Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/2/10/e1600886/DC1

Supplementary Materials for

Determination of functional collective motions in a protein at atomic resolution using coherent neutron scattering

Liang Hong, Nitin Jain, Xiaolin Cheng, Ana Bernal, Madhusudan Tyagi, Jeremy C. Smith

Published 14 October 2016, *Sci. Adv.* **2**, e1600886 (2016)
DOI: 10.1126/sciadv.1600886

This PDF file includes:

Supplementary Text

Neutron scattering fundamentals

Dynamic neutron scattering is described in terms of the intermediate scattering function, *I(q, t)*, which contains incoherent and coherent components: $I_{inc}(q,t)$ and $I_{coh}(q,t)$

$$I_{inc}(q,t) = \sum_{j}^{Z} b^{2}{}_{j,inco} \left\langle \exp[-i\vec{q}\vec{R}_j(0)] \exp[i\vec{q}\vec{R}_j(t)] \right\rangle \quad \text{(S1)}$$

$$I_{coh}(q,t) = \sum_{i}^{Z}\sum_{j}^{Z} b_{i,coh}\, b_{j,coh} \left\langle \exp[-i\vec{q}\cdot\vec{R}_i(0)] \exp[i\vec{q}\cdot\vec{R}_j(t)] \right\rangle \quad \text{(S2)}$$

where Z is the total number of atoms, $b_{j,inco}$ ($b_{j,coh}$) is the incoherent (coherent) scattering length of a given atom j, $\vec{R}_j$ is the position vector of that atom, the brackets denote an ensemble and orientation average, and q is the scattering wave vector. $I_{inc}(q,t)$ characterizes self-correlation in atomic motions, while $I_{coh}(q,t)$ probes mostly cross-correlations, i.e., inter-atomic fluctuations. When $t = 0$, the coherent intermediate scattering function, $I_{coh}(q,0)$, becomes the static structure factor, $I(q)$, characterizing the atomic structure of the system

$$I(q) = \sum_{i}^{Z}\sum_{j}^{Z} b_{i,coh}\, b_{j,coh} \exp[-i\vec{q}\cdot\vec{R}_i(0)] \exp[i\vec{q}\cdot\vec{R}_j(0)] \quad \text{(S3)}$$

The dynamic structure factor, $S(q, \Delta E)$, is the time Fourier transform of the intermediate scattering function,

$$S_{coh}(q,\Delta E) = \int_{-\infty}^{+\infty} I_{coh}(q,t)\exp(2\pi i\frac{\Delta E}{h}t)dt \quad \text{(S4)}$$

$$S_{inc}(q,\Delta E) = \int_{-\infty}^{+\infty} I_{inc}(q,t)\exp(2\pi i\frac{\Delta E}{h}t)dt \quad \text{(S5)}$$

where h is the Planck constant, and ΔE is the energy transfer between the incident and scattered neutron. $S(q, \Delta E)$ furnishes the amplitude-weighted distribution of the dynamic modes in the sample over frequency at a given q, where the frequency of the mode is defined by $\Delta E/h$.

Procedure for data analysis

Separation of contributions from coherent and incoherent scattering into the neutron signals measured on hydrogenated and deuterated proteins.

Normally, the coherent and incoherent signal from a powder sample can be expressed as: $I_{coh}=NS(q)F(q)$ and $I_{inc}=NA_{inc}$, respectively, where N is the number of protein molecule in the scattering volume, $S(q)$ is the inter-protein structure factor, defined as the ratio between the total scattering intensity on the powder sample, $I(q)$, and the form factor resulting from a single protein molecule, $F(q)$, and A_{inc} is the incoherent scattering intensity of a single protein molecule and independent on q (*32*, *33*). Here, $F(q)$ and A_{inc} are calculated based on the crystal structure of the protein molecule and $S(q)$ is measured experimentally (Fig. 3A). Hence the coherent and incoherent contributions into the static neutron scattering signal arising from the hydrogendated and deuterated samples can be derived and the results are presented in fig. S1.

