



Cite this: *Nanoscale*, 2025, **17**, 3322

## Determination of absolute intramolecular distances in proteins using anomalous X-ray scattering interferometry†

Samuel Stubhan, <sup>a,b</sup> Anna V. Baptist, <sup>a,c</sup> Caroline Körösy, <sup>b</sup> Alessandra Narducci, <sup>d</sup> Gustavo Gabriel Moya Muñoz, <sup>d</sup> Nicolas Wendler, <sup>d</sup> Aidin Lak, <sup>a</sup> Michael Sztucki, <sup>e</sup> Thorben Cordes\*<sup>d,f</sup> and Jan Lipfert <sup>\*a,b,g</sup>

Biomolecular structures are typically determined using frozen or crystalline samples. Measurement of intramolecular distances in solution can provide additional insights into conformational heterogeneity and dynamics of biological macromolecules and their complexes. The established molecular ruler techniques used for this (NMR, FRET, and EPR) are, however, limited in their dynamic range and require model assumptions to determine absolute distance or distance distributions. Here, we introduce anomalous X-ray scattering interferometry (AXSI) for intramolecular distance measurements in proteins, which are labeled at two sites with small gold nanoparticles of 0.7 nm radius. We apply AXSI to two different cysteine-variants of maltose binding protein in the presence and absence of its ligand maltose and find distances in quantitative agreement with single-molecule FRET experiments. Our study shows that AXSI enables determination of intramolecular distance distributions under virtually arbitrary solution conditions and we anticipate its broad use to characterize protein conformational ensembles and dynamics.

Received 16th August 2024,  
Accepted 1st December 2024

DOI: 10.1039/d4nr03375b

rsc.li/nanoscale

## Introduction

Atomic resolution biomolecular structures are typically determined using frozen or crystalline samples with cryo-EM<sup>1,2</sup> and X-ray methods,<sup>3–6</sup> respectively, or in aqueous solution at room temperature with NMR.<sup>7,8</sup> Measurement of intramolecular distances can provide additional insights into the structure and dynamics of biological macromolecules and their complexes. A sufficient number of intramolecular distances enables the determination of high-resolution structures and can also

provide critical information about the conformational ensemble and dynamics of macromolecules based on molecular ruler techniques such as PELDOR/DEER (EPR)<sup>9–11</sup> or single-molecule Förster resonance energy transfer (smFRET).<sup>12–27</sup>

Also small-angle X-ray scattering can provide information about intramolecular distances in biomolecules. Notably the  $P(r)$  distribution, *i.e.* the Fourier transform of the scattering intensity profile  $I(q)$ , is a histogram of pairwise distances and can be readily obtained from scattering data.<sup>28–30</sup> However,  $P(r)$  does not contain information about which specific pair contributed to a given distance. Labeling macromolecules with electron-rich labels at two positions – *e.g.* heavy atoms,<sup>31–33</sup> ions, or small gold nanocrystals – combined with SAXS as readout can overcome this limitation. In conventional X-ray scattering interferometry (XSI)<sup>34–40</sup> with gold labels, the label-label interference term is isolated from other scattering contributions by measuring multiple samples, including the double-labeled, two single-labeled, and unlabeled macromolecule.<sup>34,35</sup> An alternative approach to separating the gold–gold term, termed anomalous XSI or AXSI, uses anomalous small-angle X-ray scattering (ASAXS) and relies on the energy-dependence of the gold scattering signal.<sup>32,33,41–43</sup> A regularized Fourier transform of the gold–gold scattering term then directly provides the distribution of distances  $P(d)$  between the gold labels. (A)XSI has several advantages compared to other molecular ruler techniques: (i) it provides distance distributions on

<sup>a</sup>Department of Physics and Center for NanoScience, LMU Munich, Amalienstr. 54, 80799 Munich, Germany

<sup>b</sup>Soft Condensed Matter and Biophysics, Department of Physics and Debye Institute for Nanomaterials Science, Utrecht University, Princetonplein 1, 3584 CC Utrecht, The Netherlands. E-mail: J.Lipfert@uu.nl

<sup>c</sup>Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

<sup>d</sup>Physical and Synthetic Biology, Faculty of Biology, LMU Munich, Großhadernerstr. 2-4, 82152 Planegg-Martinsried, Germany. E-mail: cordes@bio.lmu.de

<sup>e</sup>ESRF, 71 Avenue des Martyrs, 38043 Grenoble, France

<sup>f</sup>Biophysical Chemistry, Faculty of Chemistry and Chemical Biology, Technische Universität Dortmund, Otto-Hahn-Str. 4a, 44227 Dortmund, Germany

<sup>g</sup>Institute for Physics, Augsburg University, Universitätsstrasse 1, 86159 Augsburg, Germany

† Electronic supplementary information (ESI) available: Supplementary Materials and methods, Tables S1–S3 and Fig. S1–S18. See DOI: <https://doi.org/10.1039/d4nr03375b>



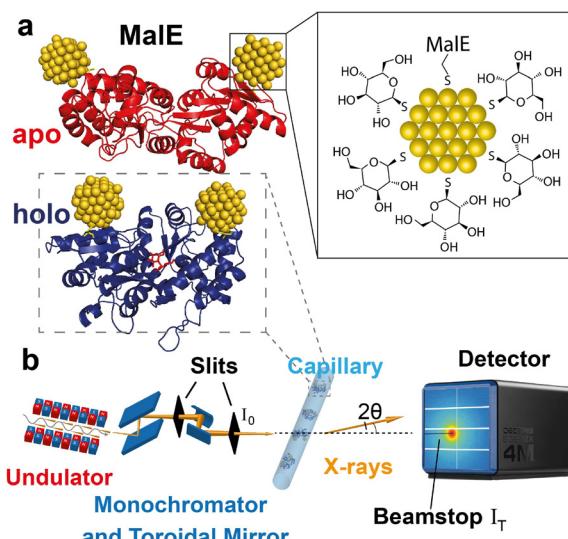
an absolute length scale, based on the fact that it is straightforward to measure the momentum transfer  $q$  ( $q = 4\pi \sin(\theta)/\lambda$ , where  $2\theta$  is the total scattering angle and  $\lambda$  the X-ray wavelength) on an absolute scale; (ii) (A)XSI can provide the full distribution of intramolecular distances (not only mean inter-label distances), without broadening through *e.g.*, photophysics, as seen in FRET or temporal dynamics, due to the short interaction time of X-ray photons with the sample;<sup>34</sup> (iii) it can readily be applied to distances  $>10$  nm, which remains very challenging for NMR, EPR,<sup>9,10,44</sup> or FRET;<sup>45</sup> (iv) finally (A)XSI distance measurements are not sensitive to label orientation or the specific label environment, unlike FRET approaches.<sup>24</sup> Insensitivity of the distance measurement to the environment is advantageous for measurements to determine conformational changes in response to *e.g.*, denaturant,<sup>46</sup> salt,<sup>47–49</sup> or ligand concentration. ASAXS-based AXSI measurements have the advantage that they in principle only require preparation of the double-labeled sample, as opposed to traditional XSI, which requires matching single-labeled constructs as well, which *e.g.* requires expression of additional protein constructs. However, so far AXSI has only been established experimentally for DNA constructs, which can be labeled in a straightforward way.<sup>43</sup>

Here, we demonstrate AXSI intramolecular distance measurements in proteins that undergo conformational changes upon ligand binding. We use MalE, the soluble periplasmic component of the maltose import system of *E. coli*,<sup>50–52</sup> which has been characterized in detail previously by smFRET experiments and other structural methods.<sup>24,50–57</sup> MalE undergoes a conformational change from an open/apo to a closed/holet state upon binding maltose with a dissociation constant<sup>24</sup>  $K_d$  of  $\sim 1\text{--}2$   $\mu\text{M}$ . We analyze AXSI data for two double-cysteine variants of MalE and extract distance distributions of the apo and holo states that exhibit sharp main peaks. The main peak position can be determined with Ångström precision and the measured distances are in good agreement, within experimental error, with quantitative distance determination *via* smFRET.

