Distribution of Pico-and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation

Shahid N Khan, Cyril Charlier, Rafal Augustyniak, Nicola Salvi, Victoire Déjean, Geoffrey Bodenhausen, Olivier Lequin, Philippe Pelupessy, Fabien Ferrage

Article

Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation

Shahid N. Khan,[1,2,3] Cyril Charlier,[1,2,3] Rafal Augustyniak,[1,2,3] Nicola Salvi,[4] Victoire Déjean,[1,2,3] Geoffrey Bodenhausen,[1,2,3,4] Olivier Lequin,[1,2,3] Philippe Pelupessy,[1,2,3] and Fabien Ferrage[1,2,3,*]

[1]Département de Chimie, École Normale Supérieure-PSL Research University, Paris, France; [2]Sorbonne Universités, UPMC Univ Paris 06, LBM, Paris, France; [3]Centre National de la Recherche Scientifique, UMR 7203 LBM, Paris, France; and [4]Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne, BCH, Lausanne, Switzerland

ABSTRACT   Intrinsically disordered proteins and intrinsically disordered regions (IDRs) are ubiquitous in the eukaryotic proteome. The description and understanding of their conformational properties require the development of new experimental, computational, and theoretical approaches. Here, we use nuclear spin relaxation to investigate the distribution of timescales of motions in an IDR from picoseconds to nanoseconds. Nitrogen-15 relaxation rates have been measured at five magnetic fields, ranging from 9.4 to 23.5 T (400–1000 MHz for protons). This exceptional wealth of data allowed us to map the spectral density function for the motions of backbone NH pairs in the partially disordered transcription factor Engrailed at 11 different frequencies. We introduce an approach called interpretation of motions by a projection onto an array of correlation times (IMPACT), which focuses on an array of six correlation times with intervals that are equidistant on a logarithmic scale between 21 ps and 21 ns. The distribution of motions in Engrailed varies smoothly along the protein sequence and is multimodal for most residues, with a prevalence of motions around 1 ns in the IDR. We show that IMPACT often provides better quantitative agreement with experimental data than conventional model-free or extended model-free analyses with two or three correlation times. We introduce a graphical representation that offers a convenient platform for a qualitative discussion of dynamics. Even when relaxation data are only acquired at three magnetic fields that are readily accessible, the IMPACT analysis gives a satisfactory characterization of spectral density functions, thus opening the way to a broad use of this approach.

INTRODUCTION

Intrinsically disordered proteins (IDPs) and regions (IDRs) lack a stable three-dimensional structure organized around a hydrophobic core (1). Such proteins nevertheless play crucial roles in many cellular processes (2). The discovery of IDPs and IDRs is a challenge for the structure-function paradigm (3) and has opened the way to new biophysical contributions to modern proteomics (4). The characterization of the conformational space of IDPs and IDRs can provide insight into the ensemble representation of their three-dimensional organization (5–8). A detailed and quantitative description of the time dependence of the exploration of the conformational space of IDPs and IDRs is required to

predict (9) and understand the molecular mechanisms underlying their biological function at the atomic scale.

NMR spectroscopy is a powerful tool for probing molecular motions at atomic resolution on a broad range of timescales in both ordered and disordered proteins (6,10,11). In particular, nuclear spin relaxation can be used to probe a diversity of motions from fast (picoseconds to nanoseconds) reorientation to slow (microseconds to milliseconds) chemical exchange (11,12). Pico- and nanosecond motions of protein backbones are most often characterized by analyzing nitrogen-15 relaxation rates, primarily the longitudinal, R_1, and transverse, R_2, relaxation rates, usually supplemented by ^{15}N-{^{1}H} nuclear Overhauser effects (NOEs). The most general level of analysis provides a map of the spectral density for reorientational motions of the internuclear ^{15}N-^{1}H vectors of the protein backbone (13). In folded proteins, a further step consists in the deconvolution of overall motion (rotational diffusion) and internal dynamics, which is possible when these two types of motions are statistically independent (14,15). The most popular framework for such an analysis is the model-free approach (14), for which the motions of each NH vector are described by a correlation time for overall motion and a correlation time and an order parameter for local motions. The so-called extended model-free approach

Shahid N. Khan and Cyril Charlier contributed equally to this work.

Rafal Augustyniak's present address is Departments of Biochemistry, Chemistry and Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

Nicola Salvi's present address is Institut de Biologie Structurale Jean-Pierre Ebel, CNRS-CEA-UJF UMR 5075, 41 rue Jules Horowitz, 38027 Grenoble Cedex, France.

was later introduced to account for internal motions characterized by two correlation times (16).

The structural flexibility of IDPs and IDRs on nanosecond timescales casts serious doubt on the separation of overall and internal motions. The very notion of a single overall motion can be challenged for IDPs. At best, an overall diffusion tensor would correspond to an average over a set of time-dependent diffusion tensors. In addition, local reorientations of bond vectors due to conformational changes that may occur on timescales similar to the instantaneous overall diffusion would make the statistical independence of internal and overall motions less plausible. New methods therefore have to be developed to describe and rationalize the dynamic properties of IDPs and IDRs (17–19).

Several approaches have been developed in the last 15 years to extract quantitative information about pico- and nanosecond dynamics in IDPs and IDRs from nuclear spin relaxation rates. Often based on spectral density mapping (20,21), most of these approaches rely on the model-free formalism, with residue-specific correlation times (22,23) (i.e., without an overall diffusion tensor), a distribution of picosecond and nanosecond correlation times (24,25), or a statistical analysis of extended model-free parameters (26). An analysis based on a distribution of correlation times necessarily introduces a physical bias, since one must choose a mathematical function to describe the distribution. In the case of the model-independent correlation (MIC) time distribution (26), the statistical independence of three types of motions is not required, since the extended model-free results are considered as a simplified representation of a continuous distribution of correlation times. However, the significance of such a statistical treatment is necessarily limited, since it provides little information about the number of modes of the actual distribution of correlation times. Neither approach seems suited to describe a distribution of correlation times that is a priori unknown. However, simply increasing the number of correlation times or distributions in either approach would be questionable, since the empirical information available from nuclear spin relaxation is limited.

Here, we introduce an approach we call interpretation of motions by a projection onto an array of correlation times (IMPACT) to analyze multiple-field relaxation data in disordered proteins. This method relies as little as possible on any particular physical model of protein motions but constitutes a mathematical reconstruction of the distribution of correlation times. We define an array of n correlation times, τ_i (or, equivalently, of reciprocal frequencies, $\omega_i = 1/\tau_i$) in a range that is effectively sampled by nitrogen-15 relaxation. The experimental spectral density function is then reproduced by a sum of n Lorentzian functions, $J_i(\omega)$, one for each correlation time τ_i. The result of this process, similar to a projection onto a basis of Lorentzian functions, is a discrete distribution of correlation times spanning a range that is relevant to rationalize relaxation. This approach is analogous to the discretization step encountered in regularization methods (27,28), but the volume of experimental data exploited in this study is too limited to use a full regularization approach. Nevertheless, the multimodal character of the distribution of correlation times can be nicely revealed, and the most relevant correlation times for backbone motions can clearly be identified.

IMPACT was originally conceived for a set of relaxation rates obtained at five magnetic fields ranging from 9.4 to 23.5 T (i.e., with proton Larmor frequencies of 400, 500, 600, 800, and 1000 MHz) and later applied to a more limited set recorded at 500, 600, and 800 MHz. Relaxation rates were recorded for a uniformly nitrogen-15-labeled sample of the protein Engrailed 2. Engrailed 2 is a transcription factor that possesses a well-folded DNA-binding homeodomain and a long, 200-residue, mostly disordered N-terminal region. The disordered region plays a crucial role in the regulation of the activity of the protein and, in particular, in binding to transcriptional regulators (29,30). We decided to study an Engrailed 2 fragment (residues 146–259) encompassing the folded homeodomain (residues 200–259) and an N-terminal 54-residue disordered region (residues 146–199) (31). The results of IMPACT show that motions with correlation times close to 1 ns dominate reorientational dynamics in the most disordered regions of the protein, which is believed to be a general property of IDPs and IDRs (32). Yet, the broad variability of correlation times of backbone motions throughout the disordered region of Engrailed stands in stark contrast with the homogeneous dynamic properties of the folded homeodomain. This study reveals a surprising richness of backbone dynamics in IDPs and IDRs on pico- and nanosecond timescales, not found in folded proteins that have been widely studied over the past three decades.

