Conformational and oligomeric states of SPOP from small-angle X-ray scattering and molecular dynamics simulations

F. Emil Thomasen[1], **Matthew J. Cuneo**[2], **Tanja Mittag**[2], **Kresten Lindorff-Larsen**[1*]

***For correspondence:**
lindorff@bio.ku.dk (KLL)

[1]Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Copenhagen, Denmark; [2]Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

Abstract

Speckle-type POZ protein (SPOP) is a substrate adaptor in the ubiquitin proteasome system, and plays important roles in cell-cycle control, development, and cancer. SPOP forms linear higher-order oligomers following an isodesmic self-association model. Oligomerization is essential for SPOP's multivalent interactions with substrates, which facilitate phase separation and localization to biomolecular condensates. Structural characterization of SPOP in its oligomeric state and in solution is, however, challenging due to the inherent conformational and compositional heterogeneity of the oligomeric species. Here, we develop an approach to simultaneously and self-consistently characterize the conformational ensemble and the distribution of oligomeric states of SPOP by combining small-angle X-ray scattering (SAXS) and molecular dynamics simulations. We build initial conformational ensembles of SPOP oligomers using coarse-grained molecular dynamics simulations, and use a Bayesian/maximum entropy approach to refine the ensembles, along with the distribution of oligomeric states, against a concentration series of SAXS experiments. Our results suggest that SPOP oligomers behave as rigid, helical structures in solution, and that a flexible linker region allows SPOP's substrate binding domains to extend away from the core of the oligomers. Additionally, our results are in good agreement with previous characterization of the isodesmic self-association of SPOP. In the future, the approach presented here can be extended to other systems to simultaneously characterize structural heterogeneity and self-assembly.

Introduction

Protein self-association is fundamental for many processes in biology (***Ali and Imperiali, 2005***; ***Marsh and Teichmann, 2015***), and it has been estimated that around half of all proteins form dimers or higher-order complexes (***Lynch, 2012***). One such protein is Speckle-type POZ protein (SPOP), a substrate adaptor in the ubiquitin proteasome system, which recruits substrates for the Cullin3-RING ubiquitin ligase (CRL3) (***Hernández-Muñoz et al., 2005***; ***Kent et al., 2006***; ***Kwon et al., 2006***). SPOP targets a range of substrates for degradation, including proteins involved in hormonal signalling, epigenetic modification, and cell-cycle control, such as the androgen receptor (AR) (***An et al., 2014***) and death-associated protein 6 (DAXX) (***Kwon et al., 2006***; ***Cuneo and Mittag, 2019***). SPOP is thus an important regulator of cellular signalling, and mutations in SPOP are associated

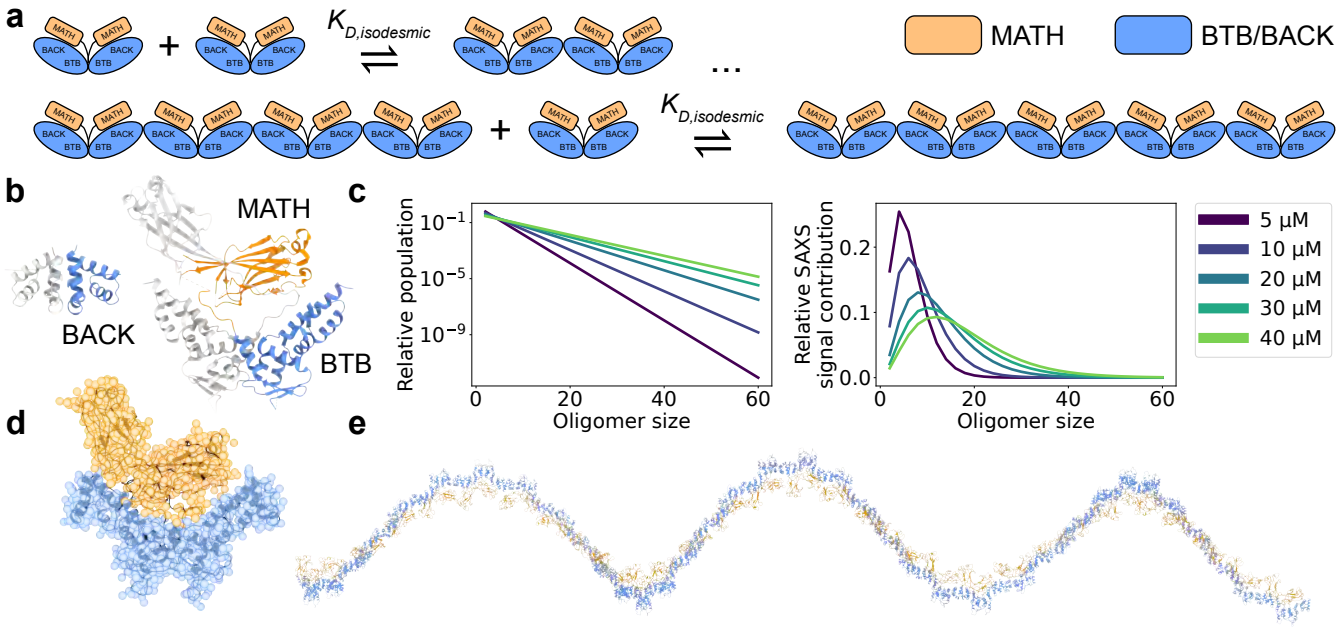

Figure 1. SPOP forms higher-order oligomers through isodesmic self-association. **a.** The SPOP BTB-BTB homodimer forms with nanomolar affinity, and is the unit of higher-order oligomerization through BACK-BACK homodimerization. Higher-order SPOP oligomerization follows an isodesmic model, where the equilibrium between oligomer i and i+1 is described by a single equilibrium constant, $K_{D,isodesmic}$, which is independent of oligomer size. **b.** Crystal structures of homodimers of the BACK (left, PDB: 4HS2) and MATH-BTB (right, PBD: 3HQI) domains of SPOP. **c.** Left: Populations of SPOP oligomers given by the isodesmic model with the previously determined $K_{D,isodesmic}$=2.4 μM for the protein concentrations used in our SAXS experiments. Note the logarithmic scale. Right: Relative contribution of each oligomer to the average SAXS signal given by the populations in left panel. **d.** Structure of a SPOP$^{28–359}$ dimer constructed based on crystal structures in panel b. The all-atom cartoon model is overlaid with the coarse-grained representation used in Martini simulations. **e.** Structure of a SPOP$^{28–359}$ 60-mer constructed based on BACK-BACK interface in panel b. MATH domains are coloured orange and BTB/BACK domains are coloured blue in all panels.

with a variety of cancers (***Le Gallo et al., 2012***; ***Kim et al., 2013***; ***Cuneo and Mittag, 2019***).

The 374-residue SPOP monomer consists of three domains. From N- to C-terminus, these are the MATH domain (i.e., the meprin and TRAF-C homology domain), the BTB domain (i.e,. the broad-complex, tramtrack, and bric-a-brac domain), and the BACK domain (i.e., the BTB and C-terminal Kelch domain). MATH is the substrate binding domain, while the BTB domain mediates interaction with CRL3 (***Zhuang et al., 2009***; ***Bosu and Kipreos, 2008***). Both the BTB and BACK domains can homodimerize, resulting in the formation of polydisperse, linear higher-order SPOP oligomers with alternating BTB-BTB and BACK-BACK interfaces (***Errington et al., 2012***; ***Van Geersdaele et al., 2013***; ***Marzahn et al., 2016***). The BTB-mediated dimer is formed with nanomolar affinity, and this dimer thus acts as the unit of higher-order oligomerization, which occurs through micromolar affinity BACK dimerization. Thus, only even-numbered SPOP oligomers are substantially populated (***Marzahn et al., 2016***) (Fig. 1).

Chemical crosslinking experiments have shown that SPOP oligomers also form inside cells (***Marzahn et al., 2016***), and analysis of SPOP homologues shows sequence co-variation across both the BTB-BTB and BACK-BACK interfaces (***Bouchard et al., 2018***), together suggesting that self-association has physiological relevance. By presenting multiple MATH domains for substrate binding, SPOP oligomers can simultaneously bind to multiple low-affinity binding motifs in a single substrate, resulting in an overall increased affinity through avidity effects (***Pierce et al., 2016***). The longer lifetimes of these complexes enable effective polyubiquitination (***Pierce et al., 2016***). This suggests that tuning SPOP's oligomerization state could act as a mechanism to regulate substrate degradation (***Errington et al., 2012***). SPOP oligomerization is also involved in phase separation. SPOP localizes to nuclear speckles in cells (***Marzahn et al., 2016***), and upon overexpression of certain

substrates, SPOP and substrate co-localize to droplets which recruit CRL3 and display active substrate ubiquitination (***Bouchard et al., 2018***). This process requires both SPOP oligomerization and substrate binding (***Marzahn et al., 2016***; ***Bouchard et al., 2018***), and it has been proposed that SPOP oligomers function as scaffolds that enable binding of substrates both within and between oligomers, resulting in filament-formation at low substrate concentrations and condensate formation at higher substrate concentrations (***Bouchard et al., 2018***; ***Schmit et al., 2019***).

The higher-order self-association of SPOP follows the isodesmic model (***Marzahn et al., 2016***), in which the equilibrium between oligomer i and i+1 is described by a single equilibrium constant independently of oligomer size (***Oosawa and Kasai, 1962***). In the case of SPOP, the BTB-mediated dimer acts as the protomer of higher-order self-association, and the isodesmic K_D thus describes BACK-BACK self-association. The isodesmic model can be used to calculate the equilibrium concentration of every oligomeric species as a function of the total protomer concentration (Fig. 1a,c). For SPOP, an isodesmic dissociation constant K_D=2.4±0.4 μM has been determined from composition gradient multi-angle light scattering experiments (CG-MALS) (***Marzahn et al., 2016***). While these insights describe the heterogeneity in oligomer sizes, the conformational heterogeneity of the higher-order oligomers have not been characterized. Previous work revealed that constitutive SPOP dimers, created via deletion of the BACK domain, have considerable conformational heterogeneity in the position of their MATH domains. These are seen docked onto the BTB dimer in the structure, but small-angle X-ray scattering (SAXS) experiments showed that they could undock from the BTB domains, enabled by a long flexible linker (***Zhuang et al., 2009***). This may enable different spacing between SPOP-binding motifs in multivalent substrates. Whether this conformational flexibility also exists in higher-order SPOP oligomers is unclear.

