Published in final edited form as:

Methods Mol Biol. 2020 ; 2141: 663–681. doi:10.1007/978-1-0716-0524-0_34.

Determining binding kinetics of intrinsically disordered proteins by NMR spectroscopy

Ke Yang¹, Munehito Arai², Peter E. Wright^{1,*}

¹Department of Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA92037

²Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Meguro, Tokyo 153-8902, Japan

Abstract

The unique structural flexibility of intrinsically disordered proteins (IDPs) is central to their diverse functions in cellular processes. Protein-protein interactions involving IDPs are frequently transient and dynamic in nature. Nuclear magnetic resonance (NMR) spectroscopy is an especially powerful tool for characterizing the structural propensities, dynamics and interactions of IDPs. Here we describe applications of the Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiment in combination with NMR titrations to characterize the kinetics and mechanisms of interactions between the intrinsically disordered proteins and their targets. We illustrate the method with reference to interactions between the activation domain of the human T-cell leukemia virus type-I (HTLV-1) basic leucine zipper protein (HBZ) and its cellular binding partner, the KIX domain of the transcriptional coactivator CBP.

Keywords

IDP; CPMG; relaxation dispersion; protein interaction; protein dynamics

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for dissecting protein molecular dynamics over a wide range of timescales at atomic resolution, allowing both global and local analysis of molecular interactions. NMR is uniquely suited for characterization of the conformational ensembles and dynamics of intrinsically disordered proteins (IDPs), providing novel insights into their structural propensities and the kinetics and mechanisms of their interactions with binding partners. IDPs play a central functional role in the cell, participating in numerous protein-protein interaction networks that control a diverse array of cellular signaling and regulatory pathways [1]. The functional versatility of IDPs arises from their conformational flexibility not only in their free forms but also during interactions with their cellular targets, the timescale of which commonly falls into the microsecond-millisecond range [2,3]. Our understanding of protein dynamics in the

*Corresponding author.

interactions of IDPs has benefited greatly from a variety of NMR experiments, including relaxation-compensated R_2 relaxation dispersion[4], rotating-frame relaxation ($R_{1\rho}$)[5] and saturation transfer experiments [6,7]. These experiments are particularly informative in revealing sparsely populated excited states that are invisible to conventional detection methods [8]. In particular, the Carr-Purcell-Meiboom-Gill (CPMG) R_2 relaxation dispersion experiment enables quantification of exchange processes that occur on a microsecond-millisecond timescale, providing novel information on the structure of weakly populated “excited” states [9], protein folding/unfolding events [10,11], the mechanisms of enzyme catalysis [12,13] and the kinetics and mechanisms of protein-protein interactions [14,3].

R_2 relaxation dispersion experiments provide important insights into the kinetics and mechanism by which an IDP binds to its target. At any given IDP:target stoichiometry, analysis of the dispersion profiles yields the populations of the different states involved in the interaction, the exchange rate (k_{ex}), and the chemical shift differences (ω) between states. By performing relaxation dispersion experiments in the settings of a titration, a detailed description of the binding process can be obtained. The kinetics and mechanism of interaction can be derived by fitting a series of relaxation dispersion experiments collected at different protein:ligand ratios to various binding models. In this chapter we will focus on the experimental implementation of backbone ^{15}N CPMG relaxation dispersion titration experiments and describe our approach to fitting CPMG titration data using the software package GLOVE [15]. To illustrate this approach, we describe how CPMG data are used to analyze the interactions between the KIX domain of the transcriptional coactivator CBP and the intrinsically disordered activation domain of the viral transcription factor HBZ (Human T-cell leukemia virus type-I (HTLV-1) basic leucine zipper protein) to determine binding kinetics and mechanism with residue-specific information.

1.1 Theoretical descriptions of chemical exchange

The dynamic process of chemical exchange exposes a nucleus to different chemical environments, resulting in distinct resonance frequencies for the different states involved in the exchange process. Consider the simplest model of chemical exchange, two-state exchange:

where k_{AB} and k_{BA} are the forward and reverse rates of reaction between state A and B. The overall exchange rate between A and B, k_{ex} , is denoted $k_{\text{ex}} = k_{\text{AB}} + k_{\text{BA}}$. The chemical shift difference between states is denoted $\Delta\omega = |\omega_{\text{A}} - \omega_{\text{B}}|$. The populations of states A and B

in the system, p_{A} and p_{B} , can be calculated using $p_{\text{A}} = \frac{k_{\text{BA}}}{k_{\text{AB}} + k_{\text{BA}}}$ and

$$p_{\text{B}} = \frac{k_{\text{AB}}}{k_{\text{AB}} + k_{\text{BA}}}.$$

The timescale of exchange in NMR spectroscopy is usually defined by comparing the exchange rate k_{ex} and the chemical shift difference ω between states. Exchange processes

on different timescales have different effects on the lineshape of peaks in the NMR spectrum. In the fast exchange regime, where $k_{\text{ex}} \gg |\Delta \omega|$, the NMR resonance of a two-state exchanging system appears at an intermediate chemical shift that is weighted by the population of states A and B. When $k_{\text{ex}} \ll |\Delta \omega|$ the exchange falls into the slow regime, where NMR resonances of both states A and B are directly observable at their distinct chemical shifts and the peak intensity of A and B reflects their populations, p_A and p_B , assuming that the relaxation rates in both states are the same ($R_{2A}^0 = R_{2B}^0$). In the intermediate exchange regime, where $k_{\text{ex}} \approx |\Delta \omega|$, the NMR resonances are broadened due to an increase in the transverse relaxation rate caused by exchange. The effect of exchange on the appearance of the NMR spectrum depends on the exchange kinetics, as illustrated in standard textbooks and many reviews [16].

