

PCCP

Physical Chemistry Chemical Physics

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: T. Karamanos, B. Hobbs, N. Limmer, G. L. Clenshaw and F. Ossa, *Phys. Chem. Chem. Phys.*, 2026, DOI: 10.1039/D6CP01833E.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Accurate interdomain contacts in a mixed folded protein from NMR-guided coarse-grained simulations

View Article Online
DOI: 10.1039/C6CP01833E

Billy Hobbs^{1‡}, Noor Limmer^{1‡}, Gwendolyn L. Clenshaw¹, Felipe Ossa¹, & Theodoros K. Karamanos^{1*}

¹Department of Life Sciences, Faculty of Natural Sciences, Imperial College London, UK
NMR spectroscopy, Intrinsically disordered proteins, Molecular dynamics simulations

‡These authors contributed equally

*Correspondence to t.karamanos@imperial.ac.uk (TKK)

Abstract

Intrinsically disordered, low-complexity regions frequently cooperate with folded domains to mediate protein–protein interactions, yet accurately describing these mixed folded–disordered systems remains challenging. To visualize these mixed folded proteins, experimentally guided coarse-grained (CG) molecular dynamics simulations are often employed to extend the timescales required to capture the complex dynamics in play. However, the minimalistic nature of these approaches often compromises structural accuracy and can lead to inaccurate inter-domain interactions. Here we introduce backbone dihedral terms directly derived from NMR chemical shift data in CG-simulations to characterize the open state of a mixed-folded construct of the anti-aggregation chaperone DNAJB6 that contains a folded J-domain and a disordered GF linker. By tuning residue-specific backbone dihedral parameters to match NMR-derived secondary-structure propensities of the linker in CG-simulations, we generate conformational ensembles that yield accurate interdomain contact maps. In agreement with analysis of NMR relaxation data, the resulting ensembles show that even in the nominally open state the linker experiences motions that resemble those of the closed state driven by hydrophobic residues in GF. More generally, we show that by expanding CG-simulations to allow them to capture both local and global structural properties, physically relevant interdomain contacts can be retrieved.

Introduction

Low complexity regions (LCRs) in proteins have attracted significant research interest as they serve crucial roles in important biological processes.^{1,2} In many cases intrinsically disordered LCRs mediate interactions between folded domains by increasing the local concentration of the latter or by acting as hotspots for protein–protein interactions.² However, and despite their importance, the atomic



level characterization of low complexity regions still represents a significant task due to their inherent dynamics and similar spectroscopic signatures that complicate analysis by the majority of structural biology methods.

Computational methods have provided crucial insights into the conformational ensembles of intrinsically disordered regions (IDRs), often guided by experimental data.³ However, the millisecond timescales required to adequately sample IDR conformational space constitute all-atom molecular dynamics (MD) simulations prohibitively expensive, motivating the development of coarse-grained (CG) models.^{4,5} While CG simulations successfully capture global ensemble properties such as chain compaction, their minimalistic nature often complicates the incorporation and/or extraction of critical sequence-specific features. In particular, residue-level secondary-structure propensities—known to be functionally important in many IDRs—are difficult to represent accurately, leading to incorrect or non-physical inter-residue and interdomain contact patterns.⁶ As a result, CG approaches frequently struggle to describe the interplay between folded domains and disordered regions.^{6,7} Here, using a J-domain chaperone as a model system, we demonstrate that incorporating experimentally derived secondary-structure propensities is sufficient to recover accurate interdomain contact maps in a mixed folded–disordered protein.

The J-domain protein (JDP) family of human molecular chaperones contain a typical, low complexity, glycine and phenylalanine-rich (GF) linker that connects their J-domain with their C-terminal, substrate binding domain and is crucial for cell viability.⁸ Normally, JDPs cooperate with Hsp70 chaperones to maintain intracellular proteostasis. However, DNAJB6, a Class-B JDP, has been established as a potent inhibitor of amyloid aggregation, independent of its interaction with Hsp70.⁹ In DNAJB6, part of the GF forms a stable α -helix (helix 5 or $\alpha 5$) that blocks the binding of Hsp70 to the J-domain creating a closed/autoinhibited state.¹⁰ Even though a number of publications investigating the closed/autoinhibited state of DNAJBs are now available,^{11,12} the characterization of the open/uninhibited state is lacking. Since the GF has been implicated in various human pathologies¹¹ and has also recently been shown to serve as a pseudo-substrate for Hsp70¹² understanding its dynamics and range of motions is crucial in determining its role in JDP-mediated proteostasis. Previously, using a construct of the J-domain followed by GF in the absence of helix 5 (JD-GF) as a proxy for the open state we showed by SAXS and NMR spectroscopy that GF partially collapses against JD.¹² Even though nominally ‘open’, the JD-GF construct shows an order of magnitude decrease in affinity for Hsp70 in comparison to JD alone, presumably due to GF- JD interactions which remain elusive in the atomic level.

In the present study we use solution NMR and CG-simulations to visualize the conformational ensemble sampled by the highly flexible JD-GF construct. Using NMR relaxation methods, we show that specific hydrophobic residues in the otherwise largely disordered GF show reduced motions in the ns timescale. To visualize the interactions between JD and GF we performed CG simulations that reproduce the overall compaction of the molecule but fail to capture the exact interdomain contacts



between JD and GF. To overcome this limitation, we introduce backbone dihedral potentials that encode residue-specific secondary-structure propensities derived from NMR chemical shifts into minimal CG simulations resulting in good agreement between the simulation derived JD – GF interdomain contacts and the NMR data. Using this combined NMR and MD approach we generate a physically relevant ensemble for a disordered region in which specific phenylalanine residues in GF form extensive contacts with the J-domain resulting in partially closed states that interfere with Hsp70 binding. More generally, we show that incorporating experimentally-derived local backbone propensities provides an avenue to improving CG descriptions of disordered proteins, enabling access to atomic-level insights that are otherwise difficult to obtain.

Methods

Protein expression and purification. DNAJB6 JD-GF, JD-GF- $\alpha 5$ and DNAJB1 JD-GF constructs were expressed and purified as described previously.¹²

Residual dipolar coupling (RDC) measurements. RDC measurements were collected at 600 MHz on a sample of 200 μM ^{13}C , ^{15}N -labelled DNAJB6 JD-GF in 20 mM sodium phosphate pH 7.4, 100 mM NaCl, aligned in 14 mg/mL bacteriophage Pf1 (ASLA Scientific) Backbone amide $^1D_{\text{NH}}$ RDCs were measured using the ARTSY pulse sequence.¹³

Paramagnetic relaxation enhancement (PRE) experiments. Solvent PRE experiments were performed by measuring the ^1H R_2 rate in the absence (diamagnetic sample) and presence of 5 mM TEMPO (paramagnetic sample) using a set of 6 relaxation delays ranging from 2 to 22 ms. The PRE Γ_2 rate is then given by $\Gamma_2 = R_2^{\text{para}} - R_2^{\text{dia}}$.

