



DOI: 10.1002/ijch.202000007

FRET Dyes Significantly Affect SAXS Intensities of Proteins

Ines Reinartz,^[a, b] Marie Weiel,^[c, d] and Alexander Schug^[e, f]

Abstract: Structural analyses in biophysics aim at revealing a relationship between a molecule's dynamic structure and its physiological function. Förster resonance energy transfer (FRET) and small-angle X-ray scattering (SAXS) are complementary experimental approaches to this. Their concomitant application in combined studies has recently opened a lively debate on how to interpret FRET measurements in the light of SAXS data with the popular example of the radius of gyration, commonly derived from both FRET and SAXS. There still is a lack of understanding in how to mutually relate and interpret quantities equally obtained from FRET or SAXS, and to what extent FRET dyes affect SAXS intensities in combined applications. In the present work, we examine the interplay of FRET and SAXS from a computational simulation perspective. Molecular simulations are a valuable complement to experimental approaches and supply instructive

information on dynamics. As FRET depends not only on the mutual separation but also on the relative orientations, the dynamics, and therefore also the shapes of the dyes, we utilize a novel method for simulating FRET-dye-labeled proteins to investigate these aspects in atomic detail. We perform structure-based simulations of four different proteins with and without dyes in both folded and unfolded conformations. In-silico derived radii of gyration are different with and without dyes and depend on the chosen dye pair. The dyes apparently influence the dynamics of unfolded systems. We find that FRET dyes attached to a protein have a significant impact on theoretical SAXS intensities calculated from simulated structures, especially for small proteins. Radii of gyration from FRET and SAXS deviate systematically, which points to further underlying mechanisms beyond prevalent explanation approaches.

1. Introduction

In the past decades, an enormous variety of protein structures has been accumulated experimentally by employing sophisticated high-resolution techniques such as X-ray crystallography or nuclear magnetic resonance spectroscopy (NMR).^[1] With cellular function, however, being dictated by the interplay between static structures and dynamic conformational changes, alternative methods have been catching up so as to elucidate the dynamic nature of the structure-function paradigm. Förster resonance energy transfer (FRET) and small-angle X-ray scattering (SAXS) are particularly popular approaches to this and complementary to the aforementioned methods. SAXS can be used to study average structures of various systems and enables even time-resolved analyses of conformational transitions in direct response to altered external conditions.^[2] A solution of biomolecules is exposed to X-rays and the integrated scattered intensity is recorded in the small-angle regime, which contains information on structural features of the solute molecules. FRET provides access to time-resolved distance information on, e.g., folding dynamics,^[3] intermediate structures,^[4,5] and function-related conformational transitions.^[6] After labeling specific molecular sites with fluorescent dyes, the distance-dependent energy transfer efficiency between them is measured.

Both FRET and SAXS are widely applied for analysis of unfolded and intrinsically disordered proteins (IDPs).^[7,8] The characteristics of such systems are of great interest due to their relevance to folding and the physiological prevalence of partly and entirely unstructured proteins as, despite lacking definite structure, IDPs fulfill important functional roles. Polymer

physics is applied to understand the dynamics of unstructured proteins with their high conformational diversity and further relate their properties to folding and function, and the validity of such approaches has been studied extensively in the context of FRET and SAXS.^[8–11]

-
- [a] *I. Reinartz*
Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
- [b] *I. Reinartz*
HIDSS4Health – Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany
- [c] *M. Weiel*
Department of Physics, Karlsruhe Institute of Technology, Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany
- [d] *M. Weiel*
Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
- [e] *A. Schug*
Institute for Advanced Simulation, Jülich Supercomputing Center, Wilhelm-Johnen-Straße, 52428 Jülich, Germany
- [f] *A. Schug*
Faculty of Biology, University of Duisburg-Essen, Germany

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/ijch.202000007>

© 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

More recently, there has been an ongoing discussion on how to interpret FRET measurements in the light of SAXS data, especially for IDPs and unfolded ensembles. A popular structural quantity equally derived from FRET and SAXS is the radius of gyration R_g , a measure of overall molecular size. Important questions are how to mutually relate and interpret derived values of R_g obtained by either of the methods, and to what extent FRET dyes influence SAXS intensities in concomitant applications. Recent studies find that FRET implies IDPs to be compacted in water by comparison with high denaturant concentrations, while this compaction could not be validated within SAXS, which is known as the so-called SAXS-FRET controversy.^[12–14] For globular proteins, theory and simulation predict the dimensions of unfolded conformations to decrease with the denaturant concentration. Whereas interpretation of single-molecule FRET data supports this prediction, SAXS data point to the opposite.^[15] Based on theoretical considerations, simulations, and new experimental data, Thirumalai et al. found that sizes of unfolded states of globular proteins have to decrease as the denaturant concentration goes down, and stated compaction of unfolded proteins to be universal.^[15] In this context, water's critical role as a solvent further comes to the fore.^[9,15,16] These findings are in accordance with results by Reddy et al., who studied the SAXS-FRET controversy in coarse-grained simulations including denaturant using the example of Ubiquitin.^[11] A possible explanation for these at first glance contradicting observations is a decoupling of size and shape fluctuations, leading to the conclusion that FRET and SAXS do not measure the same quantity but are complementary approaches.^[17] Fuertes et al. hypothesize proteins to be subject to a sequence-specific decoupling of end-to-end distance R_e measured by FRET and radius of gyration R_g deduced from SAXS, and as heteropolymers, proteins may exhibit diverse R_g - R_e relationships.^[18] Other studies assume the analysis methods to be the primary source of the apparent discrepancies.^[19] Based on combined FRET and SAXS studies of unfolded proteins and IDPs, Borgia et al. suggest SAXS measurements to be basically model-free, whereas interpretation of FRET data always relies on a model such as a Gaussian or excluded-volume chain to relate R_g and R_e .^[20] Zheng et al. performed explicit-solvent MD simulations of a 79-residue IDP, revealing potential discrepancies between FRET and SAXS for this particular system.^[21]

It however remains unclear if and, if yes, to what extent FRET dyes influence SAXS measurements, and how distinct calculation methods for R_g differ with respect to their results. Molecular simulations are the ideal tool to clarify these issues. They can be applied to study the influence of FRET dyes on scattering patterns and give access to all different variants of R_g .

Here, we illuminate the interplay of combined FRET and SAXS from a computational simulation point of view. Molecular simulations are a valuable complement to experiments and, depending on their complexity, provide insightful information up to atomistic dynamics of a system. FRET does

not directly access quantitative information about molecular distances, but measures the energy transfer efficiency between the dyes. This efficiency depends not only on the separation distance of the dyes but also on their relative orientation and dynamics, which can be observed within molecular simulations best. We consider a novel method for simulating FRET-dye-labeled proteins using native structure-based models (SBMs) on the atomistic level.^[22,23] Based on energy landscape theory and the principle of minimal frustration,^[24–27] SBMs probe dynamics arising from the system's native geometry.^[28] By this means, force field complexity is drastically decreased without loss of substantial information on the system's characteristics, resulting in improved sampling and high computational efficiency. In particular, such models enable thorough sampling of large conformational ensembles such as intrinsically disordered or unfolded systems. Using the simulation protocol by Reinartz et al.,^[23] we calculate theoretical SAXS curves from molecular simulations of four different proteins with and without dyes. By comparing these intensities, we investigate the influence of FRET dyes on scattering curves from SAXS for both folded and unfolded ensembles. Furthermore, we derive and compare different variants of the radius of gyration as a particularly popular quantity accessible in both FRET and SAXS. In doing so, we hope to make an important contribution to elucidating the relationship and interplay between the experimental methods of FRET and SAXS.

