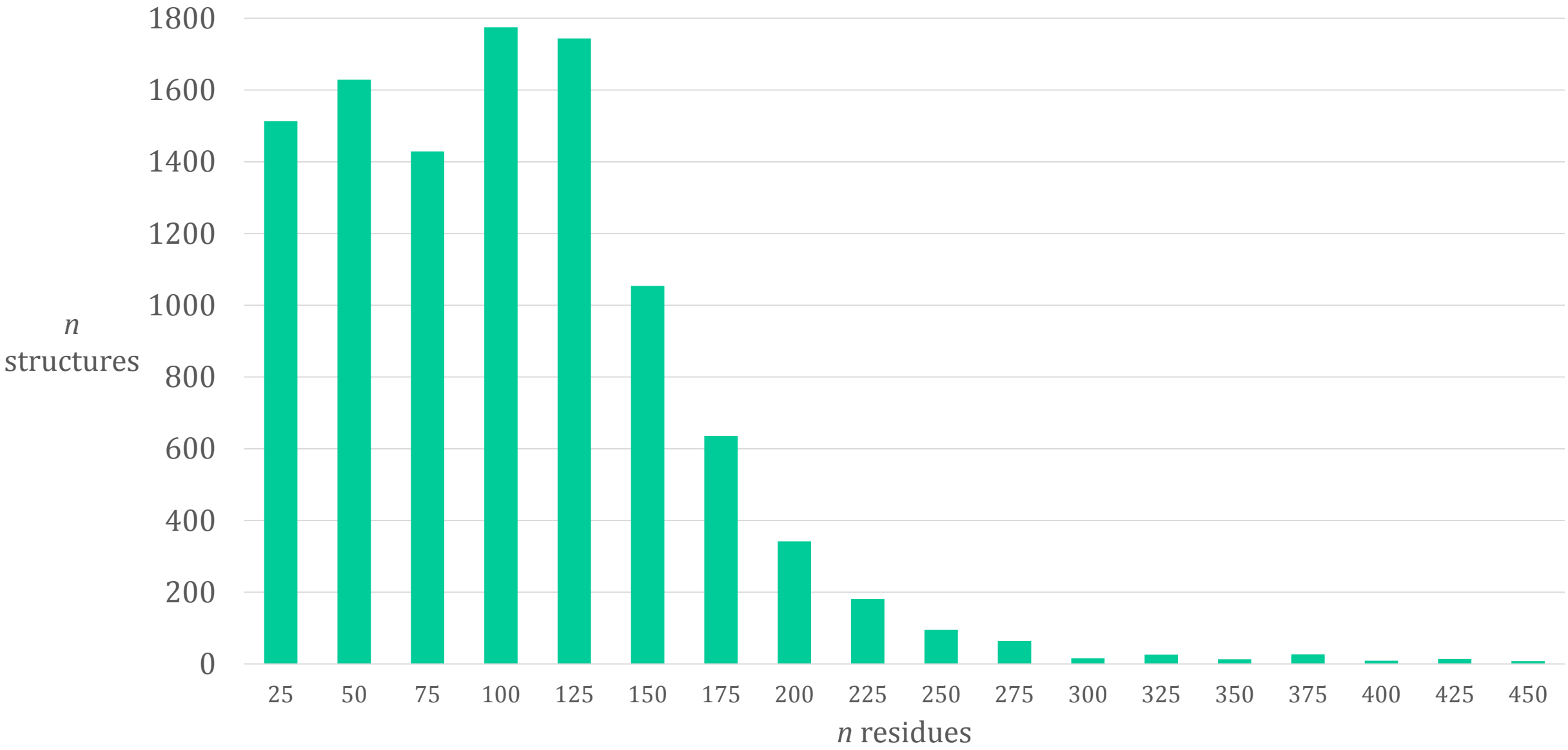
Current standing

- ≈ 11 % of current structures solved by NMR (10 618 structures, 9287 proteins)
- about 1/4 of smaller structures (<100 residues)

How many structures by NMR ?



sizes of NMR structures in protein data bank



What is coming

Background to NMR – chemistry

Calculating structures

- distance geometry
- problems with structures

For chemists: no

- chemical shifts
- 2D and higher
- residual dipole coupling, spin labels
- …

History

Younger field than X-ray

- 1 ½ Nobel prizes (Ernst, Wüthrich)

First real protein structure about 1985 or 1986

NMR from our viewpoint

A way to get structures - our focus

Can provide information on

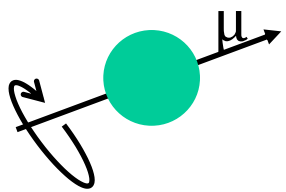
- dynamics, stability
- interactions (other proteins, small molecules)

Overview – how we get coordinates

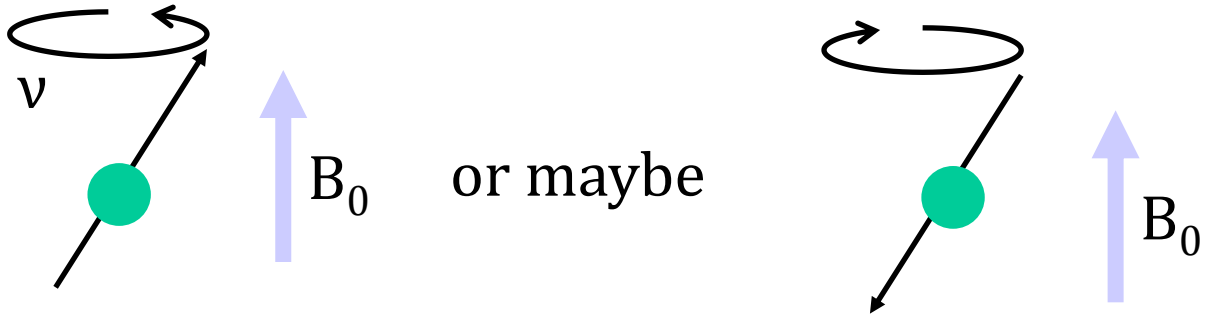
- protein in solution
- record spectra
- assign peaks to ^{1}H, ^{13}C, ^{15}N nuclei
- record some more spectra
  - distance information (mostly)
  - some internal angles
- reconstruct structure

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



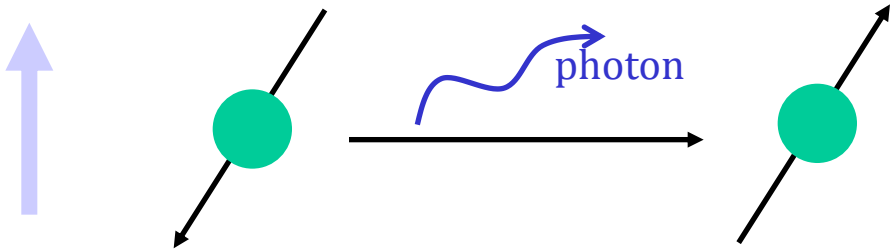
B_0 is applied field

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

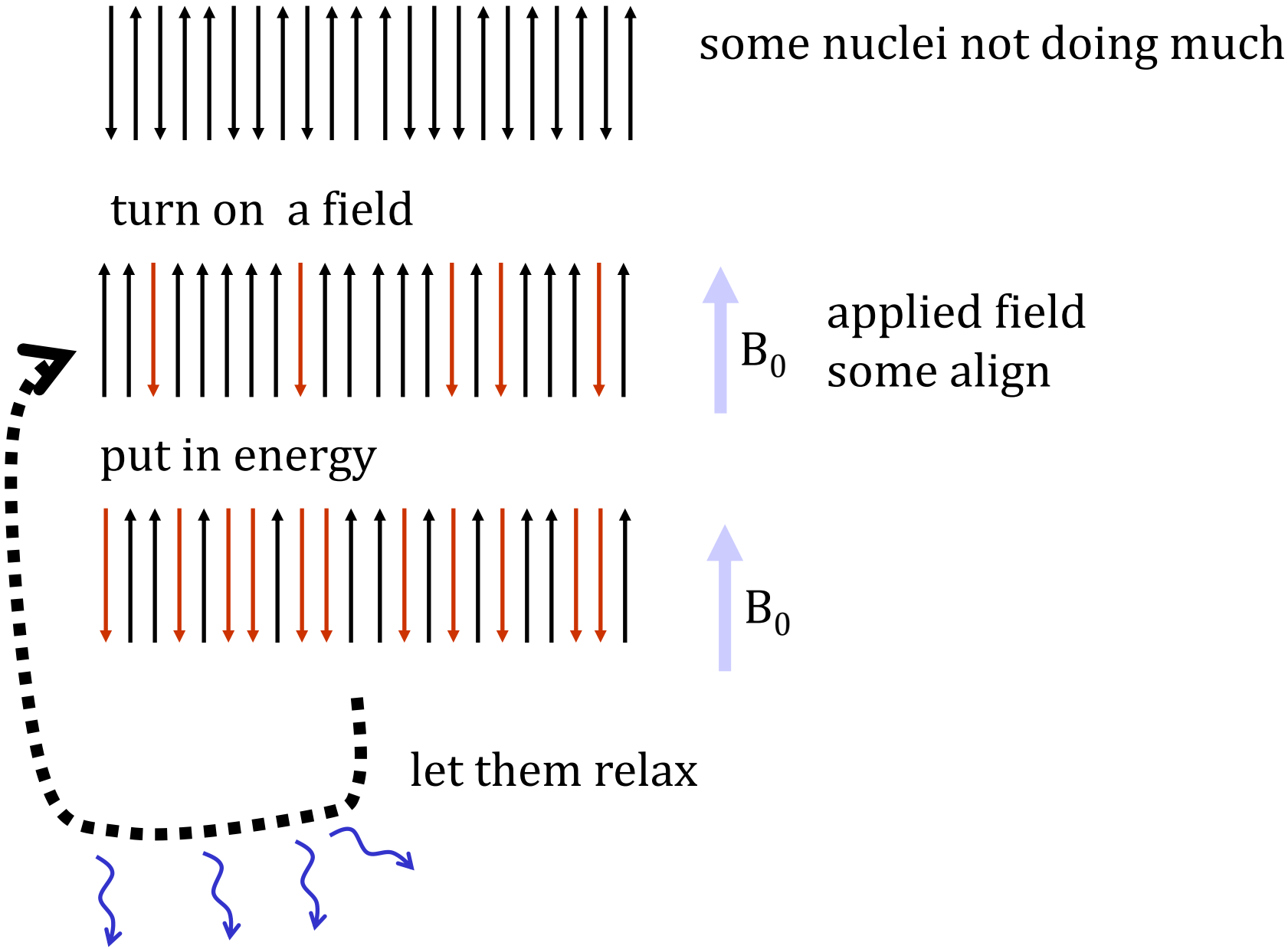
Do nuclei like fighting the field ?

