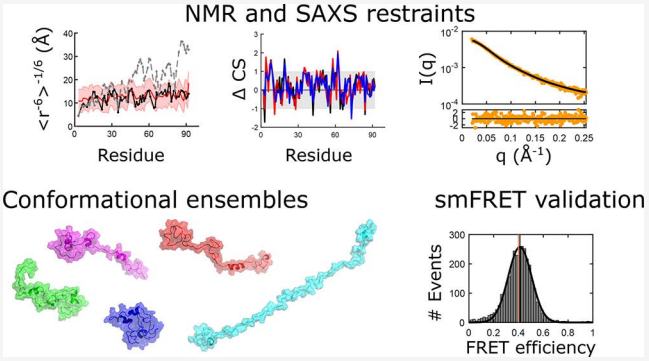
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET

Gregory-Neal W. Gomes,* Mickaël Krzeminski, Ashley Namini, Erik W. Martin, Tanja Mittag, Teresa Head-Gordon, Julie D. Forman-Kay, and Claudiu C. Gradinaru*

Cite This: J. Am. Chem. Soc. 2020, 142, 15697−15710
Read Online

ABSTRACT: Intrinsically disordered proteins (IDPs) have fluctuating heterogeneous conformations, which makes their structural characterization challenging. Although challenging, characterization of the conformational ensembles of IDPs is of great interest, since their conformational ensembles are the link between their sequences and functions. An accurate description of IDP conformational ensembles depends crucially on the amount and quality of the experimental data, how it is integrated, and if it supports a consistent structural picture. We used integrative modeling and validation to apply conformational restraints and assess agreement with the most common structural techniques for IDPs: Nuclear Magnetic Resonance (NMR) spectroscopy, Small-angle X-ray Scattering (SAXS), and single-molecule Förster Resonance Energy Transfer (smFRET). Agreement with such a diverse set of experimental data suggests that details of the generated ensembles can now be examined with a high degree of confidence. Using the disordered N-terminal region of the Sic1 protein as a test case, we examined relationships between average global polymeric descriptions and higher-moments of their distributions. To resolve apparent discrepancies between smFRET and SAXS inferences, we integrated SAXS data with NMR data and reserved the smFRET data for independent validation. Consistency with smFRET, which was not guaranteed *a priori*, indicates that, globally, the perturbative effects of NMR or smFRET labels on the Sic1 ensemble are minimal. Analysis of the ensembles revealed distinguishing features of Sic1, such as overall compactness and large end-to-end distance fluctuations, which are consistent with biophysical models of Sic1's ultrasensitive binding to its partner Cdc4. Our results underscore the importance of integrative modeling and validation in generating and drawing conclusions from IDP conformational ensembles.



1. INTRODUCTION

Under physiological conditions, the amino acid sequences of intrinsically disordered proteins (IDPs) encode for a large and heterogeneous ensemble of conformations, allowing them to perform critical biological functions.[1,2] The properties of IDP conformational ensembles are intimately related to their function in health and disease.[3] This has prompted intense efforts to develop formal and heuristic descriptions of how sequence properties relate to conformational ensembles,[4−8] and how the properties of conformational ensembles, once determined, can be mined to generate hypotheses about biological function.[9−11] Conformational ensembles are therefore central to understanding both sequence-to-ensemble and ensemble-to-function relationships in IDPs, which makes their accurate and comprehensive characterization of high importance.

To provide insights into the structural properties of IDPs, Nuclear Magnetic Resonance (NMR),[12] Small-Angle X-ray Scattering (SAXS),[13] and single-molecule Förster Resonance Energy Transfer (smFRET)[14,15] have emerged as particularly powerful techniques. Computational approaches to integrate the information from these measurements typically represent conformational ensembles as a collection of structures, each described by its atomic coordinates, and use the experimental data for constructing (e.g., restraining or reweighting) or validating the ensemble calculation.[16−18]

Despite their demonstrated complementarity,[19−24] conformational ensembles of IDPs/unfolded states which use data from all three techniques in their construction or validation are rarely reported. Aznauryan et al. reported ensembles of ubiquitin denatured in 8 M urea which are consistent with SAXS and a large number of restraints from NMR and smFRET experiments.[20] However, concerns about the mutual

Received: February 21, 2020
Published: August 25, 2020

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

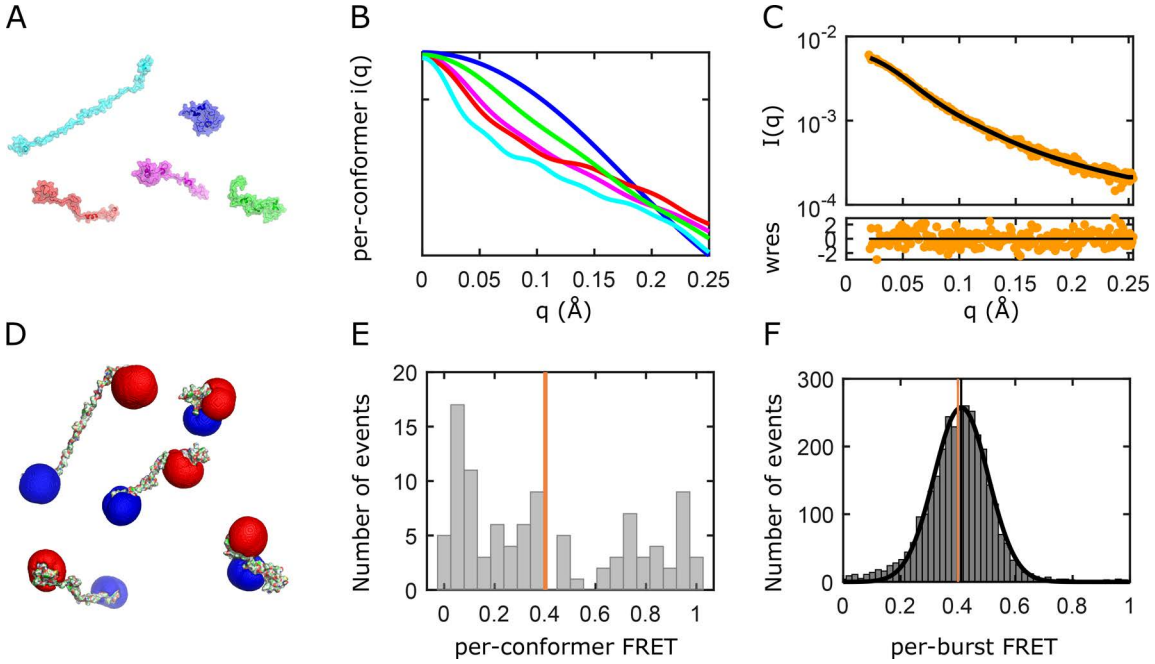


Figure 1. Schematic showing the ENSEMBLE approach for SAXS and smFRET data from an ensemble of structures. (A, B) The SAXS intensity curve of each conformation, $i(q)$, is back-calculated from the atomic coordinates using CRYSOL.[33] (C) The linear average of the CRYSOL-calculated SAXS profiles of individual conformers (black) is compared with the experimental SAXS profile (yellow). (D, E) Per-conformer FRET efficiencies are calculated assuming a quasi-static distribution of interdye distances predicted by accessible volume simulations.[34,35] (F) The ensemble-averaged transfer efficiency $\langle E \rangle_{ens}$ (orange vertical line in E and F) is compared to the mean experimental transfer efficiency $\langle E \rangle_{exp}$ (black vertical line).

consistency of smFRET and SAXS data posit that, in the absence of denaturant, the FRET fluorophores could interact with each other and/or the IDP itself.[25] Piana et al. reported ensembles of α-synuclein in physiological conditions, which are directly compared to SAXS and NMR data but are compared to distances inferred from smFRET data using an assumed homopolymer model.[26] However, it is difficult to determine which, if any, homopolymer model is appropriate for a particular heteropolymeric IDP.[27,28] Thus, the use of data from all three techniques to construct or validate conformational ensembles of an IDP (*i*) in *physiological* conditions and (*ii*) without assuming a homopolymer model would provide valuable insights into each technique's sensitivity to different aspects of IDP structure.

We therefore sought to determine conformational ensembles of an IDP in physiological conditions with conformational restraints/validation imposed by NMR, SAXS, and smFRET. Using the disordered N-terminal region of the Sic1 protein as a test case (see below), we generated new smFRET and SAXS data to complement previously published NMR data.[29,30] To combine these data sets, we used the ENSEMBLE approach (Figure 1), which selects a subset of conformations from a large starting pool of conformations to achieve agreement with experimental data.[17,31,32] Our final ensembles of Sic1 are consistent with a diverse set of experimental data suggesting that their properties can be examined with a high degree of confidence. This allowed us to examine relationships between average global polymeric descriptions of Sic1 and higher-order moments of their distributions.

The achievement of our objective of determining Sic1 ensembles consistent with all three data sets also allows us to provide additional insight into the so-called "smFRET and SAXS controversy".[36−38] Previous studies have either (i) posited attractive fluorophore interactions in the absence of denaturant[25,39] or (ii) have jointly restrained ensemble calculations using both the smFRET and SAXS data.[19,21,40] The latter approach is based on the recognition that, for heteropolymers, deviations from homopolymer chain statistics can cause smFRET and SAXS to be sensitive to different aspects of IDP structure.[21,41] For a given IDP and set of labels, both explanations for discrepant inferences are *a priori* plausible, so additional experimental information is needed. Additional experimental information in the second approach was provided by self-consistent smFRET distance inferences with labels of varying physicochemical properties[19] or self-consistent SAXS/NMR measurements of samples with and without FRET labels.[21,42] Here, instead we provide additional experimental information in the form of NMR restraints and reserve the smFRET data as an independent validation. Consistency with the smFRET data indicates that, globally, perturbative effects of PRE or smFRET labels on the Sic1 ensemble are minimal.

