Spectral Processing and Analysis with NMRPipe

frank_delaglio@agilent.com

Frank Delaglio
July 2012
Spectrometer Format Conversion
1D-4D Fourier Transform and Signal Enhancement
Spectral Visualization
1D-4D Peak Detection and Quantification: Position, Amplitude, Width and Modulation/Evolution
Spectral Assignment
Extraction of Structural Parameters
Molecular Structure Calculation
Molecular Display and Structure Verification
Exploitation of Structure
Spectral Imaging
Automation, Batch Analysis, and Screening
Web-Based Server Implementations

Ad Bax ● Joeseph Barchi ● James Chou ● Gabriel Cornilescu ● George Gray ● Alex Grishaev
Stephan Grzesiek ● Georg Kontaxis ● John Kuszewski ● Ryan McKay ● John Pfeiffer
Ben Ramirez ● Michael Shapiro ● Tobias Ulmer ● Gerteen Vuister ● Justin Wu ● Jinfu Ying
Shen Yang ● Guang Zhu ● Edward Zartler

July 2012
frank_delaglio@agilent.com
Spectral Processing Function as a UNIX Filter

\[ \text{fourierTransform} \ (< \text{in.fid} > \text{out.ft}) \]

---

Spectral Processing Scheme as a UNIX Pipeline

\[ \text{sineBell} < \text{in.fid} | \text{fourierTransform} > \text{out.ft} \]
NMRPipe: Spectral Processing as a UNIX Pipeline

1D-4D FT, LP, MEM, ML, PCA
Parallel Processing
1D-4D Peak Detection and Quantification
Spectral Graphics, Strips, Projections
Extensively Customizable
Molecular Structure Calculation

NMR Parameter Calculation (Shifts, Dipolar Couplings, PCS, etc)

Customization is via standard scripting languages (C-shell, TCL)

Created and Maintained by one developer, with contributed modules

Solaris, IRIX, HP/UX, DEC OSF, IBM AIX, IBM Blue Gene, Convex OS, Cray OS, Mac OS X, Linux, WindowsXP Interix, VMWare Player

Bottom-up Software Design
NMRWish

Customized version of TCL/TK “wish” interpreter

Script-based Interactive Spectral graphics (multi-window and PostScript)

Generic Database Engine (GDB)

Manipulate Peak Data, Assignments, NMR Parameters, and Molecular Structure
The image shows a screenshot of the NMRPipe Conversion Utility. The screen displays various parameters and options for NMR data processing.

- **Spectrometer Input:**
- **Input Protocol:** Bruker (NIH)
- **Output Protocol:** NMRPipe
- **Output Template:** fid/test%03d.fid
- **Output Script:** fid.com

**Digits Oversampling Correction:**
- During Conversion
- During Processing

**Parameters and Values:**
- **Total Points R+I:** 1024
- **Valid Points:** 512
- **Acquisition Mode:** DQD
- **Spectral Width Hz:** 8992.806
- **Observe Freq MHz:** 600.141
- **Center Position PPM:** 4.534
- **x-axis:**
  - 1024
  - 512
  - DQD
  - 8992.806
  - 600.141
  - 4.534

**y-axis:**
- 82
- 41
- Complex
- 5952.381
- 60.819
- 118.089
- 15N

**z-axis:**
- 76
- 38
- Complex
- 1457.726
- 600.130
- 50.000
- 1H

**Conversion Script Text:**
```bash
#!/bin/csh
zcat | bruk2pipe -bad 0.0 -noaswap -DMX -decim 16 -dspfvs 12 \
  -xN 1024 -yN 82 -zN 76 \
  -xT 512 -yT 41 -zT 38 \
  -xMODE DQD -yMODE Complex -zMODE Complex \
  -xSW 8992.806 -ySW 5952.381 -zSW 1457.726 \
  -xOBS 600.141 -yOBS 60.819 -zOBS 600.130 \
  -xCAR 4.534 -yCAR 118.089 -zCAR 50.000 \
  -xLAB 1H -yLAB 15N -zLAB 1H \
  -ndim 3 -aq2D States \
  -out fid/test%03d.fid -verb -ov

sleep 5
```
Analyze Titration Curve to Estimate $K_d$
About the Cover

