Accurate conformational ensembles of mixed folded proteins from NMR-guided simulations

Billy Hobbs[1‡], Noor Limmer[1‡], Gwendolyn L. Clenshaw[1], Felipe Ossa[1], & Theodoros K. Karamanos[1*]

[1]Department 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 often cooperate with folded domains to mediate a variety of protein-protein interactions. To visualize these mixed folded proteins, experimentally guided molecular dynamics simulations are employed but may struggle to simultaneously capture the properties of disordered and folded domains or to generate accurate inter-domain interactions. Here we use solution NMR combined with coarse-grained (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 dihedral angle parameters that allow CG-simulations to capture the local secondary structure propensities of the linker in coarse-grained simulations, we obtain accurate interdomain contact maps. In agreement with model-free analysis of NMR relaxation data, the resulting ensembles show that even in the open state the linker experiences motions that resemble those of the closed state driven by hydrophobic residues in GF.

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 later or by acting as hotspots for protein-protein interactions.[2] Computational methods have provided crucial insights into the conformational ensembles of intrinsically disordered regions (IDRs), often guided by experimental data.[3] Of particular interest are coarse grained (CG) molecular dynamics (MD) simulations that overcome the requirement for ms-long simulations and have been employed with significant success.[4, 5] However, these CG-based force-fields often struggle to describe the cooperation of folded domains with IDRs.[6]

J-domain chaperones contain a typical, low complexity, glycine and phenylalanine-rich (GF) linker that connects their J-domain with the C-terminal, substrate binding domain and is crucial for cell viability.[7] In DNAJB6, a highly potent aggregation inhibitor[8], part of the GF forms a stable α-helix (helix 5 or α5) that blocks the binding of Hsp70 to the J-domain creating a closed/autoinhibited state.[9] Even though a number of publications investigating the closed/autoinhibited state of DNAJBs are now available,[10, 11] the characterization of the open/uninhibited state in which the GF-linker could serve as a pseudo-substrate for Hsp70[11], is lacking. 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.[11] However, an atomic level description of the GF – JD interactions and their role in autoinhibition and/or Hsp70 binding could not be established. Here, using solution NMR and CG-simulations we investigate the conformational ensemble sampled by the highly flexible JD-GF construct. By introducing and tuning potential terms that capture the secondary structure of the linker we generate ensembles that are in good agreement with NMR data. The ensembles reveal that the open JD-GF visits conformations reminiscent of the closed state through extensive contacts of specific GF residues with the J-domain.

The dynamics of DNAJB6 constructs were probed by R_1, R_2 and NOE relaxation data at two fields and analysed by model-free analysis.[12] Conceptually, a model-free analysis of the open JD-GF construct is challenging due to the absence of a single tumbling time (τ_c) for residues in GF.[13] Indeed, for helices 1 to 4 a simple model that takes into account the slight anisotropy of the JD diffusion tensor and only includes a single ordered 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). Not surprisingly, these observations also hold for the J-domain in the context of the closed state in the presence of α5 (JD-GF-α5 construct, Figure S1). 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.[14] 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. In comparison, in the closed state the same residues show an S_s^2 of 0.48-0.75. Moving towards

the C-terminal part of GF, S2 values drop drastically as expected for an IDR, except for residues F87, E88 (S^2 ~ 0.50) and to a smaller extent F89, F91 and R94 (S^2 ~ 0.35) which show reduced motions in comparison to other GF residues. The average 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 (τ_e ~ 100 ps). We interpret these motions as transient, large-scale interactions formed by GF residues 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 for $^1D_{NH}$ residual dipolar coupling (RDC) data (Figure S2) that show non-zero values for GF residues.