2. Methods

2.1 Förster Resonance Energy Transfer

Förster resonance energy transfer (FRET)^[30] is a mechanism describing non-radiative energy transfer between two light-sensitive molecules. An electronically excited donor may transfer energy to an acceptor via non-radiative dipole-dipole coupling. The efficiency of this energy transfer depends on the sixth power of the distance between donor and acceptor. FRET consequently is extremely sensitive to small distance changes in the nanometer range and also referred to as a “spectroscopic ruler”.^[31] By labeling specific protein residues with suitable dyes as illustrated in Figure 1, different conformations become distinguishable and conformational changes can be observed directly through changes in spatial dye separation. Experimentally, the FRET efficiency E is measured, which depends on the inter-dye distance R_{DA} as^[32]

$$E = \frac{1}{1 + \left(\frac{R_{DA}}{R_0}\right)^6} . \quad (1)$$

The Förster radius R_0 is given by the donor-acceptor distance at which E equals 0.5. It depends on the relative orientations of donor and acceptor represented by the dipole orientation factor κ^2 as $R_0^6 \propto \kappa^2$.^[32] Rotational dye diffusion is usually assumed to be fast with respect to the lifetime of the

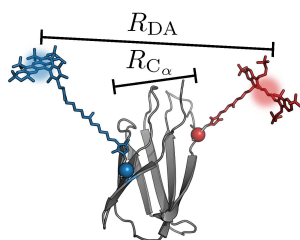


Figure 1. Tenth type III domain of fibronectin with Alexa Fluor dyes attached. Tenth type III domain of fibronectin (¹⁰FNIII, PDB code: 1TTG^[29]) with AF 546 (AF546, blue) and AF 647 (AF647, red) dyes attached at residues 11 and 86, respectively. The C_α atoms of these residues are shown as blue and red sphere, respectively. Inter-dye distance R_{DA} and C_α distance R_{C_α} are marked.

excited state, yielding a constant value of $\overline{\kappa^2} = 2/3$ in the “isotropic averaging regime”.^[32] In contrast to this, dye molecules are modeled explicitly at atomistic resolution in the structure-based protocol for simulation of dye-labeled proteins by Reinartz et al.^[23] applied here. Thus, κ^2 can be calculated directly from such simulations without further approximations.

2.2 Small-angle X-ray Scattering

Small-angle X-ray scattering (SAXS) is an efficient tool for low-resolution structural characterization of dissolved biomolecules.^[2,33] A solution of proteins is exposed to X-rays with wavelength λ . The integrated intensity from elastic scattering is measured in the small-angle regime as a function of momentum transfer $q = 4\pi\sin\theta/\lambda$ where 2θ is the scattering angle. SAXS records the averaged scattering intensity over the entire conformational ensemble and all possible orientations of the solute molecules. Ideally, this isotropic intensity distribution is proportional to the spatially averaged scattering from a single particle. The net solute scattering, in return, is related to the electron density difference between solute and solvent.

The spherically averaged scattering intensity I of a molecule modeled as a collection of elementary scatterers, e. g., atoms or amino acids, can be calculated via the Debye equation:^[34]

$$I(q) = \sum_{i,j} f_i(q)f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}} \quad (2)$$

r_{ij} is the distance between two scatterers i and j , f_i and f_j are the corresponding form factors.

Different parts of such an intensity pattern provide information about different structural features. However, it is important to note that the signal-to-noise ratio of experimentally measured intensities decreases rapidly with increasing momentum transfer q .

For small q , the intensity can be described by the *Guinier approximation*.^[35]

$$I(q) = I(0)\exp\left[-\frac{q^2 R_g^2}{3}\right] \quad (3)$$

Accordingly, R_g can be extracted from the curve slope in a Guinier plot. Note that the Guinier approximation is only valid for $qR_g < 1.3$ for globular proteins^[2] and in an even smaller range for elongated structures.

2.3 Structure-based Simulations of Dye-labeled Proteins

2.3.1 Structure-based Models

Gō-type or structure-based models (SBMs) provide a minimalistic description of biomolecular dynamics arising from the native geometry. Giving access to biologically relevant time-scales, computationally efficient SBMs provide rich information on the system’s characteristics. Successful applications cover a wide range of protein dynamics such as folding pathways^[36–41] and kinetics.^[42] SBMs are also employed for structure prediction,^[43–46] integrative structural modeling of experimental data from, e. g., SAXS^[47] or cryo-EM,^[48] and investigation of transition state ensembles.^[49,50]

Founded on energy landscape theory and the principle of minimal frustration, protein dynamics are modeled based on the assumption that native interactions are generally stabilizing, whereas non-native interactions are only included to preserve excluded volume.^[24–28] The essential part lies in the so-called contact potential. Each native contact defined by a pair interaction between atoms spatially close in the native state is assigned an attractive potential, whereas a purely repulsive excluded-volume term is included for all atom pairs. As a result, an overall energetic drive to the native structure overtops kinetic traps which would originate from non-native interactions.

We use an all-atom SBM taking into account all heavy atoms of the protein^[22] as implemented in eSBMTools.^[38] With native bond lengths r_0 , bond angles θ_0 , and proper and improper dihedral angles φ_0 and χ_0 , the simplified potential reads:^[51]

$$\begin{aligned} V_{\text{SBM}} = & \sum_{\text{bonds}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_a (\theta - \theta_0)^2 \\ & + \sum_{\substack{\text{improper} \\ \text{dihedrals}}} K_i (\chi - \chi_0)^2 \\ & + \sum_{\substack{\text{proper} \\ \text{dihedrals}}} K_d \left\{ [1 - \cos(\varphi - \varphi_0)] + \frac{1}{2} [1 - \cos(3(\varphi - \varphi_0))] \right\} \\ & + \sum_{\text{contacts}} K_c C_G(r_{ij}, r_0^{ij}) \\ & + \sum_{\substack{\text{non-native} \\ \text{contacts}}} K_{\text{nc}} \left(\frac{\tilde{\sigma}}{r_{ij}} \right)^{12} \end{aligned} \quad (4)$$

Numerical values of energetic weights K , the excluded volume for Pauli repulsion $\tilde{\sigma}$, and the functional form of the Gaussian contact potential C_G can be found in Supplementary Information S2.1 (see also Refs. [23] and [52]).

