Hydrogen Exchange Method

Hydrogen exchange (HX) techniques is described for measuring the approximate exchange rates of the more labile amide protons in a macromolecule. The exchangeable amides in proteins are:
Exchangeable Nucleotides

β-D-Ribose
XWTPMA

2′-Deoxy-β-D-Ribose
XAMWTNP

Cytosine, C
AX

Thymine, T
A₃X

Uracil, U
AX

Adenine, A
A + A

Guanine, G
A
Hydrogen-Exchange Chemistry

- Hx rate is catalyzed by OH\(^-\) and H\(^+\)
  
  \[ k_{\text{intrinsic}} = k_{\text{oH}} [\text{OH}^-] + k_{\text{H}} [\text{H}^+] + k_w \]

- All exchange rates are referenced to random coil polyAla at 0 C.

- HX rates are sensitive to pH, local chemical environment, solvent, sidechain type, neighboring amino acids and temperature.

- \( k_{\text{intrinsic}} \) for each amino acid is different

\[ pD = \text{pH}^* + 0.4 \]

A minimum \( \sim \) pH 3.5

- > 1hr at pH 3
- < 1ms at pH 10

Bai. And Englander. (1993) Proteins, 17, 75;
HX vs. Protein Structure

In proteins, HX rates can be altered:
- H-bonding
- Shielding in the center of protein
- Shielding by binding another molecules
- pH and temperature

Extremely slow exchange can be months, yrs

Protection factor $\theta_p = \frac{k_{\text{intrinsic}}}{k_{\text{obs}}}$

$\theta_p > 10^6-10^7$ for slow exchange

Amide exchange rate contains information about secondary structural elements
Hx Mechanism (Ex1/EX2)

- Hvidt & Nielsen, 1966

- Solvent penetrates protein secondary structure
- A protected amide hydrogen is ‘closed’ to exchange and becomes accessible to exchange through an ‘open’ state at the exchange rate for an unstructured peptide.

Ex1: $k_{\text{intrinsic}} \gg k_{\text{cl}} \quad k_{\text{obs}} = k_{\text{op}}$ independent of pH

Ex2: $k_{\text{intrinsic}} \ll k_{\text{cl}} \quad k_{\text{obs}} = k_{\text{op}} k_{\text{intrinsic}}$ pH dependent

Ex2 is typically encountered in proteins under conditions where folded state is stable and intrinsic exchange is relative slow
HX is an excellent way to look at the stability of proteins

- The rates of amide proton exchange for individual protons can be related to equilibrium constants for opening of individual hydrogen bonds. Knowing the equilibrium constants, one can calculate the free energy for the conformational transition which allows exchange to occur.

- When certain protons are only exposed in the completely unfolded form then the equilibrium constants and $\Delta G$s correspond to the global unfolding reaction. These protons are usually the slowest exchanging protons in the molecule.

$$\Delta G_{HX} = -RT \ln \left( \frac{k_{obs}}{k_{intrinsic}} \right)$$

- For mutation, the change of stability:

$$\Delta \Delta G_{HX} = (\Delta G_{HX})_{wt} - (\Delta G_{HX})_{mut} = -RT \ln \left( \frac{k_{ex,wt}}{k_{ex,mut}} \right)$$
Folded-state Mechanism

• "penetration" model
  – multiple, small, non-cooperative, internal fluctuations create ensembles of inter-converting conformations of varying protection, and provide transient access of solvent to buried NHs (Woodward et al., 1982).

• "local unfolding" model
  – a helix undergoes cooperative breakage of H-bonds and exchangeable species have locally unfolded secondary structure (Englander & Kallenbach, 1984).

• Miller & Dill model
  – exchange occurs from first-excited states that can have a slightly higher free energy, yet very different conformations than the native state.
Adding D$_2$O to our H$_2$O solution and take spectra at different times, signals from different amide protons will decrease in size at different rates. We look at the NH to H$\alpha$ fingerprint at different times in DQF-COSY or HSQC.

