Prediction and improved characterization of coiled-coil protein origami nanostructures by comparative modeling

by

Ratnakshi Mandal

B.Tech., University of Calcutta, 2015

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Tim Taylor Department of Chemical Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Won Min Park

Copyright

© Ratnakshi Mandal 2022.

Abstract

Coiled-coil protein origami (CCPO) is a method that connects and folds coiled-coil protein modules into well-defined nanostructures, which offer a great promise in the fields of nanotechnology and biomaterials. In the determination of atomistic details of a CCPO nanostructure, small-angle X-ray scattering (SAXS) has been used in combination with comparative protein structure modeling. The modeling utilizes a critical step of molecular dynamics (MD) optimization with simulated annealing for structure refinement, but the details of the optimization and reliable evaluation for CCPO models are not available. In this report, the effect of MD optimization on the accuracy of comparative modeling was studied by the fitting of SAXS data. Under extended MD optimization, structural models of nearly complete matches to SAXS data were built. In addition, a method of predicting the radius of gyration of comparative structure models was developed, which enabled a significantly improved evaluation of the comparative models of CCPO. It will provide great potential as a method for computational screening of CCPO designs.

Table of Contents

List of Figures

Acknowledgements

I would like to extend my sincere gratitude and thanks to the Tim Taylor Department of Chemical Engineering for their continued support and the funding opportunity for the work that is presented here. I would also thank my advisor, Dr. Won Min Park for the constant guidance and motivation to finish my project and MS Report. I would like to thank my committee members for their help with this draft. Further on, I would also thank my lab mates who have been a constant support during my time here.

Introduction

Precisely controlled nanostructures built from biological molecules have been of great interest for a variety of applications in the field of biotechnology. For example, DNA origami nanostructures, which were built from well-controlled programmable assemblies of DNA strands in various shapes,[1] have been used as novel nano-architectures for biomolecular detection[2], nanomechanical molecular devices[2], DNA nanorobots[2] for nano-plasmonics [3] and fluorescence imaging agents[4]. Such devices were based on DNA origami nano-architectures assembled in well-defined configurations, which are challenging tasks using other materials like metal nanoparticles.[5,6] Transited from DNA to protein, new origami methods have recently been developed using α-helical protein motifs[7], coiled-coil motifs, which are tunable in affinity and specificity and useful in attaining stability and assembling supramolecular nanostructures.[8] The method is called coiled-coil protein origami (CCPO), which is a method to create three-dimensional protein nanostructures in a well-defined shape by genetic fusion and folding of orthogonally interacting coiled-coil motifs.[8,9] The methods of CCPO developed were successfully used to build versatile polygonal[10] and polyhedral[9] protein nanostructures.

One of the tools to characterize CCPO nanostructures at atomic resolution is small-angle X-ray scattering (SAXS). The scattering profiles have been used to determine a radius of gyration (R_g) as well as an envelope of a CCPO nanostructure. R_g is defined as the root mean square average of the distance of the electrons from the center of the particle. It is calculated from the slope of a Guinier plot.[11] Furthermore, combined with the comparative structure modeling, SAXS scattering profiles have been used to characterize atomic-resolution structures of CCPO designs using SAXS modeling tools.[12,13] Comparative structure modeling is a technique to construct a protein structure

model from its amino acid sequence and templates that are experimentally determined structures of homologous proteins.[12,13] The modeling procedure for a CCPO nanostructure includes the steps of selection of coiled-coil templates, alignment of a CCPO design with the sequences of the templates, model building, refinement by molecular dynamics (MD) with simulated annealing,[14] followed by model evaluation, as shown in Figure 1. In the comparative modeling of CCPO nanostructures, the degree of MD refinement with simulated annealing significantly affects the resulting models due to the presence of flexible regions.[15] Hence, a comprehensive study on the degree and length of simulated annealing may help improve the accuracy of SAXS characterization that is combined with comparative modeling. In this report, the effect of the length of MD optimization with simulated annealing on the accuracy of comparative modeling of CCPO nanostructures is demonstrated. Prediction of R_g and a new method of model evaluation were also described, which can be used as a potentially useful tool for the design of new CCPO designs.

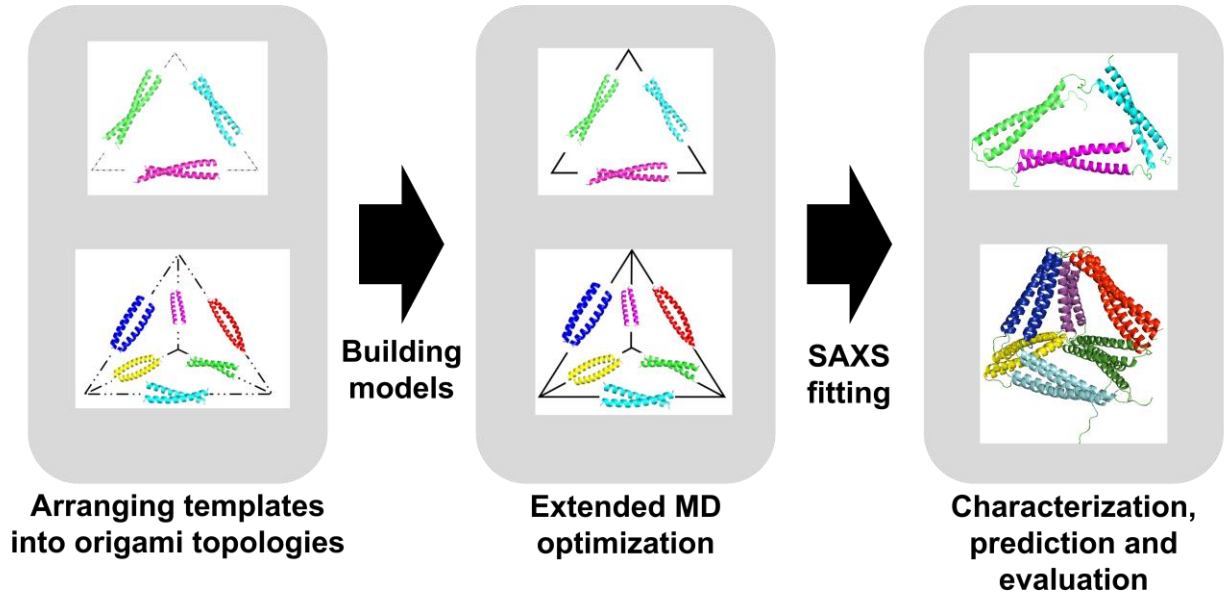

Figure 1: Scheme of comparative modeling with extended MD optimization for characterization, prediction, and evaluation of CCPO nanostructures.

