Protein Structure Determination by NMR

Hsiau-Wei Lee, Ph.D.

April 14th, 2015
History of NMR

1940s  First observation of NMR in solids and liquids (1945)
1960s  Superconducting magnet
       Pulse Fourier transform approach developed
       Nuclear Overhauser effect used for structure determination
1985   First protein structure solved by NMR
2000s  High field NMR with cryogenic probes

Unique Application of NMR: protein with disordered region, weak associations, membrane association, and in-cell observation.

NMR Recognition

• 1952 - Bloch and Purcell – Nobel Prize in Physics
• 1991 – Richard Ernst – Nobel Prize in Chemistry
• 2002 – Kurt Wuthrich – Nobel Prize in Chemistry
• 2003 – Lauterbur and Mansfield – Nobel Prize in Physiology and Medicine
What Kind of Information Can NMR Provide?

Conformational Analysis
Metal Binding Studies

Sequential Assignment
Distant Constraints for Structural Calculation

Dynamic Properties
Hydrogen Exchange
Measuring Relaxation Rate

Ligand Binding Studies
Protein-Protein Interactions
Protein-DNA Interactions

Determine the Oligomeric State of Biomolecule

- 1D $^1$H Spectrum
- 3D HNCO

$I = I_0 e^{-t/T_1}$
UCSC NMR Facility

Coming Soon!!!
NMR Sample

To be active in NMR, the nuclei need to possess a property called the Spin and is represented by the letter I

\[ ^1\text{H}: I = \frac{1}{2} \rightarrow \text{Signal}; \text{Abundance} = 99.98\% \]
\[ ^{13}\text{C}: I = \frac{1}{2} \rightarrow \text{Signal}; \text{Abundance} = 1.108\% \]
\[ ^{15}\text{N}: I = \frac{1}{2} \rightarrow \text{Signal}; \text{Abundance} = 0.365\% \]

Ideal Sample Properties:
- Protein with high expression yield
- No aggregation
- Stable at room temperature
- High Solubility
  Preferably > 500 μM
- Salt concentration, minimum if possible
- Volume
  Standard NMR Tube = 450 μL
  Shigemi NMR Tube = 260 μL
NMR Actively Nuclei

Net magnetization ($M_0$)

$\varphi = 90^\circ$

Predicted 1H NMR Spectrum

NMR Spectrum of a Protein
Proteins can be uniformly isotopically labeled by recombinant expression using defined media:

- Bacterial expression most common
  - Also yeast, and cell-free systems are being developed
- Minimal media using $^{13}$C$_6$ glucose as the sole carbon source and $^{15}$NH$_4$Cl (or -SO$_4$) as the sole nitrogen source
  - Normally >98% atom excess
  - Also labeled “rich” media ($$$)
- For larger proteins, uniform or fractional $^2$H labeling also used
  - $^2$H, $^{13}$C glucose and D$_2$O

### Uniform Isotopic Labeling of Proteins

\[
\begin{align*}
^1\text{H}: & I = \frac{1}{2} \rightarrow \text{Signal}; \text{Abundance} = 99.98\% \\
^{13}\text{C}: & I = \frac{1}{2} \rightarrow \text{Signal}; \text{Abundance} = 1.108\% \\
^{15}\text{N}: & I = \frac{1}{2} \rightarrow \text{Signal}; \text{Abundance} = 0.365\%
\end{align*}
\]
Heteronuclear Single Quantum Coherence Spectroscopy
The Need for Isotope Labeling and 3D Experiments

IKURA M; KAY LE; BAX A, (1990) BIOCHEMISTRY 29:4659-4667
Triple Resonance Assignment Strategy
With $^{15}\text{N}$ and $^{13}\text{C}$ Isotopic Labeling

**HNCA**

[Diagram of HNCA with resonance assignments and 2D/3D projections]
NMR Structure Determination

- Sequential Assignment
- NOE Distance Restraint Assignment
- Dihedral Angle Restraints (TALOS)
- Residual Dipolar Couplings (REDCAT)

Structural Calculation for the Monomer (CYANA and Xplor-NIH)
Triple Resonance Assignment Strategy

