Published in final edited form as:

J Mol Biol. 2018 July 06; 430(14): 2164–2180. doi:10.1016/j.jmb.2018.05.007.

Enhanced sampling of interdomain motion using map-restrained Langevin Dynamics and NMR: application to Pin1

Jill J. Bouchard[†], Junchao Xia^{‡,*}, David A. Case[§], and Jeffrey W. Peng^{†,*}

[†]Department of Chemistry & Biochemistry, University of Notre Dame, Notre Dame, IN, 46556

^{||}Department of Physics, University of Notre Dame, Notre Dame, IN, 46556

[‡]Center for Biophysics & Computational Biology and Department of Chemistry, Temple University, Philadelphia, PA, 19122

[§]Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, NJ, 08854

Abstract

Many signaling proteins consist of globular domains connected by flexible linkers that allow for substantial domain motion. Because these domains often serve as complementary functional modules, the possibility of functionally important domain motions arises. To explore this possibility, we require knowledge of the ensemble of protein conformations sampled by interdomain motion.

Measurements of NMR residual dipolar couplings (RDCs) of backbone NH bonds offer a per-residue characterization of interdomain dynamics, as the couplings are sensitive to domain orientation. A challenge in reaching this potential is the need to interpret the RDCs as averages over dynamic ensembles of domain conformations.

Here, we address this challenge by introducing an efficient protocol for generating conformational ensembles appropriate for flexible, multi-domain proteins. The protocol uses map-restrained self-guided Langevin dynamics simulations to promote collective, interdomain motion while restraining the internal domain motion to near rigidity. Critically, the simulations retain an all-atom description for facile inclusion of site-specific NMR RDC restraints. The result is the rapid generation of conformational ensembles consistent with the RDC data.

We illustrate this protocol on human Pin1, a two-domain peptidyl-prolyl isomerase relevant for cancer and Alzheimer's disease. The results include the ensemble of domain orientations sampled by Pin1, as well as those of a dysfunctional variant, I28A-Pin1. The differences between the ensembles corroborate our previous spin relaxation results that showed weakened interdomain

*Corresponding Authors: Junchao Xia, Phone: (215) 204-7118, tuf29141@temple.edu. Jeffrey W. Peng, Phone: (574) 631-2983, jpeng@nd.edu.

Supporting Information Available: RDC values, PALES analysis, metrics demonstrating the stability of the simulations, and further details of the GA analysis.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

contact in the I28A variant relative to wild type. Our protocol extends our abilities to explore the functional significance of protein domain motions.

Graphical abstract

Keywords

protein dynamics; NMR; ensemble description; domain motion; self-guided Langevin dynamics; residual dipolar couplings; multi-conformational fitting

INTRODUCTION

Many signaling proteins have a modular architecture, consisting of globular domains that carry out distinct and inter-related functions^{1–3}. Often the linkers enable substantial rotational motion between adjacent domains. Exploring the functional role of these motions in the signaling mechanism calls for both experimental and computational approaches.

An effective experimental approach is the measurement of NMR residual dipolar couplings (RDCs) between pairs of spin $\frac{1}{2}$ nuclei⁴. Dipolar couplings depend on the average orientation between the dipolar interaction vector (the inter-nuclear vector), and the external static magnetic field, $\mathbf{B}_0$. In standard solution-state samples, the internuclear vectors sample all orientations with equal likelihood as a result of isotropic molecular tumbling; consequently, the dipolar couplings average to zero. However, the dipolar couplings can be re-introduced by weak alignment of the protein molecules using established alignment media^{5, 6}. In this work, we focus on RDCs of backbone amide ^1H - ^{15}N bonds, which report on the average orientation of the ^1H - ^{15}N bond vectors with respect to $\mathbf{B}_0$. Backbone HN RDC measurements are particularly popular due to the availability of robust 2-D heteronuclear correlation pulse schemes for extracting HN RDC values, and the ubiquity of uniformly ^{15}N -enriched ($\text{U-}^{15}\text{N}$) proteins samples in protein NMR studies^{7, 8}. Of note, ($\text{U-}^{15}\text{N}$) protein samples offer several advantages that explain their ubiquity: (i) 2-D ^1H - ^{15}N correlation spectra can account for all residues except Proline, as their 2-D cross-peaks correspond to amide ^1H - ^{15}N bonds; (ii) 2-D ^1H - ^{15}N spectra of backbone amides generally show favorable chemical shift dispersion; and (iii) $\text{U-}^{15}\text{N}$ enrichment is relatively inexpensive.

To extract conformational information from the RDCs, one re-expresses the orientation angle between the RDC internuclear vectors and $\mathbf{B}_0$ in terms of two sets of angles. The first set describes the average orientation of the HN bond vector relative to a molecule-fixed set of axes that corresponds to the protein alignment tensor. Then, the second set of angles

describes the orientation of the tensor axes relative to $\mathbf{B}_0$. The alignment tensor describes the orientational preferences of the protein, as dictated by the chosen alignment medium and the protein's spatial distribution of mass and charge. Being a 2nd rank spherical tensor, the alignment tensor has five independent components. Provided one measures at least five independent RDCs from the same rigid domain or whole molecule (e.g. RDCs from five non-collinear HN bonds), the tensor components can be determined using singular value decomposition (SVD) ⁹. The set of backbone HN RDCs usually suffices for this purpose. Importantly, the alignment tensor, once determined, acts as a common reference frame for comparing the orientations of widely separated NH bonds, such as those in distinct globular domains. This feature endows RDCs with long-range conformational information that is lacking in more traditional NMR structural parameters, such as NOEs or through-bond scalar couplings ⁴.

The default analysis of ^1H ^{15}N RDC data assumes a single alignment tensor corresponding to a single, dominant backbone structure ⁹. This assumption is reasonable for single domain proteins. But it is questionable for modular proteins, which often consist of globular domains are connected by flexible linkers. Relative motion between the domains can give rise to multiplicity of domain orientations with an attendant multiplicity of alignment tensors. If the domain motions occur on time scales short compared to the RDC measurements (i.e. < 1 ms) ¹⁰, then the measured RDC values are not reflective of a single interdomain configuration, but instead, represent averages over the ensemble of domain orientations sampled by the domain motions. In this case, the original SVD method is not appropriate for deriving the alignment tensor in that it attempts to match the experimental RDCs to a single conformation. In contrast, model-based simulations can estimate the alignment tensor using individual conformations, such as in the PALES approach ("Prediction of ALignmEnt from Structure") ^{11, 12}, where the tensor is estimated via a Monte Carlo search of molecules in the vicinity of an infinite two dimensional plate.

In principle, the experimental RDCs can provide insight into the domain motions, if they can be quantitatively related to distributions of domain orientations contained within a suitable ensemble of protein conformations. Interpreting RDCs and other NMR parameters in terms of conformational ensembles have relied on computational approaches, such as explicit solvent molecular dynamics (MD) simulations, to generate structural ensembles ^{13–16}. A persistent challenge facing these simulations is achieving sufficiently broad conformational sampling on the time scales of protein dynamics characterized by experiment. Collective motions, such as domain re-orientation studied here, occur less frequently and thus require longer simulation trajectories. Millisecond simulation trajectories are now possible using MD-dedicated supercomputers ¹⁷. Nevertheless, such trajectories are not yet routine, and may be impractical for large-scale studies (e.g. multiple protein constructs and series of homologs). Enhancing the efficiency of conformational sampling thus remains a practical and important priority.

Two techniques for enhancing sampling efficiency are coarse-grained simulations, and restraint potentials incorporating experimental results. Coarse-graining reduces the degrees of freedom during the simulation by judicious removal of some level of atomic detail. Examples include methods that treat the protein as a set of hydrodynamic spheres in order to

focus on domain reorientation^{18–20}. Experimental restraint potentials bias the simulation to regions of configuration space pertinent to the experimental observations (e.g. nuclear Overhauser enhancements, scalar couplings, RDCs, chemical shifts)²¹. Yet, there is potential conflict between these two routes: namely, decreased atomic resolution from coarse-graining may exclude the atomic detail necessary for including site-specific NMR data.

The above considerations motivate the present work: an efficient protocol for generating conformational ensembles to interpret NMR observables reporting on whole domain motion. Our protocol is based on map-restrained self-guided Langevin dynamics (Map-SGLD)^{22, 23}. Map restraints enable trajectories emphasizing relative domain motion, while simultaneously allowing for facile inclusion of NMR restraints. The combination of Map-SGLD simulations with NMR restraints (MapSGLD-NMR, hereafter) generates a broad ensemble of domain orientations suitable for interpreting dynamic RDCs. To inter-relate the RDCs and the resulting ensemble, we use a genetic algorithm (GA) based multi-conformation fitting procedure²⁴. The GA procedure fits the observed RDCs to population-weighted averages calculated over the distribution of domain orientations contained in the Map-SGLD-NMR ensemble. The end result is a tractable set (~10) of population-weighted conformations that collectively capture the inter-domain flexibility sensed by the dynamically averaged RDCs from experiment. This GA procedure is similar to previous multi-structure approaches for RNA²⁵ and protein²⁶ ensembles.

We demonstrate the MapSGLD-NMR protocol by applying it to ¹H–¹⁵N RDC measurements of Pin1, a peptidyl-prolyl isomerase recruited by numerous signaling pathways regulating mitosis and neurodegenerative disease^{27–30}. Pin1 consists of two flexibly linked domains: an N-terminal WW domain (residues 1-39) binds protein substrates, and a larger C-terminal peptidyl-prolyl-isomerase (PPIase) domain (residues 52-163). Both domains target phospho-Ser/Thr-Pro (pS/T-P) motifs. The WW domain serves as a binding module, while the larger PPIase accelerates the cis-trans isomerization of imide pS/T-P linkage. Previous x-ray and NMR studies indicated significant interdomain motion^{31–38}; yet, an atomic-level description of the ensemble of domain configurations remains elusive. Our MapSGLD-NMR analysis of Pin1 gives a first glimpse of this ensemble.