In yeast, the disordered protein Sic1 is eliminated via ubiquitination by the SCF^{Cdc4} ubiquitin ligase and subsequent degradation by the proteasome, allowing initiation of DNA replication.[43,44] Sic1 binding to Cdc4, the substrate recognition subunit of the ubiquitin ligase, generally requires phosphorylation of a minimum of any six of the nine Cdc4 phosphodegron (CPD) sites on (full length) Sic1. This effectively sets a high threshold for the level of active G1 CDK required to initiate transition to S-phase. This ultrasensitivity with respect to G1 CDK activity ensures a coordinated onset of DNA synthesis and genomic stability.[43] The N-terminal 90 residues of Sic1 (henceforth Sic1) are sufficient for targeting to Cdc4 when highly phosphorylated (henceforth pSic1), making this region a valuable model for

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

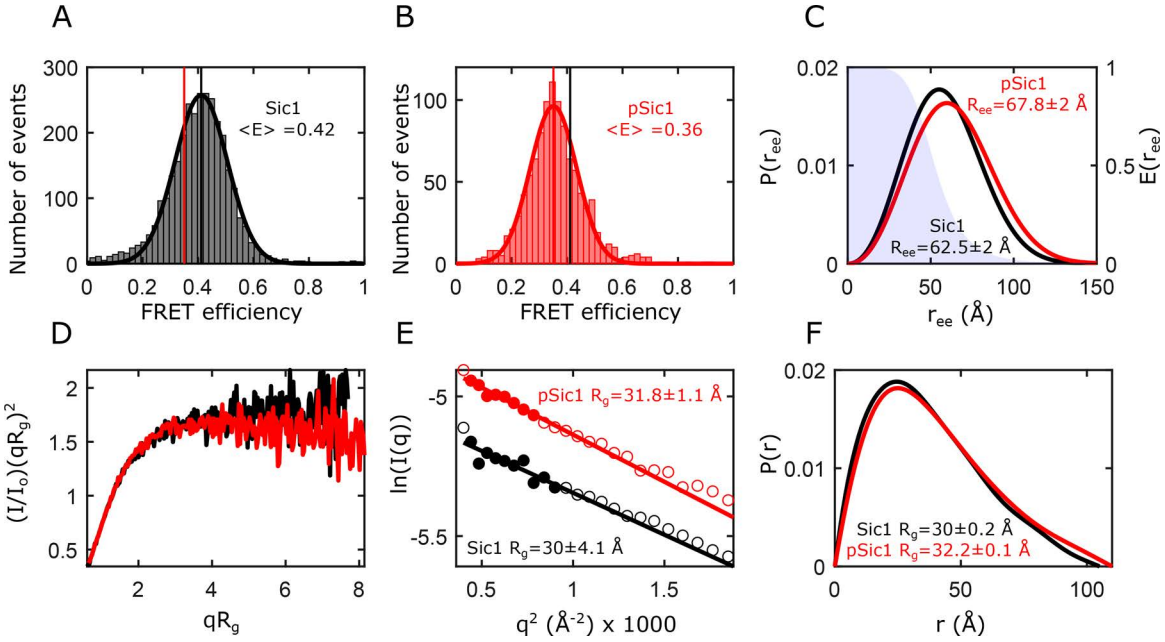


Figure 2. smFRET efficiency (E) histograms of Sic1 (A) and pSic1 (B) labeled with Alexa Fluor 488 and Alexa Fluor 647 at positions −1C and T90C in TE buffer pH 7.5 150 mM NaCl. (C) Example SAW homopolymer $P(r_{ee})$ distributions (left vertical scale) for Sic1 (black) and pSic1 (red). The shaded underlying region shows the FRET distance dependence function $E(r_{ee})$ (right vertical scale). (D) Dimensionless Kratky plots of Sic1 (black) and pSic1 (red), normalized by initial intensity I_0 and the R_g estimated from the DATGNOM[49] fit of the distance distribution function. (E) Guinier plots of Sic1 (black) and pSic1 (red). The solid circles are the data points selected for fitting a restricted range appropriate for IDPs (q_{max} R_g < 0.9) and the solid lines show the Guinier fits using these data points. (F) The normalized distance distribution function $P(r)$ estimated by DATGNOM for Sic1 (black) and pSic1 (red).

structural characterization.[45] Neither phosphorylation nor binding to Cdc4 leads to the folding of Sic1.[29,30] As the binding properties of Sic1 and pSic1 are vastly different, accurate conformational ensembles of Sic1 and pSic1 are central to the development and validation of the biophysical models of their differential binding.[46−48]

Surprisingly, previous analysis showed only subtle global changes in Sic1 upon phosphorylation, though only SAXS data were used to restrain the global dimensions. The insensitivity of global dimensions to phosphorylation is surprising given the drastic changes in charge but is consistent with proposed polyelectostatic models of ultrasensitivity.[46] These subtle changes resemble those of another yeast IDP, Ash1, and point to compensatory effects from local and long-range intrachain contacts[8] that would be difficult to quantify without an integrative approach. Our integrative modeling with new SAXS data and validation with new smFRET data allow us to examine the details of Sic1 phosphorylation at a previously unattainable level.

2. RESULTS

2.1. Measurements of R_{ee} and R_g Inferred Individually from smFRET or SAXS Provide Discrepant Descriptions of Sic1 and pSic1 Conformational Ensembles. Figure 2A−C shows smFRET data measured on the Sic1 FRET construct, which is based on Sic1(1−90) and hereafter called Sic1. This construct was labeled stochastically at its termini with the FRET donor Alexa Fluor 488 and acceptor Alexa Fluor 647 (Förster radius R_0 = 52.2 ± 1.1 Å, details in Supporting Information). The FRET histogram is fit to a Gaussian function to extract the mean transfer efficiency $\langle E \rangle_{exp}$, which reports on the interdye distance and therefore the end-to-end distance distribution $P(r_{ee})$ as a result of terminal labeling. Multisite phosphorylated Sic1 (pSic1) was generated via overnight incubation with Cyclin A/Cdk2 resulting in predominantly 6- and 7-fold phosphorylated Sic1, with a minor population of 5-fold phosphorylated Sic1 (determined by ESI mass spectrometery). Upon phosphorylation, $\langle E \rangle_{exp}$ decreases from 0.42 to 0.36 indicating chain expansion (precision ±0.005, accuracy ±0.02; see Supporting Information).

An estimate of the root-mean-squared end-to-end distance R_{ee} can be made from $\langle E \rangle_{exp}$ by assuming $P(r_{ee})$ is described by a homopolymer model (details in Supporting Information). However, the smFRET data itself does not suggest which (if any) homopolymer model is appropriate for a certain IDP. There is considerable flexibility in the choice of homopolymer model and in the rescaling of the root-mean-squared interdye distance $R_{D,A}$ to R_{ee}. This results in a range of R_{ee}, 61−65 Å for Sic1 and 66−72 Å for pSic1, which exceeds other sources of uncertainty (Table S2). If the same polymer modeling is used to analyze Sic1 and pSic1, multisite phosphorylation results in an approximately 10% increase in R_{ee}. However, the smFRET data alone cannot justify this assumption.

To infer the root-mean-squared radius of gyration R_g from R_{ee}, we require an additional assumption about the polymeric nature of the system under study, namely, the ratio $G = R_{ee}^2/R_g^2$, which cannot be determined from the smFRET experiment itself. It has recently been shown that finite-length heteropolymeric chains can take on values of G that deviate from the values derived for infinitely long homopolymers in either the θ-state (Gaussian chains, G = 6) or excluded-volume (EV)-limit (self-avoiding walks, $G \approx 6.25$).[21,40,41] Application of polymer-theoretic values of G to the smFRET inferred R_{ee} results in R_g 24−27 Å for Sic1 and 26−29 Å for pSic1 (Table S3).

Figure 2D−F shows SAXS data for Sic1 and pSic1. R_g was estimated to be approximately 30 Å for Sic1 and 32 Å for pSic1

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

Table 1. Agreement of Sic1 N_{conf} = 500 Ensembles with Experimental Data[a]

Restraints	χ^2 PRE	χ^2 $^{13}C_\alpha$ CS	χ^2 $^{13}C_\beta$ CS	χ^2 SAXS	$\langle E\rangle_{exp} - \langle E\rangle_{ens}$
TraDES RC (none)	1.51	0.514	0.518	2.03	0.12
SAXS	3.56	0.578	0.575	0.952	0.17
PRE	0.230	0.607	0.632	14.03	−0.18
SAXS+PRE	0.252	0.511	0.462	1.01	0.02
SAXS+PRE+CS	0.246	0.456	0.185	0.986	−0.003

[a] N_{conf} = 500 ensembles are derived by combining conformations from five independently calculated N_{conf} = 100 ensembles. Differences $|\langle E\rangle_{exp} - \langle E\rangle_{ens}| \leq \sqrt{\sigma^2_{E,exp} + \sigma^2_{E,ens}} \approx 0.02$ indicate no disagreement between back-calculated and experimental mean transfer efficiencies (see Materials and Methods).

using the Guinier approximation and from the distance distribution function $P(r)$ obtained using the indirect Fourier transform of the regularized scattering curve (Figure 2E,F and SI text). Though a model of chain statistics does not need to be specified, these methods are limited in describing IDPs and unfolded proteins.[13,19] For example, the expanded and aspherical conformations of IDPs lead to a reduced range of scattering angles in which the Guinier approximation can be applied without systematic error.[19] The degree of underestimation of R_g increases as the maximum scattering angle q_{max} increases, while decreasing q_{max} reduces the number of points restraining the Guinier fit, which increases the uncertainty in R_g[19] (see also, Table S4). In particular, for Sic1, the restricted range ($q_{max}R_g < 0.9$) which gives similar R_g to analysis of the full SAXS profile (see below) introduces considerable error in R_g (±4 Å).

One solution to these limitations is the modeling of the protein chain explicitly by generating ensembles of conformations. This is epitomized by the Ensemble Optimization Method (EOM)[50] and ENSEMBLE method.[31] Both approaches select a subset of conformations from an initial pool of conformations, such that the linear average of the CRYSOL-calculated SAXS profiles of individual conformers is in agreement with the full experimental SAXS profile (Figure 1A−C). However, the techniques differ in their generation of the initial pool of conformations and in the algorithm and cost-function used to minimize the disagreement with experiment (details in the Supporting Information). Despite their differences, both ensemble-based approaches fit the SAXS data equally well and resulted in nearly identical R_g values, which are similar to the "model-free" estimates (Table S5). As was seen from the smFRET data, multisite phosphorylation results in chain expansion; the SAXS data indicate an approximately 6% increase in R_g.

Riback and co-workers recently introduced another procedure for fitting SAXS data by pregenerating ensembles of conformations with different properties (specifically, the strength and patterning of inter-residue attractions) and extracting dimensionless "molecular form factors" (MFFs).[25,39] The properties of interest are then inferred from the ensemble whose MFF best fits the data. Usage of the MFFs generated from homopolymer or heteropolymer simulations resulted in R_g similar to those of the aforementioned methods (Table S6). In summary, R_g is strongly determined by the SAXS data, such that differences in the construction and refinement of models lead to minor differences in R_g.

Analysis of the full SAXS profiles using conformational ensembles shows that the smFRET and SAXS data, analyzed individually, provide discrepant inferences of Sic1 and pSic1 global dimensions. Although the various methods calculate ensembles which fit the SAXS data equally well, they have distinct values of R_{ee}, i.e., from 71 to 81 Å for Sic1 and from 71 to 87 Å for pSic1 depending on the method used (Tables S5, S6). Unlike R_g, the SAXS data do not uniquely determine R_{ee}, independent of modeling approach. Taking the ENSEMBLE SAXS-only inferred R_{ee} as representative, the inferred R_{ee} = 76.0 ± 2 Å (SEM, 5 replicates) for Sic1 is larger than the largest smFRET inferred R_{ee} = 65.4 ± 2 Å. Similarly, for homopolymer-based smFRET inferences, the largest Sic1 R_g = 26.8 ± 1.6 Å is still smaller than the SAXS inferred R_g = 30.1 ± 0.4 Å (SEM, 5 replicates) using the ENSEMBLE method.

The benefits of integrative modeling are apparent from the above analysis. Naturally, the accuracy of those aspects of the ensemble not strongly determined by the SAXS data will depend on the initial conformer generation and the optimization/selection algorithms. The wide range of SAXS-inferred R_{ee} suggests that integration of additional experimental data will improve weakly restrained structural properties, possibly reducing the discrepancy with smFRET. Likewise, $G = R^2_{ee}/R^2_g$ cannot be determined from either data set individually and must be assumed *a priori* in smFRET or influenced by assumptions inherent to each SAXS analysis method. It would therefore be desirable to back-calculate a mean FRET efficiency $\langle E\rangle_{ens}$ from a structural ensemble that is restrained by SAXS and additional experimental data, and to compare $\langle E\rangle_{ens}$ and $\langle E\rangle_{exp}$ directly. Finally, although the differences in $\langle E\rangle_{exp}$ for Sic1 and pSic1 are significant ($\Delta\langle E\rangle_{exp}$ = 0.065 ± 0.007), their R_{ee} cannot be compared with commensurate precision, since the same homopolymer model may not apply to both.