MATERIALS AND METHODS

Sample

All experiments were performed on a sample of uniformly nitrogen-15-labeled chicken Engrailed 2 (residues 146–259) at a concentration of 0.6 mM in 40 mM sodium succinate buffer at pH 6 supplemented with 1 μg/mL of each of the three protease inhibitors leupeptin, pepstatin, and AEBSF, as well as 10 μM EDTA, which allow one to increase the lifetime of the protein (33). The protein was prepared as described elsewhere (31). Note that the protein construct comprises the residues Gly-Pro-Met at the N-terminus before residue Glu[146], which remain after cleavage of the GST-tag by PreScission protease (GE Healthcare, Little Chalfont, UK). All experiments were carried out at 303 K, which was adjusted in each spectrometer to have a chemical shift difference of 1.462 ppm between the signals of the methyl and hydroxyl protons of pure methanol (4% protonated and 96% deuterated).

NMR spectroscopy

The relaxation rates were measured at five different static fields of 9.4, 11.7, 14.1, 18.8, and 23.5 T, with corresponding proton Larmor frequencies of 400, 500, 600, 800, and 1000 MHz. Three aliquots of the same sample

were used for all experiments, except at 11.7 T, which was performed on a separate, but identical, sample.

At each field, a full set of ^{15}N relaxation measurements was obtained. The longitudinal relaxation rates, $R_1(^{15}N)$, were obtained in the traditional way (34–36), with saturation of the water signal for each scan, whereas the transverse relaxation rates, $R_2(^{15}N)$, were recorded with a train of ^{15}N π-pulses (Carr-Purcell-Meiboom-Gill pulse train), interleaved with ^1H π-pulses to suppress cross-correlated relaxation effects. ^{15}N-$\{^1H\}$ NOEs were obtained by detecting the ^{15}N steady-state polarization while saturating the protons with a train of π-pulses, with suitable interpulse delays and rf amplitudes (37,38). Finally, experiments to measure the transverse and longitudinal cross-relaxation rates due to correlated fluctuations of the nitrogen-15 chemical shift anisotropy (CSA) and the dipolar coupling between the ^{15}N nucleus and the amide proton were recorded using the so-called symmetrical reconversion principle (39,40). All experiments were recorded on Bruker Avance spectrometers (Billerica, MA). Experiments at 500 MHz, 800 MHz, and 1 GHz, and the NOE at 600 MHz, have been recorded using triple-resonance indirect-detection cryogenic probes (41) equipped with z-axis pulsed-field gradients. Other experiments at 600 MHz were recorded on an indirect-detection triple-resonance probe with triple-axis gradients with detection coils at room temperature. Experiments at 400 MHz were recorded on a liquid-nitrogen-cooled cryogenic probe (Prodigy BBO, Bruker) equipped with a z-axis gradient.

Spectral density analysis

The full analysis was carried out at 11 points obtained with the reduced spectral mapping, $J(0.87\omega_H)$ and $J(\omega_N)$ at five fields and $J(\omega = 0)$ calculated from relaxation rates measured at 23.5 T. Analyses with two sets of three magnetic fields used seven points on the spectral density function; $J(\omega = 0)$ was derived from the relaxation rates measured at the highest magnetic field, i.e., 18.8 T or 23.5 T. A Monte Carlo simulation with 510 steps

was performed to evaluate the error of each parameter, A_i. All simulations were carried out with Mathematica (42).

Supporting Material

The Supporting Material includes tables of all relaxation rates used in the analysis and tables of all parameters resulting from conventional analysis with two and three correlation times as well as from our IMPACT analysis; equations relevant for reduced spectral density mapping; a plot of transverse relaxation rates, R_2, measured at 18.8 T; a comparison of Akaike's Information Criteria (AIC) for IMPACT and conventional analyses with two or three correlation times; one-dimensional plots of AIC for five- and six-correlation-time analyses; a correlation of consecutive IMPACT coefficients; plots of IMPACT coefficients and the IMPACT barcode representation of the analysis of relaxation rates based on data recorded at a set of three fields, which cover a broad range (9.4, 14.1, and 23.5 T) and at a set of three more widely accessible fields (11.7, 14.1, and 18.8 T); and IMPACT coefficients for an analysis with five correlation times and $\tau_{max} = 38$ ns.

RESULTS AND DISCUSSION

Secondary structure

Fig. 1 e displays the secondary structure propensity (SSP) (43) based on the assignment of the protein (31). The three α-helices of the homeodomain (residues 200–259) are well identified by SSP scores close to 1. Another region, including the so-called hexapeptide (residues 169–174, WPAWVY) and surrounding residues, displays SSP scores close to 0.3, thus highlighting the presence of some
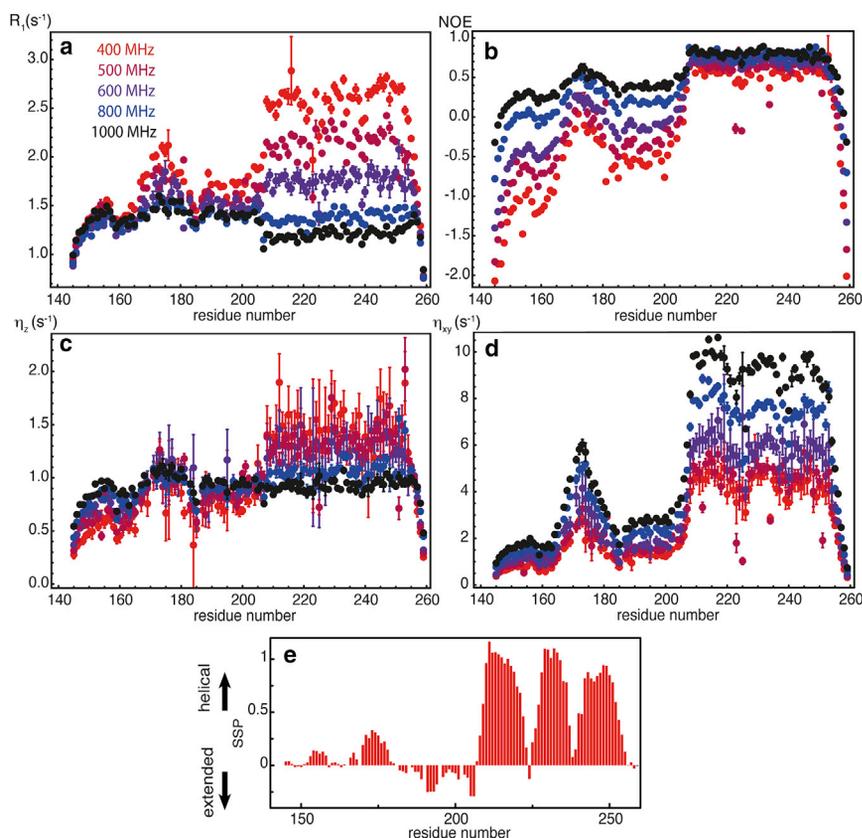

FIGURE 1 Backbone ^{15}N relaxation rates and NOEs measured in Engrailed 2 at five magnetic fields: 400 MHz (*red*), 500 MHz (*burgundy*), 600 MHz (*purple*), 800 MHz (*blue*), and 1000 MHz (*black*). (*a*) Longitudinal relaxation rates, R_1, of ^{15}N. (*b*) ^{15}N-$\{^1H\}$ NOE ratios. (*c*) Longitudinal cross-relaxation rates, η_z, due to correlated fluctuations of the ^{15}N CSA and the ^{15}N-^1H dipolar couplings. (*d*) Transverse cross-relaxation rates, η_{xy}, due to the same correlated fluctuations. (*e*) SSP calculated from the chemical shifts of carbonyl and α and β carbon-13 nuclei. To see this figure in color, go online.

(residual) structure. The region connecting the hexapeptide and the homeodomain features negative SSP scores, which suggests a trend toward extended conformations.

Relaxation experiments were carried out at 400, 500, 600, 800, and 1000 MHz (Fig. 1, a–d) to determine longitudinal R_1 nitrogen-15 relaxation rates, the steady-state ^{15}N-{^1H} NOE, as well as the longitudinal η_z and transverse η_{xy} cross-relaxation rates due to correlated fluctuations of the nitrogen-15 CSA and the dipolar coupling with the amide proton. Transverse relaxation rates, R_2, were measured using Carr-Purcell-Meiboom-Gill echo trains at 800 MHz (Fig. S1 in the Supporting Material). All experiments were analyzed with NMRPipe and the intensities were obtained from a fit of peaks with the routine nlinLS (44). In some exceptional cases, the limited resolution of spectra measured at 9.4 and 11.7 T may have led to inaccuracies in the intensities of a few poorly resolved peaks.

