Sequential Assignment Strategies in Proteins
NMR assignments

• in order to determine a structure by traditional, NOE-based $^1\text{H}^-^1\text{H}$ distance-based methods, the chemical shifts of the individual $^1\text{H}$ nuclei must be known

• older methods based on homonuclear ($^1\text{H}$) experiments only, whereas newer methods based on heteronuclear / triple resonance experiments

• triple resonance techniques generally require $^1\text{H}$, $^{13}\text{C}$ and $^{15}\text{N}$ chemical shift assignments in a protein

• chemical shifts so obtained are useful for predicting secondary structure of proteins including main chain $\phi$ and $\psi$ angles
1H methods: The Sequential Assignment Procedure

- due to Wüthrich and coworkers (early/mid 1980’s)

```
“1. Identification of amino acid side chain spin systems.  2. Identification of neighboring residues in the amino acid sequence.  3. Suitable combinations of the results from 1 and 2 which provides individual resonance assignments in the primary structure of the protein.”

K. Wüthrich  1983  Biopolymers 22, 131-138
```

1). Use $J$ correlation spectroscopy (COSY, TOCSY) to identify individual spin systems (amino acids)

2). Use NOE spectroscopy (NOESY) to link amino acids

**Book:** “NMR of Proteins and Nucleic Acids”, Kurt Wüthrich, 1986 (Wiley)
**$^1$H methods: The Sequential Assignment Procedure**

1). Identify amino acid spin systems using COSY and TOCSY

- Spin systems are classified based on the spins that can be correlated in COSY spectra, and then TOCSY spectra
  - remember, there is no way to connect amino acids in a protein by COSY or TOCSY
  - a COSY or TOCSY spectrum of a protein is comprised of the individual spectra of the amino acids in the protein
- the individual spin systems are identified as to amino acid type
  - how? Based on “amino acid specific patterns” of COSY and TOCSY cross peaks

![Diagram of amino acids]

- Asp
- Val
More examples of COSY patterns
\( ^1H \) chemical shift information

- \( ^1H \) chemical shift information is also very useful in assisting to identify spin systems

- shown are distributions of \( ^1H \) chemical shifts for amino acids in proteins

**$^1$H methods: The Sequential Assignment Procedure**

2). Link identified amino acids using NOESY data

- critical observation: it is virtually impossible to align the main chain atoms of two adjacent amino acid residues in a protein so that there is not at least one pair of interresidue distances between main chain hydrogens that is significantly less than the NOE detection limit (~5Å)

---

**dashed lines:** nonlabile protons correlated by $J$ correlation experiments

**arrows:** sequential interresidue NOEs
Sequential interresidue NOEs

- most short interproton \((^1H^N, \; ^1H\alpha, \; ^1H\beta)\) distances in proteins are between directly adjacent amino acid residues
  - intense NOEs indicate adjacent amino acids
- less intense \(i, \; i+2\) and \(i, \; i+3\) NOEs ("nonsequential") also observed and useful, particularly in well-defined secondary structures

<table>
<thead>
<tr>
<th>Table 8.1. Statistics of Short (^1H—^1H) Distances in Protein Crystal Structures.(^a)</th>
<th>Table 7.1. Short ((\leq 4.5 \ \text{Å})) Sequential and Medium-Range (^1H—^1H) Distances in Polypeptide Secondary Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (Å)</td>
<td>j-1 = 1 (%)</td>
</tr>
<tr>
<td>(d_{\alpha N}(i,j) \leq 2.4)</td>
<td>98</td>
</tr>
<tr>
<td>3.0</td>
<td>88</td>
</tr>
<tr>
<td>3.6</td>
<td>72</td>
</tr>
<tr>
<td>(d_{NN}(i,j) \leq 2.4)</td>
<td>94</td>
</tr>
<tr>
<td>3.0</td>
<td>88</td>
</tr>
<tr>
<td>3.6</td>
<td>76</td>
</tr>
<tr>
<td>(d_{\beta N}(i,j) \leq 2.4)</td>
<td>79</td>
</tr>
<tr>
<td>3.0</td>
<td>76</td>
</tr>
<tr>
<td>3.6</td>
<td>66</td>
</tr>
<tr>
<td>(d_{\alpha N}(i,j) \leq 3.6, ; d_{NN} \leq 3.0)</td>
<td>99</td>
</tr>
<tr>
<td>(d_{\alpha N}(i,j) \leq 3.6, ; d_{\beta N} \leq 3.4)</td>
<td>95</td>
</tr>
<tr>
<td>(d_{NN}(i,j) \leq 3.0, ; d_{\beta N} \leq 3.0)</td>
<td>90</td>
</tr>
</tbody>
</table>