November 4, 2010: Vol. 114, Iss. 43

Conventional (1D) electronic spectra of NO$_2$ are largely patternless, but coherent 2D spectra of NO$_2$ show numerous X-shaped clusters such as that shown here. This spectrum is displayed in 2D contour form (below, in purple and black) and as a corresponding 3D intensity surface (above, in purple and blue). Orthogonal 1D projections (orange) of the surface resemble conventional 1D spectra. The spectral data was preprocessed using a pipeline-based software system for multidimensional spectroscopy (Delaglio, F.; Gnesi, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. J. Biomol. NMR 1995, 6, 277—293), which was also used to generate images used to build the 3D scene. The figure was produced by Frank Delaglio (www.nmrscience.com), who used the interactive virtual reality environment Second Life to render this image (see page 11365).
Bax Group NMR Calculation Servers

http://spin.niddk.nih.gov/bax/nmrserver
NMRPipe: Related Programs and Features

**TALOS+**: prediction of protein backbone phi,psi torsions from chemical shifts.

**SPARTA+**: prediction of protein backbone chemical shifts from structure.

**PROMEGA**: prediction of PRO cis-trans conformation from backbone chemical shifts.

**DYNAMO/PDBUTIL**: simple structure calculation, utilities to add protons, create extended structures, etc.

**DC**: manipulation of Dipolar Couplings and NMR Homology Search (MFR).

**AXES**: prediction of Small-Angle X-Ray Scattering from PDB.
**TALOS+** predicts protein backbone torsions from chemical shifts. It can form predictions for 88% of residues on average, with an RMS of better than 15 degrees.
Tensor and NMR Homology Search Utilities

• Rotate PDB onto Tensor Frame
• Molecular Fragment Replacement
• Amino Acid Type by Chemical Shift
• Create Extended Structure
• Add Protons
• Transformations of PDB Coordinates
• List Secondary Structure, H-Bonds
• Mass, Volume, Surface Area
• Simple simulated annealing
NMRPipe and VnmrJ
NMRPipe – the Old Way

NMRPipe Conversion Utility

Spectrometer Input: /dev/fs/C/space/delaglio/vj5/pipe_data
Output Template: /dev/fs/C/space/delaglio/vj5/pipe_data
Output Script: fid.com
Other Options: -noaswap

Input Protocol: Varian
Output Protocol: NMRPipe
Dimension Count: States
2D Mode: States
Temperature (K): 298.000

Total Points: 2048
Valid Points: 1024
Acquisition Mode: Complex
Spectral Width Hz: 13020.800
Observe Freq MHz: 799.597
Center Position PPM: 4.773
Axis Label: HN

Read Parameters  Save Script  Execute Script  Hide Script  Clear Script  Update Script  Quit  Help

Conversion Script Text

#!/bin/csh
var2pipe -in /dev/fs/C/space/delaglio/vj5/pipe_data_sets/HNCA_3D.fid/fid \
    -noaswap -aq2RD 1 \n    -xN 2048 -yN 64 -zN 64 \n    -xT 1024 -yT 32 -zT 32 \n    -xMODE Complex -yMODE Rance-Kay -zMODE Complex \n    -xSW 13020.800 -ySW 6031.720 -zSW 2100.600 \n    -xOB 799.597 -yOB 201.069 -zOB 81.032 \n    -xCAR 4.773 -yCAR 56.117 -zCAR 120.144 \n    -xLAB HN -yLAB 013 -zLAB N15 \n    -ndim 3 -aq2D States \n    -out /dev/fs/C/space/delaglio/vj5/pipe_data_sets/HNCA_3D.fid/data/test%03d.fid -verb -ov

sleep 5
NMRPipe – the Old Way

Manually Select Acquisition Mode
# NMRPipe – the Old Way

![NMRPipe Conversion Utility](image)

**NMRPipe Conversion Utility**

<table>
<thead>
<tr>
<th>Spectrometer Input:</th>
<th>/fid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output Template:</td>
<td>/data/test%03d.fid</td>
</tr>
<tr>
<td>Output Script:</td>
<td>fid.com</td>
</tr>
<tr>
<td>Other Options:</td>
<td>-noaswp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input Protocol:</th>
<th>Varian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output Protocol:</td>
<td>NMRPipe</td>
</tr>
<tr>
<td>Dimension Count:</td>
<td>States</td>
</tr>
<tr>
<td>2D Mode:</td>
<td></td>
</tr>
<tr>
<td>Temperature (K):</td>
<td>298.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x-axis</th>
<th>y-axis</th>
<th>z-axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Points:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid Points:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acquisition Mode:</td>
<td>Complex</td>
<td>Range Key</td>
</tr>
<tr>
<td>Spectral Width HZ:</td>
<td>13020.800</td>
<td>6031.720</td>
</tr>
<tr>
<td>Observe Freq MHz:</td>
<td>799.597</td>
<td>201.069</td>
</tr>
<tr>
<td>Center Position PPM:</td>
<td>4773</td>
<td>6117</td>
</tr>
<tr>
<td>Axis Label:</td>
<td>HN</td>
<td>C13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Manually Select Nucleus**
Manually Select Axis Label
Traditionally, NMRPipe scripts are manually edited to set parameter values.
#!/bin/csh