Spectral density mapping. Reduced spectral density mapping was performed as in reference [14] (Supplementary methods). For well-folded residues the local rotational correlation time that contains contributions from diffusional anisotropy can be obtained from the values of $J(0)$ and $J(\omega_{\text{N}})$ as:

$$\tau_{\text{M}} = \omega_{\text{N}}^{-1} \left(\frac{J(0) - J(\omega_{\text{N}})}{J(\omega_{\text{N}})} \right)^{\frac{1}{2}}$$

With the value of τ_{m} determined, the amplitude of intramolecular motion of the N-H vector can be estimated by¹⁵:

$$S^2 = \frac{5(J(0) - J(\omega_{\text{N}})) (1 - \omega_{\text{N}}^2 \tau_{\text{M}}^2)}{2\omega_{\text{N}}^2 \tau_{\text{M}}^3}$$



For GF residues the value of τ_m was fixed to the average value obtained for residues in JD. Given a well defined τ_m value for folded residues, equation (2) can provide a good approximation of the generalized order parameter square (S^2) for internal motions in the ns timescale. For example, S^2 values obtained by equation (2) at one magnetic field show an excellent correspondence with the ones obtained by extensive spectral density mapping at five magnetic fields for the mixed folded protein GCN4^{16,17} which shares a lot of structural and dynamic features with JD-GF.

Model-free analysis. Model-free analysis was performed in Model-free 4.¹⁸ Since the closed stated in the JD-GF- $\alpha 5$ construct can be considered as a globular protein with a dynamic linker region, a global correlation time (τ_c) was used for all residues. For the open JD-GF, a global τ_c was used for JD residues. For GF residues τ_c was fixed to the value obtained for JD. For residues in the J-domain an axially symmetric diffusion tensor was used to account for the slight anisotropy of the J-domain ($D_{\parallel}/D_{\perp} \sim 1.6$). The relaxation data were fit to the following models:

$$\text{Model 1: } J(\omega) = \frac{2}{5}\tau_c \left[\frac{S^2}{1+(\omega\tau_c)^2} \right]$$

$$\text{Model 2: } J(\omega) = \frac{2}{5}\tau_c \left[\frac{S^2}{1+(\omega\tau_c)^2} + \frac{(1-S^2)((\tau_e+\tau_c)\tau_e)}{(\tau_c+\tau_e)^2+(\omega\tau_c\tau_e)^2} \right]$$

$$\text{Model 3: } J(\omega) = \frac{2}{5}\tau_c \left[\frac{S^2}{1+(\omega\tau_c)^2} + \frac{(S_f^2-S^2)((\tau_s+\tau_c)\tau_s)}{(\tau_c+\tau_s)^2+(\omega\tau_c\tau_s)^2} \right]$$

Most of the data for JD and $\alpha 5$ residues can be fitted to the simplest model 1 with only a few residues located in less structured loops showing significant fast motions and thus requiring the use model 2. All GF residues in JD-GF could only be fitted using the extended model-free formalism (model 3).¹⁹ No models that contain exchange contributions to relaxation (R_{ex}) were used as there is no evidence of such motions in JD-GF. Model selection was performed using the Bayesian Information Criterion (BIC) and errors on the fitted parameters were obtained from 500 Monte Carlo simulations.

Coarse-grained simulations. Martini3 simulations were performed with the Martini3IDP forcefield²⁰ in OpenMM for a total of 4 (4 x 1) μ s. Initial topology files were generated in martinize2²¹ followed solvation in a 10x10x10 nm box of water using the insane script. Fully coarse-grained simulations where each residue was represented as a single bead were performed within the CALVADOS framework⁵ that uses OpenMM as its underlying engine. All simulations were repeated four times, and convergence and error estimation was performed using blocking analysis (Supplementary Table 1). To allow backbone dihedral angles to be included, the HPS-SS model of Rizuan *et al.*²² (Supplementary methods) was implemented as a C++ plugin for OpenMM and is freely available at <https://github.com/karamanoslab/OpenMMDihedralPlugin>. ϵ_d represents the residue-specific parameter that determines the probability of each residue to be in a helical or extended conformation (see Supplementary methods). ϵ_d values were determined by maximizing the agreement with the observed



secondary structure propensities calculated by Talos based on the backbone $C\alpha$, $C\beta$, $C\gamma$, N , H chemical shifts of JD-GF. The optimized ε_d values are shown in Supplementary Table 2, with a negative/positive value favoring α -helical/ β -strand conformation respectively. A double well angle potential that describes the angle θ between three consecutive residues was also included as described in references [22,23]. The addition of the dihedral angle and angle potentials has little effect on the overall compaction of the ensemble as observed earlier. Since the original HPS-SS model is parametrized for $C\alpha$ dihedral angles we chose to run the CALVADOS simulations with beads centered on the $C\alpha$ atom of each residue. The structure of the folded JD fully satisfies the NMR-derived dihedral angles and there is no evidence to suggest that its backbone adopts multiple conformations.¹⁰ Thus, the backbone potential terms were only applied to the GF linker (residues 75 – 98) while the JD was restrained as a rigid body using harmonic restraints. The N-terminal part of helix 3 has a slight kink picked up by its backbone chemical shifts but is fully helical in terms of its $C\alpha$ configuration. The small helix formed by residues K70 – N74 is too short to be picked up by Talos but has short-range 1H - 1H NOEs fully consistent with a helical conformation. Default values for the stickiness parameter (λ) were used as described in references [4,6].

Analysis of trajectories. Assignment of helical and extended residues in the simulations was performed as follows: if the dihedral angle φ between 4 consecutive beads ($i, i+1, i+2, i+3$) falls within the helical ($0.1 < \varphi < 1.5$ rad) or extended ($-2.5 < \varphi < -0.5$ rad) well, then residues $i+1, i+2$ or $i+1$ were assigned as pseudo-helical, pseudo-extended respectively. If three consecutive residues in a moving tripeptide window were found to be pseudo-helical/pseudo-extended, the middle residue was assigned as helical/extended. Contact order values²⁴ for residue i were calculated as $(1/L \cdot N_i) \cdot \sum_1^{N_i} \Delta S_{i,j}$ where N_i is the total number of contacts for residue i , $\Delta S_{i,j}$ is the sequence separation between i, j residues in a contact pair and L is the total number of residues. Thus, the per residue contact orders reported here highlight residues involved in long range interactions. Contact order values were compared with long-range CSPs calculated as the absolute difference in 1H , ^{15}N peak positions between the JD-GF and JD alone constructs after removing residues within 5 Å of residue G69. Principal component analysis and projection of the trajectory on the principal component vectors was performed in MDanalysis.²⁵ Lifetimes for each contact pair in Martini3IDP simulations were calculated by dividing the total time that each contact was formed (defined by a distance of < 8 Å) by the number of release events (defined by an increase in distance to > 12 Å).²⁶

Results and Discussion

Linker dynamics by NMR relaxation. The dynamics of the DNAJB6 JD-GF were probed by R_1 , R_2 and NOE relaxation data at two fields and analyzed by spectral density mapping.¹⁴ The low complexity sequence combined with the intrinsically disordered nature of GF results in highly overlapped resonances for GF residues and therefore the relaxation data were recorded with a triple resonance