The uniform decrease of the longitudinal relaxation rates, R_1, with increasing magnetic field B_0 in the 200–259 homeodomain (Fig. 1 a) indicates motions in the nanosecond range, resulting from overall rotational diffusion. The variations $R_1(B_0)$ are much less pronounced in the disordered region, except in the 169–174 hexapeptide region. Longitudinal cross-relaxation rates, η_z (Fig. 1 c), increase with B_0 in the IDR. This reflects the very slow decay, slower than $1/\omega$, of the spectral density function in the range 40–100 MHz (i.e., the range of ^{15}N Larmor frequencies between 9.4 and 23.5 T), as the increase of the amplitude of the CSA interaction counterbalances the decay of the spectral density function with increasing frequency. The profile of transverse relaxation rates, R_2, is marked by variations along the sequence of the protein of both the distribution of picosecond-nanosecond motions and contributions of chemical exchange (Fig. S1). NOEs are sensitive markers of local order in IDPs and IDRs and have been used as such for many years (45). Indeed, the variations of NOEs along the sequence are pronounced at moderate fields (9.4–14.1 T); however, the profile of NOEs is much flatter at high fields, in particular at 23.5 T. On the other hand, transverse cross-correlation rates, η_{xy}, which depend primarily on $J(\omega = 0)$, exhibit sharp variations at all fields that are strongly correlated with SSP scores. This suggests that transverse cross-correlation rates, η_{xy}, should become the method of choice to characterize order in IDPs and IDRs.

Spectral density mapping

Most current software packages designed to characterize protein dynamics based on relaxation rates (46–50) offer a direct derivation of the parameters of local dynamics (order parameters and correlation times for local motions). This approach is efficient and reliable when the theoretical framework of the analysis has been validated. Since a general understanding of motions in intrinsically disordered
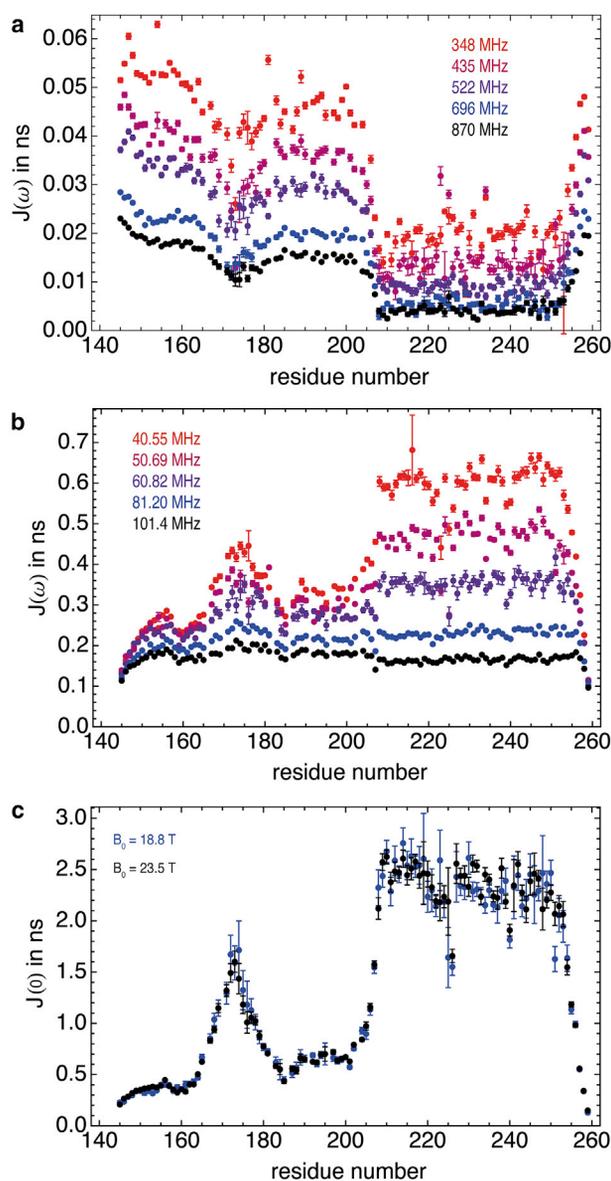

FIGURE 2 Spectral density functions for backbone NH vectors in Engrailed 2. (a) Effective spectral density near the proton Larmor frequency, $J(0.87\omega_H)$ (ns). (b) Spectral density at the Larmor frequency of nitrogen-15, $J(\omega_N)$ (ns). (c) Spectral density at zero frequency, $J(0)$ (ns). All data are color-coded as a function of the magnetic field at which the relaxation rates were recorded, with the same code as in Fig. 1. To see this figure in color, go online.

proteins is still lacking, the derivation of the spectral densities from relaxation rates provides a representation of experimental data that is more amenable to physical analysis than relaxation rates (20).

Spectral density mapping (13) can be achieved without resorting to any proton auto-relaxation rate (51,52). The effective spectral density at high frequency, $J(0.87\omega_H)$ (see Fig. 2 b), can be derived from {^1H}-^{15}N NOE and longitudinal nitrogen-15 relaxation rates for different magnetic fields according to (51)

$$J(0.87\omega_{\mathrm{H}}) = \frac{4\gamma_{\mathrm{N}}R_1(NOE - 1)}{5d^2\gamma_{\mathrm{H}}}, \quad (1)$$

with $d = (\mu_0/4\pi)(\hbar\gamma_{\mathrm{H}}\gamma_{\mathrm{N}}/\langle r_{\mathrm{NH}}^3\rangle)$. μ_0 is the permittivity of free space; γ_{H} and γ_{N} are the gyromagnetic ratios of the proton and nitrogen-15 nuclei; $\hbar$ is Planck's constant divided by 2π and $r_{\mathrm{NH}} = 1.02$ Å is the distance between the amide proton and the nitrogen-15 nucleus.

Our data recorded at five magnetic fields allowed us to fit the spectral density function at high frequency, $J(0.87\omega_{\mathrm{H}})$, to the expression

$$J(\omega) = \lambda + \frac{\mu}{\omega^2}, \quad (2)$$

in analogy to an earlier study of carbon-13 relaxation (53). The parameters λ and μ are real positive numbers. This functional form is expected to be a good approximation of the spectral density at high frequency in a folded protein, but not necessarily for a protein with significant motions with correlation times in the hundreds of picoseconds. Nevertheless, we obtain satisfactory fits for all residues in the IDR as well as in the homeodomain. This validates the self-consistency of the use of a single effective frequency, $\omega_{\mathrm{eff}} = 0.87\omega_{\mathrm{H}}$, in Eqs. 8–10 of the article by Farrow et al. (51) (see Eqs. S1–S6), where the spectral density function at high frequency was assumed to be of the form of Eq. 2 in both the folded homeodomain and the IDR. Thus, most results of spectral density mapping pertaining to disordered proteins that have been published in recent years are validated.

The results of the fit of the spectral density were used to evaluate contributions to the spectral density at higher frequencies (at $\omega_H \pm \omega_{\mathrm{N}}$) in the derivation of $J(\omega_{\mathrm{N}})$ from the rate R_1 according to the equation

$$\begin{aligned} J(\omega_{\mathrm{N}}) = {} & R_1/(3d^2/4 + c^2) - (6J(\omega_H + \omega_{\mathrm{N}}) \\ & + J(\omega_H - \omega_{\mathrm{N}}))/(3 + 4c^2/d^2), \end{aligned} \quad (3)$$

with $c = \gamma_{\mathrm{N}}B_0\Delta\sigma/\sqrt{3}$ and $\Delta\sigma = 160$ ppm is the axially symmetric CSA of the nitrogen-15 nucleus.

Overall, contributions of high-frequency terms to R_1 are small (54), so that the deviations between the values of $J(\omega_{\mathrm{N}})$ obtained from a series of approximations (51) and the current method are limited to ~2%, which is commensurate with the estimated precision (Fig. S2). Again, this validates, a posteriori, many spectral density mapping studies performed on IDPs. In addition, the low sensitivity of $J(\omega_{\mathrm{N}})$ upon the model used to describe the spectral density at high frequency shows that the enhanced accuracy expected from more sophisticated approaches, for instance, following Kaděrávek et al. (20), would be smaller than the typical precision of our measurements. The values of $J(\omega_{\mathrm{N}})$ derived at five magnetic fields are shown in Fig. 2 b.