Amide Exchange Rates

- $t = 0$ - No D$_2$O
- Add D$_2$O
- $t = t_1$
- $t = t_2$
Amide exchange rates in *Escherichia coli* acyl carrier protein: Correlation with protein structure and dynamics

**Michael Andrec, R. Blake Hill, and James H. Prestegard**
Department of Chemistry, Yale University, New Haven, Connecticut 06511

---

**Fig. 1.** Selected spectra from the amide $\text{H}_2\text{O}/\text{D}_2\text{O}$ exchange time-course. A: HMQC spectrum obtained in $\text{H}_2\text{O}$. B, C, D: Spectra obtained 9, 61, and 133 min after addition of $\text{D}_2\text{O}$ (see the Materials and methods). Peak corresponding to Ser 27 (10.07, 116.2 ppm) is visible in the spectra corresponding to A and B but is not shown in the plot limits of this figure. This signal does not appear in the spectra of C and D.

**Fig. 3.** Examples of signal loss due to $\text{D}_2\text{O}$ exchange as a function of time after addition of $\text{D}_2\text{O}$ for three selected residues. Peak height is in arbitrary units.

\[ I = r + \alpha e^{-k_{\text{ext}}t} \]
Amide Exchange Rates in ACP

- Residues at the center of helices and hydrophobic core have slow exchange rates.
- The overall protection factors (< $10^{4.5}$) are smaller than other proteins suggesting that ACP has high mobility.
- Helix II exchanges faster than helix I and helix III suggesting that Helix II is highly dynamic.

Unfolding/Folding and Misfolding

Unfolded Protein → Molten Globule Intermediate → Native Folded Protein

Misfolding → Aggregates

Disease States
- Prion diseases
- Amyloidoses
The refolding experiment involved dilution of droplets of protein denatured in 6 M GuHCl in H₂O solution into a denaturant free solution of D₂O to initiate refolding and hydrogen exchange simultaneously.

After folding completed, HX is quenched by lowing the pH.

Comparing 2D NMR spectrum of the refolded protein with that was not denatured. Residues protected early in refolding can be detected using NMR.
Competition HX of lysozyme

- Using 65 slowly exchanging amide hydrogen as probes. The majority of residues in β-domain have exchange >30%. The majority of residues in α-domain have exchange <30% suggesting that two structural domains of lysozyme are folding domains that differ significantly in the extent to which protected structure accumulates early in the folding process.

Pulsed-Label Hydrogen Exchange

• After an adjustable refolding time, $t_f$, the protein is subjected to a short high pH pulse, where exchange of the unprotected NHs is very fast. NHs protected by structure within the folding time does not exchange during the short pulse.
• After a pulse time $t_p$, the D to H exchange is quenched by rapidly lowering the pH.
• After folding completed, the pattern of NH and ND labels in the refolded protein is analyzed by 2D NMR.
• Increasing $t_f$ time, proton occupancies measured in the NMR spectrum decreases. Plotting proton occupancy vs. folding time $t_f$. 
Identifying Folding Pathway by HX Pulse-Labeling

(a) pure 2-state
All probes achieve 100% protection at the same rate in a single kinetic step.

(b) U -> I -> N sequential,
I has A&B H-bonds with the same HX constant

(c) U1 -> N(30%) 
U2 -> N(70%) two heterogeneous parallel paths

(d) U1-> N
U2->I->N contribution of intermediate and heterogeneous folding
Parallel Folding Pathway of Lysozyme

- All probes (50% of 126 amides) have one fast phase and one slow phase.
- Fast phase rates for both α- and β domains are 10 ms.
- Slow phase rates for α is 65 ms and β is 350 ms, respectively.
- Data suggest that α domains fold before β domain and different populations of molecules folded by kinetically distinct pathways.
Additional Methods for Amide-Water Exchange


CLEANEX-PM has the ability to specifically monitor water-proton exchange without 1) exchange-relayed NOE/ROE from rapidly exchanging protons (hydroxyl or amide groups) in the macromolecules, 2) intra-molecular NOE/ROE peaks from protein $\text{C}\alpha\text{H}$ protons which has chemical shifts coincident with water, or TOCSY-type interactions.

**CLEANEX-PM spin-locking sequence:**

$135^\circ(\text{x})\ 120^\circ\ (-\text{x})\ 110^\circ\ (\text{x})\ 110^\circ\ (-\text{x})\ 120^\circ(\text{x})\ 135^\circ\ (-\text{x})$

CLEANX-PM has the ability to specifically monitor water-proton exchange without 1) exchange-relayed NOE/ROE from rapidly exchanging protons (hydroxyl or amide groups) in the macromolecules, 2) intra-molecular NOE/ROE peaks from protein $\text{C}\alpha\text{H}$ protons which has chemical shifts coincident with water, or TOCSY-type interactions.
The FHSQC indicates proton signal that remain at the amide resonance through out the pulse sequence. The CLEANEX-PM indicates $^1$H signals that initiate in the $^1$H$_2$O resonance and then transfer to $^1$H amide resonance during the mixing period of the pulse sequence.