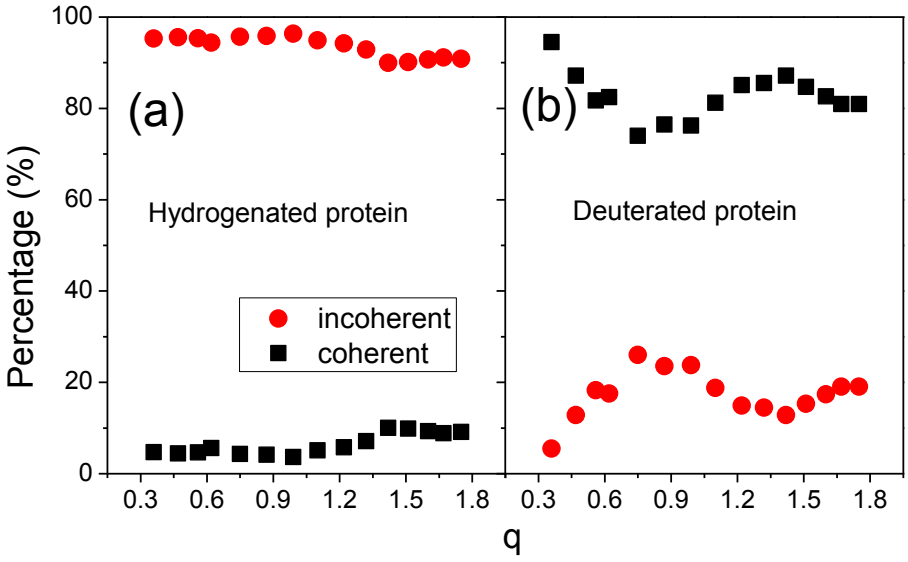

fig. S1. Coherent and incoherent contribution in to the neutron-scattering signal arising from the hydrogenated and deuterated protein samples.

Therefore, the ratio between the coherent and incoherent inelastic scattering intensity, $\gamma(q)$, defined in the main text (Fig. 1, C and D) is estimated as $\frac{In_D(q)P_{coh,D}(q)}{In_H(q)P_{inc,H}(q)}$, where $In_D(q)$ and $In_H(q)$ are the integral of the dynamic structure factor measured on the deuterated and hydrogenated protein samples in the energy window from 5 to 10 μeV, respectively, and $P_{coh,D}(q)$ ($P_{inc,H}(q)$) is the contribution in percentage of coherent (incoherent) scattering in to the deuterated (hydrogenated) samples (fig. S1). For reference, we compare $\gamma(q)$ with that of In_D/ In_H in fig. S2, and they differ only slightly as coherent scattering dominates the neutron signal in deuterated sample while the incoherent signal dominates that of the hydrogenated sample.

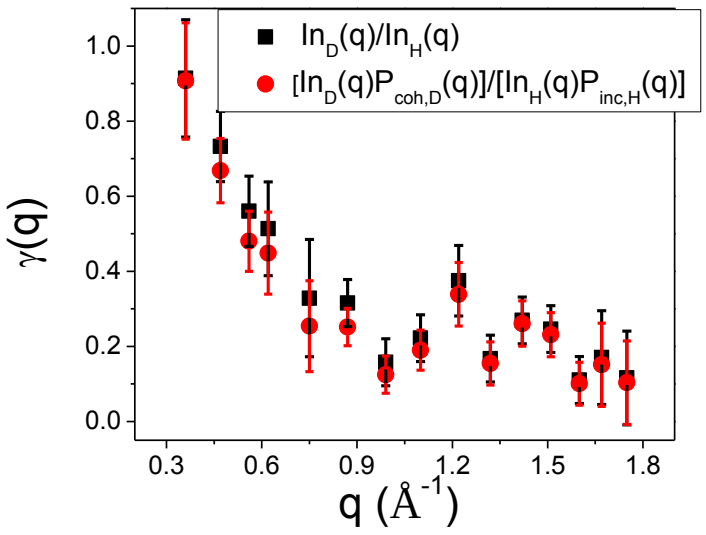

fig. S2. Comparison of $\frac{In_{\mathrm{D}}(q)P_{\mathrm{coh,D}}(q)}{In_{\mathrm{H}}(q)P_{\mathrm{inc,H}}(q)}$ **and** $\frac{In_{\mathrm{D}}(q)}{In_{\mathrm{H}}(q)}$**.**