To avoid contributions from line-broadening due to chemical exchange, we did not consider transverse relaxation rates, $R_2(^{15}\mathrm{N})$, and only used longitudinal and transverse CSA/DD cross-correlated relaxation rates, η_z and η_{xy}, to derive the spectral density $J(0)$ from $J(\omega_{\mathrm{N}})$ using (55)

$$J(0) = J(\omega_{\mathrm{N}})\frac{3}{4}\left(2\frac{\eta_{xy}}{\eta_z} - 1\right). \quad (4)$$

As can be seen in Fig. S3, measurements of η_z and η_{xy} are not precise enough at lower fields to provide reliable estimates of $J(0)$ by lack of sensitivity (in particular for η_z). However, the data recorded at 18.8 T and 23.5 T (Fig. 2 c) are very similar and do not show any of the outliers observed at lower fields. Significant chemical exchange contributions to $R_2(^{15}\mathrm{N})$ can be observed in the hexapeptide region of the disordered region and in the homeodomain (see Fig. S1). Such contributions preclude the proper derivation of $J(0)$ from $R_2(^{15}\mathrm{N})$ rates, in particular at high magnetic fields.

Principles and optimization of IMPACT

The limitations of conventional approaches to the analysis of relaxation rates in IDPs and IDRs result from the complexity of their dynamics. These span at least three orders of magnitude, so that it appears unlikely that they can be accurately described by a single distribution of correlation times or by a small number of correlation times. However, the scarcity of relaxation rates limits the number of adjustable parameters that can be determined and thus the sophistication of spectral density functions that can be postulated. Here, we significantly increase the number of correlation times by defining an array of n fixed correlation times. Only the relative coefficient of each correlation time in the distribution is fitted to experimental data, so that the number of adjustable parameters is reduced. Thus, our only assumption is that the correlation function can be approximated by a sum of exponentials. The physical content of the IMPACT model is thus limited to a minimum. IMPACT can be described as a mathematical approach that converts experimental relaxation rates (or, equivalently, spectral density mapping results) into a distribution of correlation times that is more amenable to physical interpretation than the raw experimental data. The array of n correlation times is defined as a geometric series, so that correlation times are equally spaced on a logarithmic scale (Fig. 3, a and b):

$$\tau_i = \alpha^{i-1}\tau_{\max} \quad \alpha = (\tau_{\min}/\tau_{\max})^{\frac{1}{n-1}}. \quad (5)$$

Thus, $J(\omega)$ is a sum of Lorentzian functions:

$$J(\omega) = \sum_{i=1}^{n} J_i(\omega) = \frac{2}{5}\sum_{i=1}^{n}\frac{A_i\tau_i}{1 + (\omega\tau_i)^2}, \quad (6)$$

where A_i is the coefficient of correlation time τ_i in the spectral density function. The coefficients A_i must be positive and fulfill the normalization constraint
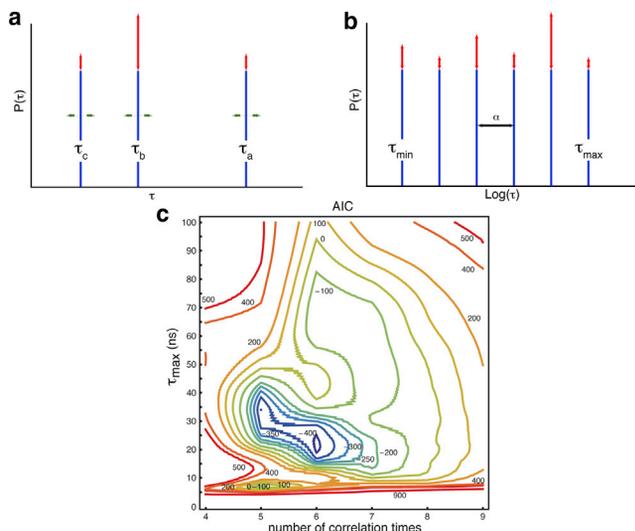
FIGURE 3 Principle and optimization of the parameters of our IMPACT analysis. (*a*) In the 3CT analysis, both the value and the relative weight of each correlation time must be adjusted. (*b*) In IMPACT, the values of the correlation times are fixed and equally spaced on a logarithmic scale, so that only their relative weights need adjusting. (*c*) Optimization of IMPACT by considering AIC. The range (τ_{min}, τ_{max}) of correlation times characterized by IMPACT was varied from (1 ps, 1 ns) to (100 ps, 100 ns) and the number of correlation times was varied in the range $n = 4$–9. Despite the solid lines shown in the contour plot (*c*), the reader should be aware that the number of correlation times is an integer. To see this figure in color, go online.

$$\sum_{i=1}^{n} A_i = 1. \qquad (7)$$

Thus, the number of free parameters is $n-1$.

A preliminary step of the IMPACT analysis is the optimization of the three parameters τ_{min}, τ_{max}, and n. The first step is to define the range of correlation times that are probed by relaxation rates. A series of correlation times could be chosen as the inverse of the Larmor frequencies at which the spectral density is mapped, in analogy to the study by LeMaster (56). Considering that the range of frequencies where a Lorentzian function varies extends beyond the inflection point, we chose a slightly different approach. We first define the range of correlation times that are sampled by various ^{15}N relaxation rates. We consider that the lowest magnetic field adapted to protein studies is 400 MHz, whereas the highest accessible field currently is 1 GHz. Thus, the lowest nonzero frequency at which the spectral density is sampled is $\omega_N/2\pi = 40$ MHz, and the highest is $0.87\omega_H/2\pi = 870$ MHz. A Lorentzian function with a correlation time of $\tau_c = 40$ ns drops to 1% of $J(0)$ at $\omega_N/2\pi = 40$ MHz. A Lorentzian with $\tau_c = 18$ ps merely decreases to 99% of $J(0)$ at $0.87\omega_H/2\pi = 870$ MHz. The resulting range spans slightly more than three orders of magnitude. Therefore, we have decided to limit the range to three orders of magnitude

$$\tau_{max}/\tau_{min} = 10^3. \qquad (8)$$

To define the optimal values of τ_{min} and τ_{max}, we carried out a series of IMPACT analyses for $[\tau_{min}, \tau_{max}] = [1$ ps, 1 ns] to [100 ps, 100 ns], as well as for $4 < n < 9$. In contrast to the approach of LeMaster (56), the number of correlation times is adjustable. The statistical relevance of each combination of parameters was evaluated from the resulting Akaike's information criteria (AIC) (57–60):

$$\text{AIC} = n_{exp}\ln\left(\sum_{k=1}^{nres} \chi_k^2/n_{exp}\right) + 2n_{model} + C. \qquad (9)$$

$n_{exp} = n_J \times nres$ is the total number of experimental data, with $n_J = 11$ points at which the spectral density function $J(\omega)$ is sampled when relaxation data at five magnetic fields are used; $nres = 108$ is the number of residues included in the analysis; and $n_{model} = (n - 1) \times nres$ is the number of free parameters in each model. Here, the constant is $C = 0$. AIC are shown in Fig. 3 *c*. Two local minima were found for $[\tau_{min}, \tau_{max}] = [34$ ps, 34 ns] with $n = 5$ and for $[\tau_{min}, \tau_{max}] = [21$ ps, 21 ns] with $n = 6$. The likelihood of the latter array of correlation times is 10^3 times higher than that of the former. Here, we will thus present the IMPACT analysis with $[\tau_{min}, \tau_{max}] = [21$ ps, 21 ns] and $n = 6$; the analysis with $[\tau_{min}, \tau_{max}] = [34$ ps, 34 ns] and $n = 5$ can be found in the Supporting Material. This result is dominated by the diverse dynamic properties of the IDR of Engrailed. Indeed, if we exclude the rigid residues of the homeodomain, the optimal parameters change slightly (residues 146–207) to $[\tau_{min}, \tau_{max}] = [42$ ps, 42 ns] with $n = 5$, whereas the optimal set of parameters for the rigid part of the homeodomain alone (residues 208–259) is significantly different, $[\tau_{min}, \tau_{max}] = [10$ ps, 10 ns] and $n = 4$.

