Recent Advances in Biomolecular NMR

Lucia Banci
CERM – University of Florence
Recent Advances in Biomolecular NMR

• Protonless NMR
  for the characterization of Unfolded proteins, Large protein assemblies, Paramagnetic systems

• In cell NMR
  For studying biomolecules in a cellular context

• Combination of Solution and Solid State NMR
  For characterization of dynamic proteins and large aggregates

• Mechanistic Systems Biology
  To describe and understand biological processes at molecular level
Why *protonless* NMR?

Inverse (i.e. through $^1$H) detection of heteronuclei was a major advancement!!

Properties of $^1$H (high $g_H$, ..)

- ☺ high $^1$H sensitivity
- ☺/☹ large dipolar interactions
- ☺/☹ efficient relax processes 
  (paramagnetic and large)
- ☹/☹ relatively low chemical shift dispersion 
  (unfolded systems)
$^{13}\text{C}$ direct detection

…with increase in sensitivity, (high $B_0$, cryo!)

direct detection of heteronuclei (low $\gamma$ nuclei) becomes accessible

Isotopic enrichment necessary anyway

$^{13}\text{C}$ direct detection is a complementary tool
$^{13}\text{C}$ direct detection, protonless NMR

A complementary tool for challenging systems

- paramagnetic proteins
- very large proteins
- parts of proteins affected by exchange processes
- unfolded systems
- high salt concentrations
Set of *exclusively heteronuclear* experiments based on C´ and Cα detection for sequence specific assignment of a protein.
C’ direct detection - IPAP

Set-up on $^{13}$C-$^{15}$N labeled Alanine

Info on the splitting!! © RDC!!!
C′ direct detection – CON-IPAP

Transfer pathway: \( F_1(CO) \rightarrow F_3(N,t_1) \rightarrow F_1(CO,t_2) \)

Correlations observed: \( N_i - C'_i - 1 \)

The delays are: \( \Delta = 9 \text{ ms}, \Delta_1 = 25 \text{ ms}, \varepsilon = t_1(0) \). The phase cycle is: \( \phi_1 = x,-x; \phi_2 = 2x,2(-x); \phi_3 = 4x,4(-x); \phi_{IPAP}(IP) = x; \phi_{IPAP}(AP) = -y; \phi_{rec} = x,(-x),x,(x),(-x),(x),x \). Quadrature detection in the \( F_1 \) dimension is obtained by incrementing \( \phi_1 \) in a States-TPPI manner.
C$^\text{13}$ direct detection – CON-IPAP

CON-IPAP

600 MHz
Prototype cryoprobe optimized for $^{13}$C sensitivity
(S/N 1400:1)

Reduced monomeric SOD (15 kDa)

161 out of the 163 expected correlations are resolved

C′ direct detection – CACO-IPAP

Transfer pathway: $F1(C^\alpha, t_1) \rightarrow F1(CO, t_2)$
Correlations observed: $C^\alpha_i - C^{\prime}_i$

CACO-IPAP - The delays are: $\Delta = 9$ ms. The phase cycle is: $\phi_{IPAP} (IP) = x, -x$ and $\phi_{rec} = x, -x$; $\phi_{IPAP} (AP) = -y, y$ and $\phi_{rec} = x, -x$. Quadrature detection in the $F_1$ dimension is obtained by incrementing $\phi_1$ in a States-TPPI manner.
**C´ direct detection –CBCACO-IPAP**

Transfer pathway: \( F_1(C^{\alpha/\beta}, t_1) \rightarrow F_1(C^\alpha, t_2) \rightarrow F_1(CO, t_3) \)

Correlations observed: \( C^\beta_i - C^\alpha_i - C´_i, C^\alpha_i - C^\alpha_i - C´_i \)

CBCACO-IPAP - The delays are: \( \Delta = 9 \text{ ms}, \Delta_1 = 8 \text{ ms} \). The phase cycle is: \( \phi_1 = x, -x; \phi_2 = 8x, 8(-x); \phi_3 = 2y, 2(-y); \phi_{\text{IPAP}(IP)} = 4(x), 4(-x); \phi_{\text{IPAP(AP)}} = 4(-y), 4(y); \phi_{\text{rec}} = x, (-x), (-x), x, x, (-x) \). Quadrature detection in the \( F_1 \) and \( F_2 \) dimensions is obtained by incrementing \( \phi_1 \) and \( \phi_3 \) in a States-TPPI manner.
C´ direct detection – S³E