Here, we aimed to determine simultaneously both the distribution of oligomeric states of SPOP and the conformational ensemble of each SPOP oligomer by combining small-angle X-ray scattering (SAXS) experiments and molecular dynamics (MD) simulations. SAXS can provide low resolution information on protein structure in solution, but reports on an ensemble average, which in the case of SPOP is both an average over different oligomeric states and the structural heterogeneity of each oligomeric state. Therefore, SAXS experiments are often combined with MD simulations to provide a full structural model of the system (***Thomasen and Lindorff-Larsen, 2022***). In the case of polydisperse systems, it is sometimes possible to deconvolute the information into contributions from a small number of individual species and analyse these individually (***Herranz-Trillo et al., 2017***; ***Meisburger et al., 2021***). We took a different approach and aimed to model explicitly every relevant configuration of SPOP in its range of oligomeric states along with the associated thermodynamic weight of each configuration. We collected SAXS data on SPOP at a range of protein concentrations and constructed initial conformational ensembles of every substantially populated oligomeric state using coarse-grained MD simulations. We then developed an approach to simultaneously and self-consistently optimize the distribution of oligomeric states, given by the isodesmic model (***Oosawa and Kasai, 1962***; ***Shemesh et al., 2021***), and refine the conformational ensemble of each oligomer against the SAXS data using Bayesian/maximum entropy (BME) reweighting (***Bottaro et al., 2020***). Our results show that SPOP forms rigid, helical oligomers in solution, and that the linker connecting the MATH and BTB domains is likely flexible, allowing for repositioning of the MATH domains during substrate binding. Our results also provide further evidence that SPOP self-association follows the isodesmic model, and we find an isodesmic K_D in good agreement with the previously determined value (***Marzahn et al., 2016***). Using SAXS experiments of a cancer variant of SPOP we also show how our approach can be used to determine changes in the level of self-association.

Results

We collected a concentration series of SAXS data on a previously used truncated version of SPOP, $SPOP^{28–359}$ (full length is 374 residues), with total protein concentrations ranging from 5 to 40 μM. In order to build structural models to refine against the SAXS data, we first needed to decide which oligomeric species to include in our modelling. To this aim we used the isodesmic self-association

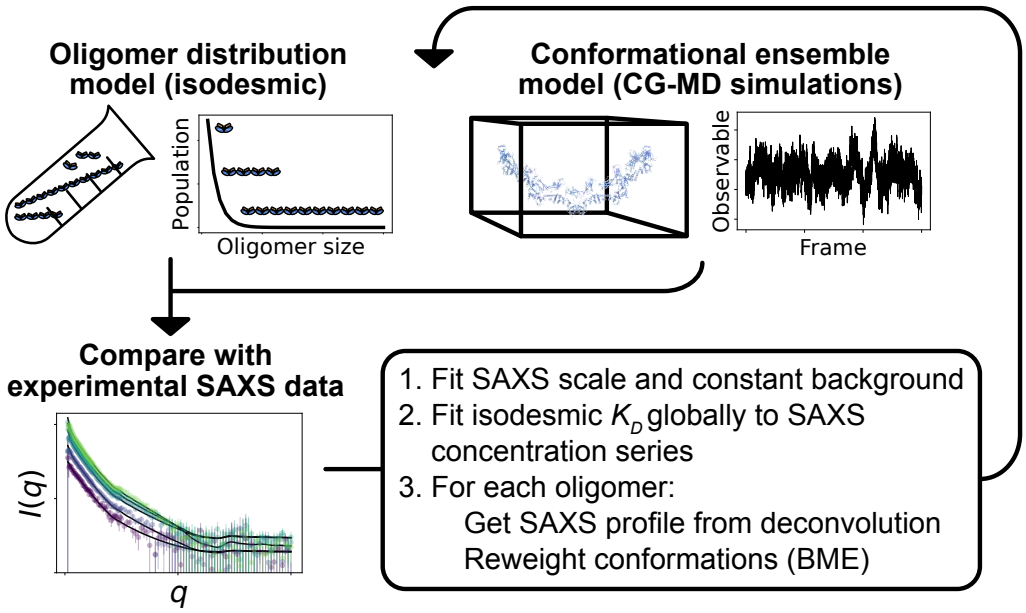

Figure 2. Overview of the self-consistent approach used to fit conformational ensembles of SPOP oligomers to SAXS data. Small-angle X-ray scattering (SAXS) data on SPOP represents an average over a range of oligomeric species present in solution. Here, the distribution of oligomeric species and the conformational ensemble of each oligomer were self-consistently fitted to a concentration series of SAXS data by iteratively fitting the scale and constant background of the SAXS data and the isodesmic K_D, followed by reweighting of the conformational ensemble of each oligomer.

model with K_D=2.4±0.4 μM, which has previously been shown to describe SPOP oligomerization well (***Marzahn et al., 2016***). Based on the isodesmic model, the population of oligomers larger than the ~30-mer should be very low at the concentration range used in our SAXS experiments. As scattering intensity is proportional to particle size squared, larger oligomers, however, make a considerable contribution to the SAXS signal despite their low concentrations (Fig. 1c). Given the concentrations from the isodesmic model and taking into account the increased scattering of larger oligomers, we decided that constructing models of oligomers up to the 60-mer should be sufficient to capture all substantial contributions to the SAXS data.

There are no crystal structures of SPOP$^{28–359}$ available, so we constructed a model of the SPOP$^{28–359}$ BTB-dimer using the crystal structure of the isolated BACK domain (4HS2) (***Van Geersdaele et al., 2013***) and the crystal structure of a truncated construct containing only the MATH and BTB domains (3HQI) (***Zhuang et al., 2009***) (Fig. 1b,d). We used this model of the BTB-dimer to construct SPOP$^{28–359}$ oligomers, which we used as starting structures for MD simulations. We ran 60 μs MD simulations of oligomers ranging from the dimer to the dodecamer; we used a coarse-grained representation of SPOP modelled using the Martini 3 force field (***Souza et al., 2021***) further modified by increasing protein-water interactions by 6% (***Thomasen et al., 2022***). It would be computationally prohibitive to run simulations of large oligomers up to the 60-mer. Instead, we relied on the assumption that the dodecamer behaves similarly to a segment of an arbitrarily long oligomer, and constructed conformational ensembles of oligomers up to the 60-mer by joining together conformers from the simulations of the dodecamer at the BACK-BACK interface (Fig. 1e).

We calculated SAXS intensities from our conformational ensembles and, given the relative population of each oligomer from the isodesmic model with previously determined K_D=2.4 μM, we calculated SAXS profiles averaged over all the oligomeric species. We found that the SAXS data calculated in this way from the ensembles generated by MD simulations convoluted with the isodesmic model were in good agreement with the experimental SAXS data, giving a reduced χ^2 to the con-

centration series of SAXS data ($\chi^2_{r,\text{global}}$) of 2.22 (Fig. 3). Despite the overall good agreement, the residuals revealed some systematic deviations to the experimental SAXS profiles. These deviations could potentially arise from inaccuracies in the distribution of oligomeric states given by the isodesmic model, from inaccuracies in the modelled conformational ensembles, or from both. As a first step, we wanted to see if we could eliminate the deviations by only tuning the distribution of oligomeric states. We globally optimized the K_D of the isodesmic model against the concentration series of SAXS data, which gave K_D=0.9±0.4 μM, in good agreement with K_D=2.4±0.4 μM determined previously, and resulted in a $\chi^2_{r,\text{global}}$ of 1.24 to the SAXS data (Fig. 3). However, this did still not fully eliminate the systematic deviations to the experimental SAXS profiles.

To improve the agreement with the experimental SAXS data further, we aimed to refine simultaneously the conformational ensemble of each oligomer and optimize the distribution of oligomeric states. We developed a self-consistent optimization scheme, in which the isodesmic K_D is optimized globally to the entire concentration series of SAXS data followed by reweighting of the conformations of each oligomer against a SAXS profile deconvoluted from the experimental SAXS data (Fig. 2) (see Methods section for details). To reweight the ensembles, we used BME reweighting, in which the population weights of the conformational ensemble are minimally perturbed with respect to the prior ensemble (generated by the MD simulations) to improve the agreement with a set of experimental data (***Bottaro et al., 2020***). This approach resulted in excellent agreement with the experimental SAXS data, giving a $\chi^2_{r,\text{global}}$ of 0.69, while only small deviations remained (Fig. 3). The isodesmic K_D was fitted to 1.3±0.5 μM, and thus also remained in good agreement with the previously determined value (***Marzahn et al., 2016***). To validate our approach and to examine the possibility of overfitting, we left out one SAXS profile recorded with 15 μM protein from the optimization. The optimized K_D and ensemble weights also improved the fit to this SAXS profile, suggesting that we had avoided overfitting (Fig. S1). These results show that SAXS data on SPOP can be explained well by conformational ensembles of linear oligomers with populations given by the isodesmic model, and thus provide further evidence that SPOP self-association follows a simple isodesmic mechanism (***Marzahn et al., 2016***).

Having generated a conformational ensemble of each SPOP oligomer in agreement with the SAXS data, we proceeded to analyze the structures. Reweighting resulted in an increase in the radius of gyration (R_g) for almost all oligomeric species, suggesting that slightly more expanded conformations than those sampled with our modified version of Martini are more consistent with the SAXS data (Fig. 4a-b and S2). This expansion can be attributed both to a slight increase in the end-to-end distance for most oligomers (Fig. 4c-d and S3), as well as a slight increase in the average distance between the MATH and BTB/BACK domains for all oligomers upon reweighting (Fig. 4e-f).