Background

1. Coiled-Coil Protein Motifs

Coiled-coil proteins, consisting of two or more α-helices wrapped around each other into a super-helical bundle, are ubiquitous structural protein motifs associated with a range of functions in nature.[16] The amino acid sequences of peptides forming coiled-coil bundles are characterized by a heptad repeating unit denoted as $(abcdefg)_n$, where n is the number of repeats.[17] Non-polar residues preferentially occur in the first (*a*) and fourth (*d*) positions of the heptad. The fifth (*e*) and seventh (*g*) positions tend to be occupied by polar/charged amino acids, together creating complementary charge pairs across the bundle, as shown in the helical diagram of a dimeric coiled-coil complex (Figure 2). These help to stabilize a coiled-coil protein via inter-strand electrostatic interactions.[17]

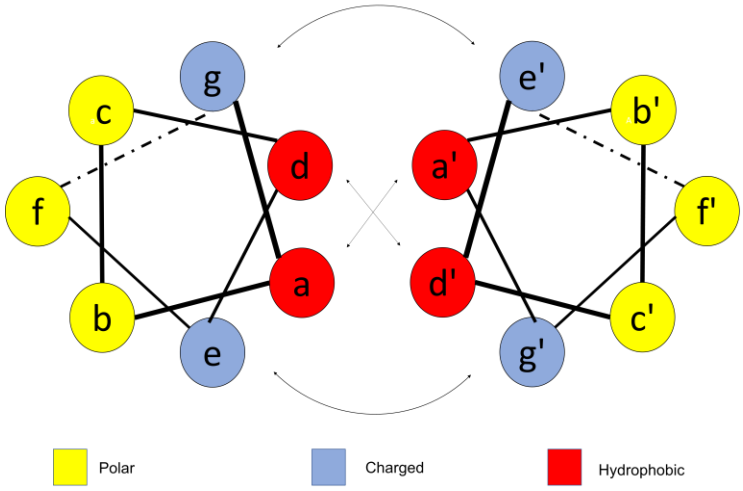

Figure 2: A helical diagram of a dimeric coiled-coil protein (left) and schematic side view of parallel and antiparallel complexes (right)[8]

The affinity and stability of coiled-coil motifs is mainly achieved by a distinctive 'knobs-into-holes' packing of the non-polar side chains into a hydrophobic core, which was first postulated by Crick.[18] Residues in the (*b*), (*c*), and (*f*) positions can be used to provide enough solubility of

the peptides, as well as to control the higher-order aggregation of oligomers via the exterior surface of a coiled-coil motif. In Figure 2, the interactions of the residues in the heptad over 2 turns of a dimeric protein complex is illustrated. Residues in the (*a*) and (*d*) positions have a hydrophobic interaction while residues in the (*e*) and (*g*) positions have electrostatic interactions. While the hydrophobic and electrostatic interactions play a critical role in stabilization of the specific binding, these interactions can be used to create different types of nanomaterials from these proteins. Also, modifying the residues in certain positions of the heptad lead to the formation of stable coiled-coil protein nanoparticles.

Recent results suggest that ionic interactions are not only important for proper alignment, orientation, and selectivity of coiled-coil motifs but can also contribute considerably to the stability, thereby modulating assembly of coiled coils in a pH-dependent manner.[19] Moreover, the coiled-coil assembly can also be modulated by phosphorylation or by interaction with ions.[20] This makes the coiled-coil system a highly versatile protein folding motif adaptable to many biological functions. As a result, coiled-coil domains are involved in molecular recognition systems and protein refolding processes and can even form ion channels.[18]

2. Coiled-Coil Protein Origami (CCPO)

In the strategies of CCPO, orientations of coiled-coil protein motifs are precisely controlled by specific intra- or intermolecular dimerization of multiple distinct coiled-coil modules, which were arranged to form polygonal or polyhedral shapes.[15] The first method of CCPO was developed by the Jerala group. It is based on linking complimentary coiled-coil strands into a single chain (Figure 3a).[9,10,21] The interaction between the complimentary coiled-coil strands folds the single fusion protein chain into the desired shape. A tetrahedron could be created from a polypeptide chain of

either three parallel and three antiparallel coiled-coil pairs or four parallel and two antiparallel coiled-coil pairs. The flexible linker used was a tetrapeptide of Ser-Gly-Pro-Gly to connect the consecutive foiled coil forming segments.[21] Using the method, various shapes such as four-sided pyramid[9], triangular prism[9], triangular bipyramids[22], and triangle[10] were also created. The design of the folding pathway[23] for a single-chain CCPO was also studied.

The Keating group developed a different method of CCPO, which is used to design and assemble multiple fusion proteins of coiled-coil modules (Figure 3b). They reported building a triangle shape[15] which was assembled from three fusion proteins. The coiled-coil fusion proteins were designed by genetically linking six orthogonally interacting heterospecific synthetic coiled-coil peptides (SYNZIPs).[24] The Woolfson group reported examples of square and triangle assembled from a pair of dimeric coiled-coil in which the length of the linker between the coiled-coil pair was varied[25].

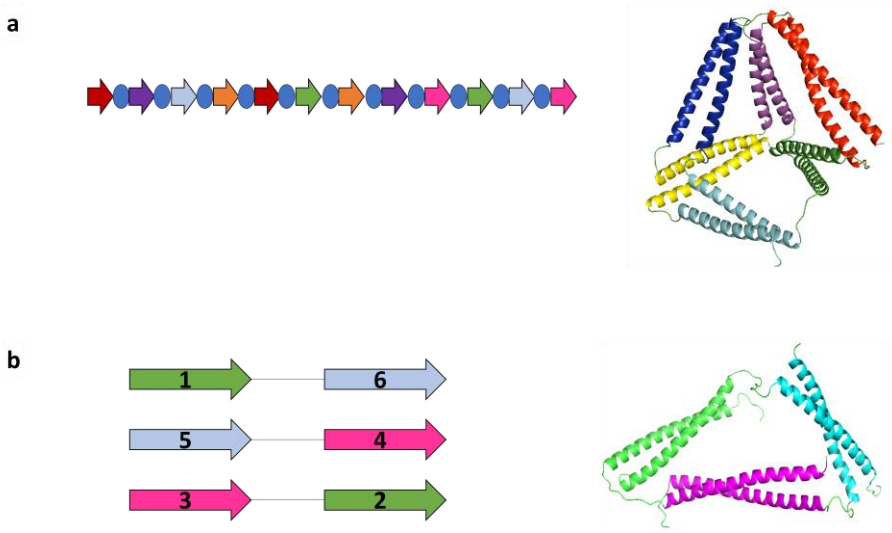

Figure 3: The methods of CCPO. (a) Design of a tetrahedron by linking coiled-coil segments in a single fusion protein strand[21]. (b) Design of a nanotriangle by genetic fusion and assembly of multiple coiled-coil protein strands.[15]