2.2. Ensembles Jointly Restrained by SAXS and NMR Data Are Consistent with Measured FRET efficiencies.

We hypothesized that jointly restraining ensembles with non-smFRET internal distance restraints and SAXS data could result in ensembles with back-calculated mean transfer efficiencies, $\langle E\rangle_{ens}$, in agreement with the experimental mean transfer efficiency $\langle E\rangle_{exp}$. In addition to independently validating the calculated ensemble, this would provide compelling evidence that the smFRET and SAXS data sets are mutually consistent.

To provide non-smFRET information for joint refinement with SAXS data, we used previously published NMR data on Sic1.[29,30] Briefly, the NMR data consist of $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts (CSs) from Sic1 and Paramagnetic Relaxation Enhancement (PRE) data from six single-cysteine Sic1 mutants using a nitroxide spin label (MTSL) coupled to cysteine residues in positions −1, 21, 38, 64, 83, and 90. We used the ENSEMBLE approach to calculate ensembles that are in agreement with the NMR and SAXS data (see Materials and

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

Methods and Supporting Information). We used fluorophore accessible volume (AV) simulations[34] to back-calculate the mean transfer efficiency $\langle E \rangle_{ens}$ from the sterically accessible space of the dye attached to each conformation via its flexible linker (details in Supporting Information). Briefly, the back-calculated $\langle E \rangle_{ens}$ are averages over the accessible interdye distances for a particular conformation, as well as averages over all conformations in an ensemble. To determine the proper time-averaging regime, we performed Monte Carlo simulations of the photon emission process and Brownian motion simulations of dye translational diffusion within the space allowed by sterics and its flexible linker. The slow interdye and end-to-end distance dynamics, relative to the donor excited state lifetime, allows $\langle E \rangle_{ens}$ to be calculated using the quasi-static averaging approximation.

Table 1 summarizes the agreement of the Sic1 ensembles under various restraint and validation combinations. The agreement of the experimental and back-calculated NMR and SAXS data were quantified using a reduced χ^2 inspired metric. This metric gives an impression of the level of agreement with the various data, though a number of assumptions required for χ^2 statistics are only approximately held. Strictly speaking, reduced $\chi^2 \sim 1$ indicates a good fit, only if the weighted residuals are standard normally distributed and the degrees of freedom can be accurately estimated. For the PRE data, we use a highly simplified treatment that restrains the distance between the C_β of the residue with the spin label to various NH positions, based on the interpretation of the loss of intensity due to an r^{-6} broadening effect (see Supporting Information). Recognizing this, we use a flat-bottom restraining potential and allow generous error margins (±5 Å); therefore, we do not expect standard normally distributed PRE residuals.

Similarly, for the CS restraints, the prediction errors derived from training and validation on folded proteins may not accurately predict errors for IDPs.[51,52] For SAXS, there are difficult to quantify back-calculation uncertainties from implicit hydration modeling.[53] For all measurements, the degrees of freedom can be smaller than the number of data points because of correlations in the data,[54] and the selection of conformers may be considered as free parameters. However, since we likely overestimate errors and degrees of freedom, $\chi^2 \gg 1$ indicates disagreement with experiment. The above concerns do not prevent the use of χ^2 for model comparison. As a structureless null-hypothesis, we also include a random coil (RC) ensemble generated with the statistical coil generator TraDES for Sic1.[55,56] Residue-by-residue fits to the NMR restraints are shown in Figures S5−S7, and fits to the full SAXS profile are shown in Figure S8.

The TraDES random coil (RC) ensemble agrees with the CS data; however, the agreement with the PRE, smFRET, and SAXS data is poor. Internal distances between specific residues are generally larger in the RC ensemble than those expected from the PRE and smFRET data. The RC ensemble has significant discrepancy in the low-q region, as it underestimates the radius of gyration (Figure S8 and Table S11). When only the SAXS data are used as a restraint, the ensemble reproduces the SAXS curve very well. However, relative to the RC ensemble, the *overall* larger inter-residue distances in the SAXS ensemble further deteriorate the agreement with data reporting on specific inter-residue distances (PRE: Figure S5 and Table 1; smFRET: Table 1 and S9).

When only the PRE data are used as a restraint, the agreement with the PRE data is achieved at the expense of disagreement with all other observables. This ensemble reproduces specific inter-residue distances encoded by the PRE data but not the overall distribution of inter-residue distances encoded by the SAXS data. A corollary of the r^{-6} PRE weighting is that the PRE ensemble average is dominated by contributions from compact conformations.[57] Consistent with this, the PRE-only ensemble is much more compact ($R_g \approx$ 22 Å) than that expected from the SAXS data. Similarly, the transfer efficiency calculated from the ensemble $\langle E \rangle_{ens}$ is larger than $\langle E \rangle_{exp}$ indicating either too short end-to-end distances overall or some conformations with strongly underestimated end-to-end distances. Although the absolute value of χ^2 suggests agreement with CSs, the PRE-only ensemble is in worse agreement with the CS data than with the TraDES RC or SAXS ensemble.

When the overall distribution of inter-residue distances from SAXS and the specific pattern of inter-residue distances from PRE are synthesized in one ensemble model, the transfer efficiency calculated from the ensemble, $\langle E \rangle_{ens}$, is in excellent agreement with the experimental transfer efficiency, $\langle E \rangle_{exp}$. The fit of the CS data (which were not used as a restraint for this ensemble) are also improved relative to the TraDES RC, the SAXS, and the PRE ensembles. As was previously observed, the generation of ensembles by satisfying tertiary structure restraints seems to place some restraints on the backbone conformations.[58] Finally, we calculated ensembles jointly restrained by SAXS, PRE, and CS data. This improves agreement with CSs, in particular, C_β CSs (Figure S7), while the agreement with the other experimental data are within the variation for SAXS+PRE calculations.

2.3. Integrative Modeling and Validation Provide a Richer Description of Global Dimensions than Those Provided by SAXS or smFRET Individually. To better understand the implications and advantages of combining multiple data sets, we calculated global descriptions of Sic1 and pSic1 conformational ensemble dimensions (radius of gyration R_g, end-to-end distance R_{ee}, and hydrodynamic radius R_h). Table S11 summarizes the global dimensions of five independently calculated ensembles with 100 conformations each (N_{conf} = 100).

The radii of gyration, including the implicit solvent layer, of the SAXS+PRE and SAXS+PRE+CS ensembles are ≈5% smaller than the SAXS-only estimates. However, no attempt was made to optimize the default solvation parameters in CRYSOL, and small differences in these parameters can result in a 5% to 10% change in R_g for the same set of protein coordinates.[53] The radius of gyration calculated directly from the C_α coordinates of the fully restrained ensembles is ≈29.5 ± 0.1 Å and ≈30.0 ± 0.1 Å for Sic1 and pSic1, respectively (SEM, 5 replicates). As Sic1 and pSic1 have larger than random coil (excluded volume) R_g (i.e., 27.9 ± 0.2 Å; Table S11), we focus on the performance of the self-avoiding walk (SAW) homopolymer models to infer end-to-end distances R_{ee}. The Gaussian chain model has a known tendency to overestimate R_{ee} when the underlying chain statistics are closer to those of an excluded volume polymer.[27,41,59] The end-to-end distance of the fully restrained/validated Sic1 and pSic1 ensembles is ≈62 ± 1 Å and ≈69 ± 1 Å, respectively (SEM, 5 replicates). The SAW homopolymer model inferences of R_{ee} agree within error, with an average percent error of 1% and −2% for Sic1 and pSic1, respectively.

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

Table 2. Nominally Universal Polymer Properties of the Calculated Ensembles[a]

		G	ρ	$\langle A \rangle$	ΔA	ΔR_{ee}
Polymer Theory	EV ($n \rightarrow \infty$)	6.254	~1.59	0.431	0.442	0.374
	EV (n = 90 − 100)	6.32	1.27−1.39	0.438	0.437	
	θ-state ($n \rightarrow \infty$)	6	~1.5	0.396		0.422
Sic1	TraDES RC	6.37	1.33	0.438	0.438	0.352
	SAXS-only	6.39	1.36	0.470	0.398	0.329
	SAXS+PRE	4.78	1.32	0.346	0.454	0.414
	SAXS+PRE+CS	4.64	1.33	0.342	0.472	0.442
pSic1	TraDES RC	6.35	1.33	0.438	0.432	0.366
	SAXS-only	5.97	1.34	0.418	0.427	0.354
	SAXS+PRE	5.26	1.31	0.369	0.428	0.398

[a]Reported values are the mean of 5 independent N_{conf} = 100 ensembles. Data are reproduced in Supporting Information (Table S12) including standard deviations (SDs) of ensemble values and references for polymer theory values. The SD range from 5 replicates for G is ±0.25−0.5, for ρ is ±0.01−0.02, for $\langle A \rangle$ is ±0.01−0.02, for ΔA is ±0.02−0.04, and for ΔR_{ee} is ±0.01−0.04.

The above analysis shows that, individually, SAXS and smFRET can accurately infer R_g and R_{ee}, respectively. However, we wish to highlight the advantages of an integrative analysis for Sic1 and pSic1. The global conformational properties of pSic1, as measured by SAXS, are very similar to those of Sic1. This is surprising, given the change in the net charge per residue from ca. 0.12 to −0.01 and −0.03 for 6- and 7-fold phosphorylated Sic1. However, this global insensitivity to phosphorylation state has been observed in a similar yeast IDP, Ash1,[8] and is required in the polyelectrostatic model of Sic1 ultrasensitive binding to Cdc4.[46] The SAXS ensembles suggest that R_{ee} is similarly insensitive to multisite phosphorylation (Table S11), while the jointly restrained SAXS+PRE ensembles show an expansion that is confirmed by a direct measurement, smFRET. Similarly, two-dimensional scaling maps (see below) point to heterogeneous changes in internal distances upon phosphorylation that could be observed/validated by future smFRET measurements.

The calculated hydrodynamic radius, R_h, was found to be highly similar for all considered ensembles ($R_h \approx 21-23$ Å). Although we can determine R_h with high precision (variation between replicates is very small <0.3 Å, Table S11), the accuracy is considerably lower. There are larger margins of error in back-calculating a dynamic quantity (R_h) from a set of static structures, and in properly modeling solvation effects. For example, calculation of R_h using the Kirkwood-Riseman approximation[60] or HYDROPRO[61] results in R_h values that differ by ca. 20%. Thus, while it is encouraging that ensemble R_h values are close to experimental values determined by NMR[30] (R_h = 21.5 ± 1.1 Å for Sic1 and R_h = 19.4 ± 1.6 Å for pSic1) and by FCS[62] (R_h = 22 ± 2 Å for Sic1), it is premature to consider this a validation of the ensemble, especially given the insensitivity of R_h to different restraint combinations (see Table S11).

2.4. Analysis of Conformational Behavior of Calculated Ensembles beyond Global Dimensions. We next sought to determine descriptions of the calculated conformational ensembles which go beyond global dimensions and enable comparison with polymer theory reference states, and with IDPs and unfolded states of varying sequence and chain length, n. To this end, we used the fact that many aspects of homopolymer behavior become universal, or independent of monomer identity, in the long chain (as $n \rightarrow \infty$) limit.[63] This allowed us to clearly identify ways in which ensembles deviate from homopolymer behavior (Table 2).