Transfer pathway: \( F_1(C_{\text{ali}}, t_1) \rightarrow F_1(C^\alpha, t_2) \rightarrow F_1(CO, t_3) \)

Correlations observed: \( C_{\text{ali}} - C^\alpha_i - C'_i, C^\alpha_i - C^\beta_i - C'_i \)

\( \delta(^{13}C') \)

The delays are: \( \Delta = 9\text{ms}, \epsilon = t_1(0) \). The phase cycle is: \( \phi_1 = x, -x; \phi_2 = 2x, 2(-x); \phi_{\text{IPAP}}(IP) = 4x, 4(-x); \phi_{\text{IPAP}}(AP) = 4(-y), 4y; \phi_{\text{rec}} = x, (-x), x, (-x), x, x, (-x). \) Quadrature detection in the \( F_1 \) and \( F_2 \) dimensions is obtained by incrementing \( \phi_1 \) and \( \phi_2 \) respectively in a States-TPPI manner.
C’ detection - Assignment strategy

Spin system identification
CACO, CBCACO, CCCO-IPAP

600 MHz Cryoprobe optimized for $^{13}$C sensitivity (S/N 1400:1)

16 scans 2-3.5 hours

the majority of the $^{13}$C spin systems could be assigned


C' detection - Assignment strategy

Sequential assignment
C´ detection - Assignment strategy

CBCACON-IPAP  

CCCON-IPAP

@600 MHz CPTXO (S/N 1400:1) on 1.5 mM $^{13}$C, $^{15}$N labeled reduced monomeric SOD. CBCACON-IPAP, 16 scans, 3 days, CCCON-IPAP, 32 scans, 4.5 days.

96 % of the $^{13}$C resonances could be identified

C’ detection - Assignment strategy

PRO 74

LYS 75

One of the powerful applications of $^{13}\text{C}$ direct detection NMR

Intrinsically disordered proteins - IDPs!
... Reduction in $^1$H chemical shifts

Cu(I)Zn(II)SOD 153 AA
Well folded

Synuclein 140 AA
IDP
$^{13}$C carbonyl direct detection – IDPs

CON of intrinsically unfolded $\alpha$-synuclein

All residues assigned
$(N, C', C^\alpha, C^\beta)$

Prolines are visible

Intrinsically unfolded α-synuclein

Sequence specific assignment
3D CBCACON-IPAP
3D COCON-IPAP

Securin – Intrinsically disordered protein

Securin
Intrinsically disordered protein (IDP!)
202 AA (>10% PROs)
Intrinsically unfolded human securin

Securin – 202 AA, 24 PRO

Oberved well resolved peaks:

**HSQC:** 122
68% of the expected
60% of the whole protein

**CON:** 165
82% of the expected
82% of the whole protein

Intrinsically unfolded human securin

193, out of the 201 expected, spin patterns are identified (96%) in CBCACON-IPAP.

Assignment and chemical shift analysis of securin

α-helical secondary structure propensity for the stretch  D^{150}-F^{159}

Human securin - other NMR observables

D\textsuperscript{150}-F\textsuperscript{159}, E\textsuperscript{113}-S\textsuperscript{127} and W\textsuperscript{174}-L\textsuperscript{178}

Can one implement all the tricks to reduce experimental time?

- Longitudinal relaxation enhancement
- Reduction in datapoints acquired in indirect dimensions
- Decrease the recycle delay
**13C direct detection – Speeding up**

Longitudinal relaxation enhancement

\(^1\)H-start, \(^1\)H-flip


Reduction in datapoints acquired in indirect dimensions

13C direct detection – *Speeding up*

(Hflip)CACO-IPAP  (Hflip)CANCO-IPAP

In-cell NMR spectroscopy
In-cell NMR spectroscopy

• In-cell NMR allows the characterization of biomolecules inside living cells.