Staphylococcal nuclease –Hwang et al., 1998
Quantitation of rapid proton-deuteron amide exchange using hadamard spectroscopy

Catherine Bougault\textsuperscript{a}, Lianmei Feng\textsuperscript{b}, John Glushka\textsuperscript{b}, Eriks Kupče\textsuperscript{c} & J.H. Prestegard\textsuperscript{b,*}

\textsuperscript{a} Structural Biology Institute, CNRS, Grenoble, France; \textsuperscript{b} Complex Carbohydrate Research Center, University of Georgia, Athens, GA, U.S.A. \textsuperscript{c} Varian Associates, Palo Alto, CA, U.S.A.

Received 14 October 2003; Accepted 20 October 2003

Key words: HSQC, $^{15}$N, NMR, protein structure, ubiquitin

Abstract

The rates of amide proton exchange in protein backbones are very useful reporters of accessibility and structural stability of specific residues and secondary structure elements. Measurement by monitoring changes in intensity of cross-peaks in standard $^{15}$N-$^1$H HSQC spectra as protons are replaced by solvent deuterons has become widely accepted. However, these methods are limited to relatively slow rates due to time limitations of the conventional 2D HSQC experiment. Here we show that a Hadamard encoded version of the HSQC, which relies on a multiplexed, frequency selective, excitation in the $^{15}$N dimension, extends application to rates that are as much as an order of magnitude faster than those previously accessible.
Conventional $[{1H,15N}]$-HSQC

A 0.5 mM15N-labeled sample was prepared in 50 mM potassium phosphate buffer in H2O/D2O 5/95, pH = 6.2 600 MHz with cryoprobe. It required approximately 21 min using 128 t1 time increments and 4 scans per increment.
Reconstructed Hadamard [1H,15N]-HSQC spectra for ubiquitin.

(A) Data in 1H2O collected with 128 t1 increments in 20 min. The sample was then lyophilized overnight and brought back to its initial volume with 99.9% 2H2O and immediately returned to the spectrometer for rapid collection of a series of Hadamard spectra. (B) First point after 1 min in 2H2O collected with 4 scans in 42 s.
Cross-peak intensities as a function of time

Lines are best fits to 
\[ I(t) = I_0(\exp(-kt) + \text{const}). \]

The precision of the data is quite high with the estimated errors for rates in the range of \( 1 \times 10^{-3} \) min\(^{-1}\) being on the order of 5%. Rates derived also show reasonable agreement with previously published rates.
Molecular recycling within amyloid fibrils

H/D exchange of PI(3)K-SH3 amyloid fibrils monitored by NMR spectroscopy.

The peak volumes in 1H-15N heteronuclear single quantum correlation (HSQC) spectra recorded after H/D exchange (V_D2O) relative to those before exchange (V_H2O) after dissolution of the fibrils (see main text) are plotted against residue number. The bars correspond to the relative peak volumes in the spectra of fibrils exchanged for 4 (orange) and 15 (blue) days. An asterisk above a bar indicates a residue whose resonance is not fully resolved; the absence of a bar indicates that the resonance of the residue is not detectable (Supplementary Fig. 2S). The locations of the β-sheet strands in the native state of PI(3)K-SH3 are indicated.