The remainder of our report proceeds as follows. The results section highlights the central features of the Map-SGLD-NMR and GA/Multi-conformational fitting protocols, with finer technical details deferred to the Materials and Methods section. We then describe the MapSGLD-NMR ensemble for wild-type (WT) Pin1, focusing on simulation parameters governing the RDC restraints and overall sampling efficiency. We also explore the validity of the MapSGLD-NMR ensemble through studies of a Pin1 variant, I28A, which carries a single Isoleucine to Alanine substitution in the WW domain $\beta 2\beta 3$ -turn. The I28A-Pin1 ensemble indicates enhanced interdomain mobility compared to WT-Pin1 – a result consistent with our previous ¹⁵N/¹³C_{METHYL} spin relaxation and chemical shift analyses of I28A-Pin1^{36, 39}. We conclude by discussing the potential of MapSGLD-NMR for the general study of interdomain motion. Notably, while this work illustrates MapSGLD-NMR using RDC data, it could in principle be used for any NMR observables sensitive to

interdomain flexibility (e.g. chemical shifts, spin relaxation rates, paramagnetic relaxation enhancements).

RESULTS

Backbone HN RDCs for Pin1

We collected two independent ^1H ^{15}N RDC data sets for wild-type (WT) Pin1, corresponding to two alignment media: (i) a 4% (w/vol)/ $r=0.87\%$ octanol C8E5/octanol mixture⁴⁰, hereafter referred to as C8E5; (ii) and 10–20 mg/mL Pf1 phage media⁴¹. Weak alignment results from steric collisions in the C8E5 mixture, and electrostatic interactions in the case of Pf1 phage. ^1H ^{15}N RDCs were measured in triplicate (i.e. three spectra recorded and dipolar splittings measured) using standard IPAP 2D ^1H ^{15}N pulse schemes⁸; data acquisition parameters are in the **Materials and Methods** subsection (*vide infra*).

We pruned the raw RDCs to those most representative of domain re-orientational motion. This meant excluding HN bonds showing signs of large amplitude, sub-nanosecond local bond librations (e.g. low order parameter S_{HN} per Eq 2), excessive exchange broadening, or resonance overlap. The pruned datasets (Fig. S1) consisted of 83 RDCs from 4% C8E5 media (11 from the WW domain, 72 from the PPIase domain), and 93 RDCs from 10 mg/mL Pf1 phage (17 from the WW domain, 76 from the PPIase domain). We used the standard deviation of the triplicate RDC values to estimate the uncertainties in the RDCs. The average uncertainties were 1.52 Hz for 4% C8E5, and 1.04 Hz for 10 mg/mL Pf1 phage.

To assess the independence of the 4% C8E5 and Pf1 phage RDC data sets, we compared the corresponding alignment tensors, using PALES (“Prediction of ALignmEnt from Structure”) ^{11, 12} (Table S1, Supporting information). In particular, we evaluated the normalized dot product between vectors consisting of the five tensor components (A_{ZZ} , $A_{\text{XX}}-A_{\text{YY}}$, A_{XY} , A_{YZ} , A_{XZ}) from the respective media^{10, 42, 43}. Orthogonal vectors (zero dot product, 90 degrees between vectors) correspond to total independence. The dot products (Table S2 in the Supporting information) indicate sufficient independence (vector angles differing by 55 to 79 degrees) between the 4% C8E5 and Pf1 phage RDC datasets. These values are consistent with the fact that alignment tensors from C8E5 and Pf1 reflect different molecular properties (steric effect and charge distribution), although the conformations might be correlated.

Map-restrained Self-Guided Langevin Dynamics

To generate Pin1 conformations relevant to backbone ^1H - ^{15}N RDCs, we pursued an MD strategy that emphasized relative domain motion. Specifically, we kept the PPIase domain fixed in space by applying strong Cartesian harmonic restraints (force constants of 1.0 kcal/mol·Å²); indeed the PPIase RMSD values were < 0.5 Å over the simulations (Fig. S2, Supporting Information). We then moved the WW domain relative to the fixed PPIase domain using map-restrained²³ self-guided Langevin dynamics (Map-SGLD)²². The map restraints were of central importance, as they enabled motion of the WW domain as a nearly rigid body, while keeping full atomic detail. Further details concerning the setup of the MapSGLD simulations are in the **Materials and Methods**.

RDC restraints

This preservation of full atomic detail presented an opportunity to extend the original MapSGLD method by including residue-specific restraint potentials based on experimental NMR RDC data ²¹

$$E_{RDC} = \kappa_{RDC} \sum_j^{N_{RDC}} (RDC_{j,EXP} - RDC_{j,CALC})^2. \quad (1)$$

Eq 1 is a sum of quadratic restraint potentials over the ¹H–¹⁵N RDCs for both domains and in both the 4% C8E5 and phage alignment media. The overall magnitude of E_{RDC} , κ_{RDC} , was set to a low value ($\kappa_{RDC} = 0.1$ kcal/Hz²) to avoid inhibition of the conformational search. The $RDC_{j,CALC}$ values were calculated from the atomic coordinates of the j^{th} NH bond using the form ²¹

$$\begin{aligned} RDC_{j,CALC} = & \frac{\mu_0}{2\pi} \left(\frac{\omega_{DD}^{HH}}{2\pi} \right) S_{HN} \oslash \left[A_{ZZ} \left(\frac{3\cos^2\phi_{Z,j}}{2} - \frac{1}{2} \right) \right] \\ & + \frac{(A_{XX} - A_{YY})}{3} \left(\frac{3\cos^2\phi_{X,j}}{2} - \frac{3\cos^2\phi_{Y,j}}{2} \right) \\ & + \frac{4}{3} \left[A_{XY} \left(\frac{3\cos\phi_{X,j}\cos\phi_{Y,j}}{2} \right) + A_{XZ} \left(\frac{3\cos\phi_{X,j}\cos\phi_{Z,j}}{2} \right) + A_{YZ} \left(\frac{3\cos\phi_{Y,j}\cos\phi_{Z,j}}{2} \right) \right]. \end{aligned} \quad (2)$$

Eq 2 assumes that the internal motions of the HN bonds are cylindrically symmetric about a well-defined average position in the molecule-fixed frame ²¹. The pre-factor contains the ¹H–¹⁵N dipolar coupling constant $\omega_{DD}^{NH} = \gamma_N \gamma_H \hbar / \langle r_{NH}^3 \rangle$. The cosine terms ($\cos\phi_{X,j}$, $\cos\phi_{Y,j}$, and $\cos\phi_{Z,j}$) are the average X, Y, and Z positions of the unit vector lying along the j^{th} HN bond in an arbitrary molecule-fixed reference frame. The A_{XX} , A_{YY} , A_{ZZ} , A_{XY} , A_{YZ} , and A_{XZ} terms are Cartesian alignment tensor components in the same molecule-fixed frame. The Lipari-Szabo scaling factor S_{HN} , ($0 < S_{HN} < 1$) ⁴⁴ accounts for the amplitude of presumed symmetric motion of the HN bond about its average position (the aforementioned cosines). The pruning of raw RDCs serves to select for those HNs for whom the S_{HN} are close to unity (low amplitude local motion). Of note, the tensor components are treated as dynamical coordinates in these simulations; they are updated each 1-femtosecond time step along with the usual atomic coordinates ²¹. Hence, the alignment tensor is determined by minimization of Eq 1, and is independent of an *a priori* physical model for the alignment mechanism.

MapSGLD-NMR Simulation Protocol

The initial Pin1 coordinates for the MapSGLD-NMR simulations came from the PDB deposition 1NMV, the NMR solution structure of apo Pin1 ³³. We chose 1NMV, as it was the

only Pin1 PDB deposition with an intact linker, thereby preserving the translational information between the domains. To promote broader conformational searching, we ran three parallel simulations starting from structures 1, 2, and 4 in the 1NMV PDB deposition, referred to as 1NMV-1, 1NMV-2, and 1NMV-4 hereafter. This trio of initial structures spanned the range of interdomain orientations defined by all ten structures within 1NMV.

The three initial structures were prepared identically for the MapSGLD-NMR simulations. Preparation included a first energy minimization (500 steps) with NMR E_{RDC} restraint potentials (Eq 1) turned off. The resulting energy reductions were typically to -3000 to -5000 kcal/mol. Next, the structures underwent a second energy minimization (100 steps), with the E_{RDC} potentials turned on in order to define the initial alignment tensor components A_{XX} , A_{YY} , A_{ZZ} , A_{XY} , A_{XZ} , and A_{YZ} for each conformer and alignment medium. This involved varying the tensor components while holding the protein domain coordinates (direction cosines) fixed ²¹. The final preparatory step defined the initial map restraints for the WW domains of the starting structures, per the methods outlined by Wu et al ²².

The three minimized structures (1NMV-1, 1NMV-2, and 1NMV-4) kicked off three 30ns MapSGLD-NMR trajectories, each using the same set of parameters listed in Table 1. Each trajectory used a 1 femtosecond time step to evaluate all forces needed to advance atomic positions, velocities, and alignment tensor components A_{XX} , A_{YY} , A_{ZZ} , A_{XY} , A_{XZ} , and A_{YZ} for each alignment medium. Notably, the atomic “positions” included both the protein coordinates and the *map* coordinates x_i , y_j , and z_k . The latter evolved according to the overdamped Langevin motion of its proxy rigid body ²². The interval for local time-averaging needed to calculate the guiding force was $t_L = 0.5$ picoseconds (500 steps of $\delta t = 1$ femtoseconds).

Each starting structure spawned a 30 nanoseconds MapSGLD-NMR trajectory, with snapshots (structural coordinates) written every 1000 steps (1 picosecond). Each 30 nanoseconds trajectory took approximately 6 days on 48 processors, using the Gyges processors at Rutgers University. Pooling the three 30 nanoseconds trajectories generated the raw MapSGLD-NMR ensemble of 90,000 conformations (snapshots) used for the subsequent analysis.