The CCPO nanostructures have been characterized using various tools. One of the tools is circular dichroism (CD) spectroscopy which measures the helicity of CCPO structures. CCPO nanostructures have been imaged using the transmission electron microscope (TEM) and atomic force microscope (AFM). The size, shape, topology of the CCPO nanostructures have been determined by the AFM and TEM imaging. Dynamic light scattering helps to determine a hydrodynamic radius (R_H) of CCPO nanostructures. Another powerful tool is SAXS. While a SAXS profile is used to determine the R_g and the structure envelope of a CCPO nanostructure, further details at atomic resolution are achievable if it is combined with SAXS computation and fitting to comparative structure models.

3. **Comparative Protein Structure Modeling and SAXS Computation**

In the CCPO design, a computational platform to build structure models of their CCPO designs was developed using a comparative modeling tool, MODELLER.[26] In comparative modeling, there are several steps. First, templates are identified, and initial alignment and alignment correction are made. Models are built by templating the coiled-coil helices, and the model structures are optimized with the variable target function method (VTFM) and refined by MD with simulated annealing.[14] Simulated annealing comprises of optimizing the system at a high temperature and then slowly lowering the temperature until the system freezes and no further changes occur.[27] This method is quite similar to structure determination by NMR spectroscopy.[26] After that, models are evaluated using score functions, such as Discrete Optimized Protein Energy (DOPE) [28] score or its normalized one (z-DOPE).

Theoretical SAXS profiles of the comparative protein structure models are computed using tools, which include the FoXS web server (http://salilab.org/foxs). The results from CCPO

structure models are fit to experimentally determined SAXS data, followed by measurement of the function χ^2, which indicates the quality of the fit. The models with the lowest χ^2 values are used to determine CCPO structures at atomic scale resolution.

Methodology

1. Comparative Protein Structure Modelling

Comparative structure models of coiled-coil protein origami were built using MODELLER[29]. The crystal structures of SYNZIP1:SYNZIP2 (PDB ID: 3HE5)[24] and SYNZIP5:SYNZIP6 (PDB ID: 3HE4)[30] and the predicted structure model of SYNZIP3:SYNZIP4[24] were used as templates to build the triangle structure models. For building the tetrahedron structure models, the predicted structure models of P3-P4, P5-P6, P7-P8, GCN_{sh}–GCN_{sh}, APH-APH, BCR-BCR were used as templates. The coiled-coil templates were arranged into the corresponding topologies of a triangle or a tetrahedron, where they were set to be in end-to-end distances (from 118.4 Å to 189 Å for the triangle and 51.8 Å to 98.6 Å for the tetrahedron) sufficient to avoid entanglement of linked-coiled-coil chains. Numerous models (from 100 to 500 models) were generated with no spatial restraints applied to the linkers (G_4SG_4S or SGPG). Then, the models were refined using slow MD optimization with simulated annealing which was programmed in MODELLER.[27] The cycle of MD optimization with simulated annealing was varied from 1 to 15. For each model built, the normalized Discrete Optimized Protein Energy (z-DOPE) score[28,31] was calculated accordingly.

2. Small Angle X-Ray Scattering (SAXS) Data

The synchrotron SAXS data of the CCPO triangle[15] and tetrahedron[21] were used for the fitting and analysis of comparative protein structure models. In the publication,[15] the SYNZIP triangle was prepared in 50 mM Tris, 150 mM NaCl, and 1 mM EDTA and at pH 7.4 for SAXS analysis. We used the scattering curve of the negatively supercharged tetrahedron (TET12(1.10)SN-f5 or TET12SN)[21] available at the Small Angle Scattering Biological Data Bank (SASBDB,

https://www.sasbdb.org). The tetrahedron sample condition was in 20 mM Tris, 150 mM NaCl, 10% glycerol, and at pH 7.5.[21]

3. Computation and Fitting of SAXS Data

Theoretical scattering profiles of the comparative protein structure models were computed and fit to the experimental SAXS data using the FoXS web server (http://salilab.org/foxs). The 1100 Theoretical scattering profiles of the comparative protein structure models were computed and fit to the experimental SAXS data using the FoXS web server (http://salilab.org/foxs)[12,13]. The 1100 resulting structure models of either the triangle or tetrahedron were fit to the corresponding SAXS data, followed by measurement of the function χ^2, which indicates the quality of the fit: resulting structure models of either the triangle or tetrahedron were fit to the corresponding SAXS data, followed by measurement of the function χ^2, which indicates the quality of the fit:

$$\chi^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left[\frac{I_{exp}(q_k) - I_{calc}(q_k)}{\sigma\left(I_{exp}(q_k)\right)}\right]^2$$

where q_i is the scattering vector, I_{exp} and I_{calc} are the experimental and computed scattered intensities, c is the scaling parameter, and σ is the experimental error. The R_g of each structure model was also calculated using the FoXS web server[12,13]. The relative absolute error (RAE) in R_g, which was used as an indicator of structure, was calculated by:

$$RAE = \frac{R_g(individual) - R_g(predicted)}{R_g(predicted)}$$

The Pearson's correlation coefficient was calculated for the linear correlations between the data sets of χ^2 and z-DOPE score or χ^2 and the RAE in R_g, which was used to assess the significance of the prediction power.