In order to investigate the global flexibility and compaction of SPOP oligomers, we fitted a power law to the average end-to-end distance ($R_{\text{E-E}}$) as a function of the number of subunits in the oligomer (N), $R_{\text{E-E}} = R_0 N^{\nu}$, where R_0 is the subunit segment size and ν is a scaling exponent (Fig. 4c). The fit gave R_0=~3.1 nm and ν=0.99, suggesting a linear growth of the end-to-end distance with the number of subunits. This result is consistent with no significant curvature or compaction of the oligomers and, along with the narrow distribution of end-to-end distances for each oligomer (Fig. S3), suggests that the SAXS data is compatible with a distribution of straight and relatively rigid SPOP oligomers, at least on length scales up to the ~180 nm of the 60-mer. The helical structure of larger oligomers, with ~16 subunits per turn, is evident as small periodic deviations from the fit (Fig. 4c).

The MATH and BTB domains are connected through a ~20 residue long linker region (Fig. 1b). We hypothesized that this linker may be flexible, allowing for reconfiguration of the MATH domains with respect to the crystal structure (***Zhuang et al., 2009***). We calculated the distances between the center-of-mass (COM) of the MATH domain and the COM of the BTB/BACK domains for every subunit of every oligomer in the ensembles. The distribution of these MATH-BTB/BACK distances reveal two populations overlapping with the two crystal structure configurations (Fig. 4e,i and 5c), where the MATH domains are in close proximity to the BTB/BACK domains (***Zhuang et al., 2009***).

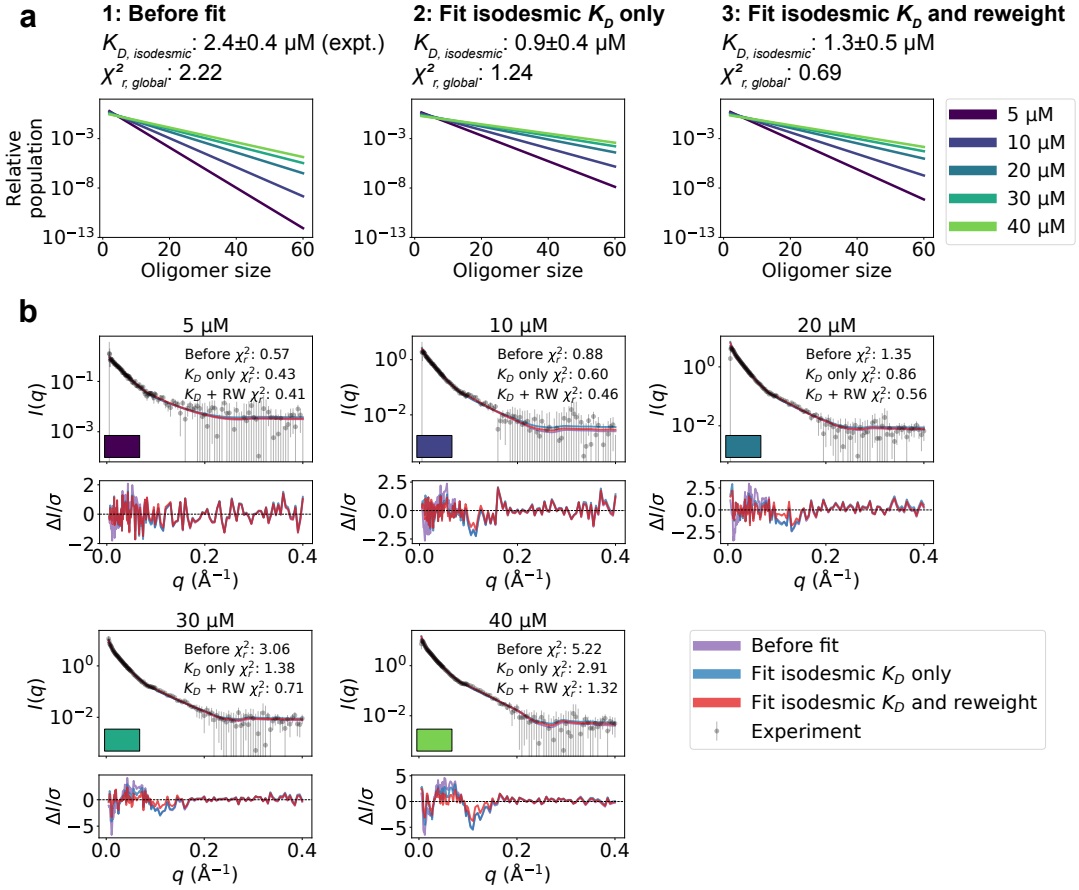

Figure 3. Refining oligomer populations and conformational ensembles against SAXS data. **a.** Relative populations of oligomers for the protein concentrations used in SAXS experiments. Note the logarithmic scale. Populations are given by the isodesmic model with the K_D noted above the plot, which is either (1) previously determined by CG-MALS or (2–3) fitted globally to the SAXS data in panel b. $\chi^2_{r,\text{global}}$ quantifies the agreement with SAXS data in panel b for the three scenarios. **b.** Agreement between experimental SAXS data and averaged SAXS data calculated from conformational ensembles of SPOP oligomers with populations given by the isodesmic model (as shown in panel a). SAXS profiles are shown for three different scenarios: (1) calculated from the conformational ensembles generated by MD simulations with the isodesmic K_D previously determined with CG-MALS, (2) calculated from the conformational ensembles generated by MD simulations with the isodesmic K_D fitted to the SAXS data, and (3) calculated from conformational ensembles refined against the SAXS data using Bayesian/MaxEnt reweighting, and with the isodesmic K_D self-consistently fitted to the SAXS data. Error-normalized residuals are shown below the SAXS profiles and χ^2_r to each SAXS profile is shown on the plot.

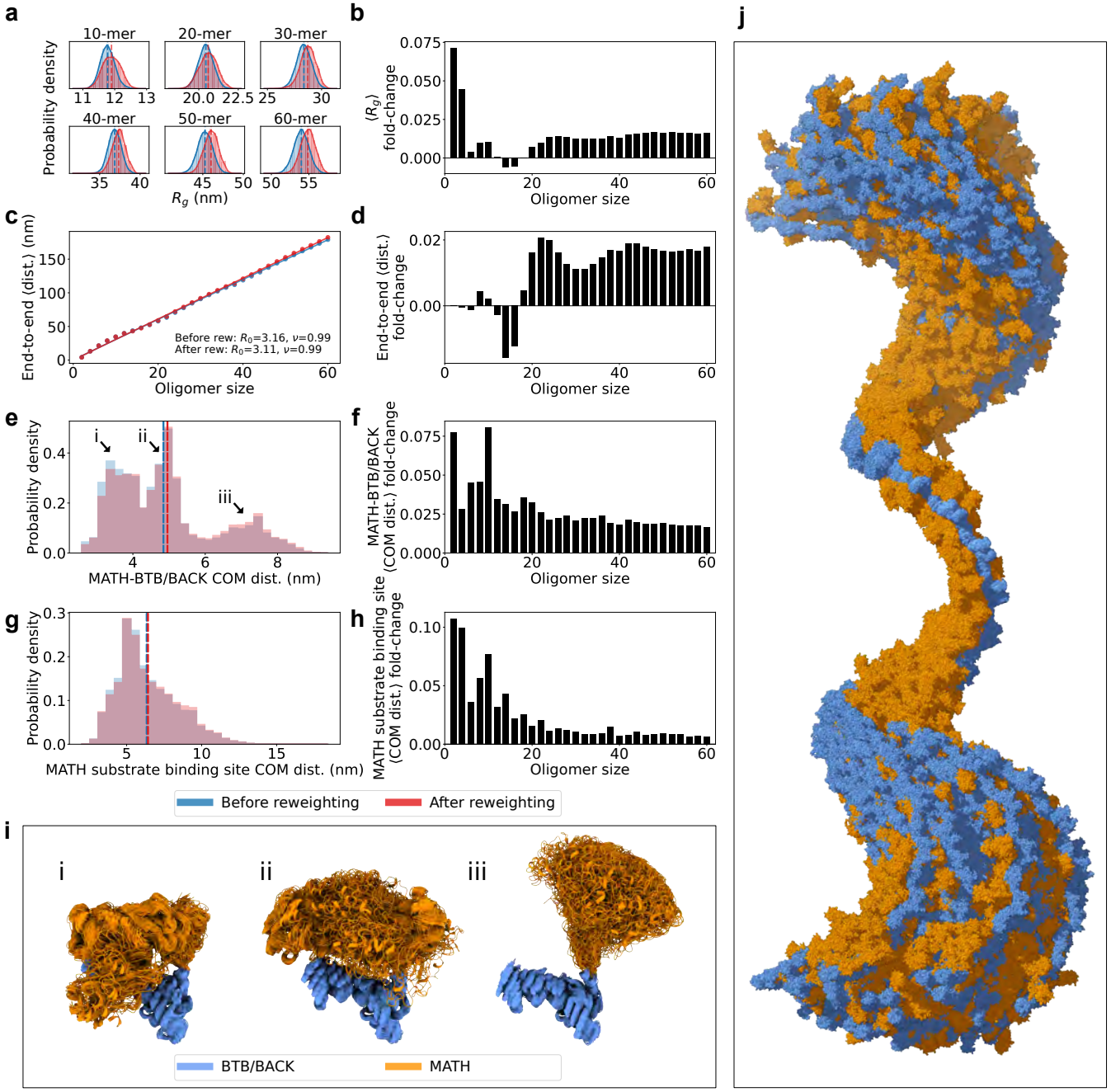

Figure 4. SPOP forms rigid, linear oligomers with flexible MATH domains in solution. **a.** Probability distribution of the radius of gyration (R_g) calculated from ensembles of six representative SPOP oligomers before and after reweighting (see Fig. S2 for R_g distributions for all oligomers). Dashed lines show the average values. **b.** The fold-change in average R_g after reweighting for all SPOP oligomers. **c.** The average end-to-end distance calculated from ensembles of SPOP oligomers before and after reweighting (see Fig. S3 for distributions for all oligomers). Solid lines show the fit of a power law: $R_{\text{E-E}} = R_0 N^{\nu}$, where $R_{\text{E-E}}$ is the average end-to-end distance, R_0 is the subunit segment size, N is the number of subunits in the oligomer, and ν is a scaling exponent. The fit gave R_0=3.16 nm, ν=0.99 before reweighting and R_0=3.11 nm, ν=0.99 after reweighting. **d.** The fold-change in average end-to-end distance after reweighting for all SPOP oligomers. **e.** Normalized histogram of distances between the center-of-mass (COM) of the MATH domain and the COM of the BTB/BACK domains in the same subunit before and after reweighting. The histogram contains the distances from every conformation of every subunit in every oligomer. **f.** The fold-change in average MATH-BTB/BACK COM distance after reweighting for all SPOP oligomers. **g.** Normalized histogram of COM distances between MATH substrate binding sites in neighbouring subunits. The histogram contains the distances from every conformation of every subunit in every oligomer. **h.** The fold-change in average COM distance between neighbouring MATH substrate binding sites after reweighting for all SPOP oligomers. **i.** Overlay of conformational ensembles corresponding to the three populations in panel e. The structures are from all non-terminal subunits of the SPOP dodecamer and are superposed on the BTB/BACK domains. **j.** Overlay of 151 randomly selected frames from the conformational ensemble of the SPOP 60-mer with atoms represented as spheres. Structures were superposed to the BTB/BACK domains in the four middle subunits. MATH domains are shown in orange and BTB/BACK domains are shown in blue.

