Transient Electrostatic Interactions Dominate the Conformational Equilibrium Sampled by Multidomain Splicing Factor U2AF65: A Combined NMR and SAXS Study

Jie-rong Huang,[†,‡,§,||] Lisa R. Warner,[⊥,#] Carolina Sanchez,[⊥,#] Frank Gabel,[†,‡,§] Tobias Madl,[⊥,#] Cameron D. Mackereth,[⊥,#] Michael Sattler,*[,⊥,#] and Martin Blackledge*[,†,‡,§]

[†]University Grenoble Alpes, [‡]CNRS, and [§]CEA, Protein Dynamics and Flexibility, Institut de Biologie Structurale, 38000 Grenoble, France
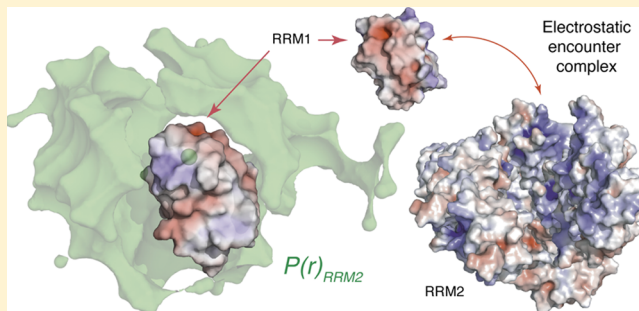
[||]Institute of Biochemistry and Molecular Biology, National Yang-Ming University, No. 155 Section 2 Linong Street Taipei, Taiwan

[⊥]Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

[#]Center for Integrated Protein Science Munich at Biomolecular NMR, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany

S *Supporting Information*

ABSTRACT: Multidomain proteins containing intrinsically disordered linkers exhibit large-scale dynamic modes that play key roles in a multitude of molecular recognition and signaling processes. Here, we determine the conformational space sampled by the multidomain splicing factor U2AF65 using complementary nuclear magnetic resonance spectroscopy and small-angle scattering data. Available degrees of conformational freedom are initially stochastically sampled and experimental data then used to delineate the potential energy landscape in terms of statistical probability. The spatial distribution of U2AF65 conformations is found to be highly anisotropic, comprising significantly populated interdomain

contacts that appear to be electrostatic in origin. This hypothesis is supported by the reduction of signature PREs reporting on expected interfaces with increasing salt concentration. The described spatial distribution reveals the complete spectrum of the unbound forms of U2AF65 that coexist with the small percentage of a preformed RNA-bound domain arrangement required for polypyrimidine-tract recognition by conformational selection. More generally, the proposed approach to describing conformational equilibria of multidomain proteins can be further combined with other experimental data that are sensitive to domain dynamics.

INTRODUCTION

Multidomain proteins comprise 80% of eukaryotic proteomes,[1] where interdomain dynamics play key roles in a large number of molecular recognition and signaling processes.[2−5] These complex dynamic modes cannot be understood from static structures of either the entire protein or the individual domains. Indeed intrinsically disordered linkers connecting folded subunits often encode the degrees of conformational flexibility essential to protein function.[6−8] While crystallography describes individual points on the free-energy landscape, nuclear magnetic resonance (NMR) and small-angle X-ray scattering (SAXS) report on averages over ensembles of interchanging conformers present in solution, and as such offer the possibility to study the conformational behavior of proteins exhibiting high levels of domain dynamics.[5,9−24]

The human U2AF65 protein plays an essential role for spliceosome assembly in pre-mRNA splicing of metazoan pre-mRNAs.[22,25−27] U2AF65 comprises three RNA Recognition Motif (RRM) domains. The first two RRMs (RRM1,2) are connected by a 32-amino acid flexible linker and are required for recognition of single-stranded poly pyrimidine-tract RNA,[22,25,28] whereas RRM3 mediates protein−protein interactions with Splicing factor 1.[29,30] A deletion mutant of RRM1,2 has been previously crystallized,[28,31] while the wild-type protein has been studied by NMR[22,32−34] and SAXS.[35,36] Paramagnetic relaxation enhancements (PREs) measured in solution suggested a dynamic equilibrium comprising a predominant compact closed structure in exchange with a minor population with a more open domain arrangement resembling the RNA-bound form.[22,34,37] This conformational equilibrium plays an important role in the recognition of a range of Py tracts found in human pre-mRNA introns, where overall binding affinity is translated into a functionally

important population shift from the closed to the open conformation. However, SAXS data indicate that the conformational ensemble is yet more complex as neither open nor closed conformations reproduce experimental scattering curves (Figure 1A).[36] RRM1,2 therefore represents a classical example of the family of dynamic multidomain proteins, which sample a highly diverse continuum of conformations in solution.
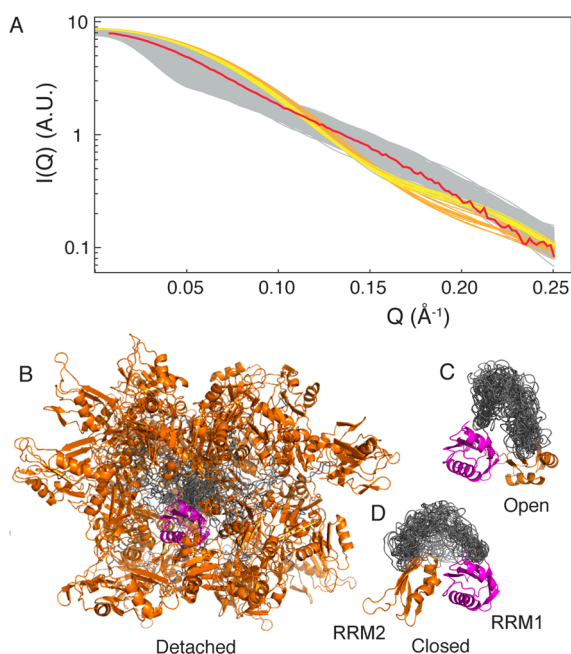

Figure 1. (A) Experimental (red) and predicted SAXS curves from the ensemble of NMR derived structures for closed (orange, PDB entry: 2YH0) and open conformations (yellow, PDB entry: 2YH1), and from the detached ensemble (gray). Representative structures of the detached (B), open (C), and closed ensembles (D) used as the prior distributions for ASTEROIDS analysis. RRM1 (purple), RRM2 (orange), and linker and flexible C-terminus (gray).