Basic Experiment
- $^{15}$N HSQC
- $^{13}$C HSQC
- HNCO

Backbone Assignments
- Good (S/N) Poor (S/N)
  - HNCA(CO)NH: HNCO
  - CBCA(CO)NH
  - HNCA
  - HBHA(CO)NH

Sidechain Assignments
- HC(CO)NH: HCCH-TOCSY
- C(CO)NH: $^{15}$N NOESY
- HCCH-TOCSY: $^{13}$C NOESY
- HBHA(CO)NH

Distance Restraints
- $^{15}$N NOESY
- $^{13}$C NOESY
  - aliphatic
  - aromatic

Structural Calculation using CYANA and XPLOR
Refinement with Residual Dipolar Couplings
Overview of Protein Structure Initiative (PSI)

**Aim:** To reduce cost and time required to determine three-dimensional protein structure

**PSI – 1 (2000 to 2005)**
- Develop methods to streamline the structure determination
- Optimizing the cell cloning, expression, and protein purification
- Approx. 1,100 protein structures were solved

**PSI – 2 (2005 to 2010)**
- Determine large numbers of structures based on the foundation developed in PSI – 1
  - Further develop the high throughput pipeline

**PSI – 3 (2010 to 2015)**
- Emphasis on the biological relevance
- High-Throughput enabled structural biology partnership
Protein Sequence of ChR158

MKHIKFTEKI TIHASAEIFF DVTQDYAQRL RWDTFLKQAE LIEGAERAGKG
VKAYCAAKNG MGMVTEYVTF NRPKATAINM TKGPYMFESF LGSWNYKHIGE
NETEVIFLYA FSLRFPFNLI WKVVENNLQR NVRQRLDLLK KYIETNLEHHH

- Count # of crosspeaks in $^{15}$N-edit HSQC spectrum
- Look for crosspeaks of Gly, Ser, and Thr
Protein Sequence of ChR145

MEIKLIAQVK TVINAPIEKV WEALVNPEII KEYMFGTTVV SDWKEGSQIVW
KGEWKGKAYE DKGTLQFNE RSILQYSHFS PLTGKPDLPE NYHVVTITLTA
LKKGVEVEILT QDNNETEKEQ KHSEDNWNTM LEGLEHHHHHH H

-V-S-A-L- Section not found
Protein Sequence of ChR158.006 (WRONG)
MKHIKFTeki TiHASAEIIIF DVTQDYAQRL RWDTFLKQAe LIEGAERAGkG
VKAYCAAkNG MGMVTEyVTf NRPKATAINM TKGPyMFESF LGSwNYKHIGe
NETeVIFLYA FSLRFPFNLI WKIVeNNLQR NrVQRaLLDLK KEYETNLEHHH
HHH
Calculated MW of $^{13}$C, $^{15}$N Labeled Protein: 19457
Molecular Weight measured by Mass-Spec: 18861

Protein Sequence of ChR145.008 (WRONG)
MEIKLIAQVK TVINAPIEKV WEALVNPEII KEYMfGTcVv SDWKEGSQIVW
KGewKgkAYe DKGTLQFNE RsILQySHFS PLTgKPDLPE NYHVVTITLTA
LkKGeVEVELT QDNNeteKEQ KHSeDnWntm LeGLeHHHHH H
Calculated MW of $^{13}$C, $^{15}$N Labeled Protein: 17547
Molecular Weight measured by Mass-Spec: 17306

Protein Sequence of ChR145
MEIKLIAQVK TVINAPIEKV WEALVNPEII KEYMfGTcVv SDWKEGSQIVW
KGewKgkAYe DKGTLQFNE RsILQySHFS PLTgKPDLPE NYHVVTITLTA
LkKGeVEVELT QDNNeteKEQ KHSeDnWntm LeGLeHKLFLk KVSaLeHHHHH
Calculated MW of $^{13}$C, $^{15}$N Labeled Protein: 18880
Molecular Weight measured by Mass-Spec: 18861
Heteronuclear: Backbone Assignment

HNCACB

CBCA(CO)NH

T55

L56

Dimension 1: $^1$H
Dimension 2: $^{13}$C
Dimension 3: $^{15}$N
Heteronuclear: Backbone Assignment

