Optimising resolution and sensitivity in 3D-4D NMR:

Multi-Dimensional decomposition (MDD) of non-linearly sampled spectra

Vladislav Orekhov
Outline

- Multidimensional Decomposition (MDD)
- Prehistory (MUNIN): application of the model to uniformly sampled 3D data. Resolving overlapped signals in NOESY’s and relaxation spectra
- General validity of the MDD model for the NMR spectroscopy
- Saving of NMR measurement time in structural proteomics
- Improving resolution in 4D NOESY spectra of large proteins
MDD Model: Spectral components (e.g. peaks) can be written as a direct products of 1D shapes

$$| G \cdot [S - \Sigma_{\beta} (a^{\beta} F_1^{\beta} \otimes F_2^{\beta} \otimes F_3^{\beta})] |^2$$

MDD approximate 3D spectrum $S$ by a sum of components, i.e. minimize the expression above

Components (3D): $a^{\beta} F_1^{\beta} \otimes F_2^{\beta} \otimes F_3^{\beta}$
Shapes (1D): $F_1, F_2, F_3$
$G$: sparse data; $G_{ijk} \in \{0, 1\}$

e.g. $^{15}$N NOESY-HSQC

component 1
component 2

(a) Reconstruction. 27 component
(b) Regular FT transform
(c) Reconstruction. Component 1
(d) Reconstruction. Component 2
15N-HSQC-NOESY: structural constraints

- one component - NOEs to one HN
- NOEs result from 1D peak picking
- distance constraints from TWD are correct and complete

Short distances identified by method B among 1000 most likely from method A

<table>
<thead>
<tr>
<th>proton</th>
<th>distance</th>
<th>shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>H46 Hα</td>
<td>2.3</td>
<td>7.29</td>
</tr>
<tr>
<td>H46 Hβ</td>
<td>2.9</td>
<td>3.30, 2.36</td>
</tr>
<tr>
<td>N47 Hα</td>
<td>3.0</td>
<td>4.74</td>
</tr>
<tr>
<td>L86 CH₃</td>
<td>3.1</td>
<td>1.00, 1.11</td>
</tr>
<tr>
<td>C112 Hα</td>
<td>3.1</td>
<td>5.76</td>
</tr>
<tr>
<td>C112 Hβ</td>
<td>3.2</td>
<td>3.41, 2.85</td>
</tr>
<tr>
<td>N47 Hβ</td>
<td>3.4</td>
<td>2.65, 2.41</td>
</tr>
<tr>
<td>W113 HN</td>
<td>3.7</td>
<td>10.03</td>
</tr>
<tr>
<td>F114 HN</td>
<td>4.1</td>
<td>10.53</td>
</tr>
<tr>
<td>L86 Hα</td>
<td>4.2</td>
<td>4.75</td>
</tr>
<tr>
<td>L33 Ch₃</td>
<td>4.3</td>
<td>0.99, 1.06</td>
</tr>
<tr>
<td>W48 HN</td>
<td>4.3</td>
<td>7.47</td>
</tr>
<tr>
<td>H46 HN</td>
<td>4.4</td>
<td>8.29</td>
</tr>
</tbody>
</table>

(not assigned: F15 Hζ, H46 Hδ1)
Relaxation: **accurate rates from overlapped signals**


Relaxation rates for all 337 assigned residues from MBP (370 residues):
reduction of errors for relaxation rates is statistically significant

Separation of strongly overlapped peaks (red dots in green box):
clean deconvolution (red and blue components)
independent relaxation rates
Theory: general applicability in N-D NMR

\[ S(\tau_1, \tau_2, \tau_3) = \langle O^\dagger | \sigma(\tau_1, \tau_2, \tau_3) \rangle = \langle O^\dagger | P(\tau_1, \tau_2, \tau_3) | \sigma(0) \rangle \]

\[ = \langle O^\dagger | E_3(\tau_3) \cdot M_3 \cdot E_2(\tau_2) \cdot M_2 \cdot E_1(\tau_1) \cdot M_1 | \sigma(0) \rangle \]

\[ = O_\alpha \cdot E_{3,\alpha,\beta} \cdot M_{3,\beta,\gamma} \cdot \exp[-E_2 \cdot \tau_2]_{\gamma,\gamma} \cdot M_{2,\gamma,\delta} \cdot E_{1,\delta,\epsilon} \cdot M_{1,\epsilon,\nu} \cdot \sigma(0)_{\nu} \]

\[ = F_{1,\gamma} \cdot F_{2,\gamma} \cdot F_{3,\gamma} \]

Assumptions

\[ E_2(t_2) = \exp[-E_2 \cdot \tau_2], \quad E_2 - \text{mean Liouvillian} \]

\[ E_2 - \text{diagonal in some basis set} \]
3D 1H NOE-NOE

General validity of the TWD model for N-D NMR
Reconstruction using 4 components
Difference between the reconstructed and original spectrum
Non-uniform (sparse) sampling allows optimization of resolution and sensitivity.