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, open JD-GF. Inspired by the success of the CALVADOS force field5 we decided to take a similar, CG-based approach by fine-tuning certain force field parameters to match the available NMR data. CALVADOS, where each residue is represented by a single bead, captures the overall compaction of JD-GF, producing an average R_g value of 1.65 nm in close agreement with the SAXS determined value of 1.73 nm (Figures 2A, 2B) without the need of any reweighting. By comparing the chemical shifts of JD alone versus those of JD-GF we have previously shown that the GF-linker interacts with the J-domain.11 These interactions are also captured by the CALVADOS simulation that shows good agreement with the experimentally determined chemical shift perturbations (Figure 2C). 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. 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. Efforts to overcome this limitation of single bead CG-simulations have been presented before by Rizuan et al.[15] 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.[15] Here we tune ε_d values to match the NMR-determined secondary structure propensity of the GF-linker (Table S1), to force it to populate an extended conformation in its C-terminus, in good agreement with the experimental data (Figure 3D). When compared to an ensemble that was generated without the use of the backbone dihedral angle potential, the extended nature of the GF linker is not captured as expected, while the calculated R_g is only 2% smaller (1.63 nm) (Figures S3A, S3B). Overall, the ensemble can now capture the compaction, long-range interactions between JD and GF, and the local GF secondary structure.

It is evident from the energy maps in Figure 3A that the GF is involved in numerous interactions either with itself or with the J-domain. In the absence of the backbone dihedral angle term, self-interactions between GF residues are more prevalent, with weak interactions involving the back of the J-domain (helix 1) also present (Figure S3C, S3D). However, when the dihedral angle potential is 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 3B, S4). Thus, the dihedral angle potential is able to define the correct interdomain contacts by modulating the local conformation of GF residues. 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 (Figure 3C, 3D). These motions largely resemble the closed/autoinhibited state and can rationalise the reduced affinity for Hsc70 of the open JD-GF in comparison to JD alone. Analysis of the contact order in the simulations that can discriminate between local and large-scale motions, shows spikes for several GF residues, most of which coincide with residues that show reduced motions in the ps – ns timescales as shown by Model-free analysis in Figure 1 (Figure 3E). It is thus intriguing to attribute the reduced motions of F87, E88, F89, F91 and R94 to long-range interactions with the J-domain as shown by the CG-simulations.

The model-free analysis 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. Removing the harmonic restraints for these residues results in a significant reduction of the interactions between the linker and the front face of the JD (Figure S5A) suggesting that the conformation of the N-terminal part of the linker is important in positioning the GF ensemble in relation to the J-domain. To further validate the observed JD-GF contacts, the GF was swapped with a fully disordered GSS linker of the same length (JD-GSS), resulting in complete loss of the interactions with JD both experimentally (Figures 3B, S4A) and computationally (Figure S5B).

Our previous findings suggested that the GF of another class-B DNAJ chaperone, DNAJB1 behaves differently to that of DNAJB6.[11] To investigate this further, we simulated the equivalent JD-GF constructs of 14 other DNAJs, including all class B members. As seen in Figure 4, the GF of DNAJA1, a class A DNAJ forms a reduced number of contacts with its corresponding J-domain. DNAJA1 cannot adopt a closed/autoinhibited state and therefore it is intriguing to hypothesise that the number of JD – GF contacts correlates with autoinhibition in DNAJs. Indeed JD – GF contacts are more prevalent for most class B DNAJs for which autoinhibition is suggested to be a common feature. DNAJB7 and the endoplasmic DNAJB11 are exceptions to this rule, potentially suggesting that they also lack a closed state. Among the rest of DNAJBs, the number of JD – GF contacts varies, with DNAJB6's GF being the stickier, even compared to that of the canonical DNAJB1, an observation that is also experimentally validated by NMR (Figure S6).

In conclusion, by tuning residue-specific backbone dihedral parameters (ε_d) we were able to obtain accurate interaction maps between a folded domain and an LCR. The ensembles can rationalise linker dynamics as observed by model-free analysis of NMR relaxation data and reveal that the open JD-GF state experiences motions that resemble the closed state.

Data availability

The dihedral potential is freely available as an OpenMM plugin at https://github.com/karamanoslab/OpenMMDihedralPlugin.