To monitor the trajectory stabilities and verify the intended rigid body motion, we followed domain-specific RMSDs and standard energy values. The energies stabilized quickly, (within ~ 250 picoseconds), and remained stable throughout the rest of simulations (Fig. S3, Supporting Information).

We also followed the trajectories progress by following the correlation between the experimental RDCs versus the predicted RDCs from PALES ¹². The predicted RDCs came from the standard “-bestFit” mode of PALES, which uses the singular value decomposition (SVD) described by Losonsci et al ⁹, to best fit the experimental RDCs. We used the PALES calculated Q factor ^{4, 45}

$$Q = \sqrt{\sum_j^{ALL} (RDC_{j,CALC} - RDC_{j,EXP})^2} / \sqrt{\sum_j^{ALL} RDC_{j,EXP}^2} \quad (3)$$

to define the agreement between the calculated versus experimental RDC values (lower Q indicates better agreement). Figure 2 compares the Q factors and linear correlation plots obtained at the conclusion of energy minimization to those obtained after 10 nanoseconds of the production-run MapSGLD-NMR simulations. Both the Q factors and linear correlation plots show clear improved consistency between the experimental RDCs and the Pin1 conformations sampled by the trajectories, for both alignment media.

Single conformer analysis of RDCs

Of main interest was the ability of the MapSGLD-NMR ensemble to reproduce the experimental ^1H - ^{15}N RDCs, and thereby investigate the extent to which those RDCs demanded a dynamic interpretation. To set a baseline, we first did the standard single conformer analysis, testing the *individual* abilities of each of 90,000 Pin1 conformations in the MapSGLD-NMR ensemble to reproduce all of the experimental RDCs. For each conformation, we used PALES¹² in the “-stPales” mode, to calculate the alignment tensor, ^1H - ^{15}N RDCs, and Q factor, per Eq 3. Of note, in contrast to the “-bestFit” mode, the “-stPales” mode determines the alignment tensor based on the structural coordinates and a physical model of its interactions with the alignment media; the experimental RDCs are not used for the prediction.

To explore the effect of the RDC restraints on conformational ensemble, we varied their number and looked for changes in the Q factors. Switching off the RDC restraints one by one raised the minimum Q, indicating the influence of experimental RDCs on the MapSGLD-NMR trajectories. We also investigated the influence of media-specific RDCs by generating three additional 90 nanosecond MapSGLD-NMR ensembles (again 90,000 conformations each): (a) an ensemble restrained exclusively by 4% C8E5 RDCs; (b) an ensemble restrained by Pf1 phage RDCs exclusively; (c), and an ensemble lacking all RDC restraints. Table 2 summarizes the results, listing the minimum, maximum, and average values for the single conformer Q factors. Generally, these Q factors were large compared to the Q factors reported in the literature to be demonstrative of high-quality fits among single domain globular proteins (e.g. $Q \sim 0.2$ to 0.3 ^{4, 46}). The high single conformer Q factors suggested that the underlying premise – that a single rigid structure should be able to account for the NH RDCs – was unsatisfactory for Pin1. Comparisons of Q factors between media-exclusive ensembles are discussed further below.

Multi-conformer analysis of RDCs

Proteins sampling a distribution of conformations on time scales up to ~ ms yield dynamically-averaged RDCs¹⁰. To account for this averaging, we used a multi-conformational fitting procedure in the calculated RDCs⁴⁷. The procedure calculates the RDC for the j^{th} NH bond vector as a population-weighted average over “M” basis structures, per Eq 4,

$$\left\langle RDC_j \right\rangle = \sum_{k=1}^M P_k RDC_{j,k}. \quad (4)$$

$RDC_{j,k}$ denotes the RDC for the j^{th} bond vector, specific to the k^{th} basis structure with fractional population P_k . Notably, Eq 4 allows each k^{th} basis structure to have its own distinct pair of alignment tensors. To calculate $\langle RDC_j \rangle$, we applied PALES to each of the “M” basis structures to calculate alignment tensor components and RDCs, and then took the P_k -weighted sum. Critically, in these PALES calculations, the aforementioned “-bestFit” (SVD) approach must be avoided, as it assumes the experimental RDCs reflect one conformation – an assumption inconsistent with conformational averaging. Instead, we used the “-stPales” mode, which bypasses the experimental RDCs, and makes predictions based purely on the protein structure and shape and an *a priori* model of the alignment media mechanism¹². For a given set of $\langle RDC_j \rangle$ values, we calculated Q_M , the multi-conformational analog of the standard Q:

$$Q_M = \sqrt{\sum_j^{N-RDCs} (\langle RDC_j \rangle - RDC_{j,EXP})^2} \sqrt{\sum_j^{N-RDCs} RDC_{j,EXP}^2} \quad (5)$$

Eq 5 simply replaces $RDC_{j,CALC}$ in the standard Q formula (Eq 3) with $\langle RDC_j \rangle$. Q_M thus measured the ability of the multi-conformational average, $\langle RDC_j \rangle$, rather than the RDC of a single conformation, to reproduce the j^{th} experimental RDC ($RDC_{j,EXP}$).

To find the optimal weights P_k , the multi-conformational fitting used a genetic algorithm (GA) approach⁴⁷, selecting for population combinations yielding lower Q_M . The GA approach for RDCs is similar to the Sparse Ensemble Selection model of Berlin *et al.*²⁶ in its emphasis on finding a small number of configurations that fit the data, in the hopes of facilitating physical insights; comparisons can also be made to earlier ensemble methods for RNA RDCs^{48, 49}, and schemes for NOE-based structure determination of bound ligands⁵⁰.

We applied the GA multi-conformational fitting to the raw Pin1 MapSGLD-NMR ensemble; this entailed an initial clustering of the 90,000 MapSGLD-NMR conformations to a smaller set of 1000 unique structures. The 1000 structures were then subjected to 1,000,000 GA steps to identify an optimal set of population weights (Further details are in **Materials and Methods**). From the 1000 conformers, the GA fit identified a set of 6 conformers with populations $> 1.0 \times 10^{-3}$ (Fig. 3A). Critically, the GA fitting allows each unique conformer to have its *own* alignment tensors; Table S3 in the Supporting Information lists the 4% C8E5 tensor components for the final set of six conformers. The GA conformer with the lowest single conformer Q ($Q = 0.236$) contributed more than 50% of the total population (Fig 3A and, Fig. S4 Supporting Information). The GA multi-conformational fitting gave a Q_M reduced by 17% relative to that obtained from the best single-structure (e.g. $Q_M = 0.196$ versus $Q = 0.236$) (Fig. 3B). These results strengthen the notion that Pin1 RDCs reflect conformational averaging.

To check whether we found a unique (global) solution corresponding to the experimental ^1H - ^{15}N RDCs, we used a procedure demonstrated previously in the first paper presenting the GA method for RDCs⁴⁷. Specifically, we reran the GA fitting procedure on the 1000 Pin1 conformers 5 times, each time with different random population weights assigned to each of the 1000 conformers. Gratifyingly, in all cases, the GA fit reproduced the same six conformations and their relative population weights (Fig. S5). This demonstrates that the GA-determined set was stable, as it could be reproduced from different randomized initial populations (Fig. S5).

We also wanted to explore the sensitivity of GA-determined populations to the precision of the experimental RDCs. In principle, this could involve a Monte Carlo approach, generating hundreds of synthetic RDC data sets using Gaussian deviates based on the experimental variances. Such a procedure was used in the first paper presenting the GA multi-conformational fit for RDCs on oligo-saccharide systems (e.g. Supporting Information in Xia et al.,⁴⁷). However, the larger scale of the present Pin1 RDC study presented challenges. Notably, one GA run of 1,000,000 steps required several days on a workstation to determine the population weights of 1000 conformers. This would have to be repeated for hundreds of synthetic datasets, and this proved to be impractical.

As an alternative approach, we evaluated sensitivity of the Q_M factors and Chi-squared values to the number of conformations in the final GA ensemble. While this was clearly not as satisfactory as a bootstrapping analysis described above, it addressed the question of the minimum number of GA conformations needed to account for the experimental RDCs within their estimated errors. Specifically, we monitored the change in Q_M factors and Chi-squared, as we successively omitted the lowest-populated GA conformation and renormalized the population weights for the remaining conformations. The results (Fig. S6, Supporting Information) suggested that conformations with original populations < 10% (5th and 6th conformations in the original six) could be dropped without significant changes in Q_M or Chi-squared.

Effects of alignment media

A persistent discrepancy emerged between the Q factors calculated from the C8E5 media RDCs versus the Pf1 phage RDCs. Including the Pf1 phage RDCs raised the Q factors for both the conventional single-conformer fits (e.g. Table 2, rows 1 and 2) and GA multi-conformational fits. We wondered if this discrepancy was related to known difficulties in predicting RDCs from protein structures in the case of electrostatically induced alignment, such as Pf1 phage⁵¹. We investigated this possibility using the three additional 90 nanosecond MapSGLD-NMR ensembles (90,000 conformations each) listed in Table 2: (a) an ensemble restrained exclusively by 4% C8E5 RDCs; (b) an ensemble restrained by Pf1 phage RDCs exclusively; (c), and an ensemble lacking all RDC restraints. We then compared the abilities of the media-exclusive ensembles to reproduce the experimental RDCs, using both the conventional single-conformer approach, and GA multi-conformational approach.