PREs are highly sensitive to transient contacts between domains in biomolecules, and can be used to detect encounter-like contacts in protein–protein and protein–nucleic acid complexes.[38,39] Residual dipolar couplings (RDCs) report on the distribution of relative orientations sampled by different structured domains,[40] as well as reporting on the conformational propensities of disordered chains.[41] SAXS, on the other hand, is sensitive to the population-weighted pairwise distribution functions averaged over the entire ensemble present in solution, providing low-resolution information about interconverting states of different shape.[12]

Over recent years, considerable efforts have been devoted to the development of molecular descriptions that account for the conformational heterogeneity in flexible multidomain proteins using NMR data. The number of degrees of conformational freedom available to multidomain proteins may be expected to outweigh independent experimental measurements, so that care must be taken to avoid overfitting molecular ensemble descriptions of such systems. Equally, it is important to consider the most appropriate means for presenting inherent uncertainty in ensemble descriptions, including both the selection of representative conformations, as well as the determination of their associated populations. In this context, Luchinat and co-workers have developed an approach to determine the maximum occurrence, or weight, of each possible

protein conformation on the basis of experimental NMR or SAXS data.[18,42] Interdomain dynamics were also recently studied using paramagnetic constraints derived from lanthanides bound to a Calmodulin-IQ recognition motif.[24] In this case, the interpretation of solution data was based on ensemble selection from existing crystal structures and free molecular dynamics (MD) simulation in a similar way to the study of the conformational equilibrium sampled by Hck tyrosine kinase using SAXS data.[43] Fushman and co-workers recently presented an approach to ensemble representation that selects non-uniformly weighted representative ensembles that best reproduce experimental data,[44] while the EROS approach uses maximum entropy weight distributions to describe multidomain equilibria from diverse types of experimental data.[45,46] Clore and co-workers studied full-length HIV-1 capsid protein using a combination of X-ray scattering and RDCs, and interpreted these data in terms of an ensemble of conformers, using rigid-body modeling to sample conformational space.[47] More commonly, replica exchange molecular dynamics ensemble simulations[48−55] are used to provide multiconformational distributions that are in agreement with experimental data.[56−58]

In parallel to the emergence of these diverse methodologies, over the past few years we have developed an approach for ensemble representations of intrinsically disordered systems (ASTEROIDS)[59,60] that uses rational sampling of conformational space, via a tested statistical coil model, and a genetic algorithm to optimize members of an ensemble whose size is determined by reproduction of independent experimental data. Here, we have adapted this approach to map the nature of the intrinsic conformational equilibrium sampled by the flexible multidomain protein U2AF65. To map the conformational-energy surface of RRM1,2, transient contact mapping from PREs, orientational averaging from RDCs, and distance distribution functions from SAXS are simultaneously integrated into the ASTEROIDS analysis. Particular attention is paid to calibration and validation of the representative ensembles.

■ ASTEROIDS ANALYSIS OF MULTIDOMAIN PROTEINS − THEORETICAL ASPECTS

While the philosophy is broadly similar to other sample-and-select approaches, some specific characteristics differentiate ASTEROIDS from existing protocols for studying multidomain proteins, meriting more in-depth discussion. In general, ensemble selection approaches must resolve two problems; the identification of representative substates, and the determination of their relative populations. With respect to the first of these, to guarantee the maximum unbiased sampling available to the system, the *flexible-meccano* statistical coil model[61,62] is used to maximally explore the phase space available to the disordered amino acids in the system of interest, in this case the flexible linker between domains RRM1 and RRM2 and the C-terminal tail. No experimental data are used in this step, the objective of which is to sample the potentially extensive conformational energy landscape accessible to a protein in the context of a simple amino-acid specific statistical coil model. In the case of the RRM1,2 tandem domains, only steric hindrance, of both folded and unfolded regions, and the amino-acid specific statistical coil model restrain the sampling, which is thereby free from bias from either experimental data, existing structures determined by alternative methods, or more complex MD force fields. The first step therefore broadly
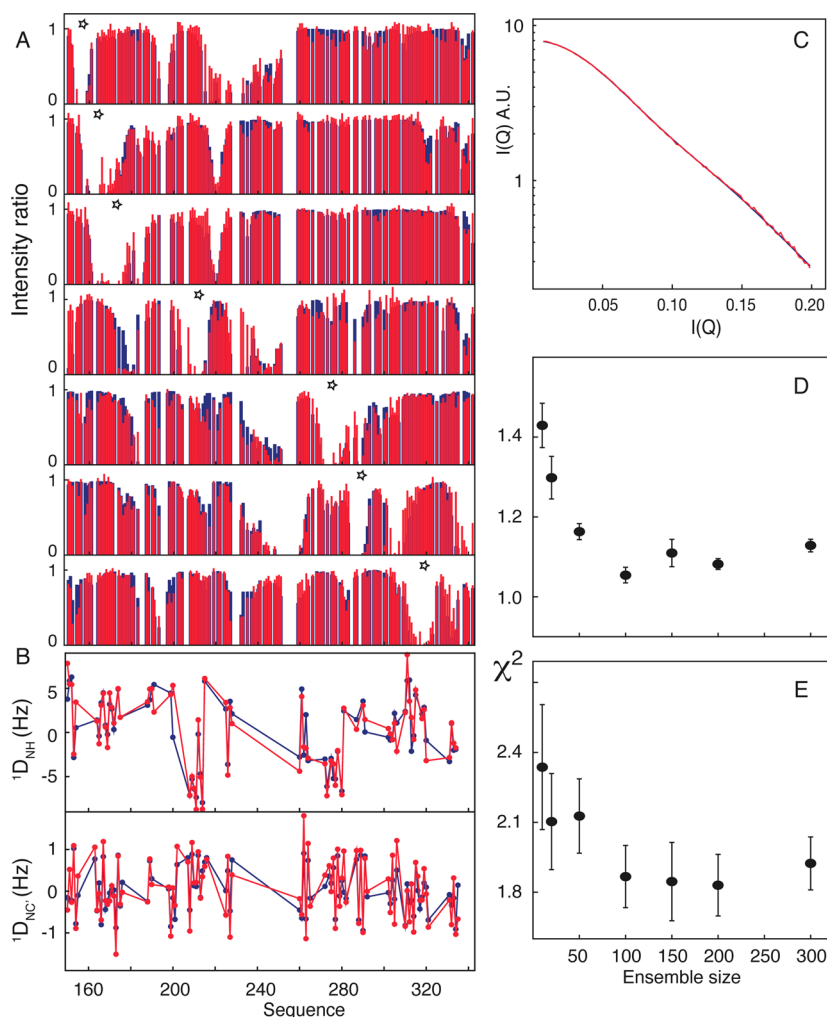
Figure 2. Comparison between experimental (red) and fitted (blue) data for PRE (intensity ratio for the paramagnetic and reduced state (I^{para}/I^{dia})) (A), RDC (B), and SAXS (C) (x-axis : Å^{-1}). In (A), spin labels (stars) are located on residue numbers 155, 164, 171, 209, 273, 287, and 318. For active (D) and passive (E) data sets as a function of ensemble size.

populates conformational space, and establishes a prior sampling, from which ensembles can be selected.

