NMR (Nuclear Magnetic Resonance Spectroscopy)

Literature / background (already in Stine)

Current standing
- ≈ 11% of current structures solved by NMR (10,618 structures, 9,287 proteins)
- about 1/4 of smaller structures (<100 residues)
How many structures by NMR?
sizes of NMR structures in protein data bank

Andrew Torda
26/10/2015
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 $^1\text{H}$, $^{13}\text{C}$, $^{15}\text{N}$ 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
\[ \nu \] 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

- Turn on a field
- Put in energy
- Let them relax

Some nuclei not doing much

Applied field
Some align
### Important nuclei (spin $\frac{1}{2}$)

<table>
<thead>
<tr>
<th>nucleus</th>
<th>sensitivity</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1$H</td>
<td>1</td>
<td>cheap and natural</td>
</tr>
<tr>
<td>$^{13}$C</td>
<td>$1.6 \times 10^{-2}$</td>
<td>expensive, but only 1% of natural abundance</td>
</tr>
<tr>
<td>$^{15}$N</td>
<td>$10^{-3}$</td>
<td>not cheap, 0.4 % natural abundance</td>
</tr>
<tr>
<td>$^{31}$P</td>
<td>$7 \times 10^{-2}$</td>
<td>DNA and other PO$_4$ chemistry, less protein</td>
</tr>
</tbody>
</table>

- 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
- $\text{NOE } \propto r^{-6}$
- usually only up to about 5 - 6 Å

Story
- two spins' dipoles interact

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$\alpha$ coupling

$\phi$  

$\text{cis} < 6 \text{ - 7 Hz}$  

$\text{trans} \sim 10 \text{ Hz}$  

$J_{H\alpha, NH}$
$^{3}J_{HN\alpha}$ coupling

formalised as

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

Problems...

Where do 6.4, 1.4, 1.9 come from?
Do not learn for Klausur

Amide NH to H\textsubscript{\(\alpha\)} coupling

- can help distinguish \(\alpha\) from \(\beta\)
- not always seen (exchange / motion)
- NH not always present
- other angles ?
  - other vicinal protons
    - \(C^{\alpha}\) to \(C^{\beta}\)
Problems with $J$-coupling

1. we have a formula

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

Measure $J$, solve for $\theta$

- 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 $-90^\circ$
Practical NMR

We have some basic methods

Real NMR
• more techniques
  • 2D and more
  • identifying specific kinds of atom
  • spreading peaks out
### Information summary

<table>
<thead>
<tr>
<th>phenomenon</th>
<th>assignments</th>
<th>structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>chemical shift</td>
<td>important</td>
<td>not much used</td>
</tr>
<tr>
<td>spin-spin ((J)) coupling</td>
<td>important</td>
<td>torsion angles</td>
</tr>
<tr>
<td>NOE</td>
<td>important</td>
<td>distances</td>
</tr>
</tbody>
</table>

More spectroscopy
- filtering according to chemistry, atom types
- \(n\)-dimensional methods

Structural information
- labels for broadening / shifting peaks
- orientation of bonds to reference..
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 $20 \, N_{res}$
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

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

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

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

implied 6 or 7 Å

5 Å
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  |  A  B  C  D  E
---|-------------
A  |  4
B  |  3  5
C  |  2  10
D  |  3
E  |  max  max  max
```

```
A
B
C
D
E

A  |  A  B  C  D  E
---|-------------
A  |  max max max
B  |  3  5 max
C  |  2  10
D  |  3
E  |  3
```
Bound smoothing / Floyd

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

\[ O(n^3) \]
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

- 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 $\sum_i (r_i - r_i^{measured})^2$ 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

1\textsuperscript{st} step

2\textsuperscript{nd} step

3\textsuperscript{rd} step

ideas from Braun and Gō, 1980s
Stepwise variable target function method

- Collect experimental data

<table>
<thead>
<tr>
<th>distance in sequence</th>
<th>residue atom 1</th>
<th>residue atom 2</th>
<th>distance in space (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 Hα</td>
<td>6 HN</td>
<td>4.0</td>
</tr>
<tr>
<td>0</td>
<td>8 Hα</td>
<td>8 Hγ</td>
<td>4.4</td>
</tr>
<tr>
<td>80</td>
<td>2 Hα</td>
<td>82 HN</td>
<td>4.5</td>
</tr>
<tr>
<td>2</td>
<td>3 Hα</td>
<td>5 Hγ</td>
<td>5.0</td>
</tr>
<tr>
<td>1</td>
<td>7 Hβ</td>
<td>8 Hγ</td>
<td>3.8</td>
</tr>
<tr>
<td>0</td>
<td>3 Hα</td>
<td>3 HN</td>
<td>5.0</td>
</tr>
</tbody>
</table>

- Sort according to distance in sequence
## Stepwise variable target function method

<table>
<thead>
<tr>
<th>distance in sequence</th>
<th>residue 1</th>
<th>atom 1</th>
<th>residue 2</th>
<th>atom 2</th>
<th>distance in space (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>Hα</td>
<td>8</td>
<td>Hγ</td>
<td>4.4</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>Hα</td>
<td>3</td>
<td>HN</td>
<td>5.0</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>Hα</td>
<td>6</td>
<td>HN</td>
<td>4.0</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Hβ</td>
<td>8</td>
<td>Hγ</td>
<td>3.8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>Hα</td>
<td>5</td>
<td>Hγ</td>
<td>5.0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>2</td>
<td>Hα</td>
<td>82</td>
<td>HN</td>
<td>4.5</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Stepwise variable target function method

<table>
<thead>
<tr>
<th>distance in sequence</th>
<th>residue</th>
<th>atom</th>
<th>residue</th>
<th>atom</th>
<th>distance in space (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>$H^\alpha$</td>
<td>8</td>
<td>$H^\gamma$</td>
<td>4.4</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>$H^\alpha$</td>
<td>3</td>
<td>$H^N$</td>
<td>5.0</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>$H^\alpha$</td>
<td>6</td>
<td>$H^N$</td>
<td>4.0</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>$H^\beta$</td>
<td>8</td>
<td>$H^\gamma$</td>
<td>3.8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>$H^\alpha$</td>
<td>5</td>
<td>$H^\gamma$</td>
<td>5.0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>2</td>
<td>$H^\alpha$</td>
<td>82</td>
<td>$H^N$</td>
<td>4.5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
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 $E_{phys}(\mathbf{r}) = \textit{bonds} + \textit{angles} + \textit{electrostatics}$ ...

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

- and... $E_{restr} = \sum_i k_i (r_{i,\text{struct}} - r_{i,\text{measured}})^2$

- 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