xyz2pipe -in _inName_ -x -verb \
| nmrPipe  -fn SOL          
| nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 2 -c __xC1__ 
| nmrPipe  -fn ZF           
| nmrPipe  -fn FT           
| nmrPipe  -fn PS -p0 __xP0_ -p1 __xP1_ -di 
| nmrPipe  -fn EXT -x1 __xEXTX1_ -xn __xEXTXN_ -sw 
| nmrPipe  -fn TP           
| nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 1 -c __yC1__ 
| nmrPipe  -fn ZF           
| nmrPipe  -fn FT           
| nmrPipe  -fn PS -p0 __yP0_ -p1 __yP1_ -di 
| nmrPipe  -fn TP           
| nmrPipe  -fn POLY -auto   
| pipe2xyz -out _auxName_  -x -to 0

xyz2pipe -in _auxName_ -z -verb \
| nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 1 -c __zC1__ 
| nmrPipe  -fn ZF           
| nmrPipe  -fn FT           
| nmrPipe  -fn PS -p0 __zP0_ -p1 __zP1_ -di 
| pipe2xyz -out _outName_  -z

Now, values determined in VnmrJ can be automatically inserted into any NMRPipe script
Automatically Create and Run NMRPipe Scripts
NMRPipe in VnmrJ

The VnmrJ “Do it All” Button
NMRPipe in VnmrJ

Strip display overview of processed result is created automatically
Strip display schemes for multiple spectra can be created automatically
CBCA(CO)NH
NUS Sampling Schedule
25% Density

$^{13}$C (10% Decay)

$^{15}$N (Uniform)
Signal Processing and NMRPipe
\[ X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt, \]
\[ X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \ldots, N - 1 \]
\[ X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} \, dt, \]

\[ X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \ldots, N - 1 \]

\[ \begin{array}{c|c}
  x & i^x \\
  \hline
  1 & i \\
  2 & -1 \\
  3 & -i \\
  4 & 1 \\
  5 & i \\
  6 & -1 \\
  7 & -i \\
  8 & 1 \\
\end{array} \]

\[ \exp(-i \, 2\pi ft) = \cos(2\pi ft) - i \sin(2\pi ft) \quad \ldots \quad \exp(i \, 2\pi ft) = \cos(2\pi ft) + i \sin(2\pi ft) \]
\[ X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} \, dt, \]

\[ X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \ldots, N - 1 \]
A One-Dimensional Absorption-Mode Time-Domain Signal

\[ x(t) = c \ w(t) \ \exp(2\pi iqt) + \varepsilon \]

\[ \exp(-2\pi at) \]

\[ \cos(2\pi qt) \]

\[ i \ \sin(2\pi qt) \]

\[ \varepsilon \]
\[ X(f) = \sum x(t) \left[ \cos \left( \frac{2\pi ft}{N} \right) - i \sin \left( \frac{2\pi ft}{N} \right) \right] \]
Time Domain Data to Fourier Transform
Time Domain (Real Part)

Fourier Term

Multiplied with Time Domain Data

Forms the Product

Sum Over Product to Form a Frequency Point:

Frequency Domain (Real Part)
/* Fourier transform of complex data tR,tI to produce fR,fI. */

void ft( float *tR, float *tI, float *fR, float *fI, int size )
{
    float vR, vI, twoPI;
    int    mid, k, n;

    twoPI = 4.0*acos( 0.0 );
    mid   = size/2;

    for( k = 0; k < size; k++ )           /* For every output freq point ... */
    {
        fR[k] = 0.0;
        fI[k] = 0.0;

        for( n = 0; n < size; n++ )       /* Sum over input times sinusoid. */
        {
            f  = twoPI*(k - mid)*n/size;

            vR = cos( f );
            vI = sin( f );

            fR[k] += tR[n]*vR - tI[n]*vI;
            fI[k] += tR[n]*vI + tI[n]*vR;
        }
    }
}
Truncated Time Domain Data to Fourier Transform
Time Domain (Real Part)

Fourier Term

Multiplied with Truncated Time Domain Data

Forms the Product

Sum Over Product to Form a Frequency Point:

Frequency Domain (Real Part)
**Time Domain**

Shifted Convolution Kernel

Multiplied With

Forms the Product

---

**Frequency Domain**

Fourier Transform of Kernel

Multiplied With

Forms the Product
**Time Domain**

Shifted Convolution Kernel

Multiplied With

Forms the Product

---

**Frequency Domain**

Fourier Transform of Kernel

Multiplied With

Forms the Product

---

Sum Over Product to Form a Convolution Point:
**Initial Domain**
Shifted Convolution Kernel