Application of IMPACT to Engrailed

The optimal parameters $n = 6$ and $[\tau_{min}, \tau_{max}] = [21$ ps, 21 ns] were employed to analyze the spectral density function in Engrailed. Note that the coefficients A_i were fitted to spectral density mapping results. In principle, relaxation rates could also be used directly as input for the IMPACT analysis. Fig. 4 illustrates the remarkable variety of dynamics found in Engrailed. In the homeodomain, the second correlation time, τ_2, lies just below the correlation time for overall rotational diffusion, which is close to 7 ns, as can be seen from the analysis based on only two correlation times (vide infra). Thus, the second coefficient is by far the most important in the homeodomain. A small amplitude A_1 of τ_1 corrects for the fact that τ_2 is shorter than the actual correlation time for the motion of the whole domain. Note that the correlation time for overall diffusion, τ_m, is well approximated by:

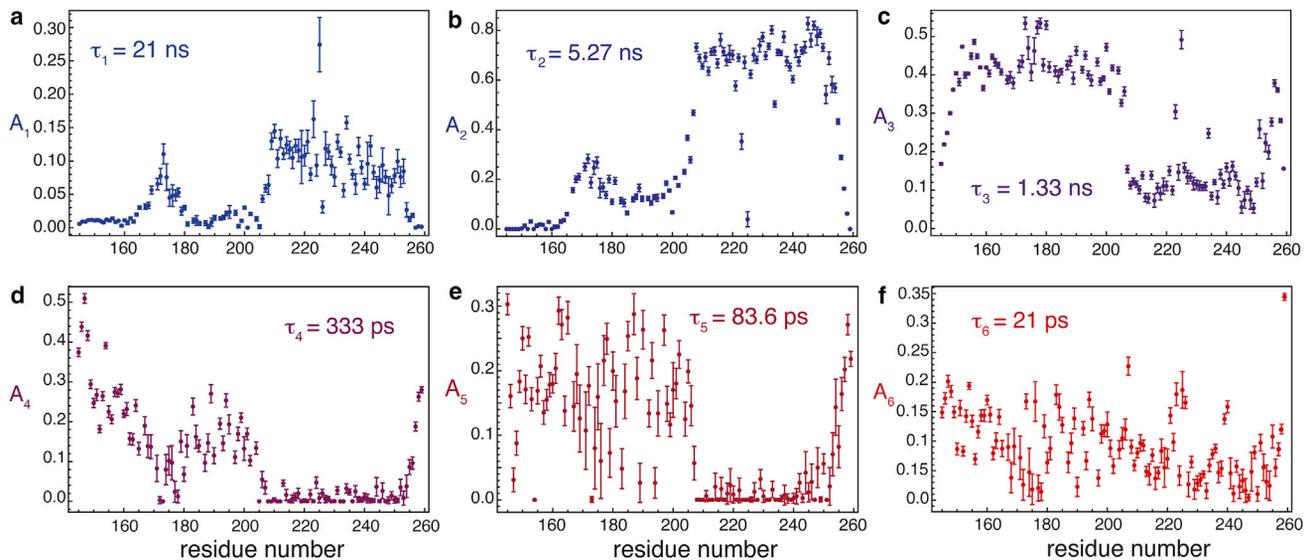$$\tau_m \approx (A_1\tau_1 + A_2\tau_2)/(A_1 + A_2). \qquad (10)$$

FIGURE 4 Plots of the six coefficients, A_i (i = 1, 2... 6) of the n = 6 correlation times, τ_i, in the range $[\tau_{min}, \tau_{max}]$ = [21 ps, 21 ns] determined by the IMPACT analysis of Engrailed: (a) τ_1 = 21 ns; (b) τ_2 = 5.27 ns; (c) τ_3 = 1.33 ns; (d) τ_4 = 333 ps; (e) τ_5 = 83.6 ps; (f) τ_6 = 21 ps. To see this figure in color, go online.

The average value over the helices of the homeodomain is $<\tau_m>$ = 7.19 ns, which is in good agreement with an estimate of the correlation time for overall diffusion (see below). Small but significant and mostly uniform contributions A_3 of the correlation time τ_3 in the three α-helices, which are also obtained in a conventional model-free analysis (vide infra), may be attributed to fluctuations of the overall diffusion tensor (61), likely due to conformational fluctuations of the IDR (residues 146–199). Enhanced values of A_3 in the loops may reflect the flexibility of these regions (41). The very small coefficients A_4 and A_5 demonstrate the presence of a gap in the distribution of correlation times, as was also observed in ubiquitin (62). Finally, the coefficients A_6 for the shortest correlation time τ_6 indicate the presence of fast motions in the tens of picoseconds range. Note that the Lorentzian function $J_6(\omega)$ drops by ~1% of $J_6(0)$ at the highest frequency explored in this analysis (i.e., $\omega/2\pi$ = 870 MHz). Thus, this last term in the spectral density function can be approximated to a constant that effectively represents all fast motions:

$$J_6(\omega) \approx \frac{2}{5} A_6 \tau_6 \approx \frac{2}{5} \int_0^{\tau_5} p(\tau) d\tau, \qquad (11)$$

where $p(\tau)$ is the probability function of correlation times, containing little information on the complexity of such motions (63).

Results obtained in the disordered region of Engrailed will be discussed with the help of Fig. 4 but also with the IMPACT barcode shown in Fig. 5. In the latter, for each residue, the width of each histogram represents the coefficient A_i associated with the correlation time τ_i that can be read on

the y axis. This graph appears to be a convenient way to display the results of the IMPACT analysis in a single figure.

For the first residues at the N-terminus and the last residues at the C-terminus, significant coefficients A_{3-6} are found for the four shortest correlation times. This seems to indicate the presence of motions that are broadly distributed over all timescales up to 1 ns. On the other hand, the two disordered regions just at the N-terminus and the C-terminus of the hexapeptide display a high density of motions around τ_3. The coefficients for the correlation time τ_4 decrease almost linearly with the distance to the N- or C-termini of the polypeptide chain in disordered regions and reach different plateaus in each disordered segment. A notable difference between the disordered region at the N-terminus and the one between the hexapeptide and the homeodomain is the slight decrease of the coefficient A_3 and a significant increase of A_2. It is difficult to assign this change to a particular process without a better characterization of the conformational space of the protein (such characterization is beyond the scope of this article and is a work in progress). Nevertheless, two effects may contribute to the presence of some orientational order beyond 1 ns. First, this IDR is short and located between a folded domain and a small hydrophobic cluster, so that its dynamics is likely influenced by the overall diffusion of both structured elements. Second, this IDR contains a majority of residues that favor extended conformations, as confirmed by SSP scores: three proline residues, nine positively charged, and only one negatively charged residue between positions 177 and 198, which should restrict the conformational space and possibly slow down reorientational dynamics.

The barcode representation of IMPACT coefficients nicely illustrates variations of the ensemble of correlation times
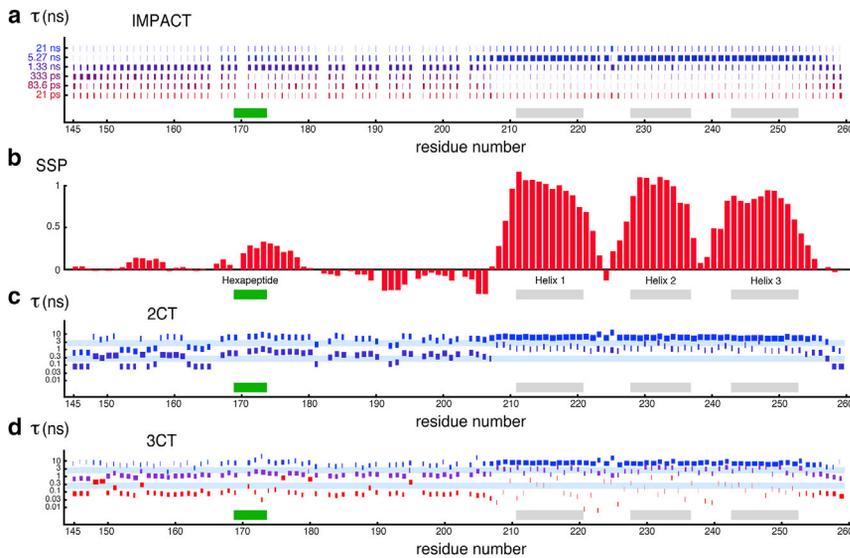
FIGURE 5 Graphical representations of (a) IMPACT, (c) 2CT, and (d) 3CT analyses of the spectral density function in Engrailed 2. Histograms are drawn for all residues and represent the contributions of (a) each of the six correlation times, τ_i ($i = 1, 2, …6$), considered in IMPACT, (c) each of the two correlation times, $\tau_{a,b}$, determined by the 2CT analysis, (d) each of the three correlation times, $\tau_{a,b,c}$, determined by the 3CT analysis. The width of each rectangle is proportional to the corresponding weights A_i in IMPACT (a), $B_{a,b}$ in 2CT (c); and $B_{a,b,c}$ in 3CT (d). In (c) and (d), the light blue horizontal bars represent the ranges of correlation times, τ, for which reciprocal frequencies lie in the constrained regions between $40 < 1/(2\pi\ \tau) < 100$ MHz or between $348 < 1/(2\pi\tau) < 870$ MHz. Gray rectangles in (a), (c), and (d) indicate rigid α-helices, and a green rectangle shows the rigid hydrophobic hexapeptide sequence. (b) As in Fig. 1 e, the SSP is shown to guide the comparison between structural and dynamic features. To see this figure in color, go online.