For very long homopolymer chains, the scaling exponent ν tends to one of only three possible limits (1/3, 1/2, 0.588), describing the poor-solvent, θ-state, and excluded volume (EV)-limit, respectively. Homopolymers in these limits have well-defined universal values for the size ratios $G = R_{ee}^2/R_g^2$ and $\rho = R_g/R_h$, the overall shape of the ensemble, as characterized by the average asphericity $\langle A \rangle$ ($A \sim 0$ for a sphere and $A \sim 1$ for a rod), the relative variance in the end-to-end distance distribution $\Delta R_{ee} = \sqrt{\langle r_{ee}^2 \rangle - \langle r_{ee} \rangle^2}/R_{ee}$, and the relative variance in the distribution of the shape of individual conformations $\Delta A = \sqrt{\langle A^2 \rangle - \langle A \rangle^2}/\langle A \rangle$. Table 2 summarizes the universal values expected for homopolymers in the θ-state or the EV-limit, in the case of very long chains (EV and θ-state $n \rightarrow \infty$) and for chains with similar length to Sic1 (EV n = 90−100). The TraDES random coil, though not a homopolymer, is constructed with only excluded volume long-range interactions; therefore it is expected to have behavior consistent with polymer theory predictions for an EV-limit polymer of similar chain-length (EV n = 90−100 Table 2).

The values of G for the Sic1 and pSic1 random coil and SAXS-only ensembles are indistinguishable from the expected value for a homopolymer in the EV-limit ($G \approx 6.3$). In contrast, ensembles jointly restrained by SAXS and PRE have G outside the range $G_\theta = 6 \leq G \leq G_{EV} \approx 6.3$ despite having apparent scaling exponents between the θ-state and EV-limits (see below). For Sic1, G = 4.7 ± 0.1, and for pSic1, G = 5.3 ± 0.1 (SEM, 5 replicates). The ratio ρ is not sensitive to deviations from homopolymer statistics at long sequence separations. The value of ρ remains ~1.3 for all ensembles inspite of large changes in R_{ee} and G. The calculated ρ values are consistent with the range of polymer-theoretic values for a finite length EV homopolymer (EV n = 90−100 Table 2).

The Sic1 and pSic1 RC and SAXS-only ensembles have an average asphericity $\langle A \rangle$ very close to the polymer-theoretic value for a homopolymer in the EV-limit. Although individual conformations are not necessarily spherical, SAXS+PRE ensembles of both Sic1 and pSic1 are on average more spherical, with significantly lower $\langle A \rangle$, despite their larger-than-RC R_g. Similar to G, the values of $\langle A \rangle$ for the SAXS+PRE ensembles (0.346 ± 0.005 for Sic1 and 0.369 ± 0.005 for pSic1, SEM 5 replicates) are outside of the θ-state and EV-limit.

The relative variance in the end-to-end distance distribution, ΔR_{ee}, is close to the EV-limit value ($\Delta R_{ee}^{EV} \approx 0.37$) for the random coil and SAXS-only restrained ensembles. In contrast,

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

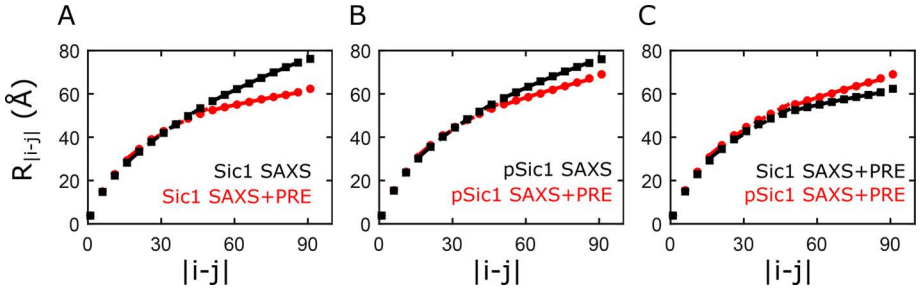


Figure 3. Internal scaling profiles calculated from 5 N_{conf} = 100 ensembles. (A) Sic1 SAXS+PRE ensembles (red circles) and Sic1 SAXS-only ensembles (black squares). (B) pSic1 SAXS+PRE ensembles (red circles) and pSic1 SAXS-only ensembles (black squares). (C) pSic1 (red circles) and Sic1 (black squares) SAXS+PRE ensembles. For all panels, fits are shown for to intermediate (dashed) and long (solid) sequence separations. For visualization, every fifth data point is shown.

Table 3. Fitting Results for the TraDES RC Ensemble, SAXS-Only Ensemble, and SAXS+PRE Ensembles Internal Scaling Profiles[a]

		Sic1		pSic1	
	TraDES RC[b]	SAXS-only	SAXS+PRE	SAXS-only	SAXS+PRE
ν (fixed l_p = 4 Å)	0.570	0.589	0.567	0.596	0.583
ν_{int}	0.566	0.601	0.524	0.569	0.517
ν_{long}	0.51	0.52	0.28	0.47	0.38
$\Delta\nu_{ends}$	−0.06 (0.03)	−0.08 (0.01)	−0.25 (0.04)	−0.09 (0.05)	−0.13 (0.03)

[a]Table results are the mean results from fitting 5 N_{conf} = 100 ensembles. SEM for ν and ν_{int} is ≈0.005 and 0.03 for ν_{long}. SEM for $\Delta\nu_{ends}$ shown in parentheses. See Materials and Methods for additional details. [b]Sic1 TraDES RC and pSic1 TraDES RC result in nearly identical fits.

Sic1 and pSic1 SAXS+PRE restrained ensembles have ΔR_{ee} values which are more consistent with the θ-state value. Although Sic1 and pSic1 R_{ee} are more compact than the RC, they exhibit larger relative variations in the end-to-end distances of their conformations. All ensembles have a relative variance in the distribution of shapes, ΔA, similar to that of an EV-limit homopolymer. The broadness of the shape distribution stresses the fact that despite being, as an ensemble of conformations, more spherical than an EV polymer, the Sic1 ensembles contains individual conformations with a large distribution of shapes.

2.5. Internal Scaling Profiles and Apparent Scaling Exponents. Recently, the focus of the smFRET and SAXS debate has moved from inferring R_g to inferring apparent scaling exponents.[25,40] To extract further insights regarding the effects of combining multiple solution data types on the statistics of internal distances in the ensembles, we calculated internal scaling profiles (ISPs, Figure 3). ISPs quantify the mean internal distances ($R_{|i-j|} = \langle\langle r_{ij}^2\rangle\rangle^{1/2}$) between all pairs of residues that are $|i - j|$ residues apart in the linear amino acid sequence (see Materials and Methods). The dependence of $R_{|i-j|}$ on sequence separation $|i - j|$ is often quantified by fitting to the power-law relation

$$R_{|i-j|} = \sqrt{2l_p b}\,|i - j|^{\nu} \quad (1)$$

where b = 3.8 Å is the distance between bonded C_α atoms, and $l_p \approx 4$ Å is the persistence length. This persistence length was found to be applicable to a broad range of denatured and disordered states.[5,21,64] Scaling laws are derived for homopolymers in the infinitely long-chain limit. For a finite-length heteropolymer, we measure merely an apparent scaling exponent ν_{app}; however, we drop the subscript to aid the clarity of the text.

ISPs highlight important differences between ensembles. If the majority of internal distances are similar in two ensembles, their R_g values will be similar, as $R_g = \sqrt{\frac{1}{2n^2}\sum_{ij}^{n}\langle r_{ij}^2\rangle}$.[21] However, if their spatial separations start to diverge at long sequence separations, the ensembles will have dissimilar R_{ee} and $\langle E\rangle_{exp}$ when terminally labeled. This decoupling of R_{ee} from R_g is illustrated by Figure 3A which shows the scaling of the SAXS-only and SAXS+PRE Sic1 ensembles, which have similar R_g, but only the SAXS+PRE ensemble is consistent with the smFRET data.

We quantify the change in scaling behavior at long sequence separations (ν_{long}, $51 < |i - j| \leq n_{res} - 5$) relative to intermediate sequence separations (ν_{int}, $15 \leq |i - j| \leq 51$) by calculating $\Delta\nu_{ends} = \nu_{long} - \nu_{int}$ (5 replicates with N_{conf} = 100, Table 3). For homopolymers in the long-chain limit we expect $\Delta\nu_{ends} = 0$; though of finite-length, the Sic1 and pSic1 random coil ensembles have $\Delta\nu_{ends} \approx 0$ ($\Delta\nu_{ends} = -0.06 \pm 0.03$, SEM 5 replicates). For Sic1, both the SAXS and SAXS+PRE ensembles show $\Delta\nu_{ends} < 0$, though the deviation from homopolymer statistics is stronger in the SAXS+PRE ensembles ($\Delta\nu_{ends} = -0.08 \pm 0.01$ and $\Delta\nu_{ends} = -0.25 \pm 0.04$, respectively, SEM 5 replicates). Internal distances in the Sic1 SAXS+PRE ensemble follow marginally good-solvent scaling at intermediate sequence separations, and transition to poor solvent scaling at larger sequence separations. Expansion of Sic1 upon phosphorylation has been attributed to transient tertiary contacts involving nonphosphorylated CPDs that are lost or weakened upon phosphorylation.[29] Consequently, while pSic1 SAXS ensembles do not identify deviations from homopolymer statistics ($\Delta\nu_{ends} = -0.09 \pm 0.05$), pSic1 SAXS +PRE ensembles identify smaller deviations than those observed for Sic1 ($\Delta\nu_{ends} = -0.13 \pm 0.03$).

Further, we compared scaling behavior determined by our integrative approach to recently published methods based only on SAXS data using "molecular form factors" (MFFs)[25,39] (Table S6) or only on smFRET data using the SAW-ν method[59] (Table S9). For Sic1, but not pSic1, there is

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

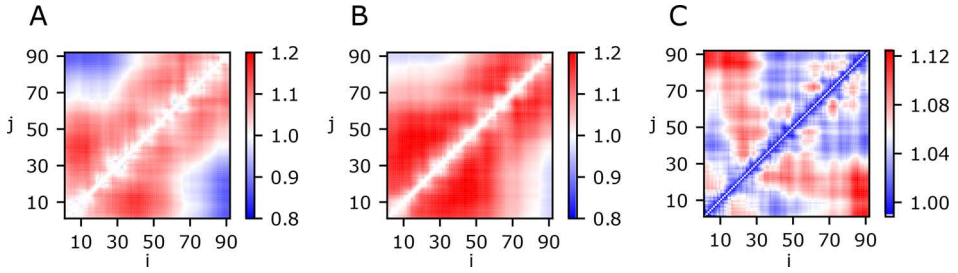


Figure 4. (A) Sic1 2D scaling map $\alpha_{ij} = \langle r_{ij}\rangle_{ens}/\langle r_{ij}\rangle_{RC}$ using the Sic1 (SAXS+PRE) N_{conf} = 500 and the Sic1 N_{conf} = 500 TraDES RC ensemble. (B) pSic1 2D scaling map $\alpha_{ij} = \langle r_{ij}\rangle_{ens}/\langle r_{ij}\rangle_{RC}$ using the pSic1 (SAXS+PRE) N_{conf} = 500 and the pSic1 N_{conf} = 500 TraDES RC ensemble. (C) pSic1 normalized by Sic1 dimensions.

agreement between global scaling determined from the SAXS data and the global scaling determined from the SAXS+PRE ensemble scaling profiles. Due to the terminal labeling positions and because $\Delta\nu_{ends} < 0$, ν inferred from smFRET is less than ν inferred from SAXS. However, neither approach using a single data type fully captures the heteropolymeric behavior of Sic1 and pSic1.