Resolution \( \sim t_{acq} \)

Uniform sampling
FFT, LP

Non-uniform sampling
MDD, MEM
Sparse detection and processing of 3D spectra

Sparse 3D FID

1D shapes in time and frequency

comp. 1: $\omega_1$, $\omega_2$, $\omega_3$

comp. 2: $\omega_1$, $\omega_2$, $\omega_3$

comp. N: $\omega_1$, $\omega_2$, $\omega_3$

Complete spectrum (sum of all components)
3D $^1$H-$^{15}$N NOESY-HSQC

$^1$H$_{(\text{NOE})}$ [$\tau_1$] - 160 complex

$^{15}$N [$\tau_2$] - 44 complex

$^1$H$_N$ [$\omega_3$] - 24 real (8.67-8.9 ppm)

25% random selection in t1, t2 with 2D exponential probability distribution
Reference vs. sparse. Representative region from the 3D spectrum

Reference

25% sparse, 27 comp.
Reference vs. sparse. Difference

Difference: 25% sparse vs. reference

25% sparse, 27 comp.
Reference vs. LP. 25% points in the 15N dimension

Reference

LP, extension 6 times
Proteomics: rapid spectra acquisition of NMR spectra


These methods improve resolution in *sample limiting* case. Sensitivity is at best preserved for a given measurement time.
13C 3D NOESY spectrum of EC0298 (69 aa)
(Prof. C Arrowsmith)

1H-13C correlation spectrum

Estimating number of components

**AA sequence:** MNKDEAGGNWKQFKGKVKEQWGKLTD DDMTIIEGKRDQLVGKIQEYRYGYQKDQAEKEVVDWETRN EYRW

<table>
<thead>
<tr>
<th></th>
<th>CH</th>
<th>CH2</th>
<th>CH3</th>
<th># in seq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>2</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>3</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td>4</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>R</td>
<td>1</td>
<td>3</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Y</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>H</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Q</td>
<td>1</td>
<td>2</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>W</td>
<td>1</td>
<td>1</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

Total: 74 114 25 69

Minimum number of components needed: 213
Maximum number of components needed: 327
Number of components needed: 270
Full data set: performance vs # components

Reference spectrum

NC=230
\( \lambda=0.005 \)

NC=250
\( \lambda=0.001 \)

NC=270
\( \lambda=0.005 \)

NC=300
\( \lambda=0.005 \)
EC02098 sparse 27% data set (NC=270) : performance vs $\lambda$ setting

Reference spectrum

$\lambda=0.01$
$I=1.7k$

$\lambda=1e^{-3}$
$I=2k$

$\lambda=1e^{-4}$
$I=2k$

$\lambda=1e^{-5}$
$I=1k$

$\lambda=1e^{-6}$
$I=1k$
Peak integrals are well reproduced in the sparse spectrum of EC02098 (only 27% of the FIDs used)

$\omega_3$ ($^1H$) region
2.46 .. -0.05 ppm

2000 strongest peaks from the sparse vs complete reference 3D 13C-NOESY
Improving resolution in 4D NOESY spectra of large proteins

malate synthase G (81.4 kDa)

sparse 4D 13C Methyl HMQC NOESY

(Prof. L Kay)
4D 13C Methyl HMQC NOESY malate synthase (81.4 kDa)

4D FID, sparse 30.8%.

Acquisition times & SW:
- $^1\text{H}$ 27 ms, 960 Hz
- $^{13}\text{C}$ 22 ms, 2211 Hz
- $^{13}\text{C}$ 16 ms, 2211 Hz

Out of 359424 FIDs 110592 were randomly selected for detection.

6.5 instead of 21 days of measurements
4D resolves ambiguities in 3D spectra

$4D \ 13C$ Methyl HMQC NOESY of malate synthase

$13C$ 3D

sparse 4D

$1H$ 3D
Conclusions: MDD features

- No loss of sensitivity, high dynamic range of signals
- Non-uniform sampling in the time domain is possible; this can be used to achieve optimal resolution and sensitivity.
- General validity of the model for multidimensional NMR data sets. Few components describe complex line shapes
- Promising method for compressed spectra presentation (processing, analysis, visualization, storage, etc.)
- Either frequency or time domain data can be used.
- Any unique property of a signal can contribute to the resolution of the components, e.g. line position, J-coupling, relaxation rate, NOE or TOCSY cross-peak patterns, etc.
Acknowledgments

Structural Proteomics
C Arrowsmith (NorthEast Structural Genomics Consortium)
A Yee (NorthEast Structural Genomics Consortium)
T Luan (SNC Göteborg)

Large Proteins (malate synthase G)
V Tugarinov & L E Kay (University Toronto)

MUNIN
M. Billeter (GU Göteborg)
D Korzhnev (GU & SNC Göteborg)
A Gutmanas, (GU Göteborg)
P Jarvoll (SNC Göteborg)

Math & Software
I. Ibragimov (SU, Germany)