The selection step performed by the ASTEROIDS algorithm[59,63,64] determines combinations of conformers from the prior sampling pools that are in agreement with the experimental data within their estimated uncertainty. ASTEROIDS uses a specifically adapted genetic algorithm, where each gene represents one conformer in the ensemble, and whose success is measured against a target function gauging the ability to reproduce experimental data. Random selection, mutations, and crossings of conformers are used to evolve generations toward the final ensemble. The size of the chromosome represents the size of the representative ensemble (n), which is constant and is experimentally determined a priori using cross-validation against randomly selected, independent data sets that are not included in the target function.

Crucially, the populations of different conformers (i) are given by $p_i = 1/n$ and are not optimized in the selection procedure. If a given state requires a higher population to fulfill experimental data, this is achieved by the genetic algorithm oversampling additional substates presenting the required characteristics with respect to the experimental data. The optimal number of structures is determined by cross-validation of independent data, and depends on a number of factors, of which two are clearly identifiable: (1) the volume, accuracy, and diversity of experimental data, which necessarily dictate the complexity of the ensemble required to reproduce all data, and (2) the complexity of the underlying dynamic equilibrium. ASTEROIDS therefore contrasts to approaches that both identify representative substates and subsequently optimize their populations. In such cases, fewer structures may be used to reproduce the experimental data, but their differential weights must be determined in a second step. We note that these individual populations, or weights, are often estimated to percentile precision, which is implicitly equivalent to a much higher number of equally populated states. Populations determined in this way may also be correlated, so that the uncertainty of populations of selected conformers may be difficult to determine, due to the complex parametric surface presenting multiple local minima. Although for combinatorial reasons, the ASTEROIDS single-step selection algorithm requires more equally populated conformers, it avoids this additional optimization problem, and crucially allows for robust noise-based Monte Carlo error analysis providing straightforward experimental error propagation (vide infra). The ASTEROIDS procedure is presented in the form of a flowchart in Figure S1 of the Supporting Information.

Table 1. Statistical Analysis of Data Reproduction Using the ASTEROIDS Approach to Ensemble Selection from Pools DET (Detached), OP (Open), and CL (Closed) (Compare Figure 1)

	PRE+RDC+SAXS			PRE+RDC+SAXS[g]		
	(1184)			no selection (1184)		
pool[a]	DET,CL,OP	DET	CL,OP	OP	CL	DET
χ^{2b}	1502	1577	23 321	28 130	30 341	13 376
χ^2/N^c	1.27	1.33	19.70	23.8	25.6	11.3
DET (%)[d]	75.0	100	0	0	0	100
CL (%)[e]	24.3	0	18.3	0	100	0
OP (%)[f]	0.7	0	81.7	100	0	0

[a]Type of pool used in the selection. DET (detached), OP (open), and CL (closed). [b]$\chi^2 = \sum(\text{param}_{calc} - \text{param}_{exp})^2/\sigma^2$ summed over all experimentally measured points. [c]Value in b divided by the number of measured points. [d]Percentage of conformers derived from pool DET (detached) in the final selection. [e]Percentage of conformers derived from pool CL (closed) in the final selection. [f]Percentage of conformers derived from pool OP (open) in the final selection. [g]Reproduction of experimental data by the pools (no selection is performed).

RESULTS AND DISCUSSION

The conformational space theoretically available to the RRM1,2 tandem domains of U2AF65 was sampled using the statistical coil *flexible-meccano* model to describe the linker region and the flexible tail,[61,62] while known structures were used to describe domains RRM1 and RRM2.[22] Conformers calculated in this way comprise the "detached" ensemble (Figure 1B). Additional ensembles randomly sampling the linker region were created from the previously determined NMR-derived "open" and "closed" conformations[22] using XPLOR-NIH[65] sampling for the linker (Figure 1C,D), with the relative positions of the folded domains being predefined. Although selection is from the "detached" ensemble alone, these additional ensembles were included in additional analysis for comparison with the statistical coil approach, because they had been previously proposed on the basis of "open" like and "closed" like experimental PRE profiles. These three pools represent the unbiased prior distributions from which ensembles are selected on the basis of experimental data. Experimental SAXS data from the RRM1,2 tandem domains of U2AF65 are compared to predicted curves from "open", "closed", and "detached" ensembles in Figure 1. The more compact dimensions of the "open" and "closed" conformations as compared to the experimental curve are evident, suggesting the existence of a significant population of more extended conformations, which may be represented in the "detached" ensemble.

RDCs were calculated for each member of the ensemble using a steric repulsion model.[66] PREs were calculated by sampling side-chain conformations using published rotameric distributions and built explicitly for each spin-label site of each structure in the pools.[59] Paramagnetic relaxation effects were averaged over all sampled conformers as described previously.[59] Small-angle scattering curves were predicted using the program CRYSOL.[67]

The conformational-energy landscape was mapped by selecting multiconformer ensembles in agreement with the experimental data (Figure S1, Supporting Information).[60] Typical experimental data reproduction is shown in Figure 2A–C. Two types of calculation were performed, either selecting from the detached pool alone, or from a mixture of open, closed, and detached pools. 10% of all data were withdrawn from the analysis and used as passive reporters on the predictive nature of the approach. While this procedure rigorously addresses the ability to predict independent PREs and RDCs, we note that due to the inherently low resolution of SAXS data, randomly removing 10% of the measured data points from a single curve may not provide for truly independent *R*-free measurement.[68] However, we also note that the number of structures required to reproduce all independent data is entirely dominated by the PRE and RDC data sets as compared to the SAXS data. This analysis was therefore used to determine the optimal number of conformers present in each representative ensemble (Figure 2D,E). In this case, 200 structures were found to best reproduce experimental data, although no differences in the conformational energy landscape could be identified when the entire procedure was performed using 100 conformers (data not shown). The resulting reduced χ^2 is approximately unity (Table 1), justifying the combination of noise-based Monte Carlo simulations with ASTEROIDS to assess the precision of the proposed conformational description.

Typical maps of conformational space based on ensemble descriptions are shown in Figure 3. The spheres represent the centers of mass (*com*) of RRM2 conformers, when all domains of RRM1 are superimposed (Figure 3A–C). The density of populated substates is also shown as a contour map projected onto *XY* and *XZ* planes (Figure 3E–H). Monte Carlo simulations based on experimental noise show population distributions that are effectively identical to the analysis of experimental data (Supporting Information Figure S2), demonstrating statistical robustness. This procedure populates the conformational space within the estimated experimental error, providing a map of the conformational-energy landscape, and a measure of confidence in the substate populations.