## Results

### Labeling of MalE mutants *via* thiol–gold chemistry

(A)XSI measurements for computing intramolecular distances necessitate site-specific attachment of gold-labels. For this purpose, we used variants of the 42.4 kDa maltose-binding protein MalE that comprise two cysteines in the two different lobes of the protein (ESI Materials and methods: over-expression, isolation and refolding of MalE proteins†), either at positions 36 and 352 (MalE<sub>31–212</sub>; Fig. 1) or at positions 31 and 212 (MalE<sub>36–352</sub>).<sup>24,52</sup> The pairs of label positions were chosen to be situated in the different lobes of the protein and use positions that have a high probability of being successfully labeled based on their physico-chemical properties,<sup>58</sup> but otherwise selected randomly. All four selected positions are at the surface of the protein, but have average B-factors and are



**Fig. 1** Schematic of anomalous X-ray scattering measurements to determine intramolecular distance distributions. (a) Illustrations of double-labeled MalE in the apo and holo state with gold labels at amino acid positions 36 and 352 (rendered from PDB ID 1OMP (red – apo) and 1ANF (blue – holo), respectively. Gold nanocrystals are positioned using FPS calculations<sup>19</sup>). The zoom depicts the thioglucose shell on the gold NPs as well as the S–Au attachment to the protein. (b) Illustration of the SAXS experiment. The undulator and X-ray optics at the synchrotron beam line provide X-rays with tunable energy. The monochromator is used to select energies and collimated X-rays are scattered by the sample in a quartz capillary. The incident intensity  $I_0$  is measured in front of the capillary and the transmitted intensity  $I_T$  is measured at the beamstop. Scattered photons are collected in an Eiger2X 4M pixel detector.

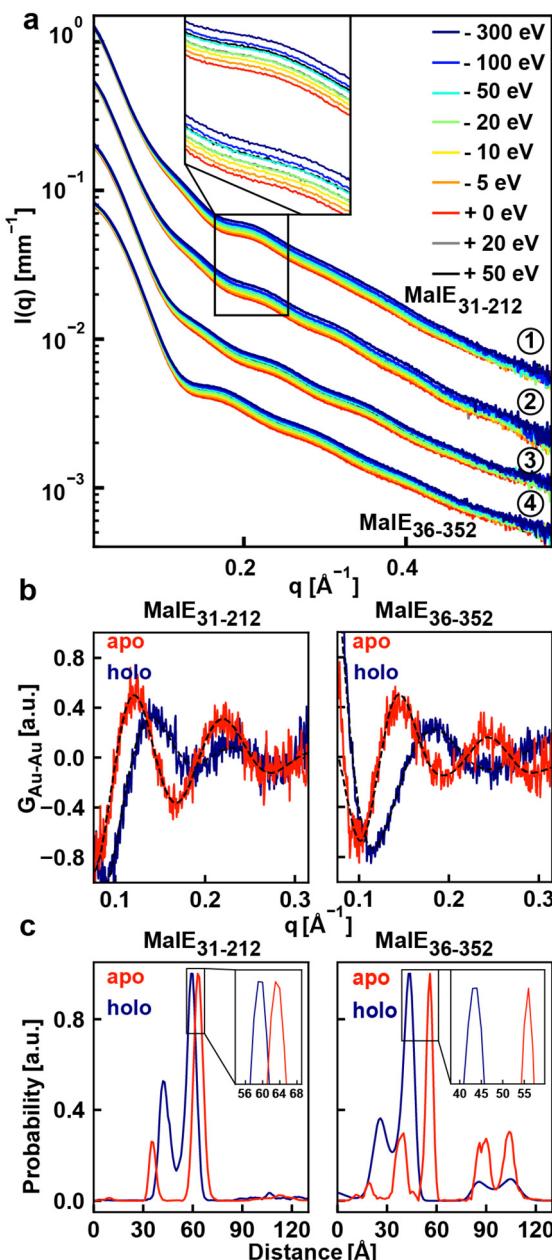
in secondary structure elements (residues 31, 212, and 352 are in  $\alpha$ -helices and 36 in a  $\beta$ -sheet). MalE lacks native cysteines, thus allowing site-specific attachment of gold labels *via* chemical coupling to thiols<sup>52</sup> (Fig. 1a). Thioglucose coated gold nanoparticles (NPs) were synthesized following a one-phase Brust–Schiffirin method<sup>59</sup> (ESI Materials and methods: synthesis of gold nanoparticles†) and exhibit a radius of 0.7 nm with high monodispersity<sup>34,35,40</sup> (Fig. S1 and S2†). Proteins were labeled with the NP on nickel-sepharose columns<sup>11,52,60</sup> and subsequently purified by size exclusion chromatography (Fig. S3 and S4†).

### SAXS and ASAXS measurements of MalE constructs

We carried out SAXS measurements at both fixed and variable X-ray energies at beamline ID02 of the European Synchrotron Radiation Facility (ESRF;<sup>61–64</sup> and ESI Materials and methods: ASAXS measurements†; Fig. 1). Control measurements of the unlabeled protein at fixed energy reveal SAXS profiles indicative of a monodisperse sample and show systematic, but subtle changes upon addition of 10 mM maltose, with radii of gyration in good agreement with predictions from the crystal structures in the open and closed conformations (Fig. S5†).