HNCO readout. Raw relaxation rates shown in Figures 1A – 1C are consistent with a highly ordered JD and a dynamic GF although certain residues in the latter seem to be more ordered than others. To gain more insights into the complex dynamics of the GF region the relaxation data were analysed by spectral density mapping¹⁴ in order to extract the values of $J(0)$, $J(\omega_N)$ and $J(\omega_H)$ (see Methods) which are shown in Figures 1D – 1F. Large $J(0)$ and small $J(\omega_H)$ values consistent with a stably folded domain are observed for JD residues as expected. Interestingly, residues K70 – N74 in GF also exhibit the same behaviour of those in JD suggestive of a high degree of order for the N-terminus of GF. The patterns of $J(0)$ and $J(\omega_H)$ are reversed for the rest of GF with small and large values observed respectively in agreement with a highly dynamic region (Figures 1D – 1F). $J(\omega_N)$ values for residues D85 - N93 display a rise and fall in amplitude indicative of a sequence dependent variation in nanosecond dynamics ($\sim 1/\omega_N$). $J(\omega_H)$ values which report on faster, picosecond dynamics ($\sim 1/\omega_H$) and are not contaminated by contributions from the global tumbling of the molecule are lower for residues F87, E88, F89, F91, F93 and R94 especially at 600 MHz indicating reduced motions of these residues in the picosecond timescale. Since $\omega_H\tau_m \gg 1$ for JD-GF, $J(\omega_H)$ values are mainly affected by internal motions and strongly depend on the order parameter squared (S^2) which can be estimated purely from $J(0)$ and $J(\omega_N)$ as shown in Figure 1G. Indeed, S^2 values for F87, E88, F89, F91, F93 are all above 0.2 consistent with a smaller amplitude of motion for these largely hydrophobic residues.

To gain a more comprehensive understanding of the complex dynamics of the GF region, relaxation measurements at multiple magnetic fields can be performed and provided that the measured spectral densities are linear functions of ω , the amplitudes and timescales of motions at various timescales can be extracted with high precision.^{17,27} Alternatively, a more complicated model-free analysis could be performed, although this is conceptually challenging for fully disordered regions due to the absence of a single tumbling time (τ_c). However, for a system like JD-GF where a relatively short IDR is tethered to and makes contacts with a well-folded domain,¹² model-free derived motional parameters have been shown to correlate well with those from spectral density mapping at multiple fields.^{17,19} Based on previous small angle X-ray scattering (SAXS) and chemical shift perturbation (CSP) data¹² it is safe to assume that GF and JD tumble with the same global correlation time (τ_c) and thus, model-free analysis²⁸ can be performed to extract the effective timescales of the internal motions. For helices 1 to 4 a simple model that takes into account the slight anisotropy of the JD diffusion tensor²⁹ and only includes a single order parameter squared (S^2) can adequately describe the relaxation data. An effective correlation time for the internal motions (τ_e) in the order of 50 – 100 ps is only needed for loop residues in JD (Figure 1). For GF residues however, a more complicated extended model-free formalism that includes a slow (S_S^2) and a fast (S_F^2) order parameter is required.¹⁹ For residues K70 – N74 in the GF both S_S^2 and S_F^2 are >0.85 in the open state suggesting a remarkable rigidity of the N-terminal part of the GF-linker. These residues also show ¹H-¹H NOEs in the open state that are consistent with a helical structure as seen in various DNAJs (Figure S2A).¹² Moving towards the C-



terminal part of GF, S^2 values drop drastically as expected for an IDR, except for residues F87, E88 ($S^2 \sim 0.50$) and to a smaller extent F89, F91 and R94 ($S^2 \sim 0.35$) which show reduced motions in comparison to other GF residues. A good correlation is observed between the model-free and spectral density mapping derived S^2 values with a Pearson correlation coefficient of 0.98 and a root-mean-square deviation of 0.06 (Figure S1F). The average effective correlation time for the internal GF motions is on the order of 1 ns, significantly slower than the one observed for the well folded J-domain ($\tau_e \sim 100$ ps). Given the complex dynamics of GF, the obtained parameters are not a literal decomposition of their motions, but can be interpreted as transient, large-scale interactions which can obviously take place at various timescales but are captured by a single effective correlation time. Further support that the GF-linker does not adopt a fully random-coil structure stems from $^1D_{NH}$ residual dipolar coupling (RDC) data (Figure S2B) that show non-zero values for GF residues.

It is instructive to compare the obtained motional parameters for GF residues in the context of the closed state with $\alpha 5$ present (JD-GF- $\alpha 5$ construct, Figure S1). This construct can be considered as a globular protein with a large, disordered linker (Figure 2A) and thus a global correlation time for all residues is appropriate and does not need to be assumed. As expected, the J-domain in the context of JD-GF- $\alpha 5$ behaves similarly to JD-GF (Figures S1 and S3) with helix 5 residues also showing reduced motions consistent with a highly ordered region. GF residues still require an extended model-free formalism to explain their dynamics and show a characteristic drop-off and rise in their S^2 values (Figure S3F) with a correlation time for the internal motions of around 1.2 ns. Notably, residues K70 – N74 are significantly more dynamic in the closed state showing an S^2 of 0.48-0.75. Taken together, analysis of the NMR relaxation data suggests that hydrophobic GF residues show reduced dynamics in JD-GF, presumably due to interactions with the J-domain.

CG-simulations do not capture the details of the interaction between JD and GF. To structurally visualize the large-scale motions of GF we were interested in generating a conformational ensemble that captures the key properties of the mixed-folded JD-GF. In the presence of experimental data that report on long-range interactions, such as RDCs and paramagnetic relaxation enhancement (PRE) data, such ensembles have been traditionally constructed by fitting the experimental data using an ensemble approach.³⁰ However, in the JD-GF- $\alpha 5$ strong solvent PREs are observed for GF linker residues suggesting a preferential interaction of the PRE tag with this region of the protein (Figure S4). Thus, we specifically avoided collecting intramolecular PRE data on JD-GF since the highly hydrophobic PRE tags would likely interfere with the hydrophobic interactions that seem most likely to modulate GF motions. In theory, specific interproton NOEs between JD and GF resonances could also guide the generation of an ensemble but they require a relatively stable interaction and therefore were only observed for N- but not the C-terminus of GF (Figure S2A). Inspired by the success of CG MD simulations for disordered proteins, we decided to take a similar approach to visualise the GF conformational ensemble. Although full-atom MD simulations are possible for systems of this size,