Similarly, the intensity of elastic peak of coherent dynamics structure factor, $S_{coh}(q,\Delta t)$, should present some difference from that measured directly on deuterated sample, $S_D(q,\Delta t)$, both normalized by the corresponding value at 4 *K*. The connection between them can be expressed as

$$
\begin{aligned}
S_D(q,\Delta t) &\approx \frac{I_{coh,D}(q,\Delta t)+I_{inc,D}(q,\Delta t)}{I_{coh,D}(q)+I_{inc,D}(q)} \\
&= \frac{I_{coh,D}(q,\Delta t)/I_{coh,D}(q)+\dfrac{P_{inc,D}(q)}{P_{coh,D}(q)}I_{inc,D}(q,\Delta t)/I_{inc,D}(q)}{1+\dfrac{P_{inc,D}(q)}{P_{coh,D}(q)}} \\
&\approx \frac{S_{coh}(q,\Delta t)+\dfrac{P_{inc,D}(q)}{P_{coh,D}(q)}S_{inc}(q,\Delta t)}{1+\dfrac{P_{inc,D}(q)}{P_{coh,D}(q)}}
\end{aligned}
\tag{S6}
$$

where $I_{coh,D}(q)$ and $I_{inc,D}(q)$ are the coherent and incoherent elastic scattering intensities from the deuterated protein, while $I_{coh,D}(q,\Delta t)$ and $I_{inc,D}(q,\Delta t)$ are the intermediate scattering functions at time Δt arising from the deuterated protein. $S_{inc}(q,\Delta t)$ is the intensity of the elastic peak in the incoherent dynamic structure factor when scaled by the value at 4 *K*, and $P_{inc,D}(q)$ and $P_{coh,D}(q)$ are the contributions of incoherent and coherent scattering into the deuterated samples at a given *q*, respectively (see fig. S1b). Similarly, the relation between $S_{inc}(q,\Delta t)$ and $S_H(q,\Delta t)$ can be expressed as

$$
S_H(q,\Delta t) \approx \frac{S_{inc}(q,\Delta t)+\dfrac{P_{coh,H}(q)}{P_{inc,H}(q)}S_{coh}(q,\Delta t)}{1+\dfrac{P_{coh,H}(q)}{P_{inc,H}(q)}} \tag{S7}
$$

where $P_{inc,H}(q)$ and $P_{coh,H}(q)$ are the contributions in percentage of incoherent and coherent scattering into the hydrogenated samples, respectively (see fig. S1a). Combining eqs. S6 and S7, one can get

$$S_{coh}(q,\Delta t) \approx \frac{(1+\frac{P_{inc,D}(q)}{P_{coh,D}(q)})S_D(q,\Delta t) - \frac{P_{inc,D}(q)}{P_{coh,D}(q)}(1+\frac{P_{coh,H}(q)}{P_{inc,H}(q)})S_H(q,\Delta t)}{1-\frac{P_{inc,D}(q)}{P_{coh,D}(q)}\frac{P_{coh,H}(q)}{P_{inc,H}(q)}} \quad \text{(S8)}$$

$$S_{inc}(q,\Delta t) \approx \frac{(1+\frac{P_{coh,H}(q)}{P_{inc,H}(q)})S_H(q,\Delta t) - \frac{P_{coh,H}(q)}{P_{inc,H}(q)}(1+\frac{P_{inc,D}(q)}{P_{coh,D}(q)})S_D(q,\Delta t)}{1-\frac{P_{inc,D}(q)}{P_{coh,D}(q)}\frac{P_{coh,H}(q)}{P_{inc,H}(q)}} \quad \text{(S9)}$$

In fig. S3, we compare $S_{coh}(q,\Delta t)$ with $S_D(q,\Delta t)$, and they differ only slightly.