• It relies on high resolution NMR experiments to obtain information at atomic resolution on biomolecule structure, folding and interactions.

• It has a high biological relevance, as the biomolecules are monitored in a cellular environment.
In-cell NMR: what’s new?

• With traditional *in vivo* NMR, few small metabolites were monitored in living cells. With high resolution in-cell NMR a biomacromolecule, such as a protein is characterized at atomic level.

• To overcome the sensitivity limit of NMR, the protein needs to be concentrated, often above the physiological levels (that is why “in-cell” and not “*in vivo*”, but the cells are alive!).

• Uniform or selective *isotopic labelling* of the molecule of interest with NMR-active nuclei is needed to record heteronuclear multidimensional NMR experiments.
Effects of the cellular environment

The cellular environment can have protein-specific and/or general, non-specific effects

Non-specific effects:

- The *microscopic viscosity* slows down the tumbling rate of the protein, increasing its rotational correlation time $\tau_c$, and increasing the linewidth of the protein NMR signals.
- Weak interactions with small cellular components have the same effect.
- Crowding and ionic strength can cause small chemical shift differences between in-cell and *in vitro*.

Protein-specific effects:

- Some proteins can interact strongly with high molecular weight cellular components (e.g. DNA, heat-shock proteins, folding machineries, cell membranes).
- This is a protein-dependent effect, which can cause line broadening up to the complete loss of the NMR signals.

Different living organisms

• Different cells have been and are used: bacteria, oocytes and mammalian cells. Different techniques are exploited to obtain high protein concentration: overexpression and injection/insertion.

• Prokaryotic cells are more commonly used. Indeed, the bacterial cytoplasm is a good model of the eukaryotic one, in terms of molecular crowding, pH and redox potential.

• Eukaryotic cells have been used to monitor protein interactions with specific cellular components, such as kinases. They have the machineries and chaperones for the correct maturation of eukaryotic proteins.
Prokaryotes: *E. coli* cells

NMR in *E. coli* cells allows to investigate the folding state of a nascent protein, to monitor maturation steps which do not require specific chaperones, to have insights on interactions with the environment, and to investigate the structural features and interaction network of a protein,

Interactions can be studied by sequentially expressing two or more proteins, with only one labelled.

- **Pros**: fast growing, easy to handle, good overexpression, ease of labelling.

- **Cons**: no cellular compartments, no machineries for protein maturation, difficult to insert external proteins or small molecules.
**Protein expression in *E. coli* cells**

*E. coli* cells allow for overexpression of the protein of interest. The strategy to obtain selective isotopic labelling of the protein consists of switching the culture medium before the induction of the protein with a medium containing isotopically labelled nutrients.

In cell NMR on eukaryotic cells allows the characterization of an eukaryotic protein in its true environment, and therefore it is more physiologically relevant.

The effects of post-translational modifications and localization in different cellular compartments can be monitored directly.

**Pros:** more physiologically relevant, correct protein maturation occurs, sub-cellular localization of the protein can be monitored.

**Cons:** slower to grow, more difficult to handle than bacterial cells, difficult to obtain a high amount of protein inside the cells.
Two kinds of cells have been used: mammalian cells strains (e.g. CHO, HeLa), and *Xenopus laevis* oocytes.

For mammalian cell cultures, protein insertion is achieved by using cell-penetrating peptides to deliver a fusion protein, or porins to permeabilize the cells.

To insert the protein into *X. laevis* oocytes the microinjection technique is used.

---

Fig. From: D.S. Burz, A. Shekhtman, *Nature*, 458 (2009).


Virtually any solution NMR pulse sequence can be used for in-cell NMR experiments, BUT:

- The low sensitivity of NMR requires high protein concentration, not always obtained;
- The viability of the cell sample is limited to few hours;

Therefore fast and sensitive experiments are often needed:

- Fast pulsing experiments: 2D SOFAST-HMQC, 3D BEST-triple resonance experiments;
- Sparsed sampling experiments: non-uniform sampling, projection spectroscopy.
The $^1$H-$^{15}$N SOFAST-HMQC$^{(1)}$ is often used for in-cell NMR. It is the fast equivalent of the $^1$H-$^{15}$N HMQC.