H/D exchange of PI(3)K-SH3 amyloid fibrils analysed by ESI-MS

- Mass spectra (8+ and 7+ charge states) showing the relative populations of Ppd and Pfd species after exchange for different times. The spectral intensities are all plotted relative to the same intensity of the Ppd 8+ charge state.  
- Plot of the relative fraction of Pfd molecules in the sample as a function of time of exchange. Error bars indicate standard deviations resulting from duplicate samples. The inset shows the expansion of the spectrum corresponding to the 7+ charge states of the Ppd and Pfd species after 42 days. The peak widths of the Ppd and Pfd species (3.0 ± 0.14 and 2.9 ± 0.14 Da, respectively) are comparable to that observed in spectra of solutions in which the soluble monomeric PI(3)K-SH3 protein was allowed to exchange in 95% DMSO-d6/5% D2O solvent (2.9 ± 0.07 Da).
Electron micrographs and H/D exchange data for fibril suspensions

- a–c, Images of fibrils collected before exchange (a) and after 49 days of exchange during which the samples were subjected to repetitive pipetting (b) or sonication (c). d–f, Images of fibrils collected after 7 days of repeated sonication (d), and 7 days (e) and 48 days (f) after the end of the sonication procedure. Scale bars, 100 nm. g, Histogram of the percentage of fully deuterated species, Pfd, as a function of time of exchange for fibril suspensions that were untreated (black bars), subjected to repetitive pipetting (light grey bars) or sonication (dark grey bars). Note that the x axis represents the times at which different samples were analysed and therefore is nonlinear. Error bars indicate standard deviations resulting from duplicate samples.
The molecular recycling model

- **a**, Schematic representation of the recycling process involving a single amyloid fibril. A protonated protein molecule (open circle) dissociates from the fibril. Once in solution, hydrogen exchange takes place rapidly and the molecule is subsequently reincorporated in a fully deuterated state (filled circle).
- **b**, Schematic representation of the recycling mechanism for a distribution of amyloid fibrils at different times of exchange. At a given time point the dissociation of a molecule from a fibril (above) is counteracted by the re-association of that molecule in another fibril (below).
- **c**, Simulated equilibrium distributions corresponding to average degrees of polymerization.
- **d**, Simulations of the time dependence of the incorporation of deuterated molecules into fibrils using the molecular recycling model as a function of the degree of polymerization shown in **c**;
Life in a Crowded World-
Applications of In Cell NMR Spectroscopy

Posttranslational modifications (left), conformational changes (middle), and binding events (right) can be monitored relative to the unaffected protein (top) through changes in the resonance positions of peaks in NMR spectra.
References for In-cell NMR

- Hubbard, J., MacLachlan, LK, King G., Jones, J and Fosberry, A. Nuclear magnetic resonance spectroscopy reveals the functional state of the signaling protein CheY in vivo in E. coli. 2003, Molecular Microbiology, 49(5):1191-1200
[\textsuperscript{15}N,\textsuperscript{1}H]-HSQC of Calmodulin with \textsuperscript{15}N-labeled Lys

In-cell    In vitro
$^{15}\text{N,}^{1}\text{H]}$-HSQC spectrum of the N-terminal domain of the bacterial mercury-detoxification protein MerA

In-cell

In vitro
Selective Observation of a Protein Inside Living Cells

13C, 15N and 19F

- Using the same cells for protein overexpression and subsequent NMR experiments would lead to a huge background signal caused by labeling not only the protein of interest with the NMR-active isotope, but also virtually every cellular component?
  - Rifampicin, a drug that inhibits the bacterial RNA polymerase but does not block the RNA polymerase of the bacteriophage T7, was used.

Instead of background signals, the overexpression level of the protein is the most important parameter for the spectral quality of in-cell NMR experiments.
In-cell NMR spectra of NmerA

• Comparison of the quality of in-cell NMR spectra of NmerA which were obtained by protein expression in (A) 15N-labeled minimal medium and (B) 98% 15N-labeled, 97% deuterated rich medium (Celtone-dN, Martek).
Sample preparation procedures

For conventional in vitro NMR spectroscopy, cells are grown and induced in isotopically labeled media. The protein is then purified and placed in an NMR tube or HRMAS rotor.

For in-cell NMR, the cells are first grown in unlabeled media and then transferred to isotopically labeled media prior to induction. Intact cells harboring the overexpressed protein are then prepared as a slurry (20-30%) for NMR studies.
The Rotational Correlation Time of Proteins Inside Cells.