The C8E5-only ensemble included conformers with lower PALES-predicted Q factors. As before, the GA multi-conformational search found a reduced set of 6 weighted conformers

reproducing experimental C8E5 RDCs. By contrast, the Pf1-only ensemble had conformers with typically higher PALES-predicted Q factors, and GA multi-conformational search failed to find a reduced ensemble to reproduce the experimental Pf1 RDC values. We also investigated the cross-fitting capabilities of the media-specific ensembles. We found that the Pf1-only ensemble gave a poorer fit of the 4% C8E5 RDCs compared to the 4% C8E5-only ensemble. Similarly, the 4% C8E5-only ensemble gave a poorer fit of the phage RDCs compared to the phage-only ensemble.

The above results are consistent with the notion that the discrepancies stemmed from greater difficulties in model-based predictions of electrostatic (Pf1-phage) as opposed to collision-induced (4% C8E5) alignment. For electrostatic media, alignment tensor predictions rely on knowledge of the charge distribution over the whole protein; that distribution varies with force fields and electrostatic models, and the optimal choice is not obvious.

It is worth noting here that the uneven success of model-based predictions alignment tensors media underscores an advantage of the MapSGLD-NMR simulations. For MapSGLD-NMR, the simulations optimize the alignment tensor components (and RDC values) directly at each time step via minimization of the RDC restraint potential (Eq 1). As such, the resulting MapSGLD-NMR tensors are independent from *a priori* model assumptions concerning the physical mechanism of the alignment media. This independence is appealing when considering the standard imperative that getting a reliable conformational ensemble requires RDCs from numerous different alignment media, some of which may be incompatible with protein function or stability, or perturb the distribution of conformations sampled in different ways.

While the GA multi-conformational search had difficulties in reproducing the Pf1 RDCs, it worked well with the C8E5 RDCs, producing a “basis set” of six population-weighted conformers that capture the essential features of the overall C8E5 ensemble. This is very encouraging as it suggests a reasonable conformational ensemble can be obtained without measurements across a large number (> 5) of alignment media. This reduced GA description should prove more tractable for future modeling studies involving Pin1 binding modes.

Description of the Pin1 Interdomain Ensemble

To describe the distribution of interdomain orientations, we used a fictitious interdomain vector connecting the geometric centers of the catalytic pocket PPIase domain catalytic site to the WW domain substrate-binding loop (residues 16-21) (Fig. 4A). We defined the vector coordinates (r, θ, ϕ) in the principal axes frame of the PPIase domain inertia tensor, which remained immobile throughout the MapSGLD-NMR simulations. We defined the polar angle θ such that it ends to 180° as the domains adopt a more extended configuration, with the WW domain along the Z principal axis of the PPIase inertia tensor. We then polled the (r, θ, ϕ) values, as well as the radius of gyration for the 90,000 conformers of the raw MapSGLD-NMR ensemble, resulting in the histograms of Figs 4C–F.

The histograms for “ r ” and the radius of gyration showed both compact and extended conformations, with the compact conformations being more prevalent. The compact conformations do not show direct interdomain contact, nor are they as compact as depicted

in the crystal structures 1PIN³¹ or 1F8A³². However, we note that the crystal structures were actually Pin1 *complexes*, such as ala-cis-pro dipeptide in the PPIase active site in 1PIN³¹ and a doubly phosphorylated CTD peptide bound to the WW domain³².

It is also worth emphasizing that the RDCs do not directly restrain the interdomain distance; they report mainly on HN bond orientation. This also influences the prevalence of extended versus compacted domains within the ensemble. Conceivably, adding distinct NMR restraints based on parameters more sensitive to interdomain proximity (e.g. PREs⁵²) would produce conformers with closer interdomain contact. We also mapped the interdomain vector coordinates of the six GA-selected ensemble for WT-Pin1 (Fig. 4B). Their values for (r , θ , ϕ) and the radius of gyration correspond to the green diamonds in Figs 4C–F.

Factors Influencing Conformational Searching

We investigated the influence of the NMR RDC restraints on the trajectories, by comparing the range of r and θ values for 20 nanoseconds MapSGLD simulations with and without NMR-RDC restraints (i.e. MapSGLD in Fig. 5A versus MapSGLD-NMR in Fig. 5B). Including the RDC restraints clearly reduced the number of conformations with smaller interdomain distances and angles.

We also compared the conformational space explored by MapSGLD versus conventional explicit-solvent simulations. We examined the breadth of sampling by the aforementioned 20ns MapSGLD simulation (~5 days CPU time, Fig. 5B) versus that of a 400 nanoseconds standard explicit-solvent simulation (~1 month CPU time, Fig. 5C), both in the absence of any RDC restraints. Per Figs 5B, C, the 20 nanoseconds of rigid-body MapSGLD simulations (implicit solvent) covered a substantially broader r and θ range (broader range of motion between the domains) than even 400 nanoseconds of standard explicit-solvent MD. The increased search space by MapSGLD is noteworthy, given its nearly 6-fold reduction in CPU time compared to the explicit solvent simulation.

Wild-Type Pin1 versus I28A Ensembles

We wanted to investigate the validity of the MapSGLD-NMR ensembles. Previous validation strategies have used a cross-validation approach – generating conformational ensembles from a subset of a large and diverse set of RDCs and then evaluating the ensemble's ability to predict RDCs excluded from ensemble generation (i.e. that were not used as restraints)^{53–55}.

Unfortunately, we lacked a sufficiently large set of RDCs for this type of cross-validation. We therefore pursued an indirect approach, investigating the consistency of the MapSGLD-NMR ensembles with the findings from our previous studies of Pin1 versus I28A interdomain motion³⁶. Critically, those findings were based on comparisons of NH chemical shift perturbations and ¹⁵N spin relaxation parameters excluded from the present ensemble generation. Moreover, those previous studies showed that I28A, a single site substitution in the WW domain β 2- β 3 hairpin (see Fig. 1), decreased the transient interdomain contact compared to that observed in wild-type (WT) Pin1. Here, we investigated whether an I28A-Pin1 MapSGLD-NMR ensemble based on I28A-Pin1 RDCs,

would have features consistent with the reduced interdomain contact. This query addresses cross-validation, in that it examines the consistency of the RDC-based ensemble with results from non-RDC parameters excluded from ensemble generation.

Toward this end, we subjected I28A-Pin1 to the same experiments and ensemble analysis carried out for WT-Pin1. Specifically, we measured I28A-Pin1 NH RDCs (83 RDCs 4% C8E5 media, 94 RDCs Pf1 phage) and verified dataset independence via a dot product analysis (Table S2, Supporting Information). We then generated a MapSGLD-NMR ensemble of 90,000 conformations for I28A-Pin1. The raw ensemble furnished a basis for the GA based multi-conformational fit, which identified seven population-weighted I28A-Pin1 conformers (cutoff population = 0.001), each with their own alignment tensor (Table S4, Supporting Information). We then compared the raw I28A- and WT-Pin1 ensembles (Fig. 4), with attention to the same interdomain vector depicted in Fig. 4A and the radius of gyration.

The histograms for length (r) and the radius of gyration were particularly revealing. They showed that the I28A-Pin1 ensemble preferred conformations that were more extended (longer “ r ” and a longer radius of gyration), than those in the WT-Pin1 ensemble. Greater extension in I28A-Pin1 was also apparent in the θ (polar angle) histogram, which skewed to values closer to 180° , indicating a higher preference for extended interdomain conformations than WT-Pin1. These differences between the I28A- and WT-Pin1 ensembles were consistent with the decreased levels of interdomain contact inferred from our previous ^{15}N spin-relaxation studies^{36, 39}. The shift to more extended conformations apparently decreased the uniformity of the ϕ . In particular, the I28A-Pin1 ϕ distribution was less broad than WT, and noticeably depopulated in the range ($50^\circ < \phi < 120^\circ$). The “dot clouds” of Fig. 6 give another view of the differences between the I28A-Pin1 and WT-Pin1 ensembles. The dots, gold for I28A-Pin1 and blue for WT-Pin1, are the tips of the interdomain vector that start in the PPIase domain active site and terminate at the WW domain substrate-binding loop, for various ensemble snapshots. The dots help visualize the distribution of WW orientations relative to the fixed PPIase domain for both I28A-Pin1 and WT-Pin1. Non-overlapping regions of dots highlight differences in orientational preferences for the two constructs.

Closer examination of the less populated conformations within both the WT-Pin1 and I28A-Pin1 ensembles showed I28A-Pin1 could become more compact than WT-Pin1. A plausible explanation for this is that in WT-Pin1, greater side chain steric repulsion (I instead of A) discourages these more compact conformations. We return to this issue in the **Discussion** subsection below.

Collectively, the above differences suggest (i) I28A-Pin1 samples a broader span of conformations than WT-Pin1; (ii) the more populated conformations of I28A-Pin1 are more extended than those of WT-Pin1. These results are consistent with the decreased level of interdomain contact and greater interdomain mobility inferred from our previous ^{15}N spin-relaxation for I28A vs. WT-Pin1³⁶.

DISCUSSION

Multi-domain signaling proteins such as Pin1 consist of flexibly linked globular domains that can sample a multiplicity of conformations involving different domain orientations. NMR RDCs measured for these domains reflect averages over their orientational fluctuations. Interpreting their RDCs thus requires a description in terms of conformational ensembles. The present work addressed this challenge by introducing a new protocol: MapSGLD-NMR simulations. MapSGLD-NMR promotes efficient and broad conformational sampling of interdomain motion via four main routes: (i) use of an implicit solvent model; (ii) self-guiding forces that enhance slower and more collective motion; (iii) map-restraints that enable quasi-rigid body motion of domains, while retaining a full-atom description of the protein; (iv) inclusion of site-specific NMR restraints. Our results show that the MapSGLD-NMR simulations give an ensemble of conformations consistent with NMR RDCs, with the more frequently visited structures corresponding to the more relevant structures. The MapSGLD-NMR ensemble provides raw material for fitting the experimental RDCs in terms of a reduced set of population-weighted conformers. In particular, the GA/multi-conformational search⁴⁷ can define a set of population-weighted conformations derived from the raw ensemble.