However, there is also a third population in which the MATH domains are extended away from the BTB/BACK domains, suggesting that the MATH-BTB linker is flexible and allows for movement out of the configurations observed in the crystal structure. Reweighting slightly increased the population of this extended state (Fig. 4e-f). This flexibility in the configuration of the MATH domains gives rise to a broad distribution of distances between the substrate binding sites in neighbouring MATH domains, which is also slightly increased upon reweighting for all oligomers (Fig. 4g-h). Both the overall rigidity of the oligomers and the flexibility of the MATH domains are also evident from visual inspection of the conformational ensemble of the 60-mer (Fig. 4j).

To examine further whether the SAXS data support the observed flexibility of the MATH-BTB linker and repositioning of the MATH domains, we generated new ensembles of SPOP oligomers following the same protocol as above, but this time restraining the MATH domains to the BTB-BACK domains based on the configuration in the crystal structure using the elastic network model implemented in Martini. We calculated SAXS data from the generated ensembles and again fitted the isodesmic K_D globally to the SAXS data, resulting in K_D=0.2±0.2 μM. The agreement with the SAXS data was substantially worse than for the original ensembles with the MATH domains unrestrained ($\chi^2_{r,\text{global}}$=4.38 and $\chi^2_{r,\text{global}}$=1.24 respectively), and the systematic deviations from the experimental SAXS profiles were clearly exacerbated (Fig. 5). These results suggest that, first, the resolution of the SAXS data is high enough to distinguish between different configurations of the MATH domains and, second, that the SAXS data are indeed in better agreement with a model where the MATH-BTB linker is flexible. Taken together, our results support a model where, in solution, SPOP oligomers behave as rigid, helical structures with flexible MATH domains that can extend away from the BTB/BACK domains.

The comparison between the conformational ensembles generated with the MATH domains free or restrained also suggests that accurate determination of the isodesmic K_D relies on accurate conformational ensembles. Fitting the SAXS data using ensembles with the MATH domains restrained resulted in a lower isodesmic K_D, and calculating the agreement with SAXS for a range of isodesmic K_D values revealed that there was no clear minimum in the $\chi^2_{r,\text{global}}$ for K_D values >0, which is also reflected in the large error range for the fitted K_D (Fig. S5). In line with this observation, validation with SAXS data at 15 μM protein revealed that improving the accuracy of the ensembles by reweighting also improved the accuracy of the fitted isodesmic K_D independently of the fitted ensemble weights (Fig. S1d).

Having analyzed the conformational properties of wild type SPOP and shown that the SAXS data are sensitive to the degree of self-association, we next wished to test whether our approach could capture the effects of mutations on SPOP self-association. We collected a concentration series of SAXS data on the SPOP mutant R221C, which has been identified in melanoma (***Krauthammer et al., 2012***) and colorectal cancer (***Giannakis et al., 2016***). R221C is located in the BTB-BTB interface, so we hypothesized that it may affect SPOP's propensity to self-associate. We used the same approach as for wild type to fit the isodesmic K_D globally to the SAXS data, but without reweighting the conformational ensembles. For R221C, the isodesmic K_D was fitted to 8.2±2.3 μM, which resulted in a reasonable fit to the SAXS data with χ^2_{global}=1.79 (Fig. S6), suggesting that the mutation results in a decreased propensity to self-associate compared with wild type (K_D=0.9 μM using a comparable approach or 1.3 μM when also reweighting the ensemble). Because R221C is located at the BTB-BTB interface, the 6–9 fold increase of the isodesmic K_D (which relates to BACK-BACK dimerization) is perhaps surprising. While a long-range effect of R221C cannot be ruled out, an alternative mechanism may involve shifting the equilibrium of the BTB-BTB dimer, thus effectively decreasing the concentration of dimeric species available for self-association.

Discussion

The ability of SPOP, a cancer-associated substrate adaptor in the ubiquitination machinery, to self-associate is important for its role in biology and disease. Characterizing the conformational ensemble of flexible and self-associating proteins such as SPOP from ensemble-averaged experiments is,

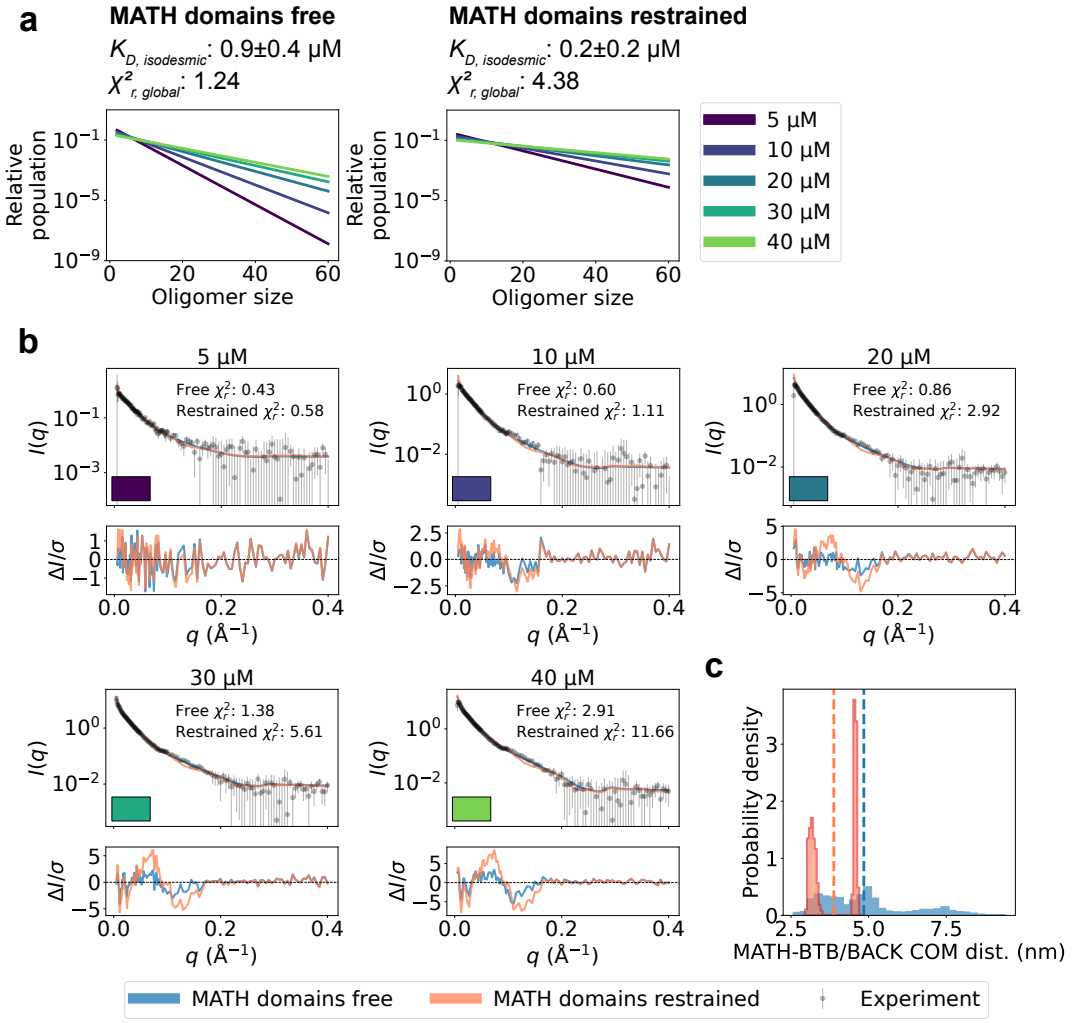

Figure 5. Unrestrained MATH domains give better agreement with SAXS data. Comparison of conformational ensembles with MATH domains either unrestrained (blue) or restrained to BTB/BACK domains based on the configuration in the crystal structure (orange). **a.** Relative populations of oligomers for the protein concentrations used in SAXS experiments. Note the logarithmic scale. Populations are given by the isodesmic model with the K_D noted above the plot. K_D was fitted globally to the SAXS data in panel b. $\chi^2_{r,global}$ quantifies the agreement with SAXS data in panel b for the two setups. **b.** Agreement between experimental SAXS data and averaged SAXS data calculated from conformational ensembles of SPOP oligomers generated with the two setups. Oligomer populations are given by the isodesmic model (as shown in panel a). Error-normalized residuals are shown below the SAXS profiles and χ^2_r to each SAXS profile is shown on the plot. **c.** Histogram of center-of-mass distances between MATH and BTB/BACK domains in the same subunit calculated from all conformations of all subunits of all oligomers. Average values are shown as dashed lines.

however, difficult due to conformational and compositional heterogeneity. In one approach, SAXS data of mixtures may be attempted to be decomposed into contributions of individual components that may then be analysed separately (***Herranz-Trillo et al., 2017***; ***Meisburger et al., 2021***). Here, we have developed an alternative 'forward modelling' approach to characterize proteins that undergo polydisperse oligomerization by self-consistently and globally fitting the distribution of oligomeric species and reweighting the conformational ensembles of the oligomers against SAXS data. A similar idea has recently been applied to study the self-association of tubulin using static structures as input (***Shemesh et al., 2021***). We recorded a concentration series of SAXS data on SPOP, which is known to form linear higher-order oligomers, and combined MD simulations with our approach to simultaneously refine conformational ensembles of thirty oligomeric states of SPOP along with their relative populations.