- In Yeast cell, an intracellular viscosity twice that of water was observed through measurements of 19F relaxation times of selectively tryptophan-fluorinated enzymes.
- In fibroblasts, viscosities as low as 1.2-1.4 times that of water determined by fluorescence polarization experiments.
- In higher cells using 13C relaxation, 1H line width, and electron spin resonance also showed an upper limit of the intracellular viscosity for most cell types of twice that of water.
- Due to the linear relationship between the viscosity, the rotational correlation time, and the molecular mass of a protein, this 2-fold increase in the viscosity leads to a 2-fold increase in the apparent molecular mass of a macromolecule.
- The intracellular viscosity is not a limiting factor for in-cell NMR experiments (??).
# Comparison In-Cell NMR Techniques

Using either Magic Angle Spinning or Liquid-State NMR

<table>
<thead>
<tr>
<th></th>
<th>high-resolution magic angle spinning NMR</th>
<th>high-resolution liquid-state NMR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>host organism</strong></td>
<td><em>R. solanacearum</em></td>
<td><em>E. coli</em> BL21</td>
</tr>
<tr>
<td><strong>biological macromolecule</strong></td>
<td>cyclic osmoregulated periplasmic glucan</td>
<td>proteins: NmerA, calmodulin</td>
</tr>
<tr>
<td><strong>cellular compartment</strong></td>
<td>periplasm</td>
<td>cytoplasm</td>
</tr>
<tr>
<td><strong>heteronucleus</strong></td>
<td>$^{13}\text{C}$</td>
<td>$^{15}\text{N}$</td>
</tr>
<tr>
<td><strong>cell density</strong></td>
<td>~80%</td>
<td>20-30%</td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td>Guy Lippens</td>
<td>Volker Dotsch</td>
</tr>
</tbody>
</table>
FlgM is structured in *E. coli*.

Free FlgM is mostly unstructured in dilute solution, but its C-terminal half forms a transient helix. One signature of protein structure can be the absence of crosspeaks in HSQC NMR spectra because of conformational exchange. NMR studies in dilute solution indicate that the C-terminal half of FlgM becomes structured on binding 28, as shown by the disappearance of crosspeaks from residues in the C-terminal half of FlgM in the FlgM-28 complex.

Glucose and BSA induce FlgM Structure

Glucose induces structure in FlgM. Overlaid HSQC spectra of FlgM in dilute solution (red) and in 450 g/liter 2.5 M glucose (black).

BSA induces structure in FlgM. HSQC spectrum of FlgM in 400 g/liter BSA.
In-Cell NMR of CheY

A: in vitro purified protein
B: in vivo
C: Extracellular solution
D: Lysed cell without purification

*In vivo* signals can be obtained but the line width is broader than the purified proteins.

Hubbard JA., Fosberry AP. Molecular Microbiology, 2003 49(5), 1191.
BRL-16492PA induced chemical shift changes

Positions with large change

A: in vitro; B: in vivo
Black: without BRL-16492PA
Red: with BRL-16492PA

Hubbard JA., Fosberry AP. Molecular Microbiology, 2003 49(5), 1191.
A cool way to make proteins

- The cold shock expression system is repressed at 37 °C and induced by transfer of the bacteria to low temperature. Cellular protein production stops after transfer to the 15 °C labeling medium. NMR spectra obtained from centrifuged (100,000g) lysates after sonication are very similar to those of the purified protein. The spectra can be used to obtain details about mutants or sequence variants of proteins of known structure.
Calmodulin spectra at 10 °C.

Qing et al., 2004, Nat Biotechnol, 877
Calmodulin in Cell vs. cell lysate

Blue: in-cell
Red: supernatant

600 MHz

Yang, JJ, Unpublished data
Yang JJ, unpublished data
In-cell NMR of CD2

- The *in vivo* signals are in a narrow region possibly due to protein aggregation or high viscosity.
- The supernatants of both CD2 and 6D31 after cell lysis resemble wild type-like dispersion.
(a) In-cell NMR of interacting proteins. Target protein was over-expressed in uniformly labeled [U-15N] medium and a sample containing no interactor protein was prepared for in-cell NMR (top). Cells were then washed and resuspended in label-free medium. Samples were taken as the interacting protein was overexpressed (bottom). Changes in the target protein structure are measured as the concentration of the interactor is increased. (b) SDS-PAGE of ubiquitin and STAM2 sequential expression. Rosetta(DE3) cells were induced with L-arabinose and then with IPTG for the indicated amounts of time to overexpress ubiquitin and STAM2, respectively. Lane 1, uninduced Rosetta(DE3) cells. Note that the ubiquitin level remains essentially constant as STAM2 overexpression increases.
NMR spectra and interaction maps of ubiquitin-ligand complexes.