CBCA(CO)NH

Dimension 1: $^1\text{H}$
Dimension 2: $^{13}\text{C}$
Dimension 3: $^{15}\text{N}$

HBHA(CO)NH

Dimension 1: $^1\text{H}$
Dimension 2: $^1\text{H}$
Dimension 3: $^{15}\text{N}$
Heteronuclear: Sidechain Assignment

C(CO)NH

H(CCO)NH
$^{15}\text{N}-^{1}\text{H}$ HSQC Spectrum of ChR145
Homonuclear: Degeneracy of a NOESY Spectrum
Heteronuclear Edited NOESY

\[ ^{15}\text{N-NOESY} \]

- Dimension 1: \(^{1}\text{H} \)
- Dimension 2: \(^{1}\text{H} \)
- Dimension 3: \(^{15}\text{N} \)

\[ ^{13}\text{C-NOESY} \]

- Dimension 1: \(^{1}\text{H} \)
- Dimension 2: \(^{1}\text{H} \)
- Dimension 3: \(^{13}\text{C} \)
NOE Experimental Restraints

**NOE Distance Restraint:**
- Sequential NOE (2° structure)
- Calibration with well defined $^1$H-$^1$H distance
- NOE intensities $\rightarrow$ approximate distance
  - Strong $\quad 1.8 – 2.7 \text{ Å}$
  - Medium $\quad 1.8 – 3.3 \text{ Å}$
  - Weak $\quad 1.8 – 5.0 \text{ Å}$
  - Very weak $\quad 1.8 – 6.0 \text{ Å}$
NOE Experimental Restraints

Intra-strand NOE between V5 and L15 of ubiquitin
Manual NOE Cross Peak Calibration

NOE S/N Scale for NOESY 150

<table>
<thead>
<tr>
<th>S/N</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 150</td>
<td>2.20 0.40 0.80</td>
</tr>
<tr>
<td>150 &gt; x &gt; 80</td>
<td>2.40 0.60 1.00</td>
</tr>
<tr>
<td>80 &gt; x &gt; 60</td>
<td>2.60 0.80 1.00</td>
</tr>
<tr>
<td>60 &gt; x &gt; 40</td>
<td>3.00 1.20 1.00</td>
</tr>
<tr>
<td>40 &gt; x &gt; 20</td>
<td>3.20 1.40 1.60</td>
</tr>
<tr>
<td>&lt; 20</td>
<td>4.00 2.20 2.00</td>
</tr>
</tbody>
</table>

Y93Hδ-Y93Hε:
S/N is 91
2.40 0.40 0.40
## NOE Experimental Restraints

**TABLE 7.1.** Short (≤ 4.5 Å) Sequential and Medium-Range ¹H–¹H Distances in Polypeptide Secondary Structures

<table>
<thead>
<tr>
<th>Distance</th>
<th>α-helix</th>
<th>β-ribbon</th>
<th>β-sheet</th>
<th>turn I⁺</th>
<th>turn II⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>d_{αN}</td>
<td>3.5</td>
<td>3.4</td>
<td>2.2</td>
<td>2.2</td>
<td>3.4</td>
</tr>
<tr>
<td>d_{αN}(i+2)</td>
<td>4.4</td>
<td>3.8</td>
<td></td>
<td>3.6</td>
<td>3.3</td>
</tr>
<tr>
<td>d_{αN}(i+3)</td>
<td>3.4</td>
<td>3.3</td>
<td></td>
<td>3.1-4.2</td>
<td>3.8-4.7</td>
</tr>
<tr>
<td>d_{αN}(i+4)</td>
<td>4.2</td>
<td></td>
<td></td>
<td>2.6</td>
<td>4.5</td>
</tr>
<tr>
<td>d_{iN}</td>
<td>2.8</td>
<td>2.6</td>
<td>4.3</td>
<td>4.2</td>
<td>2.6</td>
</tr>
<tr>
<td>d_{iN}(i+2)</td>
<td>4.2</td>
<td>4.1</td>
<td></td>
<td>3.8</td>
<td>4.3</td>
</tr>
<tr>
<td>d_{iN}(i+3)</td>
<td>2.5-4.1</td>
<td>2.9-4.4</td>
<td>3.2-4.5</td>
<td>3.7-4.7</td>
<td>3.6-4.6</td>
</tr>
<tr>
<td>d_{iB}(i+3)</td>
<td>2.5-4.4</td>
<td>3.1-5.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 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_{αN}(i+3) corresponds to the distances adopted if θ₁ is varied between −180 and 180°.