\(^a\) For the turns, the first of two numbers applies to the distance between residues 2 and 3, the second to that between residues 3 and 4 (Fig. 7.12). The range indicated for \(d_{\alpha N}(i,i+3)\) corresponds to the distances adopted if \(\psi_i\) is varied between \(-180\) and \(180^\circ\).  
\(^b\) The ranges given correspond to the distances adopted by a \(\beta\)-methine proton if \(\chi^1\) is varied between \(-180\) and \(180^\circ\).
Interresidue NOEs

- examples of NOE connectivities
- graphical means to summarize NOEs and demonstrate self-consistency of assignments
Problems in sequential assignment procedure

- spectral degeneracy can preclude complete COSY/TOCSY pattern identification

DQF COSY

ubiquitin, 8.5 kDa

cytochrome c, 12.5 kDa
Problems in sequential assignment procedure

- degeneracy of COSY patterns
  - normally, only 8-12 species can be distinguished from among the 20 amino acids prior to NOE-based sequence alignment
- plasticity of COSY patterns
  - In proteins, local environment effects on $J$-coupling and relaxation/dynamics lead to variability in COSY patterns
Main Chain Directed (MCD) approach

• due to Wand/Englander and coworkers (mid/late 1980’s)

1). use $J$ correlated and spectra to first identify $^1H_N-^1H^\alpha-^1H^\beta$ ("NAB") units
   - the $^1H_N$ region of COSY spectra is usually less crowded
   - individual amino acid type identification not attempted at this point

2). next align the NAB units sequentially using NOESY spectra
   - pattern recognition routines employed to search for well-established patterns of NOEs
   - amino acid type identification is then substantially restricted

• no initial reliance on identity of amino acid to establish connectivity
• amenable to automation (pattern matching algorithms)
Triple resonance approach

• the triple resonance approach has largely replaced $^1$H-only methods in cases where the protein can be isotopically labeled
  - is the only method currently available for large proteins

• experiments provide selective chemical shift correlation of main chain (plus $^1$H$^\beta$ and $^{13}$C$^\beta$) nuclei in adjacent amino acids
  - these correlations permit links between individual amino acids to be established, and thus assignment of the main chain
  - the chemical shifts of side chain nuclei are then correlated with assigned main chain nuclei to complete side chain assignments

• in the ideal case, no other information is required
  - in the ideal case, and in theory, amino acid type need not be established initially
  - in practice, many other types of information, including chemical shift/amino acid type information, NOE distance information, etc., play important roles
Triple resonance approach: a simple example

2D HNCA projection

3D HNCA

\[\begin{array}{ccc}
   c & a & b \\
   8.24 & 7.71 & 8.40 \\
   117.1 & 122.8 & 123.8 \\
   68.43 & 61.32 & 58.52 \\
   61.32 & 55.03 & 68.43 \\
\end{array}\]
Triple resonance approach: a simple example
Triple resonance approach: a simple example

- link the correlated shifts numerically….

\[
\begin{align*}
\text{H} & \quad \text{O} & \quad \text{H} \\
\text{C} & \quad \text{C} & \quad \text{N} & \quad \text{C} \\
55.03 & \quad 8.24 & \quad 117.1 & \quad 68.43 \\
R & \quad \text{R} & \quad \text{a} & \quad \text{b} \\
\end{align*}
\]

- or visually

\[
\begin{align*}
\text{H} & \quad \text{O} & \quad \text{H} & \quad \text{H} \\
\text{C} & \quad \text{C} & \quad \text{N} & \quad \text{C} \\
55.03 & \quad 8.24 & \quad 7.71 & \quad 8.40 \\
R & \quad \text{R} & \quad \text{a} & \quad \text{b} \\
\end{align*}
\]
Triple resonance approach: HNCA/HN(CO)CA example