1.2 Relaxation-compensated CPMG dispersion experiments for studying chemical exchange

As mentioned above, exchange at a rate comparable to the chemical shift differences between states causes line broadening in NMR spectra. When the exchange involves interconversion between a major and a minor populated state, exchange broadening and low population often make the minor state invisible to conventional NMR methods. The relaxation-compensated CPMG experiment developed by Loria et al. [4] allows measurements of exchange rates between 200 to 2000 s⁻¹, a frequency range where many protein conformational fluctuations and ligand binding processes occur. The CPMG experiment utilizes a series of spin-echo pulses ($\tau_{\text{CP}} - \pi - \tau_{\text{CP}}$, where τ_{CP} is a delay before and after the π pulse) to refocus magnetization. In the presence of exchange, the magnetization cannot be fully refocused, leading to broadening of the NMR signal, as illustrated elsewhere [16]. The effective transverse relaxation rate R_2^{eff} can be experimentally determined by comparing peak intensities of spectra collected with CPMG pulse train, $I(\tau_{\text{CP}})$, and a reference spectrum where τ_{CP} is set to zero, I_0 :

$$R_2^{\text{eff}} = -\frac{1}{T_{\text{CPMG}}} \ln\left(\frac{I(\tau_{\text{CP}})}{I_0}\right), \quad (2)$$

where T_{CPMG} is the length of the CPMG pulse train.

The effective line-broadening can be quantified as a function of the pulsing frequency of the refocusing pulses. Modulation of the refocusing pulsing frequencies in a CPMG pulse train allows deconvolution of the observed R_2^{eff} , into the intrinsic relaxation rate R_2^0 and relaxation due to chemical or conformational exchange, R_{ex} (Fig. 1).

1.3 The Carver and Richards equation

The original Carver and Richards equation [17] is one of the most widely used equations to quantitatively analyze the effective R_2 rate for all exchange regimes in the two-state exchange model:

$$\begin{aligned}
 R_2^{\text{eff}} &= \frac{1}{2} \left\{ R_{2A}^0 + R_{2B}^0 + k_{AB} + k_{BA} - \frac{1}{\tau_{CP}} \cosh^{-1} [D_+ \cosh(\eta_+) - D_- \cosh(\eta_-)] \right\} \\
 D_{\pm} &= \frac{1}{2} \left[\pm 1 + \frac{\Psi + 2\Delta\omega^2}{\sqrt{\Psi^2 + \xi^2}} \right] \\
 \eta_{\pm} &= \tau_{CP} \sqrt{\frac{1}{2} \pm \Psi + \sqrt{\Psi^2 + \xi^2}} \\
 \Psi &= (R_{2A}^0 - R_{2B}^0 + k_{AB} - k_{BA})^2 - \Delta\omega^2 + 4k_{AB}k_{BA} \\
 \xi &= 2\Delta\omega(R_{2A}^0 - R_{2B}^0 + k_{AB} - k_{BA})
 \end{aligned} \tag{3}$$

where R_{2A}^0 and R_{2B}^0 are the intrinsic transverse relaxation rates of state A and state B. It is usually assumed in data fitting procedure that $R_{2A}^0 = R_{2B}^0$.

1.4 GLOVE

GLOVE is a software package developed by Sugase et al. [15] for fitting of CMPG dispersion data using a variety of fitting models derived from the Carver and Richards equation among others. A version of GLOVE that incorporates features used in this protocol can be obtained through NMRbox [18] or from the Wright lab web page (https://www.scripps.edu/wright/?page_id=17).

1.5 Kinetic models of protein-protein interactions

For a two-state binding model:

where P and L represent the protein and ligand, respectively, and P:L represents the protein:ligand complex. [P] denotes the concentration of unbound (free) protein, which can be calculated as:

$$[P] = \frac{1}{2} \left\{ -K_D - [L]_0 + [P]_0 + \sqrt{(K_D + [L]_0 - [P]_0)^2 + 4[P]_0 K_D} \right\} \tag{5}$$

$$K_D = \frac{k_{\text{off}}}{k_{\text{on}}} = \frac{[P][L]}{[P:L]} \tag{6}$$

where $[P]_0$ and $[L]_0$ represent the total concentration of protein and ligand, respectively.

For this model, the R_2^{eff} can be calculated with Carver and Richards equation where ξ and k_{ex} are denoted by linear combinations of k_{on} and k_{off} :

$$\begin{aligned}
 R_2^{\text{eff}} &= R_2^0 + \frac{1}{2} \left\{ k_{\text{ex}} - \frac{1}{\tau_{\text{CP}}} \cosh^{-1} [D_+ \cosh(\eta_+) - D_- \cosh(\eta_-)] \right\} \\
 D_{\pm} &= \frac{1}{2} \left[\pm 1 + \frac{\Psi + 2\Delta\omega^2}{\sqrt{\Psi^2 + \xi^2}} \right] \\
 \eta_{\pm} &= \tau_{\text{CP}} \sqrt{\frac{1}{2} \left(\pm \Psi + \sqrt{\Psi^2 + \xi^2} \right)} \\
 \Psi &= k_{\text{ex}}^2 - \Delta\omega^2 \\
 \xi &= 2\Delta\omega([P]k_{\text{on}} - k_{\text{off}}) \\
 k_{\text{ex}} &= [P]k_{\text{on}} + k_{\text{off}}
 \end{aligned} \tag{7}$$

It is thus possible to obtain information on binding kinetics from the fits using this modified Carver and Richards equation. As can be seen from the equation, for a system where $[P]_0$ and $[L]_0$ are given, the on-rate k_{on} and the concentration of free protein $[P]$ are in one term $[P] k_{\text{on}}$. It is possible, however, to deconvolute $[P]$ and k_{on} by performing a CPMG titration experiment with varying $[P]_0$.