The selective $^1$H pulse excites only the amide protons, allowing faster longitudinal relaxation between the scans: shorter interscan delays.

The pulse can be set at the Ernst angle $\alpha$ (120° instead of 90°), to maximize sensitivity.

Application: In-cell NMR of hSOD1

- Human superoxide dismutase 1 (hSOD1): 32 kDa dimeric enzyme.

- The mature form binds one copper ion and one zinc ion per monomer (Cu,Zn-hSOD1). It has an intramolecular disulfide bridge.

- The \textit{in vivo} copper loading and disulfide formation involves the copper chaperone for SOD (CCS).

Banci, L., Barbieri, L., Bertini, I., Cantini, F., Luchinat, E., Nascent SOD1 analyzed by in-cell NMR, Submitted
A Cross-talk between Solid-State (SS) and Solution NMR
Apo SOD1 – a partially disordered molecule

Solid-state (crystals/microcrystals)

- Cu, Zn dimer chemical shifts ($^{13}$C, $^{15}$N)
  - Pintacuda et al, Angew. Chem. 2007

Solution-state

- Solution NMR assignment serves as the starting points of SSNMR data assignment
- SSNMR assignment can help to resolve some uncertainties in solution NMR assignment of partially unfolded species

TALOS

Average Structural Differences between solid-state and solution ($\beta$ propensity)

Hints on the structures of fibril-ready states

NMR spectra of ApoSOD1 in Microcrystals and Solution

- Similarity in spectral patterns permits the integrative analysis of both SSNMR and solution NMR data

Red: $^{13}\text{C} - ^{13}\text{C}$ 2D TOCSY spectrum of apoSOD1 in solution

Blue: $^{13}\text{C} - ^{13}\text{C}$ 2D DARR spectrum of apoSOD1 in microcrystals

Hot Spots for ApoSOD1 Amyloidosis

- In solution loops IV and VII gain transiently high β-propensity
- SSNMR and solution NMR are complementary methods
- SSNMR facilitates the use of solution NMR data for understanding the mechanism of amyloidosis at residue specific level

Mechanistic Systems
Biology
Mechanistic Systems Biology

Complex living systems should be studied in their integral state.

Functional processes need to be described based on the 3D structural and dynamic interactions of the various players.

…A system-wide perspective requires the identification of all the players in the studied process and within the “system” under analysis.

Proteins must be framed within their cellular context.
A relatively small-scale, physiologically central system:

The InterMembrane Space of the Mitochondrion

Structure of a mitochondrion

O$_2$ flows in and over 50 kg/day of ATP are produced
Mitochondria derive from parasitic Gram-negative bacteria: they contain 1000 proteins but only 15 are produced *in situ*.

The large majority of them must be imported, including those involved in copper trafficking.
Some mitochondrial pathways involved in protein import and copper transport

- **COX17**  
  CcO copper chaperone

- **COX11**  
  CcO assembly protein ctaG

- **SCO**  
  Synthesis of CcO

- **Cu_B**  
  subunit 1 of CcO

- **Cu_A**  
  subunit 2 of CcO

- **Mia40**  
  Mitochondrial intermembrane space Import and Assembly protein 40

- **ALR**  
  Augmenter of Liver Regeneration

- **Cytc**  
  cytochrome c

- **Tim**  
  mitochondrial import Traslocase  
  Inner Membrane
Solution structure of Mia40, a key protein in IMS protein import

E₀ ≤ -0.34 V

Oxidized Mia40

Partially Reduced Mia40

Banci, Bertini, Cefaro, Ciofi Baffoni, Gallo, Sideris, Tokatlidis Nat Struct Mol Biol 2009
Cox17 mitochondrial import

Reduced apoCox17 is unstructured.