Our results suggest that SPOP oligomers are rigid, helical structures in solution and that the MATH-BTB linker is flexible, allowing for the extension of MATH domains away from the oligomer core. This is consistent with SPOP's proposed role in phase separation, as reconfiguration of the MATH domains could facilitate binding of substrates across multiple MATH domains and between different SPOP oligomers (***Pierce et al., 2016***; ***Bouchard et al., 2018***). Our results also provide orthogonal evidence that SPOP self-association is described well by the isodesmic model, and that the isodesmic K_D for BACK-BACK mediated self-association is in the low micromolar range, in agreement with previous measurements by CG-MALS (***Marzahn et al., 2016***). We also collected SAXS data and fitted the isodesmic K_D for the SPOP mutant R221C. Our results suggest that SPOP R221C has a 6–9 fold decreased propensity to self-associate.

The approach presented here to study SPOP can be extended to other polydisperse systems to characterize the distribution of oligomeric states and their conformational properties. However, there are a few limitations to be aware of; SAXS is a low resolution technique, and may not be able to distinguish between all relevant conformations, a problem that is likely exacerbated here, as the contribution of many species to the SAXS signal may average out distinct features in the profile. One way to mitigate this problem is to construct multiple structural models, and test whether they show any difference in the agreement with the SAXS data. In the case of SPOP we used this approach to examine the flexibility of the MATH domain in SPOP$^{28–359}$.

Another limitation of the approach is the correlation between the fitted distribution of oligomeric states and the conformational properties of the oligomers. Here, we observed that a low isodesmic K_D with large uncertainty was fitted when using more compact structures (MATH domains restrained) than when using expanded structures (MATH domains free), which suggests that the model can compensate for the underestimated dimensions of the proteins by increasing the populations of larger oligomers. Therefore, it is important to use prior conformational ensembles that are as accurate as possible. Additionally, it is important to include all the oligomeric species that make a substantial contribution to the SAXS data in the modelling. In the future it might be relevant to include independent data reporting on the distribution of oligomeric species, such as from CG-MALS, when fitting the SAXS data.

In the case of SPOP, we described the distribution of oligomers using the isodesmic self-association model, but this can be replaced by any model that describes the populations of the species in solution — with the caveat that there should not be too many free parameters to fit to the SAXS data. Similarly, the approach to generate prior conformational ensembles is not limited to MD simulations, and can be varied based on the system at hand. This flexibility in the modelling approach will make it useful to study other polydisperse systems in the future.

Methods

Protein expression and purification

The SPOP gene encoding residues 28–359 (His-SUMO-SPOP$^{28–359}$) was expressed and purified as previously described (***Bouchard et al., 2018***). Briefly, His-SUMO-SPOP$^{28–359}$ was transformed into

BL21-RIPL cells and expressed in auto-induction media (***Studier, 2005***). Cells were harvested, lysed, and cell debris was pelleted by centrifugation. The clarified supernatant was applied to a gravity Ni Sepharose resin equilibrated in resuspension buffer (30 mM imidazole, 1 M NaCl, pH 7.8). After washing with wash buffer (75 mM imidazole, 1 M NaCl, pH 7.8), the protein was eluted with a buffer containing 300 mM imidazole, 1 M NaCl, pH 7.8. One milligram of TEV protease was added to the eluted protein and the reaction was left to dialyze into pH 7.8, 300 mM NaCl and 5 mM DTT at 4 °C overnight. The cleaved protein was then further purified using a Superdex S200 size-exclusion chromatography column equilibrated with pH 7.8, 300 mM NaCl and 5 mM DTT.

Small-angle X-ray scattering

SAXS experiments were performed at the LIX-beamline (16-ID) of the National Synchrotron Light Source II (Upton, NY) (***DiFabio et al., 2016***). Data were collected at a wavelength of 1.0 Å, yielding an accessible scattering angle range of $0.006 < q < 3.2$ Å^{-1}, where q is the momentum transfer, defined as $q = 4\pi \sin(\theta)/\lambda$, where λ is the X-ray wavelength and 2θ is the scattering angle. Data with $q < 0.4$ Å^{-1} were used for all analyses. Prior to data collection, SPOP was dialyzed into 20 mM Tris pH 7.8, 150 mM NaCl and 5 mM DTT. Samples were loaded into a 1-mm capillary for ten 1-s X-ray exposures. Data were reduced at the beamline using the Python package *py4xs*.

Molecular dynamics simulations

We ran coarse grained molecular dynamics simulations of six SPOP$^{28–359}$ oligomers ranging from the dimer to dodecamer (in steps of dimeric protomer subunits) using a beta version (3.0.4.17) of the Martini 3 force field (https://github.com/KULL-Centre/papers/tree/main/2020/TIA1-SAS-Larsen-et-al/Martini) (***Souza et al., 2021***) and Gromacs 2020 (***Abraham et al., 2015***). We built the SPOP monomer structure using Modeller (***Šali and Blundell, 1993***) based on the crystal structure of the MATH and BTB domains (PDB: 3HQI) (***Zhuang et al., 2009***) and a crystal structure of the BACK domain (PDB: 4HS2) (***Van Geersdaele et al., 2013***). We built the dimer structure by superposing two monomer structures to the crystal structure of the BTB-BTB dimer interface in 3HQI. We then built larger oligomers by iteratively adding dimer structures to the linear oligomer. Dimers were added by superposing the terminal BACK domain of the oligomer and a terminal BACK domain of the dimer to the structure of the BACK-BACK dimer (4HS2).

The starting structures were coarse grained using the Martinize2 python script. Elastic network restraints of 500 kJ mol^{-1} nm^{-2} between backbone beads within a 1.2 nm cut-off were applied with Martinize2 to keep folded domains intact and to hold oligomer subunits together. In the 'MATH free' model, we removed all elastic network restraints between MATH and BTB/BACK domains, between MATH and MATH domains, and in the linker region between MATH and BTB/BACK domains, while in the 'MATH restrained' model, we only removed elastic network restraints between MATH and MATH domains and in the linker region between MATH and BTB/BACK domains, but kept restraints between MATH and BTB/BACK domains. We added dihedral and angle potentials between side chains and backbone beads with the *-scfix* flag in Martinize2. Using Gromacs *editconf*, we placed the dimer and tetramer in a dodecahedral box. To keep the box volume small, larger oligomers were aligned with the principal axis of the system and placed in triclinic boxes that were thus elongated along the x-axis. To keep these oligomers from rotating and self-associating across the periodic boundary, we added soft harmonic position restraints of 5 J mol^{-1} nm^{-2} along the y- and z-axis to the backbone beads of the terminal BTB/BACK domains. We solvated the systems using the Insane python script (***Wassenaar et al., 2015***) and added 150 mM NaCl along with Na^+ ions to neutralize the systems. In the 'MATH free' system, we rescaled the ϵ of the Lennard-Jones potentials between all protein and water beads by a factor 1.06 to favour extension of the MATH domains into solution (***Thomasen et al., 2022***), while the unmodified Martini 3 beta v.3.0.4.17 was used for the 'MATH restrained' model.

Energy minimization was performed using steepest descent for 10,000 steps with a 30 fs time-step. Simulations were run in the NPT ensemble at 300 K and 1 bar using the Velocity-Rescaling

thermostat (***Bussi et al., 2007***) and Parinello-Rahman barostat (***Parrinello and Rahman, 1981***). Non-bonded interactions were treated with the Verlet cut-off scheme. The cut-off for Van der Waals interactions and Coulomb interactions was set to 1.1 nm. A dielectric constant of 15 was used. We equilibrated the systems for 10 ns with a 2 fs time-step and ran production simulations for 60 μs with a 20 fs time-step, saving a frame every 1 ns.

After running the simulations, molecule breaks over the periodic boundaries were treated with Gromacs *trjconv* using the flags *-pbc mol -center*. Simulations were backmapped to all-atom using a modified version of the Backward algorithm (***Wassenaar et al., 2014***), in which simulation runs are excluded and energy minimization is shortened to 200 steps (***Larsen et al., 2020***). Every fourth simulation frame was backmapped for a total of 15,000 conformers in each backmapped ensemble.

Constructing ensembles of larger SPOP oligomers

We constructed conformational ensembles of larger SPOP$^{28-359}$ oligomers with up to 60 subunits by joining together conformers from the all-atom backmapped ensembles of the SPOP dodecamer. Using ensembles of two input SPOP oligomers (SPOP 1 and SPOP 2) we started by removing the last subunit of SPOP 1 and the first subunit of SPOP 2 to ensure that the newly joined subunits were internal and not terminal. We then removed additional subunits from SPOP 2 to reach the desired length of the output oligomer. Then, we selected a random frame from SPOP 1 and SPOP 2, superposed the BTB/BACK domains of the last two subunits of SPOP 1 to the BTB/BACK domains of the first two subunits of SPOP 2, and deleted the first two subunits of SPOP 2. Next, we checked for clashes between the newly joined subunits (shortest interatomic distance<0.4 Å), and rejected the new frame if there was a clash. This approach ensured that the terminal subunits in the constructed oligomer were also the terminal subunits in the MD simulation of the dodecamer, while all internal subunits in the constructed oligomer were also internal in the MD simulation. This approach was repeated to create 15,000 structures of each larger oligomer.