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



What NMR records



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}	not cheap, 0.4 % natural abundance
^{31}P	7×10^{-2}	DNA and other PO_4 chemistry, less protein

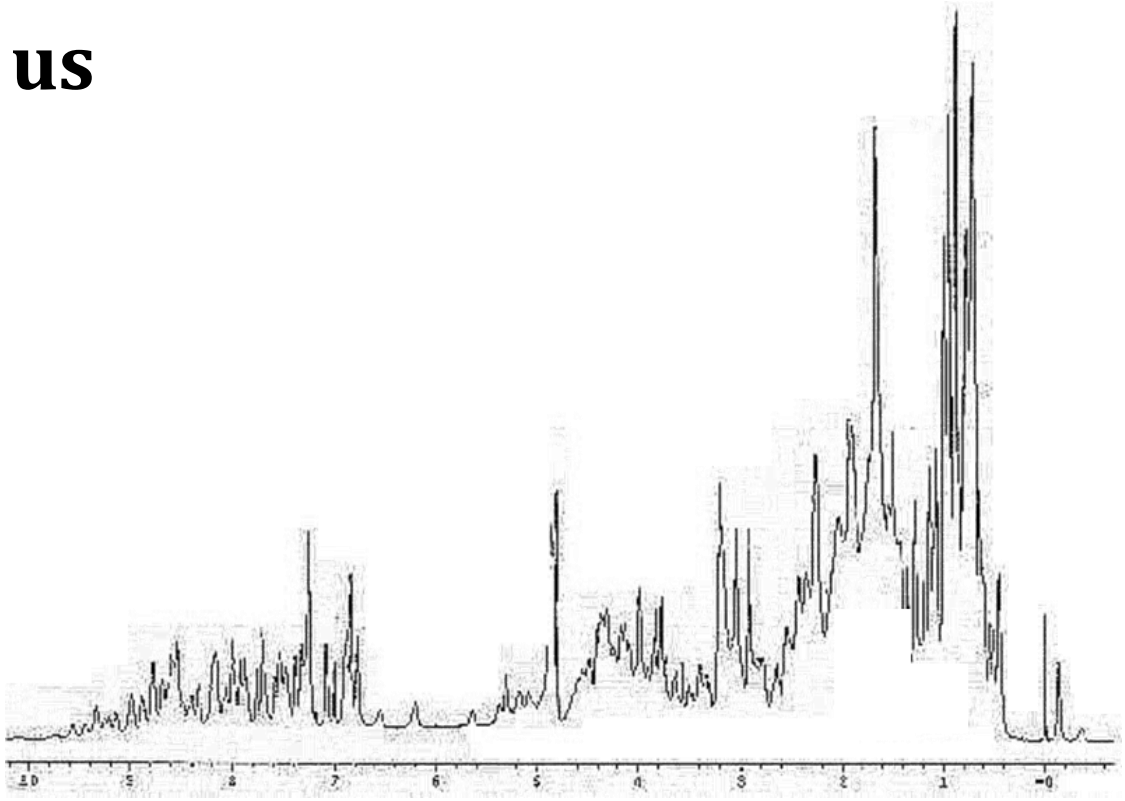
- but the natural isotopes are ^{12}C and ^{14}N
  - (usually) these isotopes require labelling
- Proteins
  - ^{1}H, ^{13}C, ^{15}N

NMR for us

- You get a spectrum (1D, 2D, ..)
- Where are the peaks ?
  - For chemists – not this course

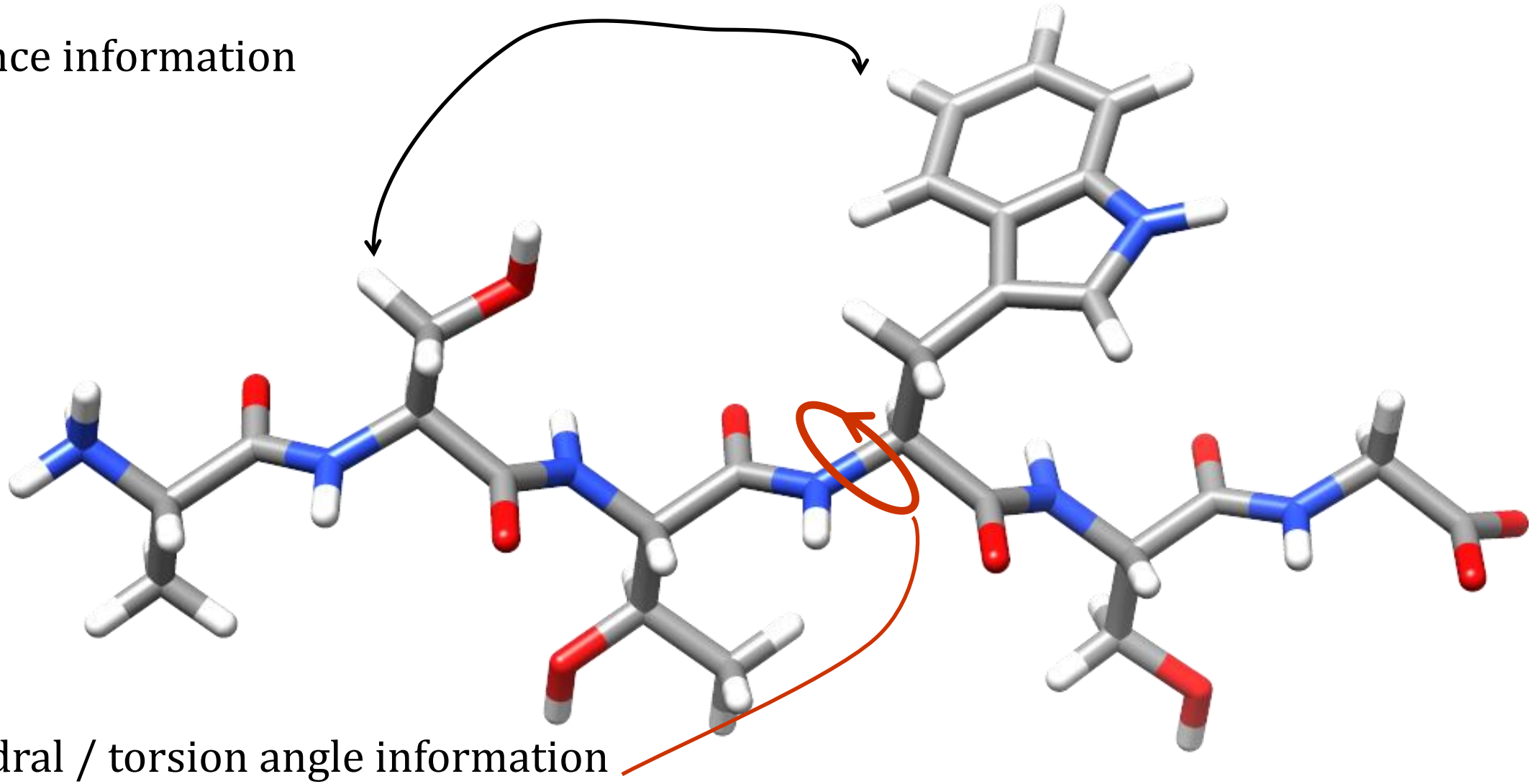
- We care about structural information
- This nucleus affects that nucleus
  - (field splitting, relaxation, …)

  - Can be related back to structure



To calculate structures ?

1. distance information



2. dihedral / torsion angle information

Distance information / the NOE

Most important (NOE = nuclear overhauser effect)

- an effect which depends on how close in space nuclei are
- NOE $\propto r^{-6}$
- usually only up to about 5 - 6 Å

Story

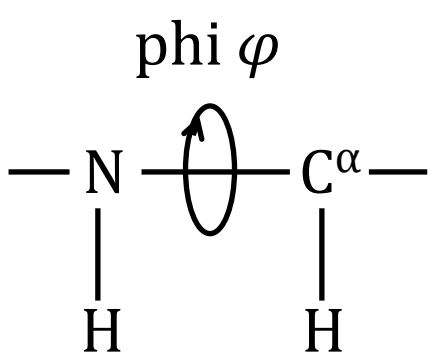
- two spins' dipoles interact
- cross relaxation phenomenon