apoCox17 $^2S$-S

Mia40

IMS

matrix

OM

IM

cytosol
Cox17 is unfolded in the cytoplasm detected in living cells

The protein folding state depends on the cellular compartment

Mitochondrial Oxidative Folding Mechanism by $^{13}$C NMR

cytosol  \[ \text{apoCox17}_{6\text{SH}} \]  Mitochondria

Mitochondrial Oxidative Folding Mechanism by $^{13}$C NMR

$^{13}$C NMR

$^{1H}$-$^{13}$C HSQC apoCox17$_{4\text{SH}}$

$^{1H}$-$^{13}$C HSQC apoCox17$_{2\text{SH1S-S}}$

$^{1H}$-$^{13}$C HSQC apoCox17$_{2\text{SH}}$

Mitochondrial Oxidative Folding Mechanism by $^{13}$C NMR
Cox17 in mitochondria has a CHCH fold

Cox17 fully reduced

Cox17 partially reduced

Cox17 fully oxidized
The first step in Cox17 folding

A hydrophobic cleft on Mia40 is the interaction site for Cox17

Upon intermolecular S-S bond formation, the first helix is formed

Banci, Bertini, Cefaro, Ciofi, Gallo, Sideris, Tokatlidis *Nature Struct Mol Biol* 2009
Final steps in the maturation of Cox17

Then the first intramolecular S-S bond and the second helix are formed

O$_2$ can now rapidly form the second disulphide bond

Cox17$_{1S-S}$

Disulphide formation

Cox17$_{2S-S}$

Banci, Bertini, Cefaro, Ciofi Baffoni, Gallo, Sideris, Tokatlidis *Nat Struct Mol Biol* 2009
Oxidative folding reaction between Mia40 and Cox17
Oxidative folding processes in IMS

Mia40 acts as a chaperon
A general folding process for CHCH proteins

Schematic overview of mitochondrial pathways

- COX17
  - CcO copper chaperone

- Mia40
  - Mitochondrial intermembrane space Import and Assembly protein 40

- ALR
  - Augmenter of Liver Regeneration
There are two splicing variants for ALR: in mitochondria there is the long one.

Hydrophobic residues in the N-terminus are the crucial molecular mediators to efficiently guide the electron transfer process from Mia40 to FAD in ALR.
Hydrophobic interactions between Mia40 and the N-terminal domain of ALR mediate efficient electron transfer from Mia40 to FAD in ALR.
Electron shuttling mechanism

Efficient transfer: The reaction of sf-ALR with MIA40 proceeds to completion at 0.5:1 molar ratio

Copper incorporation in CcO
Towards systems biology of copper

Schematic overview of mitochondrial pathways

- **COX17**
  - CcO copper chaperone
  - Cu(I)

- **COX11**
  - CcO assembly protein ctaG
  - Cu(I)

- **SCO**
  - Synthesis of CcO
  - Cu(I)

- **Cu_B**
  - Subunit 1 of CcO

- **Cu_A**
  - Subunit 2 of CcO

- **CcO**
  - Cytochrome c Oxidase Complex IV

- **Mia40**
  - Mitochondrial intermembrane space Import and Assembly protein 40

- **ALR**
  - Augmenter of Liver Regeneration

- **CytC**
  - Cytochrome c
Mia40
Mitochondrial intermembrane space Import and Assembly protein 40

COX17
CcO copper chaperone

COX11
CcO assembly protein ctaG

SCO
Synthesis of CcO

Cu(I)

ALR
Augmenter of Liver Regeneration

Cu(I)

CytC
cytochrome c

Ccu
Cytochrome c Oxidase Complex IV

Cu(I)

Tim
mitochondrial import Traslocase Inner Membrane

CHCH
coiled coil-helix-coiled coil-helix

CIAPIN1
Cytokine Apoptosis Induced inhibitor 1

Cu(I)

Cu(I)

Cu(I)

Cu(I)

Cu(I)

Cu(I)

Cu(I)

Cu(I)

Cu(I)
CERM – Magnetic Resonance Center

Bio-Labs

Conference Room

Library

Workstations

He Liquefier

Biobank

700

800

900*

950*

Relaxometer 0.01-100

Department of Chemistry (offices, bio-labs, instruments..)

* EQUIPPED WITH CRYOPROBES
Applications are welcome for:

**Undergraduate Students**
(Chemistry of Biological Molecules)

**Graduate Students**
(International PhD in Structural Biology)

**Postdoc**