they still require significant resources in order to achieve timescales relevant for IDRs. On the other hand, CG simulations allow sampling of the conformational ensemble of systems that vary in size in a straightforward manner. The NMR relaxation data in Figure 1 suggested that residues K70 - N74 are as rigid as those in the JD and therefore, these were fixed to their native conformations in the simulations. We tested the two most popular force-fields for IDPs, Martini3IDP which uses a multiple-bead representation and a sophisticated energy function, and CALVADOS that is fully coarse-grained with a single bead per-amino acid (Figures S5, S6). Martini3 that was initially developed for folded proteins³¹ samples significantly more compacted conformations (R_g 1.47 ± 0.02 nm, Figure 2B) than those observed experimentally by SAXS (1.72 ± 0.2 nm) even with the IDP modification in its energy function (Figures 2B, S5, S6).²⁰ On the other hand, CALVADOS which relies on an empirically determined energy function but was trained directly on IDPs captures the overall compaction of JD-GF, producing an average R_g value of 1.65 nm in close agreement with SAXS (Figures 2A, 2B). As seen in Figure S7A, Martini3IDP predicts some relatively prominent long-range contacts between the back or side of the JD and GF resulting in a compact ensemble (Figure 2B). This is in contrast with the CALVADOS simulations that produce various long-range contacts of the GF with the front and back faces of the J-domain, and also significant contacts of the GF with itself (Figure 2C). Experimentally, by comparing the chemical shifts of JD alone versus those of JD-GF, we have previously shown that the GF-linker is involved in long range interactions with the front of the J-domain (Figure S8A).¹² To gain a more quantitative description of the long-range interactions in the simulations, we calculated contact order values for each residue, a metric that emphasizes the long-range versus the short-range interactions that each residue is involved in (see Methods).²⁴ On top of any interdomain contacts, the observed chemical shift perturbations between the JD and JD-GF constructs contain contributions from short-range effects (such as the truncation of the protein), and possible allosteric effects. However, the latter are not expected to be significant for this system with the exception of residues 1 – 15 that show CSPs arising from the addition of any linker to helix 4 (Figure S8B). Many of the residues in this area, including, Y4, V7, L8, V10 and A14 are not surface exposed (Figure S8B), suggesting that the CSPs originate from subtle structural reorganisation of the JD domain mediated by residues in helix 4 (especially Y65) that make significant contacts with residues 1 to 15 in JD. Short-range effects can be accounted for by excluding any CSPs arising from residues in close proximity to the truncation site, resulting in a CSP dataset that reflects almost purely the long-range JD – GF interactions and can be directly compared to the contact order values calculated from the simulations. As seen in Figure 2D, both Martini3IDP and CALVADOS simulations cannot reproduce the long-range interactions observed by NMR. Overall, CALVADOS is outperforming Martini3 in sampling relevant, extended conformations for JD-GF but it cannot capture the details of interdomain communication between the folded JD and disordered GF.



Expanding CG-simulations to capture local structure. The inability of the CALVADOS forcefield to generate accurate interdomain contacts is perhaps not surprising as it was trained using experimental SAXS and PRE data that only report on the global compaction of the system.⁵ Given the good agreement with SAXS and the versatility of the fully coarse-grained CALVADOS simulations we set out increase their ability to produce correct residue-residue contacts. To this end, we hypothesized that including new terms in the CG simulation that describe local backbone properties would allow us to generate a more accurate ensemble. The backbone chemical shifts of the GF-linker in the open state are not fully random coil but point towards a significant propensity for more extended, β -strand-like conformations.¹² CALVADOS simulations fail to capture these local structural propensities since the underlying energy function does not contain any terms that bias the backbone towards specific conformations (Figure S9). Capturing secondary structure propensities in single-bead CG-simulations is not straightforward as multiple beads are normally required in order to calculate backbone dihedral angles. Efforts to overcome this limitation of single bead CG-simulations have been presented before by Rizuan et al.²² who have used a double well energy potential that allows transition between an α -helical energy well to that of an extended conformation by tuning a residue specific parameter (ϵ_d). If ϵ_d is set to match the known propensities for secondary structure of the various amino acids, this approach can capture the α -helical propensities for a number of IDPs.²² However, it is often the case that secondary structure propensities in IDRs cannot be accounted for just purely by their sequence, but are a consequence of the structural intricacies of the system as is the case for JD-GF. Therefore, we decided to tune ϵ_d values to match the NMR-determined secondary structure propensity of the GF-linker (Table S2) calculated from its previously determined backbone chemical shifts. In practice this approach could be useful for the plethora of IDRs that have their backbone chemical shifts determined by NMR. Including the backbone dihedral angle term with the optimized ϵ_d values in CALVADOS alongside a double well angle term²³ forces GF to populate an extended conformation in its C-terminus, in good agreement with the experimental data (Figure 3A). When compared to the ensemble that was generated without the use of the backbone potentials the calculated R_g is not significantly smaller (1.63 versus 1.65 nm, 2% difference) (Figures 2B, 3B) showing that the new terms do not significantly impact the global compaction of the ensemble. As we have shown previously,¹² the main point of contact between JD and GF is the nearby helix 4. These short-range interactions are also captured by the new simulation that shows good agreement with the experimentally determined chemical shift perturbations (Figure 3C). Overall, the simulations can now capture both the global compaction, the local secondary structure propensities and short-range interactions of GF.

Hydrophobic GF residues promote partially closed states. We then asked whether the new, enhanced simulations can perform better in terms of generating more accurate long-range interdomain contacts between JD and GF. As seen from the energy maps of Figure 4A when the backbone angle potentials are included the main interdomain interactions involve exposed hydrophobic residues facing



the front of the J-domain (helices 2 and 3) in good agreement with small but significant chemical shift perturbations in the same region caused by GF (Figures 4B, S8). The new simulations show a good agreement between the calculated contact order values for residues in helices 2 and 3 that all have their side-chains facing the front of the J-domain and the long-range CSPs between JD and JF (Figure 4C) suggesting that the introduced potential terms are able to define the correct interdomain contacts by modulating the local conformation of GF residues. Indeed, in the absence of the dihedral and angle potentials, GF partially collapses on itself and can form interactions with the nearby regions of the J-domain (Figure 2C). However, when the GF backbone is forced to adopt an extended conformation in order to satisfy the new backbone potential terms, it can now reach further onto the front of the J-domain. Principal component analysis shows that the compaction of GF onto the front of helices 2 and 3 is the main component of the trajectory, accounting for ~ 37% of the total motion (Figures 4D, 4E). These motions of the GF relative to the J-domain create states that largely resemble the closed/autoinhibited state even under nominally open conditions, in the absence of helix 5 (Figure 2A) rationalizing the reduced affinity for Hsc70 of the open JD-GF (~ 400 μ M) in comparison to JD alone (~3 μ M).¹² Increased contact order values are also observed for several GF residues (Figure 4C), most of which coincide with residues that show reduced motions in the ps – ns timescales as shown by spectral density mapping of NMR data in Figure 1. It is thus intriguing to attribute the reduced motions of F87, E88, F89, F91, F93 and R94 to long-range interactions with the J-domain observed in the CG-simulations. Martini3IDP simulations can be used to estimate the lifetime of the contacts formed as shown in Figure S7B. Although the intra-GF contacts are fast (< 100 ps), some of the long-range interdomain contacts involving residues F87, E88, F89, F91, F93 are slower, on the ns timescale (Figure S7B), as predicted by the model-free analysis in Figure S1.