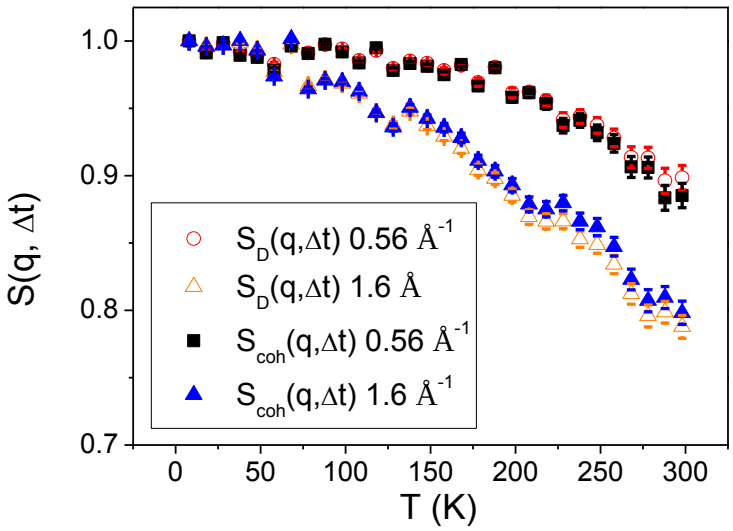

fig. S3. Comparison of $S_{coh}(q,\Delta t)$ with $S_D(q,\Delta t)$

Extraction of the mean-squared atomic displacement

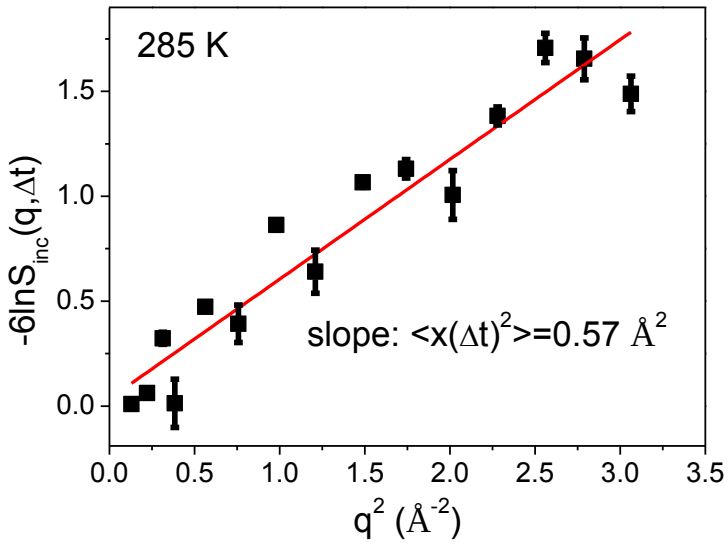

fig. S4. q^2 dependence of the intensity of the elastic peak in the incoherent dynamic structure factor, $S_{inc}(q,\Delta t)$ (eq. S9), at 285 K. The red line serves as a linear fit.

The experimental data obeys a linear fit, indicating that Gaussian Approximation holds (*34*), i.e., that $S_{inc}(q,\Delta t) = exp(-1/6q^2\langle x^2(\Delta t)\rangle)$. $\langle x^2(\Delta t)\rangle$ is obtained from the slope of the linear fit.

Normal mode analysis

Normal mode analysis was performed using the elastic network model (ENM) (*35*). In ENM, residues are modeled as point masses connected by springs of equal strength if both residues are located within a cutoff distance. A recent improvement of NMA based

on the rotational translational block (RTB) method has allowed it to be used in bio-molecular assemblies of ~10,000 residues (*36*). The major assumption behind the RTB is that low-frequency normal modes of proteins can be described as pure rigid-body motions of blocks of consecutive amino-acid residues.

Here, NMA on camphor-bound CYP101 enzyme was performed with the webserver elNémo (*20*). As the current study focuses on the largest amplitude and the lowest frequency modes of the enzyme, the ENM-RTB method is an appropriate choice. The initial structures were taken from the Protein Data Bank [PDB code: 3L63 (*21*)]. The cut-off distance for the atomic interaction was 8 Å, and each block consisted of a single residue. The ten lowest-frequency non-trivial normal modes, and the associated eigenvectors and eigenvalues ω_α^2 were used in the analysis. For simplicity, the eigenvectors, $\vec{e}_a$, were normalized, $\sqrt{\sum_i^N |\vec{e}_a|^2} = \sqrt{\sum_i^N \left|\vec{d}_{i\_norm}{}^{\alpha}\right|^2} = 1$ Å, where $\vec{d}_{i\_norm}{}^{\alpha}$ is the displacement of atom *i* in Mode α when its eigenvector is normalized.