Redistribution of States As Compared to Available Sampling. There is a clear redistribution of population density in the experimental free-energy landscape as compared to the degrees of freedom available to the system as defined by the detached pool. A high density of states is found in the vicinity of the "closed" conformation (black sphere, Figure 3B,F), even though this region of conformational space is weakly represented in the prior detached distribution (Figure 3A,E). To quantify the population distribution in the vicinity of the known structures, we have determined the number of conformers in each ensemble whose RRM2 *com* is within a given distance of the RRM2 *com* of the known structure when RRM1 is superimposed (procedure illustrated in Supporting Information Figure S3). 20% of conformers in the selection from only the detached pool were found to have *com* distances of less than 32 Å as compared to the "closed" form[22] (as compared to 9% from the nonbiased sampling). This is also true of the "open" conformation, where the *com* distance of 22% of conformers is found to be within 32 Å of the previously determined "open", RNA-bound conformation,[22] as compared
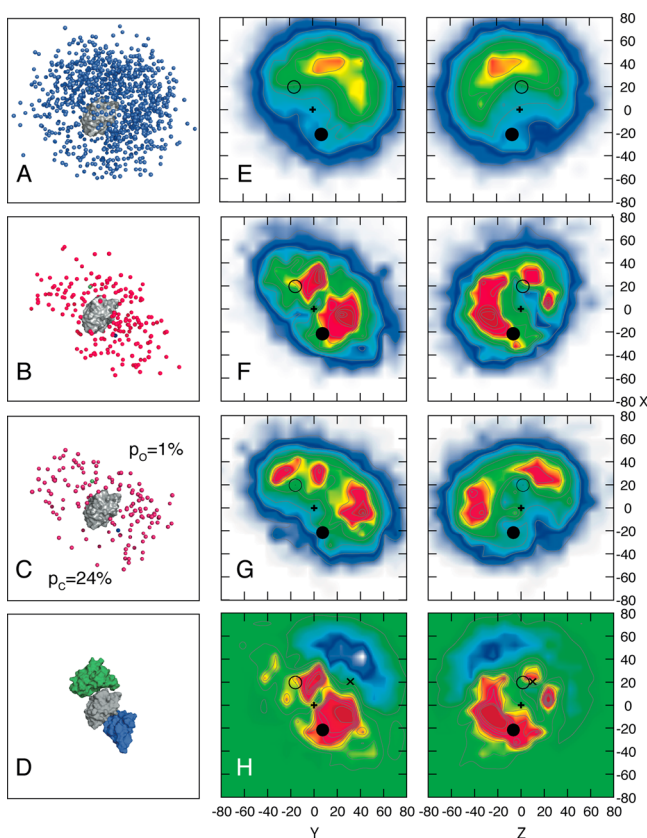
Figure 3. Comparison of distribution of *com*s of RRM2 relative to RRM1 (gray surface). (A) Representation of conformers in a pool of detached conformations, (B) distribution of conformations selected from detached ensemble, and (C) distribution of conformations selected from open, closed, and detached ensembles. (D) Position of open (2YH1-green) and closed (2YH0-blue) conformations of RRM2 relative to RRM1. (E–G) 2D projections onto *XY* and *XZ* planes showing populations (derived from 100 Monte Carlo simulations; $\chi^2/N \approx 1$) associated with calculations shown in (A)–(C) (red/blue highly/weakly populated). "+" indicates *com* of RRM1; closed and open circles 2YH0 and 2YH1, respectively. "×" position of crystal structure 2G4B. (H) Difference in populations between (E) and (F) (red, positive; green, zero; blue, negative). All coordinates are shown in Angstroms.

to only 11% in the detached prior pool. Mapping of the conformational space using a combination of open, closed, and detached prior pools resulted in a similar population distribution as compared to selection from the detached prior pool alone, with 24.3 ± 0.6% of conformers represented identically by the "closed" conformation and 0.75 ± 0.25% by the "open" conformation. However, as in the case of the selection from the detached prior pool only, 22% of conformers have *com* distance within 32 Å of this position (Figure 3C,G). Selection from prior pools containing only the more compact closed and open ensembles results in χ^2/N more than 15 times higher than selection from all three, or from the detached prior pool alone, which are approximately equivalent, while individual pools without selection reproduce the data with χ^2/N 19, 20, and 10 times higher for open, closed, and detached pools, respectively (Supporting Information Figures S4–S6, Table 1).

More extended conformers are significantly less populated in the experimental selection as compared to the detached prior pool. Further analysis indicates that 54% of the selected

conformations are more compact than the underlying bias-free sampling would predict (Supporting Information Figure S7). This is also demonstrated in the difference plots between the prior and selected distributions (Figure 3H). Notably, the previously reported crystal structure of a linker deletion variant of RRM1,2 bound to a poly uridine RNA oligonucleotide ("×" in Figure 3) lies in a weakly populated region, suggesting that this is an unfavorable conformation, likely induced by crystal packing effects.

Compactness Correlates with a Favorable Electrostatic Interdomain Interaction. We have further investigated the origin of the redistribution as compared to the prior detached pool, in particular the enhanced propensity to populate compact conformations. Figure 4 shows the three-
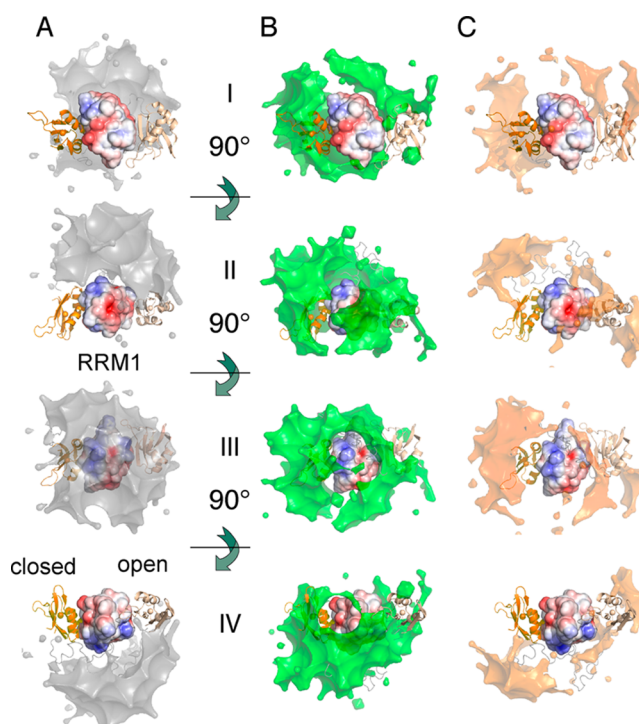