Results

1. Structure Refinement by Extended MD Optimization

To investigate the effect of MD optimization on the refinement of comparative models, 100 models of the triangle[15] and tetrahedron[21] were built, while the cycles of thorough MD optimization was varied from 1 to 15. The MD optimization with simulated annealing was performed in MODELLER.[29] Experimentally determined SAXS data of the triangle[15] or tetrahedron[21] were fit to the theoretical SAXS profiles calculated from each model. The parameter χ^2 indicates how a computational structure compares to the SAXS data. A value of 1 indicates that the corresponding computational model excellently matches the SAXS data and can be considered as a reliable model. All the 100 models from various cycles of MD optimization were fit to the data. Figure 4 shows the distribution and average of χ^2 values for the triangle. The result indicates that the fraction of high-quality models ($\chi^2 < 1.5$) significantly increased to 37% when 5 cycles of MD optimization were conducted, and it remains almost constant after more cycles. Also, the average χ^2 value went down to 2.2 ± 1.44 after 5 cycles and remained unchanged after more cycles up to 15. In the following study for the tetrahedron, similar results were also obtained. In Figure 3, the fraction of models with relatively low χ^2 values ($\chi^2 < 5$) increased after 5 cycles, and a significant decrease in an average value was not observed after 5 cycles of MD optimization (average of χ^2 was 21.4 ± 15.29). The range of the average values of χ^2 for the tetrahedron models was significantly higher (from 20.0 to 44.5) than the average values for the triangle models (from 2.20 to 3.99). This might be because of the low errors from the SAXS profiles and more entangled chains in the CCPO design.

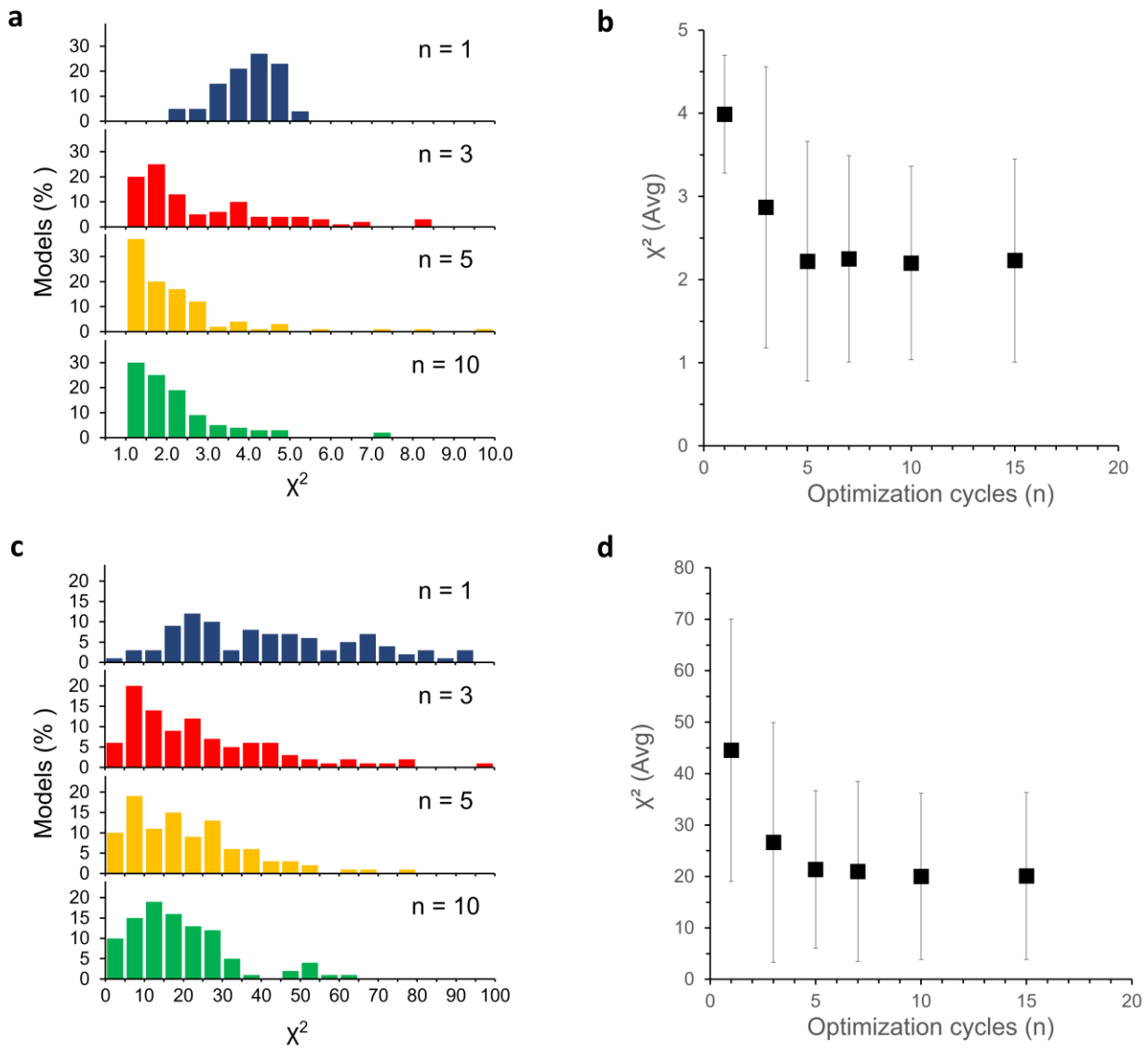

Figure 4: The quality of SAXS fitting of the comparative models from various cycles of MD optimization, *n*. (a) Distribution of χ^2 for the triangle models. (b) Average of χ^2 for the triangle models. (c) Distribution of χ^2 for the tetrahedron. (d) Average of χ^2 for the tetrahedron.

Next, we also investigate the effect of the number of models on SAXS fitting and quality. The number of models from 10 cycles of MD optimization was varied from 100 to 500, and all models were fit to the SAXS data. In both the triangle and tetrahedron, no significant differences in the distribution and average of when the total number of models were doubled or five-folded (Figure 5).