Multiplied With

Forms the Product

**Inverse Domain**
Fourier Transform of Kernel

Sum Over Product to Form a Convolution Point:
\[ X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} \, dt, \quad X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \ldots, N - 1 \]

\[ \exp(-i 2\pi ft) = \cos(2\pi ft) - i \sin(2\pi ft) \]

**First Point Acquisition Delay**
\[ X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} \, dt, \]

\[ X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i kn/N} \quad k = 0, \ldots, N - 1 \]

\[ \exp(-i 2\pi ft) = \cos(2\pi ft) - i \sin(2\pi ft) \]
$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} \, dt,$$

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \ldots, N - 1$$

$$\exp(-i 2\pi ft) = \cos(2\pi ft) - i \sin(2\pi ft)$$
Special Methods in Signal Processing
Linear Prediction

\[ q_1 \cdot a + q_2 \cdot b + q_3 \cdot c + q_4 \cdot d = e \]
\[ q_1 \cdot b + q_2 \cdot c + q_3 \cdot d + q_4 \cdot e = f \]
\[ q_1 \cdot c + q_2 \cdot d + q_3 \cdot e + q_4 \cdot f = g \]
\[ q_1 \cdot d + q_2 \cdot e + q_3 \cdot f + q_4 \cdot g = h \]
\[ q_1 \cdot e + q_2 \cdot f + q_3 \cdot g + q_4 \cdot h = i \]
Forward and Backward Coefficients Can be Averaged for Stability

Extending the Length of the Input Data Allows More LP Coefficients, Hence More Signals
xyz2pipe -in fid/test%03d.fid -x -verb            \  # Process the Directly-detected X-Axis.
| nmrPipe  -fn SOL                                    
| nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 2 -c 0.5    
| nmrPipe  -fn ZF -auto                                
| nmrPipe  -fn FT                                      
| nmrPipe  -fn PS -p0 43 -p1 0.0 -di                  
| nmrPipe  -fn EXT -x1 11.5ppm -xn 5.5ppm -sw          
| pipe2xyz -out lp/x%03d.ft1 -x

xyz2pipe -in lp/x%03d.ft1 -z -verb                    \  # Process the Indirectly-detected Z-Axis.
| nmrPipe  -fn SP -off 0.5 -end 0.95 -pow 1 -c 0.5     
| nmrPipe  -fn ZF -auto                                
| nmrPipe  -fn FT                                      
| nmrPipe  -fn PS -p0 0.0 -p1 0.0 -di                  
| pipe2xyz -out lp/xz%03d.ft2 -z

xyz2pipe -in lp/xz%03d.ft2 -y -verb                  \  # Linear Predict and Process the
| nmrPipe  -fn LP -fb -ord 12                          \  # Indirectly-Detected Z-Axis.
| nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 1 -c 1.0     
| nmrPipe  -fn ZF -auto                                
| nmrPipe  -fn FT                                      
| nmrPipe  -fn PS -p0 -135 -p1 180 -di                 
| pipe2xyz -out lp/xyz%03d.ft3 -y

xyz2pipe -in lp/xyz%03d.ft3 -z -verb                   \  # Inverse Process, Linear Predict,
| nmrPipe  -fn HT -auto                                 \  # and Re-Process the Z-Axis
| nmrPipe  -fn PS -inv -hdr                              
| nmrPipe  -fn FT -inv                                   
| nmrPipe  -fn ZF -inv                                  
| nmrPipe  -fn SP -inv -hdr                              
| nmrPipe  -fn LP -fb                                   
| nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5      
| nmrPipe  -fn ZF -auto                                 
| nmrPipe  -fn FT                                       
| nmrPipe  -fn PS -hdr -di                              
| pipe2xyz -out lp/test%03d.ft3 -z
Non-Uniform Sampling on a Uniform Grid
Non-Uniform Sampling: Skip a Fraction of the Points
Time Domain

NUS Time Domain Data

Non-Uniform Sampling
Non-Uniform Sampling:
for Fourier Transform, Replace Missing Points with Zeros
NUS Time Domain Data to Fourier Transform
*Time Domain*

Fourier Term

Multiplied with NUS Time Domain Data

Forms the Product

Sum Over Product to Form a Frequency Point:

*Frequency Domain*
Time Domain

NUS Time Domain Data

Frequency Domain
Time Domain

NUS Time Domain Data

Discrete Fourier Transform

Frequency Domain
Maximum Entropy Methods
Maximum Entropy Methods
Threshold Methods
Threshold Methods
Maximum Entropy Method