More complicated interaction mechanisms involving three-state binding models have been discussed in detail in the literature [19,15,14,3] (*see* Note 1 for ways to construct user-defined binding models in the software package GLOVE).

2. Materials

2.1 Sample and buffer

1. NMR buffer: 50 mM Tris-HCl (pH 6.8), 50 mM NaCl, 10% (v/v) D₂O
2. Uniformly ¹⁵N-labeled ligand (HBZ, expressed and purified as previously described [20]). Isotope enrichment close to 100% is desired. Sample concentrations typically range from ~300 μM to ~1 mM
3. Unlabeled target protein (KIX domain, residues 586–672 of CBP, prepared as previously described [3]).

2.2 Spectrometer and software

1. Bruker 500- and 800-MHz NMR spectrometers (*see* Note 2)
2. Bruker TopSpin software
3. nmrPipe software [21]
4. GLOVE software package for fitting of relaxation dispersion data

2.3 Pulse sequence for the relaxation-compensated constant-time CPMG experiment

Relaxation-compensated constant-time CPMG pulse sequence with a constant time delay of 40 ms [22]. Data should be acquired as an interleaved pseudo 3D experiment. Non-uniform sampling (NUS) can be applied to data collection to reduce data collection time or increase signal to noise.

3. Methods

When using relaxation dispersion to determine the kinetics of binding of an IDP (ligand) to a target protein, it is preferable to isotope label the IDP to avoid potential complications in data analysis (see Note 3). Assignments of the IDP (ligand) resonances are required for data interpretation. Backbone amide assignments are typically made using standard triple resonance experiments with ^{15}N , ^{13}C -labeled protein.

3.1 Sample preparation

Lyophilized proteins are resuspended in NMR buffer, buffer-exchanged, and concentrated to prepare protein stocks. To ensure that NMR samples for data acquisition at two fields are identical, it is advisable to prepare 1 mL of NMR sample and split it into two 500 μL samples for simultaneous data acquisition on the two spectrometers. (see Note 4).

At least four samples with different protein:ligand concentration ratios are required for analysis of binding kinetics using relaxation dispersion experiments. To ensure that the ligand concentration remains constant and only the protein concentration varies, samples of the beginning point and the end point of the titration, i.e., the lowest and highest concentration ratios, are prepared. After the acquisition of CPMG dispersion data, the samples are mixed in different ratios to make intermediate points of the titration. The pH value of each sample should be checked to ensure a constant pH across all samples. To obtain the representative dispersion profiles shown in the figures, samples were prepared with KIX: ^{15}N HBZ (protein:ligand) concentration ratios of 0.010:1, 0.015:1, 0.020:1 and 0.025:1.

3.2 Identification of the appropriate stoichiometric range for the CPMG titration experiment

The general idea is to identify a range of protein:ligand concentration ratios where high-quality dispersion data can be obtained. For instance, severe line broadening in spectra could indicate exchange on the microsecond to millisecond timescale but might impair data quality due to low peak intensity. It is possible to simulate the correlation between R_2^{eff} and the concentration ratio of protein and ligand with reasonable estimates of the on- and off-rates, k_{on} and k_{off} , and the dissociation constant of the protein-ligand interaction, K_D [23,24]. For a system for which prior knowledge of binding kinetics and dynamics is lacking, it is advisable to perform test experiments that include HSQC titrations on the protein-ligand interaction of interest, as well as CPMG experiments on the protein in both the free and the fully-bound states. Analysis of the lineshape and peak intensity in the HSQC titration and the CPMG dispersion profiles at different protein-ligand ratios should provide a useful guide in the experimental design and optimization of the concentration ratios at which CPMG titrations data are collected. It is desirable, however, to start the titration from an experimental condition where there is no dispersion in the free ligand to avoid complication in data analysis. When there is dispersion in the free ligand, even if one observes increased dispersion in the bound state that is dependent on the protein:ligand concentration ratio, data analysis may be complicated. The dispersion data, in this case, reflect a complex multi-state process: the intrinsic exchange process in the ligand of interest plus another process with

different kinetics from the exchange caused by protein interactions. Deconvolution of this multi-state process is very complicated. One cannot simply subtract the dispersion profile of the free ligand because the binding event may alter the intrinsic exchange process in the target ligand. Here, we discuss three possible scenarios and corresponding experimental strategies (assuming that the IDP (ligand) is the labeled species here; *see* Notes 3 and 5):

1. The HSQC spectrum of the IDP is well-resolved in both the free and bound form.

No dispersion is observed for the free IDP (ligand), but visible dispersion is observed when it is fully bound to the target protein. This indicates that motions in the free IDP are likely too fast to be detected by CPMG, while the bound IDP shows exchange in the μs – ms timescale, suggesting that the interaction might have both relatively slow on- and off-rates. In this case, consider performing a titration at concentration ratios around the stoichiometry of the fully bound complex [3,14] (for example, for a 1:1 protein:ligand complex where the IDP ligand is the labeled species, collect CPMG titration data at [protein]:[ligand] = 0.90:1, 0.95:1, 1.00:1, 1.05:1 etc).

2. The HSQC spectrum of the IDP is well-resolved in both the free and bound form.

No dispersion is observed for the free IDP nor when it is fully bound, indicating that the interaction might have a very fast on-rate and slow off-rate. Sub-stoichiometric protein concentrations can be considered in this case [24]. A full HSQC titration can be performed and peak intensities extracted from the HSQC experiments to identify protein:ligand ratios where an increased R_2^{eff} is observed.