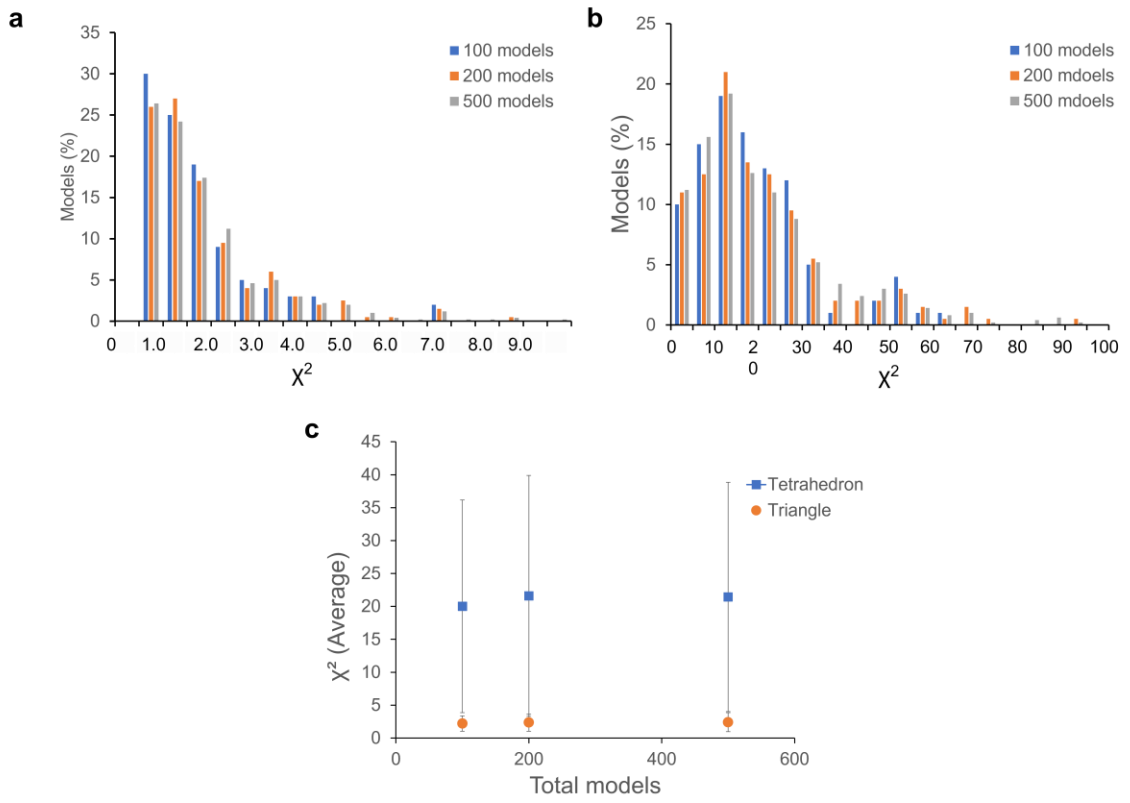

Figure 5: The quality of SAXS fitting of various total numbers of the comparative models (cycles of MD optimization = 10 and total model numbers = 100, 200, and 500). (a) Distribution of χ^2 for the triangle. (b) Distribution of χ^2 for tetrahedron. (c) Average of χ^2.

2. Improved SAXS Fitting and Structural Characterization

Building models with χ^2 that are very close to 1 is critical for reliable characterization of CCPO nanostructure using SAXS data. Increasing the cycles of MD optimization or the total number of models at a fixed cycle resulted in the best models with the lowest χ^2 for both the triangle and tetrahedron (Figure 6). In modeling for the triangle, the lowest χ^2 of the 100 models was 2.17 at a single cycle of MD optimization, while the lowest χ^2 value of 1.01 was obtained from 15 cycles. In modeling of the tetrahedron, the lowest χ^2 value was 4.11 at a single cycle of MD optimization, while 15 cycles of MD optimization resulted in the lowest χ^2 value of 1.01. Similarly, the lowest

χ^2 decreased when the total number of comparative models from 10 cycles of MD optimization increased from 100 to 200 or 500.

The extended cycles of MD optimization or the increase in the total number of models led to improvement in the quality of SAXS fitting for CCPO nanostructure characterization. The best triangle model with a χ^2 value of 1.01 was built out of 500 comparative models, which is comparable to a previously reported model that was fitted to the same SAXS data (χ^2 = 1.01).[15] Interestingly, the atomic-resolution structure of this new model is distinct from the previously reported model (Figure 7). Moreover, building total tetrahedron 500 models with 10 cycles of MD optimization resulted in the best model with a χ^2 value of 0.95. This is even lower than the value of the reported tetrahedron model fitted to the same SAXS data (χ^2 = 1.21 calculated by FoXS). This new model showed distinct orientations of the coiled-coil edges within a tetrahedron topology (Figure 7b). These new models nearly completely matched the SAXS data but were distinct in the coiled-coil orientations, which provides additional information about CCPO folding.

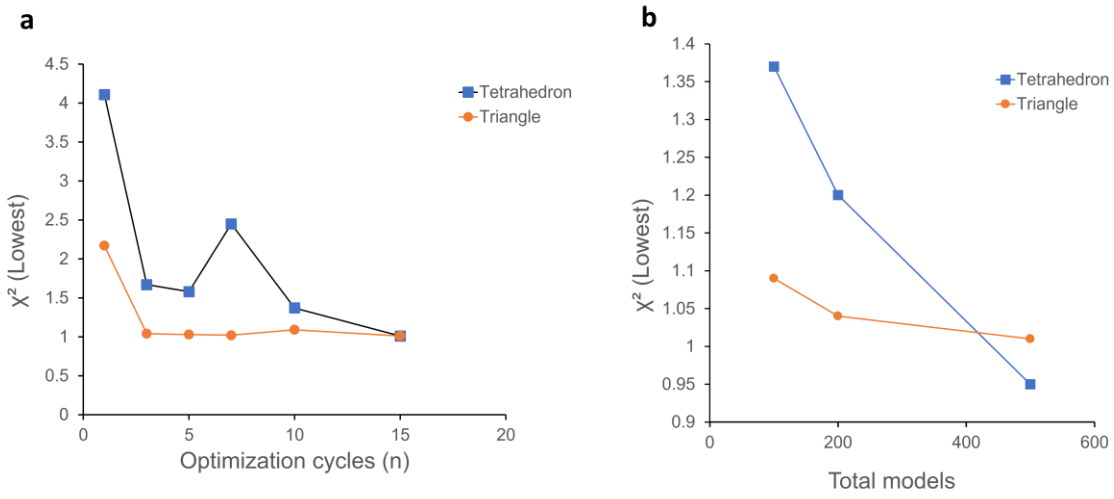

Figure 6: (a) The lowest χ^2 for different lengths of optimization cycles for both triangle and tetrahedron. (b) The lowest χ^2 for different total number of models built for the triangle and the tetrahedron.