* The ranges given correspond to the distances adopted by a β-methine proton if χ₁ is varied between −180 and 180°.

**TABLE 8.1.** Statistics of Short ¹H–¹H Distances in Protein Crystal Structures

<table>
<thead>
<tr>
<th>Distance (Å)</th>
<th>j-1 = 1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d_{αN}(i,j)</td>
<td>≤ 2.4</td>
</tr>
<tr>
<td></td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>72</td>
</tr>
<tr>
<td>d_{αN}(i,j)</td>
<td>≤ 3.0</td>
</tr>
<tr>
<td></td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>76</td>
</tr>
<tr>
<td>d_{αN}(i,j)</td>
<td>≤ 3.6</td>
</tr>
<tr>
<td></td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>66</td>
</tr>
<tr>
<td>d_{iN}(i,j)</td>
<td>≤ 3.6</td>
</tr>
<tr>
<td></td>
<td>99</td>
</tr>
<tr>
<td>d_{iN}(i,j)</td>
<td>≤ 3.6, d_{iN} ≤ 3.0</td>
</tr>
<tr>
<td></td>
<td>95</td>
</tr>
<tr>
<td>d_{iN}(i,j)</td>
<td>≤ 3.6, d_{iN} ≤ 3.4</td>
</tr>
<tr>
<td></td>
<td>90</td>
</tr>
</tbody>
</table>

Obtain the Dihedral Angle $\phi$ and $\varphi$

HNHA Spectra of CaM-CD2-III-5G

\[ \frac{I_{\text{cross}}}{I_{\text{diagonal}}} = -\tan^2(2\pi J\zeta) \]

\[ J(\Phi) = A \cos^2(\Phi-60) + B \cos(\Phi-60) + c \]

Torsion Angle Likelihood Obtained from Shift and Sequence Similarity (TALOS)

\[ \alpha \text{-helix} \quad \beta \text{-sheet} \]

Cornilescu et al. (1999), *J. Biomol. NMR*, 13, 289
TALOS’ Prediction for ChR145
Residual Dipolar Coupling

Poly-ethylene glycol ether
- Nonionic bilayer
- Readily available (Sigma, Fluka)
- Inexpensive
- Stable for long time
- Doped with CTAB / SDS

Bicelle
- Negatively charged particles

Bacteriophage

Poly-acrylamide gels
- Channels in gel contain protein
Stretched Polyacrylamide Gels
an orientation medium for RDC and RCSA measurement

Gel Pellets, Gels, Gels in Tubes

- Dry
- Cast
- Swell
- Stretched
- Isotropic
Overview of Structure Calculation

Molecular topology information

Experimental structural restraints (NOE, coupling constants, H-bond)

Template structure

Distance geometry

Simulated annealing

Regularization

Simulated annealing refinement

Structure selection

Analysis/validation (RMSD, “Procheck”, back calculation)

Adapted from Brunger, A. T., “X-PLOR Version 3.1.”
# Program for Structure Calculation

**CNS**  
[http://cns.csb.yale.edu/v1.1/](http://cns.csb.yale.edu/v1.1/)

**X-PLOR**  
[http://xplor.csb.yale.edu/xplor/](http://xplor.csb.yale.edu/xplor/)

**X-PLOR-NIH**  

**CYANA/DYANA**  

**ARIA**  
[http://www.pasteur.fr/recherche/unites/Binfs/aria](http://www.pasteur.fr/recherche/unites/Binfs/aria)


  - Güntert, P. *Prog. NMR Spectrosc.* 2003, 43, 105-125.