Acknowledgments

We thank Kresten Lindroff-Larsen and Gulio Tessei for help-ful 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.

Figures

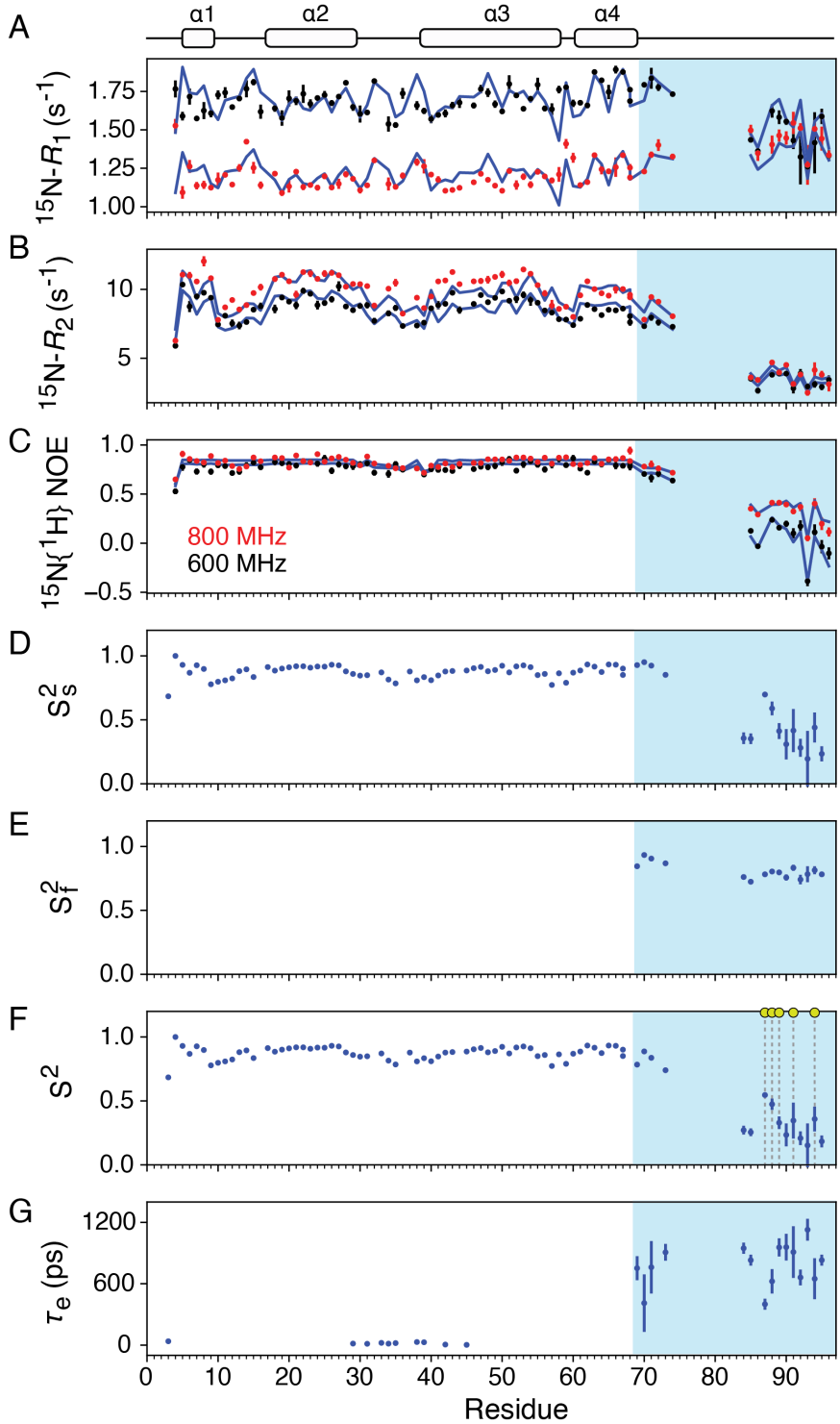

Figure 1: Backbone dynamics of the open JD-GF state. ^{15}N R_1 (A), R_2 (B) and ^{15}N{^{1}H} NOE data (C) collected at 600 (black dots) or 800 (red dots) MHz with model-free predicted values shown as blue lines. Calculated order parameter squared for the slow (S_s^2), fast (S_f^2), the generalised order parameter $S^2 = S_s^2 S_f^2$ and the correlation time for the internal motion (τ_e) are shown in D, E, F and G respectively. When an extended model-free formalism was used τe specifically refers to the slow motion ($\tau_e = \tau_s$). GF residues are shown as a cyan box and residues with $S^2 > 0.3$ are highlighted with dashed lines/yellow dots.