**Time Domain**
- Inverse-Transformed MEM
- Multiplied by Distortion
- Original Time-Domain
- Original - Distorted MEM

**Frequency Domain**
- Proposed MEM Spectrum
- MEM Spectrum with Distortion
- Original Spectrum
- Transformed Residual
- Residual with Threshold
- Exponential Amplification

![Graphs and plots](image)
Maximum Likelihood Frequency Map

Model: Fourier Term \((A)\)

Model Scaled by \(c\) to Minimize RMS \((\text{scale } c = \sum \frac{A B}{A A})\)

Subtracted from this Time Domain Data \((B)\)

Forms the Difference \((B - cA)\)

Find RMS to Form a Frequency Map Point:
Maximum Likelihood Frequency Map of NUS Data

Model: NUS-Adjusted Fourier Term  \((A)\)

Model Scaled by \(c\) to Minimize RMS  \((\text{scale } c = \sum \frac{A B}{A A})\)

Subtracted from NUS Time Domain Data \((B)\)

Forms the Difference \((B - cA)\)

Find RMS to Form a Frequency Map Point:
Fourier Transform vs Maximum Likelihood Frequency Map

Discrete Fourier Transform

Offset-Adjusted Maximum Likelihood Frequency Map

Frequency Domain
CBCA(CO)NH
NUS Sampling Schedule
25% Density

13C (10% Decay)

15N (Uniform)
CBCA(CO)NH 15N 38* Points, 13C 41* Points
Linearity – Reconstructed Peak Height vs Peak Height in FT Spectrum

MEM

ML
Spectral Matrix Decomposition
Matrix Decomposition: Each Multidimensional NMR Signal is the Product of 1D Vectors
NUS – Missing Information in a Deleted Point is Also Contained in the Same Row or Column
Spectral Matrix Decomposition by Principal Component Analysis (PCA)
Spectral Matrix Decomposition by Principal Component Analysis (PCA)
Structural Data from NMR
Structural Data from NMR

Chemical Shift

- **B₀**
- **downfield**
- **upfield**
- **TMS Signal**

- COOH
- Aldehyde Proton
- Benzene Protons
- Hydroxyl Proton (alcohol)
- Alkane Protons

External Magnetic Field Strength Scanned from high at 0 ppm to low field

J-Coupling

NOE Distance

J = Δ - B₀Cosφ + CC₀S²φ
Identify many H-H short range NOE distances
Supplement with torsions from J-Coupling values
Assume standard peptide geometry
Use simulated annealing to find a structure which matches distances
NOE distances are only qualitative.

A given peak might be the only evidence of an interaction.

A mis-assigned peak can be similarly problematic.
Alternate Approaches to NMR Structure

Chemical Shifts
Use chemical shifts directly rather than secondary shifts (SS).

For a given residue in the target:

   For a given residue type in the database:

      Compute chemical shift distance.
      Use Gaussian to estimate a P value

Find avg P value for residue type
Normalize over all residue types

<table>
<thead>
<tr>
<th>Residue Type Probability from Chemical Shifts</th>
</tr>
</thead>
<tbody>
<tr>
<td>DINI SS Method D81 Residues</td>
</tr>
<tr>
<td>50   14  4  2  3  (8)</td>
</tr>
<tr>
<td>61.7% 17.3% 4.9% 2.4% 3.7% 9.9%</td>
</tr>
<tr>
<td>DINI Direct Method 81 Residues:</td>
</tr>
<tr>
<td>64   11  3  2  1  0</td>
</tr>
<tr>
<td>79.0% 13.6% 3.7% 2.5% 1.2% 0%</td>
</tr>
<tr>
<td>FBP 170 residues:</td>
</tr>
<tr>
<td>129  20  10  6  2  (3)</td>
</tr>
<tr>
<td>75.9% 11.8% 5.9% 3.5% 1.2% (1.8%)</td>
</tr>
<tr>
<td>V-Alpha 114 residues:</td>
</tr>
<tr>
<td>66   18  14  2  3  (11)</td>
</tr>
<tr>
<td>57.9% 15.8% 12.3% 1.8% 2.6% (9.6%)</td>
</tr>
<tr>
<td>Gamma Crystalin 170 residues:</td>
</tr>
<tr>
<td>87   24  20  12  8  (19)</td>
</tr>
<tr>
<td>51.2% 14.1% 11.8% 7.1% 4.7% 11.2%</td>
</tr>
</tbody>
</table>
Subtract Residue-Specific *Random Coil Shift* to form *Secondary Shift*
Match database triplet with target, based on sum-of-squares difference in chemical shifts, plus residue type homology term.