One of the most unexpected results of the analysis of the relaxation data of Figure 1 was the apparent rigidification of residues K70 – N74 in the open state in comparison to the closed one. To investigate the functional role of this collapse we removed the harmonic restraints for these residues to allow them to move freely as the rest of the GF. These new simulations show a significant reduction of the interactions between the linker and the front face of the JD (Figure S10A) suggesting that the conformation of the N-terminal part of the linker is important in positioning the GF ensemble to the front of the J-domain. This process places the hydrophobic GF residues in a position to interact with the exposed hydrophobic residues in JD and could be important in reestablishing autoinhibition after it gets released.

In order to validate the approach and the role of hydrophobic residues in the long-range interdomain interactions, instead of performing multiple single point mutations on JD-GF we decided to swap GF with a fully disordered Glycine-Serine-Serine (GSS) linker of the same length (JD-GSS construct) as a negative control. As seen in Figure S10B JD-GSS shows an almost complete loss of long-range interdomain interactions in the CG -simulations. This is precisely what we observe experimentally by NMR where the chemical shift perturbations caused on JD resonances by

View Article Online
DOI: 10.1039/D6CP01833E



interdomain interactions with GF have almost entirely disappeared in the presence of GSS (Figures S8A, 4B).

Conclusions

Here, we visualize the transient interdomain contacts between a folded domain and an IDR by integrating readily available NMR data that report on local structural features into minimal CG simulations. This NMR-driven approach is used to generate accurate interdomain contact maps for the mixed-folded JD-GF construct that lead to a significant decrease in affinity for Hsc70. Analysis of NMR relaxation data revealed reduced dynamics for specific hydrophobic residues at both the N- and C-termini of the disordered GF linker, indicating sequence-dependent interactions with the folded J-domain. By tuning residue-specific backbone dihedral parameters (ϵ_d) to match secondary-structure propensities derived from NMR chemical shifts, accurate interdomain interaction maps between the folded domain and the low-complexity GF region can be obtained. In the absence of explicit backbone dihedral/angle terms, hydrophobic residues within the IDR preferentially collapse onto themselves or nearby sequence segments, preventing access to their native interaction interfaces on the J-domain. Enforcing experimentally determined backbone propensities promotes more extended conformations, thereby enabling these residues to engage hydrophobic surface patches that are otherwise inaccessible computationally. The resulting conformational ensembles rationalize the linker dynamics observed by NMR relaxation and reveal that the nominally open JD-GF construct samples partially closed states which resemble the autoinhibited conformation of JD-GF- α 5.

Since backbone chemical shifts are typically among the first NMR observables obtained for intrinsically disordered regions, this strategy provides an experimentally accessible route to improving CG simulations of disordered proteins and mixed folded-disordered systems. The combination of NMR-driven restraints directly³² or indirectly (this study) into fully coarse-grained simulations is particularly attractive as it allows the generation of physically-relevant ensembles for challenging systems in a matter of minutes. We note that other force fields such as PRIMO³³, UNRES³⁴, HyRes³⁵ and AWSEM³⁶ have been designed to capture transient secondary structures in IDPs. These models achieve this through either increased backbone resolution (PRIMO, HyRes) or knowledge-based/statistical potentials (UNRES, AWSEM). In contrast, our approach introduces an explicitly tunable multiple well dihedral and angle potentials that enable direct incorporation of experimentally derived backbone conformational preferences as determined by NMR. The fact that backbone propensities can be encoded in the most minimal coarse-grained representation (single backbone bead) with the lowest computational cost allows for application of our method on systems of increasing size and complexity. For instance, regulatory IDRs in signaling proteins, transcription factors, and phase-separating or amyloid proteins are likely to exhibit similar sensitivity to local backbone properties that would be difficult to capture in full-atom simulations due to their increased size. Thus, improved CG



models that can reproduce both global and local structural features of the ensemble may broadly improve the accuracy of predicted interdomain interaction maps.

View Article Online
DOI:10.1039/D6CP01833E

Data availability

The dihedral potential is freely available as an OpenMM plugin at <https://github.com/karamanoslab/OpenMMDihedralPlugin> and the raw data for the study are available at <https://figshare.com/s/55a6459b127d40d02f2c>.

Acknowledgments

We thank Kresten Lindorff-Larsen and Gulio Tesei for helpful discussions. This work was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 223268/Z/21/Z) to TKK. Access to the 700 MHz spectrometer was provided by the MRC Biomedical NMR Centre at the Francis Crick Institute, which receives core funding from Cancer Research UK Grant FC001029; Medical Research Council Grant FC001029; and Wellcome Trust Grant FC001029.