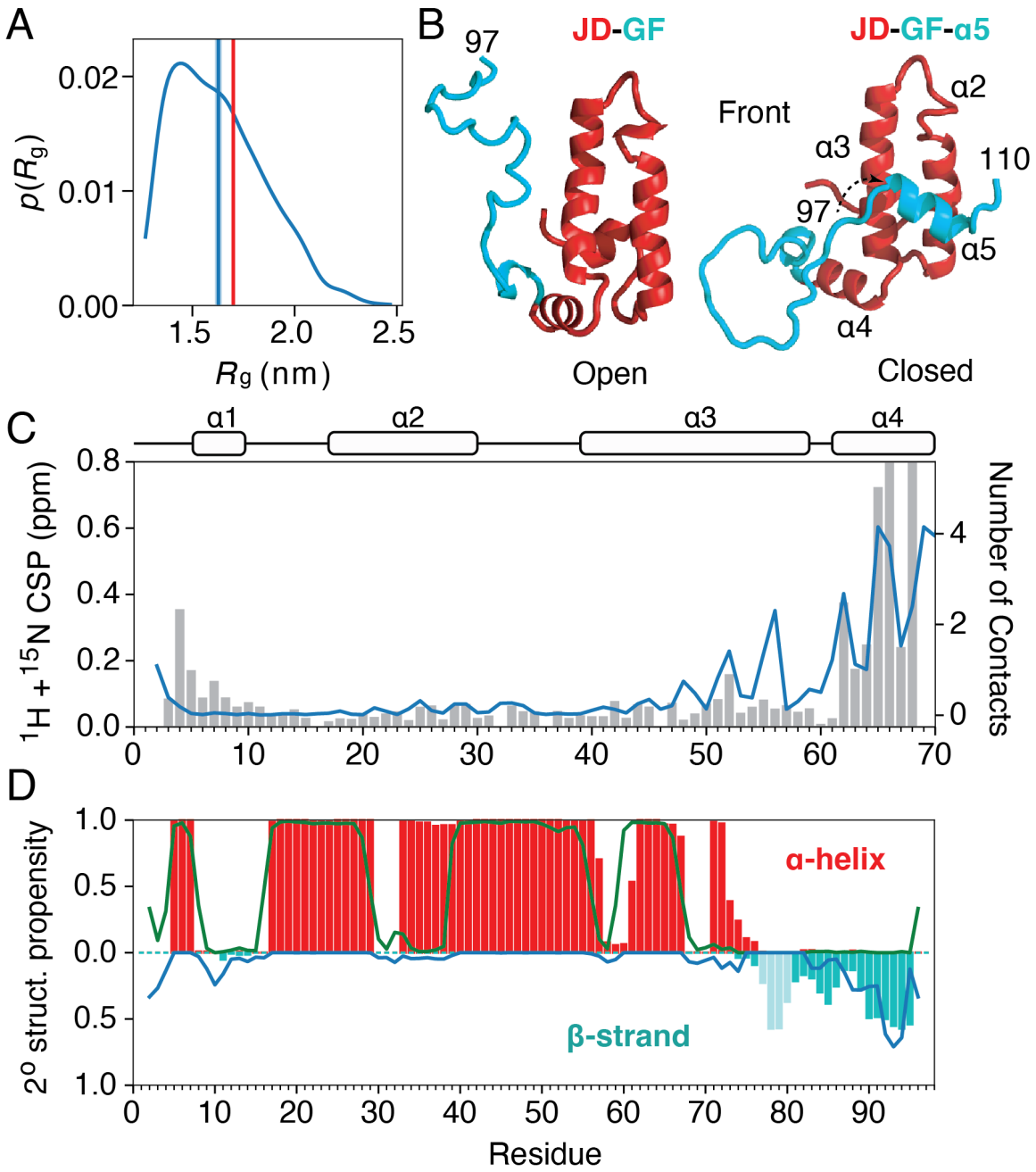

Figure 2: Agreement with experimental data. (A) Distribution of the radius of gyration from the simulations with the average value and the SAXS determined value indicated as a cyan/red vertical line respectively. (B) Structures of the open (left) and closed (right) states (JD in red, GF in cyan). (C) Combined ^{1}H, ^{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. (D) 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 where only applied to the GF linker (residues 75 – 98) (see Methods). Experimental data for residues 75 - 80 are missing due to signal overlap.