ASAXS data were recorded for double-labeled, single-labeled, and unlabeled MalE constructs by recording scattering profiles at 9 energies around the gold L-III absorption edge





**Fig. 2** ASAXS data and distance distribution for MalE labeled at position 36 and 352 and MalE labeled at position 31 and 212. (a) ASAXS measurements of both double-labeled MalE mutants with and without maltose measured at 9 energies around the gold L-III absorption edge. ① MalE<sub>31-212</sub> ② MalE<sub>31-212</sub> with 10 mM maltose ③ MalE<sub>36-352</sub> and ④ MalE<sub>36-352</sub> with 10 mM maltose. Data are vertically offset for clarity (by scaling factors ①:10, ②:6, ③:2 and ④:1). Indicated energies are relative to the Au L-III edge. Alternative representations of the data are shown in Fig. S6.† (b) Gold–gold scattering interference terms for MalE<sub>36-352</sub> (left) and MalE<sub>31-212</sub> (right) in the absence and presence of 10 mM maltose after correction for 60% single-labeled and 30% unlabeled contributions. (c) Distance distributions  $P(d)$  obtained by maximum entropy inversion from the interference terms in panel b. The insets show a zoom on the main peaks in the distance distributions.

at 11.919 keV (Fig. 2 and Fig. S6†). Ascorbic acid was added to the buffer for all ASAXS measurements to reduce radiation damage for high signal-to-noise measurements (Fig. S7†). The

scattering profiles at different energies for the double-labeled MalE<sub>31-212</sub> and MalE<sub>36-352</sub> constructs both show oscillations (Fig. 2a), in particular in the range  $q = 0.1\text{--}0.2 \text{ \AA}^{-1}$ , which are absent in the unlabeled data (Fig. S5†), indicative of the gold–gold interference contribution.<sup>40</sup> The oscillations shift upon addition of maltose (Fig. 2a, inset), suggesting a modulation of gold–gold interference term upon addition of maltose. Further, the scattering profiles show systematic changes with X-ray energy: the intensity decreases when approaching the L-III absorption edge.

HPLC purification of the double-labeled sample removes dimers and aggregates from the monomer peak, however, single-labeled and unlabeled species remain in the solution, visible as bands in gel electrophoretic analysis of the sample. From the gel, we estimate ~60% unlabeled, ~30%, single-labeled, and ~10% double-labeled sample, which agrees with a protein:gold NP concentration ratio of 1:0.5 (Fig. S3 and S4†). Despite of the relatively low labeling efficiency, we still get a robust signal, since the gold particles scatter strongly with ~85 atoms and a ~40 times higher electron density contrast compared to proteins. Additionally, in principle, only the double-labeled sample will contribute a gold–gold term to the scattering pattern. However, the presence of unlabeled and single-labeled species might deteriorate the signal. Since we conducted ASAXS measurements also for single-labeled mutants MalE<sub>31</sub>, MalE<sub>36</sub>, MalE<sub>212</sub>, MalE<sub>352</sub> and unlabeled samples, we can subtract their scattering contributions from the double-labeled sample (Fig. S8†). We tested and refined the influence of subtracting single- and unlabeled protein contributions in the ASAXS data analysis (see below).

#### Determination of the gold label–gold label distance distribution from ASAXS data

We analyzed the ASAXS data and determined the gold–gold scattering contribution from the corrected and energy-dependent scattering data with a matrix inversion approach described previously.<sup>42,43</sup> The approach takes into account the atomic scattering factor of gold and the form factor of the gold spheres (Fig. S1 and S2†) and exploits the fact that the atomic scattering factors for non-gold atoms show minimal energy dependence within the chosen energy range (Fig. S9†). The matrix inversion yields the gold–gold structure factor  $G_{\text{Au–Au}}$ , which was corrected by a constant offset by subtracting the mean<sup>43</sup> (Fig. 2b) and Fourier transformed with a maximum entropy algorithm<sup>35,40,43,65</sup> to obtain the real-space distance distributions  $P(d)$  (Fig. 2c).

Using this procedure, we determined the gold label–gold label distance distributions  $P(d)$  for the two MalE variants, MalE<sub>31-212</sub> and MalE<sub>36-352</sub>, both in absence and presence of a high (10 mM) concentration of maltose (Fig. 2c). Under all conditions, the  $P(d)$  distributions exhibit a major peak and additional, smaller peaks at smaller and larger distances. We find that peaks at smaller and larger distances ( $>80 \text{ \AA}$ ) are variable from data set to data set and are sensitive to details of the single- and unlabeled subtraction and maximum entropy procedure and are likely due to imperfections of the experimental data<sup>40</sup> and possibly due to the presence of a small fraction of



dimeric samples (Fig. S3†). In contrast, the positions of the main peak in either condition are robust (Tables S1, S2 and Fig. S10†).

### Analysis of uncertainty in AXSI measurements

To determine the uncertainties of the measured distance distributions, we analyzed the variability introduced by several factors. First, we tested the uncertainty introduced by the maximum entropy algorithm used to compute the  $P(d)$  distributions. For each measured gold–gold term  $G_{\text{Au–Au}}$ , we carried out 20 repeat runs of the maximum entropy algorithm and fitted the main peak with a Gaussian to determine its position and standard deviation (Fig. S11†). In each run the MemSys5-based<sup>65</sup> maximum entropy algorithm is randomly selecting and deleting 10% of the scattering data<sup>34</sup> in contiguous blocks of 2%. We find only small deviations between repeat runs of the maximum entropy algorithm: the main peak positions exhibit standard deviation of 0.2–0.7 Å computed from 20 repeat runs of the maximum entropy algorithm each for different MalE mutants and conditions (Tables S1 and S2†). In addition, we performed control calculation using a different inversion scheme based on Tikhonov regularization (Fig. S12 and S13†).

We again find very similar values (within <1 Å) for the position of the main peak, but a larger width of the central peak and more pronounced secondary peaks at larger and smaller distances.

Next, we test the sensitivity to differences in background subtraction, in particular from subtracting unlabeled and single-labeled contributions to the scattering pattern. We subtracted varying quantities of single-labeled and unlabeled protein contributions from the scattering pattern and computed  $P(d)$  functions as described above. We find that the secondary peaks in the distance distribution are smallest if 60% unlabeled and 30% single-labeled contributions are subtracted (red curve in Fig. S9† and data shown in Fig. 2) for both mutants MalE<sub>36–352</sub> and MalE<sub>31–212</sub>, which agrees with the estimated fractions from gel analysis (Fig. S3†). Importantly, adjusting the amount of single and unlabeled contributions subtracted or even using the scattering data without subtraction affects only the level and position of the secondary peaks and changes the main peak position by at most 1 Å (all measured peak positions in Tables S1 and S2†). Thus, our analysis suggests that while the unlabeled and single-labeled contributions add to the level of experimental noise and affect the exact shape of the  $P(d)$  function, the main peak corresponding to the gold label–gold label distance is robust against these perturbations.

Finally, we tested the reproducibility of the AXSI measurement by performing a repeat measurement for each sample. We find that the mean peak position for repeat measurements only varies by at most 0.7 Å (all measured peak positions in Tables S1 and S2†). Taken together, our error analysis suggests that we can determine the center of the main peak in the  $P(d)$  distributions to better than 1 Å (Tables S1 and S2†), in agreement with previous analyses of (A)XSI measurements for

nucleic acids.<sup>34,40,43</sup> As a control, we determined the gold–gold structure factor  $G_{\text{Au–Au}}$  using conventional XSI,<sup>34,35,40</sup> *i.e.* with measurements at only a single X-ray energy of the double-labeled, both single-labeled, and the unlabeled protein samples with SAXS. The control measurement for MalE<sub>36–352</sub> in the absence of maltose gives a main peak in the distance distributions in excellent agreement with the results of the AXSI analysis (Fig. S14†).