Figures



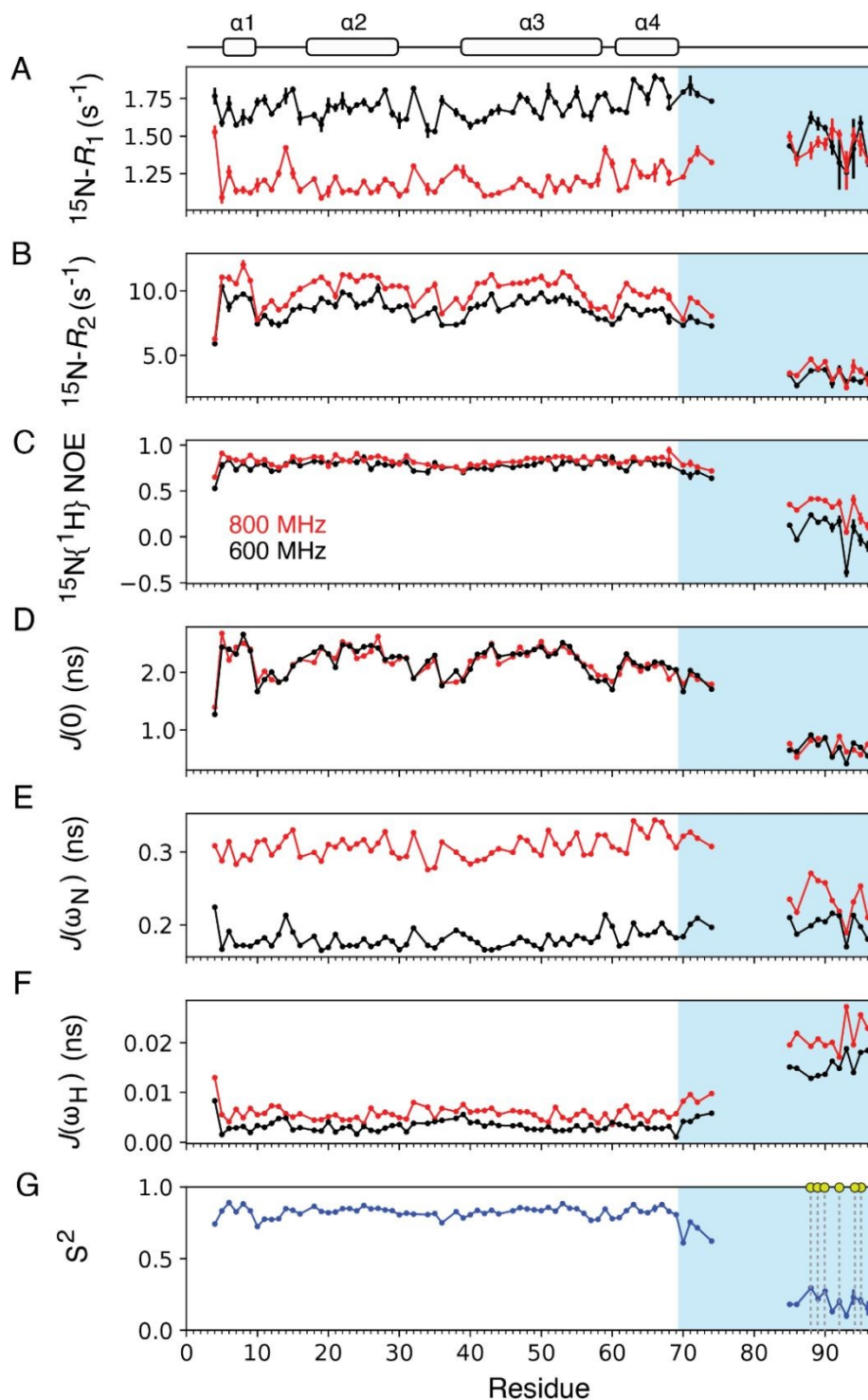


Figure 1: Backbone dynamics of the open JD-GF state. ^{15}N R_1 (A), R_2 (B) and $^{15}\text{N}\{^1\text{H}\}$ NOE data (C) collected at 600 (black dots) or 800 (red dots) MHz. $J(0)$ (D), $J(\omega_{\text{N}})$ (E) and $J(\omega_{\text{H}})$ (F) values are derived from reduced spectra density mapping. Calculated order parameter squared S^2 from $J(0)$ and $J(\omega_{\text{N}})$ at 800 MHz (eq. 2) are shown in G. GF residues are shown as a cyan box and residues with $S^2 > 0.3$ are highlighted with dashed lines/yellow dots.



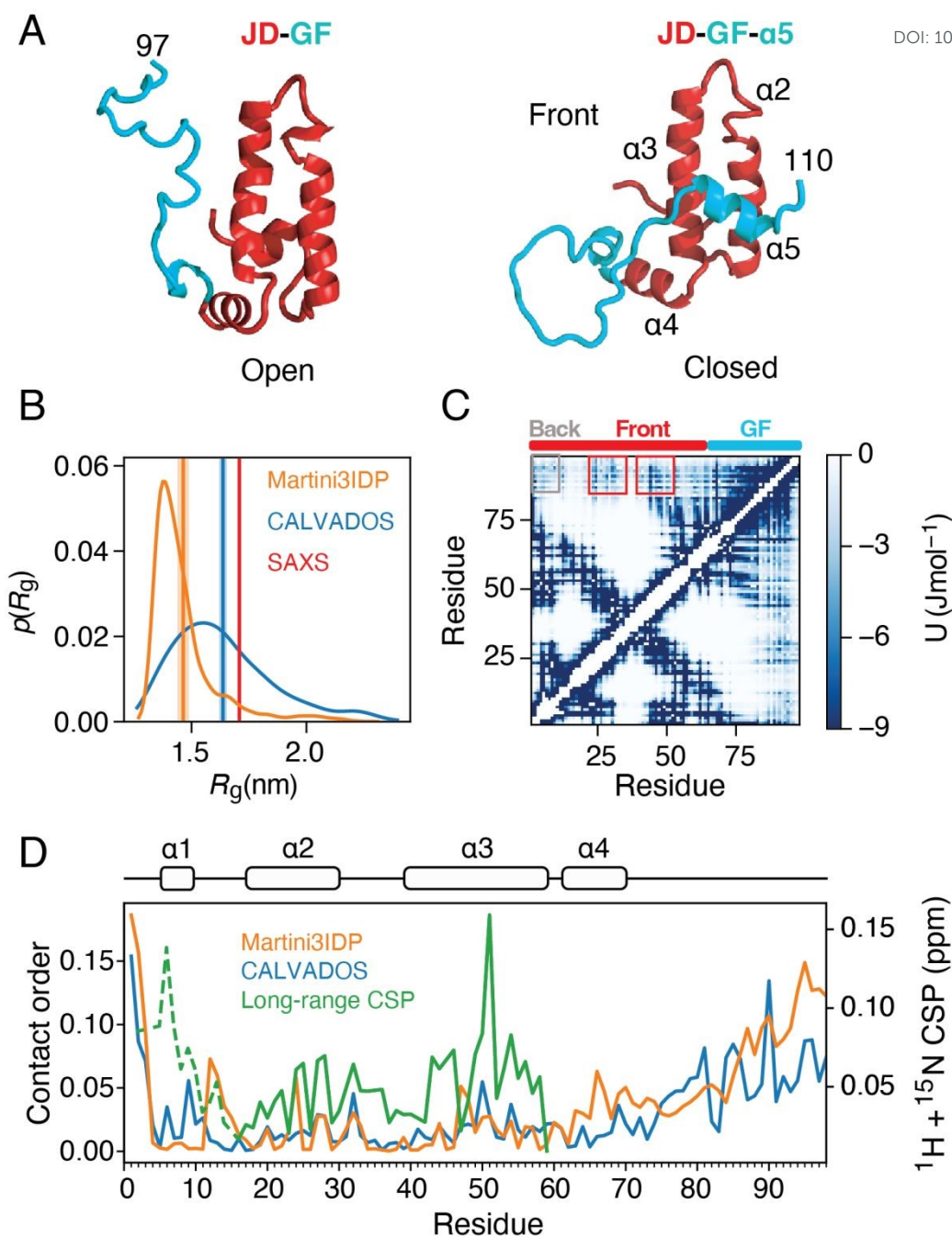


Figure 2: Typical CG-simulations do not reproduce long-range interactions. (A) Structures of the JD-GF (left) and JD-GF- $\alpha 5$ (right) (JD in red, GF in cyan). (B) Distribution of the radius of gyration from the Martini3IDP (orange) and CALVADOS (cyan) simulations with the average value and the SAXS determined value indicated as a cyan/red vertical line respectively. (C) Interaction energy maps of the Ashbaugh-Hatch term in the CALVADOS simulations. Grey/red boxes highlight the back (residues 1-15)/or the front (residues 25-30 and 45-50) of the JD. (D) Contact order values that emphasize long-range interactions for Martini3IDP (orange) and CALVADOS (cyan) overlaid with the observed long-range CSPs between JD-GF and JD alone (green) (used as a proxy for interdomain contacts, see Methods). CSPs for residues 1 – 15 that do not report on long-range interactions are shown as a dashed line.