between successive structural elements. Thus, for instance, the decrease of motions in the subnanosecond range in the hexapeptide is accompanied by an increase in the suprananosecond range. Similarly, variations of the coefficients for correlation times at the N- and C-termini of the homeodomain (residues 200–210 and 254–260) illustrate the smooth transition of motional properties along the sequence of the polypeptide. Finally, even in the homeodomain, where a classical analysis of relaxation data should be most appropriate, the dynamic transitions between helices and the two loops are clearly visible and quantitatively characterized by the IMPACT approach. Loop $\alpha1$-$\alpha2$ features enhanced dynamics in both the 1 ns and tens of picoseconds ranges, as expected from the motions demonstrated by paramagnetic relaxation enhancement studies (41), whereas loop $\alpha2$-$\alpha3$ shows a significant but more moderate enhancement of motions. Overall, the IMPACT representation offers an elegant view of the correlation of structural and dynamic features, as can be seen from the SSP scores.

Comparison with conventional analyses based on two or three correlation times

For the sake of comparison, we also fit two simple models where the spectral density function is assumed to consist of a sum of two and three Lorentzians in the manner of the familiar model-free and extended model-free approaches. However, as discussed in the Introduction, the core hypotheses of the model-free formalism cannot be fulfilled a priori, since the longest correlation time is probably an effective correlation time rather than the correlation time of overall rotational diffusion. The spectral density, J_{2CT}, assuming two correlation times (2CT) can be written as

$$J_{2CT}(\omega) = \frac{2}{5}\left[S^2\tau_a\big/\left(1 + (\omega\tau_a)^2\right)\right.$$
$$\left. + \left(1 - S^2\right)\tau'_b\big/\left(1 + \left(\omega\tau'_b\right)^2\right)\right], \tag{12}$$

where $\tau'^{-1}_b = \tau^{-1}_a + \tau^{-1}_b$, τ_a is the long correlation time, τ_b is the short effective correlation time, and S^2 is similar to the model-free order parameter. The spectral density function J_{3CT}, assuming three correlation times (3CT), can be defined as

$$J_{3CT}(\omega) = \frac{2}{5}\left[S^2\tau_a\big/\left(1 + (\omega\tau_a)^2\right) + \left(S^2_f - S^2\right)\tau'_b\big/\right.$$
$$\left(1 + \left(\omega\tau'_b\right)^2\right) + \left(1 - S^2_f\right)\tau'_c\big/\left(1 + \left(\omega\tau'_c\right)^2\right)\bigg], \tag{13}$$

where $\tau'^{-1}_c = \tau^{-1}_a + \tau^{-1}_c$, $\tau_a > \tau_b > \tau_c$, and S^2_f is equivalent to the extended model-free order parameter for fast processes with a correlation time τ_c. The two functions were fitted to the experimental spectral density, and a simple model selection was based on the comparison of the second order variant of AIC (see the Supporting Material), with $n_J = 11$ and $n_{model} = 3$ for the 2CT analysis and 5 for the 3CT analysis.

Results of this analysis are shown in Fig. 6. From a statistical point of view, the 2CT analysis is found to be sufficient to describe the motions in the mostly rigid homeodomain (residues 200–259), except at the flexible N- and C termini. With few exceptions, the longest correlation time yields a reliable measure of overall rotational diffusion, and the average value over the helices of the homeodomain is $<\tau_a>_{hom} = 7.08$ ns, in good agreement with the IMPACT analysis, with $<\tau_m> = 7.19$ ns. Interestingly, the second
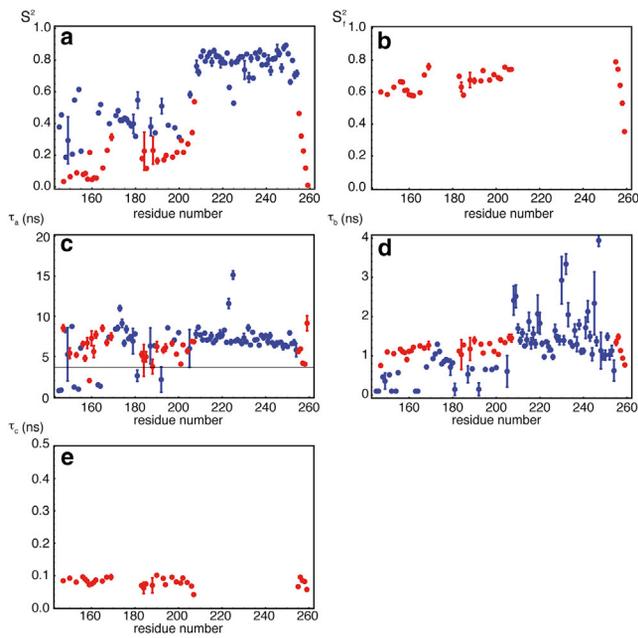
FIGURE 6 Results obtained for a conventional analysis with 2CT (*blue*) or 3CT (*red*). (*a*) Order parameter S^2. (*b*) Order parameter S^2_f for the fastest motion in the 3CT analysis. (*c*) Longest correlation time, τ_a. (*d* and *e*) Intermediate correlation time, τ_b (*d*), and shortest correlation time, τ_c (*e*). Either the 2CT or the 3CT model was selected based on the lowest AIC. To see this figure in color, go online.

correlation time, τ_b, found for most residues in the rigid homeodomain lies in the range $0.9 < \tau_b < 1.8$ ns. This is in agreement with the IMPACT analysis and is possibly due to fluctuations of the overall diffusion tensor resulting from conformational transitions in the disordered N-terminal region on timescales between 1 and 100 ns (61,64). In the disordered region (residues 146–199), the 2CT and 3CT analyses seem equally probable, with no particular pattern along the sequence, except in the hydrophobic hexapeptide cluster (residues 169–174), where the 2CT analysis is more satisfactory. The random patterns of 2CT versus 3CT selection seems to point to some instability of the model-selection step in the fit procedure. A 2CT or 3CT analysis can be performed with no model selection (see Fig. 5). The built-in absence of site-specific model selection in IMPACT shields this analysis from such a drawback. Low order parameters S^2 are found throughout the IDR, with a significant increase in order in the hydrophobic cluster. The long correlation times in the disordered regions have a broad distribution (standard deviation of 2.5 ns), but the average value, $\langle\tau_a\rangle_{IDR} = 5.9$ ns, is similar to what was found in unfolded proteins (32) and IDPs (65) and very close to the correlation time for overall tumbling of the homeodomain. The intermediate correlation time, which corresponds to the dominant term in the spectral density function, lies in the range $0.1 < \tau_b < 1.4$ ns, in agreement with the IMPACT analysis. The shortest correlation time is rather uniform and lies in the range $40 < \tau_c < 120$ ps.

One should be careful with the physical interpretation of these observations in the IDR of Engrailed. The three correlation times obtained are clearly separated in the time domain, which indicates a broad range of dynamic processes. The results should not be considered a priori as actual correlation times of particular motions, but rather as the best rendition of experimental results using two or three correlation times. This is illustrated by the jumps of order parameters and correlation times observed between the 2CT and 3CT models in the IDR, which illustrates the effective character of the fitted correlation times in this region, at least in the 2CT analysis. For instance, it is difficult to assign the longest correlation time to any particular dynamical process in the absence of complementary experimental or computational information. Such a process could be a single well-defined type of motion, such as the rotational diffusion of an IDR segment. Alternatively, the longest correlation time might account for the tail of a continuous distribution of correlation times and reflect slower motions in parts of the conformational space of the IDR. Interestingly, the correlation times obtained in a 3CT analysis often correspond to reciprocal frequencies ($\omega = 1/\tau$) that lie outside regions where the spectral density function can be adequately sampled (i.e., below 40 MHz, between 100 MHz and 348 MHz, and above 870 MHz). This is particularly true in the flexible region between the rigid hexapeptide and the rigid homeodomain and at the C-terminus of the protein. The regions where the spectral density function is most sensitive to the choice of correlation times correspond to ranges where we lack experimental constraints. This would be expected in the presence of a broad distribution of correlation times that would lead to a smooth decay of the spectral density function.