2.3.2 Simulation of Dye-labeled Proteins

To simulate protein systems with dye pairs attached, we apply a novel structure-based simulation protocol developed by Reinartz et al.^[23] In this method, quantum-chemical calculations are initially carried out to obtain three-dimensional dye structures from available chemical structures. Subsequently, linkers are added to bind the dyes to the protein. The dyes are parametrized for inclusion into the SBM, where the only interaction considered is excluded-volume repulsion.^[23] In a last step, they are attached preferably orthogonally to the protein surface. Simulations are run in GROMACS v4.5.4^[53] using the structure-based potential introduced in Eq. (4) and molecular dynamics parameters as described in Ref. [23] (see also Supplementary Information S2).

2.4 Studied Systems

2.4.1 Proteins

As a first test system, we use the 94-residue tenth type III module of *fibronectin* (¹⁰FNIII, PDB code: 1TTG^[29]) depicted in Figure 1. Fibronectin is a homodimeric glycoprotein of the extracellular matrix. It plays a major role in cell adhesion, growth, migration, and differentiation, and is important for wound healing and embryonic development.^[54] Altered expression, degradation, and organization of this protein have been associated with several pathologies, including cancer and fibrosis.^[55]

Chymotrypsin inhibitor 2 (CI-2, PDB code: 2CI2^[56]) is a widely studied and well-understood 83-residue serine protease inhibitor from barley seeds. It was among the first proteins to have its folding/unfolding transition state extensively characterized by the protein engineering method.^[57,58] Its denatured state and folding were subsequently characterized by NMR and hydrogen exchange.^[59–61]

We study the globular 66-residue *cold shock protein* from *Thermotoga maritima* (CspTm, PDB code: 1G6P^[62]) as a third system. Upon rapid temperature decrease, many bacteria produce small cold shock proteins. During cold shock, the efficiency of transcription and translation is reduced due to stabilization of nucleic acid secondary structure. Cold shock proteins are thought to counteract this by preventing the formation of messenger RNA secondary structure at low temperature as nucleic acid chaperones.

Cytolysin A (ClyA) of *Escherichia coli* is a pore-forming hemolytic toxin. This protein exists as a monomer of 303 residues (PDB code: 1QOY^[63]) and undergoes a conformational change to the protomer before assembling into a dodecameric pore (PDB code: 2WCD).^[64]

2.4.2 Dyes

We use two pairs of the *Alexa Fluor* (AF) family of fluorescent dyes,^[65] which are frequently applied as cell and tissue labels in fluorescence microscopy. The excitation and emission spectra of the AF series cover the visible spectrum and extend into the infrared. Individual members are numbered according to their approximated excitation maxima (in nm). We use the AF 488 dye with C₅-linker (AF488) and AF 594 dye with C₅-linker (AF594), and the AF 546 dye with C₅-linker (AF546) and AF 647 dye with C₂-linker (AF647).

Additionally, we use the Biotium dye CF680R (B680) for simulations with three dyes.^[66] Figures 1 and 2 show examples of the studied systems. A detailed list and depictions of all composite systems can be found in Supplementary Information S1. For structures and parameters of the dyes, see Ref. [23].

2.5 Calculation of SAXS Profiles from Structural Models

From a computational simulation perspective, theoretical calculation of accurate scattering patterns from atomic positions is a key factor for successful analysis and interpretation of SAXS data. Existing methods can be divided into either implicit- or explicit-solvent. One drawback of the computationally more efficient and widely used implicit-solvent methods is their dependence on several non-trivial free parameters with the most prominent example of the solvation shell's excess density. Given experimental data, the latter can be determined by a least-squares fit of the forwardly calculated curve at the risk of overfitting. Otherwise, it is set to 10% to 15% of the bulk water electron density.^[67] At this point, it is important to note that it may have different values for folded

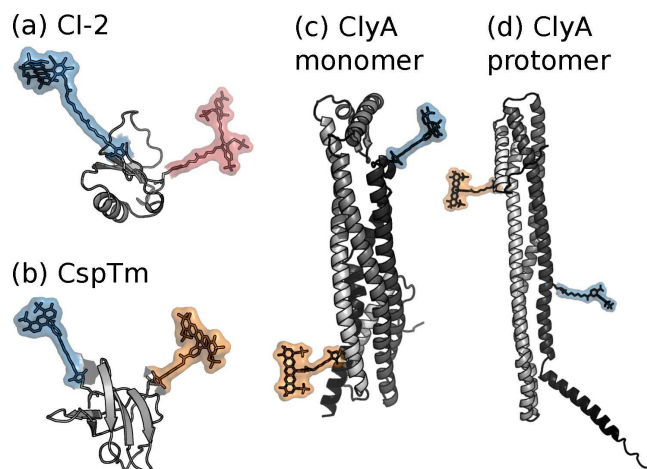


Figure 2. Example systems of three different proteins. (a) CI-2 with AF546 (blue) and AF647 (red) at positions 20 and 78. (b) CspTm with AF488 (blue) and AF594 (orange) at positions 2 and 68. ClyA monomer (c) and protomer (d) configuration with AF488 (blue) and AF594 (orange) at positions 56 and 252. More details on all systems can be found in Supplementary Information S1.

and unstructured proteins depending on their specific solvation properties.^[68] We apply the popular implicit-solvent method CRY SOL, which uses multipole expansion to evaluate spherically averaged scattering patterns from biomolecular structures.^[67] To simulate the primary hydration layer, the solvation shell is approximated by a border layer of 3 Å effective thickness and excess density $\delta\rho$ with respect to the average density of free bulk water $\rho_0 = 0.334 \text{ e}\text{\AA}^{-3}$.^[67] According to Henriques et al., $\delta\rho$ can substantially influence SAXS curves forwardly calculated from structural models, especially for unfolded proteins, and small variations to $\delta\rho$ can change computed radii of gyration by 5% to 10%.^[68] They report that the CRY SOL default value of $0.03 \text{ e}\text{\AA}^{-3}$ yields suboptimal results and generally suggest lower solvation shell contrasts between $0.01 \text{ e}\text{\AA}^{-3}$ and $0.02 \text{ e}\text{\AA}^{-3}$. Whilst a value of $0.0125 \text{ e}\text{\AA}^{-3}$ is recommended for folded proteins, specifying a single density contrast is not valid for disordered proteins.^[68] To assess the influence of $\delta\rho$ on R_g for the systems studied here, we conduct a sensitivity analysis and compare derived values of R_g for different values of $\delta\rho$ in the range of $0.00 \text{ e}\text{\AA}^{-3}$ to $0.03 \text{ e}\text{\AA}^{-3}$. Results can be found in Supplementary Information S6. As expected, different values slightly affect SAXS-derived R_g , which generally increase with $\delta\rho$. With the exception of $\delta\rho = 0.00 \text{ e}\text{\AA}^{-3}$ neglecting the solvation shell completely, we find the overall trend discussed in Section 3 to be preserved among different values of $\delta\rho$.

2.6 Radius of Gyration

A popular structural feature derived from both FRET and SAXS is the radius of gyration R_g , a measure of a molecule's spatial extent. GROMACS calculates it as^[53]

$$R_g = \sqrt{\frac{\sum_i r_i^2 m_i}{\sum_i m_i}}, \quad (5)$$

where r_i is the distance of atom i to the molecular center of mass and m_i is the atomic mass deposited in a GROMACS parameter file.