Use central residue as predictor of phi and psi.
The SPARTA Program of Shen and Bax ...

\[ \text{Δδ}_i = \delta_i - \delta_{i, \text{coil}} - \delta_{i, \text{prev}} - \delta_{i, \text{next}} - \delta_{i, \text{ring}} \]

Predicted chemical shifts \( \delta_{r,i} \) for residue \( i \)

\( r = ^{15}\text{N}, ^{1}\text{H}^\text{N}, ^{1}\text{H}^\alpha, ^{13}\text{C}^\alpha, ^{13}\text{C}^\beta, ^{13}\text{C}^\gamma \)
Alternate Approaches to NMR Structure
Residual Dipolar Couplings
Alignment by Liquid Crystal

\[ D_{ij} = D_{\text{max}} \left[ D_a (3 \cos^2 \theta - 1) + \frac{3}{2} D_r \sin^2 \theta \cos 2\phi \right] \]

\[ D_{\text{max}} = -\mu_0 (\hbar/2\pi) \gamma y \Phi (4\pi r^2 r^3_{ij}) \]
$^1$H-$^{15}$N HSQC spectra of ubiquitin

isotropic

5% w/v bicelles

8% w/v bicelles

Alignment: 0 % 
0.1%
0.16%
\[ D_{ij} = D_{\text{max}} \left[ D_a(3\cos^2 \theta - 1) + \frac{3}{2} D_r \sin^2 \theta \cos 2\phi \right] \]

\[ D_{ab} = D_{\text{max}} \left[ (s1) \cdot 0.5 \cdot (3.0 \cdot z_{IJ} \cdot z_{IJ} - 1.0) \right. \]
\[ + (s2) \cdot 0.5 \cdot (x_{IJ} \cdot x_{IJ} - y_{IJ} \cdot y_{IJ}) \]
\[ + (s3) \cdot 2.0 \cdot x_{IJ} \cdot y_{IJ} \]
\[ + (s4) \cdot 2.0 \cdot x_{IJ} \cdot z_{IJ} \]
\[ + (s5) \cdot 2.0 \cdot y_{IJ} \cdot z_{IJ} \left. \right] \]
• Search PDB for small fragments whose simulated dipolar couplings and shifts match the observed values.
• Use the fragment information to reconstitute larger structural elements.
• Also: Sequential NOEs, J values, etc
• Nucleic Acid Applications

1ubq.pdb  2  - 17
1bii.pdb   189 - 204
1cel.pdb  15  - 30
1gtm.pdb  40  - 55
Initial Structure from Average Phi and Psi of Fragment Ensemble

1ubq vs MFR phi/psi refined structure
MFR Estimation of Tensor Parameters

- Magnitude
- Rhombicity
- Orientation (Euler Angles)
MFR Fragment Tensor Magnitudes Reveal Dynamics
Gamma S

177 Residues, two similar domains, homologous structure is known.

179 Amide-Amide NOEs, 70 Methyl-Methyl NOEs, including 6 inter-domain

DC Medium 1: 144 HN-N, 111 CA-CB, 150 CA-C’, 134 N-C’

DC Medium 2: 147 HN-N, 135 CA-CB, 153 CA-C’, 139 N-C’

Side-chain $\chi_1$ angles from $^3J_{NC\gamma}$ and $^3J_{C'C\gamma}$ couplings, $\chi_2$ from $^3J_{C\gamma}C\delta$
• Conduct MFR Search with SVD (free tensor)
• Conduct second MFR Search with fixed tensor Da, Rh, and relative orientation
• Refine all fragments with fixed tensor Da, Rh to yield Phi and Psi for 90% of residues; 50% have better than 5 degree RMS consensus; 33% are 3 degree RMS or better.
• MFR Torsions Preserve Secondary Structure During High Temperature Phase  
• During Cooling, MFR Torsion Restraint Force Constant is Decreased  
• DC Force Constant is Increased as Ideal Fold is Approached

dynReadGMC -gmc $gmcDir -pdb $pdbName

for {set i 1} {$i <= $count} {incr i} \
{
    dynSimulateAnnealing -graph -print 50 -rasmol 500 \
    -sa stepCount init 100 \
    stepCount high 24000 \
    stepCount cool 8000 \
    timeStep all 3 \
    temperature all 4000 \
    temperature coolEnd 0 \
    -fc dc coolEnd 2.0 \
    torsion all 50 \
    torsion coolEnd 10 \
    noe all 25 \
    noe coolEnd 100 \
    radGyr all 0.0

    set outName [format $outTemplate $i]

    dynWrite -pdb -src $dynInfo(gmc,pdb) -out $outName -rem $dynInfo(energyText)
    dynRead -pdb -src $dynInfo(gmc,pdb) -in $pdbName

    incr iseed 111
    srand $iseed
}
Backbone RMSD $\gamma$S(NMR) and $\gamma$B-crystallin (X-ray)