For a period of time, Δt, the system moves along a given normal mode, α, by an amplitude of A_m as $\vec{R}_\alpha = \vec{R}_0 + A_m \vec{e}_\alpha$, where $\vec{R}_0$ and $\vec{R}_\alpha$ are the initial and final sets of atomic coordinates, respectively, and Δt = 1 *ns* for HFBS experiment. The mean-squared atomic displacement resulting from such process can be expressed as

$$< x^2(\Delta t) >_{NM} = \sum_i^Z \left|\vec{d}_i{}^{\alpha}\right|^2 \Big/ Z = \sum_i^N \left|A_m \vec{d}_{i\_norm}{}^{\alpha}\right|^2 \Big/ Z = A_m{}^2 / Z \text{ Å}^2 \qquad \text{(S10)}$$

where $\vec{d}_i^{\,\alpha}$ is the displacement of atom *i* in the process and *Z* is number of atoms in the system. Here, the mean-squared atomic displacement for a given normal mode thus generated is noted as $< x^2(\Delta t) >_{NM}$, being distinct from the mean-squared atomic displacements obtained from incoherent neutron scattering experiment or from MD. In the present work, $< x^2(\Delta t) >_{NM}$ is used as a quantitative measure of the amplitude of a given normal mode. Assuming the atomic displacement is small, the resulting $\zeta((q, \Delta t)_{\mathrm{coh}}$ at time Δt can be estimated as

$$
\begin{aligned}
\zeta_{coh}^{\alpha}(q,\Delta t) &= -\frac{\ln(I_{coh}(q,\Delta t)/I(q))}{q^2\Delta t} \\
&= -\ln\frac{\sum_{i,j} b_i b_j \exp(i\vec{q}(\vec{r}_i(\Delta t)-\vec{r}_j(0)))}{\sum_{i,j} b_i b_j \exp(i\vec{q}(\vec{r}_i(0)-\vec{r}_j(0)))} \Big/ q^2\Delta t \\
&\approx -\ln\left(\frac{\sum_{i,j} b_i b_j \exp(i\vec{q}(\vec{r}_i(0)-\vec{r}_j(0)))(1-\tfrac{1}{2}(\vec{q}\cdot\vec{d}_i^{\,a})^2)}{\sum_{i,j} b_i b_j \exp(i\vec{q}(\vec{r}_i(0)-\vec{r}_j(0)))}\right) \Big/ q^2\Delta t \\
&= -\ln\left(1-\tfrac{A_m^2}{2}\frac{\sum_{i,j} b_i b_j \exp(i\vec{q}(\vec{r}_i(0)-\vec{r}_j(0)))(\vec{q}\cdot\vec{d}_{i\_norm}^{\,a})^2}{\sum_{i,j} b_i b_j \exp(i\vec{q}(\vec{r}_i(0)-\vec{r}_j(0)))}\right) \Big/ q^2\Delta t \\
&\approx Z < x^2(\Delta t) >_{NM} \Big/ 2q^2\Delta t \frac{\sum_{i,j} b_i b_j \exp(i\vec{q}(\vec{r}_i(0)-\vec{r}_j(0)))(\vec{q}\cdot\vec{d}_{i\_norm}^{\,a})^2}{\sum_{i,j} b_i b_j \exp(i\vec{q}(\vec{r}_i(0)-\vec{r}_j(0)))} \\
&= < x^2(\Delta t) >_{NM} \zeta_{coh\_norm}^{\alpha}(q,\Delta t)
\end{aligned}
\qquad \text{(S11)}
$$

where $\zeta_{coh\_norm}^{\alpha}(q,\Delta t)$ is the decay of the coherent intermediate scattering function resulting from Mode α when $< x^2(\Delta t) >_{NM}$ = 1 Å^2. The values of $\zeta_{coh\_norm}^{\alpha}(q,\Delta t)$ for the first-three normal modes are presented in Fig. 4A.