Figure 4. 3D density maps (5% population contour) of RRM2 distribution with respect to RRM1. (A) Prior sampling pool of all detached conformations, (B) ASTEROIDS selection using all experimental data from detached conformations, and (C) ASTEROIDS selection using all experimental data from a combination of open, closed, and detached conformations. RRM1 (surface) is colored according to surface charge (Adaptive Poisson–Boltzmann Solver); positive (blue) and negative (red) RRM2 domains in closed and open conformations are shown in orange and magnolia, respectively. Hydrophobic surfaces are shown in Supporting Information Figure S8, and coincide approximately with the white surfaces shown here.

dimensional topology of population isocontours for the prior and experimental free-energy landscapes in comparison to electrostatic charge distribution[69] (hydrophobic surfaces are shown in Supporting Information Figure S8). This representation has the advantage of highlighting the sampling derived from the experimental constraints with respect to the physical properties of the protein surfaces. Hydrophobic patches on the surface of RRM1 (orientations I and III) are relatively infrequently found to interact with RRM2, as was also highlighted in orientation III (Figure 3H), where highly populated sampling in the prior distribution is strongly
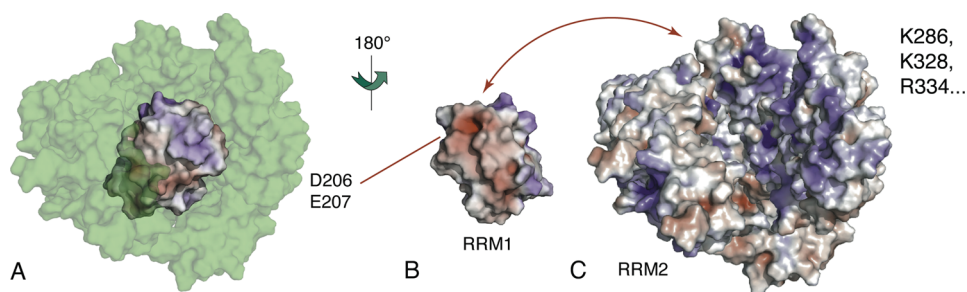
7072

dx.doi.org/10.1021/ja502030n | *J. Am. Chem. Soc.* 2014, 136, 7068–7076

Figure 5. Origin of the observed encounter complex is electrostatic. (A) Green surface shows RRM2 domains from detached selection (Figure 3B,F and Figure 4B) whose *com* values are less than 28 Å from RRM1 (colored surface) *com*. (B) Charge distribution of RRM1 face in most frequent contact with RRM2, indicating acidic patch (orientation of RRM1 rotated 180° with respect to A). (C) Electrostatic distribution on RRM2 surfaces shown in (A). Most basic region coincides with positions in closest contact with the acidic patch on RRM1 (B).

depleted, corresponding to the negative region in the population difference.

By contrast, the acidic patch (orientation II) constitutes the main source of positive population redistribution, with a clear occlusion of this surface in the selected ensemble as compared to the prior sampling. Investigation of this interaction from the RRM2 perspective (Figure 5) reveals that the RRM2 interfaces in closest proximity to D206/E207 exhibit a predominantly positive electrostatic charge (involving K286, K328, R334). This further suggests that the bias toward an ensemble of compact states is electrostatic in origin.

Contribution of Different Experimental Data Types to Ensemble Selection. It is instructive to analyze the contributions of the different types of experimental data for the selected ensembles, by removing individual types of experimental data or pairs of data types from the selection. The results (Supporting Information Table S1, Figure S9) indicate that all three data types (PRE, RDCs, and SAXS) are required to accurately map the different conformations. Not surprisingly, removal of SAXS data from the target function results in more compact ensembles, while fulfillment of PREs only requires conformers in close contact. The angular dependence of RDCs appears to be responsible for the distribution of the population into two higher density nodes, as these already appear in RDC-only selections (Supporting Information Figure S9B). These findings indicates that the system remains underdetermined, despite the volume of complementary data, and that while small, random selections of data can be removed and correctly predicted, removing all of one data type (RDC, SAXS, or PRE) destabilizes the ensemble selection.

Experimental Verification of Predicted Transient Contact. To provide further evidence for the validity of the ASTEROIDS ensemble description, we recorded PRE data as a function of salt concentration for RRM1,2 with a spin label attached to residue 318 in RRM2, which induces strong PRE effects for a number of residues in the RRM1 domain that are located in the vicinity of the observed interface. The data in Figure 6 show that the interdomain PRE is notably reduced with increasing salt concentration, in particular the contact in the vicinity of D206/E207. This observation is consistent with destabilization of encounter-like compact states in the RRM1,2 ensemble involving the two charged interfaces.

■ CONCLUSION

Interdomain dynamics play important mechanistic roles in regulating a vast range of biomolecular processes that can best
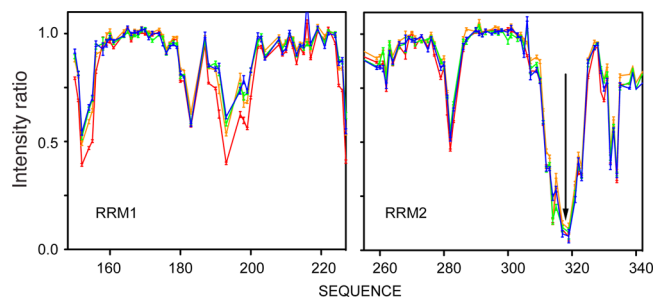

Figure 6. PRE measurement (ratio of intensities for the paramagnetic (I^{para}) and reduced (I^{dia}) sample) for spin-labeled A318C RRM1,2 at different salt concentrations (red, 0 mM; orange, 50 mM; green, 150 mM; blue, 200 mM). Negligible salt-dependent effects are seen on RRM2 (containing spin label), while there is clear reduction of PREs in RRM1 with increasing salt.

be understood at the molecular level by developing descriptions of the free-energy landscape sampled in solution. We have adapted the ASTEROIDS ensemble selection procedure, based on rational sampling of conformational space using a sequence-dependent stochastic sampling algorithm to account for the sequence specificity of interdomain linkers, and employed data-driven mapping to select representative ensembles within this distribution. This provides a statistical description of the conformational space sampled by the system, and estimates confidence levels and parametric uncertainties associated with the free-energy description that can be determined irrespective of the scarcity of experimental data.