red relaxing (jumping to lower energy)
affects black

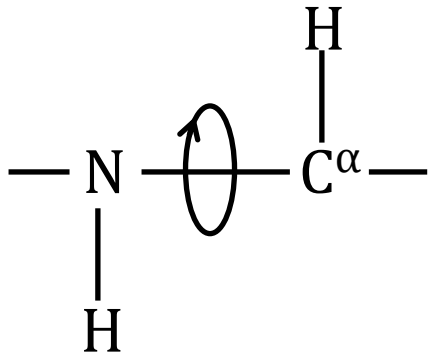
Other structural information

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

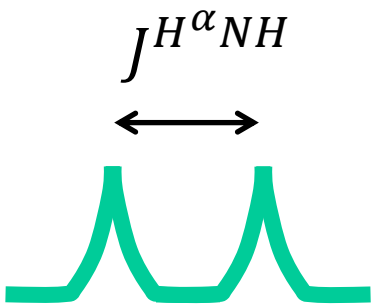
Amide NH to H^{α} coupling



cis < 6 - 7 Hz



trans ~ 10 Hz

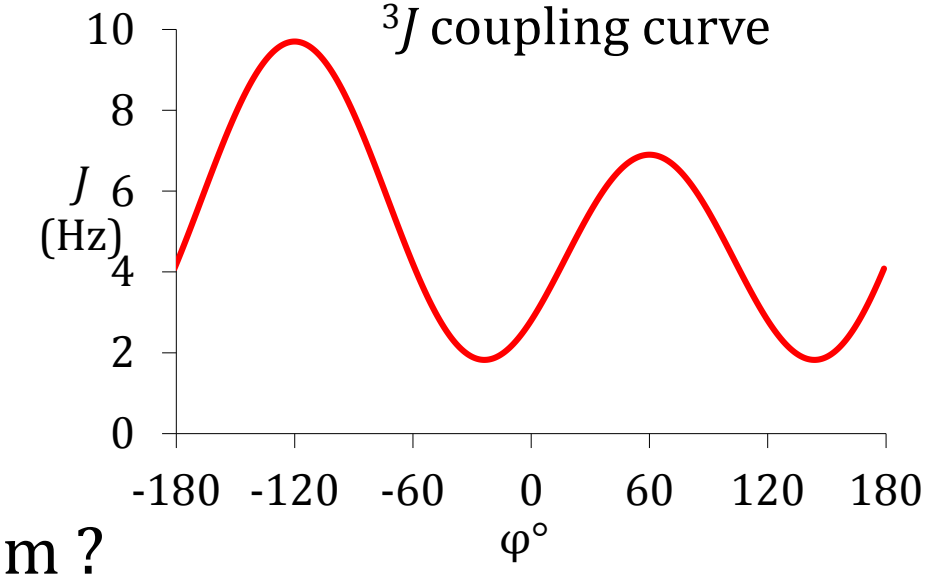


$^3J_{HN\alpha}$ coupling

formalised as

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

Problems…

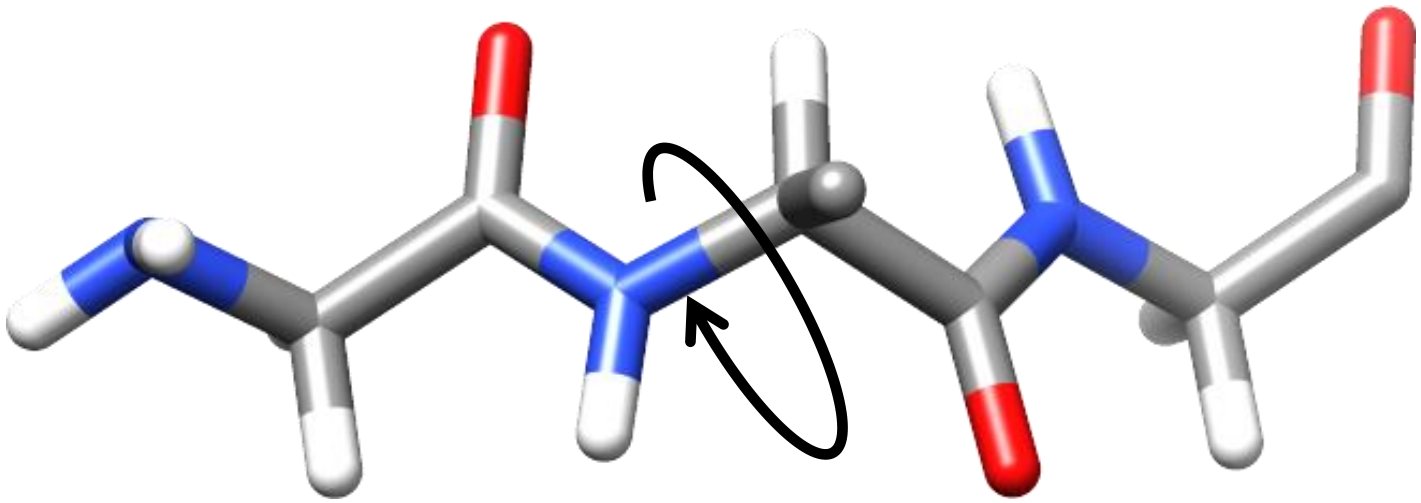


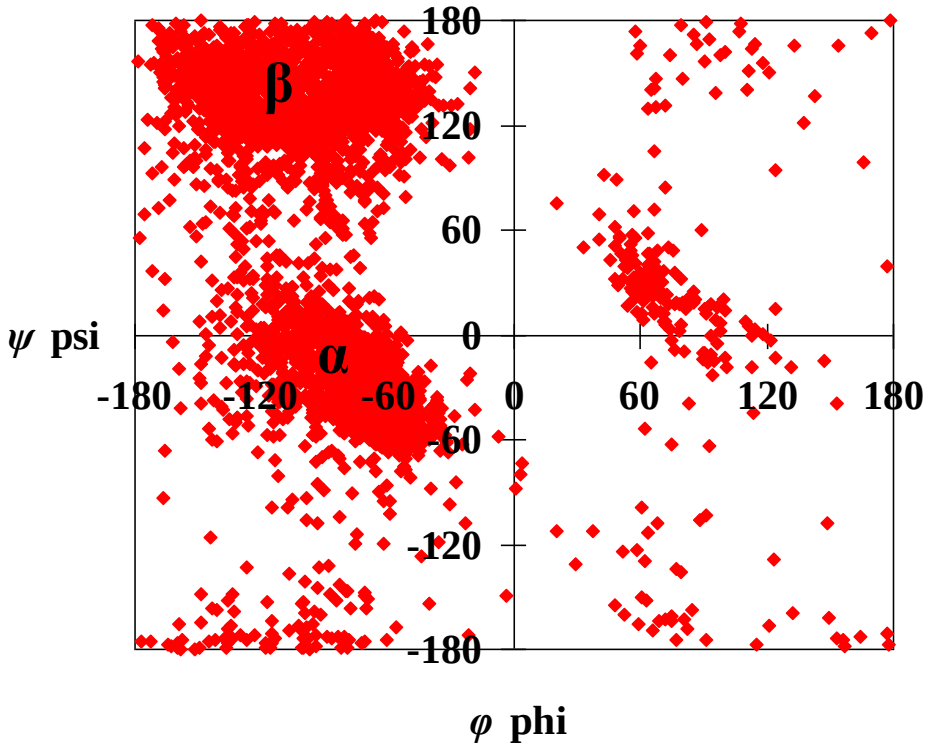
Where do 6.4, 1.4, 1.9 come from ?

Do not learn for Klausur

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}

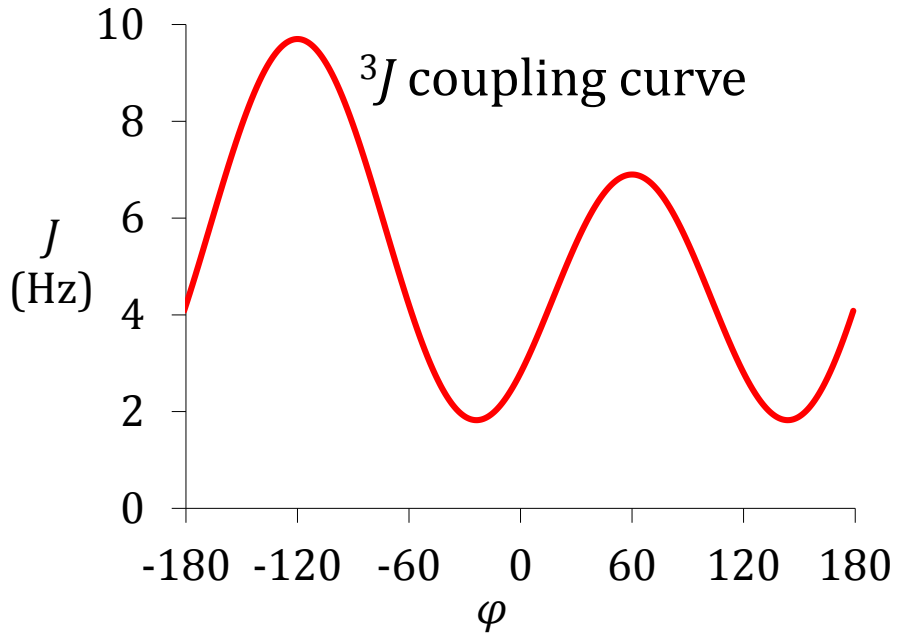
Problems with *J*-coupling

1. we have a formula

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

Measure J, solve for θ

- Most of the time, there is more than one solution
- Only use big J values



2. dynamics & errors in J measurement more serious than they appear ! look around −90°

Practical NMR

We have some basic methods

Real NMR

- more techniques
  - 2D and more
  - identifying specific kinds of atom
  - spreading peaks out

Information summary

phenomenon	assignments	structure	
chemical shift	important	not much used	not in Folien
spin-spin (*J*) coupling	important	torsion angles	
NOE	important	distances	

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

Available information

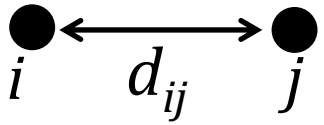
- distances
  - short (5 to 6 Å)
  - incomplete
- some dihedral / torsion angles
- does this define a structure ?
  - strictly no

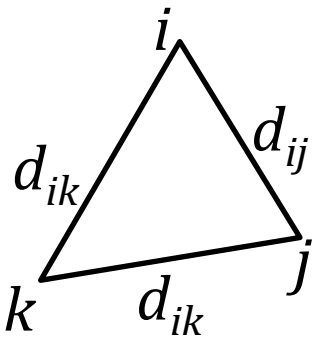
Coming

- distances in 2D and 3D
- Distance geometry – two versions

Determining distances (ideal)

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

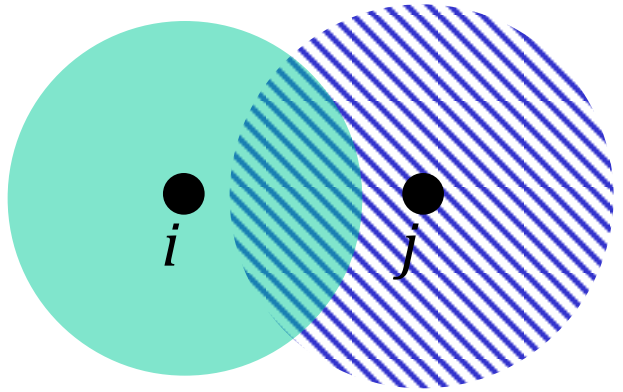




Underdetermined distances

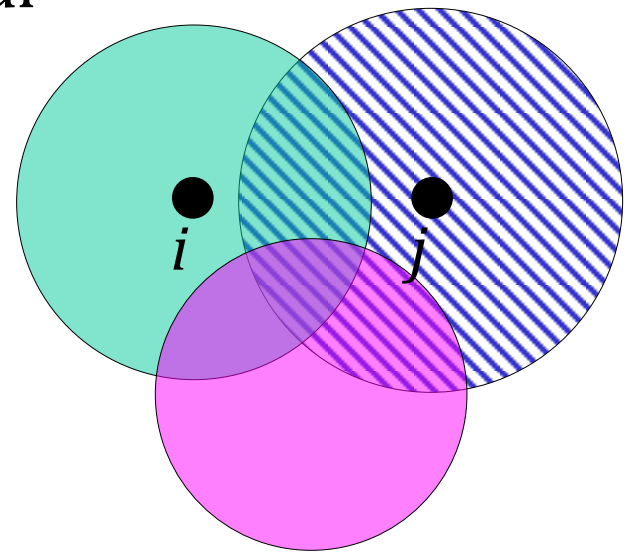
Think in terms of triangles …

- $d_{ik} < 6$ Å, $d_{jk} < 6$ Å
- where is k ?