C-terminal domain: 1.09 A  
N-terminal domain: 0.63 A
Consistent blind protein structure generation from NMR chemical shift data

*Proc Natl Acad Sci USA, (2008) 105, 4685-4690*

Yang Shen
Oliver Lange
Frank Delaglio
Paolo Rossi
James M. Aramini
Gaohua Liu
Alexander Eletsky
Yibing Wu
Kiran K. Singarapu
Alexander Lemak
Alexandr Ignatchenko
Cheryl H. Arrowsmith
Thomas Szyperski
Gaetano T. Montelione
David Baker
Ad Bax
Using SPARTA Chemical Shift Prediction to Improve ROSETTA Scoring Function

- Original ROSETTA Energy Score
- CS-ROSETTA Energy Score
### CS-ROSETTA performance on nine structural genomics proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Number of Residues</th>
<th>PDB ID</th>
<th>RMSD Å (backbone)</th>
<th>RMSD Å (all)</th>
<th>% NOE Peak Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RpT7</td>
<td>65</td>
<td>2jtv</td>
<td>0.64</td>
<td>1.29</td>
<td>69</td>
</tr>
<tr>
<td>StR82</td>
<td>69</td>
<td>2jt1</td>
<td>0.57</td>
<td>1.14</td>
<td>65</td>
</tr>
<tr>
<td>RhR95</td>
<td>72</td>
<td>2jvm</td>
<td>0.66</td>
<td>1.18</td>
<td>55</td>
</tr>
<tr>
<td>NeT4</td>
<td>73</td>
<td>2jv8</td>
<td>0.70</td>
<td>1.42</td>
<td>57</td>
</tr>
<tr>
<td>TR80</td>
<td>78</td>
<td>2jxt</td>
<td>0.69</td>
<td>1.27</td>
<td>67</td>
</tr>
<tr>
<td>VfR117</td>
<td>80</td>
<td>2jvw</td>
<td>0.60</td>
<td>1.40</td>
<td>37</td>
</tr>
<tr>
<td>PsR211</td>
<td>100</td>
<td>2jva</td>
<td>2.07</td>
<td>2.34</td>
<td>57</td>
</tr>
<tr>
<td>AtR23</td>
<td>101</td>
<td>2jya</td>
<td>1.10</td>
<td>1.81</td>
<td>60</td>
</tr>
<tr>
<td>NeR45A</td>
<td>147</td>
<td>2jxn</td>
<td>2.03</td>
<td>2.85</td>
<td>53</td>
</tr>
</tbody>
</table>

![StR82](image1.png)  ![PsR211](image2.png)  ![NeR45A](image3.png)
Structures of two designed proteins with high sequence identity

NMR structures of Ga88 and Gb88

NMR structures vs csRosetta models

Patrick A. Alexander, Yanan He, Yihong Chen, John Orban, and Philip N. Bryan
PNAS, 2007, 104:11963-11968
PNAS, 2008, 105:14412-14417

Mean-to-mean backbone RMSD
1.31Å
1.07Å
NMR Applications in Drug Discovery
Applications of NMR in the Drug Discovery Process

NMR Spectral Series:
Two Approaches

SAR by NMR (Abbott Labs)

Observe Ligand Signals

Observe Protein Signals

-binding Surface
-screen
-optimize
-link $K_{d_{total}} = K_{d_1} \times K_{d_2}$
Analyze Titration Curve to Estimate $K_d$
Entire spectrum is a single object in multidimensional space.

Coordinates of the object are the spectral intensities.

Similar spectra cluster together.

Spectra with similar features lie along lines and curves.
Correct Structure

Mis-Folded Structure with Secondary Structure Conserved
Useful Graphics Strategies
Graphics Strategies of Edward Tufte

www.edwardtufte.com

Above all, Show the Data
Show Cause and Effect
Represent Data and Scale Faithfully
Maximize “Data Ink” and Data Density, Minimize “Chart Junk”
Shrink Graphics - Integrate Text, Values, and Graphics
Be Multivariate
Use Layers – Use Macro and Micro Interpretations - Clarify by Adding Detail
Conserve Color Space
Use Small Multiples
Find Ways to Show All of the Data
Treat Design as a Solved Problem, then Find the Best Examples
Microsoft PowerPoint “with new and improved tools for video and photo editing, dramatic new transitions, and realistic animation, you can add polish to presentations that will captivate your audience.” - from microsoft.com

Adobe Photoshop “redefines digital imaging with powerful new photography tools and breakthrough capabilities for complex image selections, realistic painting, and intelligent retouching.” - from adobe.com

In the content-creation arena, amazon.com lists Microsoft Office and Adobe Photoshop as highest-sellers. It could be claimed, Photoshop is about the content, while PowerPoint is often about the scaffold …
Graphic Design and PowerPoint Abuse: ordered and unordered lists

We rely on PowerPoint to communicate. But there are many enticements to abuse PowerPoint. Beautiful 3D graphics can be used to decorate a presentation and help unify its contents. But in some cases, presentation graphics are used to hide lack of content, or in the worst case, to disguise or misrepresent data. The presentation graphics themselves often have no actual relation to the information being conveyed – they are just ways of dressing up a list or a sequence.