### Determination of MalE intramolecular distances by smFRET

To provide a reference and enable direct comparison of the results with an established technique, we performed distance measurements on MalE<sub>31–212</sub> and MalE<sub>36–352</sub> by single-molecule FRET using alternating laser excitation (ALEX)<sup>66–68</sup> (Fig. 3a and ESI Methods†). We employed a data analysis approach similar to a recent multi-lab FRET comparison study<sup>24</sup> to extract mean interprobe distances between donor (Alexa 555) and acceptor fluorophore (Alexa 647) from intensity-based single-molecule measurements<sup>20,22,66–70</sup> based on the Förster relation (Fig. 3b).

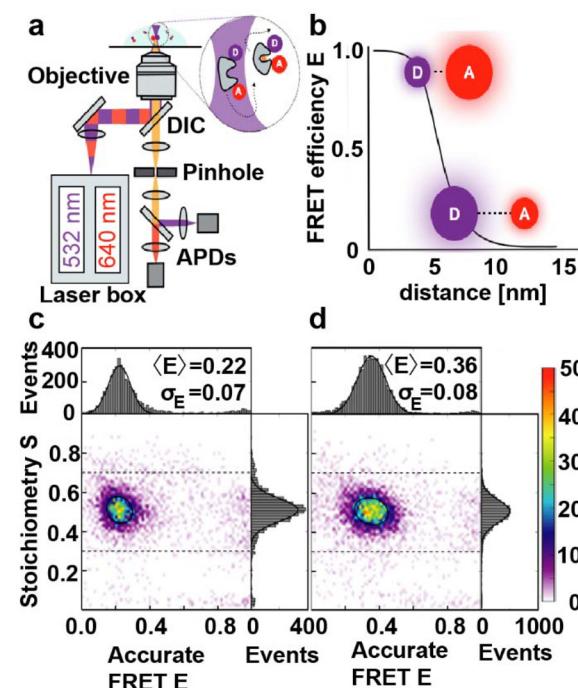
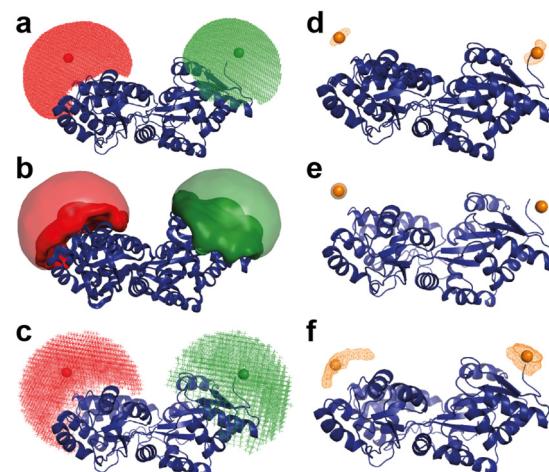


Fig. 3 Monitoring conformational changes in MalE by smFRET with ALEX. (a) Schematic overview of an ALEX confocal microscopy setup with two fiber-coupled modulated laser sources for alternating excitation of donor (D) and acceptor (A) dyes. The laser is expanded, collimated, and directed into an objective with a high numerical aperture through a dichroic mirror (DIC). The objective is used for excitation and detection of the fluorescence of individual proteins in a diffraction limited excitation spot. Subsequently, a pinhole spatially filters the fluorescence before it undergoes spectral separation into green and red detection channels. (b) Schematic plot of the FRET efficiency ( $E$ ) as a function of the distance between a donor fluorophore and an acceptor. (c and d) Corrected accurate ES-histograms of MalE<sub>31–212</sub> labeled with Alexa555 and Alexa647 depicting the ligand-free open (c) and the liganded closed (d) conformation.

Both  $\text{MalE}_{31-212}$  and  $\text{MalE}_{36-352}$  are designed to show changes from larger to smaller interprobe distances upon maltose binding. We observe shifts from low to intermediate FRET efficiency values  $E$  from the apo to holo state, as expected (Fig. 3c, d and Fig. S15†). The mean  $E$  values change from 0.22 to 0.36 for  $\text{MalE}_{31-212}$  and from 0.4 to 0.67 for  $\text{MalE}_{36-352}$ , corresponding to a reduction in the calculated interprobe distances (Table 1).

### Modeling the label geometry for FRET and AXSI measurements

AXSI and FRET give mean distances that are in good agreement (Table 1). However, we note that both AXSI and FRET measure the distances between the respective labels and not directly between the positions of the labeled residues. Therefore, the attachment, size, and flexibility of the organic dyes or gold labels need to be considered when interpreting distance measurements.<sup>20,71</sup> Taking into account the label geometries is particularly relevant in light of the very high resolution of AXSI measurements, where distances are determined to better than 1 Å (Table 1), which is much smaller than the label and linker sizes and also smaller than the spatial resolution of FRET<sup>24,72</sup> which is on the order of 2–5 Å. Taking the crystallographic structures of MalE as a starting point, we simulate label positions using label parameters summarized in ESI Table S3† and following several different approaches (Fig. 4). First, we calculated the label distances with FPS (“FRET positioning and screening”),<sup>19,73</sup> which generates accessible



**Fig. 4** Simulations of label positions for FRET dyes and gold nanoparticles using FPS,<sup>19</sup> FRETraj,<sup>74</sup> and a coarse-grained computation of accessible volumes. All panels show  $\text{MalE}_{36-352}$  in the apo state (PDB ID: 1OMP). (a)–(c) Positions for the FRET dyes Alexa555 and Alexa647 are shown as colored clouds as computed with FPS (a), FRETraj (b), and coarse-grained simulations (c). (d)–(f) Positions of the gold nanoparticles used for AXSI measurements computed with FPS (d), FRETraj (e), and coarse-grained simulations (f). The geometrical parameters used in the calculations are in Table S3.† The resulting mean label–label distances and their standard deviations are in Table 1.

**Table 1** Intramolecular distances in  $\text{MalE}_{31-212}$  and  $\text{MalE}_{36-352}$  in the absence (apo state) and presence of 10 mM maltose (holo state). Values reported for AXSI are the mean  $\pm$  standard deviation of the peak position averaged over repeat measurements and different levels of background subtractions (all values in ESI Tables S1 and S2†). Values for FRET are the mean  $\pm$  standard deviation from three technical repeats, using the Förster radius of  $R_0 = 5.1$  nm and quantum yield determined previously for this dye pair.<sup>11</sup> Errors in the distances from uncertainties in  $R_0$  and quantum yield<sup>11</sup> are 0.3–0.5 nm and, therefore, smaller than statistical errors from technical repeats. Simulated values are the mean label distances  $\pm$  the standard deviations of simulated distances. For FRET labels we report the results of the established procedures FPS and FRETraj; the coarse-grained simulations are very similar to the FPS results. For AXSI labels we conversely report values from FPS and the coarse-grained method, since FRETraj gave very similar values to FPS (Fig. 4 and Fig. S16†)