2.6.1 Determination of Different R_g Variants

To analyze different R_g variants in the context of FRET and SAXS, we first calculate a “true” reference value $R_{g,\text{gmx}}$ from the molecular model of the protein without dyes using GROMACS (see Eq. (5)). Second, we consider the corresponding value $R_{g,\text{gmx}}^{+\text{dyes}}$ computed from the molecular model with dyes.

Analysis of the Guinier region in SAXS provides two additional values $R_{g,\text{saxs}}$ and $R_{g,\text{saxs}}^{+\text{dyes}}$. Due to the occasionally narrow Guinier region, these may contain errors, in particular

for large elongated systems as the unfolded monomer and protomer (also see Supplementary Information S5).

For unfolded proteins, we calculate R_g as done previously in experimental work.^[20] The proteins are assumed to behave like excluded-volume chains^[23,69] (see also Supplementary Information S3). To mimic dyes attached to the chain termini, we analyze truncated systems. Based on simulations of respective full systems, we only consider atoms of residues between the dye positions, neglecting remaining residues in all calculations. We then extract the end-to-end C_α distance R_e (corresponding to the C_α distance between dye-labeled residues) to calculate the apparent radius of gyration

$$R_{g,C_\alpha}^{\text{app}} = \sqrt{\langle R_e^2 \rangle} / \sqrt{6.26} \quad (6)$$

for excluded volume chains.^[20] FRET is often assumed to measure the distance between dye-labeled C_α atoms, thus ignoring contributions of the dyes' linkers. Modeling the linkers as chain extensions of a certain additional sequence length L , the inter-dye distance R_{DA} can be rescaled to a corresponding C_α distance via^[20]

$$f = \left(\frac{N_{\text{inter-dye}}}{N_{\text{inter-dye}} + L} \right)^\nu \quad (7)$$

with the number of considered residues between dye-labeled sites $N_{\text{inter-dye}}$, scaling exponent ν , and additional sequence length L for the dye pairs from [23]. With this factor, we obtain

$$R_{g,R_{\text{DA}}}^{\text{app}} = \sqrt{\langle R_{\text{DA}}^2 \rangle} \cdot f / \sqrt{6.26}. \quad (8)$$

All variants of R_g reflect an average over the simulated ensemble and are compared for the different systems in the following in Figures 5, 6, and 7.

3. Results and Discussion

3.1 Influence of FRET Dyes on SAXS Measurements

To start with, we examine the direct impact of FRET dyes on SAXS measurements and derived quantities. Recapitulate that experimental SAXS data always reflect an average over all possible solute conformations and orientations. To mimic this ensemble average, we compute a representative intensity curve for each system by considering 5 000 structures equidistantly distributed over one simulation. Having calculated the individual intensity of each structure with CRY SOL,^[67] we determine mean and standard deviation of the resulting ‘array’ of intensities to obtain an ensemble-averaged curve and assess the degree of agreement of SAXS curves from different conformations. We proceed accordingly for simulations of all folded and unfolded systems, each with and without dyes.

From these representative intensity curves, SAXS-derived radii of gyration are computed and compared to a “true”

reference given by the average $R_{g,gmx}$ of each simulated ensemble, where single-frame values are calculated according to Eq. (5). To further investigate how the dyes influence the systems, we additionally study $R_{g,gmx}$ distributions, discussed in Supplementary Information S4 in detail.

As shown in Figure 3a, the resulting intensity curves of $^{10}\text{FNIII}$ with and without dyes exhibit considerable differences for the folded ensemble. Because dyes change both size and shape of a system, this is to be expected. Although rather

small, a difference is still observable for the unfolded ensemble due to an increased chain size with dyes attached. Similar curve shapes indicate that the dyes have a minor but still visible influence in this case. For the larger system of monomer, differences in Figure 3b are almost insignificant in accordance with our expectations.

All derived plots for $^{10}\text{FNIII}$ are shown in Figure 4. The Guinier approximation in Figure 4a is only valid in a certain region where $\ln(I(q)/I(0))$ versus q^2 can legitimately be

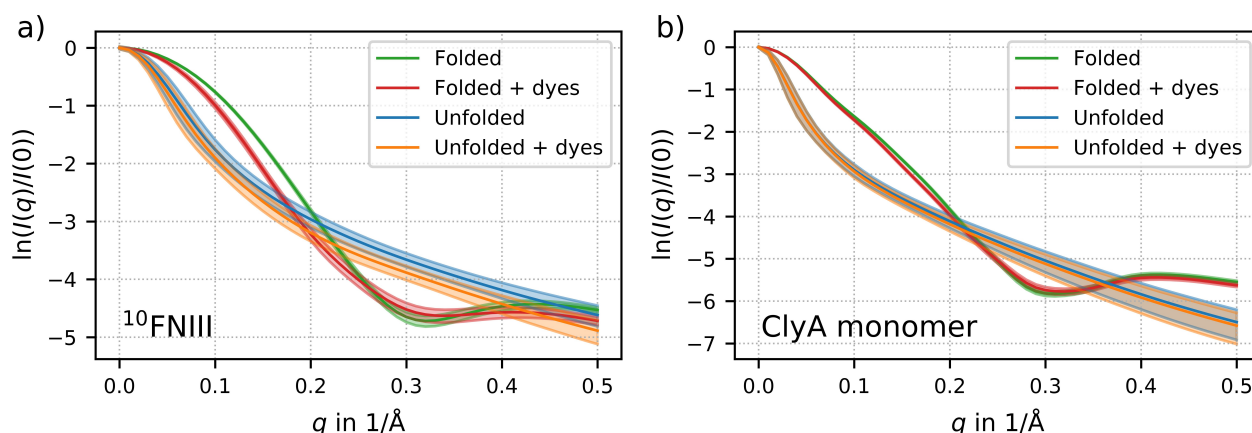


Figure 3. SAXS intensities without and with dyes for (a) $^{10}\text{FNIII}$ and (b) ClyA monomer with dyes at positions 56 and 252. More details on the systems can be found in Supplementary Information S1. Average intensities (solid lines) versus momentum transfer q along with each intensity-curve distribution's standard deviation over corresponding single-frame intensities directly calculated from individual simulation snapshots (shaded area). It is important to note that this standard deviation is to be interpreted as the intensity distribution width at a particular q point rather than an actual "error" in the sense of statistical uncertainties or systematic deviations as they would occur in experimental data. In accordance with R_g distributions (see Supplementary Information S4), the standard deviation may be considered a measure of conformational heterogeneity in the underlying simulated ensemble, which also shows in the fact that standard deviations for unfolded systems are consistently larger than those for folded systems. Curves are depicted for both systems in the folded states without (green) and with dyes (red) and in the unfolded states without (blue) and with dyes (orange).