3. The HSQC spectrum of the IDP is well-resolved in the free form, but severe line-broadening is observed in the bound form. In this case, CPMG titrations at very low, sub-stoichiometric protein concentrations are recommended (e.g. [protein]:[ligand] = 0.05:1, 0.10:1, 0.15:1, etc) (*see* Note 6).

3.3 CPMG titration data acquisition

3.3.1. Set up a vdlst—A vdlst is an input delay list used by a CPMG pulse program containing a list of different numbers of refocusing pulses to apply in the 40 ms T_{CPMG} delay. An example of the vdlst used in our experiments is given below:

0.0	20.0	4.0	80.0	8.0	40.0	16.0	32.0	12.0	48.0	0.0	80.0
4.0	24.0	8.0	64.0								

In a constant-time experiment with $T_{\text{CPMG}} = 40$ ms, the list above is equivalent to a series of $1/\tau_{\text{CP}}$ values of (s^{-1}):

0	500	100	2000	200	1000	400	800	300	1200	0	2000
100	600	200	1600								

3.3.2. Calibrate the ^1H hard and soft $\pi/2$ pulses with a standard proton 1D experiment. Use a 2π pulse to find a null point.

3.3.3. Set carrier frequencies and spectral widths. These are typically the same as used in HSQC experiments of the system of interest.

3.3.4. Set receiver gain with rga, and acquire CPMG data. Repeat the same procedure for all titration points.—Typical values of parameters used in experiments on a Bruker 800-MHz spectrometer are given in Table 1.

3.4 Dispersion data fitting using the GLOVE software

The overview of workflow is shown in Fig. 3.

3.4.1. Initial data processing—The goal of initial data processing is to convert the pseudo 3D CPMG experiment to a set of 2D spectra from which the peak intensities can be determined. Depending on the acquisition mode of the experiment, it might be necessary to split the pseudo 3D data file into multiple 2D data files. In our experiments, each 2D plane is processed using the NMRPipe [21] software. It is advisable to go through every 2D spectrum at this stage to ensure that the spectra are well-phased (*see* Note 7), that the positions of peaks do not shift between spectra, and that the spectra are free of artifacts that might interfere with determination of peak intensity.

3.4.2. Extract peak intensities using pkfit or fudaFIT—pkfit [15] and FuDA [25] are programs that integrate peak intensities from 2D spectra. Pkfit is included in the GLOVE software package, and FuDA can be downloaded from the website of the Hansen lab (<https://www.ucl.ac.uk/12ansen-lab/fuda/>).

Pkfit collects peak intensity based on peak position specified in the peak list file. An input file for pkfit contains the strength of magnetic field in MHz at which the experiments are acquired, box size (1×1 , 3×3 or 5×5 grid points centered on the peak top), the name of a peak list file, position tolerance in Hz in two dimensions (typically 12 Hz), and T_{CPMG} (s) and $1/\tau_{\text{CP}}$ (s^{-1}) for each 2D plane in nmrView format. An output file of pkfit contains magnetic field strength, peak position and peak intensity information, which can be subsequently reformatted by the program *chop* in GLOVE to generate an intensity file for the program *cpmg2glove*. In GLOVE, it is more convenient to run a shell script named *pkfity* to sequentially run pkfit and chop with the command below:

```
pkfity -I input_filename -o output_filename -v
```

Alternatively, the program FuDA can be used to determine peak intensities from the 2D planes. FuDA fits peaks in a spectrum to pure Gaussian or Lorentzian lineshapes (or a mixture of the two) and extracts peak intensities from the fitted peaks. Additional information and documentation on the usage of FuDA can be found on the Hansen lab website (<https://www.ucl.ac.uk/ansen-lab/fuda/>).

3.4.3. Convert peak intensity to R_2^{eff} using cpmg2glove—cpmg2glove converts peak intensity information from the intensity files generated by pkfit or FuDA to R_2^{eff} that can be used by GLOVE to calculate exchange parameters. A binding model and parameters of global fitting need to be declared in the input file for cpmg2glove, which will be discussed in further detail in the sections to follow. The output file from cpmg2glove is named glove.in by default, containing the R_2^{eff} values for each $1/\tau_{\text{CP}}$ (s^{-1}) used by GLOVE for fitting of CPMG data as well as the fitting model, fitting method, error estimation methods, data plotting instructions and global parameters that can be personalized by the user by modifying the glove.in file.

3.4.4. Fitting of titration points at each concentration ratio—After initial processing and generating intensity files from each CPMG experiment in the titration series, we recommend fitting each titration point to a basic two-state Carver and Richards equation (model named RICHARDS in GLOVE) to examine the exchange profiles of the labeled ligand and identify clusters of exchanging residues that share similar exchange behavior. In this step, we perform individual fits where a separate set of exchange parameters (k_{ex} , p_{B} and $\Delta \omega$) are calculated for each residue that undergoes exchange.

An example of input file for cpmg2glove using model RICHARDS for individual fitting of each residue is given below:

```
temp1="301"
model="RICHARDS"
global=
cpmg2glove -t $model $global \
-ise ./Int_field_1 $temp1 0 [percentile error] \
-ise ./Int_field_2 $temp1 0 [percentile error] \
```

where the flag `-ise` represents temperature (K, value denoted in temp1), spin (0 for ^{15}N) and percentile error (%), can vary based on the field strength and sensitivity of the spectrometer) set by the user (see Note 8). Int_field_1 and Int_field_2 are intensity files from the same sample collected at two spectrometer frequencies.