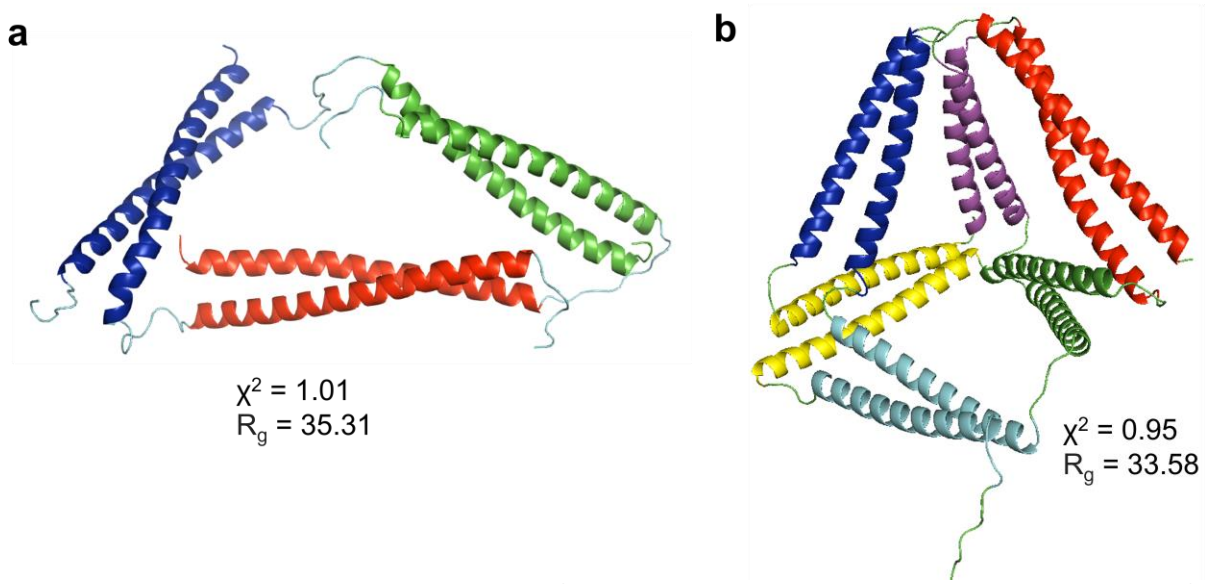

Figure 7: The best models with the lowest χ^2 from SAXS fitting. (a) A triangle model with χ^2 of 1.01, (b) A tetrahedron model with χ^2 of 0.95.

3. Prediction of R_g and Evaluation of CCPO Nanostructures

We calculated the R_g of each model using FoXS, and the average R_g of 100 models of the triangle or tetrahedron varying the cycles of MD optimization with simulated annealing was plotted (Figure 8a and b). The results indicate that, after 5 cycles, the average of the calculated R_g of the triangle became close to the experimentally determined value of 36.0 ± 2.5 Å[15] (< 1% in error). For the tetrahedron, a similar trend was observed that 5 or more cycles of simulated annealing resulted in the average R_g of 100 models from the experimentally determined value of 34 Å (< 3% in error)[9]. The average R_g values computed from the models refined by 15 cycles of MD optimization were 36.0 ± 3.1 Å (triangle) and 33.2 ± 1.1 Å (tetrahedron), respectively. Hence, the R_g of a CCPO topology is predictable using this method of comparative modeling performed with extended cycles of MD optimization. To further investigate how R_g is predicted, the R_g and χ^2 of 1100 models were plotted (Figure 8c, 8d). The results show that R_g values of the high-quality models with the relatively low values of χ^2 (close to 1) are around the average value. Hence, calculation

of the average R_g values of the models from extended MD optimization (~15 cycles) enables the prediction of accurate R_g of a CCPO nanostructure.