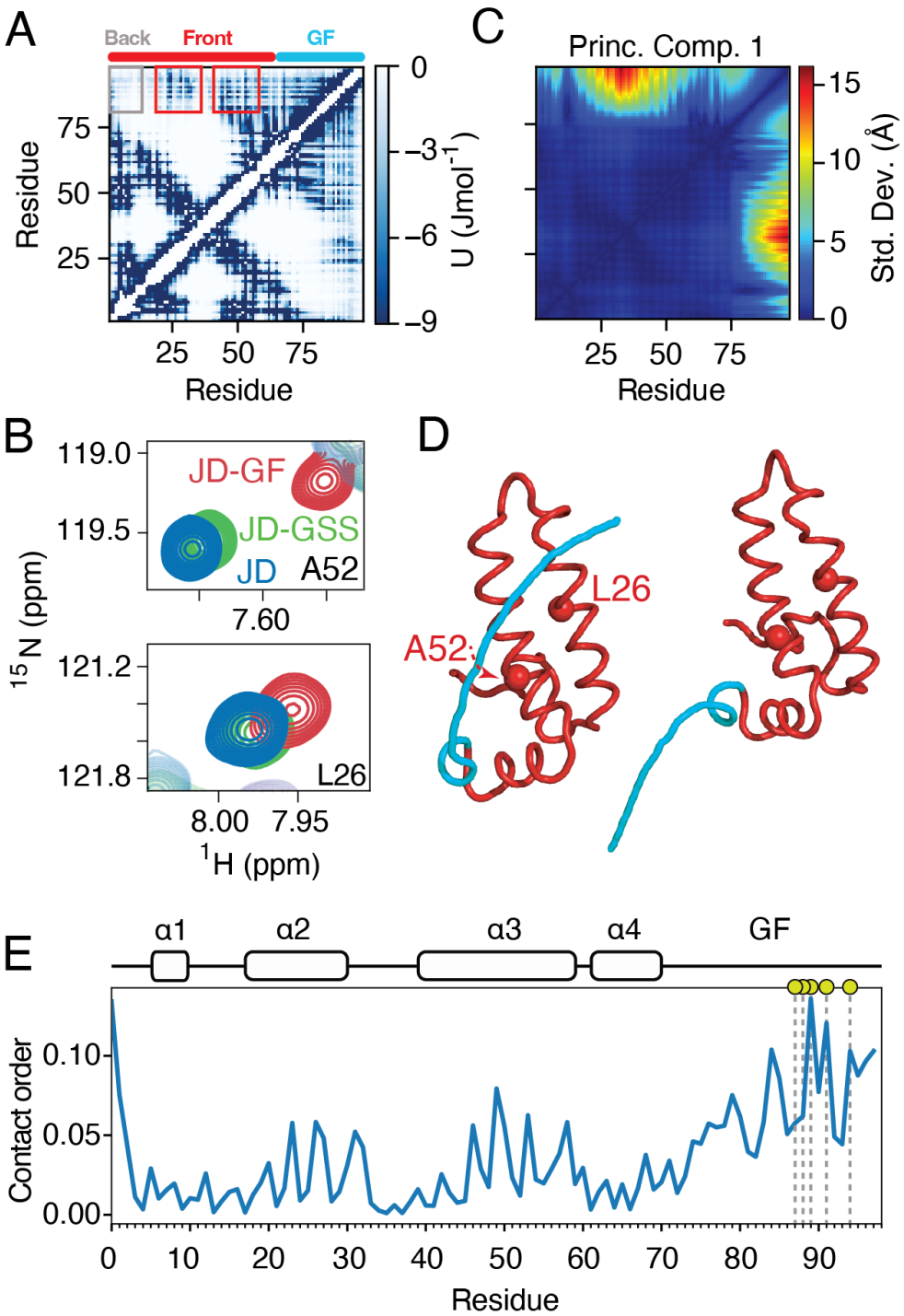

Figure 3: Analysis of the conformational ensemble with the dihedral term included. (A) Energy maps of the Ashbaugh-Hatch term. Grey/red boxes highlight the back (residues 1-15)/or the front (residues 25 -30) of the JD. (B) Regions of the ^{1}H-^{15}N of JD-GF (red), JD alone (blue) or JD-GSS (green). (C) 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 as shown in (D). L26 and A52 are shown as red spheres. (E) Contact order values that emphasise long-range interactions (see Methods). Yellow spheres and dashed lines represent the GF residues with reduced motions identified by model-free analysis (Figure 1).

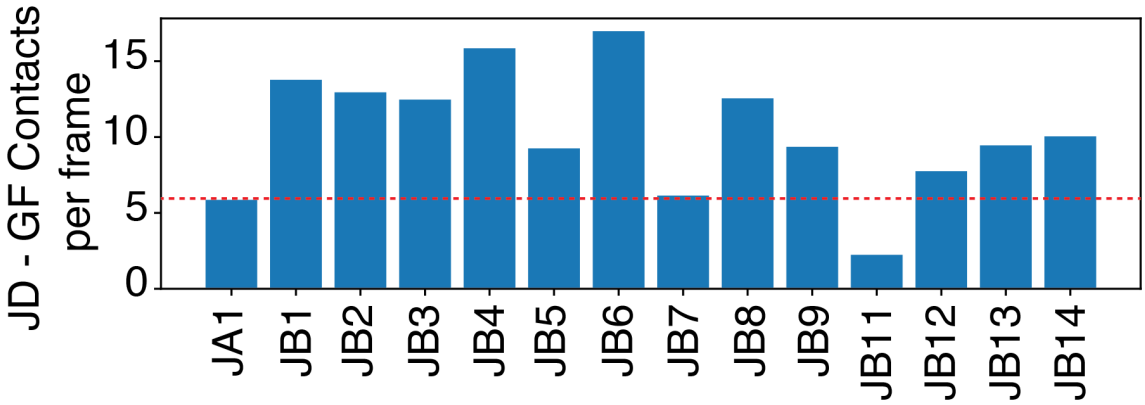

Figure 4: Number of JD – GF contacts per frame for all class-B DNAJs and that of the open DNAJA1 (class-A, red dashed line).

References

(1) Ibanez de Opakua, A.; Geraets, J. A.; Frieg, B.; Dienemann, C.; Savastano, A.; Rankovic, M.; Cima-Omori, M. S.; Schroder, G. F.; Zweckstetter, M. Molecular interactions of FG nucleoporin repeats at high resolution. *Nat. Chem.* **2022**, *14* (11), 1278-1285. DOI: 10.1038/s41557-022-01035-7

(2) Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A. P.; Kim, H. J.; Mittag, T.; Taylor, J. P. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **2015**, *163* (1), 123-133. DOI: 10.1016/j.cell.2015.09.015

(3) Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555* (7694), 61-66. DOI: 10.1038/nature25762

(4) Tesei, G.; Lindorff-Larsen, K. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Res Eur* **2022**, *2*, 94. DOI: 10.12688/openreseurope.14967.2.