	$\text{MalE}_{31-212}$ apo	$\text{MalE}_{31-212}$ holo	$\text{MalE}_{36-352}$ apo	$\text{MalE}_{36-352}$ holo
<b>Measurements (mean <math>\pm</math> error) [Å]</b>				
AXSI	$63.7 \pm 0.5$	$59.4 \pm 0.5$	$56.0 \pm 0.3$	$43.1 \pm 0.6$
FRET	$63.0 \pm 2.4$	$56.3 \pm 2.2$	$54.1 \pm 2.1$	$44.9 \pm 1.7$
<b>Simulations (mean <math>\pm</math> standard deviation) [Å]</b>				
FPS AXSI	$67.9 \pm 0.5$	$59.7 \pm 0.4$	$60.7 \pm 1.2$	$44.6 \pm 0.7$
Coarse-grained AXSI	$64.4 \pm 2.8$	$56.6 \pm 2.0$	$59.0 \pm 4.4$	$42.7 \pm 4.5$
FPS FRET	$68.6 \pm 8.3$	$56.7 \pm 9.5$	$60.5 \pm 11.3$	$44.3 \pm 11.3$
FRETraj FRET	$63.0 \pm 9.6$	$53.8 \pm 10.0$	$56.8 \pm 11.6$	$42.2 \pm 11.8$
$C_\beta$ distance	51.5	42.6	50.6	39.8

volumes for each dye and computes the distance distribution assuming random sampling of the accessible volume (Table 1, “FPS”). Second, we use FRETraj<sup>74</sup> to compute distances based on accessible contact volumes (ACVs) (Table 1, “FRETraj”), which have been shown to provide a better estimate of label–label distances than the full accessible volume if the dyes interact with the protein surface.<sup>24</sup> Finally, we compute accessible volumes using a simple coarse-grained sampling (see ESI Methods†) and calculate the mean and standard deviation of distances in the sampled positions (Table 1, Fig. 4 and Fig. S16†). In all three computational approaches the crystal structure is treated as fixed.

To model the label positions for AXSI measurements, we adopted the same procedures considering the size and attachment of the gold nanoparticles. The gold nanoparticles, in contrast to fluorescent dyes, are directly attached to the sulfur atom of the cysteine residues. Therefore, the  $C_\beta$ –Au distance is only  $\sim 3$  Å. The attached gold NP have a radius of 7 Å, resulting in a distance from attachment atom  $C_\beta$  to the center of the label of  $\sim 10$  Å. The much more confined attachment of the gold nanoparticles compared to the fluorescent dyes used for FRET results in much more narrow predicted distance distributions from the modeling of the label and its linker: the predicted distributions of label positions for FRET labels have standard deviations of  $\sim 10$  Å in contrast to only  $\sim 0.5$ –4 Å for our AXSI labels (Table 1).

### Comparison of distances from AXSI and FRET and modeling

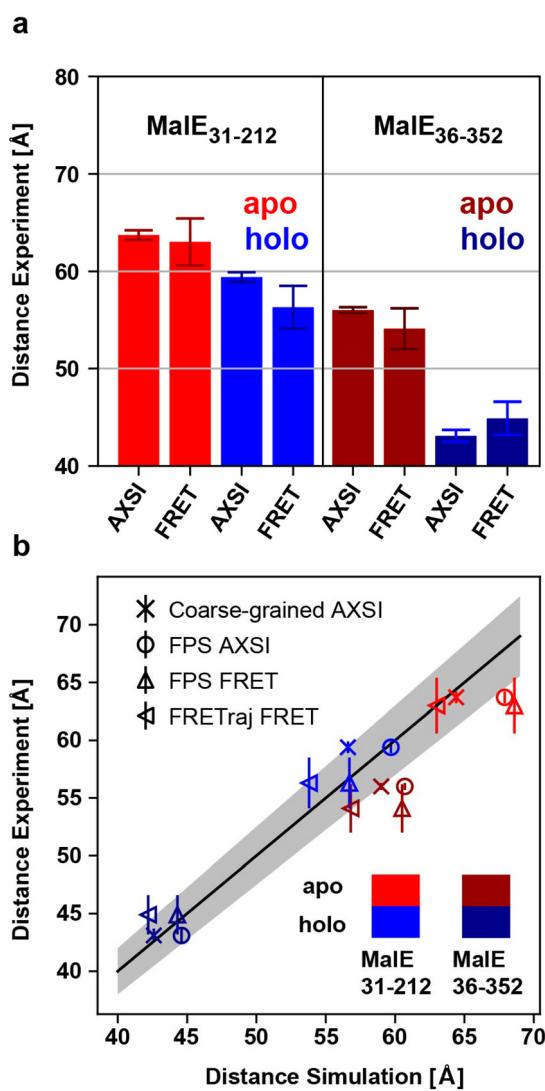
The intramolecular distances determined experimentally by AXSI and FRET agree for  $\text{MalE}_{31-212}$  and  $\text{MalE}_{36-352}$  in both the



apo and holo state within error (Table 1 and Fig. 5). This close agreement, despite the different labels used, suggests that the different physico-chemical properties and their differences in geometry do not significantly affect the mean positions of the labels and do not bias or perturb the conformations of the protein. Comparing the experimentally determined distances to the modeled distance distributions, we find good agreement of the mean distances in the closed or holo state for both MalE<sub>31-212</sub> and MalE<sub>36-352</sub>. In the closed state, all of the approaches to model the mean distances give fairly similar results, in particular for MalE<sub>36-352</sub>, due to the direction of the

label attachment being approximately perpendicular to the vector connecting the label centers (Fig. 1a). In contrast, in the open or apo state the predicted distances tend to be larger than the experimentally determined values, in particular for MalE<sub>36-352</sub> (Fig. 5). We note that for the open conformation the details of the modeling play a larger role, compared to the closed conformation. In particular, the predictions based on the accessible surface volume using FRETraj fit the experimental data for FRET better than the accessible volume-based predictions, suggesting that the fluorescent labels might have some tendency to stick to the protein's surface.<sup>24,73</sup> The differences in the apo state might also indicate that the protein conformations in solution could deviate from the crystal structure. As a plausibility test, we used a simple elastic network model approach<sup>75</sup> to deform the structure of the open conformation of MalE (Fig. S17†) and find that a simple deformation of the protein along the first normal mode could explain the observed difference between measured and predicted distances.

Unlike in FRET measurements, AXSI provides, at least in principle, the full distance distribution.<sup>34,43</sup> The mean peaks of the  $P(d)$  distributions are fairly narrow, with a width of typically  $\sim 2\text{--}5 \text{ \AA}$  ("peak width"-column in Tables S1 and S2†), which is similar to the distributions obtained by modeling the relatively inflexibly attached gold nanoparticles, assuming otherwise static protein structures, suggesting that the labeled residues of MalE and adopt a fairly static conformation in solution, as might be expected for a well-folded, globular protein. However, we note that the width of the  $P(d)$  is less well defined and more difficult to determine from the experimental data than the main peak position, suggesting that further improvements of the method are necessary to accurately determine and interpret the full molecular ensemble. In particular, a higher labeling efficiency would be desirable, in particular since previous (A)XSI measurements on nucleic acids with a high labeling efficiency have successfully provided full distance distribution information.<sup>43</sup>



**Fig. 5** Comparison of AXSI and FRET measurements and structural modeling. (a) Experimentally determined distances from AXSI and FRET for both MalE variants. Errorbars depict experimental errors (see main text). (b) Comparison of experimentally determined distances and the structural models: coarse-grained for AXSI labels (crosses), FPS for AXSI labels (circles), FPS for FRET labels (upward triangles) and FRETraj for FRET labels (left triangles). Experimental uncertainties of the respective techniques are shown. The solid line marks a 1:1 relation and the grey area indicates 5% deviation.