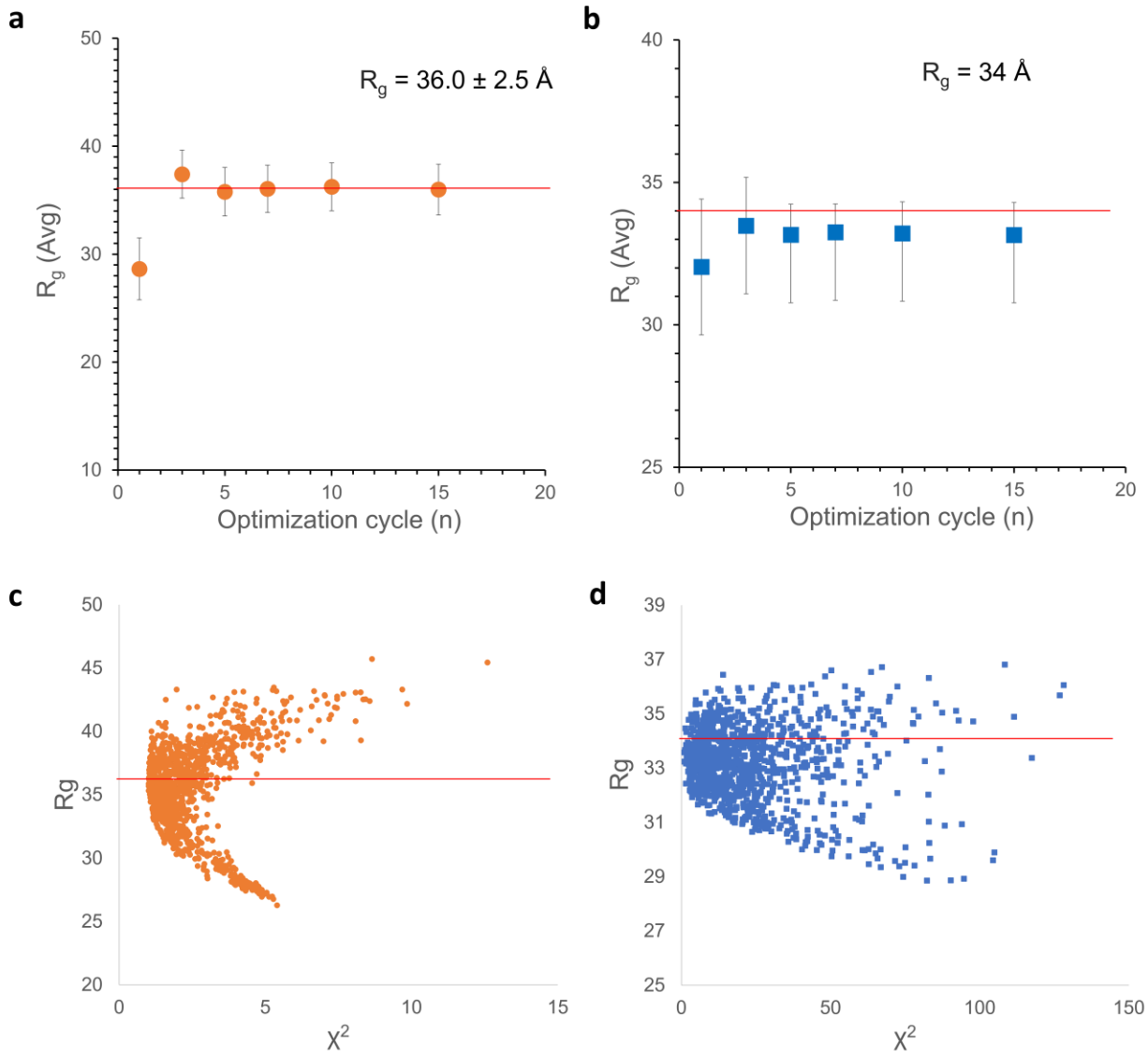

Figure 8: Prediction of R_g. (a, b) Average R_g for the triangle (a) and for the tetrahedron at varied cycles of MD optimization. (c, d) R_g values plotted against χ^2 of 1100 triangle (c) and tetrahedron models (d). The redlines indicate the R_g values from SAXS profiles.

The prediction of R_g provides a new method of model evaluation for CCPO nanostructures. In comparative modeling, DOPE or z-DOPE score is considered as the best performing function

to evaluate models and indicate how close a model is to its native state[28]. The lower DOPE score indicates the better model. However, z-DOPE (or DOPE) scores from 1100 comparative models for either the triangle or tetrahedron did not show good correlations with χ^2 (Figure 9a and b). The Pearson correlation coefficient (r), which measures the strength of a linear relationship between two variables, was 0.354 for the triangle models, and the tetrahedron models showed a correlation with a coefficient of 0.309. As a new model evaluation method, we used the RAE in R_g, which showed improved correlations with χ^2 (Figure 9c and d). In the correlation of χ^2 and RAE in R_g of the triangle models, the Pearson correlation coefficient was 0.625, which was significantly higher than the coefficient with z-DOPE. Also, RAE in R_g, of the tetrahedron models showed a correlation of χ^2 with a coefficient of 0.467. The R_g values of the CCOP nanostructures used for the RAE calculation were predicted without SAXS data (Figure 8).

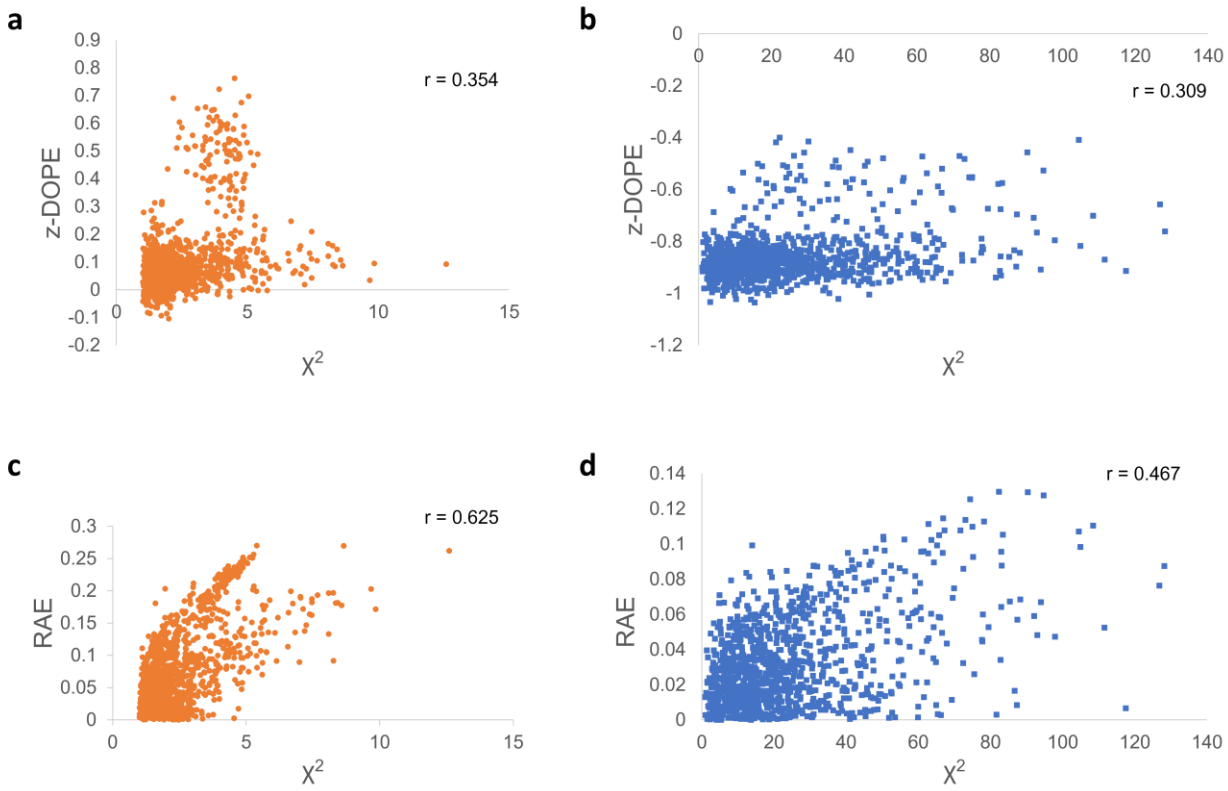

Figure 9**:** Model evaluation and agreement with SAXS data. (a, b) z-DOPE scores plotted against χ^2 values of triangle (a) and tetrahedron models (b). (c, d) RAE plotted against χ^2 values of triangle (c) and tetrahedron models (d). The Pearson correlation coefficient (r) is indicated in each panel.

Discussion

The atomistic models of CCPO nanostructures can be easily built by comparative modeling because of the availability of coiled-coil templates. However, the linkers placed between coiled-coil modules are flexible and lead to significant variations in the overall shape and conformation of the resulting models. In this study, we focused on the degree and length of MD optimization with simulated annealing for improved structure refinement. The results showed that at least 10 to 15 cycles of MD optimization in comparative modeling are needed for obtaining high-quality models. A computational design flatform for the single-chain CCPO, CoCoPOD[9], provides a convenient tool to build atomistic models that can be fit to SAXS data. However, the information on MD optimization is not known enough, which may limit its use for a new design of CCPO. Moreover, the platform has been developed only for CCOP structures that employ a single coiled-coil fusion protein chain and does not provide reliable evaluation scores but DOPE. In this work, structural refinement through extended cycles of MD optimization using MODELLER was demonstrated. The presented study provides useful information for a broad range of coiled-coil-based structure modeling and various kinds of CCPO methods. It also demonstrated the generation of new models of the triangle and tetrahedron that are nearly complete matches to the SAXS data. The new models from this study provide information that potentially sheds light on the fundamentals of CCPO.