A few more distances…

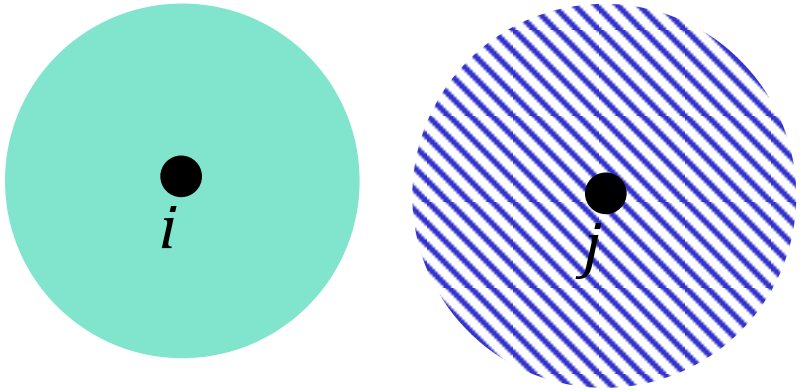
- more and more distances are useful



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)

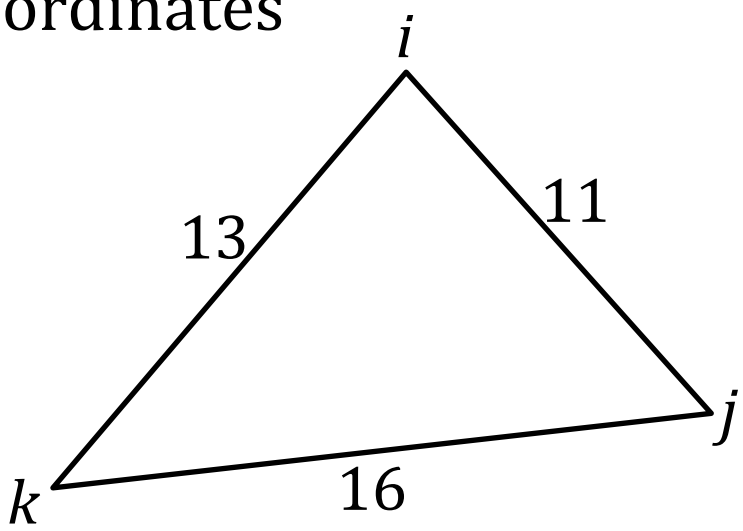
- somehow generate coordinates

- start with toys and triangles

Structures from distance information

Start in two dimensions..

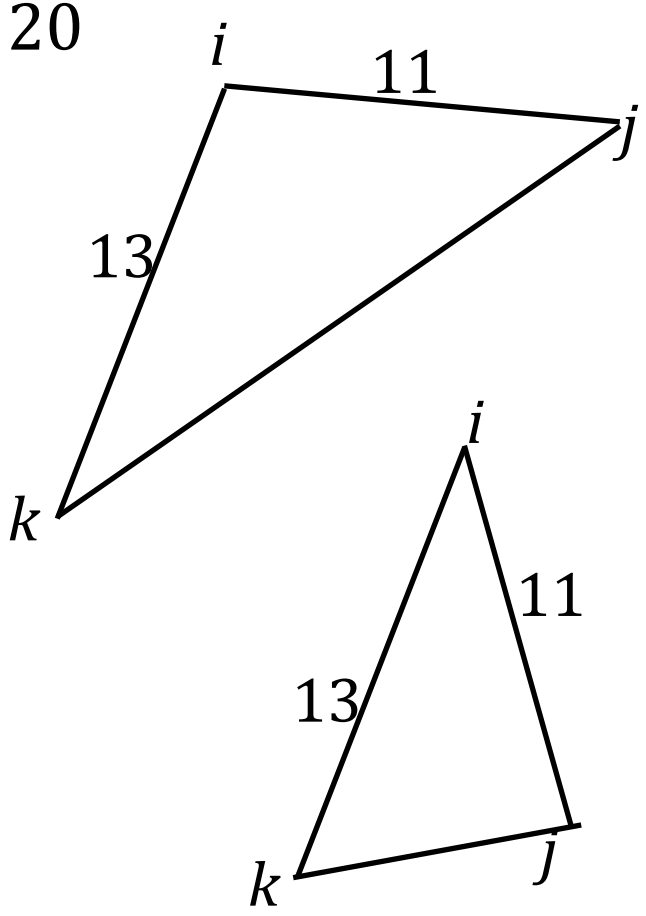
- ein freundliches Dreieck
  - $d_{ij} = 11 \quad d_{ik} = 13 \quad d_{jk} = 16$
- fix i, put j on x-axis and make coordinates
- solve analytically



Underdetermined data

- $d_{ij} = 11 \quad d_{ik} = 13 \quad d_{jk} = 12 - 20$

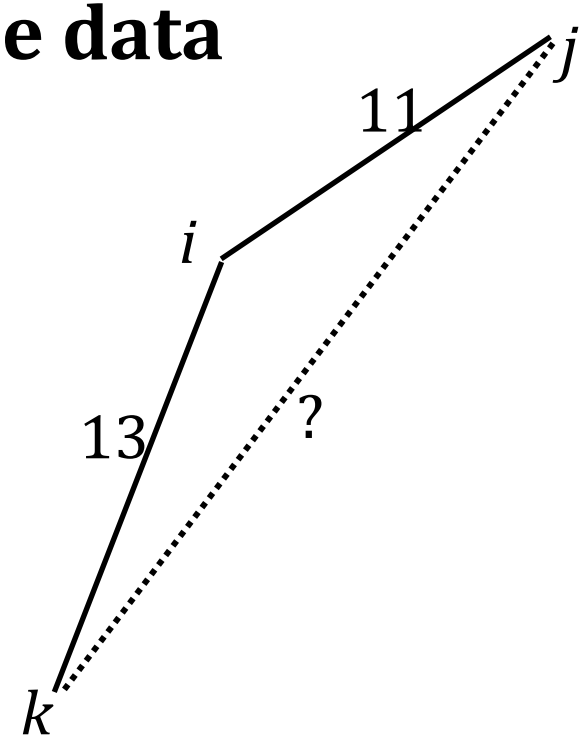
- more like NMR data
- unique solution ?
  - no



Impossible data

distance too big

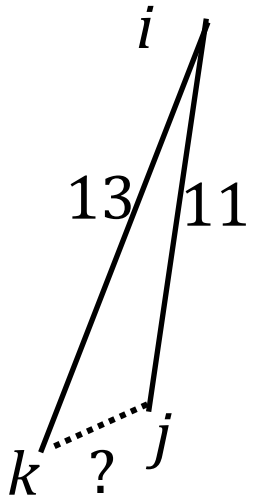
$d_{ij} = 11 \quad dik = 13 \quad djk = 25$



distance too small

$d_{ij} = 11 \quad dik = 13 \quad djk = 1$

no 3D structure



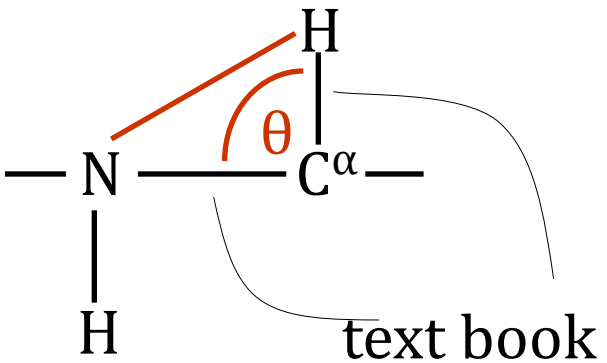
Gathering data

- 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



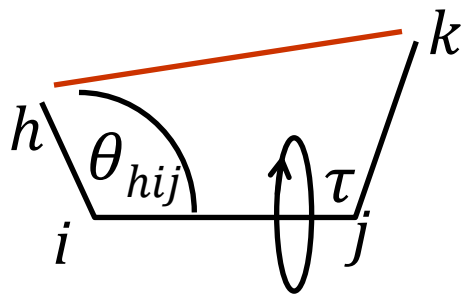
$$d_{hk}^2 = \left(d_{ij} - d_{hi}\cos\theta_{hij} - d_{jk}\cos\theta_{ijk}\right)^2 + \left(d_{hi}\sin\theta_{hij} - d_{jk}\sin\theta_{ijk}\cos\tau_{hijk}\right)^2 + \left(d_{jk}\sin\tau_{hijk}\right)^2$$