2.6. Two Dimensional Scaling Maps Reveal Regional Biases for Expansion and Compaction. To better describe the heteropolymeric nature of Sic1, a normalized two-dimensional (2D) scaling map was constructed (Figure 4). In the first step, the ensemble-averaged distances between the C_α atoms of every unique pair of residues in the sequence is calculated for the experimentally restrained ensemble ($\langle r_{ij}\rangle_{ens}$) and for the respective TraDES random coil (RC) ensemble ($\langle r_{ij}\rangle_{RC}$). Experimentally restrained distances are normalized by the RC distances and displayed as a 2D scaling map.

The normalized 2D scaling map for Sic1 (Figure 4A) displays regional biases for expansion ($\alpha_{ij} > 1$) and compaction ($\alpha_{ij} < 1$). Short internal distances $|i - j| \lesssim 45$ show expansion relative to the RC, while $|i - j| \gtrsim 60$ show compaction. The expansion, however, is heterogeneous. For example, the ∼40 residue N-terminal region is more expanded than the ∼40 residue C-terminal region. Similar distinctions between the RC and pSic1 ensembles were observed (Figure 4B).

To compare Sic1 and pSic1 ensembles, the pSic1 ensemble was normalized by the Sic1 ensemble (Figure 4C). This map describes the heterogeneous modulation of Sic1 dimensions upon multisite phosphorylation. Expansion is clustered around CPD sites, particularly those of the C-terminus and in the vicinity of Y14, previously implicated in tertiary interactions with CPDs[30] (see below).

2.7. Y14A Mutation and Phosphorylation Disrupt Tertiary Contacts in Sic1. We next sought to determine whether specific long-range interactions leading to compact end-to-end distances in Sic1 and pSic1 could be identified and disrupted. PRE effects link CPDs with Y14 and ^{15}N relaxation experiments on Sic1 identified maxima in the R_2 rates near Y14.[30] Furthermore, the substitution Y14A led to an expansion in R_h of ∼20% in pSic1.[30] We hypothesized that if Y14 engages in specific π-π and cation-π interactions throughout the chain, then removal of its π-character by mutation to alanine will disrupt these interactions, leading to larger R_{ee} and lower $\langle E\rangle_{exp}$.

We performed smFRET experiments for the Y14A mutants of Sic1 and pSic1 (Figure 5 and Table S9). Y14A mutation decreases Sic1 $\langle E\rangle_{exp}$ by approximately 7% (ca. 0.42 to 0.40, a small but reproducible shift). Phosphorylation of the Y14A mutant decreases its $\langle E\rangle_{exp}$ by approximately 16% (ca. 0.40 to 0.33). At this time, we cannot rule out that the observed FRET changes may be due (in part) to a different phosphorylation pattern for the Y14A mutant. However, these experiments suggest that the π-group of Y14 participates in long-range contacts which maintain more compact R_{ee} in Sic1 and pSic1 than those expected for a homopolymer with similar R_g. These contacts are likely key for the globally compact conformations required in the polyelectrostatic model of pSic1:Cdc4 binding.[46] This demonstrates how smFRET can be used to test structural hypotheses generated from integrative modeling.

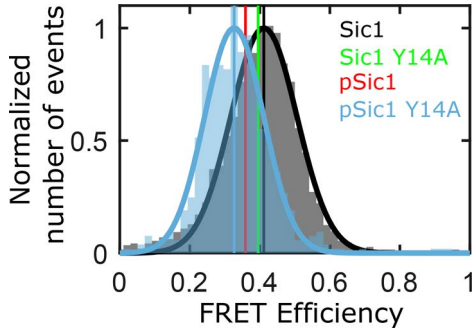


Figure 5. Y14A mutation and phosphorylation results in a shift to lower $\langle E\rangle_{exp}$ (more expanded conformations). Each histogram is normalized so that each Gaussian fit has a maximum of one.

3. DISCUSSION

We generated SAXS and smFRET data on Sic1 and pSic1 and resolved their apparently discrepant inferences by joint refinement of the SAXS data with PRE data. The ensembles restrained by SAXS and PRE data are, in addition, consistent with the smFRET data, chemical shift data, and hydrodynamic data (PFG-NMR and FCS). We used smFRET transfer efficiencies directly as validation rather than derived distances from the data via polymer theory assumptions. Our final ensembles of Sic1 and pSic1 can be examined with a high degree of confidence given their agreement with a diverse set experimental data acting as both restraints and validation. This was important since the changes in Sic1 upon phosphorylation are quite subtle.

The picture that emerges when the entirety of the experimental data on Sic1 and pSic1 is considered is that their conformational ensembles cannot be described by statistics derived for infinitely long homopolymers. Although this is unsurprising, given that Sic1 and pSic1 are finite-length heteropolymers, ensembles restrained only by the SAXS data are congruent with the set of homopolymer descriptions and scaling relationships for excluded volume homopolymers.

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

Neither the SAXS nor smFRET data, individually, suggest deviations from homopolymer statistics. Our results therefore provide a strong impetus for integrative modeling and validation approaches over homopolymer approaches whenever multiple data types exist. On this note, it is important to acknowledge that quantitative interpretation of PREs for IDPs within such integrative approaches is challenging due to the convolution of distance information with dynamics.[65] The linking of interpretation of PRE data to molecular dynamics simulation[66,67] or the combination of NMR relaxation data with the explicit modeling of the spin label[68] are likely to improve structural inferences, albeit at a higher computational cost.

We emphasize that the SAXS+PRE ensembles were not constructed by reweighting or selecting ensembles specifically to achieve agreement with $\langle E \rangle_{exp}$. In our approach, it was not guaranteed *a priori* that $\langle E \rangle_{ens}$ would match $\langle E \rangle_{exp}$, especially if either the introduction of PRE spin labels or smFRET fluorophores had perturbed the IDP ensemble. If negative values of $\Delta\nu_{ends}$ are common for IDPs and for unfolded proteins under refolding conditions, smFRET on terminally labeled samples will infer smaller ν than would SAXS. When experiments are analyzed individually, $\Delta\nu_{ends} < 0$ is consistent with both fluorophore-driven interactions and heteropolymer effects. In both cases, $\Delta\nu_{ends}$ would approach zero in high concentrations of denaturant, which would disrupt both spurious fluorophore interactions and native long-range tertiary interactions.[21,25] An integrative approach is required to decide between fluorophore-interactions and heteropolymer effects. An additional consistency check is to measure $\langle E \rangle_{exp}$ for a different pair of dyes with different physicochemical properties. In a previous publication,[62] we measured $\langle E \rangle_{exp}$ for Sic1 using smaller, less charged, and more hydrophobic dyes (TMR and Atto647N). We recalculated the expected $\langle E \rangle_{ens}$ for the N_{conf} = 500 Sic1 SAXS+PRE ensemble with TMR and Atto647N accessible volumes (Table S7) and the measured Förster radius R_0 = 60 ± 2 Å for this pair. The resulting $\langle E \rangle_{ens}$ = 0.51 ± 0.02 agrees with the measured $\langle E \rangle_{exp}$ = 0.47 ± 0.02 (Figure 7D of ref 62).

3.1. Conformation-to-Function Relationships. For soluble post-translationally modified IDPs, approximate, good-solvent scaling may be unsurprising. The balance between chain−chain and chain-solvent interactions is a driving force for aggregation[69] and phase separation,[70,71] and polymer theory predicts that proteins with overall good-solvent scaling in native-like conditions should remain soluble. At short-to-intermediate sequence separations, good-solvent scaling provides read/write access of substrate motifs to modifying enzymes (e.g., phosphorylation and ubiquitination for Sic1).

Good-solvent scaling also confers advantages to dynamic complexes, as internal friction increases with increasing chain compaction.[72] Low internal friction and fast chain reconfigurations provide more opportunities for unbound Cdc4 phosphodegrons (CPDs) to (re)bind before pSic1 diffuses out of proximity of Cdc4.[47,48,73] In the polyelectrostatic model, fast reconfiguration dynamics facilitate pSic1's dynamic interactions with Cdc4 through electrostatic averaging effects.[30,46]

The crossover to poor-solvent scaling at long sequence separations implies that unbound CPDs that are sequence-distant from a bound CPD are on average closer to the WD40 binding pocket than they would be for an EV-chain. This effectively decreases the solvent screening of electrostatic interactions and is predicted to lead to sharp transitions in the fraction bound with respect to the number of phosphorylations.[46] An increase of the effective concentration of CPDs in the vicinity of the binding pocket may also increase the probability for any CPD to rebind before diffusive exit.

Large amplitude fluctuations in the shape (ΔA) and size (ΔR_{ee}) of Sic1, effectively and rapidly sampling many different conformations, could allow CPDs in Sic1 to rapidly sample either the primary or secondary WD40 binding pocket. These fluctuations could facilitate electrostatic averaging, permitting a mean-field treatment as assumed in the polyelectrostatic model.[46]

4. CONCLUSIONS

Our work provides a description of the conformational ensembles of Sic1 and pSic1 which is consistent with experimental data reporting on a wide range of spatial and sequence separation scales, with biophysical models for Sic1 function. Our results show that there are clear advantages to combining multiple data sets and quantitative polymer-physics-based characterization of experimentally restrained ensembles allows the description and classification of IDPs as heteropolymers. The chain length independence of many of these properties facilitates comparison between different IDPs and unfolded states of proteins.

Our results suggest that, for Sic1 and our dye pair, discrepant inferences between SAXS and smFRET cannot *a priori* be assumed to arise from "fluorophore-interactions." The impact of the fluorophores (or spin-labels) will of course depend on the physicochemical properties of the specific IDP sequence and the fluorophores (or spin-labels) used. Robustness to perturbation (e.g., labels or phosphorylation) may be built into Sic1's sequence via its patterning of charged and proline residues.[8] Further understanding of the discriminatory power of FRET, and the utility of different restraint types for characterizing types of structure in IDPs, will come from recently developed Bayesian procedures.[74,75] In this regard, an integrative use of multiple experiments probing disparate scales, computational modeling, and polymer physics, will provide valuable insights into IDPs/unfolded states and their biological functions.

5. MATERIALS AND METHODS

5.1. Sic1 Samples. The Sic1 N-terminal region (1−90, henceforth Sic1) was expressed recombinantly as a Glutatione S-transferase (GST) fusion protein in *Escherichia coli* BL21 (DE3) codon plus cells and purified using glutathione-Sepharose affinity chromatography and cation-exchange chromatography. The correct molecular mass of the purified protein was verified by electrospray ionization mass spectrometry (ESI-MS). A double cysteine variant of Sic1 (−1C-T90C) for smFRET experiments was generated via site directed mutagenesis from a single-cysteine mutant produced previously for PRE measurements.[29,30] This construct was purified as above, and the correct molecular mass of the purified protein was verified by ESI-MS. A Y14A mutant Sic1 (-1C-T90C-Y14A) was generated via site directed mutagenesis from the aforementioned double-cysteine mutant and was expressed, purified, and characterized using the same protocol. Phosphorylated samples were prepared by treatment of Sic1 with Cyclin A/Cdk2 (prepared according to Huang et al.,[76]) at a kinase:Sic1 ratio of 1:100 in the presence of 50 fold excess of ATP and 2.5 mM $MgCl_2$ overnight at 30 °C. The yield of phosphorylation reaction was determined by ESI-MS. Under these conditions, the dominant species are 6- and 7-fold phosphorylated Sic1 (10195 and 10274 Da, respectively) with a small fraction of 5-fold phosphorylated

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

Sic1. After phosphorylation, the samples were buffer exchanged into PBS buffer pH 7.4 with 3 M GdmCl to prevent aggregation, denature kinase, and denature any phosphatases which may have inadvertently entered the solution. The samples were kept on ice in 4 °C and measured within 24 h.