Application of this procedure to map the free-energy landscape of the tandem RRM domains of U2AF65 reveals a complex equilibrium comprising highly populated states with RRM1 and RRM2 in close proximity, resembling previously proposed "closed" and "open" states.[22] Inclusion of these states explicitly in the prior distribution results in one-quarter of conformers adopting these conformations (mostly in the "closed" form), while sampling of only "detached" conformers results in a population redistribution that oversamples these regions of three-dimensional space. On the basis of the ensemble description, the proximity of the two domains in the region resembling the "closed" state appears to be driven by electrostatic interactions. This prediction is supported by the reduction of signature PREs with increasing salt concentration, implying a weakening of transient contacts in this interface. It is worth mentioning that, although the observation of a range of inactive domain arrangements in unbound U2AF65 is consistent with physical considerations, for example, charge

complementarity of the two interfaces of RRM1 and RRM2, it derives solely from analysis of the experimental NMR and SAXS data. No force field is used in the free-energy mapping procedure, which is fully data-driven. The resulting ensembles also confirm the presence of a significant population of substates in close proximity to the "open conformation". The similarity of this region of conformational space with the RNA-bound state of RRM1,2 supports the suggested role for conformational selection from the free-state ensemble, which represents an ensemble of "inactive" conformations that are not competent for high-affinity RNA binding.[5,22] Although in very different conformations, both "open" and "closed" conformations lie within a continuous ensemble envelope (Figure 7), suggesting a possible "pathway" of available states that can channel between the two states without evoking large-scale jumps.
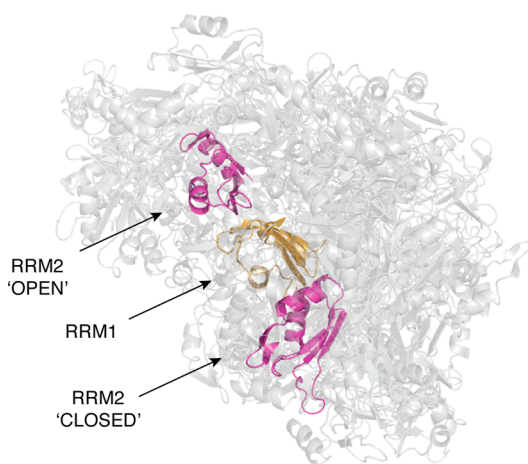

Figure 7. Position of the "open" and "closed" conformations of RRM2 (purple) with respect to the ensemble of RRM2 conformers representing the free-energy landscape sampled by the protein (gray).

The ensemble described here is distinct from previously developed models,[22,28,31,36] differing essentially in the relatively high proportion of "detached" conformations whose presence is clearly identified by the SAXS data. Paramagnetic restraints and RDCs are only weakly sensitive to these "detached" conformers. The resulting ensemble reflects conformational properties probed by all three data types, and underlines the importance of maximizing complementary information from different restraints when considering the conformational space sampled by multidomain proteins.

Although we have no direct thermodynamic evidence, the existence of a range of dynamically interconverting inactive (closed) conformations of unbound U2AF65 RRM1−RRM2 may imply a reduced cost of conformational entropy to keep U2AF65 in a predominantly inactive state, without enforcing a single unique and specific closed domain arrangement. Together with the unbiased confirmation of the existence of a minor conformation that resembles the "active" domain arrangement seen in the RNA-bound protein, our findings underline the possible contribution of conformational selection in the molecular recognition of Py tract RNA by U2AF65. In the future, it will be important to unravel the conformational shifts induced by the presence of Py tract RNA ligands of varying strength, that is, binding affinity, to understand the

molecular mechanisms employed by U2AF65 during spliceo-some assembly.

■ MATERIALS AND METHODS

Experimental Data Collection. Measurement of RDCs and PREs shown in Figure 2 has been previously described.[22] Uncertainty of $^1D_{NH}$ and $^1D_{CN}$ RDC measurements is 1 and 0.5 Hz, respectively, 0.15 for PRE intensities, and experimental errors for SAXS data were used.

For measuring salt-dependent PREs, the A318C mutant of U2AF65 RRM1,2 was expressed, purified, and labeled with 3-(2-iodoacetami-do)-2,2,5,5,tetramethyl-1-pyrrolidinyloxy radical (IPSL; Sigma-Aldrich) as described previously.[22] Briefly, the purified protein was reduced by incubation with 10 mM dithiothreitol overnight, followed by extensive dialysis into 1 M Tris (pH 8.0) and 200 mM NaCl. Next, 10 mol equiv of IPSL was added and allowed to react overnight in the dark at 4 °C. The labeled protein was removed from the excess IPSL and exchanged into 20 mM MES/Bis-tris (pH 6.5) buffer with a PD10 desalting column (GE Healthcare Life Sciences). To analyze the salt dependence of the PREs on RRM1,2 A318C, equal volumes of salt solutions were added to 20 mM MES/Bis-tris (pH 6.5) protein solutions to give final concentrations of 0, 50, 150, and 200 mM NaCl such that the final dilutions were identical.

PREs were determined using a ratio of peak intensities from $^1H,^{15}N$-HSQC and/or -TROSY spectra without (I^{para}) and with the addition of 10 mol equiv of ascorbic acid (I^{dia}). Data were analyzed as described previously.[34,70]

Construction of Prior Sampling Pools. The pool for the "detached" conformation was constructed using the statistical coil model based program, *flexible-meccano*.[61,62,71] A random-coil linker was built starting from the C-terminal end of the RRM1 domain; the RRM2 domain was constructed by *flexible-meccano* but with fixed dihedral angles. The C-terminal tail was also constructed according to statistical coil model. This pool contains 10 000 conformers.

The pools of the "closed" and "open" conformation were generated using XPLOR-NIH.[65] One of the deposited NMR-solved structures for each conformation was used as the starting structure (PDB entries: 2YH0 and 2YH1).[22] The coordinates of the RRM domains (residue number 148−227 and 260−335) were fixed with the linker (residues 228−259) and the C-terminal tail (residues 336−342) unrestrained during the simulated annealing protocol. The calculation was started from 3500 K for 20 ps dynamics run and followed by cooling to 300 K in 50 K decrements. At each decrement step, a 0.5 ps dynamics run was used for equilibration. Only the standard potentials, that is, bond, angle, improper torsion, steric repulsion from van der Waals, and the Ramachandran energy surface, were used during the simulated annealing to maintain the covalent geometry. The calculation was repeated 5000 times for both "closed" and "open" conformations to ensure that the conformational space for the linker is well sampled.