We demonstrated MapSGLD-NMR by applying it to the human peptidyl-prolyl isomerase Pin1, an essential mitotic regulator consisting of a catalytic PPIase domain and a WW domain (binding module). Their connection via a disordered linker, in principle, allows for a substantial range of domain motion. However, the nature of the conformations sampled by motion has remained obscure. The MapSGLD-NMR ensemble provides a first glimpse of range of Pin1 interdomain orientations at atomic resolution. In turn, this provides a basis for understanding the biological implications of Pin1 interdomain motion, in particular, how it may affect its role as an essential mitotic regulator^{27, 56–58}. We discuss this below.

Pin1 regulates the activity of numerous cell cycle phosphoproteins, including the oncoprotein c-Myc⁵⁹ and the p53 tumor suppressor⁶⁰. A recent tally indicates Pin1 substrates include over 40 oncogenes and 20 tumor suppressors^{30, 61, 62}. Some of these interactions occur in the nucleus (nuclear speckle) while others are in the cytoplasm. In all cases, Pin1 targets phospho-Ser/Thr-Pro (pS/T-P) motifs of its protein substrates; both the WW and PPIase domains recognize these motifs and can be simultaneously occupied by pS/T-P motifs³⁵. A natural and open question persists: what atomic properties of Pin1 enable to interact with so many different targets in different cell milieus?

The MapSGLD-NMR ensemble of Pin1 suggests that interdomain flexibility as one of those properties. The ensemble samples a broad range of domain orientations that generally disfavor highly collapsed conformations. Relative domain flexibility, and the preference for more open conformation make sense in light of the physical phenotypes of Pin1 target sites. In particular, Pin1 target sites (pS/T-P motifs) typically reside within intrinsically disordered regions (IDR) of the substrate proteins^{38, 63}. Extensive interdomain flexibility could help Pin1 adapt to these “moving targets”. Moreover, many Pin1 substrates have multiple pS/T-P sites and Pin1 interdomain motion could allow for multivalent recognition. As we have speculated previously, it is conceivable Pin1 could dock at one pS/T-P motif first via its WW

domain; interdomain flexibility would then enable the PPIase domain to search for proximal pS/T-P sites^{38, 39} in a manner resembling “fly-casting”⁶⁴.

The MapSGLD-NMR ensembles also give insight into the breadth of domain orientations supporting Pin1 function. This was illustrated in the comparison of MapSGLD-NMR ensembles of WT-Pin1 versus I28A-Pin1. As stated, the single I-to-A substitution in the WW domain β 2- β 3 hairpin (Fig. 1) that gives rise to weaker interdomain contact and a modest increase of PPIase isomerase activity^{36, 39}. Fig. 4 shows that the most populated conformations for I28A reside in a different region of τ , θ and ϕ space from WT. Per Fig 4c–h, the more populated I28A-Pin1 conformations are more extended than those of WT-Pin1. Yet, Fig. 4c–h also shows that the less populated I28A-Pin1 conformations could be more compact than those of the WT-Pin1. Closer inspection of these compact I28A-Pin1 conformations do not establish the interdomain contact of WT; in particular, the WW domain β 2/ β 3 turn is improperly oriented away from the PPIase domain interface. In effect, the I28A-Pin1 ensemble suggests that the linker flexibility is such that there are multiple ways for the two domains to become “compact” without properly forming interdomain contact. Moreover, the I28A substitution may decrease WT interdomain contact, not only by distance effects, but also by inhibiting proper interdomain orientation. This would be consistent with our results in Fig 4d–e showing that I28A samples smaller and larger values of θ , and ϕ than WT Pin1. The RDC restraints alone are insufficient to distinguish distance versus angular effects. For that, we would need to enhance our simulations with complementary NMR parameters more sensitive to distance effects, such as Paramagnetic Relaxation Enhancements (PRE).

It is also worth mentioning the functional implications of our difficulties in finding a conformational ensemble compatible with both the 4% C8E5 and Pf1 phage RDCs. As stated, these difficulties may reflect limitations of the PALES predictions for phage RDCs. However, another possibility is that Pin1 samples a slightly different distribution of domain orientations in the Pf1 phage versus 4% C8E5 media. Indeed, some NH resonances of the WW domain showed slight differences in chemical shifts and line-broadening between the two media. This accounted for the difference in the number of NHs pruned from the raw RDC data on account of overlap or poor signal to noise. These differences, combined with the rather different alignment tensor principal values (e.g. A_{zz}), suggest Pin1 may adopt slightly different conformational ensembles between the two media, and that our simulations may not have been long enough to capture the breadth of conformations in Pf1 phage.

With MapSGLD-NMR, it should be possible to extend our findings here. For example, point mutations such as I28A are straightforward, experimentally and *in silico*, and thus provide a simple perturbative means to gain more insights about interdomain mobility. MapSGLD-NMR can also account for the effects of ligand interactions implicitly, by including bound-state NMR data into the simulation. Also, relaxation of the E_{MAP} restraints would allow greater intradomain mobility; in principle, this would allow us to explore the possible interplay between intra- and interdomain dynamics.

Larger systems may also benefit from the MapSGLD-NMR method because E_{map} greatly reduces the computational expense for moving domains, and multiple E_{map} constraints can

be used in a single simulation.²² The efficiency gain may be even greater for larger protein complexes because a larger portion of atoms may be included in the rigid-bodies instead of treated atomically, thus requiring less computing power.

In conclusion, we have demonstrated the MapSGLD-NMR method, which combines simulation and experimental NMR parameters to extend current capabilities for investigating interdomain motions of modular proteins. MapSGLD-NMR provides a straightforward and efficient method for determining the conformational ensemble underlying NMR-RDC data. Moreover, it may be extended to other experimental restraints that are sensitive to domain motion (e.g. chemical shifts, spin relaxation rates, paramagnetic relaxation enhancements), which may be essential for investigating more complicated domain motion, or for finer distinction of dynamic behavior.

MATERIALS AND METHODS

Pin1 Sample Preparation

U-¹⁵N Pin1 was over-expressed and purified as previously described.⁶⁵ The final NMR buffer was 30 mM imidazole-D4, pH 6.6 with 30 mM NaCl, 0.03% NaN₃, 5 mM DTT-D10 and 90% H₂O/10% D₂O. Partial alignment for residual dipolar couplings (RDCs) was induced with a mixture of C8E5/octanol⁴⁰ or 10 mg/ml Pf1 phage⁶⁶. A 10% stock solution of C8E5 media was prepared by dissolving C8E5 in NMR buffer, pH-adjustment, sterile filtering, and adding octanol to a molar ratio of 0.87. Stock solutions of each media were added to stock solutions of Pin1 to a final concentration of 250 μM protein and 4% (w/vol) C8E5 media or 10–20 mg/mL Pf1 phage. In Pf1 phage samples, we increased the buffer ionic strength 100 mM imidazole-D4 and 100 mM NaCl to reduce interactions with phage.

NMR Measurements of HN RDCs

Backbone NH RDCs were measured in triplicate using the standard in-phase/antiphase (IPAP) 2D ¹H ¹⁵N HSQC pulse scheme⁶⁷. All NMR measurements were at a nominal temperature of 295 K on Bruker Avance spectrometers (16.4 T or 18.8 T) equipped with TCI cryogenic probes. The 2D datasets (¹⁵N dimension (ω_1/t_1): 64 complex points in t_1 , sweep width of 35.2351 ppm, carrier at 118 ppm, and 32 scans per t_1 ; ¹H dimension (ω_2/t_2): 1024 points in t_2 , 15.0031 ppm, carrier at 4.703 ppm) were apodized and Fourier-transformed using Topspin 2.1 (Bruker Biospin, Inc.). The raw ¹H-¹⁵N couplings were measured from the frequency separation of ¹⁵N doublet members along the ¹⁵N dimension, using Sparky 3 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). The NH RDCs were calculated as the Hertz difference between the raw couplings observed in aligned media versus non-aligned media. Statistical uncertainties were estimated by taking the standard deviation among RDC values measured in triplicate. The average uncertainties were 1.52 Hz for 4% C8E5, and 1.04 Hz for Pf1 phage. The initial alignment tensors to assess the independence of RDC datasets were obtained using the Prediction of ALignmEnt from Structure (PALES) software from Zweckstetter and Bax^{11, 12} (e.g. Fig. S1).

Map-restrained Self-Guided Langevin Dynamics

To generate Pin1 conformations relevant to NH RDCs, we pursued an MD strategy that emphasized relative domain motion, rather than intradomain motion. Specifically, we fixed the PPIase domain in space, and then moved the WW domain relative to it as a nearly rigid body (Fig. 1). We immobilized the PPIase domain atoms using strong Cartesian harmonic restraints (force constants of 1.0 kcal/mol·Å²); indeed, the PPIase RMSD values < 0.5 Å over the simulations (Fig. S3). To move the WW domain as a nearly rigid body, we used self-guided Langevin dynamics⁶⁸ together with map-restraints²³ – a method previously described as MapSGLD²². Each “ith” atom obeyed the equations of motion

$$\frac{d\mathbf{p}_i}{dt} = \left(-\vec{\nabla}_i U - \gamma_i \mathbf{p}_i + \mathbf{R}_i \right) + \mathbf{g}_i + \mathbf{f}_{i,MAP}. \quad (6)$$

The right-hand-side of Eq 6 denotes three types of forces. The first type, in the parentheses, are standard Langevin dynamics (LD) forces which include spatial gradients of the AMBER12 potential function (FF12SB)^{69, 70}, a solvent drag force $-\gamma \mathbf{p}_i$, and a delta-correlated random force $\mathbf{R}_i$. The drag and random forces constitute the implicit solvent model.