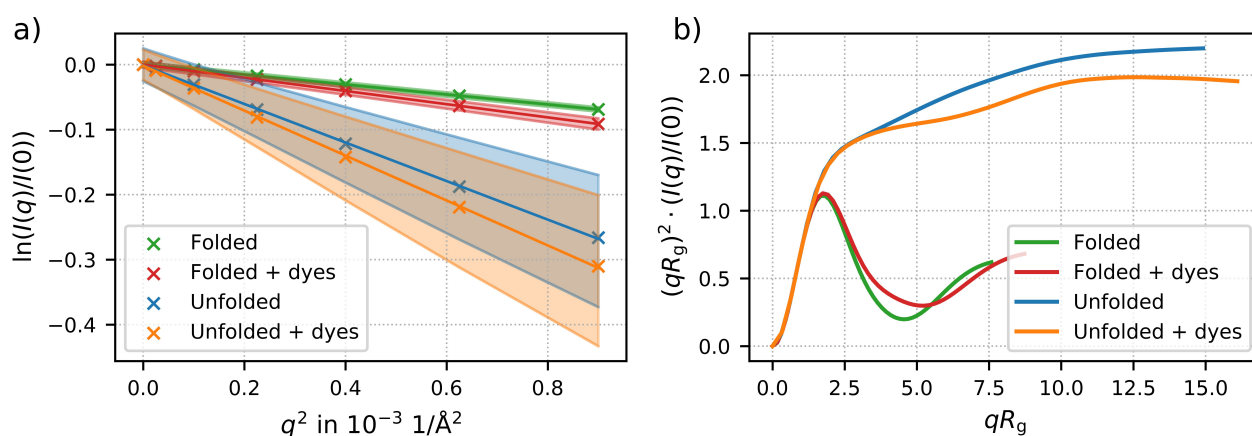


Figure 4. Guinier plot (a) and dimensionless Kratky plot (b) for $^{10}\text{FNIII}$. (a) The Guinier plot shows characteristic SAXS profiles in the Guinier region (crosses) along with linear fits (solid lines). Shaded areas indicate standard deviations (i.e. intensity distribution widths) propagated from Figure 3 to the Guinier representation. R_g errors resulting from the linear regression in the Guinier analysis are in the order of 10^{-2} to $10^{-4}R_g$ and can be found in Supplementary Information S5. (b) The dimensionless Kratky plot gives information about the protein's conformations. Both plots are shown for $^{10}\text{FNIII}$ in folded states (green, red) and unfolded states (blue, orange) without and with dyes, respectively. More details on the systems can be found in Supplementary Information S1.

approximated by a straight line in a linear fit. The slopes, from which respective radii of gyration are extracted, are different for both folded and unfolded states as well as for the system with and without dyes. The Kratky plot in Figure 4b exhibits a distinct peak for the folded ensemble and a plateau for the unfolded ensemble, thus giving a perfect example of how this kind of analysis can be used to study molecular folding. In the folded case, the broader peak for the system with dyes suggests a less compact structure compared to the purely proteinic system.

3.2 Comparison of R_g Variants

Ratios of all R_g variants to the “true” $R_{g,gmx}$ are presented in Figure 5 for the folded systems. $R_{g,gmx}$ naturally depends on the protein only. With dyes attached, the systems appear to be larger, manifesting in a greater $R_{g,gmx}^{+dyes}$ with respect to $R_{g,gmx}$. As expected, the smaller the system, the more significant is this effect. SAXS-derived R_g values show a similar shift, i.e. $R_{g,saxs}^{+dyes}$. Exempt from this is ClyA protomer with dyes at positions 56/252, where $R_{g,gmx}^{+dyes}$ is actually lower than $R_{g,gmx}$. This can be explained by a center-of-mass shift due to the dyes in favor of a reduced radius of gyration (see Figure 2). As evident from CspTm and ClyA monomer and protomer, different dye positions affect R_g only marginally. In contrast, dye types seem to have a more pronounced effect, as shown for CI-2. We find that R_g variants derived from SAXS apparently overestimate R_g for small systems, whereas underestimating it for larger systems in some cases. For the smaller proteins 10 FNIII, CI-2, and CspTm, SAXS-derived R_g values are consistently larger than those calculated with the mass-

weighted formula in Eq. (5). This is true for systems with and without dyes as well as for folded and unfolded conformations and in full accordance with the expectations. This overestimation could be triggered by the CRY SOL method for calculating SAXS profiles from structural models, which takes into account the hydration shell, or arise from neglecting hydrogen atoms in the molecular model.

The only exception from this typical behavior is ClyA in its elongated monomer and protomer configuration. Here, all values are located in a very narrow range, and SAXS-derived R_g are similar to or slightly smaller than corresponding $R_{g,gmx}$. We assume this counter-intuitive behavior to be caused by a rather narrow Guinier region, which likely results in a greater error in the linear regression.

Analogous results for the unfolded systems are depicted in Figure 6. As apparent from CI-2 and CspTm, $R_{g,gmx}$ is affected by both dye types and positions here, suggesting a subtle but however perceivable influence of the dyes on the chain dynamics. Just as for the folded systems, $R_{g,gmx}^{+dyes}$ is consistently larger than $R_{g,gmx}$. This effect is related to the dyes' distance in the protein sequence and can be illustrated using the examples of monomer and protomer. With dyes attached to the termini affecting the occupied volume to a greater extent than if attached in the middle, the observed shift increases with the dye separation. The more peripheral the dye-labeling positions in a protein sequence, the more the dyes with their linkers increase the dimensions of a system as reflected by R_g , in particular for completely elongated unfolded conformations. For the smaller systems CI-2, CspTm, and 10 FNIII, $R_{g,saxs}$ and $R_{g,saxs}^{+dyes}$ show the expected tendency just as for the folded case. For larger ClyA monomer and protomer, we find the SAXS-

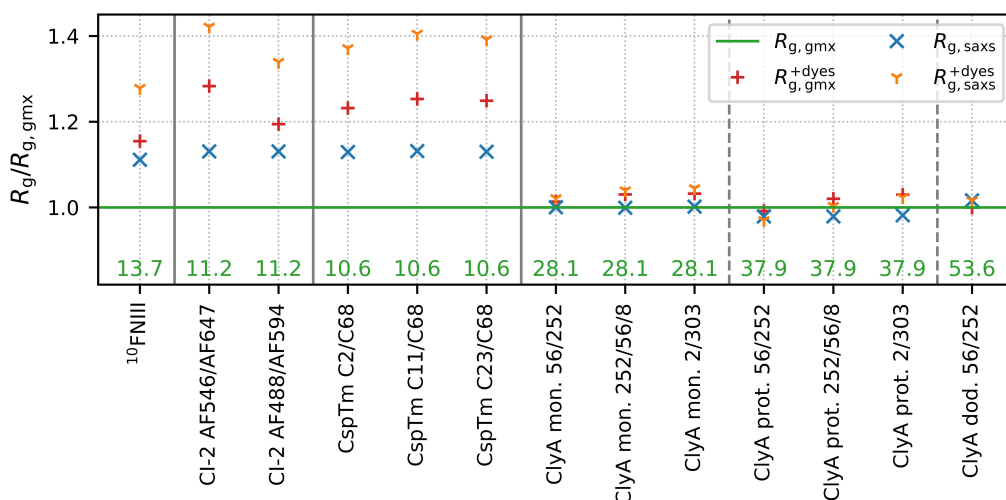


Figure 5. R_g variants for the folded systems with respect to $R_{g,gmx}$ (green line) given at the bottom in Å. We study 10 FNIII, CI-2 with two different dye pairs (AF546/AF647 and AF488/AF594), CspTm with AF488/AF594 at three different labeling positions, and ClyA monomer, protomer, and dodecamer with AF488/AF594/(B680) at different labeling sites. More details on the systems can be found in Supplementary Information S1. R_g values calculated from atomic structures with dyes ($R_{g,gmx}^{+dyes}$, red) and those derived from SAXS curves without ($R_{g,saxs}$, blue) and with dyes ($R_{g,saxs}^{+dyes}$, orange) are depicted. R_g errors derived from the Guinier linear regression are listed in Supplementary Information S5.