Calculating SAXS intensities from conformational ensembles

We calculated SAXS intensities from each of the 15,000 conformers in each of our all-atom ensembles of SPOP oligomers using Pepsi-SAXS (***Grudinin et al., 2017***). To avoid overfitting to the experimental SAXS data, we used fixed values for the parameters that describe the contrast of the hydration layer, $\delta\rho$=3.34 e/nm^3, and the volume of displaced solvent, r_0/r_m=1.025, that have been shown to work well for intrinsically disordered and multidomain proteins (***Pesce and Lindorff-Larsen, 2021***). The forward scattering ($I(0)$) was set equal to the number of subunits in the oligomer, in order to scale the SAXS intensities proportionally to the particle volume.

The isodesmic self-association model and averaging of SAXS intensities

The experimental SAXS profiles of SPOP report on the average of a polydisperse mixture of oligomeric species in solution. The concentration of each oligomer should follow the isodesmic model where the concentration of the smallest subunit, the BTB-BTB dimer, is given by:

$$c_1 = \frac{2c_{tot}K_A + 1 - \sqrt{4c_{tot}K_A + 1}}{2c_{tot}K_A^2} \quad (1)$$

The concentration c_i of any larger oligomer with i subunits can be calculated given c_1 and the concentration of oligomer i-1, c_{i-1}:

$$c_i = K_A c_{i-1} c_1 \quad (2)$$

K_A is the isodesmic association constant and c_{tot} is the total concentration of protomers. Here we assume that the SPOP BTB-BTB dimer is always fully formed (***Marzahn et al., 2016***) and c_{tot} in Eq. 1 is thus half of the total protein concentration reported for the SAXS experiments, which refers to the

SPOP monomer concentration. Given the concentration c_i of each oligomer i from the isodesmic model, we can calculate the volume fraction ϕ_i of the oligomer:

$$\phi_i = \frac{ic_i}{\sum_i^N ic_i} \tag{3}$$

The average SAXS intensities from the mixture of oligomers $\langle I \rangle_{\text{mix}}$ are then given by:

$$\langle I \rangle_{\text{mix}} = \sum_i^N \langle I \rangle_{i,\text{ensemble}} \phi_i \tag{4}$$

where $\langle I \rangle_{i,\text{ensemble}}$ is the conformationally averaged SAXS intensity of oligomer i. Note that the magnitude of the SAXS intensities calculated with Pepsi-SAXS were set to be proportional to the number of subunits in the oligomer, so given Eqs. 3 and 4 the total contribution of each oligomer to the averaged SAXS intensity is proportional to $i^2 c_i$.

Self-consistent optimization of isodesmic model parameters and conformational ensemble weights

The algorithm we developed to self-consistently optimize the isodesmic distribution of oligomer concentrations and reweight the conformational ensemble of each oligomer against SAXS data consists of three iterative steps: (1) fitting the scale and constant background of the SAXS data, (2) fitting the isodesmic K_{A}, and (3) reweighting the conformational ensemble of each oligomer using BME reweighting. We used a concentration series of SAXS experiments, to which the isodesmic K_{A} was fitted globally, and only subsequently transformed the K_{A} to the K_{D} ($K_D = 1/K_{\text{A}}$) for reporting our results.

Step 1: Fitting the SAXS scale and constant background

The following step was repeated for each SAXS experiment in the concentration series. The concentration of each oligomer was calculated using the isodesmic model with the given c_{tot} (Eqs. 1 and 2). The average SAXS intensities $\langle I \rangle_{\text{mix}}$ from all oligomers were then calculated using Eqs. 3 and 4. The scale and constant background (cst) of $\langle I \rangle_{\text{mix}}$ were fit to the experimental SAXS intensities, I_{exp}, using least-squares linear regression weighted by the experimental errors (*LinearRegression* function in scikit-learn (***Pedregosa et al., 2011***)):

$$I_{\text{exp}} = scale \langle I \rangle_{\text{mix}} + cst \tag{5}$$

In practice, to avoid modifying the SAXS scale and constant background for every conformer in our ensembles, we instead performed the inverse operation on the experimental SAXS profile:

$$I_{\text{exp,fit}} = \frac{I_{\text{exp}} - cst}{scale} \tag{6}$$

and propagated the experimental errors σ_{exp} accordingly:

$$\sigma_{\text{exp,fit}} = \frac{\sigma_{\text{exp}}}{|scale|} \tag{7}$$

Step 2: Fitting the isodesmic model

The isodesmic K_{A} was fitted globally to the concentration series of SAXS experiments using Metropolis Monte Carlo (***Metropolis et al., 1953***) with simulated annealing. For each Monte Carlo step, we generated a new random K_{A} with a Gaussian probability distribution centered around the previous K_{A}, calculated new oligomer concentrations and corresponding $\langle I \rangle_{\text{mix}}$ for each SAXS experiment in the concentration series using Eqs. 1–4, and for each SAXS experiment calculated the reduced χ^2, χ_r^2, as:

$$\chi_r^2 = \frac{1}{m}\sum_j^m \frac{(\langle I\rangle_{j,mix} - I_{j,exp})^2}{\sigma_{j,exp}^2} \tag{8}$$

where m is the number of SAXS intensities j in the SAXS profile. We then calculated the average of the χ_r^2-values across the SAXS concentration series to get the global χ_r^2, $\chi_{r,\text{global}}^2$, as we had the same number of intensities in each SAXS profile. Next, we evaluated the acceptance criterion by calculating:

$$\alpha = exp\left(-\frac{\chi_{\text{new,r,global}}^2 - \chi_{\text{old,r,global}}^2}{T}\right) \tag{9}$$

where $\chi_{\text{new,r,global}}^2$ and $\chi_{\text{old,r,global}}^2$ are the from the current and previous Monte Carlo step respectively and T is the simulated annealing temperature. If $\alpha > 1$, we accepted the new K_A. If $\alpha \leq 1$, we generated a random number, $rand$, between 0 and 1, and if $\alpha > rand$, accepted the new K_A. Otherwise, we kept the K_A from the previous Monte Carlo step. Finally, we decreased T for the next Monte Carlo step.

Step 3: Reweighting the conformational ensemble

The following step was repeated for each SAXS experiment in the concentration series. We calculated the oligomer concentrations using the isodesmic model given the new K_A determined in step 2. For each oligomer i, we extracted a SAXS profile for BME reweighting from the experimental profile using the following method: we calculated the average SAXS profile from the ensembles as in Eq. 4 but leaving out oligomer i from the sum to get $\langle I\rangle_{\text{mix,rest}}$. Next, we determined the contribution of species i to the experimental SAXS intensity as:

$$\langle I\rangle_{i,extr} = \frac{I_{\text{exp}} - \langle I\rangle_{\text{mix,rest}}}{\phi_i} \tag{10}$$

where I_{exp} is the experimental SAXS intensity and ϕ_i is the volume fraction of oligomer i. We then propagated the error $\sigma_{i,extr}$ on $\langle I\rangle_{i,extr}$ from both the errors on the experimental SAXS intensities and the errors on the calculated SAXS intensities, which we determined using block error analysis (***Flyvbjerg and Petersen, 1989***). The propagated errors were given by:

$$\sigma_{i,extr} = \frac{\sqrt{\sigma_{\text{exp}}^2 + \sum_r^N (\sigma_{r,block}\phi_r)^2}}{\phi_i} \tag{11}$$

where the sum r to N runs over all oligomers that contributed to $\langle I\rangle_{\text{mix,rest}}$, σ_{exp} is the error on the experimental SAXS intensity, $\sigma_{r,block}$ is the error on the average SAXS intensity calculated from the ensemble of oligomer r prior to reweighting using block error analysis (https://github.com/fpesceKU/BLOCKING), and ϕ_i is the volume fraction of oligomer i. The conformational ensemble of oligomer i was then reweighted against this extracted SAXS profile using BME reweighting (***Bottaro et al., 2020***), in which a set of ensemble weights w are obtained by minimizing the function:

$$\mathcal{L}(w_1 ... w_n) = \frac{m}{2}\chi_r^2(w_1 ... w_n) - \theta S_{rel}(w_1 ... w_n) \tag{12}$$

where n is the number of ensemble conformations, m is the number of experimental observables (in this case the number of SAXS intensities in the profile), χ_r^2 quantifies the agreement between $\langle I\rangle_{\text{ensemble}}$ and $\langle I\rangle_{extr}$, S_{rel} is the relative Shannon entropy that quantifies the deviation of the new weights from the initial weights, w^0, and θ is a scaling parameter that quantifies the confidence in the experimental data versus the prior ensemble. χ_r^2 is given by:

$$\chi_r^2(w_1 ... w_n) = \frac{1}{m}\sum_j^m \frac{\sum_k^n (w_k I_{j,k,\text{ensemble}} - \langle I\rangle_{j,extr})^2}{\sigma_{j,extr}^2} \tag{13}$$

where $I_{j,k,\text{ensemble}}$ is the SAXS intensity j calculated from the conformer k of the ensemble. S_{rel} is given by:

$$S_{rel} = -\sum_{k}^{n} w_k log\left(\frac{w_k}{w_k^0}\right) \quad (14)$$

Using the ensemble weights obtained from BME reweighting, we calculated new weighted average SAXS intensities, $\langle I \rangle_{i,\text{ensemble}}$, from the ensemble of oligomer i. The process of extracting a SAXS profile followed by BME reweighting was repeated for each oligomer.

Optimization parameters

The three steps described above were repeated iteratively to converge on self-consistent values of the SAXS scale and constant background, the isodesmic K_A, and the ensemble weights for each oligomer species. As the SAXS profile against which the ensemble of each oligomer was reweighted is a function of the ensemble weights of all other oligomeric species, we wished to reweight the ensembles only slightly in initial iterations, and then gradually increase the degree of reweighting as the conformational weights and isodesmic K_A converged. We achieved this by starting with a high value of θ (Eq. 12) and then gradually decreasing θ each iteration. The fraction of effective frames, ϕ_{eff}, given by $\exp(S_{rel})$, provides a measure of the fraction of the initial ensemble that is retained after reweighting. At every iteration, we checked whether the ensemble of each oligomer had reached a ϕ_{eff} below a set cut-off, after which θ was no longer decreased for that specific oligomer. Thus, the overall degree of reweighting could be tuned through selection of this ϕ_{eff}-cut-off.