The Sic1 smFRET construct was labeled stochastically with Alexa Fluor 488 C_5 Maleimide (ThermoFisher Scientific, Invitrogen, A10254) and Alexa Fluor 647 C_2 Maleimide (ThermoFisher Scientific, Invitrogen, A20347). After being labeled with Alexa Fluor 647, cation-exchange chromatography was used to separate species with a single acceptor label, from doubly acceptor labeled and unlabeled species. The single-labeled species sample was then labeled with Alexa Fluor 488, and cation-exchange chromatography was used to separate doubly heterolabeled from acceptor only species. The correct mass of the doubly labeled sample was confirmed by mass spectrometry. The final FRET labeled sample was concentrated and buffer exchanged into PBS buffer pH 7.4 with 3 M GdmCl and 2 mM DTT and stored at −80 °C. Additional details regarding protein expression, purification, and labeling are available in the Supporting Information.

5.2. Single-Molecule Fluorescence. Single-molecule fluorescence experiments were performed on a custom-built multiparameter confocal microscope with microsecond alternating laser excitation. This instrumentation allows the simultaneous detection of the intensity, anisotropy, lifetime, and spectral properties of individual molecules and for the selection of fluorescence bursts in which both dyes are present and photophysically active. The acquired data were subjected to multiparameter fluorescence analysis[77,78] and ALEX filtering.[79] The burst search was performed using an All Photon Burst Search (APBS)[80,81] with M = 10, T = 500 μs, and L = 50. Transfer efficiencies were determined burst-wise and corrected for differences in the quantum yields of the dyes and detection efficiencies, as described in further detail in the Supporting Information.

Immediately prior to measurement, samples were diluted to ∼50 pM in either (i) PBS buffer: 10 mM sodium phosphate and 140 mM NaCl pH 7.0, 1 mM EDTA (to replicate NMR measurement buffer of ref[29]) or (ii) Tris buffer: 50 mM Tris and 150 mM NaCl, pH 7.5. (to replicate SAXS measurement buffer). No difference in $\langle E \rangle_{exp}$ was detected when comparing buffer conditions, and results are shown for Tris buffer conditions. Dilution of the smFRET samples from stock concentration in 3 M GdmCl to single-molecule concentration results in approximately 60 nM residual concentration of GdmCl. Additionally, the SAXS measurements include 5 mM DTT and 2 mM TCEP to scavenge radicals and prevent radiation damage, which are detrimental to fluorophore performance, while the smFRET measurements use 143 mM 2-mercaptoethanol (BME, 1:100 v/v dilution) and 5 mM 2-mercaptoethylamine (MEA) for photoprotection and increased brightness. The smFRET samples also contain 0.001% Tween 20 for surface passivation.

The Förster radius R_0 was calculated assuming a relative dipole orientation factor $\kappa^2 = 2/3$ and the refractive index of water n = 1.33. The assumption of $\kappa^2 = 2/3$ is supported by subpopulation-specific steady-state anisotropies for the donor in the presence of the acceptor (Table S1). The overlap integral J was measured for each sample and was found not to change upon phosphorylation or Y14A mutation. The minimal variation in donor-only lifetimes τ_{D0} suggested minimal variation in the donor-quantum yield ϕ_D. R_0 was therefore calculated to be $R_0 = 52.2 \pm 1.1$ Å for all samples and variation between samples within this uncertainty.

We estimate the precision for $\langle E \rangle_{exp}$ to be ca. 0.005 (for measurements performed on the same day, with approximately equal sample dependent calibration factors). We estimate the accuracy of $\langle E \rangle_{exp}$, $\sigma_{E,exp}$, to be ca. 0.02 (due to uncertainty in the instrumental and sample dependent calibration factors). Further details about the instrumentation, photoprotection, laser excitations, burst detection, filtering, and multiparameter fluorescence analysis can be found in the Supporting Information.

5.3. Small-Angle X-ray Scattering. Small angle X-ray scattering data were collected at beamline 12-ID-B at the Argonne National Laboratory Advanced Photon Source. Protein samples were freshly prepared using size exclusion chromatography (GE Life Sciences, Superdex 75 10/300 GL) in a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT, and 2 mM TCEP. Fractions were loaded immediately after elution without further manipulation. Buffer collected one column volume after protein elution from the column was used to record buffer data before and after each protein sample. SAXS data were acquired manually; protein samples were loaded and then gently refreshed with a syringe pump to prevent X-ray damage. A Pilatus 2 M detector provided q-range coverage from 0.015 to 1.0 Å^{-1}. Wide-angle X-ray scattering data were acquired with a Pilatus 300k detector and had a q range of 0.93−2.9 Å^{-1}. Calibration of the q-range was performed with a silver behenate sample. Twenty sequential images were collected with 1 s exposure time per image with each detector. Data were inspected for anomalous exposures, and mean buffer data were subtracted from sample data using the WAXS water peak at q ∼ 1.9 Å^{-1} as a subtraction control. Details about the SAXS data analysis can be found in the Supporting Information.

5.4. ENSEMBLE. ENSEMBLE 2.1[31] was used to determine a subset of conformations from an initial pool of conformers created by the statistical coil generator TraDES.[55,56] All modules were given equal rank, and all other ENSEMBLE parameters were left at their default values.

To achieve a balance between the concerns of overfitting (under-restraining) and under-fitting (over-restraining), we performed multiple independent ENSEMBLE calculations with 100 conformers, N_{conf} = 100, as suggested by ref 58, and averaged the results from independent ensemble calculations or combined them to form ensembles with larger numbers of conformers (e.g., N_{conf} = 500). To address the possibility that changing the ensemble size could affect the structural properties of the ensemble, or its agreement with experimental observables, we reperformed the Sic1 SAXS+PRE ensemble calculations but varied the ensemble size, N_{conf} (details in the Supporting Information). The determination of polymer properties and the agreement with experimental observables is robust in a range of N_{conf} from ca. 50−100. Below $N_{conf} \approx 50$, agreement with restraining data (SAXS and PRE) is worsened, and the ensembles do not agree with validating data (smFRET and CSs). Above $N_{conf} \approx 150$, ensembles are in agreement with the experimental observables, though increased ensemble-to-ensemble variation suggests that 5 replicates (independently calculated ensembles with same set of restraints) are insufficient to ensure convergence. Larger ensembles are calculated faster (>72 h for N_{conf} = 20 vs ca. 1 h for N_{conf} = 100). Ensembles with 100 conformers were chosen to minimize the computational cost per ensemble calculation and ensemble-to-ensemble variation.

NMR data were obtained from BMRB accession numbers 16657 (Sic1) and 16659 (pSic1).[29] A total of 413 PRE restraints were used with a typical conservative upper- and lower-bound on PRE distance restraints of ±5 Å.[57,82] This tolerance was used in computing the χ^2 metric for the PRE data. CSs were back-calculated using the SHIFTX calculator,[51] and a total of 90 C_α CSs and 85 C_β CSs were used. The CS χ^2 metric was computed using the experimental uncertainty σ_{exp} and the uncertainty in the SHIFTX calculator (σ_{SHIFTX} = 0.98 ppm for C_α CSs and σ_{SHIFTX} = 1.10 ppm for C_β CSs[51]). CRYSOL[33] with default solvation parameters was used to predict the solution scattering from individual structures from their atomic coordinates. A total of 235 data points from q = 0.02 to q = 0.254 Å^{-1} were used in SAXS-restrained ensembles. The SAXS χ^2 metric was computed using the experimental uncertainty in each data point.

Accessible volume (AV) simulations[34,35] were used to predict the sterically accessible space of the dye attached to each conformation via its flexible linker (Figure 1D). These calculations were performed using the AvTraj[34] v0.0.9 and MDTraj[83] v1.9.3 packages in Python 3.7.6. In the quasi-static approximation, the interdye distance dynamics within the AVs for a particular conformation are quasi-static on the time scale of the donor excited state ($\tau_{DA} \leq \tau_{D0}$ = 3.7 ns). The per-conformer mean FRET efficiency is therefore $e = \int E(r_{DA}) P(r_{DA}) dr_{DA}$, where $P(r_{DA})$ is the distribution of interdye distances resulting from the AV simulation for a particular conformation and $E(r_{DA}) = (1 + (r_{DA}/R_0)^6)^{-1}$. End-to-end distance reconfiguration

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

times for IDPs and unfolded proteins are typically in the range 50–150 ns,[84] and so the end-to-end distance is also quasi-static on the time scale of τ_{DA}. The back-calculated ensemble-averaged $\langle E\rangle_{ens}$ is calculated as the linear average of the per-conformer FRET efficiencies $\langle E\rangle_{ens} = \langle e\rangle$. The quasi-static approximation gives the same $\langle E\rangle_{ens}$ within error as a more computationally demanding method which considers Monte Carlo simulations of the photon emission process and Brownian motion simulations of dye translation diffusion within the accessible volume (detail in the Supporting Information). Further support for the quasi-static averaging approach used comes from multiparameter E vs τ_{DA} histograms (Figure S2) which provide complementary information on interdye distances and dynamics but with different experimental integration times.

The uncertainty in $\langle E\rangle_{ens}$, $\sigma_{E,ens}$, is ca. 0.01, which is a combination of SEM and uncertainty in R_0. Differences $|\langle E\rangle_{exp} - \langle E\rangle_{ens}| \leq \sqrt{\sigma^2_{E,exp} + \sigma^2_{E,ens}} \approx 0.02$ indicate no disagreement between back-calculated and experimental mean transfer efficiencies. A comprehensive description of the ENSEMBLE calculations, restraints, and back-calculations can be found in the Supporting Information.

5.5. Polymer Scaling Analysis. The distance $R^2_{|i-j|} = \langle\langle r^2_{ij}\rangle\rangle_{ens}$ between C_α atoms is an average first over all pairs of residues that are separated by $|i - j|$ residues and then over all conformations in the ensemble. The apparent scaling exponent ν was estimated by fitting an ISP calculated for each N_{conf} = 100 ensemble to the following expression

$$\ln(R_{|i-j|}) = \nu \ln(|i - j|) + A_0 \quad (2)$$

Equation 2 is derived for homopolymers in the infinitely long chain limit. Following Peran and co-workers,[40] for finite-length chains, a lower bound of $|i - j| > 15$ was used to exclude deviations from infinitely long-chain scaling behavior at short sequence-separations, and an upper bound of $|i - j < |n_{res} - 5$ was used to exclude deviations due to "dangling ends." With these restrictions, finite-length homopolymers are expected to be well fit by eq 2. Evenly spaced points in log–log space were used during fitting. Fitting of the the entire $15 < |i - j| < n_{res} - 5$ range was used to obtain ν. A_0 was either fixed at log (5.51) (l_p = 4 Å) or left as a free fitting parameter. To test for differences in scaling behavior at intermediate and long sequence separations, the $15 < |i - j < |n_{res} - 5$ range was evenly divided into intermediate ν_{int} ($15 \leq |i - j| \leq 51$) and long ν_{long} regimes ($51 < |i - j| \leq n_{res} - 5$).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.0c02088.