A direct comparison between the results of the 3CT analysis and our IMPACT approach is illustrated in Fig. 5. For the sake of comparison, we define in Fig. 5 *a* the coefficients $B_{a,b,c}$ associated with the correlation times $\tau_{a,b,c}$, as

$$B_a = S^2, B_b = S^2_f - S^2, B_c = 1 - S^2_f. \qquad (14)$$

The statistical significance of the fit resulting from our IMPACT analysis is often better than with either the 2CT or the 3CT analysis in the disordered regions, although it is somehow comparable to that of the 2CT model in the rigid homeodomain (Fig. S4). In particular, the almost complete absence of abnormally elevated χ^2 values in the IDR of Engrailed shows that a faithful set of fitted A_i parameters can be obtained with diverse dynamical features. Interestingly, as can be seen from the schematic representation of correlation times that correspond to reciprocal frequencies that can be determined by spectral density mapping, the inflection points of most of the Lorentzian functions often lie in frequency regions where spectral density mapping does not yield any results. This is particularly true in the IDR between the hexapeptide and the homeodomain and at the C-terminus of the protein. This seems to indicate that the

decay of the spectral density function is smoother than can be described by a sum of three Lorentzian functions. The fit pushes the inflection points of individual contributions to the spectral density function beyond the areas that benefit from rich experimental constraints. In addition, the absence of a residue-specific model selection in IMPACT provides results that are directly comparable, residue by residue, which allows for a better qualitative description of dynamic properties along the protein sequence. Admittedly, model selection can be omitted in the 2CT or 3CT analysis, as in Fig. 5, *a* and *b*.

A potential concern of the IMPACT analysis lies in the fact that the correlation functions $J_i(\omega)$ are not independent, since they suffer from significant overlap. Hence, it is possible that different ensembles of coefficients A_i can describe the same experimental data. To test the sensitivity of our analysis to this potential flaw, we have plotted correlations of consecutive coefficients A_i (i.e., A_i as a function of A_{i+1}) for all 510 steps of the Monte Carlo procedure employed in the fit. Typical results are shown in Fig. S6. There is a small anticorrelation between consecutive coefficients in several instances. This will give rise to a broader distribution of individual coefficients and thus lead to a decrease of the precision of these coefficients. In the worst case, a potential decrease of accuracy due to the interdependence of consecutive coefficients will be accompanied with a decrease in precision.

A potential concern of the IMPACT analysis is the risk of overinterpretation of the results. Here, we should be clear and provide a set of rules that should be followed by users of this approach.

1) The correlation times, τ_i, are not physical correlation times of the system a priori. The range of correlation times is defined by the experimental observables, but the individual values τ_i are derived from a statistical analysis, not a physical analysis.
2) A nonzero coefficient A_1, with τ_1 the longest correlation time, means not that some motions with a correlation time τ_1 were detected but that the distribution of correlation times is larger than zero for some correlation times larger than τ_2.
3) Similarly, as mentioned in Eq. 11, the coefficient A_n of the shortest correlation time τ_n is an effective representation of all fastest motions.
4) If the coefficient A_i is larger than zero, the distribution of correlation times is larger than zero somewhere between τ_{i+1} and τ_{i-1}.
5) If the coefficients A_i and A_{i+1} are both zero, the distribution of correlation times is expected to be zero at least between τ_{i+1} and τ_i.

Finally, very few relaxation studies have compared rates at five or more magnetic fields (66,67). We have tested whether an IMPACT analysis of relaxation rates recorded at only three fields could give meaningful results, using either a broad range of magnetic fields (9.4, 14.1, and 23.5 T) or a narrow range of readily accessible magnetic fields (11.7, 14.1, and 18.8 T). In either case, this requires about two weeks of experimental time. The results of the analysis of relaxation rates at 9.4, 14.1, and 23.5 T shown in Figs. S7 and S8 are remarkably similar to those presented in Figs. 4 and 5. When the range of magnetic fields is restricted, with relaxation rates measured at 11.7, 14.1 and 18.8 T (Figs S9 and S10), some significant changes of IMPACT coefficients can be observed, but the overall description of the distribution of correlation times is very similar to what is obtained with relaxation rates at five magnetic fields. Hence, IMPACT can be applied to many proteins at a moderate cost in experimental time, and does not necessarily require exceptionally large data sets or magnetic fields as high as 23.5 T.

CONCLUSIONS

We have presented a set of nitrogen-15 relaxation rates in the 114-residue, partially disordered protein Engrailed 2 recorded at five different magnetic fields. The transverse cross-correlated rate, η_{xy}, was shown to be the most sensitive to the extent of order and disorder at all magnetic fields. The analysis validates the reduced spectral density mapping approach originally developed for folded proteins and already extensively applied to IDPs and IDRs. The spectral density functions can be fitted reasonably well with two or three correlation times, although such results may be difficult to interpret. We have introduced an approach to the analysis of spectral density functions, which we call IMPACT. This provides a better quantitative description of spectral density functions in IDRs as found in Engrailed than an analysis with three correlation times with the same number of adjustable parameters. We also introduce a barcode representation of IMPACT, which provides a condensed graphical representation of large amounts of data in a single figure. This representation lends itself to a qualitative discussion of order and disorder in proteins. IMPACT can also be useful for analyzing a smaller set of relaxation rates recorded at only three magnetic fields. IMPACT provides a unique framework for the description of the timescales of motions in IDPs and IDRs. Our approach is complementary to the determination of conformational ensembles (7,68). Insight into the dynamics of IDPs and IDRs should greatly benefit from a combined analysis.

SUPPORTING MATERIAL

AUTHOR CONTRIBUTIONS

F.F., O.L., and P.P. designed the research; S.N.K., C.C., R.A., N.S., O.L., P.P., and F.F. performed the research; C.C., N.S., V.D., G.B., and P.P.

REFERENCES

1. van der Lee, R., M. Buljan, …, M. M. Babu. 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114:6589–6631.

2. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208.

3. Uversky, V. N., and A. K. Dunker. 2010. Understanding protein non-folding. *Biochim. Biophys. Acta.* 1804:1231–1264.

4. Dunker, A. K., J. D. Lawson, …, Z. Obradovic. 2001. Intrinsically disordered protein. *J. Mol. Graph. Model.* 19:26–59.

5. Jensen, M. R., P. R. L. Markwick, …, M. Blackledge. 2009. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure.* 17:1169–1185.

6. Jensen, M. R., M. Zweckstetter, …, M. Blackledge. 2014. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* 114:6632–6660.

7. Marsh, J. A., and J. D. Forman-Kay. 2009. Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol.* 391:359–374.

8. Tompa, P. 2012. On the supertertiary structure of proteins. *Nat. Chem. Biol.* 8:597–600.

9. Lindorff-Larsen, K., N. Trbovic, …, D. E. Shaw. 2012. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* 134:3787–3791.

10. Mittermaier, A., and L. E. Kay. 2006. New tools provide new insights in NMR studies of protein dynamics. *Science.* 312:224–228.

11. Palmer, 3rd, A. G. 2004. NMR characterization of the dynamics of bio-macromolecules. *Chem. Rev.* 104:3623–3640.

12. Palmer, 3rd, A. G. 2014. Chemical exchange in biomacromolecules: past, present, and future. *J. Magn. Reson.* 241:3–17.

13. Peng, J. W., and G. Wagner. 1992. Mapping of spectral density-functions using heteronuclear NMR relaxation measurements. *J. Magn. Reson.* 98:308–332.

14. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.

15. Halle, B. 2009. The physical basis of model-free analysis of NMR relaxation data from proteins and complex fluids. *J. Chem. Phys.* 131:224507.

16. Clore, G. M., A. Szabo, …, A. M. Gronenborn. 1990. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112:4989–4991.

17. Eliezer, D. 2009. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 19:23–30.

18. Jensen, M. R., R. W. H. Ruigrok, and M. Blackledge. 2013. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.* 23:426–435.

19. Konrat, R. 2014. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* 241:74–85.

20. Kaderávek, P., V. Zapletal, …, L. Žídek. 2014. Spectral density mapping protocols for analysis of molecular motions in disordered proteins. *J. Biomol. NMR.* 58:193–207.

21. Bussell, Jr., R., and D. Eliezer. 2001. Residual structure and dynamics in Parkinson's disease-associated mutants of α-synuclein. *J. Biol. Chem.* 276:45996–46003.

22. Houben, K., L. Blanchard, …, D. Marion. 2007. Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P. *Biophys. J.* 93:2830–2844.