We ran the optimization scheme for 1,000 iterations starting with θ=100 and decreasing θ by 2% every iteration. The simulated annealing of the isodesmic K_A was run from T=10 to T=0.1 every iteration, with T decreased by 30% every Monte Carlo step, and with a standard deviation of 0.1 μM^{-1} for the Gaussian probability distribution used to generate the new K_A. The step was repeated if $K_A \leq 0$ was generated.

Preventing overfitting

We ran the optimization with a range of ϕ_{eff}-cut-offs from 0.1 to 1. To prevent overfitting, we aimed to choose a value of ϕ_{eff} that retained as much of the prior ensemble as possible (high ϕ_{eff}) while not sacrificing substantial improvement in the fit to the SAXS data (low $\chi^2_{r,\text{global}}$). As an additional approach to prevent overfitting, we left out the SAXS experiment recorded with 15 μM protein from the optimization, and used it as validation for the determined weights (averaged as explained in the next section) and isodesmic K_A at different values of the ϕ_{eff}-cut-off. For each ϕ_{eff}-cut-off, we fitted only the SAXS scale and constant background to the 15 μM SAXS experiment. We tested the effect of using the fitted K_A and ensemble weights in combination, but also the effect of using only the fitted K_A or ensemble weights independently. Although in all cases the fitted K_A and ensemble weights combined improved the fit to the SAXS data compared with the initial weights and K_A, ϕ_{eff}-cut-off=0.4 was the lowest value of ϕ_{eff} where the fit was not improved by replacing the fitted weights with uniform weights in combination with the fitted K_A (Fig. S1). Thus, we selected the conformational weights and isodesmic K_A determined with ϕ_{eff}-cut-off=0.4 to avoid overfitting the ensemble weights.

Averaging the conformational weights from different SAXS experiments

The optimization scheme outputs a set of conformational weights for each SAXS experiment in the concentration series. We combined these conformational weights to obtain a single set of weights for further analysis, under the assumption that the conformational properties of each SPOP oligomer are independent of protein concentration. The distribution of oligomeric species from the isodesmic model depends on the protein concentration. Thus, each SAXS experiment does not contain the same amount of information on every oligomer; SAXS experiments at lower

concentrations have a relatively smaller contribution from large oligomers and vice versa. Therefore, we weighted the averaging of the conformational weights to reflect this mismatch in information. The average weight of conformation k of oligomer i was calculated as:

$$\langle w \rangle_{k,i} = \sum_{l}^{o} w_{k,i,l} \rho_{i,l} \tag{15}$$

where $w_{k,i,l}$ is the weight of conformer k of oligomer i from reweighting against SAXS experiment l and $\rho_{i,l}$ is the contribution of oligomer i to SAXS experiment l relative to the contribution of oligomer i to the other SAXS experiments in the concentration series, given by:

$$\rho_{i,l} = \frac{1}{\sum_{l}^{o} \frac{i^2 c_{i,l}}{\sum_{i}^{N} i^2 c_{i,l}}} \frac{i^2 c_{i,l}}{\sum_{i}^{N} i^2 c_{i,l}} \tag{16}$$

where $c_{i,l}$ is the concentration of oligomer i in SAXS experiment l given by the isodesmic model. For a plot of the contributions $\rho_{i,l}$, see Fig. S7.

Determining the error of the fitted isodesmic K_D

To determine the uncertainty of the isodesmic K_D fitted with our optimization scheme, we scanned a range of K_D values around the fitted K_D and determined the $\chi^2_{r,\text{global}}$ to the concentration series of SAXS data. We used the same ensemble weights for every value of K_D, and only fitted the scale and constant background to the SAXS data. We then defined the error of the fitted K_D to include all K_D values that gave a $\chi^2_{r,\text{global}}$ to the SAXS data within 10% of the minimum $\chi^2_{r,\text{global}}$ (Fig. S5).

Analysis of conformational ensembles

R_g was calculated from ensembles using the *gyrate* function in Gromacs. End-to-end distances were calculated from ensembles as the distances between the center-of-mass (COM) of the BTB/BACK domains in the terminal subunits using the *compute_center_of_mass* function in MDTraj (**McGibbon et al., 2015**) and the *linalg.norm* function in NumPy (**Harris et al., 2020**). We fitted ensemble averaged end-to-end distances against oligomer size (number of subunits) with a power law: $R_{\text{E-E}} = R_0 N^{\nu}$, where R_0 is the subunit segment size, N is the number of subunits in the oligomer, and ν is a scaling exponent, using the *curve_fit* function in SciPy (**Virtanen et al., 2020**). MATH-BTB/BACK COM distance was calculated from ensembles as the distance between the COM of the MATH domain and BTB/BACK domains in every subunit using the *compute_center_of_mass* function in MDTraj and the *linalg.norm* function in NumPy. The histogram of MATH-BTB/BACK COM distances shows values for all conformations of all subunits of all oligomers. The COM distance between substrate binding sites in neighbouring MATH domains was calculated from ensembles using the *distance* function in Gromacs. The MATH substrate binding site was defined as residue Arg70, Tyr87, Ser119, Tyr123, and Lys129-Phe133. The histogram of MATH binding site COM distances shows values for all conformations of all subunits of all oligomers. Structures for Fig. 4i were selected by fitting three Gaussians to the histogram in Fig. 4e (after reweighting) using SciPy *curve_fit* and for each Gaussian selecting conformers within 0.1σ of the mean. All visualizations of protein structures were made with ChimeraX (**Pettersen et al., 2021**).

Data availability

Data, simulations and code generated for this paper will be available at https://github.com/KULL-Centre/_2022_Thomasen_SPOP.

Acknowledgments

This work was supported by the Lundbeck Foundation BRAINSTRUC structural biology initiative (R155-2015-2666, to K.L.-L.), the PRISM (Protein Interactions and Stability in Medicine and Genomics) centre funded by the Novo Nordisk Foundation (NNF18OC0033950, to K.L.-L.) and by NIH grant

R01GM112846 (to T.M.) and by the American Lebanese Syrian Associated Charities (to T.M.). We acknowledge access to computational resources from the ROBUST Resource for Biomolecular Simulations (supported by the Novo Nordisk Foundation; NNF18OC0032608), the Danish National Supercomputer for Life Sciences (Computerome), and the Biocomputing Core Facility at the Department of Biology, University of Copenhagen. We thank Melissa R. Marzahn and Erik W. Martin for the generation of preliminary data. We thank Shirish Chodankar for assistance with SAXS data collection and reduction. The LiX beamline is part of the Center for BioMolecular Structure (CBMS), which is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1605010). LiX also received additional support from NIH Grant S10 OD012331. As part of NSLS-II, a national user facility at Brookhaven National Laboratory, work performed at the CBMS is supported in part by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Program under contract number DE-SC0012704.

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015; 1-2:19–25. doi: 10.1016/j.softx.2015.06.001.

Ali MH, Imperiali B. Protein oligomerization: How and why. Bioorganic and Medicinal Chemistry. 2005; 13(17):5013–5020. doi: 10.1016/j.bmc.2005.05.037.

An J, Wang C, Deng Y, Yu L, Huang H. Destruction of Full-Length Androgen Receptor by Wild-Type SPOP, but Not Prostate-Cancer-Associated Mutants. Cell Reports. 2014 feb; 6(4):657–669. https://doi.org/10.1016/j.celrep.2014.01.013, doi: 10.1016/j.celrep.2014.01.013.

Bosu DR, Kipreos ET. Cullin-RING ubiquitin ligases: Global regulation and activation cycles. Cell Division. 2008; 3:1–13. doi: 10.1186/1747-1028-3-7.

Bottaro S, Bengtsen T, Lindorff-Larsen K. Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach. Methods in Molecular Biology. 2020; 2112:219–240. doi: 10.1007/978-1-0716-0270-6_15.

Bouchard JJ, Otero JH, Scott DC, Szulc E, Martin EW, Sabri N, Granata D, Marzahn MR, Lindorff-Larsen K, Salvatella X, Schulman BA, Mittag T. Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. Molecular Cell. 2018 oct; 72(1):19–36.e8. https://doi.org/10.1016/j.molcel.2018.08.027, doi: 10.1016/j.molcel.2018.08.027.

Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. Journal of Chemical Physics. 2007; 126(1):1–7. doi: 10.1063/1.2408420.

Cuneo MJ, Mittag T. The ubiquitin ligase adaptor SPOP in cancer. FEBS Journal. 2019; 286(20):3946–3958. doi: 10.1111/febs.15056.

DiFabio J, Chodankar S, Pjerov S, Jakoncic J, Lucas M, Krywka C, Graziano V, Yang L. The life science x-ray scattering beamline at NSLS-II. AIP Conference Proceedings. 2016 jul; 1741(1):30049. https://aip.scitation.org/doi/abs/10.1063/1.4952872, doi: 10.1063/1.4952872.

Errington WJ, Khan MQ, Bueler SA, Rubinstein JL, Chakrabartty A, Privé GG. Adaptor Protein Self-Assembly Drives the Control of a Cullin-RING Ubiquitin Ligase. Structure. 2012 jul; 20(7):1141–1153. https://doi.org/10.1016/j.str.2012.04.009, doi: 10.1016/j.str.2012.04.009.

Flyvbjerg H, Petersen HG. Error estimates on averages of correlated data. The Journal of Chemical Physics. 1989; 91(1):461–466. doi: 10.1063/1.457480.

Giannakis M, Mu X, Shukla S, Qian Z, Cohen O, Nishihara R, Bahl S, Cao Y, Amin-Mansour A, Yamauchi M, Sukawa Y, Stewart C, Rosenberg M, Mima K, Inamura K, Nosho K, Nowak J, Lawrence M, Giovannucci E, Chan A, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. Cell Reports. 2016; 15(4):857–865. https://www.sciencedirect.com/science/article/pii/S2211124716303643, doi: https://doi.org/10.1016/j.celrep.2016.03.075.

Grudinin S, Garkavenko M, Kazennov A. Pepsi-SAXS: An adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. Acta Crystallographica Section D: Structural Biology. 2017; 73(5):449–464. doi: 10.1107/S2059798317005745.