Automated NOE Assignment

**ARIA**

*ARIA: automated NOE assignment and NMR structure calculation*

Jens P. Linge, Michael Habeck, Wolfgang Rieping and Michael Nilges

Unité de Bio-Informatique Structurale, Institut Pasteur, 25-28 rue du docteur Roux, F-75015 Paris, France

Received on July 25, 2002; revised on September 5, 2002; accepted on September 9, 2002

**CYANA**

- Amino acid sequence
- Sequence-specific assignments
- NOESY cross peak positions and volumes

Find NOE assignments

Evaluate NOE assignments

Structure calculation

NOE assignments

3D Structure

NMR Experiment

NOE Spectrum

Iterative NOE assignment and structure calculation

Processing

ARIA
Automated NOE Assignment

< 1500 NOEs

2000 NOEs

2700 NOEs

CYANA

Amino acid sequence
Sequence-specific assignments
NOESY cross peak positions and volumes

Find NOE assignments
Evaluate NOE assignments
Structure calculation

NOE assignments
3D Structure
NOE Experimental Restraints
Applications of RDCs

<table>
<thead>
<tr>
<th>NeSG</th>
<th>bmrB</th>
<th>pdb nmr</th>
<th>pdb xray</th>
<th>alignment media</th>
<th>#residues</th>
<th>xray Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeR31</td>
<td>15702</td>
<td>2k2e</td>
<td>3cpk</td>
<td>phage</td>
<td>150</td>
<td>0.28</td>
</tr>
<tr>
<td>CsR4</td>
<td>15317</td>
<td>2jr2</td>
<td>2ota</td>
<td>peg (and peg+ctab)</td>
<td>68</td>
<td>0.32</td>
</tr>
<tr>
<td>CtR107</td>
<td>16097</td>
<td>2kcu</td>
<td>3e0h</td>
<td>phage (and peg)</td>
<td>158</td>
<td>0.30</td>
</tr>
<tr>
<td>GmR137</td>
<td>15844</td>
<td>2k5p</td>
<td>3cwi</td>
<td>peg</td>
<td>70</td>
<td>0.21</td>
</tr>
<tr>
<td>HR3646E</td>
<td>16250</td>
<td>2khn</td>
<td>3fia</td>
<td>polyacrylamide gel</td>
<td>110</td>
<td>0.29</td>
</tr>
<tr>
<td>MbR242E</td>
<td>16368</td>
<td>2kk0</td>
<td>3gw2</td>
<td>peg</td>
<td>100</td>
<td>0.29</td>
</tr>
<tr>
<td>Pfr193A</td>
<td>16385</td>
<td>2kl6</td>
<td>3idu</td>
<td>phage</td>
<td>114</td>
<td>0.30</td>
</tr>
<tr>
<td>SgR42</td>
<td>15604</td>
<td>2jz2</td>
<td>3c4s</td>
<td>peg</td>
<td>58</td>
<td>0.23</td>
</tr>
<tr>
<td>SoR77</td>
<td>15456</td>
<td>2juw</td>
<td>2qti</td>
<td>polyacrylamide gel</td>
<td>72</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Structure Validation

\[
Q = \left[ \frac{\sum (D_{obs} - D_{calc})^2}{\sum D_{obs}^2} \right]^{1/2}
\]

Refine NOE Based Structure

\[
E_{RDC} = (D_{obs} - D_{calc})^2
\]
Summary of Structural Statistics of the Ensemble of 10 ChR145 Solution Structures

# of NOE-derived distance restraints
- intra-residues: 424
- sequential: 768
- medium range: 508
- long range: 928
- total: 2628

# of backbone dihedral angle restraints: 255

<table>
<thead>
<tr>
<th>RDC</th>
<th>Q-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH Gel</td>
<td>0.32</td>
</tr>
<tr>
<td>NH C12E5 PEG</td>
<td>0.27</td>
</tr>
<tr>
<td>NH Phage</td>
<td>0.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rmsd (Å)</th>
<th>All</th>
<th>Ordered</th>
</tr>
</thead>
<tbody>
<tr>
<td>backbone</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>all heavy atoms</td>
<td>1.1</td>
<td>1.0</td>
</tr>
</tbody>
</table>

backbone dihedral angle distribution
- % in most favored region: 91%
- % in additionally allowed region: 8.1%
- % in generously allowed regions: 0.8%
- % in disallowed regions: 0.1%

Protein Structure Validation Suite (PSVS)
URL: psvs-1_5-dev.nesg.org