(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* **2021**, *118* (44). DOI: 10.1073/pnas.2111696118

(6) Cao, F.; von Bulow, S.; Tesei, G.; Lindorff-Larsen, K. A coarse-grained model for disordered and multi-domain proteins. *Protein Sci.* **2024**, *33* (11), e5172. DOI: 10.1002/pro.5172 Kim, Y. C.; Hummer, G. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.* **2008**, *375* (5), 1416-1433. DOI: 10.1016/j.jmb.2007.11.063 Mugnai, M. L.; Chakraborty, D.; Nguyen, H. T.; Maksudov, F.; Kumar, A.; Zeno, W.; Stachowiak, J. C.; Straub, J. E.; Thirumalai, D. Sizes, conformational fluctuations, and SAXS profiles for intrinsically disordered proteins. *Protein Sci.* **2025**, *34* (4), e70067. DOI: 10.1002/pro.70067 Zhang, Y.; Liu, X.; Chen, J. Toward accurate coarse-grained Simulations of disordered proteins and their dynamic interactions. *J. Chem. Inf. Model.* **2022**, *62* (18), 4523-4536. DOI: 10.1021/acs.jcim.2c00974

(7) Kampinga, H. H.; Craig, E. A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (8), 579 EP-592, DOI: 10.1074/jbc.M600842200.

(8) Kakkar, V.; Månsson, C.; de Mattos, E. P.; Bergink, S.; van der Zwaag, M.; van Waarde, M. A. W. H.; Kloosterhuis, N. J.; Melki, R.; van Cruchten, R. T. P.; Al-Karadaghi, S.; 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* **2016**, *62* (2), 272-283, DOI: 10.1016/j.molcel.2016.03.017.

(9) Karamanos, T. K.; Clore, G. M. Large chaperone complexes through the lens of Nuclear Magnetic Resonance spectroscopy. *Annu Rev. Biophys.* **2022**, *51*, 223-246. DOI: 10.1146/annurev-biophys-090921-120150 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.* **2019**, *116* (43), 21529-21538, 10.1073/pnas.1914999116. DOI: 10.1073/pnas.1914999116.

(10) 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.* **2023**, *14* (1), 7066. DOI: 10.1038/s41467-023-42735-z Adupa, V.; Ustyantseva, E.; Kampinga, H. H.; Onck, P. R. Tertiary structure and conformational dynamics of the anti-amyloidogenic chaperone DNAJB6b at atomistic resolution. *Nat. Commun.* **2024**, *15* (1), 3285. DOI: 10.1038/s41467-024-46587-z

(11) Hobbs, B.; Limmer, N.; Ossa, F.; Knupling, E.; Lenton, S.; Fodera, V.; Kalverda, A. P.; Karamanos, T. K. A low-complexity linker as a driver of intra- and intermolecular interactions in DNAJB chaperones. *Nat. Commun.* **2025**, *16* (1), 5070. DOI: 10.1038/s41467-025-60063-2

(12) 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* (17), 4546-4559. DOI: 10.1021/ja00381a009.

(13) 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.* **2016**, *18* (8), 5839-5849. DOI: 10.1039/c5cp06197k

(14) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. Deviations from the simple 2-parameter model-free approach to the interpretation of ^{15}N nuclear magnetic-relaxation of proteins. *J. Am. Chem. Soc.* **1990**, *112* (12), 4989-4991. DOI: 10.1021/ja00168a070.

(15) 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.* **2022**, *62* (18), 4474-4485. DOI: 10.1021/acs.jcim.2c00450