Most importantly, we demonstrated that comparative modeling with extended MD optimization can be used to predict R_g of CCPO designs without SAXS data. Our results fall within 3% of errors from experimentally determined values. Using the predicted R_g, we calculated RAE in R_g and introduced it as a new evaluation tool. Its correlation with χ^2 was significantly improved than with z-DOPE scores. This might be because the CCPO structures are artificial and distinct

from native states of many natural proteins which DOPE was developed from. This new evaluation gives us an advantage of predicting “good” models, which will enable a fast screen of promising design candidates of a new CCPO nanostructure.

Conclusion

A comprehensive study on the degree and length of MD optimization for comparative modeling of CCPO nanostructures was conducted. The effective optimization on structure refinement was described, which is informative for comparative modeling for a broad range of CCPO designs. New atomistic models of CCPO triangle and tetrahedron with high quality of SAXS data fitting were obtained, which demonstrates improved accuracy of SAXS characterization assisted by comparative modeling. The extended MD optimization enabled the prediction of R_g and resulted in the introduction of a new method for model evaluation using the relative errors in R_g, which was much more effective than z-DOPE score. This is a potentially useful tool for the prediction of structure models and screening design candidates of a new CCPO nanostructure.

Bibliography

1 P. W. K. Rothemund, *Nature*, 2006, **440**, 297–302.
2 A. Kuzuya and Y. Ohya, *Acc. Chem. Res.*, 2014, **47**, 1742–1749.
3 A. Kuzyk, R. Jungmann, G. P. Acuna and N. Liu, *ACS Photonics*, 2018, **5**, 1151–1163.
4 Jungmann, Ralf, et al. "Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami." *Nano letters* 10.11 (2010): 4756-4761..
5 C. A. Mirkin, R. L. Letsinger, R. C. Mucic and J. J. Storhoff, *Nature*, 1996, **382**, 607–609.
6 A. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez and P. G. Schultz, *Nature*, 1996, **382**, 609–611.
7 F. Lapenta, J. Aupič, Ž. Strmšek and R. Jerala, *Chem. Soc. Rev.*, 2018, **47**, 3530–3542.
8 Park, Won Min. "Coiled-coils: The molecular zippers that self-assemble protein nanostructures." *International Journal of Molecular Sciences* 21.10 (2020): 3584.
9 A. Ljubetič, F. Lapenta, H. Gradišar, I. Drobnak, J. Aupič, Ž. Strmšek, D. Lainšček, I. Hafner-Bratkovič, A. Majerle, N. Krivec, M. Benčina, T. Pisanski, T. Ć. Veličković, A. Round, J. M. Carazo, R. Melero and R. Jerala, *Nature Biotechnology*, 2017, **35**, 1094–1101.
10 S. Božič Abram, H. Gradišar, J. Aupič, A. R. Round and R. Jerala, *ACS Chem. Biol.*, 2021, **16**, 310–315.
11 K. H. Choi and M. Morais, *Methods Mol Biol*, 2014, **1138**, 241–252.
12 D. Schneidman-Duhovny, M. Hammel, J. A. Tainer and A. Sali, *Nucleic Acids Res*, 2016, **44**, W424-429.
13 D. Schneidman-Duhovny, M. Hammel, J. A. Tainer and A. Sali, *Biophys J*, 2013, **105**, 962–974.
14 A. Sali and T. L. Blundell, *J Mol Biol*, 1993, **234**, 779–815.
15 W. M. Park, M. Bedewy, K. K. Berggren and A. E. Keating, *Scientific Reports*, 2017, **7**, 10577.
16 A. Rose and I. Meier, *Cellular and molecular life sciences: CMLS*, 2004, **61**, 1996–2009.
17 Y. Wu and J. H. Collier, *Wiley Interdiscip Rev Nanomed Nanobiotechnol*, , DOI:10.1002/wnan.1424.
18 P. Burkhard, J. Stetefeld and S. V. Strelkov, *Trends in Cell Biology*, 2001, **11**, 82–88.
19 K. Suzuki, T. Doi, T. Imanishi, T. Kodama and T. Tanaka, *Biochemistry*, 1997, **36**, 15140–15146.
20 L. Szilák, J. Moitra and C. Vinson, *Protein Sci*, 1997, **6**, 1273–1283.
21 H. Gradišar, S. Božič, T. Doles, D. Vengust, I. Hafner-Bratkovič, A. Mertelj, B. Webb, A. Šali, S. Klavžar and R. Jerala, *Nature Chemical Biology*, 2013, **9**, 362–366.
22 F. Lapenta, J. Aupič, M. Vezzoli, Ž. Strmšek, S. Da Vela, D. I. Svergun, J. M. Carazo, R. Melero and R. Jerala, *Nat Commun*, 2021, **12**, 939.
23 J. Aupič, Ž. Strmšek, F. Lapenta, D. Pahovnik, T. Pisanski, I. Drobnak, A. Ljubetič and R. Jerala, *Nature Communications*, 2021, **12**, 940.
24 A. W. Reinke, R. A. Grant and A. E. Keating, *J Am Chem Soc*, 2010, **132**, 6025–6031.
25 Boyle, Aimee L., et al. "Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design." *Journal of the American Chemical Society* 134.37 (2012): 15457-15467.
26 B. Webb and A. Sali, in *Functional Genomics: Methods and Protocols*, eds. M. Kaufmann, C. Klinger and A. Savelsbergh, Springer, New York, NY, 2017, pp. 39–54.
27 S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Science*, 1983, **220**, 671–680.
28 M. Shen and A. Sali, *Protein Sci*, 2006, **15**, 2507–2524.
29 B. Webb and A. Sali, *Current Protocols in Bioinformatics*, 2014, **47**, 5.6.1-5.6.32.
30 K. E. Thompson, C. J. Bashor, W. A. Lim and A. E. Keating, *ACS Synth. Biol.*, 2012, **1**, 118–129.
31 D. Eramian, N. Eswar, M.-Y. Shen and A. Sali, *Protein Sci*, 2008, **17**, 1881–1893.