set $\tau = 0$

- minimum

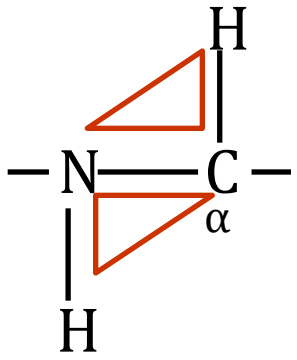
$\tau = \pi$

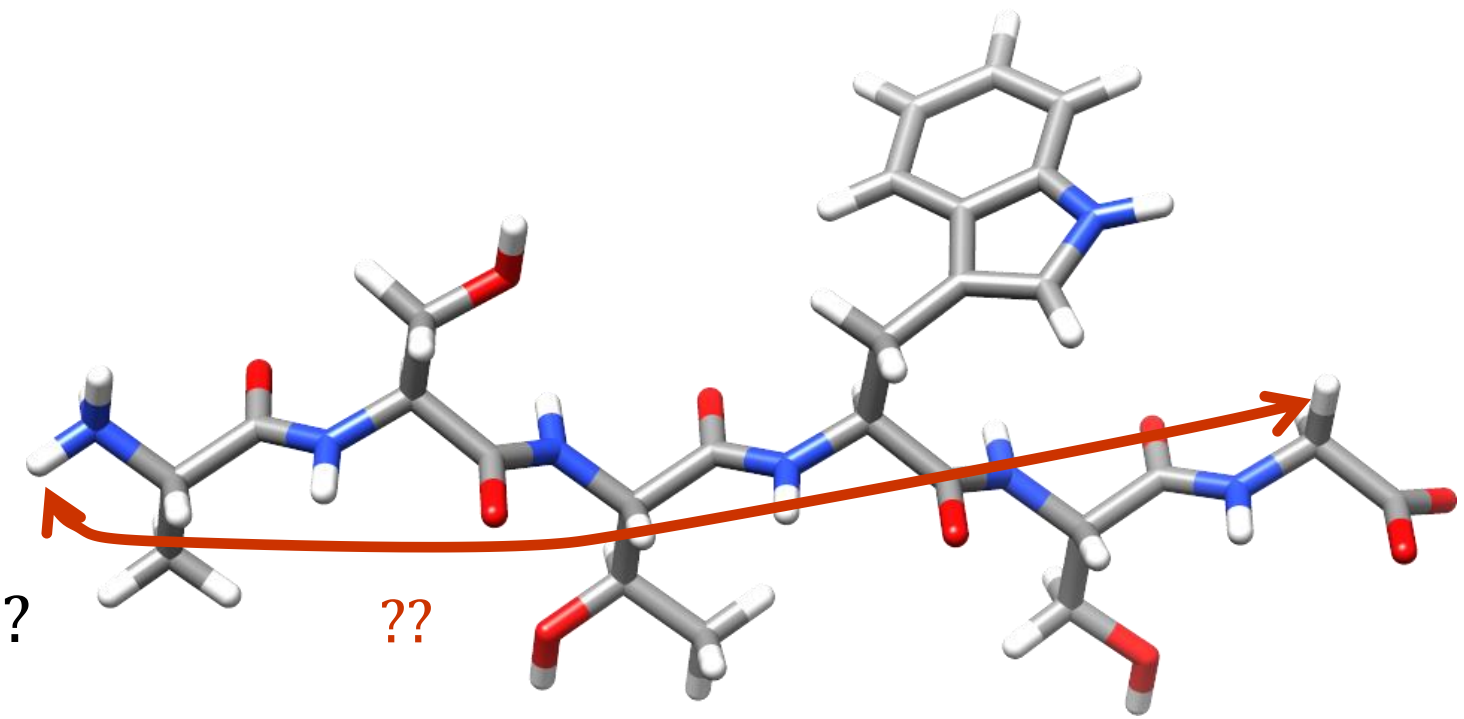
- maximum



How to get more distance information

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





Can one formalise this ?

More general / triangle inequality

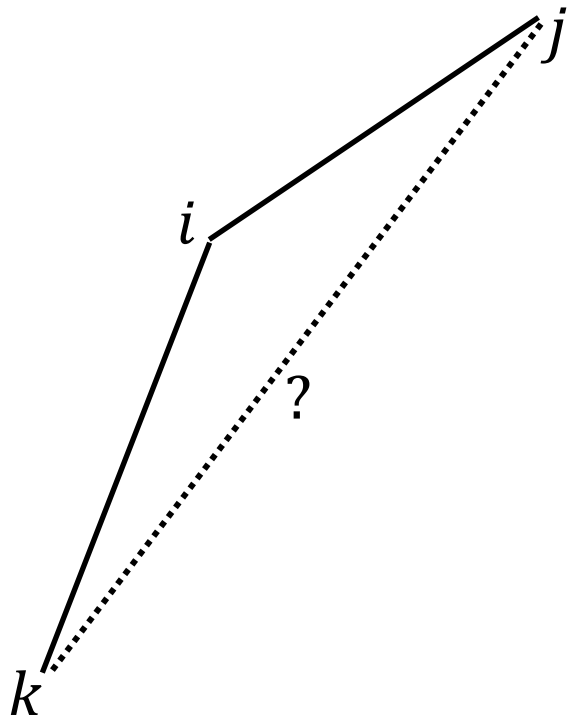
What limits can be worked out ?

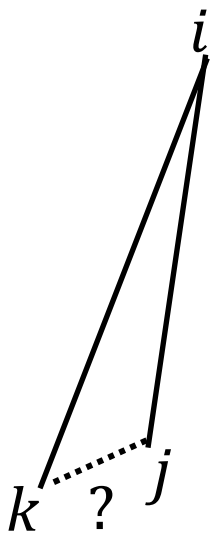
upper bound

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

lower bound

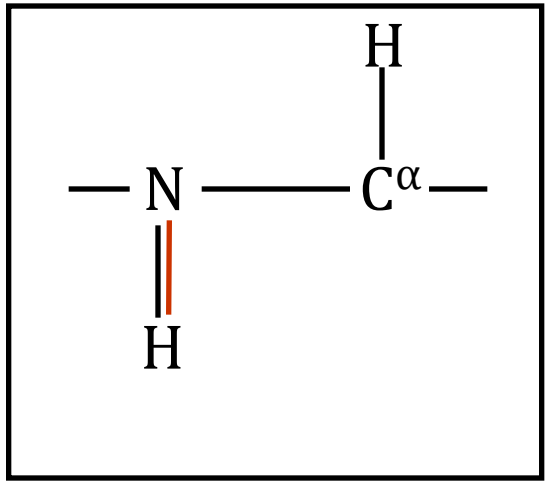
$$d_{jk} \geq |d_{ij} - d_{ik}|$$

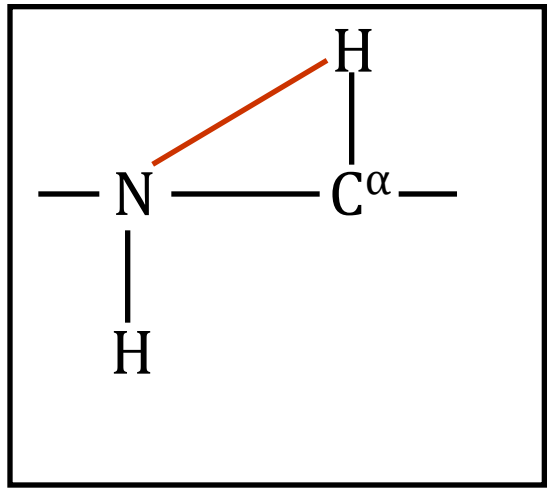


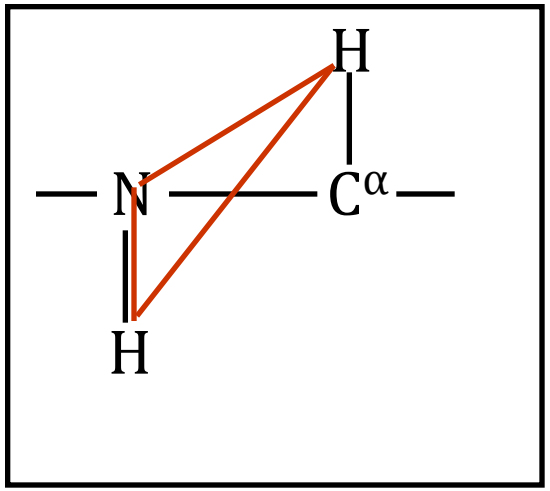


Where to use triangle inequality

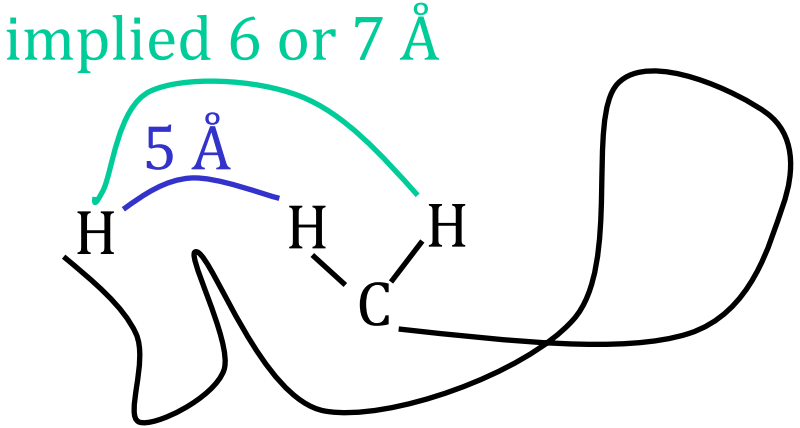
One could avoid some ugly trigonometry







more general

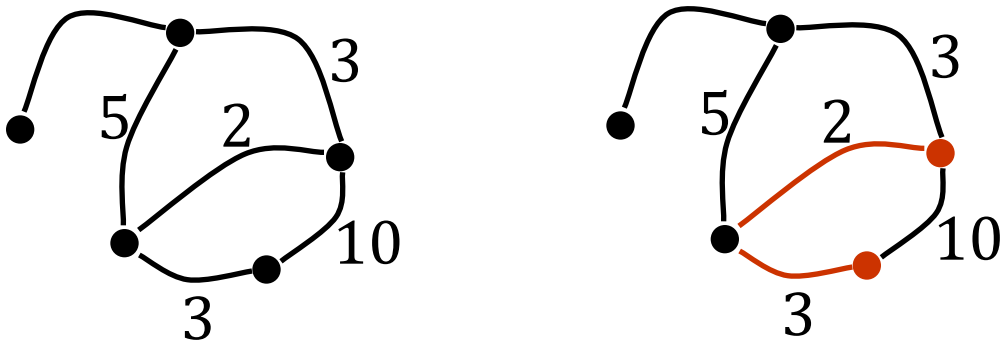


Most general triangle bound inequality

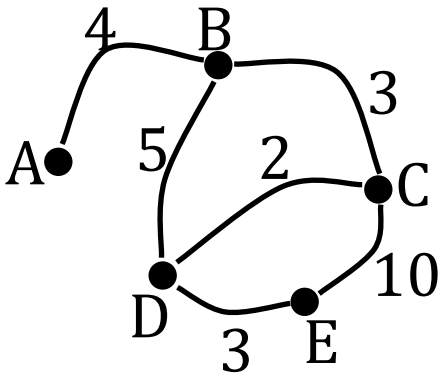
Triangle bound should be satisfied by any three points