The second type of force is the guiding-force “ $\mathbf{g}_i$,” that distinguishes SGLD from standard LD. The guiding force promotes low frequency, collective motions, such as domain re-orientation. It consists of two parts

$$\mathbf{g}_i(t) = \lambda \gamma \langle \mathbf{p}_i(t) \rangle_L - \xi \gamma \mathbf{p}_i(t). \quad (7)$$

The first part is the local time average of the momentum, $\langle \mathbf{p}(t) \rangle_L$, which filters for lower frequency collective motions. Local time-averaging occurs over a duration $t_L = 0.5$ picoseconds (10^{-12} sec) that far exceeds the basic simulation time-step, $\delta\tau = 1$ femtoseconds (10^{-15} sec). The second term has a ξ factor that compensates for the spurious energy increase that would otherwise come from $\mathbf{g}_i$. λ sets the overall strength of $\mathbf{g}_i$.

The third force term on the far-right term of Eq 1, $\mathbf{f}_{i,MAP}$ refers to a protein map function (‘map’, hereafter) and its restraint potential²³. The map is what allows a group of atoms to move as a nearly rigid structure, while maintaining its full atomic description. More specifically, the map is an effective mass density $\rho(x_j, y_j, z_k)$, defined on a 3-d lattice of points, and created from atomic structural coordinates²². The deviation from complete rigidity is controlled by the assumed resolution of the density map. We defined the Pin1 WW map from its coordinates in PDB deposition 1NMV³³. The map and its average value, $\bar{\rho}$, were incorporated into the restraint potential,

$$E_{MAP} = -C_{MAP} \sum_l^N m_l \{ \rho(x_l, y_l, z_l) - \bar{\rho} \} / \delta\rho, \quad (8)$$

where C_{MAP} sets the overall magnitude ($C_{MAP} = 0.1 \text{ kcal/amu}$). The sum includes all atoms of the WW domain, and a normalized map is used where $\delta\rho = 1 \text{ \AA}^{-3}$ ²². Other critical parameters values are in Table 1. The essential feature of Eq 8 is that E_{MAP} decreases when the WW domain coordinates adopt configurations reproducing the initial map density. The $\mathbf{f}_{i,MAP}$ terms are the spatial gradients of E_{MAP} that push WW domain atoms to positions reducing E_{MAP} , thereby maintaining its internal atomic structure. In this manner, E_{MAP} and $\mathbf{f}_{i,MAP}$ promote nearly rigid body motion of the WW domain while maintaining its full-atom description. We chose parameters that kept the internal structure of the WW domain within about 1 Å of its initial structure (see Fig. S3, Supporting Information).

Map restraints and MapSGLD-NMR Simulations

Both MapSGLD-NMR and standard explicit solvent simulations were conducted in AMBER 12⁶⁹ using the ff12SB force field. For the initial Pin1 coordinates, we used three structures from the PDB deposition 1NMV, the results of the NMR structure determination of apo Pin1³³. For each initial structure, we generated AMBER topology and coordinate files using the auxiliary program, “tleap”. Each initial set of coordinates then underwent 500 steps of energy minimization (without NMR restraint potentials), using the improved constant pH generalized Born implicit solvent with a 12 Å cut-off^{71, 72}. Typical energy reductions were to −3000 – −5000 kcal/mole. To generate the RDC restraint potentials, we used the AMBER auxiliary program, make_dip_RST.protein. The resulting restraint potentials E_{RDC} were of the form given in Eq 1 above.

The MapSGLD-NMR simulations moved the Pin1 WW domain as a rigid body relative to a fixed PPIase domain using E_{MAP} per Eq 8 above²². The map function is an effective density function $\hat{\rho}(x_i, y_j, z_k)$ defined on a 3-d grid of $N_X N_Y N_Z$ points, where

$$\hat{\rho}(x_i, y_j, z_k) = \{ \rho(x_i, y_j, z_k) - \bar{\rho} \} / \delta\rho \quad (9)$$

and

$$\left\langle \rho \right\rangle = \frac{1}{N_X N_Y N_Z} \sum_{i,j,k=1}^{N_X, N_Y, N_Z} \rho(x_i, y_j, z_k), \quad \rho\delta = \sqrt{\frac{1}{N_X N_Y N_Z} \sum_{i,j,k=1}^{N_X, N_Y, N_Z} (\rho(x_i, y_j, z_k) - \langle \rho \rangle)^2} \quad (10)$$

The summation indices are over the grid points defining the map function. The map function $\rho(x,y,z)$ describes the internal shape of the WW domain to be maintained. The WW domain rigidity was adjusted using the map resolution parameter, R_{MAP} , which has dimensions of length. Smaller R_{MAP} describes the targeted conformation with finer grid spacing. Thus, E_{MAP} potentials with smaller R_{MAP} increase internal domain rigidity; this includes bond vector mobility that would affect the scaling parameter S_{HN} in Eq 2. We used a small value of $R_{\text{MAP}} = 5 \text{ \AA}$, which led to a WW domain RMSD of $\sim 2.0 \text{ \AA}$.

Explicit Solvent MD simulations

We ran standard MD simulations in explicit solvent using the TIP4P water model⁷³ and the same AMBER ff12SB force field. The initial Pin1 structure for these simulations was the first structure (1NMV-1) from the NMR solution structure 1NMV³³. The initial structure was energy minimized as follows: the protons were first minimized for 1000 steps, then the solvent was minimized for 1000 steps, then the whole system was minimized for 2500 steps. The system was then equilibrated at constant volume for 40 picoseconds and at constant pressure for 40 picoseconds, while the protein was restrained. The production run was a 400 nanoseconds simulation using Particle Mesh Ewald Molecular Dynamics (PMEMD) using an NVIDIA GPU C2050 card on a local machine at Rutgers University.

Trajectory Analysis

For all simulations, we monitored kinetic and potential energies using the publicly available Perl script, `proc_out_md.perl`. The potential energies included the standard AMBER protein potential,⁶⁹ and those of the E_{MAP} and NMR E_{RDC} restraint potentials. To calculate histograms of selected angles and distances, and RMSDs, we used the CPPTRAJ tool in AMBER12.

Genetic Algorithm/Multi-conformational Search

The genetic algorithm (GA) multi-conformational search⁴⁷ was performed on raw MapSGLD-NMR ensembles (90,000 conformations). Prior to GA fitting, we trimmed the raw MapSGLD-NMR ensemble by clustering conformations according to the coordinates interdomain vector coordinates of Fig. 4. We assigned conformations to different “boxes” within a 3-d grid, each box with edges of $2 \text{ \AA} \times 2 \text{ deg} \times 2 \text{ deg}$. Then, all conformers in the same box were replaced by the single conformation with the lowest individual Q_{RDC} factor, as calculated using PALES in the “-stPales” mode. For the MapSGLD-NMR ensemble based on both alignment media, this approach reduced the raw 90,000 conformations to 1000 conformations. We then performed 1,000,000 GA steps on these reduced ensembles, and converged to a final ensemble of six conformers (Fig’s 3, 4) with fractional populations > 0.001 .

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH Grant No. RO1-GM083081 to J.W.P., NIH training grant T32GM075762 to J.J.B., and NIH Grant GM103297 to D.A.C. Computing resources at Rutgers were funded in part by NIH grant S10OD012346-01A1. We thank Dr. Xiongwu Wu, Dr. In-Suk Joung, Dr. Pawel Janowski, Dr. Brendan Mahoney, Ms. Meiling Zhang, Dr. Kimberly Wilson, Dr. Xingsheng Wang, and Dr. Thomas Frederick for valuable suggestions and useful discussions.

ABBREVIATIONS

SGLD	self-guided Langevin Dynamics
PPIase	peptidyl-prolyl isomerase
RDC	residual dipolar coupling
GA	genetic algorithm
WT	wild-type
IPAP	in-phase antiphase
PALES	“Prediction of ALignmEnt from Structure”

References

1. Pufall MA, Graves BJ. Autoinhibitory domains: modular effectors of cellular regulation. *Annu Rev Cell Dev Biol.* 2002; 18:421–62. [PubMed: 12142282]
2. Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA. Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual Review of Biochemistry.* 2006:655–80.
3. Lim WA. Designing customized cell signalling circuits. *Nature Reviews Molecular Cell Biology.* 2010; 11:393–403. [PubMed: 20485291]
4. Bax A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* 2003; 12:1–16. [PubMed: 12493823]
5. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. NMR evidence for slow collective motions in cyanometmyoglobin. *Nature Structural Biology.* 1997; 4:292–7. [PubMed: 9095197]
6. Tjandra N, Bax A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science.* 1997; 278:1111–4. [PubMed: 9353189]
7. Tolman JR, Prestegard JH. A quantitative J-correlation experiment for the accurate measurement of one-bond amide ^{15}N - ^1H couplings in proteins. *J Magn Reson B.* 1996; 112:245–52. [PubMed: 8921605]
8. Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson.* 1998; 131:373–8. [PubMed: 9571116]
9. Losonczi JA, Andrec M, Fischer MWF, Prestegard JH. Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance.* 1999; 138:334–42. [PubMed: 10341140]
10. Tolman JR, Ruan K. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem Rev.* 2006; 106:1720–36. [PubMed: 16683751]
11. Zweckstetter M, Bax A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *Journal of the American Chemical Society.* 2000; 122:3791–2.
12. Zweckstetter M. NMR: prediction of molecular alignment from structure using the PALES software. *Nature Protocols.* 2008; 3:679–90. [PubMed: 18388951]