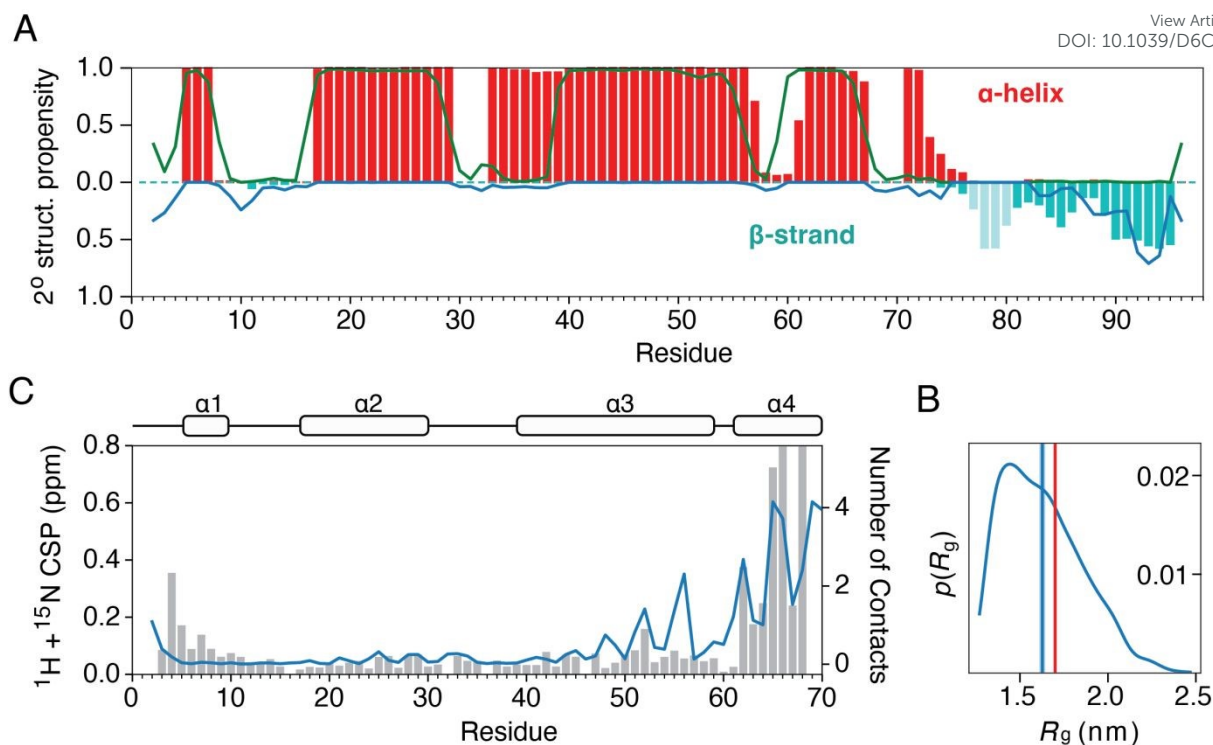
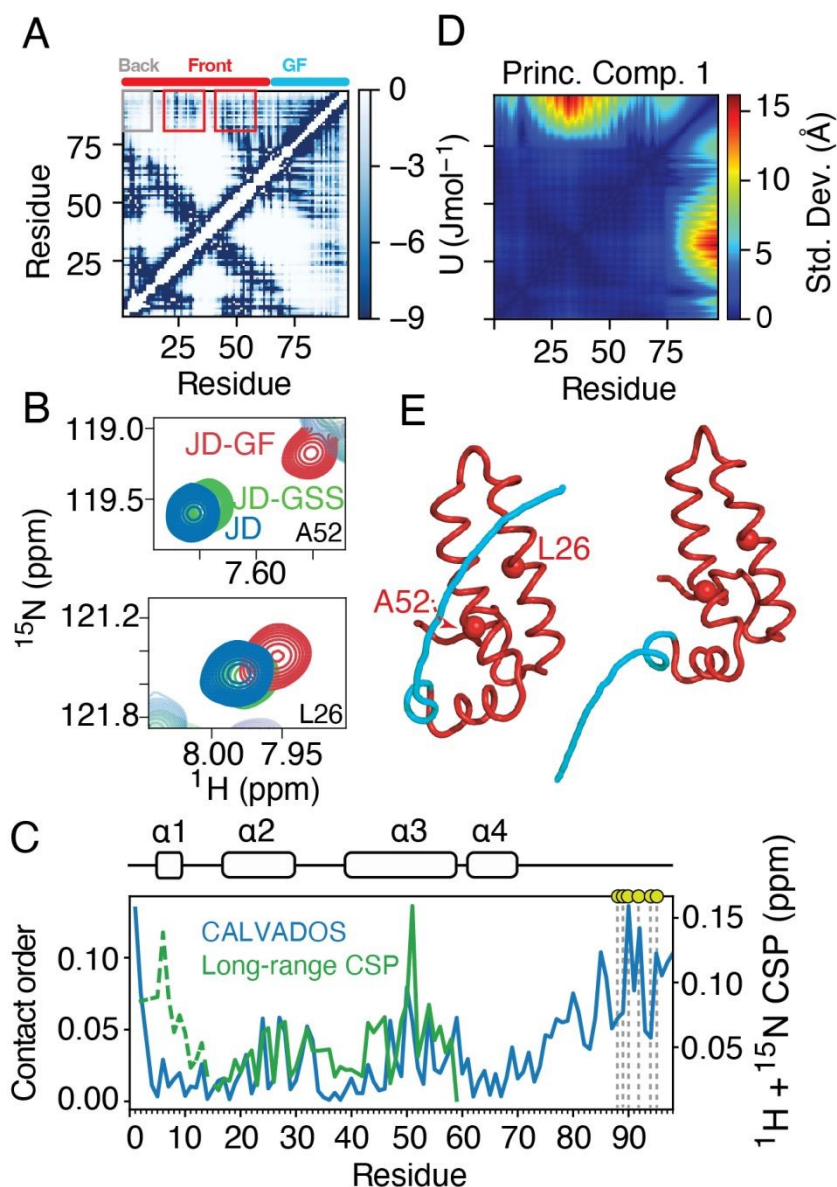


Figure 3: Agreement of the enhanced CALVADOS simulations with experimental data. (A) Secondary structure propensities of JD-GF calculated in the simulation are shown as bars (α -helix, red/ β -strand, cyan). Talos-derived secondary structure based on the JD-GF backbone shifts are shown as lines (α -helix, green/ β -strand, blue). The JD was restrained as a rigid body and backbone dihedral angles were only applied to the GF linker (residues 75 – 98) (see Methods). Experimental data for residues 75 - 80 are missing due to signal overlap of the five consecutive glycine residues. (B) Distribution of the radius of gyration from the simulations with the average value and the SAXS determined value indicated as a cyan/red vertical line respectively. (C) Combined ^1H , ^{15}N chemical shift perturbations for J-domain resonances in a JD alone versus a JD-GF construct. The number of contacts (<10 Å) each JD residue makes with the GF in the simulation is shown as a cyan line.