- chemists
  - triangle bound smoothing

- informatik
  - all points shortest path problem



All points shortest path (Floyd)

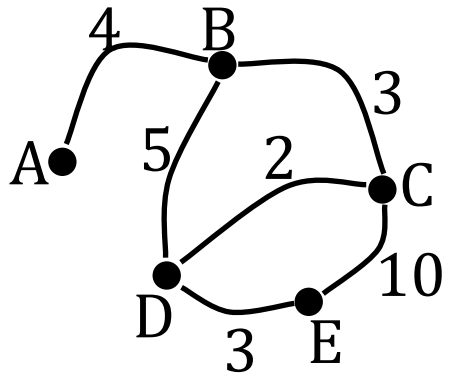


	A	B	C	D	E
A		4			
B			3	5	
C				2	10
D					3
E					

	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

Metric matrix method

- get best lower bounds + upper bounds
  - guess distances between

    → trial distance matrix

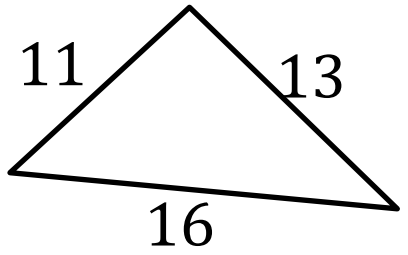
- repeat *n* times
  - get *n* guesses
- some OK, some bad
- repeat until you have 20 or 100 OK structures

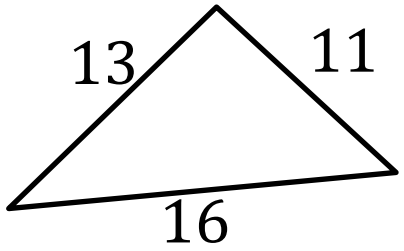
- OK = agrees with experimental data + chemically OK

Chirality

2D version

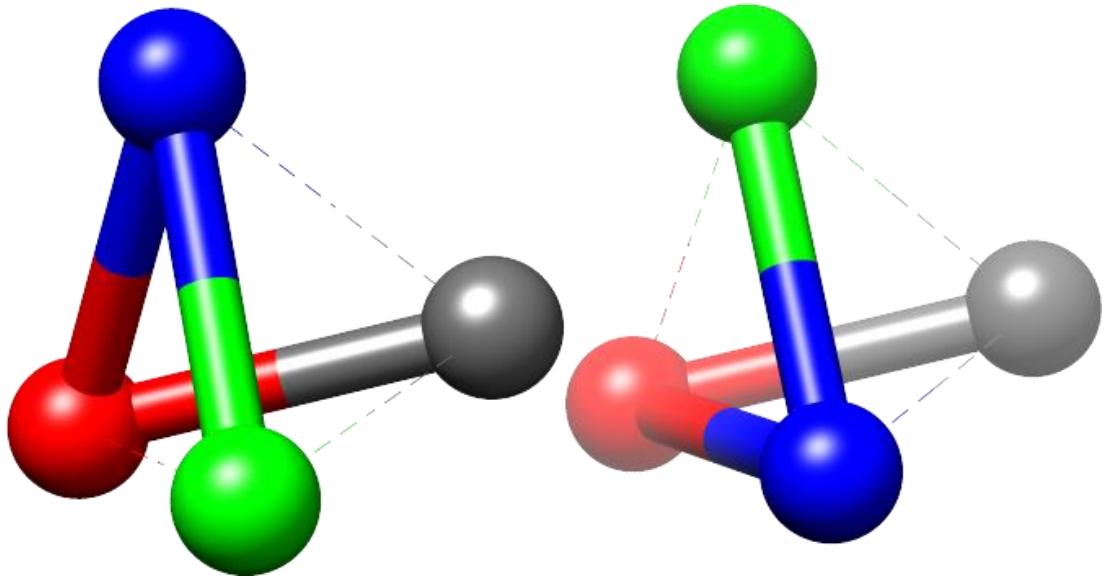
- can *not* be rotated on to each other
- can not be distinguished by distances





3D

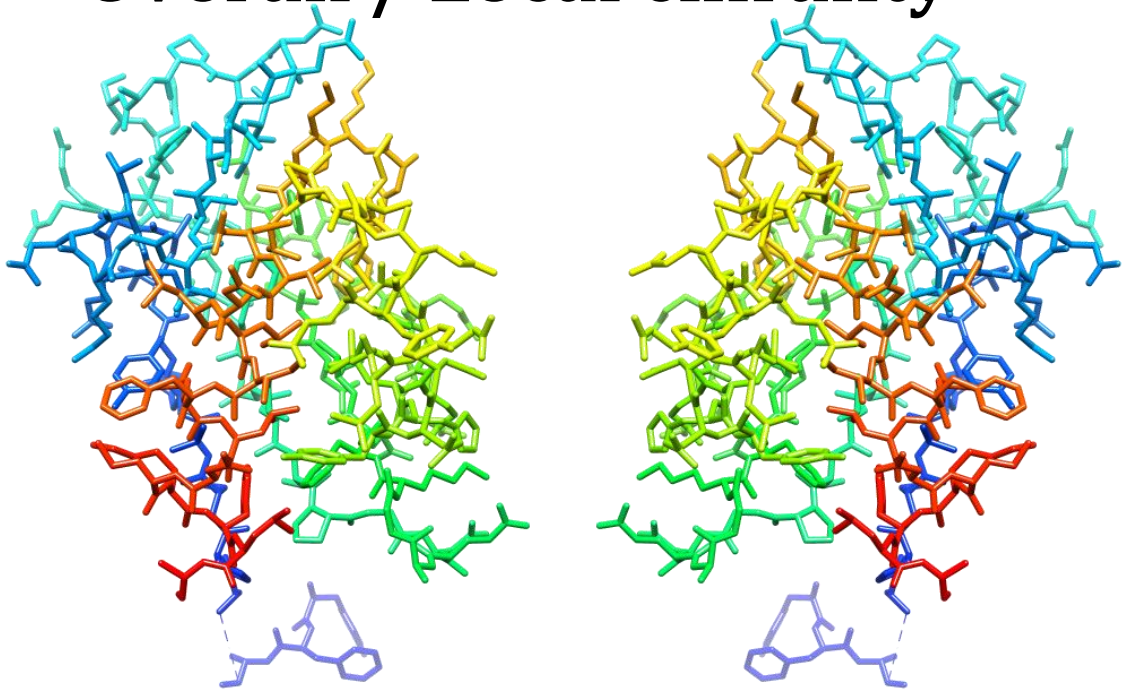
- chirality is random
- problem ? no
  - flip all coordinates and check



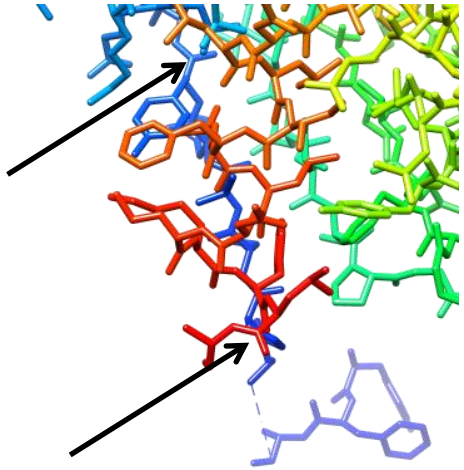
Local chirality …

Overall / Local chirality

overall chirality



local chirality



- some points correct
- some wrong
- If you invert a site, will damage other parts of structure

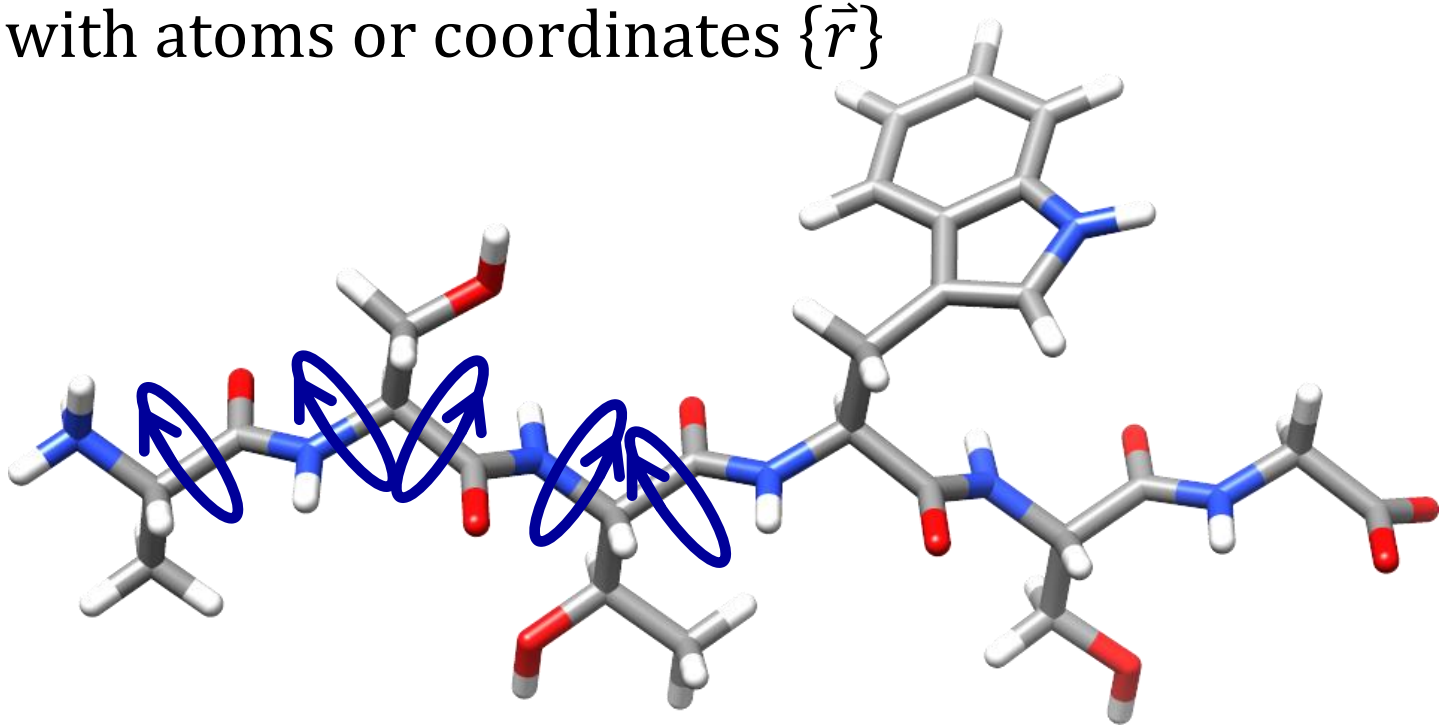
The Optimisation problem

Find the coordinates that put atoms so they agree with experimental data

- cost c is $\sum_i \left(r_i - r_i^{measured}\right)^2$ for each measured distance r