## Discussion

In summary, we demonstrate accurate intramolecular distance measurements using AXSI for a protein that undergoes ligand-induced conformational motion. Mean distances can be determined very precisely (within  $<1 \text{ \AA}$ ) and we find excellent agreement with distances measured experimentally by quantitative FRET. The good agreement between FRET, AXSI, and modeled structures strongly suggest that the MalE constructs used in this work are not significantly perturbed or denatured by our labeling approach, consistent with the observation that the binding affinity of labeled MalE agrees closely, within experimental error, with measurements on wildtype MalE<sup>24,50,52,53,76</sup> (Fig. S18†). In the future, improved labeling and purification procedures and more sophisticated modeling approaches, *e.g.* based on molecular dynamics simulations, should enable improved estimates of the full distance distributions and allow for full quantitative comparisons to molecular models. A range

of different labeling and functionalization approaches is also desirable to select functionalization approaches that minimally perturb the protein of interest.<sup>77–80</sup> Simulations for distributions with the same mean but different widths suggest that even considerably broader ensembles should be resolvable using AXSI (Fig. S19†). In addition, (A)XSI has the potential to discern different sub-populations in the  $P(d)$  distribution, as has been demonstrated for a Holliday junction using XSI.<sup>81</sup> Simulation with different size particles suggest that in principle smaller gold nanoparticles could be used, however the signal-to-noise deteriorates with decreasing gold nanoparticles (Fig. S20†).

## Conclusions

The introduction of AXSI for proteins opens up exciting possibilities in structural biology and beyond. The ability to determine intramolecular distance distributions for proteins in free solution and under virtually arbitrary solution conditions has the potential to address questions regarding partially folded conformations and natively disordered proteins.<sup>46,82–84</sup> In principle, our method could be used on dehydrated or frozen samples<sup>85</sup> and our labeling approach might be adapted for marking structures in cryo-EM measurements.<sup>86,87</sup> Beyond intramolecular distance measurements, AXSI might be extended to studying protein–protein interactions, *e.g.* by monitoring inter-molecular label–label distances in complexes or even assemblies of single-labeled proteins. In addition, our approach might be used to probe ligand binding kinetics and signaling pathways. Finally, the integration of computational modeling could establish a platform for refining structural predictions.

## Author contributions

T.C. and J.L. designed the study; A.N., G.G.M.M., and N.W. prepared protein samples; S.S., A.V.B., C.K., and A.L. prepared gold-labeled samples; S.S., A.V.B., C.K., and M.S. performed ASAXS measurements; A.N., G.G.M.M., and N.W. performed FRET measurements; S.S., A.V.B., and G.G.M.M. analyzed the data; T.C. and J.L. supervised research; S.S., T.C., and J.L. wrote the manuscript with input from all authors.

## Data availability

Data for this article, including ASAXS scattering profiles, gold label–gold label interference terms and distance distributions, are available in the public repository YODA at <https://doi.org/10.24416/UU01-CKHVKZ>.

## Conflicts of interest

The authors declare no competing financial interest.

## Acknowledgements

This work was supported by the Center for NanoScience (CeNS), through ERC Consolidator Grant “ProForce” (to J.L.), the Deutsche Forschungsgemeinschaft (SFB863 projects A11 to J.L. and A13 to T.C.; Sachbeihilfe CO879/4-1 to T.C.) and the BMBF (KMU innovative “quantum FRET” to T.C.). We thank Sebastian Doniach, Pehr Harbury, Rebecca Mathew, and Theyencheri Narayanan for useful discussions, Thomas Zettl for initial work on protein labeling, Thomas Nicolaus for laboratory assistance, and Savannah J. Turner for help with high-resolution TEM imaging. We acknowledge the ESRF for the provision of synchrotron radiation facilities.

## References

- 1 X.-C. Bai, G. McMullan and S. H. W. Scheres, *Trends Biochem. Sci.*, 2015, **40**, 49–57.
- 2 R. M. Glaeser, *Nat. Methods*, 2016, **13**, 28–32.
- 3 J. S. Richardson and D. C. Richardson, *Biophys. J.*, 2014, **106**, 510–525.
- 4 C. D. Putnam, M. Hammel, G. L. Hura and J. A. Tainer, *Q. Rev. Biophys.*, 2007, **40**, 191–285.
- 5 H. W. Wang and J. W. Wang, *Protein Sci.*, 2017, **26**, 32–39.
- 6 Y. Shi, *Cell*, 2014, **159**, 995–1014.
- 7 K. Wüthrich, *Europophys. News*, 1986, **17**, 11–13.
- 8 A. Cavalli, X. Salvatella, C. M. Dobson and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9615–9620.
- 9 G. Jeschke, *ChemPhysChem*, 2002, **3**, 927–932.
- 10 R. Ward, A. Bowman, E. Sozudogru, H. El-Mkami, T. Owen-Hughes and D. G. Norman, *J. Magn. Reson.*, 2010, **207**, 164–167.
- 11 M. F. Peter, C. Gebhardt, R. Mächtel, G. G. Moya Muñoz, J. Glaenzer, A. Narducci, G. H. Thomas, T. Cordes and G. Hagelueken, *Nat. Commun.*, 2022, **13**, 4396.
- 12 U. B. Choi, P. Strop, M. Vrljic, S. Chu, A. T. Brunger and K. R. Weninger, *Nat. Struct. Mol. Biol.*, 2010, **17**, 318–324.
- 13 J. Nagy, D. Grohmann, A. C. M. Cheung, S. Schulz, K. Smollett, F. Werner and J. Michaelis, *Nat. Commun.*, 2015, **6**, 6161.
- 14 S. Kilic, S. Felekyan, O. Doroshenko, I. Boichenko, M. Dimura, H. Vardanyan, L. C. Bryan, G. Arya, C. A. M. Seidel and B. Fierz, *Nat. Commun.*, 2018, **9**, 235.
- 15 X. A. Feng, M. F. Poyton and T. Ha, *Curr. Opin. Struct. Biol.*, 2021, **70**, 26–33.
- 16 A. T. Brunger, P. Strop, M. Vrljic, S. Chu and K. R. Weninger, *J. Struct. Biol.*, 2011, **173**, 497–505.
- 17 A. Muschiolok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer and J. Michaelis, *Nat. Methods*, 2008, **5**, 965–971.
- 18 M. Dimura, T.-O. Peulen, H. Sanabria, D. Rodnин, K. Hemmen, C. A. Hanke, C. A. M. Seidel and H. Gohlke, *Nat. Commun.*, 2020, **11**, 5394.
- 19 S. Kalinin, T. Peulen, S. Sindbert, P. J. Rothwell, S. Berger, T. Restle, R. S. Goody, H. Gohlke and C. A. M. Seidel, *Nat. Methods*, 2012, **9**, 1218–1225.