Extended description of smFRET and SAXS experiment/analysis, ENSEMBLE methods, and additional tables (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Gregory-Neal W. Gomes − *Department of Physics, University of Toronto, Toronto, Ontario M5G 1X8, Canada; Department of Chemical and Physical Sciences, University of Toronto Mississauga, Mississauga, Ontario L5L 1C6, Canada;* Email: gregory.gomes@mail.utoronto.ca

Claudiu C. Gradinaru − *Department of Physics, University of Toronto, Toronto, Ontario M5G 1X8, Canada; Department of Chemical and Physical Sciences, University of Toronto Mississauga, Mississauga, Ontario L5L 1C6, Canada;* orcid.org/0000-0002-1897-313X; Email: claudiu.gradinaru@utoronto.ca

Authors

Mickaël Krzeminski − *Molecular Medicine Program, Hospital for Sick Children, Toronto, Ontario M5S 1A8, Canada; Department of Biochemistry, University of Toronto, Toronto, Ontario M5G 1X8, Canada*

Ashley Namini − *Department of Physics, University of Toronto, Toronto, Ontario M5G 1X8, Canada; Department of Chemical and Physical Sciences, University of Toronto Mississauga, Mississauga, Ontario L5L 1C6, Canada*

Erik W. Martin − *Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, United States;* orcid.org/0000-0002-8938-5001

Tanja Mittag − *Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, United States;* orcid.org/0000-0002-1827-3811

Teresa Head-Gordon − *Departments of Chemistry, Bioengineering, Chemical and Biomolecular Engineering University of California, Berkeley, California 94720, United States;* orcid.org/0000-0003-0025-8987

Julie D. Forman-Kay − *Molecular Medicine Program, Hospital for Sick Children, Toronto, Ontario M5S 1A8, Canada; Department of Biochemistry, University of Toronto, Toronto, Ontario M5G 1X8, Canada;* orcid.org/0000-0001-8265-972X

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.0c02088

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada (Grant No. RGPIN 2017−06030 to C.C.G. and Grant No. RGPIN-2016-06718 Fund 490974 to J.F.-K.). J.F.-K., and T.H.G thank the National Institutes of Health for support under Grant 5R01GM127627-03. T.M. was supported by funding from St. Jude Children's Research and the American Lebanese Syrian Associated Charities. We thank Dr. Taehyung Chris Lee for help in preparing the Sic1 Y14A sample, and S. Chakravarthy, J. Hopkins, and all BioCAT beamline staff at the Advanced Photon Source for assistance with SAXS measurements. Use of the Advanced Photon Source was supported by the U.S. Department of Energy under contract DE-AC02-06CH11357.

■ REFERENCES

(1) Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signaling and Regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (1), 18−29.

(2) Dunker, A.K.; Lawson, J.D.; Brown, C. J; Williams, R. M; Romero, P.; Oh, J. S; Oldfield, C. J; Campen, A. M; Ratliff, C. M; Hipps, K. W; Ausio, J.; Nissen, M. S; Reeves, R.; Kang, C.; Kissinger, C. R; Bailey, R. W; Griswold, M. D; Chiu, W.; Garner, E. C; Obradovic, Z. Intrinsically disordered protein. *J. Mol. Graphics Modell.* **2001**, *19* (1), 26−59.

(3) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu. Rev. Biophys.* **2008**, *37* (1), 215−246.

(4) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why Are "Natively Unfolded" Proteins Unstructured under Physiologic Conditions? *Proteins: Struct., Funct., Genet.* **2000**, *41* (3), 415−427.

(5) 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.; Kragelund, B. B.; Best, R. B.; Schuler, B. Extreme disorder

in an ultrahigh-affinity protein complex. *Nature* **2018**, *555* (7694), 61−66.
(6) Das, R. K.; Pappu, R. V. Conformations of Intrinsically Disordered Proteins Are Influenced by Linear Sequence Distributions of Oppositely Charged Residues. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (33), 13392−13397.
(7) 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.; Kragelund, B. B.; Best, R. B.; Schuler, B. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555* (7694), 61−66.
(8) 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.; Kragelund, B. B.; Best, R. B.; Schuler, B. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555* (7694), 61−66.
(9) 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.; Kragelund, B. B.; Best, R. B.; Schuler, B. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555* (7694), 61−66.
(10) Milles, S.; Mercadante, D.; Aramburu, I. V.; Jensen, M. R.øb.; Banterle, N.; Koehler, C.; Tyagi, S.; Clarke, J.; Shammas, S. L.; Blackledge, M.; Grater, F.; Lemke, E. A. Plasticity of an Ultrafast Interaction between Nucleoporins and Nuclear Transport Receptors. *Cell* **2015**, *163* (3), 734−745.
(11) Nath, Abhinav; Sammalkorpi, Maria; DeWitt, David C.; Trexler, Adam J.; Elbaum-Garfinkle, Shana; O'Hern, Corey S.; Rhoades, Elizabeth The Conformational Ensembles of α-Synuclein and Tau: Combining Single-Molecule FRET and Simulations. *Biophys. J.* **2012**, *103* (9), 1940−1949.
(12) Sormanni, P.; Piovesan, D.; Heller, G. T; Bonomi, M.; Kukic, P.; Camilloni, C.; Fuxreiter, M.; Dosztanyi, Z.; Pappu, R. V; Babu, M M.; Longhi, S.; Tompa, P.; Dunker, A K.; Uversky, V. N; Tosatto, S. C E; Vendruscolo, M. Simultaneous Quantification of Protein Order and Disorder. *Nat. Chem. Biol.* **2017**, *13*, 339−342.
(13) Kikhney, A. G.; Svergun, D. I. A Practical Guide to Small Angle X-Ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589* (19), 2570−2577.
(14) Gomes, G.-N.; Gradinaru, C. C. Insights into the Conformations and Dynamics of Intrinsically Disordered Proteins Using Single-Molecule Fluorescence. *Biochim. Biophys. Acta, Proteins Proteomics* **2017**, *1865* (11), 1696−1706.
(15) Schuler, B.; Soranno, A.; Hofmann, H.; Nettels, D. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annu. Rev. Biophys.* **2016**, *45*, 207−231.
(16) Jensen, M. R.øb.; Zweckstetter, M.; Huang, J.-r.; Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* **2014**, *114* (13), 6632−6660.
(17) Marsh, J. A.; Forman-Kay, J. D. Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins: Struct., Funct., Genet.* **2012**, *80* (2), 556−572.
(18) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of Protein Structural Ensemble Determination. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106−116.
(19) Borgia, A.; Zheng, W.; Buholzer, K.; Borgia, M. B.; Schuler, A.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A.; Best, R. B.; Schuler, B. Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. *J. Am. Chem. Soc.* **2016**, *138* (36), 11714−11726.
(20) Aznauryan, M.; Delgado, L.; Soranno, A.; Nettels, D.; Huang, J.-r.; Labhardt, A. M.; Grzesiek, S.; Schuler, B. Comprehensive Structural and Dynamical View of an Unfolded Protein from the Combination of Single-Molecule FRET, NMR, and SAXS. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (37), E5389−E5398.
(21) Fuertes, G.; Banterle, N.; Ruff, K. M.; Chowdhury, A.; Mercadante, D.; Koehler, C.; Kachala, M.; Estrada Girona, G.; Milles, S.; Mishra, A.; Onck, P. R.; Grater, F.; Esteban-Martin, S.; Pappu, R. V.; Svergun, D. I.; Lemke, E. A. Decoupling of Size and Shape Fluctuations in Heteropolymeric Sequences Reconciles Discrepancies in SAXS vs. FRET Measurements. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (31), E6342−E6351.
(22) Voith von Voithenberg, L.; Sanchez-Rico, C.; Kang, H.-S.; Madl, T.; Zanier, K.; Barth, A.; Warner, L. R.; Sattler, M.; Lamb, D. C. Recognition of the 3′ splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E7169.
(23) Delaforge, E.; Milles, S.; Bouvignies, G.; Bouvier, D.; Boivin, S.; Salvi, N.; Maurin, D.; Martel, A.; Round, A.; Lemke, E. A.; Ringkjøbing Jensen, M.; Hart, D. J.; Blackledge, M. Large-Scale Conformational Dynamics Control H5N1 Influenza Polymerase PB2 Binding to Importin α. *J. Am. Chem. Soc.* **2015**, *137* (48), 15122−15134.
(24) Mockel, C.; Kubiak, J.; Schillinger, O.; Kuhnemuth, R.; Della Corte, D.; Schroder, G. F.; Willbold, D.; Strodel, B.; Seidel, C. A. M.; Neudecker, P. Integrated NMR, Fluorescence, and Molecular Dynamics Benchmark Study of Protein Mechanics and Hydrodynamics. *J. Phys. Chem. B* **2019**, *123* (7), 1453−1480.
(25) Riback, J. A.; Bowman, M. A.; Zmyslowski, A. M.; Plaxco, K. W.; Clark, P. L.; Sosnick, T. R. Commonly Used FRET Fluorophores Promote Collapse of an Otherwise Disordered Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 8889−8894.
(26) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119* (16), 5113−5123.
(27) Song, J.; Gomes, G.-N.; Gradinaru, C. C.; Chan, H. S. An Adequate Account of Excluded Volume Is Necessary To Infer Compactness and Asphericity of Disordered Proteins by Förster Resonance Energy Transfer. *J. Phys. Chem. B* **2015**, *119* (49), 15191−15202.
(28) O'Brien, E. P.; Morrison, G.; Brooks, B. R.; Thirumalai, D. How Accurate Are Polymer Models in the Analysis of Förster Resonance Energy Transfer Experiments on Proteins? *J. Chem. Phys.* **2009**, *130* (12), 124903.
(29) Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D. Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* **2010**, *18* (4), 494−506.
(30) Mittag, T.; Orlicky, S.; Choy, W.-Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D. Dynamic Equilibrium Engagement of a Polyvalent Ligand with a Single-Site Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (46), 17772−17777.
(31) Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W.-Y.; Forman-Kay, J. D. Characterization of Disordered Proteins with ENSEMBLE. *Bioinformatics* **2013**, *29* (3), 398−399.
(32) Marsh, J. A.; Forman-Kay, J. D. Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *J. Mol. Biol.* **2009**, *391* (2), 359−374.
(33) Svergun, D.; Barberato, C.; Koch, M. H. J. CRYSOL - a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28* (6), 768−773.
(34) Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J; Berger, S.; Restle, T.; Goody, R. S; Gohlke, H.; Seidel, C. A M A Toolkit and Benchmark Study for FRET-Restrained High-Precision Structural Modeling. *Nat. Methods* **2012**, *9* (12), 1218−1225.
(35) Sindbert, S.; Kalinin, S.; Nguyen, H.; Kienzler, A.; Clima, L.; Bannwarth, W.; Appel, B.; Muller, S.; Seidel, C. A. M. Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *J. Am. Chem. Soc.* **2011**, *133* (8), 2463−2480.
(36) Fuertes, G.; Banterle, N.; Ruff, K. M.; Chowdhury, A.; Pappu, R. V.; Svergun, D. I.; Lemke, E. A. Comment on "Innovative