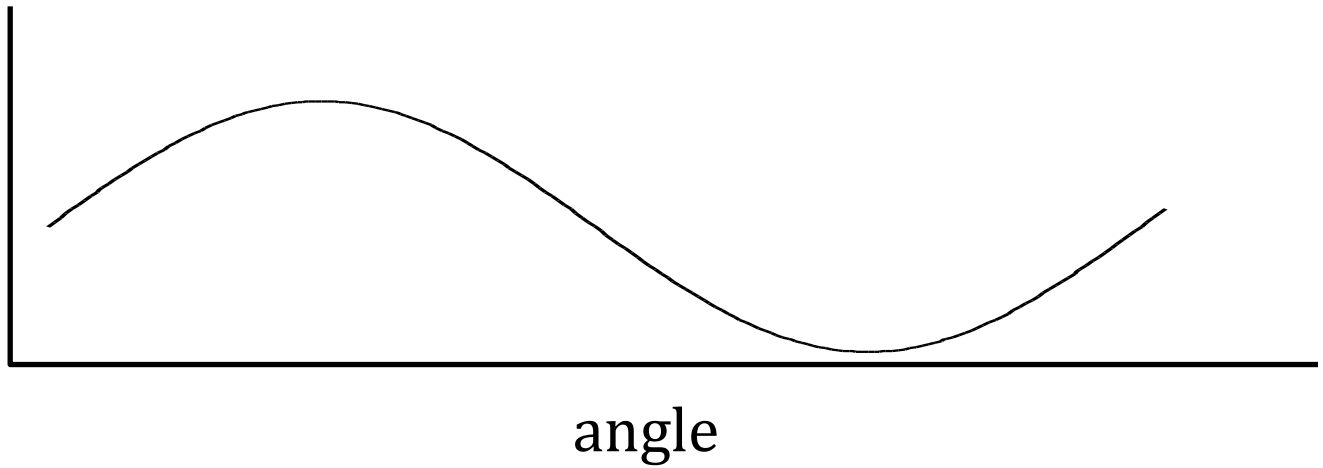
Maybe we do not work directly with atoms or coordinates $\{\vec{r}\}$

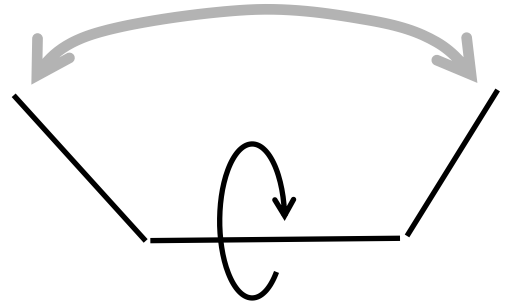
work with angles



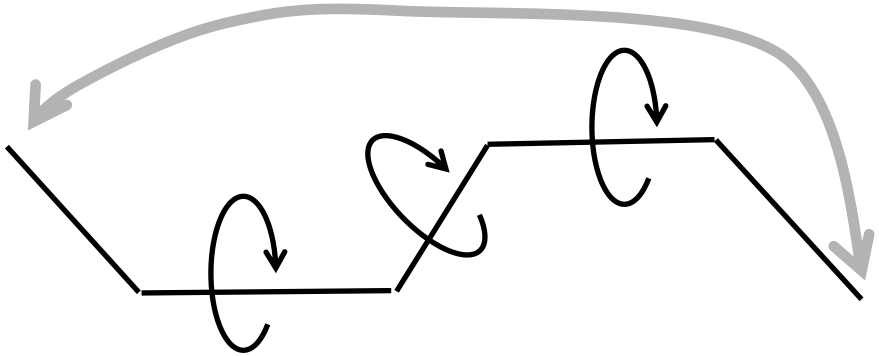
Distances and angles

One angle is easy





longer distances depend on several angles

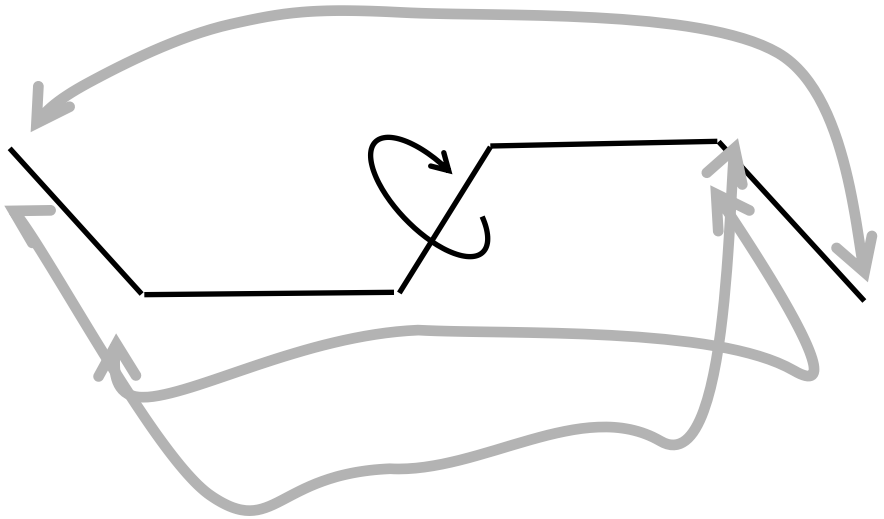


Distances and angles

Each angle affects many distances

What does one know ?

- simple optimisation will not work



Optimisation Strategy

Start

- concentrate on distances with few angles in between
- shorter distances become correct

Add in more distances

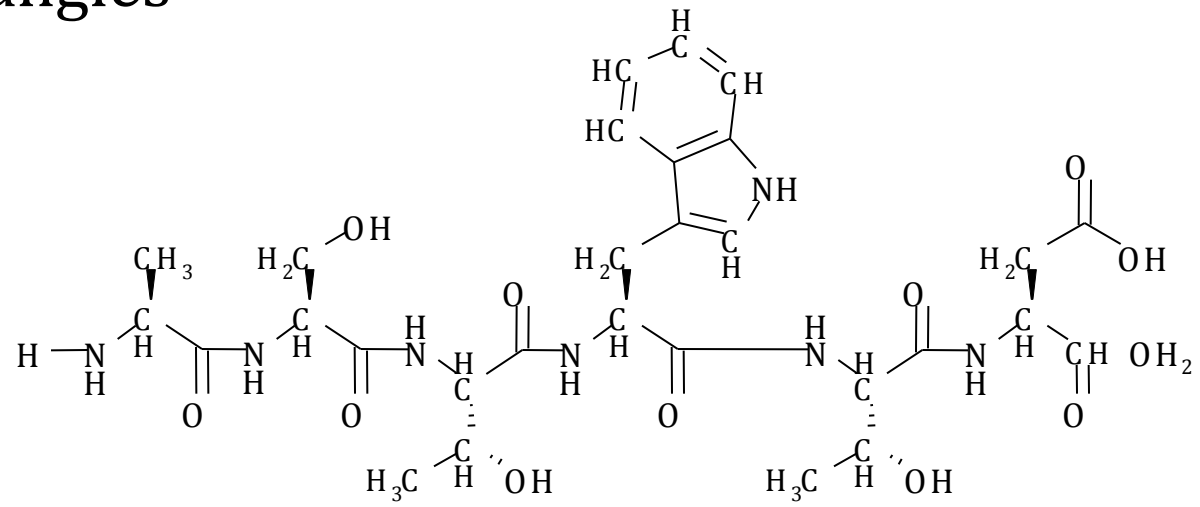
- re-optimise

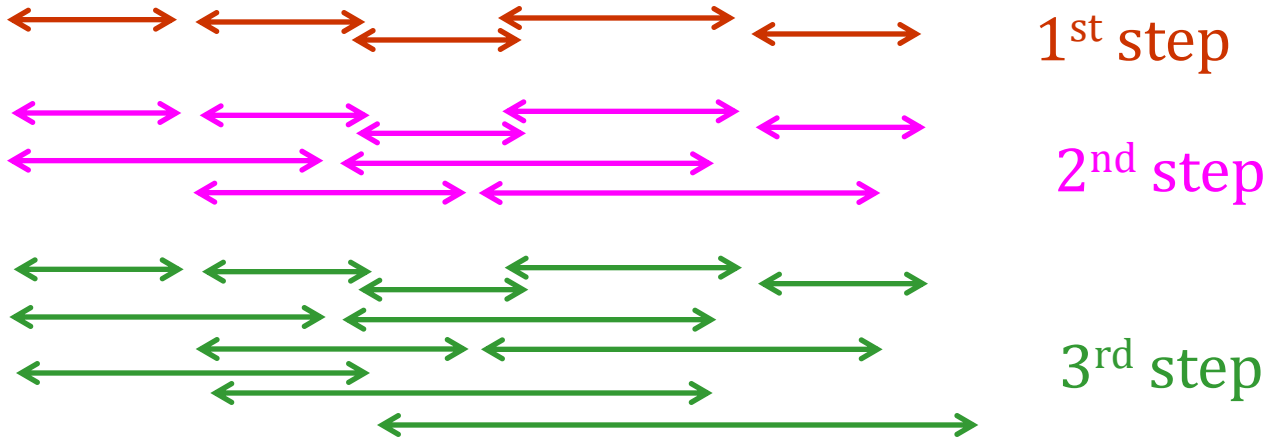
Add in more distances

- …

Variable target function

Work with torsion angles





ideas from Braun and Gō, 1980s

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

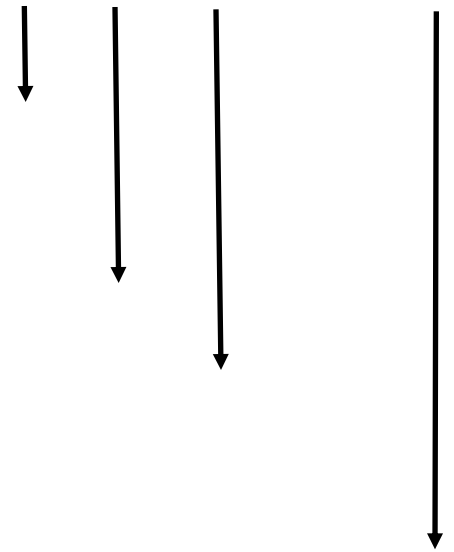
- Sort according to distance in sequence