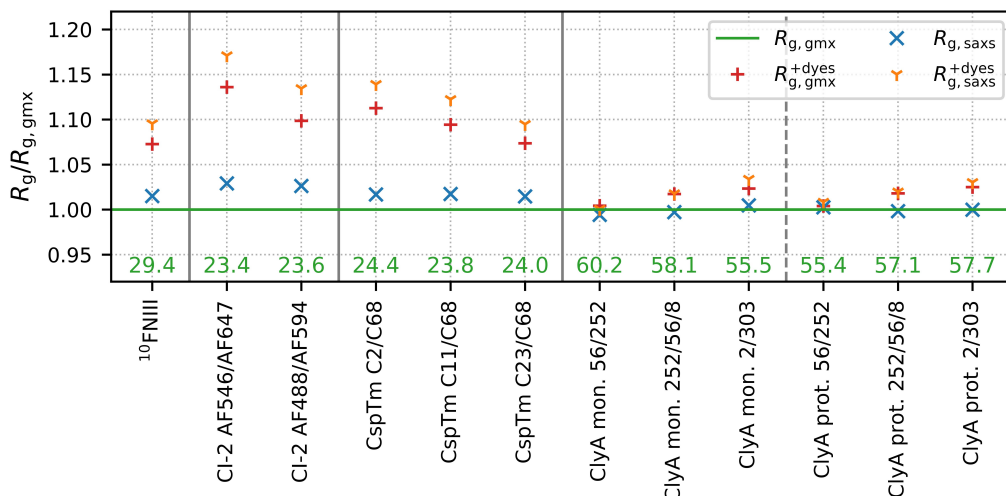


Figure 6. R_g variants for the unfolded systems with respect to $R_{g,gmx}$ (green line) given at the bottom in Å. The systems studied are $^{10}\text{FNIII}$, CI-2 with two different dye pairs (AF546/AF647 and AF488/AF594), CspTm with AF488/AF594 at three different labeling positions, and ClyA monomer and protomer with AF488/AF594/(B680) at different labeling sites. More details on the systems can be found in Supplementary Information S1. R_g values calculated from atomic structures with dyes ($R_{g,gmx}^{+dyes}$, red) and those derived from SAXS curves without ($R_{g,saxs}$, blue) and with dyes ($R_{g,saxs}^{+dyes}$, orange) are depicted. R_g errors derived from the Guinier linear regression are listed in Supplementary Information S5.

derived values of $R_{g,saxs}$ and $R_{g,saxs}^{+dyes}$ to be almost identical to the respective references of $R_{g,gmx}$ and $R_{g,gmx}^{+dyes}$.

Finally, we analyze R_g variants obtained from end-to-end distances presented in Figure 7. Here, we only consider the residues between the dye positions to mimic labeling at the termini. The ratios of $R_{g,gmx}^{+dyes}$ and $R_{g,saxs}^{+dyes}$ with respect to $R_{g,gmx}$ are in good agreement and both shifted to higher values as

before (see Figure 6). $R_{g,saxs}$, $R_{g,C\alpha}^{app}$, and $R_{g,R_{DA}}^{app}$ are all very similar to $R_{g,gmx}$. $R_{g,R_{DA}}^{app}$ is consistently larger than $R_{g,saxs}$, pointing to a small systematic difference in the quantities accessible to FRET and SAXS. Note that investigating IDPs in varying denaturant concentrations as done experimentally^[20] and in explicit-solvent MD simulations^[21] is not yet possible within the structure-based simulation protocol. However, this

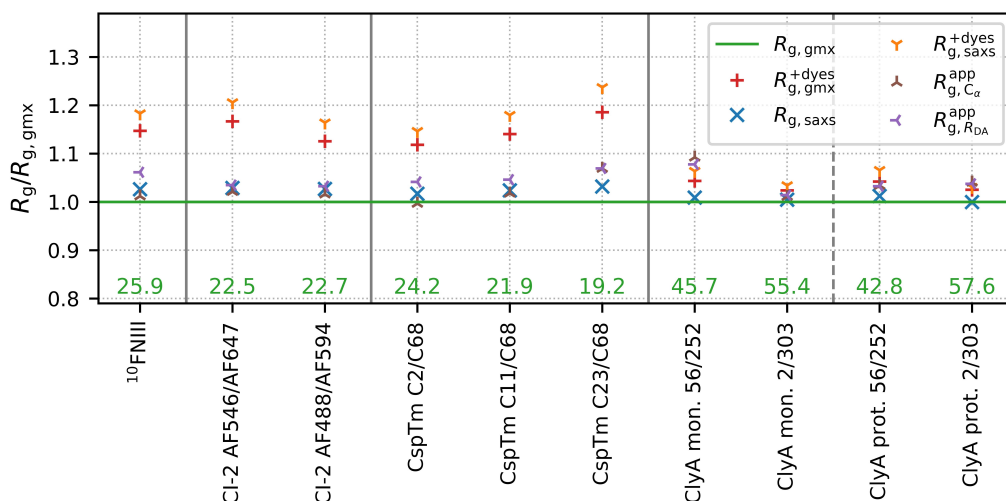


Figure 7. R_g variants for different truncated systems in the unfolded states with respect to $R_{g,gmx}$ (green line) given at the bottom in Å. We study $^{10}\text{FNIII}$, CI-2 with two different dye pairs (AF546/AF647 and AF488/AF594), CspTm with AF488/AF594 at three different labeling positions, and ClyA monomer and protomer with AF488/AF594 at different labeling sites. More details on the systems can be found in Supplementary Information S1. R_g values calculated from atomic structures with dyes ($R_{g,gmx}^{+dyes}$, red), those derived from SAXS curves without ($R_{g,saxs}$, blue) and with dyes ($R_{g,saxs}^{+dyes}$, orange), and apparent values calculated from C_α end-to-end distance ($R_{g,C\alpha}^{app}$, brown) and inter-dye distance ($R_{g,R_{DA}}^{app}$, purple) are shown. R_g errors derived from the Guinier linear regression are listed in Supplementary Information S5.

could be included by, e.g., implementing a varying overall attractive potential.