Fitting methods in GLOVE include GRID (grid search of global parameters in the parameter space that yields the best fit for both the global and local parameters), RANDOM (randomly choose initial values from parameter space for minimization), MCMIN (Monte Carlo-minimization using the current best fit as initial values), ONE (separated, alternative single-point minimization of local and global parameters from the lower limit of the range of

parameters) and ONEEX (simultaneous single-point minimization of both local and global parameters from the initial values of parameters) that can be executed in any order and combination. For the purpose of this step we find that a combination of RANDOM (50 iterations) + 3 × MCMIN (50 iterations each, with scale of 0.1, 0.01 and 0.001 sequentially) typically yields good fits within minutes (*see* Note 9 and 10). In addition to user-specified fitting methods, GLOVE applies an initial fitting method (named INIT) that utilizes the user-defined parameter values or the lower limit of parameter ranges to calculate a global minimum. After specifying fitting methods in the glove.in file, GLOVE is executed by command:

```
glove -vv -i glove.in -o glove.out -xmgr
```

that calculates the fits with the option of reporting the reduced χ^2 values in real time during the fitting process, storing the results after each round of calculation (-vv, an extended verbose mode), and writing out Grace files for plotting data (-xmgr). After successfully running an individual fit, dispersion profiles can be plotted for each residue using the software mplot as shown in Fig. 2. Detailed inspection of the individually fitted parameters (k_{ex} , p_B and $\Delta \omega$) helps identify residues that can be clustered for global fitting of data. In this example, we chose residues 21, 23, 25 and 31 as a cluster for global fitting (*see* Note 11).

3.4.5. Fitting of a titration series to the binding model of choice—Binding models incorporated in GLOVE include a two-state binding model and two distinct three-state binding models (*see* Note 1 and 12). Here we focus on the use of the two-state binding model, named CPMG_BIND2C or BIND2C depending on the version of GLOVE. CPMG_BIND2C uses the aforementioned modified Carver and Richards equation with

$$[P] = \frac{1}{2} \left\{ -K_D - [L]_0 + a[P]_0 + \sqrt{(K_D + [L]_0 - a[P]_0)^2 + 4a[P]_0 K_D} \right\} \quad (8)$$

where a is the concentration ratio of protein used for each individual dataset compared to the dataset used by the program for calculating the protein concentration [3,14]. For the representative dispersion curves shown in Fig. 3 we use [KIX]:[HBZ] = 0.025:1 (where KIX is considered to be the protein and HBZ the ligand) as the dataset where a is arbitrarily set to 1. a values for titration points at 0.010:1, 0.015:1, 0.020:1 are hence 0.4, 0.6 and 0.8, respectively.

An example of the input file for cpmg2glove using model BIND2C is given below:

```
templ="301"
model="BIND2C"
global="-gpar KON -gpar KOFF -gpar PB"
cpmg2glove -t $model $global \
-ice ./Int_field_1 $templ [conc. Ratio 1] [percentile error] \
-ice ./Int_field_2 $templ [conc. Ratio 1] [percentile error] \
-ice ./Int_field_1 $templ [conc. Ratio 2] [percentile error] \
```

```

-ice ./Int_field_2 $temp1 [conc. Ratio 2] [percentile error] \
-filt X1-hn
-filt X2-hn
-filt X3-hn
-filt X4-hn

```

where the flag `-ice` represents temperature (K), concentration ratio in regard to the dataset of which the a value is set as 1 in equation (8), and error (%). The flag `-filt` filters out residues except for the ones specified after it; in this case only residues X1, X2, X3 and X4 are included in the cluster. Here k_{on} , k_{off} and p_B are set as global parameters that are applied to all residues in the cluster. Initial values, ranges of global parameters and grid sizes for parameter search can be defined in the `glove.in` file, as well as whether each parameter is set as fixed or variable. Fits using model BIND2C were initially carried out on each residue to find a reasonable initial value for the global parameters. Global fits of all residues in the cluster are carried out multiple times using different initial parameter values and fitting strategies to obtain the best fit (with the smallest reduced χ^2). Fitting errors of global fits are estimated by Monte Carlo methods (see Note 13). Fitted curves are plotted using `mplot`. Representative dispersion curves for the KIX:HBZ interaction are shown in Fig. 4.

For quality control, fitted results are compared to experimental data derived from other biophysical methods. For example, the calculated K_D of the KIX:HBZ interaction is close to the K_D measured by isothermal titration calorimetry (ITC), and the calculated chemical shift difference $|\delta|$ (in ppm) values correlate with the equilibrium chemical shift differences $|\Delta\delta|$ between the free and the bound form of HBZ, indicating that the minor-populated state detected in the CPMG experiment represents the bound form of HBZ (see Note 14). Fits of the dispersion data to several binding models should be tested. In the example of Fig. 4, only BIND2C gave reasonable fits, indicating that the interaction between KIX and HBZ possibly follows a two-state mechanism or a mechanism where there is conformational selection with fast pre-equilibrium conformational exchange [14].

4. Notes

1. A general schematic representation of three-state binding models is shown in Fig. 5, where there can be either one (Fig. 5a) or two bound forms (Fig. 5b) among the three possible species and all three species interconvert.