Stepwise variable target function method

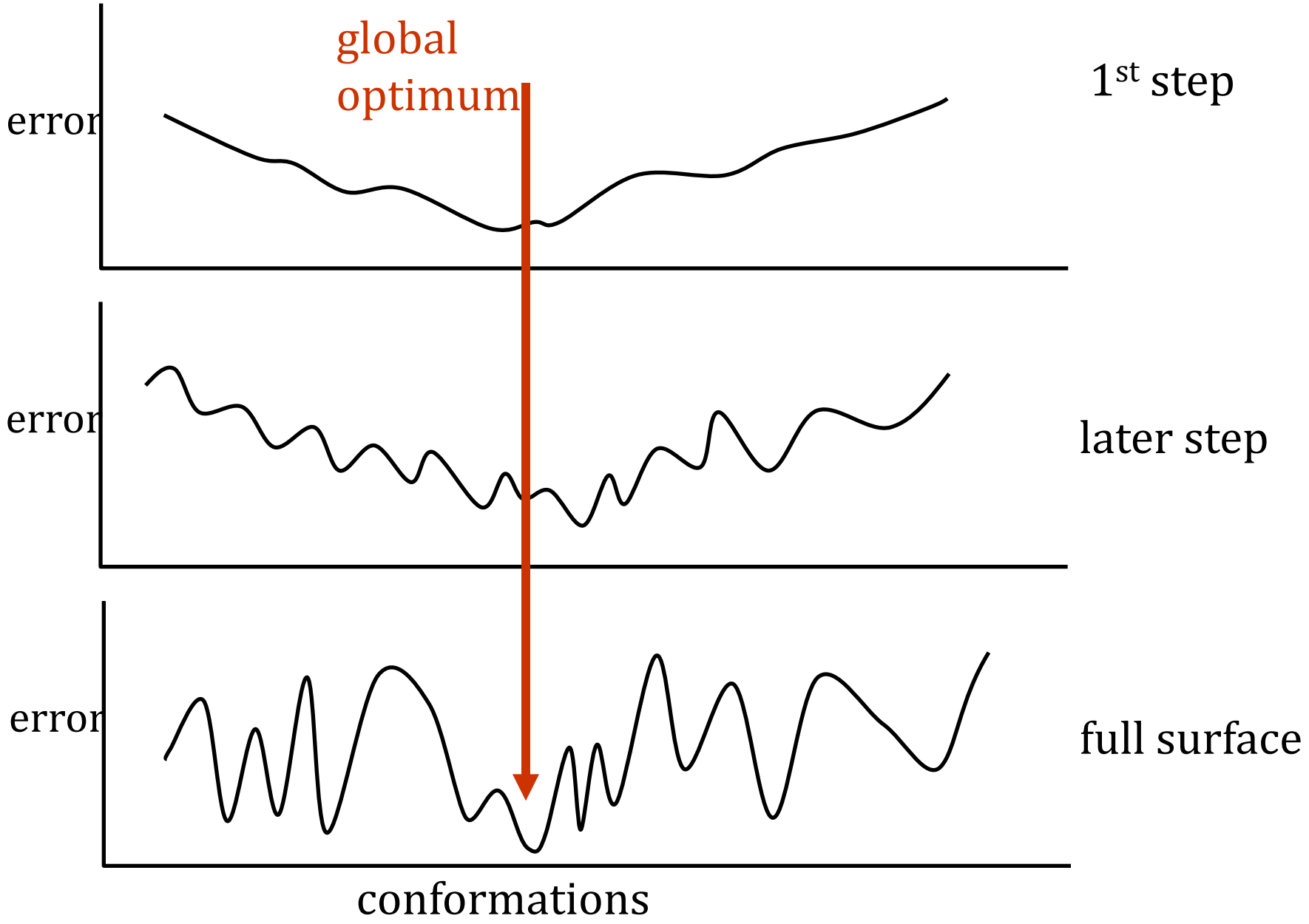
distance in sequence	residue 1	atom 1	residue 2	atom 2	distance in space (Å)
0	8	H^{α}	8	H^{γ}	4.4
0	3	H^{α}	3	H^{N}	5.0
1	5	H^{α}	6	H^{N}	4.0
1	7	H^{β}	8	H^{γ}	3.8
2	3	H^{α}	5	H^{γ}	5.0
…					
80	2	H^{α}	82	H^{N}	4.5
…	…				

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 *versus* variable target function

- proponents of both

variable target function probably more popular

- no problems with chirality

Real implementations of distance geometry

- not small programs
- Input ?
  - 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

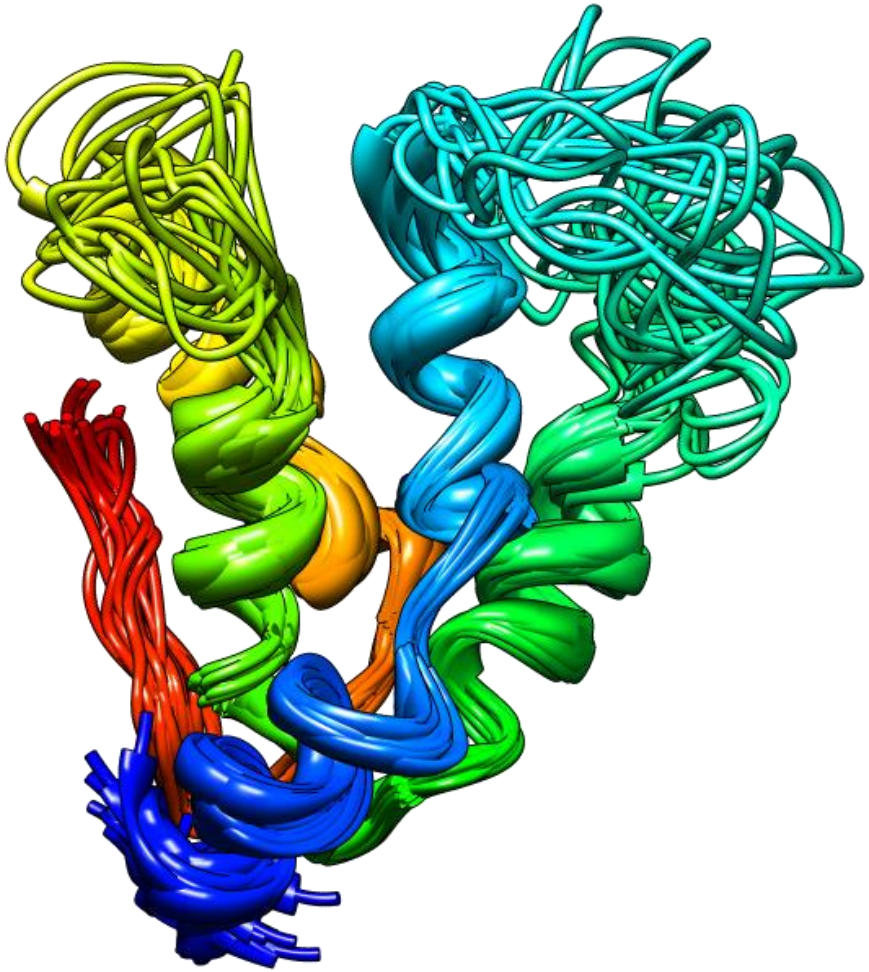
Output from programs

Structure impossible ?

- program dies or
- best possible solution

Structure not determined ?

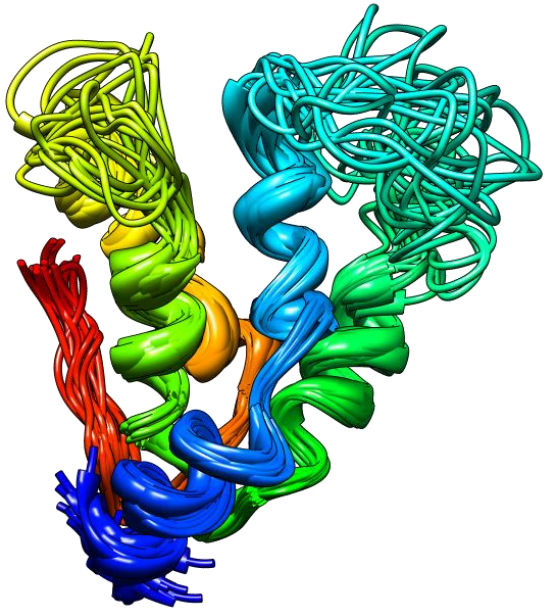
- set of possible conformations (10 to 100)



example 1sm7

Lots of models in a PDB file

- big difference compared to 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



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

- is this part of the protein mobile ?

Interactions

- add small molecule – which parts of spectrum change ?

Still more structural information

- residual dipolar coupling
- spin labels

Summary

- What information does one have ?
- Is it enough ? Is it consistent ?
- Two methods to generate structures
- Differences in handling chirality