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, et al. Array programming with NumPy. Nature. 2020 Sep; 585(7825):357–362. https://doi.org/10.1038/s41586-020-2649-2, doi: 10.1038/s41586-020-2649-2.

Hernández-Muñoz I, Lund AH, van der Stoop P, Boutsma E, Muijrers I, Verhoeven E, Nusinow DA, Panning B, Marahrens Y, van Lohuizen M. Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. Proceedings of the National Academy of Sciences. 2005; 102(21):7635–7640. https://www.pnas.org/doi/abs/10.1073/pnas.0408918102, doi: 10.1073/pnas.0408918102.

Herranz-Trillo F, Groenning M, Van Maarschalkerweerd A, Tauler R, Vestergaard B, Bernadó P. Structural analysis of multi-component amyloid systems by chemometric SAXS data decomposition. Structure. 2017; 25(1):5–15.

Kent D, Bush EW, Hooper JE. Roadkill attenuates Hedgehog responses through degradation of Cubitus interruptus. Development. 2006 may; 133(10):2001–2010. https://doi.org/10.1242/dev.02370, doi: 10.1242/dev.02370.

Kim MS, Je EM, Oh JE, Yoo NJ, Lee SH. Mutational and Expressional Analyses Of SPOP, A Candidate Tumor Suppressor Gene, In Prostate, Gastric and Colorectal Cancers. Apmis. 2013; 121(7):626–633. doi: 10.1111/apm.12030.

Krauthammer M, Kong Y, Ha BH, Evans P, Bacchiocchi A, McCusker JP, Cheng E, Davis MJ, Goh G, Choi M, Ariyan S, Narayan D, Dutton-Regester K, Capatana A, Holman EC, Bosenberg M, Sznol M, Kluger HM, Brash DE, Stern DF, et al. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. Nature Genetics. 2012; 44(9):1006–1014. https://doi.org/10.1038/ng.2359, doi: 10.1038/ng.2359.

Kwon JE, La M, Oh KH, Oh YM, Kim GR, Seol JH, Baek SH, Chiba T, Tanaka K, Bang OS, Joe CO, Chung CH. BTB Domain-containing Speckle-type POZ Protein (SPOP) Serves as an Adaptor of Daxx for Ubiquitination by Cul3-based Ubiquitin Ligase. Journal of Biological Chemistry. 2006 may; 281(18):12664–12672. http://www.jbc.org/content/281/18/12664.abstract, doi: 10.1074/jbc.M600204200.

Larsen AH, Wang Y, Bottaro S, Grudinin S, Arleth L, Lindorff-Larsen K. Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. PLOS Computational Biology. 2020 apr; 16(4):e1007870. https://doi.org/10.1371/journal.pcbi.1007870.

Le Gallo M, O'Hara AJ, Rudd ML, Urick ME, Hansen NF, O'Neil NJ, Price JC, Zhang S, England BM, Godwin AK, Sgroi DC, Hieter P, Mullikin JC, Merino MJ, Bell DW, Program NIHISCNCS. Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. Nature Genetics. 2012; 44(12):1310–1315. https://doi.org/10.1038/ng.2455, doi: 10.1038/ng.2455.

Lynch M. The Evolution of Multimeric Protein Assemblages. Molecular Biology and Evolution. 2012 may; 29(5):1353–1366. https://doi.org/10.1093/molbev/msr300, doi: 10.1093/molbev/msr300.

Marsh JA, Teichmann SA. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. Annual Review of Biochemistry. 2015; 84(1):551–575. doi: 10.1146/annurev-biochem-060614-034142.

Marzahn MR, Marada S, Lee J, Nourse A, Kenrick S, Zhao H, Ben-Nissan G, Kolaitis R, Peters JL, Pounds S, Errington WJ, Privé GG, Taylor JP, Sharon M, Schuck P, Ogden SK, Mittag T. Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles . The EMBO Journal. 2016; 35(12):1254–1275. doi: 10.15252/embj.201593169.

McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang LP, Lane TJ, Pande VS. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophysical Journal. 2015; 109(8):1528 – 1532. doi: 10.1016/j.bpj.2015.08.015.

Meisburger SP, Xu D, Ando N. REGALS: a general method to deconvolve X-ray scattering data from evolving mixtures. IUCrJ. 2021; 8(2):225–237.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics. 1953 jun; 21(6):1087–1092. https://doi.org/10.1063/1.1699114, doi: 10.1063/1.1699114.

627 **Oosawa F**, Kasai M. A theory of linear and helical aggregations of macromolecules. Journal of Molecular Biology.
628 1962; 4(1):10–21. doi: 10.1016/S0022-2836(62)80112-0.

629 **Parrinello M**, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal
630 of Applied Physics. 1981 dec; 52(12):7182–7190. https://doi.org/10.1063/1.328693, doi: 10.1063/1.328693.

631 **Pedregosa F**, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R,
632 Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine
633 Learning in Python. Journal of Machine Learning Research. 2011; 12:2825–2830.

634 **Pesce F**, Lindorff-Larsen K. Refining conformational ensembles of flexible proteins against small-angle x-ray
635 scattering data. Biophysical Journal. 2021; 120(22):5124–5135. https://www.sciencedirect.com/science/article/
636 pii/S0006349521008286, doi: https://doi.org/10.1016/j.bpj.2021.10.003.

637 **Pettersen EF**, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX:
638 Structure visualization for researchers, educators, and developers. Protein Science. 2021 jan; 30(1):70–82.
639 https://doi.org/10.1002/pro.3943, doi: https://doi.org/10.1002/pro.3943.

640 **Pierce WK**, Grace CR, Lee J, Nourse A, Marzahn MR, Watson ER, High AA, Peng J, Schulman BA, Mittag T.
641 Multiple Weak Linear Motifs Enhance Recruitment and Processivity in SPOP-Mediated Substrate Ubiquiti-
642 nation. Journal of Molecular Biology. 2016; 428(6):1256–1271. http://dx.doi.org/10.1016/j.jmb.2015.10.002,
643 doi: 10.1016/j.jmb.2015.10.002.

644 **Šali A**, Blundell TL. Comparative Protein Modelling by Satisfaction of Spatial Restraints. Journal of Molecu-
645 lar Biology. 1993; 234(3):779–815. http://www.sciencedirect.com/science/article/pii/S0022283683716268, doi:
646 https://doi.org/10.1006/jmbi.1993.1626.

647 **Schmit JD**, Bouchard JJ, Martin EW, Mittag T. Protein network structure enables switching between liquid and
648 gel states. Journal of the American Chemical Society. 2019; 142(2):874–883.

649 **Shemesh A**, Ginsburg A, Dharan R, Levi-Kalisman Y, Ringel I, Raviv U. Structure and energetics of gtp-and
650 gdp-tubulin isodesmic self-association. ACS Chemical Biology. 2021; 16(11):2212–2227.

651 **Souza PCT**, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grünewald F, Patmanidis I, Abdizadeh H, Bruininks
652 BMH, Wassenaar TA, Kroon PC, Melcr J, Nieto V, Corradi V, Khan HM, Domański J, Javanainen M, Martinez-
653 Seara H, Reuter N, Best RB, et al. Martini 3: a general purpose force field for coarse-grained molecular
654 dynamics. Nature Methods. 2021; 18(4):382–388. doi: 10.1038/s41592-021-01098-3.

655 **Studier FW**. Protein production by auto-induction in high-density shaking cultures. Protein Expression and
656 Purification. 2005; 41(1):207–234. https://www.sciencedirect.com/science/article/pii/S1046592805000264, doi:
657 https://doi.org/10.1016/j.pep.2005.01.016.

658 **Thomasen FE**, Lindorff-Larsen K. Conformational ensembles of intrinsically disordered proteins and flexible
659 multidomain proteins. Biochemical Society Transactions. 2022 feb; p. BST20210499. https://doi.org/10.1042/
660 BST20210499, doi: 10.1042/BST20210499.

661 **Thomasen FE**, Pesce F, Roesgaard MA, Tesei G, Lindorff-Larsen K. Improving Martini 3 for Disordered and
662 Multidomain Proteins. Journal of Chemical Theory and Computation. 2022 apr; 18(4):2033–2041. https:
663 //doi.org/10.1021/acs.jctc.1c01042, doi: 10.1021/acs.jctc.1c01042.

664 **Van Geersdaele LK**, Stead MA, Harrison CM, Carr SB, Close HJ, Rosbrook GO, Connell SD, Wright SC. Structural
665 basis of high-order oligomerization of the cullin-3 adaptor SPOP. Acta Crystallographica Section D: Biological
666 Crystallography. 2013; 69(9):1677–1684. doi: 10.1107/S0907444913012687.

667 **Virtanen P**, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser
668 W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E,
669 Carey CJ, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;
670 17:261–272. doi: 10.1038/s41592-019-0686-2.

671 **Wassenaar TA**, Ingólfsson HI, Böckmann RA, Tieleman DP, Marrink SJ. Computational lipidomics with insane:
672 A versatile tool for generating custom membranes for molecular simulations. Journal of Chemical Theory
673 and Computation. 2015; 11(5):2144–2155. doi: 10.1021/acs.jctc.5b00209.

674 **Wassenaar TA**, Pluhackova K, Böckmann RA, Marrink SJ, Tieleman DP. Going Backward: A Flexible Geometric
675 Approach to Reverse Transformation from Coarse Grained to Atomistic Models. Journal of Chemical Theory
676 and Computation. 2014 feb; 10(2):676–690. https://doi.org/10.1021/ct400617g, doi: 10.1021/ct400617g.

677 **Zhuang M**, Calabrese MF, Liu J, Waddell MB, Nourse A, Hammel M, Miller DJ, Walden H, Duda DM, Seyedin
678 SN, Hoggard T, Harper JW, White KP, Schulman BA. Structures of SPOP-Substrate Complexes: Insights into
679 Molecular Architectures of BTB-Cul3 Ubiquitin Ligases. Molecular Cell. 2009 oct; 36(1):39–50. https://doi.org/
680 10.1016/j.molcel.2009.09.022, doi: 10.1016/j.molcel.2009.09.022.