GLOVE has two distinct three-state binding models BIND3C_1F2B (for two bound forms and one free form) and BIND3C_2F1B (for two free and one bound forms) that uses the largest eigenvalue of the matrix below to calculate R_2^{eff} :

$$\begin{pmatrix} R \left[\exp\left(-A \frac{\tau_{CP}}{2}\right) \exp\left(-A^* \frac{\tau_{CP}}{2}\right) \right] & I \left[\exp\left(-A \frac{\tau_{CP}}{2}\right) \exp\left(-A^* \frac{\tau_{CP}}{2}\right) \right] \\ I \left[\exp\left(-A \frac{\tau_{CP}}{2}\right) \exp\left(-A^* \frac{\tau_{CP}}{2}\right) \right] & -R \left[\exp\left(-A \frac{\tau_{CP}}{2}\right) \exp\left(-A^* \frac{\tau_{CP}}{2}\right) \right] \end{pmatrix} \quad (9)$$

In (9), functions $R[\ ]$ and $I[\ ]$ extract the real and imaginary parts of the complex matrix. $\mathbf{A}^*$ is the complex conjugate of the matrix $\mathbf{A}$. For BIND3C_2F1B,

$$\mathbf{A} = \begin{pmatrix} k_{12} + [\text{P}]k_{\text{on}1} & -k_{21} & -k_{\text{off}1} \\ -k_{12} & k_{21} + [\text{P}]k_{\text{on}2} - i\Delta\omega_{\text{F1F2}} & -k_{\text{off}2} \\ -[\text{P}]k_{\text{on}1} & -[\text{P}]k_{\text{on}2} & k_{\text{off}1} + k_{\text{off}2} - i\Delta\omega_{\text{F1B}} \end{pmatrix} \quad (10)$$

and for BIND3C_1F2B,

$$\mathbf{A} = \begin{pmatrix} [\text{P}](k_{\text{on}1} + k_{\text{on}2}) & -k_{\text{off}1} & -k_{\text{off}2} \\ -[\text{P}]k_{\text{on}1} & k_{\text{off}1} + k_{12} - i\Delta\omega_{\text{FB1}} & -k_{21} \\ -[\text{P}]k_{\text{on}2} & -k_{12} & k_{\text{off}2} + k_{21} - i\Delta\omega_{\text{FB2}} \end{pmatrix} \quad (11)$$

To construct customized three-state binding models where certain pathways in Fig. 5 can be eliminated, fix the exchange parameters in these pathways to 0. For instance, in BIND3C_2F1B (Fig. 5a), when $k_{\text{on}1}$ and $k_{\text{off}1}$ are set to zero, the model becomes a conformational selection model, where P interacts with a weakly populated, “excited” state conformation L^* [14]. Similarly, in BIND3C_1F2B, when $k_{\text{on}1}$ and $k_{\text{off}1}$ (or $k_{\text{on}2}$ and $k_{\text{off}2}$) are set to zero, the model becomes an induced-fit model, where conformational change occurs after binding.

When the three-state models are used, it is usually not easy to obtain good fits, especially when only a small data set is used [26]. A global fit of dispersion curves derived from many peaks may improve the accuracy of fitting. The following stepwise fitting procedure may also solve the problem [14]. First, local fits for each residue are performed to search for the best fit by a GRID search. Then, the median of the rate constants obtained for all residues is calculated and is used as an initial value for the subsequent global fit. In the global fit, a GRID search around the medians of local fits is performed. If the full global fit gives a reasonable fit, a subglobal fit, in which some residues are clustered, is performed using the results of the full global fit as initial parameter values. For quality control, the following conditions should also be checked: the population of the major form at equilibrium (usually the bound form or free form) is larger than that of other forms, and the bound form is more structured than other forms.

2. The use of a cryo-probe is recommended, especially when severe line-broadening is observed.

3. Isotopic labeling of the IDP (ligand) is recommended for 1) better data quality, 2) potentially less complication caused by inter-domain motions in bigger proteins, and 3) easier data processing and interpretation. If the primary purpose of the experiment is to obtain binding kinetics, labeling either the target protein or ligand should in principle yield comparable results. However, the presence of μs -ms time scale exchange processes that cause relaxation dispersion of target protein resonances in the free state will greatly complicate data analysis. In such cases, the dispersion profiles of the complex may reflect intrinsic target protein exchange kinetics plus kinetics of the binding event, which will be difficult to deconvolute. In contrast, it is

unlikely that a highly flexible IDP will exhibit relaxation dispersion in the free state, thus simplifying analysis of the dispersion profiles.

4. Accurate measurement of protein and ligand concentrations is important for obtaining accurate binding parameters. However, if protein or ligand concentration is not accurately measured (e.g., when dithiothreitol is present in UV absorption measurement), correction for protein or ligand concentration is possible during GLOVE fitting. In model BIND2C, BIND3C_1F2B and BIND3C_2F1B, parameters A_0 and B_0 represent the total concentrations of protein and ligand, respectively. Setting one of them as a variable enables correction for the concentration. If a corrected concentration obtained by a fit deviates more than two-fold from the original value, the fitting model may not be appropriate. Note that fitting may be unstable, if both A_0 and B_0 are set as variables [14].

5. R_2 relaxation dispersion is not observed when chemical exchange does not occur on a microsecond-millisecond timescale. However, changing experimental conditions, such as pH, temperature and salt concentration of the sample solution may affect binding/dissociation kinetics, leading to observation of R_2 relaxation dispersion. Chemical exchange on a faster microsecond timescale may be measured by $R_{1\rho}$ relaxation dispersion experiments [27,28].

6. The presence of an excess of the IDP ligand relative to the protein target might result in multiple ligand molecules binding to a single target molecule. HSQC titrations can resolve this possibility.

7. Proper phasing of the 2D planes is crucial for obtaining accurate peak intensities. It is tempting to use the same phasing parameters for all CPMG experiments collected on the same instrument; however, it is advisable to phase the HSQC spectra at each titration point independently to make sure the peak intensity is not biased due to the spectra being out of phase.