20 E. Lerner, A. Barth, J. Hendrix, B. Ambrose, V. Birkedal, S. C. Blanchard, R. Borner, H. Sung Chung, T. Cordes, T. D. Craggs, A. A. Deniz, J. Diao, J. Fei, R. L. Gonzalez, I. V. Gopich, T. Ha, C. A. Hanke, G. Haran, N. S. Hatzakis, S. Hohng, S. C. Hong, T. Hugel, A. Ingargiola, C. Joo, A. N. Kapanidis, H. D. Kim, T. Laurence, N. K. Lee, T. H. Lee, E. A. Lemke, E. Margeat, J. Michaelis, X. Michalet, S. Myong, D. Nettels, T. O. Peulen, E. Ploetz, Y. Razvag, N. C. Robb, B. Schuler, H. Soleimaninejad, C. Tang, R. Vafabakhsh, D. C. Lamb, C. A. Seidel and S. Weiss, *eLife*, 2021, **10**, e60416.

21 M. Dimura, T. O. Peulen, C. A. Hanke, A. Prakash, H. Gohlke and C. A. Seidel, *Curr. Opin. Struct. Biol.*, 2016, **40**, 163–185.

22 E. Lerner, T. Cordes, A. Ingargiola, Y. Alhadid, S. Chung, X. Michalet and S. Weiss, *Science*, 2018, **359**, eaan1133.

23 E. D. Holmstrom, A. Holla, W. Zheng, D. Nettels, R. B. Best and B. Schuler, *Methods Enzymol.*, 2018, **611**, 287–325.

24 G. Agam, C. Gebhardt, M. Popara, R. Mächtel, J. Folz, B. Ambrose, N. Chamachi, S. Y. Chung, T. D. Craggs, M. de Boer, D. Grohmann, T. Ha, A. Hartmann, J. Hendrix, V. Hirschfeld, C. G. Hübner, T. Hugel, D. Kammerer, H.-S. Kang, A. N. Kapanidis, G. Krainer, K. Kramm, E. A. Lemke, E. Lerner, E. Margeat, K. Martens, J. Michaelis, J. Mitra, G. G. Moya Muñoz, R. B. Quast, N. C. Robb, M. Sattler, M. Schlierf, J. Schneider, T. Schröder, A. Sefer, P. S. Tan, J. Thurn, P. Tinnefeld, J. van Noort, S. Weiss, N. Wendler, N. Zijlstra, A. Barth, C. A. M. Seidel, D. C. Lamb and T. Cordes, *Nat. Methods*, 2023, **20**, 523–535.

25 B. Hellenkamp, P. Wortmann, F. Kandzia, M. Zacharias and T. Hugel, *Nat. Methods*, 2017, **14**, 174–180.

26 H. Mazal and G. Haran, *Curr. Opin. Biomed. Eng.*, 2019, **12**, 8–17.

27 L. A. Metskas and E. Rhoades, *Annu. Rev. Phys. Chem.*, 2020, **71**, 391–414.

28 O. Glatter, *J. Appl. Crystallogr.*, 1977, **10**, 415–421.

29 M. H. Koch, P. Vachette and D. I. Svergun, *Q. Rev. Biophys.*, 2003, **36**, 147–227.

30 J. Lipfert and S. Doniach, *Annu. Rev. Biophys. Biomol. Struct.*, 2007, **36**, 307–327.

31 H. B. Stuhrmann, *Adv. Polym. Sci.*, 1985, **67**, 123–163.

32 H. B. Stuhrmann and H. Notbohm, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 6216–6220.

33 R. C. Miake-Lye, S. Doniach and K. O. Hodgson, *Biophys. J.*, 1983, **41**, 287–292.

34 R. S. Mathew-Fenn, R. Das and P. A. Harbury, *Science*, 2008, **322**, 446–449.

35 R. S. Mathew-Fenn, R. Das, J. A. Silverman, P. A. Walker and P. A. Harbury, *PLoS One*, 2008, **3**, e3229.

36 X. Shi, K. A. Beauchamp, P. B. Harbury and D. Herschlag, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E1473–E1480.

37 X. Shi, D. Herschlag and P. A. Harbury, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, E1444–E1451.

38 G. L. Hura, C. L. Tsai, S. A. Claridge, M. L. Mendillo, J. M. Smith, G. J. Williams, A. J. Mastroianni, A. P. Alivisatos, C. D. Putnam, R. D. Kolodner and J. A. Tainer, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 17308–17313.

39 X. Shi, L. Huang, D. M. Lilley, P. B. Harbury and D. Herschlag, *Nat. Chem. Biol.*, 2016, **12**, 146–152.

40 T. Zettl, R. Das, P. A. B. Harbury, D. Herschlag, J. Lipfert, R. S. Mathew and X. Shi, *Curr. Protoc. Nucleic Acid Chem.*, 2018, **73**, e54.

41 H. B. Stuhrmann, *Q. Rev. Biophys.*, 1981, **14**, 433–460.

42 V. J. Pinfield and D. J. Scott, *PLoS One*, 2014, **9**, e95664.

43 T. Zettl, R. S. Mathew, S. Seifert, S. Doniach, P. A. B. Harbury and J. Lipfert, *Nano Lett.*, 2016, **16**, 5353–5357.

44 Z. Yang, G. Jiménez-Osés, C. J. López, M. D. Bridges, K. N. Houk and W. L. Hubbell, *J. Am. Chem. Soc.*, 2014, **136**, 15356–15365.

45 G. Krainer, A. Hartmann and M. Schlierf, *Nano Lett.*, 2015, **15**, 5826–5829.

46 H. M. Watkins, A. J. Simon, T. R. Sosnick, E. A. Lipman, R. P. Hjelm and K. W. Plaxco, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 6631–6636.

47 H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb and L. Pollack, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 799–804.

48 A. Y. Sim, J. Lipfert, D. Herschlag and S. Doniach, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2012, **86**, 021901.

49 K. Hübner, M. Raab, J. Bohlen, J. Bauer and P. Tinnefeld, *Nanoscale*, 2022, **14**, 7898–7905.

50 F. A. Quiocho, J. C. Spurlino and L. E. Rodseth, *Structure*, 1997, **5**, 997–1015.

51 P. G. Telmer and B. H. Shilton, *J. Biol. Chem.*, 2003, **278**, 34555–34567.

52 M. de Boer, G. Gouridis, R. Vietrov, S. L. Begg, G. K. Schuurman-Wolters, F. Husada, N. Eleftheriadis, B. Poolman, C. A. McDevitt and T. Cordes, *eLife*, 2019, **8**, e44652.