4. Conclusion

We find that FRET dyes attached to a protein significantly affect SAXS measurements on that system, as the dyes change both its size and shape. This effect is particularly pronounced for small proteins. In the case of unfolded ensembles, the difference is small however observable, while almost insignificant for larger systems. Systems appear to be larger with dyes than without, manifesting in a larger radius of gyration. In line with our expectations, the smaller the protein, the more significant is this effect. Dye types also show an effect on R_g . For unfolded ensembles, dye positions further affect the values derived, and our findings suggest a subtle but observable influence of FRET dyes on the chain dynamics. This means that, when performing both FRET and SAXS measurements on the same system, respective effects have to be taken into account in the data analysis methods applied. We find that R_g variants derived from SAXS apparently overestimate R_g for small systems, whereas underestimating it slightly for some of the larger systems. As expected, SAXS-derived variants shift to higher values for systems with dyes attached. All R_g values derived by FRET and SAXS are in good agreement, consistent with prior work suggesting that the analysis methods are the primary source of the discrepancies observed.^[19,20] However, we find the FRET-derived R_g variant to be consistently larger than the SAXS-derived value, pointing to a small systematic difference in the quantities accessible to FRET and SAXS.

Acknowledgements

We thank Pascal Friederich for the calculation of the structure of the Biotium B680 dye. The present contribution is supported by the Helmholtz Association under the joint research school “HIDSS4Health – Helmholtz Information and Data Science School for Health” and the Helmholtz Association Initiative and Networking Fund under project number ZT–I-0003. The authors gratefully acknowledge support by DFG under the GRK 2450. Open access funding enabled and organized by Projekt DEAL.

References

- [1] S. K. Burley, H. M. Berman, C. Christie, J. M. Duarte, Z. Feng, J. Westbrook, J. Young, C. Zardecki, *Protein Sci.* **2018**, 316–330.
- [2] C. D. Putnam, M. Hammel, G. L. Hura, J. A. Tainer, *Q. Rev. Biophys.* **2007**, *40*, 191–285. doi: 10.1017/S0033583507004635.
- [3] B. Schuler, H. Hofmann, *Curr. Opin. Struct. Biol.* **2013**, *23*, 36–47, doi: 10.1016/j.sbi.2012.10.008.
- [4] R. Rieger, A. Kobitski, H. Sielaff, G. U. Nienhaus, *ChemPhysChem* **2011**, *12*, 627–633. doi: 10.1002/cphc.201000693.
- [5] R. Rieger, G. U. Nienhaus, *Chem. Phys.* **2012**, *396*, 3–9. doi: 10.1016/j.chemphys.2011.03.026.
- [6] Y. Gambin, A. Schug, E. A. Lemke, J. J. Lavinder, A. C. M. Ferreón, T. J. Magliery, J. N. Onuchic, A. A. Deniz, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10153–10158.
- [7] A. G. Kikhney, D. I. Svergun, *FEBS Lett.* **2015**, *589*, 2570–2577. doi: 10.1016/j.febslet.2015.08.027.
- [8] B. Schuler, A. Soranno, H. Hofmann, D. Nettels, *Annu. Rev. Biophys.* **2016**, *45*, 207–231. ISSN 1936–122X. doi:10.1146/annurev-biophys-062215-010915. URLhttp://www.annualreviews.org/doi/10.1146/annurev-biophys-062215-010915.
- [9] A. S. Holehouse, R. V. Pappu, *Annu. Rev. Biophys.* **2018**, *47*, 19–39.
- [10] E. P. O’Brien, G. Morrison, B. R. Brooks, D. Thirumalai, *J. Chem. Phys.* **2009**, *130*, 124903.
- [11] G. Reddy, D. Thirumalai, *J. Phys. Chem. B* **2017**, *121*, 995–1009.
- [12] J. A. Riback, M. A. Bowman, A. M. Zmyslowski, C. R. Knoverek, J. M. Jumper, J. R. Hinshaw, E. B. Kaye, K. F. Freed, P. L. Clark, T. R. Sosnick, *Science* **2017**, *358*, 238–241. doi: 10.1126/science.aan5774.
- [13] J. A. Riback, M. A. Bowman, A. M. Zmyslowski, C. R. Knoverek, J. M. Jumper, E. B. Kaye, K. F. Freed, P. L. Clark, T. R. Sosnick, *Science* **2018**, *361*, eaar7949. doi:10.1126/science.aar7949.
- [14] H. Maity, G. Reddy, *J. Am. Chem. Soc.* **2016**, *138*, 2609–2616.
- [15] D. Thirumalai, H. S. Samanta, H. Maity, G. Reddy, *Trends Biochem. Sci.*, **2019**.
- [16] P. L. Clark, K. W. Plaxco, T. R. Sosnick, *J. Mol. Biol.*, **2020**.
- [17] G. Fuertes, N. Banterle, K. M. Ruff, A. Chowdhury, D. Mercadante, C. Koehler, M. Kachala, G. Estrada Girona, S. Milles, A. Mishra, P. R. Onck, F. Gräter, S. Esteban-Martin, R. V. Pappu, D. I. Svergun, E. A. Lemke, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E6342–E6351. doi: 10.1073/pnas.1704692114.
- [18] G. Fuertes, N. Banterle, K. M. Ruff, A. Chowdhury, R. V. Pappu, D. I. Svergun, E. A. Lemke, *Science* **2018**, *361*, eaau8230. doi: 10.1126/science.aau8230.
- [19] R. B. Best, W. Zheng, A. Borgia, K. Buholzer, M. B. Borgia, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, et al. *Science* **2018**, *361*, eaar7101.
- [20] A. Borgia, W. Zheng, K. Buholzer, M. B. Borgia, A. Schüler, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, R. B. Best, B. Schuler, *J. Am. Chem. Soc.* **2016**, *138*, 11714–11726. doi: 10.1021/jacs.6b05917.
- [21] W. Zheng, A. Borgia, K. Buholzer, A. Grishaev, B. Schuler, R. B. Best, *J. Am. Chem. Soc.* **2016**, *138*, 11702–11713. doi: 10.1021/jacs.6b05443.
- [22] P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu, J. N. Onuchic, *Proteins Struct. Funct. Bioinf.* **2009**, *75*, 430–441. doi: 10.1002/prot.22253.
- [23] I. Reinartz, C. Sinner, D. Nettels, B. Stucki-Buchli, F. Stockmar, P. T. Panek, C. R. Jacob, G. U. Nienhaus, B. Schuler, A. Schug, *J. Chem. Phys.* **2018**, *148*, 123321. doi: 10.1063/1.5010434.
- [24] J. D. Bryngelson, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7524–7528.
- [25] P. E. Leopold, M. Montal, J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8721–8725.
- [26] H. Frauenfelder, S. G. Sligar, P. G. Wolynes, *Science* **1991**, *254*, 1598–603.
- [27] J. N. Onuchic, P. G. Wolynes, *Curr. Opin. Struct. Biol.* **2004**, *14*, 70–75. doi: 10.1016/j.sbi.2004.01.009.