13. Case DA. Molecular dynamics and NMR spin relaxation in proteins. *Acc Chem Res.* 2002; 35:325–31. [PubMed: 12069616]
14. van Gunsteren WF, Brunne RM, Gros P, van Schaik RC, Schiffer CA, Torda AE. Accounting for molecular mobility in structure determination based on nuclear magnetic resonance spectroscopic and X-ray diffraction data. *Methods Enzymol.* 1994; 239:619–54. [PubMed: 7530320]
15. Bruschweiler R, Case DA. Collective NMR relaxation model applied to protein dynamics. *Phys Rev Lett.* 1994; 72:940–3. [PubMed: 10056573]
16. Bruschweiler R, Roux B, Blackledge M, Griesinger C, Karplus M, Ernst RR. Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution. *J Am Chem Soc.* 1992; 114:2289–302.
17. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-level characterization of the structural dynamics of proteins. *Science.* 2010; 330:341–6. [PubMed: 20947758]
18. Garcia de la Torre J, Ortega A, Perez Sanchez HE, Hernandez Cifre JG. MULTHYDRO and MONTEHYDRO: conformational search and Monte Carlo calculation of solution properties of rigid or flexible bead models. *Biophys Chem.* 2005; 116:121–8. [PubMed: 15950824]
19. Bernado P, Fernandes MX, Jacobs DM, Fiebig K, de la Torre JG, Pons M. Interpretation of NMR relaxation properties of Pin1, a two-domain protein, based on Brownian dynamic simulations. *Journal of Biomolecular Nmr.* 2004; 29:21–35. [PubMed: 15017137]
20. Garcia de la Torre J, Navarro S, Lopez Martinez MC, Diaz FG, Lopez Cascales JJ. HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. *Biophys J.* 1994; 67:530–1. [PubMed: 7948671]
21. Tsui V, Zhu LM, Huang TH, Wright PE, Case DA. Assessment of zinc finger orientations by residual dipolar coupling constants. *Journal of Biomolecular Nmr.* 2000; 16:9–21. [PubMed: 10718608]
22. Wu X, Subramaniam S, Case DA, Wu KW, Brooks BR. Targeted conformational search with map-restrained self-guided Langevin dynamics: application to flexible fitting into electron microscopic density maps. *J Struct Biol.* 2013; 183:429–40. [PubMed: 23876978]
23. Wu XW, Milne JLS, Borgnia MJ, Rostapshov AV, Subramaniam S, Brooks BR. A core-weighted fitting method for docking atomic structures into low-resolution maps: Application to cryo-electron microscopy. *Journal of Structural Biology.* 2003; 141:63–76. [PubMed: 12576021]
24. Xia JC, Margulis CJ, Case DA. Searching and Optimizing Structure Ensembles for Complex Flexible Sugars. *Journal of the American Chemical Society.* 2011; 133:15252–5. [PubMed: 21863822]
25. Salmon L, Bascom G, Andricioaei I, Al-Hashimi HM. A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for interhelical motions revealed. *J Am Chem Soc.* 2013; 135:5457–66. [PubMed: 23473378]
26. Berlin K, Castaneda CA, Schneidman-Duhovny D, Sali A, Nava-Tudela A, Fushman D. Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *J Am Chem Soc.* 2013; 135:16595–609. [PubMed: 24093873]
27. Lu KP, Hanes SD, Hunter T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature.* 1996; 380:544–7. [PubMed: 8606777]
28. Lu PJ, Wulf G, Zhou XZ, Davies P, Lu KP. The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein. *Nature.* 1999; 399:784–8. [PubMed: 10391244]
29. Liao XH, Zhang AL, Zheng M, Li MQ, Chen CP, Xu H, et al. Chemical or genetic Pin1 inhibition exerts potent anticancer activity against hepatocellular carcinoma by blocking multiple cancer-driving pathways. *Sci Rep.* 2017; 7:43639. [PubMed: 28262728]
30. Zhou XZ, Lu KP. The isomerase PIN1 controls numerous cancer-driving pathways and is a unique drug target. *Nat Rev Cancer.* 2016; 16:463–78. [PubMed: 27256007]
31. Ranganathan R, Lu KP, Hunter T, Noel JP. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell.* 1997; 89:875–86. [PubMed: 9200606]

32. Verdecia MA, Bowman ME, Lu KP, Hunter T, Noel JP. Structural basis for phosphoserine-proline recognition by group IVWW domains. *Nature Structural Biology*. 2000; 7:639–43. [PubMed: 10932246]
33. Bayer E, Goettisch S, Mueller JW, Griewel B, Guiberman E, Mayr LM, et al. Structural analysis of the mitotic regulator hPin1 in solution - Insights into domain architecture and substrate binding. *Journal of Biological Chemistry*. 2003; 278:26183–93. [PubMed: 12721297]
34. Jacobs DM, Saxena K, Vogtherr M, Bernado P, Pons M, Fiebig KM. Peptide binding induces large scale changes in inter-domain mobility in human Pin1. *Journal of Biological Chemistry*. 2003; 278:26174–82. [PubMed: 12686540]
35. Namanja AT, Wang XDJ, Xu BL, Mercedes-Camacho AY, Wilson KA, Etkorn FA, et al. Stereospecific gating of functional motions in Pin1. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:12289–94. [PubMed: 21746900]
36. Wilson KA, Bouchard JJ, Peng JW. Interdomain interactions support interdomain communication in human Pin1. *Biochemistry*. 2013; 52:6968–81. [PubMed: 24020391]
37. Wang X, Mahoney B, Zhang M, Zintsmaster JS, Peng JW. Negative Regulation of Peptidyl-Prolyl Isomerase Activity by Interdomain Contact in Human Pin1. *Structure*. 2015 in press.
38. Peng JW. Investigating Dynamic Interdomain Allostery in Pin1. *Biophys Rev*. 2015; 7:239–49. [PubMed: 26495045]
39. Wang X, Mahoney BJ, Zhang M, Zintsmaster JS, Peng JW. Negative Regulation of Peptidyl-Prolyl Isomerase Activity by Interdomain Contact in Human Pin1. *Structure*. 2015; 23:2224–33. [PubMed: 26602185]
40. Ruckert M, Otting G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society*. 2000; 122:7793–7.
41. Hansen MR, Hanson P, Pardi A. Pf1 filamentous phage as an alignment tool for generating local and global structural information in nucleic acids. *J Biomol Struct Dyn*. 2000; 17(Suppl 1):365–9. [PubMed: 22607445]
42. Moltke S, Grzesiek S. Structural constraints from residual tensorial couplings in high resolution NMR without an explicit term for the alignment tensor. *J Biomol NMR*. 1999; 15:77–82. [PubMed: 20703963]
43. Tolman JR. A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *Journal of the American Chemical Society*. 2002; 124:12020–30. [PubMed: 12358549]
44. Lipari G, Szabo A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J Am Chem Soc*. 1982; 104:4546–59.
45. Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *J Am Chem Soc*. 1998; 120:6836–7.
46. Montalvao RW, De Simone A, Vendruscolo M. Determination of structural fluctuations of proteins from structure-based calculations of residual dipolar couplings. *J Biomol NMR*. 2012; 53:281–92. [PubMed: 22729708]
47. Xia J, Margulis CJ, Case DA. Searching and optimizing structure ensembles for complex flexible sugars. *J Am Chem Soc*. 2011; 133:15252–5. [PubMed: 21863822]
48. Fisher CK, Zhang Q, Stelzer A, Al-Hashimi HM. Ultrahigh resolution characterization of domain motions and correlations by multialignment and multireference residual dipolar coupling NMR. *J Phys Chem B*. 2008; 112:16815–22. [PubMed: 19367865]
49. Frank AT, Stelzer AC, Al-Hashimi HM, Andricioaei I. Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Res*. 2009; 37:3670–9. [PubMed: 19369218]
50. Pearlman DA. FINGAR: A new genetic algorithm-based method for fitting NMR data. *J Biomol NMR*. 1996; 8:49–66. [PubMed: 21136323]
51. Zweckstetter M, Hummer G, Bax A. Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophysical journal*. 2004; 86:3444–60. [PubMed: 15189846]

52. Iwahara J, Tang C, Marius Clore G. Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J Magn Reson.* 2007; 184:185–95. [PubMed: 17084097]
53. Clore GM, Schwieters CD. How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *Journal of the American Chemical Society.* 2004; 126:2923–38. [PubMed: 14995210]
54. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KF, Becker S, et al. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science.* 2008; 320:1471–5. [PubMed: 18556554]
55. Blackledge M. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog Nucl Magn Reson Spectr.* 2005; 46:23–61.
56. Yaffe MB, Schutkowski M, Shen M, Zhou XZ, Stukenberg PT, Rahfeld JU, et al. Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. *Science.* 1997; 278:1957–60. [PubMed: 9395400]
57. Shen M, Stukenberg PT, Kirschner MW, Lu KP. The essential mitotic peptidyl-prolyl isomerase Pin1 binds and regulates mitosis-specific phosphoproteins. *Genes Dev.* 1998; 12:706–20. [PubMed: 9499405]
58. Stukenberg PT, Kirschner MW. Pin1 acts catalytically to promote a conformational change in Cdc25. *Mol Cell.* 2001; 7:1071–83. [PubMed: 11389853]
59. Helander S, Montecchio M, Pilstal R, Su Y, Kuruvilla J, Elven M, et al. Pre-Anchoring of Pin1 to Unphosphorylated c-Myc in a Fuzzy Complex Regulates c-Myc Activity. *Structure.* 2015; 23:2267–79. [PubMed: 26655473]
60. Follis AV, Llambi F, Merritt P, Chipuk JE, Green DR, Kriwacki RW. Pin1-Induced Proline Isomerization in Cytosolic p53 Mediates BAX Activation and Apoptosis. *Mol Cell.* 2015; 59:677–84. [PubMed: 26236013]
61. Wei S, Kozono S, Kats L, Nechama M, Li W, Guarnerio J, et al. Active Pin1 is a key target of all-trans retinoic acid in acute promyelocytic leukemia and breast cancer. *Nat Med.* 2015; 21:457–66. [PubMed: 25849135]
62. Liou YC, Zhou XZ, Lu KP. Prolyl isomerase Pin1 as a molecular switch to determine the fate of phosphoproteins. *Trends Biochem Sci.* 2011; 36:501–14. [PubMed: 21852138]
63. Lippens G, Landrieu I, Smet C. Molecular mechanisms of the phospho-dependent prolyl cis/trans isomerase Pin1. *Febs J.* 2007; 274:5211–22. [PubMed: 17892493]
64. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proceedings of the National Academy of Sciences of the United States of America.* 2000; 97:8868–73. [PubMed: 10908673]
65. Namanja AT, Peng T, Zintsmaster JS, Elson AC, Shakour MG, Peng JW. Substrate recognition reduces side-chain flexibility for conserved hydrophobic residues in human Pin1. *Structure.* 2007; 15:313–27. [PubMed: 17355867]
66. Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology.* 1998; 5:1065–74. [PubMed: 9846877]
67. Ottiger M, Delaglio F, Marquardt JL, Tjandra N, Bax A. Measurement of dipolar couplings for methylene and methyl sites in weakly oriented macromolecules and their use in structure determination. *Journal of Magnetic Resonance.* 1998; 134:365–9. [PubMed: 9761712]
68. Wu XW, Brooks BR. Self-guided Langevin dynamics simulation method. *Chemical Physics Letters.* 2003; 381:512–8.
69. Cheatham, TE., III, Simmerling, CL., Wang, J., Duke, RE., Luo, R., Walker, RC., Zhang, W., Merz, KM., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, AW., Kolossváry, I., Wong, KF., Paesani, F., Vanicek, J., Wolf, RM., Liu, J., Wu, X., Brozell, SR., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M-J., Cui, G., Roe, DR., Mathews, DH., Seetin, MG., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., Kollman, PA. Case DADTA. AMBER 12. University of California; San Francisco: 2012.
70. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, et al. The Amber biomolecular simulation programs. *J Comput Chem.* 2005; 26:1668–88. [PubMed: 16200636]