- problems:
  - $^{13}\text{C}^\alpha$ chemical shift degeneracy in proteins
  - $^{13}\text{C}^\alpha$ linewidths/resolution
    - these preclude complete linkage via $^{13}\text{C}^\alpha$ alone
    - the same is true for $^{13}\text{C}^\beta$, $^{13}\text{C}'$
$	ext{13C}^\alpha$ chemical shifts

- $\text{13C}^\alpha$ chemical shift degeneracy limits the utility of experiments that link amino acids via $\text{13C}^\alpha$

- $\text{13C}^\alpha$ chemical shifts can be useful for identifying some amino acids (Gly, Ile/Val)

- $\text{13C}^\alpha$ chemical shifts are also good indicators of 2° structure (providing the amino acid type is known)

$\text{13C}^\alpha$ chemical shifts in proteins
- dark bars, alpha helices
- light bars, beta sheets
HNCACB / CBCA(CO)NH, connectivity via $^{13}\text{C}^\beta$

- $^{13}\text{C}^\beta$ correlations permit resolution of ambiguities in connectivities due to $^{13}\text{C}^\alpha$ degeneracy
\( ^{13}\text{C}^\beta \) chemical shifts

- \( ^{13}\text{C}^\beta \) chemical shift degeneracy, on average, is not as severe as \( ^{13}\text{C}^\alpha \) degeneracy
- \( ^{13}\text{C}^\beta \) chemical shifts are also useful for amino acid identification (Ser/Thr, Ala)
- \( ^{13}\text{C}^\alpha \) chemical shifts are also good indicators of 2\(^\circ\) structure (providing the amino acid type is known)

\( ^{13}\text{C} \) chemical shifts in proteins
HNCO / HN(CA)CO, connectivity via $^{13}C'$ (carbonyl)

- $^{13}C'$ correlations permit resolution of ambiguities in connectivities due to $^{13}C^\alpha$ and $^{13}C^\beta$ degeneracy

- $^{13}C'$ chemical shifts somewhat degenerate, but good indicators of 2° structure

$^{13}C'$ chemical shifts in proteins
- dark bars, alpha helices
- light bars, beta sheets
Sidechain assignments

- side chain $^1$H and $^{13}$C resonances are assigned based on correlations with main chain $^1$HN, $^{15}$N, $^1$H$^\alpha$, and $^1$H$^\beta$ and $^{13}$C$^\beta$ chemical shifts
- $^1$HN, $^{15}$N resolved TOCSY experiments are often used, but don’t provide for direct correlation of side chain $^{13}$C and $^1$H nuclei with one another
- these experiments are also limited by $^1$HN, $^{15}$N degeneracy

-HCCH TOCSY and HCCH COSY experiments provide direct side chain $^1$H, $^{13}$C connectivities
- NOESY data are also used to assist in assignment of side chain resonances
Other (main chain) assignment procedures

• automated/semiautomated assignment procedures
  - performance (usually) highly dependent on data quality, completeness, and peak picking
  - usually require, as input, correlated spins from many triple resonance experiments ($^1$H$^N_i$, $^{15}$N$^N_i$, $^{13}$C$^\alpha_i$, $^{13}$C$^\alpha_{i-1}$, $^{13}$C$^\beta_i$, $^{13}$C$^\beta_{i-1}$, $^{13}$C$'_{i}$, $^{13}$C$'_{i-1}$)
  - other types of useful information permitted (amino acid type if known, etc.)
  - usually require, or work best with, some interactive user intervention
• “no assignment” procedures
  - residual dipolar coupling method for simultaneous structure determination and resonance assignment (4, Prestegard group)
  - measured dipolar couplings and chemical shifts limit connectivity ambiguity via $^{13}$C$^\alpha$ permitting

Some uses of chemical shifts

- predicting secondary structure
  - because $^{1}H^\alpha$, $^{13}C^\alpha$, $^{13}C^\beta$, and $^{13}C'$ chemical shifts depend on secondary structure, these shifts can be used to predict secondary structure
  - Chemical Shift Index ("CSI" (1))

- predicting main chain dihedral angles ($\phi$ and $\psi$)
  - "TALOS" (2): uses chemical shift and sequence information / database matching to predict reliable values for $\phi$ and $\psi$

   also http://www.pence.ca/software/csi/latest/csi.html
   also http://spin.niddk.nih.gov/bax/software/TALOS