8. Precise estimation of experimental errors is necessary for obtaining reasonable reduced χ^2 values. If experimental errors are not manually given, cpmg2glove automatically estimates experimental errors, when multiple measurements are performed with the same $1/\tau_{CP}$ value (for example, the vdlst of 3.3.1 shows that duplicate measurements are performed with a $1/\tau_{CP}$ of 0, 100, 200 and 2000). Data points with a low signal-to-noise ratio (a default cut-off value is 2.5) and with negative intensity are removed when glove.in is created.

9. Model ONEEX is time consuming and therefore not recommended at early stages of parameter search and optimization.

10. Non-linear least-squares fitting requires good estimates of initial values. If the fitting using RANDOM and MCMIN does not give a good fit, we recommend using GRID to perform many fitting runs from various initial parameter values.

11. Empirical criteria for identifying residues that share collective behavior, i.e. have the same k_{ex} and p_B , have been discussed [12,29,3]. Generally, the criterion used to determine whether a given residue belongs within a cluster is that $\chi^2_{cluster}/\chi^2_{individual}$

2, where $\chi^2_{\text{individual}}$ is the χ^2 value for an independent fit of the dispersion profile for that residue and χ^2_{cluster} is the χ^2 value when fitted to a common value of k_{ex} and p_{B} applied to all residues in the cluster.

12. Choice of binding model: We recommend starting with the simplest binding model unless there is strong evidence suggesting that a more complex model is appropriate.

13. In GLOVE, standard deviations of fitting parameters are calculated using the covariance matrix method by default. The Monte Carlo method gives more reliable error estimates, but it requires a number of (e.g., 100 times more) calculations.

14. When there is more than one state from which the equilibrium chemical shift differences can be calculated, the state that gives the best correlation between $|\Delta\omega|$ and $|\Delta\delta|$ is most likely the minor populated state.

Acknowledgements

We thank Gerard J Kroon and David Oyen for expert assistance with the NMR experiments, and Rebecca Berlow for insightful discussions and proofreading. This work was supported by grants CA096865 and CA214054 from the National Institutes of Health and the Skaggs Institute for Chemical Biology.

Reference

1. Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16 (1):18–29. doi:10.1038/nrm3920 [PubMed: 25531225]
2. Schneider R, Maurin D, Communie G et al. (2015) Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. *J Am Chem Soc* 137 (3):1220–1229. doi:10.1021/ja511066q [PubMed: 25551399]
3. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447 (7147):1021–1025. doi:10.1038/nature05858 [PubMed: 17522630]
4. Loria JP, Rance M, Palmer AG (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J Am Chem Soc* 121 (10):2331–2332. doi:DOI 10.1021/ja983961a
5. Trott O, Palmer AG 3rd, (2002) R1rho relaxation outside of the fast-exchange limit. *J Magn Reson* 154 (1):157–160. doi:10.1006/jmre.2001.2466 [PubMed: 11820837]
6. Fawzi NL, Ying J, Ghirlando R et al. (2011) Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR. *Nature* 480 (7376):268–272. doi:10.1038/nature10577 [PubMed: 22037310]
7. Vallurupalli P, Bouvignies G, Kay LE (2012) Studying “invisible” excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* 134 (19):8148–8161. doi:10.1021/ja3001419 [PubMed: 22554188]
8. Bouvignies G, Vallurupalli P, Hansen DF et al. (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477 (7362):111–114. doi:10.1038/nature10349 [PubMed: 21857680]
9. Whittier SK, Hengge AC, Loria JP (2013) Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. *Science* 341 (6148):899–903. doi:10.1126/science.1241735 [PubMed: 23970698]
10. Neudecker P, Zarrine-Afsar A, Davidson AR et al. (2007) Phi-Value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc Natl Acad Sci U S A* 104 (40):15717–15722. doi:10.1073/pnas.0705097104 [PubMed: 17898173]