71. Mongan J, Case DA, McCammon JA. Constant pH molecular dynamics in generalized born implicit solvent. *J Comput Chem.* 2004; 25:2038–48. [PubMed: 15481090]
72. Simonson T, Carlsson J, Case DA. Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models. *Journal of the American Chemical Society.* 2004; 126:4167–80. [PubMed: 15053606]
73. Jorgensen W, JDM. Temperature and size dependence of Monte Carlo simulations of TIP4P water. *Mol Phys.* 1985; 56:1381–92.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HIGHLIGHTS

- Efficient generation of conformational ensembles for NMR probes of domain motion
- Protocol combines self-guided Langevin dynamics with map and NMR RDC restraints
- NH RDCs fit to weighted populations of domain conformations; applied to human Pin1
- RDC-based ensembles support findings of other Pin1 NMR parameters

Fig. 1.

Human Pin1, a two-domain mitotic regulator with PPIase (cyan) and WW (magenta) domains; the structure is a snapshot from a 30ns MapSGLD-NMR simulation started from a conformer of 1NMV³³. The $-Z$ and $-X$ axes are the principal axes of PPIase domain inertia tensor (origin at its center of mass). Linker flexibility (green curved arrow) enables relative domain motion. Residue I28 lies in the WW domain (Yellow), the site of the I28A substitution. The strategy for simulating interdomain motion is: (i) strong Cartesian restraints hold the PPIase fixed in space, (ii) while the WW domain moves as a nearly rigid body by map restraints (green curved arrow).

Fig. 2.

(A) RDC Q factors (blue and maroon bars) for assessing agreement between ^1H - ^{15}N RDCs predicted from AMBER generated conformations with those from the NMR experiments. The Q factors are averages over the three parallel production-run simulations started from structures 1NMV-1, 1NMV-2, and 1NMV-4 from the solution structure deposition PDB 1NMV³³; the error bars are the corresponding standard deviations of Q factors. The predicted RDCs came from singular value decomposition (SVD) fitting to the experimental RDCs⁹, using the “-bestFit” option in PALES¹². Pairs of blue and maroon bars indicate Q factors for 4% C8E5/octanol media (blue), and 10 mg/ml Pf1 phage (maroon), respectively. Q factors for full-length Pin1, the WW domain, and the PPIase domain are indicated. The dashed-dotted line distinguishes two sets of Q factors. Left of the line are Q-factors just after energy minimization and before the production-run simulations. Right of the line are the analogous Q factors after 10 ns of the three production-run AMBER simulations, using the MapSGLD-NMR protocol. The reduction in Q factors after 10 ns (1/3 of the total simulation time per starting structure) is apparent. (B, C) Representative correlation plots between the ^1H - ^{15}N RDCs predicted by AMBER conformations (x-axes) versus experiment (y-axes). Black circles are RDCs in and 4% C8E5; red squares are RDCs in 10 mg/ml Pf1 phage. In (B), the standard linear regression coefficient is $r = 0.88 \pm 0.07$ for 4% C8E5 (black circles) and $r = 0.87 \pm 0.03$ for 10 mg/ml Pf1 phage (red squares); (C) after 10 nanoseconds of production-run simulations, the standard linear regression coefficient is $r = 0.991 \pm 0.002$ for 4% C8E5 (black circles) and $r = 0.984 \pm 0.008$ for 10 mg/ml Pf1 phage (red squares).

Fig. 3.

Identification of reduced conformational ensembles using the genetic algorithm (GA) approach for multi-conformational fit⁴⁷. The bars indicate conformations with a fractional population > 0.001 for WT-Pin1 (**A**) and the single-substitution mutant, I28A-Pin1 (**C**). Linear correlation plots of predicted versus experimental ^1H - ^{15}N RDCs recorded in 4% C8E5 media for WT-Pin1 (**B**) and I28A-Pin1 (**D**). Open circles (O) are RDCs predicted by the multi-conformational fit (WT $Q_M = 0.196$, I28A-Pin1 $Q_M = 0.202$); turquoise crosses + are RDCs predicted by the GA ensemble conformation yielding the lowest single conformer Q factor (best single conformational fit): WT-Pin1 $Q = 0.236$, and I28A-Pin1 $Q = 0.280$.

Fig. 4.

WT-Pin1 conformations sampled by three 30ns AMBER simulations, using the MapSGLD-NMR protocol. The simulations started from three different structures (1NMV-1, 1NMV-2, and 1NMV-3) producing a total of 90,000 conformations. (A) The relative domain orientation is specified by an interdomain vector (green arrow) that starts at the geometric center of the PPIase domain active site and ends at the geometric center of the WW domain substrate-binding loop (residues 16-21). The interdomain vector coordinates (length r and angles θ and ϕ) are in the principle axis frame of the PPIase domain inertia tensor. (B) The six conformers identified by the GA/Multi-conformational search, and their corresponding interdomain vectors (green arrows defined in (A)). Panels (C–E) show histograms for the spherical coordinates of the interdomain vector, including (C) the radial (distance) coordinate r ; (D) the polar angle θ ; and (E) the azimuthal angle ϕ . Panel (F) is the average and the radius of gyration. Black bars represent the 90,000 conformations of the raw WT MapSGLD-NMR ensemble; the red trace represents the analogous 90,000 conformations for the raw I28A MapSGLD-NMR ensemble. Green diamonds (◆) indicate the six WT conformers; red x's (✱) indicate the seven I28A conformers identified by GA/multi-conformational search; stacked diamonds/x's are populations in nearly the same bins. Panels (G and H) compare conformations sampled by the raw WT-Pin1 *versus* I28A-Pin1 ensembles (i.e. MapSGLD-NMR simulations producing 90,000 conformations). Yellow-orange ribbons indicate the most frequently sampled conformation; magenta indicates a less populated conformation.

Fig. 5.

2-d histograms comparing the conformational space sampled by **(A)** 20 nanoseconds rigid-body Map-SGLD-NMR simulations, implicit solvent model, and map and RDC restraints enforced; **(B)** 20 nanoseconds Map-SGLD simulation, implicit solvent model, map restraints enforced but RDC restraints absent; **(C)** depicts 400 nanoseconds explicit-solvent simulations lacking both map and RDC restraints. Comparison of **(A)** and **(B)** indicates certain conformations are excluded by the RDC data; Comparison of **(B)** and **(C)** indicates many conformations are not sampled in the explicit-solvent simulation. All simulations were started from the same structure, marked by the star.

Fig. 6.

Comparison of interdomain vector orientations (r , Θ , ϕ) sampled by WT-Pin1 and I28A-Pin1. The PPIase domains are superimposed and fixed (cyan); the black lines are the $-X$ and $-Z$ principal axes of its inertia tensor. The dots represent the tips of the interdomain vectors starting at the geometric center of the PPIase domain active site, and terminating at the geometric center of the WW domain-binding loop. Blue dots are the WT-Pin1 ensemble; gold dots are the I28A-Pin1 ensemble. The interdomain vectors (red lines) and WW domains (dark blue) for conformers 4 and 6 from the WT GA ensemble are shown.

Table 1

Parameters for MapSGLD-NMR Simulations

Parameter	Significance	Value
k_{RDC}	E_{RDC} Constant	0.1 kcal/mol-Hz ²
C_{MAP}	E_{MAP} Constant	0.1 kcal/amu,
R_{MAP}	Map Resolution	5 Å
γ_{MAP}	map friction coefficient	0.1 ps ⁻¹
γ	atomic friction coefficient	5 ps ⁻¹
δt	time-step	1 fs
t_{L}	window for local time-averaging	0.5 ps (500 time steps)
λ	guiding factor	1.0
T_{SG}	guiding force enhancement	400K

Table 2

Distribution of Single Conformer Q Factors from Different MapSGLD-NMR Simulations

RDC restraints Enforced	Min Q	Max Q	Average Q
Both Media RDCs (83+93)	0.236 [*]	1.054 [*]	0.858 [*]
Exclusive 4% C8E5 RDCs (83)	0.257	1.039	0.774
Exclusive phage RDCs (93)	0.560	1.054	0.915
No RDC restraints	0.527	1.053	0.930

^{*} Q factors calculated using on the 4% C8E5 experimental RDCs