Predicted Experimental Values. For PREs, side-chain conformations were sampled using published rotameric distributions[72] and built explicitly for each spin-label site of each structure in each of the pools. 600 sterically allowed side-chain conformers were calculated for each spin label of each conformer. Relaxation effects were averaged over these conformers as described previously.[59]

The RDCs were calculated on the basis of the steric alignment tensor of every conformer.[66] In the "detached" pool, the RRMs (residues 148−227 and 260−335) were replaced with NMR solved structures (1YH0 and 1YH1) to reduce the peptide plane geometry difference between *flexible-meccano* generated plane and deposited structures. This replacement is only used in the RDC calculation.

For SAXS, side chains were added using SCCOMP.[73] The predicted SAXS curves were calculated using CRYSOL[67] with default settings, except the number of points in theoretical curve (ns = 132) and the maximum scattering vector (σ_μ = 0.2508) were changed to fit the proper range of experimental data.

ASTEROIDS Selection. The procedure of ASTEROIDS selection was described previously.[59,63] To prevent bias from different types of pool, the predicted parameters for "closed" and "open" conformations were repeated four times in the pool for selection (i.e., 20 000

conformers for each pool). In fact, even when 80% of conformers in the pool are in "closed" or "open" conformations, the selected results were not biased to these two types of structures as compared to using pools comprising 50% of these conformers (25% selected structures derive from pools B and C in both cases).

To optimize the size of ensemble for selection, different sizes, ranging from 10 to 300 conformers, were used for ASTEROIDS selection. 10% of experimental data were left out randomly for cross-validation. This procedure was repeated 100 times to provide good statistical sampling and the basis of the error bars in Figure 2E. The χ^2 between experimentally measured data and back-calculated values for active data (90% data used as restraints) or passive data (10% left-out data) were converged when the size of ensemble is around 200 conformers.

Statistic Significance Assessment. One set of ASTEROIDS-selected back-calculated experimental parameters was used as a starting point ("Set A") for statistical significance assessment. At each new ASTEROIDS selection, the "experimental" restraints were derived from the "Set A" added with a value according to the Gaussian distribution of experimental errors. This simulation was repeated 400 times. The distribution of χ^2 behaves as Gaussian, with a mean value (μ) of 1241.6 and the square root of variance (σ) as 54.8. Therefore, the structures with 95% confidence were taken from those selections with χ^2 between $\mu \pm 2\sigma$ and with 68% confidence were taken from the selections with χ^2 between $\mu \pm \sigma$.

Measurement of SAXS. The SAXS measurements were recorded on the new ID14-3 BioSAXS beamline at the European Synchrotron Radiation Facility (ESRF Grenoble, France). Sample−detector distance was 2.6 m yielding a Q-range from 0.005 to 0.56 Å^{-1} for an X-ray wavelength of 0.931 Å (13.32 keV). 50 μL of RRM12 solutions (at four different protein concentrations of 1.19, 2.39, 4.78, and 9.48 mg/mL) were loaded in a flow-through quartz capillary cell at 20 °C. The sample volume exposed to the X-ray beam was about 10 μL. The radiation damage was checked by 10 successive exposure times of 10 s each. Final exposure time was 100 s for all samples/buffers. The last three frames were not used in the averaging due to effects of radiation damage being present. The 2D diffraction patterns were normalized to an absolute scale and azimuthally averaged to obtain the intensity profiles $I(Q)$, within BSxCuBE (ESRF beamline data collection software). Solvent contributions (buffer backgrounds collected before and after every protein sample) were averaged and subtracted from the associated protein sample using the program PRIMUS.[74]

ASSOCIATED CONTENT

S **Supporting Information**

Table showing comparison of data fitting in the absence of one or two types of data (RDCs, SAXS, or PREs). Figures illustrating the ensembles resulting from these selections. Figures showing how well the different pools reproduce the experimental data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors
sattler@helmholtz-muenchen.de
martin.blackledge@ibs.fr

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Vogel, C.; Bashton, M.; Kerrison, N. D.; Chothia, C.; Teichmann, S. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 208.
(2) Smock, R. G.; Gierasch, L. M. *Science* **2009**, *324*, 198.
(3) Bahar, I.; Chennubhotla, C.; Tobi, D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633.
(4) Tzeng, S.-R.; Kalodimos, C. G. *Curr. Opin. Struct. Biol.* **2011**, *21*, 62.
(5) Mackereth, C. D.; Sattler, M. *Curr. Opin. Struct. Biol.* **2012**, *22*, 287.
(6) Shamoo, Y.; Abdul-Manan, N.; Williams, K. R. *Nucleic Acids Res.* **1995**, *23*, 725.
(7) Ma, B.; Tsai, C.-J.; Haliloglu, T.; Nussinov, R. *Structure* **2011**, *19*, 907.
(8) Tripsianes, K.; Madl, T.; Machyna, M.; Fessas, D.; Englbrecht, C.; Fischer, U.; Neugebauer, K. M.; Sattler, M. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1414.
(9) Henzler-Wildman, K.; Lei, M.; Thai, V.; Kerns, S.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913.
(10) Bernadó, P.; Blackledge, M. *Nature* **2010**, *468*, 1046.
(11) Camilloni, C.; Robustelli, P.; De Simone, A.; Cavalli, A.; Vendruscolo, M. *J. Am. Chem. Soc.* **2012**, *134*, 3968.
(12) Bernado, P.; Mylonas, E.; Petoukhov, M.; Blackledge, M.; Svergun, D. *J. Am. Chem. Soc.* **2007**, *129*, 5656.
(13) Bernadó, P.; Fernandes, M. X.; Jacobs, D. M.; Fiebig, K.; García de la Torre, J.; Pons, M. *J. Biomol. NMR* **2004**, *29*, 21.
(14) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789.
(15) Baber, J. L.; Szabo, A.; Tjandra, N. *J. Am. Chem. Soc.* **2001**, *123*, 3953.
(16) Ryabov, Y. E.; Fushman, D. *J. Am. Chem. Soc.* **2007**, *129*, 3315.
(17) Clore, G. M.; Iwahara, J. *Chem. Rev.* **2009**, *109*, 4108.
(18) Bertini, I.; Giachetti, A.; Luchinat, C.; Parigi, G.; Petoukhov, M. V.; Pierattelli, R.; Ravera, E.; Svergun, D. I. *J. Am. Chem. Soc.* **2010**, *132*, 13553.
(19) Rozycki, B.; Kim, Y. C.; Hummer, G. *Structure* **2011**, *19*, 109.
(20) Rambo, R. P.; Tainer, J. A. *Curr. Opin. Struct. Biol.* **2010**, *20*, 128.
(21) Bashir, Q.; Volkov, A. N.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2010**, *132*, 241.
(22) Mackereth, C. D.; Madl, T.; Bonnal, S.; Simon, B.; Zanier, K.; Gasch, A.; Rybin, V.; Valcarcel, J.; Sattler, M. *Nature* **2011**, *475*, 408.
(23) Rezaei-Ghaleh, N.; Klama, F.; Munari, F.; Zweckstetter, M. *Angew. Chem., Int. Ed.* **2013**, *52*, 11410.
(24) Russo, L.; Maestre-Martinez, M.; Wolff, S.; Becker, S.; Griesinger, C. *J. Am. Chem. Soc.* **2013**, *135*, 17111.
(25) Zamore, P. D.; Patton, J. G.; Green, M. R. *Nature* **1992**, *355*, 609.
(26) Banerjee, H.; Rahn, A.; Davis, W.; Singh, R. *RNA* **2003**, *9*, 88.
(27) Wahl, M. C.; Will, C. L.; Lührmann, R. *Cell* **2009**, *136*, 701.
(28) Sickmier, E. A.; Frato, K. E.; Shen, H.; Paranawithana, S. R.; Green, M. R.; Kielkopf, C. L. *Mol. Cell* **2006**, *23*, 49.
(29) Berglund, J. A.; Abovich, N.; Rosbash, M. *Genes Dev.* **1998**, *12*, 858.
(30) Selenko, P.; Gregorovic, G.; Sprangers, R.; Stier, G.; Rhani, Z.; Krämer, A.; Sattler, M. *Mol. Cell* **2003**, *11*, 965.
(31) Thickman, K. R.; Sickmier, E. A.; Kielkopf, C. L. *J. Mol. Biol.* **2007**, *366*, 703.
(32) Ito, T.; Muto, Y.; Green, M. R.; Yokoyama, S. *EMBO J.* **1999**, *18*, 4523.
(33) Mackereth, C. D.; Simon, B.; Sattler, M. *ChemBioChem* **2005**, *6*, 1578.
(34) Simon, B.; Madl, T.; Mackereth, C. D.; Nilges, M.; Sattler, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 1967.
(35) Jenkins, J. L.; Shen, H.; Green, M. R.; Kielkopf, C. L. *J. Biol. Chem.* **2008**, *283*, 33641.
(36) Jenkins, J. L.; Laird, K. M.; Kielkopf, C. L. *Biochemistry* **2012**, *51*, 5223.
(37) Madl, T.; Felli, I. C.; Bertini, I.; Sattler, M. *J. Am. Chem. Soc.* **2010**, *132*, 7285.