11. Meinhold DW, Wright PE (2011) Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion. *Proc Natl Acad Sci U S A* 108 (22):9078–9083. doi:10.1073/pnas.1105682108 [PubMed: 21562212]
12. Boehr DD, McElheny D, Dyson HJ et al. (2006) The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* 313 (5793):1638–1642. doi:10.1126/science.1130258 [PubMed: 16973882]
13. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, Benkovic SJ, Wright PE (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* 332 (6026):234–238. doi:10.1126/science.1198542 [PubMed: 21474759]
14. Arai M, Sugase K, Dyson HJ et al. (2015) Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc Natl Acad Sci U S A* 112 (31):9614–9619. doi:10.1073/pnas.1512799112 [PubMed: 26195786]
15. Sugase K, Konuma T, Lansing JC et al. (2013) Fast and accurate fitting of relaxation dispersion data using the flexible software package GLOVE. *J Biomol NMR* 56 (3):275–283. doi:10.1007/s10858-013-9747-5 [PubMed: 23754491]
16. Mittermaier AK, Kay LE (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* 34 (12):601–611. doi:10.1016/j.tibs.2009.07.004 [PubMed: 19846313]
17. Carver JP, Richards RE (1972) General 2-Site Solution for Chemical Exchange Produced Dependence of T2 Upon Carr-Purcell Pulse Separation. *J Magn Reson* 6 (1):89-&. doi:10.1016/0022-2364(72)90090-X
18. Maciejewski MW, Schuyler AD, Gryk MR et al. (2017) NMRbox: A Resource for Biomolecular NMR Computation. *Biophys J* 112 (8):1529–1534. doi:10.1016/j.bpj.2017.03.011 [PubMed: 28445744]
19. Korzhnev DM, Salvatella X, Vendruscolo M et al. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430 (6999):586–590. doi:10.1038/nature02655 [PubMed: 15282609]
20. Yang K, Stanfield RL, Martinez-Yamout MA et al. (2018) Structural basis for cooperative regulation of KIX-mediated transcription pathways by the HTLV-1 HBZ activation domain. *Proc Natl Acad Sci U S A* 115 (40):10040–10045. doi:10.1073/pnas.1810397115 [PubMed: 30232260]
21. Delaglio F, Grzesiek S, Vuister GW et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6 (3):277–293 [PubMed: 8520220]
22. Vallurupalli P, Hansen DF, Kay LE (2008) Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc Natl Acad Sci U S A* 105 (33):11766–11771. doi:10.1073/pnas.0804221105 [PubMed: 18701719]
23. Eisenmesser EZ, Bosco DA, Akke M et al. (2002) Enzyme dynamics during catalysis. *Science* 295 (5559):1520–1523. doi:10.1126/science.1066176 [PubMed: 11859194]
24. Sugase K, Lansing JC, Dyson HJ et al. (2007) Tailoring relaxation dispersion experiments for fast-associating protein complexes. *J Am Chem Soc* 129 (44):13406–13407. doi:10.1021/ja0762238 [PubMed: 17935336]
25. Hansen DF, Yang D, Feng H et al. (2007) An exchange-free measure of 15N transverse relaxation: an NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. *J Am Chem Soc* 129 (37):11468–11479. doi:10.1021/ja072717t [PubMed: 17722922]
26. Neudecker P, Korzhnev DM, Kay LE (2006) Assessment of the effects of increased relaxation dispersion data on the extraction of 3-site exchange parameters characterizing the unfolding of an SH3 domain. *J Biomol NMR* 34 (3):129–135. doi:10.1007/s10858-006-0001-2 [PubMed: 16604422]
27. Palmer AG 3rd, Massi F (2006) Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* 106 (5):1700–1719. doi:10.1021/cr0404287 [PubMed: 16683750]
28. Bhabha G, Ekiert DC, Jennewein M et al. (2013) Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat Struct Mol Biol* 20 (11):1243–1249. doi:10.1038/nsmb.2676 [PubMed: 24077226]
29. Farber PJ, Mittermaier A (2015) Relaxation dispersion NMR spectroscopy for the study of protein allostery. *Biophys Rev* 7 (2):191–200. doi:10.1007/s12551-015-0166-6 [PubMed: 28510170]

Fig. 1. CPMG relaxation dispersion experiment.

Simulation showing a peak that broadens at lower refocusing pulse frequencies (longer τ_{CP} delays) and the effective transverse relaxation rates (R_2^{eff}) as a function of pulsing frequency. R_2^0 represents the intrinsic transverse relaxation rate that can be measured at high pulsing frequencies. R_{ex} can be calculated by extracting R_2^0 from experimentally determined R_2^{eff} .

Fig. 2. Representative ^{15}N relaxation dispersion profiles of HBZ at $[\text{KIX}]:[\text{HBZ}] = 0.02:1$. Dispersion data for ^{15}N -labeled HBZ were collected at ^{15}N frequencies of 50.65 MHz (black dots, collected on a 500 MHz spectrometer) and 81.20 MHz (red dots, collected on an 800 MHz spectrometer). The lines represent the fits of the relaxation dispersion profiles for each individual residue.

Fig. 3.
Overview of workflow of dispersion data fitting using GLOVE.

Fig. 4. Globally fitted ^{15}N relaxation dispersion titration profiles of HBZ.

Superposition of dispersion curves for ^{15}N -labeled HBZ for all titration points, with [KIX]:[HBZ] ratios of 0.010:1, 0.015:1, 0.020:1, and 0.025:1. Dispersion data were acquired on 500-MHz (^{15}N frequencies of 50.65 MHz) and 800-MHz (^{15}N frequencies of 81.20 MHz) spectrometers.

Fig. 5. Schematic representation of three-state binding models.

(a) Three-state binding model where there are two free forms and one bound state. L and L* (colored in blue) represent two exchanging forms of the free ligand. The protein-ligand complex is shown as [P:L] (colored in pink). k_{12} and k_{21} are rate constants of the forward and reverse reaction between the two free forms of ligand. ω_{F1F2} represents the chemical shift difference between the two free forms. ω_{F1B} is the chemical shift difference between the bound and the most highly populated form of the free ligand (L). (b) Three-state binding model where there is one free form (L, blue) and two bound forms (P:L* and P:L, pink). k_{12} and k_{21} are rate constants for the exchange between the two bound forms. ω_{FB1} and ω_{FB2} are the chemical shift differences between the free form and the two bound forms.

0.0	20.0	4.0	80.0	8.0	40.0	16.0	32.0	12.0	48.0	0.0	80.0
4.0	24.0	8.0	64.0								

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

0	500	100	2000	200	1000	400	800	300	1200	0	2000
100	600	200	1600								

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Summary of typical experimental parameters in constant-time CPMG experiments on a Bruker 800-MHz spectrometer.

Parameters	Annotations	Typical Values
TD (F ₃)	Size of FID in the ¹ H dimension (F ₃)	2048
TD (F ₂)	Size of FID in the ¹⁵ N dimension (F ₂)	256 (154 if use 60% NUS)
TD (F ₁)	Number of planes	16
SW (F ₃)	Spectral width (F ₃)	16 ppm
SW (F ₂)	Spectral width (F ₂)	32 ppm
SW (F ₁)	Spectral width (F ₁)	10 ppm (dummy value)
NS	Number of scans	16
DS	Number of dummy scans	16
TD0	Loop count for td0	1
NBL	Number of blocks	16
O1P	Carrier frequency offset of ¹ H	4.7 ppm
O3P	Carrier frequency offset of ¹⁵ N	118.0 ppm

Note that the value of NBL needs to be equal to, or an integer multiple of the number of entries in the vdlst.