View Article Online
DOI: 10.1039/D6CP01833E

Figure 4: Analysis of the conformational ensemble with the dihedral term included in CALVADOS simulations. (A) Interaction energy maps of the Ashbaugh-Hatch term. Grey/red boxes highlight the back (residues 1-15)/or the front (residues 25 -30) of the JD. (B) Regions of the ^1H - ^{15}N of JD-GF (red), JD alone (blue) or JD-GSS (green, negative control). (C) Contact order values (cyan) that emphasize long-range interactions overlaid with the observed long-range CSPs between JD-GF and JD alone (green). CSPs for residues 1 – 15 that do not report on long-range interactions are shown as a dashed line. Yellow spheres and dashed lines represent the GF residues with reduced motions identified by spectral density mapping of the NMR relaxation data (Figure 1). (D) Principal component analysis of the resulting trajectory. The principal component vectors are shown as a contact map depicting the change of each distance in relation to all other distances. Principal component 1 describes the compaction of the GF-linker onto the front face of the JD in the open state as shown in (E). L26 and A52 are shown as red spheres.



References

1. Ibanez de Opakua, A. et al. Molecular interactions of FG nucleoporin repeats at high resolution. *Nat. Chem.* **14**, 1278-1285 (2022).
2. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123-33 (2015).
3. Borgia, A. et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61-66 (2018).
4. Tesei, G. & Lindorff-Larsen, K. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Res. Eur.* **2**, 94 (2022).
5. Tesei, G., Schulze, T.K., Crehuet, R. & Lindorff-Larsen, K. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U S A* **118**(2021).
6. Cao, F., von Bulow, S., Tesei, G. & Lindorff-Larsen, K. A coarse-grained model for disordered and multi-domain proteins. *Protein Sci.* **33**, e5172 (2024).
7. Thomassen, F.E. et al. Rescaling protein-protein interactions improves Martini 3 for flexible proteins in solution. *Nat. Commun.* **15**, 6645 (2024).
8. Kampinga, H.H. & Craig, E.A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.* **11**, 579 EP-592 (2010).
9. Kakkar, V. et al. The S/T-Rich Motif in the DNAJB6 chaperone delays polyglutamine aggregation and the onset of disease in a mouse model. *Mol. Cell* **62**, 272-283 (2016).
10. Karamanos, T.K., Tugarinov, V. & Clore, G.M. Unraveling the structure and dynamics of the human DNAJB6b chaperone by NMR reveals insights into Hsp40-mediated proteostasis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21529-21538 (2019).
11. Abayev-Avraham, M., Salzberg, Y., Gliksberg, D., Oren-Suissa, M. & Rosenzweig, R. DNAJB6 mutants display toxic gain of function through unregulated interaction with Hsp70 chaperones. *Nat. Commun.* **14**, 7066 (2023).
12. Hobbs, B. et al. A low-complexity linker as a driver of intra- and intermolecular interactions in DNAJB chaperones. *Nat. Commun.* **16**, 5070 (2025).
13. Fitzkee, N.C. & Bax, A. Facile measurement of ^1H - ^{15}N residual dipolar couplings in larger perdeuterated proteins. *J. Biomol. NMR* **48**, 65-70 (2010).
14. Farrow, N.A., Zhang, O., Szabo, A., Torchia, D.A. & Kay, L.E. Spectral density function mapping using ^{15}N relaxation data exclusively. *J. Biomol. NMR* **6**, 153-62 (1995).
15. Lefevre, J.F., Dayie, K.T., Peng, J.W. & Wagner, G. Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochemistry* **35**, 2674-86 (1996).



16. Bracken, C., Carr, P.A., Cavanagh, J. & Palmer, A.G., 3rd. Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol.* **285**, 2133-46 (1999). View Article Online
DOI: 10.1039/D6CP01833E
17. Gill, M.L., Byrd, R.A. & Palmer, A.G., III. Dynamics of GCN4 facilitate DNA interaction: a model-free analysis of an intrinsically disordered region. *Phys. Chem. Chem. Phys.* **18**, 5839-49 (2016).
18. Mandel, A.M., Akke, M. & Palmer, A.G., 3rd. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* **246**, 144-63 (1995).
19. Clore, G.M. et al. Deviations from the simple 2-parameter model-free approach to the interpretation of ¹⁵N-nuclear magnetic-relaxation of proteins. *J. Am. Chem. Soc.* **112**, 4989-4991 (1990).
20. Wang, L., Brasnett, C., Borges-Araujo, L., Souza, P.C.T. & Marrink, S.J. Martini3-IDP: improved Martini 3 force field for disordered proteins. *Nat. Commun.* **16**, 2874 (2025).
21. Kroon, P.C. et al. Martinize2 and Vermouth provide a unified framework for molecular topology generation. *Elife* **12**(2025).
22. Rizuan, A., Jovic, N., Phan, T.M., Kim, Y.C. & Mittal, J. Developing bonded potentials for a coarse-grained model of intrinsically disordered proteins. *J. Chem. Inf. Model.* **62**, 4474-4485 (2022).
23. Kim, Y.C. & Hummer, G. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.* **375**, 1416-33 (2008).
24. Plaxco, K.W., Simons, K.T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985-994 (1998).
25. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B. & Beckstein, O. MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319-27 (2011).
26. Galvanetto, N. et al. Extreme dynamics in a biomolecular condensate. *Nature* **619**, 876+ (2023).
27. Mayo, K.H., Daragan, V.A., Idiyatullin, D. & Nsmelova, I. Peptide internal motions on nanosecond time scale derived from direct fitting of ¹³C and ¹⁵N NMR spectral density functions. *J. Magn. Reson.* **146**, 188-95 (2000).
28. 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.* **104**, 4546-4559 (1982).
29. Tjandra, N., Feller, S.E., Pastor, R.W. & Bax, A. Rotational diffusion anisotropy of human ubiquitin from ¹⁵N NMR relaxation. *J. Am. Chem. Soc.* **117**, 12562-12566 (1995).



30. Clore, G.M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* **109**, 4108-39 (2009). View Article Online
DOI: 10.1039/D6CP01833E
31. Souza, P.C.T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382-388 (2021).
32. Cullen, M. et al. Integrating NMR restraints into coarse-grained simulations: toward accurate conformational ensembles of complex protein systems. *J. Am. Chem. Soc.* **148**, 13160-13173 (2026).
33. Kar, P., Gopal, S.M., Cheng, Y.-M., Predeus, A. & Feig, M. PRIMO: A transferable coarse-grained force field for proteins. *J. Chem. Theory Comput.* **9**, 3769-3788 (2013).
34. Krupa, P. et al. Prediction of protein structure by template-based modeling combined with the UNRES force field. *J. Chem. Inf. Model.* **55**, 1271-1281 (2015).
35. Liu, X. & Chen, J. HyRes: a coarse-grained model for multi-scale enhanced sampling of disordered protein conformations. *Phys. Chem. Chem. Phys.* **19**, 32421-32432 (2017).
36. Davtyan, A. et al. AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **116**, 8494-8503 (2012).



Data availabilityView Article Online
DOI: 10.1039/D6CP01833E

The dihedral potential is freely available as an OpenMM plugin at <https://github.com/karamanoslab/OpenMMDihedralPlugin> and the raw data for the study are available at <https://figshare.com/s/55a6459b127d40d02f2c>.