23. Brutscher, B., R. Brüschweiler, and R. R. Ernst. 1997. Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by ^{1}H, ^{13}C, and ^{15}N nuclear magnetic resonance spectroscopy. *Biochemistry.* 36:13043–13053.

24. Buevich, A. V., and J. Baum. 1999. Dynamics of unfolded proteins: incorporation of distributions of correlation times in the model free analysis of NMR relaxation data. *J. Am. Chem. Soc.* 121:8671–8672.

25. Buevich, A. V., U. P. Shinde, …, J. Baum. 2001. Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. *J. Biomol. NMR.* 20:233–249.

26. Modig, K., and F. M. Poulsen. 2008. Model-independent interpretation of NMR relaxation data for unfolded proteins: the acid-denatured state of ACBP. *J. Biomol. NMR.* 42:163–177.

27. Sternin, E. 2007. Use of inverse theory algorithms in the analysis of biomembrane NMR data. *Methods Mol. Biol.* 400:103–125.

28. Aster, R. C., B. Borchers, and C. H. Thurber. 2012. Parameter Estimation and Inverse Problems, 2nd ed. Academic Press, New York.

29. Foucher, I., M. L. Montesinos, …, A. Trembleau. 2003. Joint regulation of the MAP1B promoter by HNF3β/Foxa2 and Engrailed is the result of a highly conserved mechanism for direct interaction of homeoproteins and Fox transcription factors. *Development.* 130:1867–1876.

30. Peltenburg, L. T. C., and C. Murre. 1996. Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.* 15:3385–3393.

31. Augustyniak, R., S. Balayssac, …, O. Lequin. 2011. ^{1}H, ^{13}C and ^{15}N resonance assignment of a 114-residue fragment of Engrailed 2 homeoprotein, a partially disordered protein. *Biomol. NMR Assign.* 5:229–231.

32. Xue, Y., and N. R. Skrynnikov. 2011. Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, ^{15}N relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? *J. Am. Chem. Soc.* 133:14614–14628.

33. Augustyniak, R., F. Ferrage, …, G. Bodenhausen. 2011. Methods to determine slow diffusion coefficients of biomolecules: applications to Engrailed 2, a partially disordered protein. *J. Biomol. NMR.* 50:209–218.

34. Kay, L. E., L. K. Nicholson, …, D. A. Torchia. 1992. Pulse sequences for removal of the effects of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T_1 and T_2 values in proteins. *J. Magn. Reson.* 97:359–375.

35. Palmer, III, A. G., N. J. Skelton, …, M. Rance. 1992. Suppression of the effects of cross-correlation between dipolar and anisotropic chemical-shift relaxation mechanisms in the measurement of spin spin relaxations rates. *Mol. Phys.* 75:699–711.

36. Ferrage, F. 2012. Protein dynamics by ^{15}N nuclear magnetic relaxation. *Methods Mol. Biol.* 831:141–163.

37. Ferrage, F., D. Cowburn, and R. Ghose. 2009. Accurate sampling of high-frequency motions in proteins by steady-state ^{15}N-^{1}H nuclear Overhauser effect measurements in the presence of cross-correlated relaxation. *J. Am. Chem. Soc.* 131:6048–6049.

38. Ferrage, F., A. Reichel, …, R. Ghose. 2010. On the measurement of ^{15}N-^{1}H nuclear Overhauser effects. 2. Effects of the saturation scheme and water signal suppression. *J. Magn. Reson.* 207:294–303.

39. Pelupessy, P., G. M. Espallargas, and G. Bodenhausen. 2003. Symmetrical reconversion: measuring cross-correlation rates with enhanced accuracy. *J. Magn. Reson.* 161:258–264.

40. Pelupessy, P., F. Ferrage, and G. Bodenhausen. 2007. Accurate measurement of longitudinal cross-relaxation rates in nuclear magnetic resonance. *J. Chem. Phys.* 126:134508.

41. Carlier, L., S. Balayssac, …, O. Lequin. 2013. Investigation of homeodomain membrane translocation properties: insights from the structure determination of engrailed-2 homeodomain in aqueous and membrane-mimetic environments. *Biophys. J.* 105:667–678.

42. Wolfram Research. 2012. Wolfram Mathematica. Wolfram Research, Champaign, IL.

43. Marsh, J. A., V. K. Singh, …, J. D. Forman-Kay. 2006. Sensitivity of secondary structure propensies to sequence differences between α- and γ-synuclein: implications for fibrillation. *Protein Sci.* 15:2795–2804.

44. Delaglio, F., S. Grzesiek, …, A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6:277–293.

45. Dyson, H. J., and P. E. Wright. 2004. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* 104:3607–3622.

46. Mandel, A. M., M. Akke, and A. G. Palmer, 3rd. 1995. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 246:144–163.

47. Cole, R., and J. P. Loria. 2003. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR.* 26:203–213.

48. Fushman, D., S. Cahill, and D. Cowburn. 1997. The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of 15N relaxation with monomer/dimer equilibration. *J. Mol. Biol.* 266:173–194.

49. Dosset, P., J. C. Hus, …, D. Marion. 2000. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR.* 16:23–28.

50. Bieri, M., E. J. d'Auvergne, and P. R. Gooley. 2011. relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and μs motion of proteins. *J. Biomol. NMR.* 50:147–155.

51. Farrow, N. A., O. Zhang, …, L. E. Kay. 1995. Spectral density function mapping using 15N relaxation data exclusively. *J. Biomol. NMR.* 6:153–162.

52. Ishima, R., and K. Nagayama. 1995. Protein backbone dynamics revealed by quasi spectral density function analysis of amide N-15 nuclei. *Biochemistry.* 34:3162–3171.

53. Hill, R. B., C. Bracken, …, A. G. Palmer. 2000. Molecular motions and protein folding: Characterization of the backbone dynamics and folding equilibrium of α_2D using ^{13}C NMR spin relaxation. *J. Am. Chem. Soc.* 122:11610–11619.

54. Ropars, V., S. Bouguet-Bonnet, …, C. Roumestand. 2007. Unraveling protein dynamics through fast spectral density mapping. *J. Biomol. NMR.* 37:159–177.

55. Kroenke, C. D., J. P. Loria, …, A. G. Palmer, III. 1998. Longitudinal and transverse ^{1}H-^{15}N dipolar/^{15}N chemical shift anisotropy relaxation interference: unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *J. Am. Chem. Soc.* 120:7905–7915.

56. LeMaster, D. M. 1995. Larmor frequency selective model free analysis of protein NMR relaxation. *J. Biomol. NMR.* 6:366–374.

57. Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Int. Symp. Information Theory.* B. N. Petrov, and F. Csáki, editors. Akadémiai Kiadó, Budapest, Hungary. 267–281.

58. Sugiura, N. 1978. Further analysis of data by Akaike's information criterion and finite corrections. *Commun. Stat. Theory Methods.* 7:13–26.

59. Hurvich, C. M., and C. L. Tsai. 1989. Regression and time-series model selection in small samples. *Biometrika.* 76:297–307.

60. Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach. Springer, New York.

61. Wong, V., D. A. Case, and A. Szabo. 2009. Influence of the coupling of interdomain and overall motions on NMR relaxation. *Proc. Natl. Acad. Sci. USA.* 106:11016–11021.

62. Prompers, J. J., and R. Brüschweiler. 2002. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* 124:4522–4534.

63. Calandrini, V., D. Abergel, and G. R. Kneller. 2010. Fractional protein dynamics seen by nuclear magnetic resonance spectroscopy: relating molecular dynamics simulation and experiment. *J. Chem. Phys.* 133:145101.

64. Ryabov, Y. E., and D. Fushman. 2007. A model of interdomain mobility in a multidomain protein. *J. Am. Chem. Soc.* 129:3315–3327.

65. Parigi, G., N. Rezaei-Ghaleh, …, C. Luchinat. 2014. Long-range correlated dynamics in intrinsically disordered proteins. *J. Am. Chem. Soc.* 136:16201–16209.

66. Hall, J. B., and D. Fushman. 2006. Variability of the ^{15}N chemical shielding tensors in the B3 domain of protein G from ^{15}N relaxation measurements at several fields. Implications for backbone order parameters. *J. Am. Chem. Soc.* 128:7855–7870.

67. Charlier, C., S. N. Khan, …, F. Ferrage. 2013. Nanosecond time scale motions in proteins revealed by high-resolution NMR relaxometry. *J. Am. Chem. Soc.* 135:18665–18672.

68. Ozenne, V., F. Bauer, …, M. Blackledge. 2012. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics.* 28:1463–1470.