7075

dx.doi.org/10.1021/ja502030n | *J. Am. Chem. Soc.* 2014, 136, 7068−7076

(38) Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18945.

(39) Tang, C.; Iwahara, J.; Clore, G. M. *Nature* **2006**, *444*, 383.

(40) Tolman, J.; Ruan, K. *Chem. Rev.* **2006**, *106*, 1720.

(41) Jensen, M. R.; Ruigrok, R. W.; Blackledge, M. *Curr. Opin. Struct. Biol.* **2013**, *23*, 426.

(42) Bertini, I.; Gupta, Y. K.; Luchinat, C.; Parigi, G.; Peana, M.; Sgheri, L.; Yuan, J. *J. Am. Chem. Soc.* **2007**, *129*, 12786.

(43) Yang, S.; Blachowicz, L.; Makowski, L.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15757.

(44) Berlin, K.; Castañeda, C. A.; Schneidman-Duhovny, D.; Sali, A.; Nava-Tudela, A.; Fushman, D. *J. Am. Chem. Soc.* **2013**, *135*, 16595.

(45) Rozycki, B.; Kim, Y. C.; Hummer, G. *Structure* **2011**, *19*, 109.

(46) Francis, D. M.; Rozycki, B.; Koveal, D.; Hummer, G.; Page, R.; Peti, W. *Nat. Chem. Biol.* **2011**, *7*, 916.

(47) Deshmukh, L.; Schwieters, C. D.; Grishaev, A.; Ghirlando, R.; Baber, J. L.; Clore, G. M. *J. Am. Chem. Soc.* **2013**, *135*, 16133.

(48) Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2007**, *368*, 1448.

(49) Wu, K.-P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. *J. Mol. Biol.* **2009**, *391*, 784.

(50) Terakawa, T.; Takada, S. *Biophys. J.* **2011**, *101*, 1450.

(51) Knott, M.; Best, R. B. *PLoS Comput. Biol.* **2012**, *8*, e1002605.

(52) Zhang, W.; Ganguly, D.; Chen, J. *PLoS Comput. Biol.* **2012**, *8*, e1002353.

(53) Narayanan, C.; Weinstock, D. S.; Wu, K.-P.; Baum, J.; Levy, R. M. *J. Chem. Theory Comput.* **2012**, *8*, 3929.

(54) Wang, Y.; Chu, X.; Longhi, S.; Roche, P.; Han, W.; Wang, E.; Wang, J. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E3743.

(55) Mittal, J.; Yoo, T. H.; Georgiou, G.; Truskett, T. M. *J. Phys. Chem. B* **2013**, *117*, 118.

(56) Roux, B.; Weare, J. *J. Chem. Phys.* **2013**, *138*, 084107.

(57) Cavalli, A.; Camilloni, C.; Vendruscolo, M. *J. Chem. Phys.* **2013**, *138*, 094112.

(58) Im, W.; Jo, S.; Kim, T. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 252.

(59) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407.

(60) Guerry, P.; Salmon, L.; Mollica, L.; Ortega Roldan, J.-L.; Markwick, P.; van Nuland, N. A. J.; McCammon, J. A.; Blackledge, M. *Angew. Chem., Int. Ed.* **2013**, *52*, 3181.

(61) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002.

(62) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.-R.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. *Bioinformatics* **2012**, *28*, 1463.

(63) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908.

(64) Ozenne, V.; Schneider, R.; Yao, M.; Huang, J.-R.; Salmon, L.; Zweckstetter, M.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2012**, *134*, 15138.

(65) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65.

(66) Zweckstetter, M. *Nat. Protoc.* **2008**, *3*, 679.

(67) Svergun, D.; Barberato, C.; Koch, M. *J. Appl. Crystallogr.* **1995**, *28*, 768.

(68) Rambo, R. P.; Tainer, J. A. *Nature* **2013**, *496*, 477.

(69) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037.

(70) Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355.

(71) Huang, J.-R.; Gentner, M.; Vajpai, N.; Grzesiek, S.; Blackledge, M. *Biochem. Soc. Trans.* **2012**, *40*, 989.

(72) Sezer, D.; Freed, J. H.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 5755.

(73) Eyal, E.; Najmanovich, R.; McConkey, B. J.; Edelman, M.; Sobolev, V. *J. Comput. Chem.* **2004**, *25*, 712.

(74) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *J. Appl. Crystallogr.* **2003**, *36*, 1277.