53 R. Mächtel, A. Narducci, D. A. Griffith, T. Cordes and C. Orelle, *Res. Microbiol.*, 2019, **170**, 321–337.

54 X. Duan, J. A. Hall, H. Nikaido and F. A. Quiocho, *J. Mol. Biol.*, 2001, **306**, 1115–1126.

55 B. H. Shilton, M. M. Flocco, M. Nilsson and S. L. Mowbray, *J. Mol. Biol.*, 1996, **264**, 350–363.

56 A. J. Sharff, L. E. Rodseth, J. C. Spurlino and F. A. Quiocho, *Biochemistry*, 1992, **31**, 10657–10663.

57 D. Svergun, C. Barberato and M. H. J. Koch, *J. Appl. Crystallogr.*, 1995, **28**, 768–773.

58 C. Gebhardt, P. Bawidamann, K. Schütze, G. G. Moya Muñoz, A.-K. Spring, D. A. Griffith, J. Lipfert and T. Cordes, *bioRxiv*, 2023, preprint, DOI: [10.1101/2023.06.12.544586](https://doi.org/10.1101/2023.06.12.544586).

59 M. Brust, M. Walker, D. Bethell, D. J. Schiffrin and R. Whyman, *J. Chem. Soc., Chem. Commun.*, 1994, 801–802.

60 G. Gouridis, G. K. Schuurman-Wolters, E. Ploetz, F. Husada, R. Vietrov, M. De Boer, T. Cordes and B. Poolman, *Nat. Struct. Mol. Biol.*, 2015, **22**, 57–64.



61 M. Sztucki, E. Di Cola and T. Narayanan, *J. Appl. Crystallogr.*, 2010, **43**, 1479–1487.

62 M. Sztucki, E. Di Cola and T. Narayanan, *J. Phys.: Conf. Ser.*, 2011, **272**, 012004.

63 M. Sztucki, E. Di Cola and T. Narayanan, *Eur. Phys. J.: Spec. Top.*, 2012, **208**, 319–331.

64 T. Narayanan, M. Sztucki, T. Zinn, J. Kieffer, A. Homs-Puron, J. Gorini, P. Van Vaerenbergh and P. Boesecke, *J. Appl. Crystallogr.*, 2022, **55**, 98–111.

65 S. F. Gull and J. Skilling, *Quantified maximum entropy MemSys5 users' manual*, Maximum Entropy Data Consultants Ltd., Suffolk, 1999, pp. 1–108.

66 J. Hohlbein, T. D. Craggs and T. Cordes, *Chem. Soc. Rev.*, 2014, **43**, 1156–1171.

67 A. N. Kapanidis, N. K. Lee, T. A. Laurence, S. Doose, E. Margeat and S. Weiss, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 8936–8941.

68 N. K. Lee, A. N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright and S. Weiss, *Biophys. J.*, 2005, **88**, 2939–2953.

69 B. Ambrose, J. M. Baxter, J. Cully, M. Willmott, E. M. Steele, B. C. Bateman, M. L. Martin-Fernandez, A. Cadby, J. Shewring, M. Aaldering and T. D. Craggs, *Nat. Commun.*, 2020, **11**, 5641.

70 S. Wolf, B. Sohmen, B. Hellenkamp, J. Thurn, G. Stock and T. Hugel, *Chem. Sci.*, 2021, **12**, 3350–3359.

71 T. Eilert, E. Kallis, J. Nagy, C. Röcker and J. Michaelis, *J. Phys. Chem. B*, 2018, **122**, 11677–11694.

72 B. Hellenkamp, S. Schmid, O. Doroshenko, O. Opanasyuk, R. Kühnemuth, S. Rezaei Adariani, B. Ambrose, M. Aznauryan, A. Barth, V. Birkedal, M. E. Bowen, H. Chen, T. Cordes, T. Eilert, C. Fijen, C. Gebhardt, M. Götz, G. Gouridis, E. Gratton, T. Ha, P. Hao, C. A. Hanke, A. Hartmann, J. Hendrix, L. L. Hildebrandt, V. Hirschfeld, J. Hohlbein, B. Hua, C. G. Hübner, E. Kallis, A. N. Kapanidis, J.-Y. Kim, G. Krainer, D. C. Lamb, N. K. Lee, E. A. Lemke, B. Levesque, M. Levitus, J. J. McCann, N. Naredi-Rainer, D. Nettels, T. Ngo, R. Qiu, N. C. Robb, C. Röcker, H. Sanabria, M. Schlierf, T. Schröder, B. Schuler, H. Seidel, L. Streit, J. Thurn, P. Tinnefeld, S. Tyagi, N. Vandenberk, A. M. Vera, K. R. Weninger, B. Wünsch, I. S. Yanez-Orozco, J. Michaelis, C. A. M. Seidel, T. D. Craggs and T. Hugel, *Nat. Methods*, 2018, **15**, 669–676.

73 C. Gebhardt, M. Lehmann, M. M. Reif, M. Zacharias, G. Gemmecker and T. Cordes, *ChemPhysChem*, 2021, **22**, 1566–1583.

74 F. D. Steffen, R. K. O. Sigel and R. Borner, *Bioinformatics*, 2021, **37**, 3953–3955.

75 M. M. Tirion, *Phys. Rev. Lett.*, 1996, **77**, 1905–1908.

76 T. Ferenci, M. Muir, K.-S. Lee and D. Maris, *Biochim. Biophys. Acta, Biomembr.*, 1986, **860**, 44–50.

77 C.-A. J. Lin, C.-H. Lee, J.-T. Hsieh, H.-H. Wang, J. K. Li, J.-L. Shen, W.-H. Chan, H.-I. Yeh and W. H. Chang, *J. Med. Biol. Eng.*, 2009, **29**, 276–283.

78 M.-E. Aubin-Tam and K. Hamad-Schifferli, *Biomed. Mater.*, 2008, **3**, 034001.

79 M.-E. Aubin-Tam and K. Hamad-Schifferli, *Langmuir*, 2005, **21**, 12080–12084.

80 M.-E. Aubin-Tam, W. Hwang and K. Hamad-Schifferli, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4095–4100.

81 T. Zettl, X. Shi, S. Bonilla, S. M. Sedlak, J. Lipfert and D. Herschlag, *Nucleic Acids Res.*, 2020, **48**, 8090–8098.

82 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 and T. R. Sosnick, *Science*, 2017, **358**, 238–241.

83 R. B. Best, W. Zheng, A. Borgia, K. Buholzer, M. B. Borgia, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev and B. Schuler, *Science*, 2018, **361**, eaar7101.

84 A. H. Larsen, Y. Wang, S. Bottaro, S. Grudinin, L. Arleth and K. Lindorff-Larsen, *PLoS Comput. Biol.*, 2020, **16**, e1007870.

85 S. P. Meisburger, M. Warkentin, H. Chen, J. B. Hopkins, R. E. Gillilan, L. Pollack and R. E. Thorne, *Biophys. J.*, 2013, **104**, 227–236.

86 C. T. Beales and O. Medalia, *J. Struct. Biol.*, 2022, **214**, 107880.

87 N. Groysbeck, V. Hanss, M. Donzeau, J. M. Strub, S. Cianférani, D. Spehner, M. Bahri, O. Ersen, M. Eltsov and P. Schultz, *Small Methods*, 2023, **7**, 2300098.