- [28] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, *Proteins Struct. Funct. Genet.* **1995**, *21*, 167–195. doi: 10.1002/prot.340210302.
- [29] A. L. Main, T. S. Harvey, M. Baron, J. Boyd, I. D. Campbell, *Cell* **1992**, *71*, 671–678. doi: 10.1016/0092-8674(92)90600-H.
- [30] T. Förster, *Ann. Phys.* **1948**, *437*, 55–75.
- [31] L. Stryer, *Annu. Rev. Biochem.* **1978**, *47*, 819–846. doi: 10.1146/annurev.bi.47.070178.004131.
- [32] E. Sisamakris, A. Valeri, S. Kalinin, P. J. Rothwell, C. A. M. Seidel. Chapter 18 - Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection. In *Single Molecule Tools, Part B: Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods*, volume 475 of *Methods in Enzymology*, pages 455–514. Academic Press, **2010**. doi: [https://doi.org/10.1016/S0076-6879\(10\)75018-7](https://doi.org/10.1016/S0076-6879(10)75018-7).
- [33] D. I. Svergun, M. H. J. Koch, *Rep. Prog. Phys.* **2003**, *66*, 1735–1782. doi: 10.1088/0034-4885/66/10/R05.
- [34] P. Debye, *Ann. Phys.* **1915**, *351*, 809–823. doi: 10.1002/andp.19153510606.
- [35] A. Guinier, G. Fournet, K. L. Yudowitch, Small-angle scattering of X-rays 1955.
- [36] C. Clementi, H. Nymeyer, J. N. Onuchic, *J. Mol. Biol.* **2000**, *298*, 937–953. doi: 10.1006/jmbi.2000.3693.
- [37] M. Fernanda Rey-Stolle, M. Enciso, A. Rey, *J. Comput. Chem.* **2009**, *30*, 1212–1219.
- [38] B. Lutz, C. Sinner, G. Heuermann, A. Verma, A. Schug, *Bioinformatics* **2013**, *29*, 2795–2796. doi: 10.1093/bioinformatics/btt478.
- [39] C. Sinner, B. Lutz, S. John, I. Reinartz, A. Verma, A. Schug, *Isr. J. Chem.* **2014**, *54*, 1165–1175. doi: 10.1002/ijch.201400012.
- [40] J. I. Sulkowska, J. K. Noel, J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17783–17788.
- [41] Y. Levy, J. N. Onuchic, *Acc. Chem. Res.* **2006**, *39*, 135–142. doi: 10.1021/ar040204a.
- [42] L. L. Chavez, J. N. Onuchic, C. Clementi, *J. Am. Chem. Soc.* **2004**, *126*, 8426–32. doi: 10.1021/ja049510+.
- [43] A. Schug, M. Weigt, J. N. Onuchic, T. Hwa, H. Szymant, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 22124–22129. doi: 10.1073/pnas.0912100106.
- [44] A. E. Dago, A. Schug, A. Procaccini, J. A. Hoch, M. Weigt, H. Szymant, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1733–E1742.
- [45] E. De Leonardis, B. Lutz, S. Ratz, S. Cocco, R. Monasson, A. Schug, M. Weigt, *Nucleic Acids Res.* **2015**, *43*, 10444–10455.
- [46] J. I. Sulkowska, F. Morcos, M. Weigt, T. Hwa, J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10340–10345.
- [47] M. Weiel, I. Reinartz, A. Schug, *PLoS Comput. Biol.* **2019**, *15*, e1006900.
- [48] P. C. Whitford, A. Ahmed, Y. Yu, S. P. Hennesly, F. Tama, C. M. T. Spahn, J. N. Onuchic, K. Y. Sanbonmatsu, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18943–18948.
- [49] C. Clementi, P. A. Jennings, J. N. Onuchic, *J. Mol. Biol.* **2001**, *311*, 879–890. doi: 10.1006/jmbi.2001.4871.
- [50] H. Lammert, A. Schug, J. N. Onuchic, *Proteins Struct. Funct. Bioinf.* **2009**, *77*, 881–891. doi: 10.1002/prot.22511.
- [51] J. K. Noel, A. Schug, A. Verma, W. Wenzel, A. E. Garcia, J. N. Onuchic, *J. Phys. Chem. B* **2012**, *116*, 6880–6888. doi: 10.1021/jp212623d.
- [52] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, J. N. Onuchic, *Nucleic Acids Res.* **2010**, *38*, W657–W661. doi: 10.1093/nar/gkq498.
- [53] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, H. J. C. Berendsen, *Gromacs User Manual version 4.5.6*. URL: <http://www.gromacs.org>, 2010.
- [54] R. Pankov, K. M. Yamada, *J. Cell Sci.* **2002**, *115*, 3861–3863.
- [55] C. M. Williams, A. J. Engler, R. D. Slone, L. L. Galante, J. E. Schwarzbauer, *Cancer Res.* **2008**, *68*, 3185–3192.
- [56] C. A. McPhalen, M. N. G. James, *Biochemistry* **1987**, *26*, 261–269.
- [57] S. E. Jackson, N. elMasry, A. R. Fersht, *Biochemistry* **1993**, *32*, 11270–11278.
- [58] L. S. Itzhaki, D. E. Otzen, A. R. Fersht, *J. Mol. Biol.* **1995**, *254*, 260–288.
- [59] T. R. Killick, S. M. V. Freund, A. R. Fersht, *FEBS Lett.* **1998**, *423*, 110–112.
- [60] J. L. Neira, L. S. Itzhaki, D. E. Otzen, B. Davis, A. R. Fersht, *J. Mol. Biol.* **1997**, *270*, 99–110.
- [61] S. L. Kazmirski, K.-B. Wong, S. M. V. Freund, Y.-J. Tan, A. R. Fersht, V. Daggett, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4349–4354.
- [62] W. Kremer, B. Schuler, S. Harrieder, M. Geyer, W. Gronwald, C. Welker, R. Jaenicke, H. R. Kalbitzer, *Eur. J. Biochem.* **2001**, *268*, 2527–2539. doi: 10.1046/j.1432-1327.2001.02127.x.
- [63] A. J. Wallace, T. J. Stillman, A. Atkins, S. J. Jamieson, P. A. Bullough, J. Green, P. J. Artymiuk, *Cell* **2000**, *100*, 265–76.
- [64] M. Mueller, U. Grauschopf, T. Maier, R. Glockshuber, N. Ban, *Nature* **2009**, *459*, 726–730.
- [65] URL: <http://www.thermofisher.com>.
- [66] URL: <https://biotium.com>.
- [67] D. I. Svergun, C. Barberato, M. H. J. Koch, *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
- [68] J. Henriques, L. Arleth, K. Lindorff-Larsen, M. Skepö, *J. Mol. Biol.* **2018**, *430*, 2521–2539.
- [69] P. J. Flory, *J. Chem. Phys.* **1949**, *17*, 303–310.

Manuscript received: January 31, 2020
Revised manuscript received: May 26, 2020
Version of record online: July 22, 2020