It's useful for us to be able to detect this sort of problem when we see it, avoid this in our own presentations, and avoid it in the User Interface design of our software.

“Imress your audience with these professional and pre-designed 3D PowerPoint Graphics” - from presentationload.com
Avoiding inappropriate presentation modes and unneeded 3D effects

**Microsoft PowerPoint** – “as you can see, the blue-greenish quarter was about the same as the blue one.”

There are no numbers given, and the trend over time, which shows falling sales, is hidden. Also, the 3D perspective might actually distort the apparent values:

**Apple Keynote** – “as you can see, our maple year was only 1% different than our teak year.”

The segments are at least nicely labeled with numbers, but the wood textures, while really cool, are impossible to decipher. Likewise, the trend with time is hidden, and 3D perspective potentially distorts the apparent values.

The wrong graphical paradigm can make data hard to interpret, or even misleading.
Morton Thiokol engineers debated the problem of O-ring failure due to low temperature for several hours the night before the launch, and made the company’s only no-launch request in 12 years. Their presentation of evidence did not convince NASA management. The shuttle blew up 73 seconds after ignition.
### HISTORY OF O-RING EROSION ON SRM FIELD JOINTS

<table>
<thead>
<tr>
<th>Cross Sectional View</th>
<th>Top View</th>
<th>Docking Location (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SRM No.</strong></td>
<td><strong>Max Erosion (In.)</strong></td>
<td><strong>Length Of Erosion (In.)</strong></td>
</tr>
<tr>
<td><strong>Erosion Depth</strong> (In.)</td>
<td><strong>Erosion Perimeter</strong> (In.)</td>
<td><strong>Nominal Diia.</strong> (In.)</td>
</tr>
<tr>
<td>4121-40 Center Field***</td>
<td>71A</td>
<td>None</td>
</tr>
<tr>
<td>4121-40 Center Field***</td>
<td>22A</td>
<td>None</td>
</tr>
<tr>
<td>5102-70 Forward Field***</td>
<td>35A</td>
<td>0.010</td>
</tr>
<tr>
<td>5102-70 Center Field (prim)***</td>
<td>155</td>
<td>0.008</td>
</tr>
<tr>
<td>5102-70 Center Field (sec)***</td>
<td>155</td>
<td>None</td>
</tr>
<tr>
<td>4121-40 Forward Field</td>
<td>31B</td>
<td>0.020</td>
</tr>
<tr>
<td>4121-40 Alt Field*</td>
<td>11A</td>
<td>None</td>
</tr>
<tr>
<td>4121-40 Forward Field</td>
<td>10A</td>
<td>0.040</td>
</tr>
<tr>
<td>4121-40 Alt Field</td>
<td>7A</td>
<td>0.055</td>
</tr>
</tbody>
</table>

---

*That gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

**OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLE IN PUTTY AND NO Soot NEAR OR BEYOND THE PRIMARY O-RING.**

**SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO Soot BLOWHOLE. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLE IN PUTTY.**
History of O-Ring Damage in Field Joints

O-Ring Temp (°F)

Development Motor Number
1 2 3 4 5

O-Ring Temp (°F)

Qualification Motor Number
1 2 3 4

Code

S = Heating of Secondary O-Ring
B = Primary O-Ring Blowby
E = Primary O-Ring Erosion
H = Heating of Primary O-Ring

= No Damage

STATIC TEST MOTORS

• HORIZONTAL ASSEMBLY
• SOME PUTTY REPAIRED
O-ring damage index, each launch

26°–29° range of forecasted temperatures (as of January 27, 1986) for the launch of space shuttle Challenger on January 28

Temperature (°F) of field joints at time of launch
Napolean’s Route: 422,000 Men to 10,000 Men, Five Dimensions
UE de la température en degrés du thermomètre de Réaumur au dessus de

Bax Group Figure: 18 values
Weather Statistics: 1,800+ Values, Four Variables, Notations
The diagram shows the relationship between atomic number and atomic volume. The labels include Li (3), Na (11), K (19), Rb (37), Cs (55), and Fr (87). The dots represent the positions of elements, with the rare earths indicated by the dashed line.