Scattering Analysis Shows That Hydrophobic Disordered Proteins Are Expanded in Water. *Science* **2018**, *361* (6405), eaau8230.
(37) Best, R. B.; Zheng, W.; Borgia, A.; Buholzer, K.; Borgia, M. B.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A.; Schuler, B. Comment on "Innovative Scattering Analysis Shows That Hydrophobic Disordered Proteins Are Expanded in Water. *Science* **2018**, *361* (6405), eaar7101.
(38) Riback, J. A.; Bowman, M. A.; Zmyslowski, A.; Knoverek, C. R.; Jumper, J.; Kaye, E. B.; Freed, K. F.; Clark, P. L.; Sosnick, T. R. Response to Comment on "Innovative Scattering Analysis Shows That Hydrophobic Disordered Proteins Are Expanded in Water. *Science* **2018**, *361* (6405), 45145.
(39) Riback, J. A.; Bowman, M. A.; Zmyslowski, A. M.; Knoverek, C. R.; Jumper, J. M.; Hinshaw, J. R.; Kaye, E. B.; Freed, K. F.; Clark, P. L.; Sosnick, T. R. Innovative Scattering Analysis Shows That Hydrophobic Disordered Proteins Are Expanded in Water. *Science* **2017**, *358* (6360), 238−241.
(40) Peran, I.; Holehouse, A. S.; Carrico, I. S.; Pappu, R. V.; Bilsel, O.; Raleigh, D. P. Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 12301.
(41) Song, J.; Gomes, G.-N.; Shi, T.; Gradinaru, C. C.; Chan, H. S. Conformational Heterogeneity and FRET Data Interpretation for Dimensions of Unfolded Proteins. *Biophys. J.* **2017**, *113* (5), 1012−1024.
(42) Sanchez-Rico, C.; Voith von Voithenberg, L.; Warner, L.; Lamb, D. C.; Sattler, M. Effects of Fluorophore Attachment on Protein Conformation and Dynamics Studied by spFRET and NMR Spectroscopy. *Chem. - Eur. J.* **2017**, *23* (57), 14267−14277.
(43) Nash, P.; Tang, X.; Orlicky, S.; Chen, Q.; Gertler, F. B.; Mendenhall, M. D.; Sicheri, F.; Pawson, T.; Tyers, M. Multisite Phosphorylation of a CDK Inhibitor Sets a Threshold for the Onset of DNA Replication. *Nature* **2001**, *414* (6863), 514−521.
(44) Verma, R.; McDonald, H.; Yates, J. R.; Deshaies, R. J. Selective degradation of ubiquitinated Sic1 by purified 26S proteasome yields active S phase cyclin-Cdk. *Mol. Cell* **2001**, *8* (2), 439−448.
(45) Verma, R.; Feldman, R. M.; Deshaies, R. J. SIC1 Is Ubiquitinated in Vitro by a Pathway That Requires CDC4, CDC34, and Cyclin/CDK Activities. *Mol. Biol. Cell* **1997**, *8* (8), 1427−1437.
(46) Borg, M.; Mittag, T.; Pawson, T.; Tyers, M.; Forman-Kay, J. D.; Chan, H. S. Polyelectrostatic Interactions of Disordered Ligands Suggest a Physical Basis for Ultrasensitivity. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (23), 9650−9655.
(47) Klein, P.; Pawson, T.; Tyers, M. Mathematical Modeling Suggests Cooperative Interactions between a Disordered Polyvalent Ligand and a Single Receptor Site. *Curr. Biol.* **2003**, *13* (19), 1669−1678.
(48) Csizmok, V.; Orlicky, S.; Cheng, J.; Song, J.; Bah, A.; Delgoshaie, N.; Lin, H.; Mittag, T.; Sicheri, F.; Chan, H. S.; Tyers, M.; Forman-Kay, J. D. An Allosteric Conduit Facilitates Dynamic Multisite Substrate Recognition by the SCF Cdc4 Ubiquitin Ligase. *Nat. Commun.* **2017**, *8*, 13943.
(49) Petoukhov, M. V.; Konarev, P. V.; Kikhney, A. G.; Svergun, D. I. ATSAS 2.1 − towards Automated and Web-Supported Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2007**, *40* (s1), s223−s228.
(50) Tria, G.; Mertens, H. D. T.; Kachala, M.; Svergun, D. I. Advanced Ensemble Modelling of Flexible Macromolecules Using X-Ray Solution Scattering. *IUCrJ* **2015**, *2* (2), 207−217.
(51) Neal, S. Rapid and Accurate Calculation of Protein 1H, 13C and 15N Chemical Shifts. *J. Biomol. NMR* **2003**, *26* (3), 215−240.
(52) Crehuet, R.; Buigues, P. J.; Salvatella, X.; Lindorff-Larsen, K. Bayesian-Maximum-Entropy Reweighting of IDP Ensembles Based on NMR Chemical Shifts. *Entropy* **2019**, *210* (9), 898.
(53) Henriques, J.; Arleth, L.; Lindorff-Larsen, K.; Skepo, M. On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J. Mol. Biol.* **2018**, *4300* (16), 2521−2539.
(54) Vestergaard, B.; Hansen, S. Application of Bayesian analysis to indirect Fourier transformation in small-angle scattering. *J. Appl. Crystallogr.* **2006**, *390* (6), 797−804.
(55) Feldman, H. J.; Hogue, C. W.V. A fast method to sample real protein conformational space. *Proteins: Struct., Funct., Genet.* **2000**, *390* (2), 112−131.
(56) Feldman, H. J.; Hogue, C. W.V. Probabilistic Sampling of Protein Conformations: New Hope for Brute Force? *Proteins: Struct., Funct., Genet.* **2002**, *460* (1), 8−23.
(57) Ganguly, D.; Chen, J. Structural Interpretation of Paramagnetic Relaxation Enhancement-Derived Distances for Disordered Protein States. *J. Mol. Biol.* **2009**, *3900* (3), 467−477.
(58) Marsh, J. A.; Forman-Kay, J. D. Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins: Struct., Funct., Genet.* **2012**, *800* (2), 556−572.
(59) Zheng, W.; Zerze, G. H.; Borgia, A.; Mittal, J.; Schuler, B.; Best, R. B. Inferring Properties of Disordered Chains from FRET Transfer Efficiencies. *J. Chem. Phys.* **2018**, *1480* (12), 123329.
(60) Kirkwood, J. G.; Riseman, J. The Intrinsic Viscosities and Diffusion Constants of Flexible Macromolecules in Solution. *J. Chem. Phys.* **1948**, *160* (6), 565−573.
(61) Ortega, A.; Amorós, D.; García de la Torre, J. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* **2011**, *1010* (4), 892−898.
(62) Liu, B.; Chia, D.; Csizmok, V.; Farber, P.; Forman-Kay, J. D.; Gradinaru, C. C. The Effect of Intrachain Electrostatic Repulsion on Conformational Disorder and Dynamics of the Sic1 Protein. *J. Phys. Chem. B* **2014**, *1180* (15), 4088−4097.
(63) Schäfer, L. *Excluded Vol. Effects in Polymer Solutions as Explained by the Renormalization Group*; Springer: Berlin, 1999.
(64) Zhou, H.-X. Polymer Models of Protein Stability, Folding, and Interactions. *Biochemistry* **2004**, *430* (8), 2141−2154.
(65) 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.* **2009**, *1090* (9), 4108−4139.
(66) Sasmal, S.; Lincoff, J.; Head-Gordon, T. Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid-β. *Biophys. J.* **2017**, *1130* (5), 1002−1011.
(67) Xue, Y.; Skrynnikov, N. R. Motion of a Disordered Polypeptide Chain as Studied by Paramagnetic Relaxation Enhancements, 15N Relaxation, and Molecular Dynamics Simulations: How Fast Is Segmental Diffusion in Denatured Ubiquitin? *J. Am. Chem. Soc.* **2011**, *1330* (37), 14614−14628.
(68) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.øb.; Zweckstetter, M.; Blackledge, M. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *1320* (24), 8407−8418.
(69) Pappu, R. V.; Wang, X.; Vitalis, A.; Crick, S. L. Polymer Physics Perspective on Driving Forces and Mechanisms for Protein Aggregation. *Arch. Biochem. Biophys.* **2008**, *4690* (1), 132−141.
(70) Lin, Y.-H.; Chan, H. S. Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated. *Biophys. J.* **2017**, *1120* (10), 2043−2046.
(71) Dignon, G. L.; Zheng, W.; Best, R. B.; Kim, Y. C.; Mittal, J. Relation between Single-Molecule Properties and Phase Behavior of Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *1150* (40), 9929−9934.
(72) Zheng, W.; Hofmann, H.; Schuler, B.; Best, R. B. Origin of Internal Friction in Disordered Proteins Depends on Solvent Quality. *J. Phys. Chem. B* **2018**, *1220* (49), 11478−11487.
(73) Locasale, J. W. Revisited: An Analysis of Multisite Phosphorylation and Substrate Rebinding. *J. Chem. Phys.* **2008**, *1280* (11), 115106.
(74) Bottaro, S.; Bengtsen, T.; Lindorff-Larsen, K. Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2020**, *2112* (0), 219−240.

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710

(75) Lincoff, J.; Krzeminski, M.; Haghighatlari, M.; Teixeira, J. M. C.; Gomes, G.-N. W.; Gradinaru, C. C.; Forman-Kay, J. D.; Head-Gordon, T. Extended Experimental Inferential Structure Determination Method for Evaluating the Structural Ensembles of Disordered Protein States. *Commun. Chem.* **2020**, *3*, 1−12.
(76) Huang, Y.; Yoon, M.-K.; Otieno, S.; Lelli, M.; Kriwacki, R. W. The Activity and Stability of the Intrinsically Disordered Cip/Kip Protein Family Are Regulated by Non-Receptor Tyrosine Kinases. *J. Mol. Biol.* **2015**, *4270* (2), 371−386.
(77) Sisamakis, E.; Valeri, A.; Kalinin, S.; Rothwell, P. J.; Seidel, C. A. M.; Walter, N. G. Chapter 18 - Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection. In *Methods in Enzymology*, Single Molecule Tools, Part B: Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods, Elsevier: Amsterdam, Netherlands, **2010**; Vol. 475, 455−514 DOI: 10.1016/S0076-6879(10)75018-7
(78) Kudryavtsev, V.; Sikor, M.; Kalinin, S.; Mokranjac, D.; Seidel, C. A. M.; Lamb, D. C. Combining MFD and PIE for Accurate Single-Pair Förster Resonance Energy Transfer Measurements. *ChemPhysChem* **2012**, *130* (4), 1060−1078.
(79) Kapanidis, A. N.; Laurence, T. A.; Lee, N. K.; Margeat, E.; Kong, X.; Weiss, S. Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **2005**, *380* (7), 523−533.
(80) Eggeling, C.; Berger, S.; Brand, L.; Fries, J.R.; Schaffer, J.; Volkmer, A.; Seidel, C.A.M. Data Registration and Selective Single-Molecule Analysis Using Multi-Parameter Fluorescence Detection. *J. Biotechnol.* **2001**, *860* (3), 163−180.
(81) Nir, E.; Michalet, X.; Hamadani, K. M.; Laurence, T. A.; Neuhauser, D.; Kovchegov, Y.; Weiss, S. Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments. *J. Phys. Chem. B* **2006**, *1100* (44), 22103−22124.
(82) Gillespie, J. R.; Shortle, D. Characterization of Long-Range Structure in the Denatured State of Staphylococcal Nuclease. I. Paramagnetic Relaxation Enhancement by Nitroxide Spin labels. *J. Mol. Biol.* **1997**, *2680* (1), 158−169.
(83) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *1090* (8), 1528−1532.
(84) Soranno, A.; Holla, A.; Dingfelder, F.; Nettels, D.; Makarov, D. E.; Schuler, B. Integrated View of Internal Friction in Unfolded Proteins from Single-Molecule FRET, Contact Quenching, Theory, and Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *1140* (10), E1833−E1839.

https://dx.doi.org/10.1021/jacs.0c02088
J. Am. Chem. Soc. 2020, 142, 15697−15710