

Electronic Supplementary Information (ESI)

## Coiled-Coil Protein Origami Structure Modeling for Improved Characterization and Prediction

Ratnakshi Mandal,<sup>a</sup> Mostafa Bedewy,<sup>b</sup> and Won Min Park<sup>a\*</sup>

<sup>a</sup>. Tim Taylor Department of Chemical Engineering, Kansas State University, 1701A Platt Street, Manhattan, Kansas 66506, USA.

<sup>b</sup>. Department of Industrial Engineering, University of Pittsburgh, 3700 O'Hara Street, Pittsburgh, Pennsylvania 15261, USA.

\*Correspondence to: [wmpark@ksu.edu](mailto:wmpark@ksu.edu)

**Triangle (a heterotrimer of SYNZIP1-SYNZIP6, SYNZIP5-SYNZIP4, and SYNZIP3-SYNZIP2)**

```

1      10      20      30      40      50      60      70      80      90      100
SYNZIP1-SYNZIP6  MNLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEGGGGSGGGGSKENAKLENIVARLENDNANLEKDIANLEKDIANLERDVALEHHHHHH
SYNZIP1:         -NLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIE-----KENAKLENIVARLENDNANLEKDIANLEKDIANLERDVA-----
SYNZIP6:         -----

1      10      20      30      40      50      60      70      80      90      100
SYNZIP5-SYNZIP4  MNTVKELKNYIQELEERNAELKNLKEHLKFAKAELEFELAAHKFEGGGSGGGGSNRNEQLKNKVEELKNNRNAYLNKELATLENEVARLENDVALEHHHHHHH
SYNZIP5:         -NTVKELKNYIQELEERNAELKNLKEHLKFAKAELEFELAAHKFE-----NRNEQLKNKVEELKNNRNAYLNKELATLENEVARLENDVAE-----
SYNZIP4:         -----

1      10      20      30      40      50      60      70      80      90      100
SYNZIP3-SYNZIP2  MNEVTTLENDAAFLENENAYLEKEIARLRKEKAALRNRLAHKGGGGSGGGGSARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVALEHHHHHH
SYNZIP3:         -NEVTTLENDAAFLENENAYLEKEIARLRKEKAALRNRLAHKK-----ARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVA-----
SYNZIP2:         -----

```

**Tetrahedron (TET12SN)**

```

1      10      20      30      40      50      60      70      80      90      100
TET12SN          MLEEEELKQLEEEELQAIIEEQLAQLQWKAQARKEKLAQLKEKLSGPGSPEDIQQLEEEISQLEQKNSLKEKNQELKYGGPGDIEQELERAKESIRRLEQ
APH:APH          -LEEEELKQLEEEELQAIIEEQLAQLQWKAQARKEKLAQLKEKL-----SPEDEIQQLEEEISQLEQKNSLKEKNQELKY-----DIEQELERAKESIRRLEQ
P3:P4            -----
BCR:BCR          -----
GCNh:GCNh       -----
P7:P8            -----
P5:P6            -----

101     110     120     130     140     150     160     170     180     190     200
TET12SN          EVNQERSRMQYLQTLLEKSGPGQLEDKVEELLKSNYHLENEVERLKKLVGSGPGLLEEELKQLEEEELQAIIEEQLAQLQWKAQARKEKLAQLKEKLSGPGSP
APH:APH          -LEEEELKQLEEEELQAIIEEQLAQLQWKAQARKEKLAQLKEKL-----LEEEELKQLEEEELQAIIEEQLAQLQWKAQARKEKLAQLKEKL-----
P3:P4            -----
BCR:BCR          EVNQERSRMQYLQTLLEK-----QLEDKVEELLKSNYHLENEVERLKKLV-----
GCNh:GCNh       -----
P7:P8            -----
P5:P6            -----

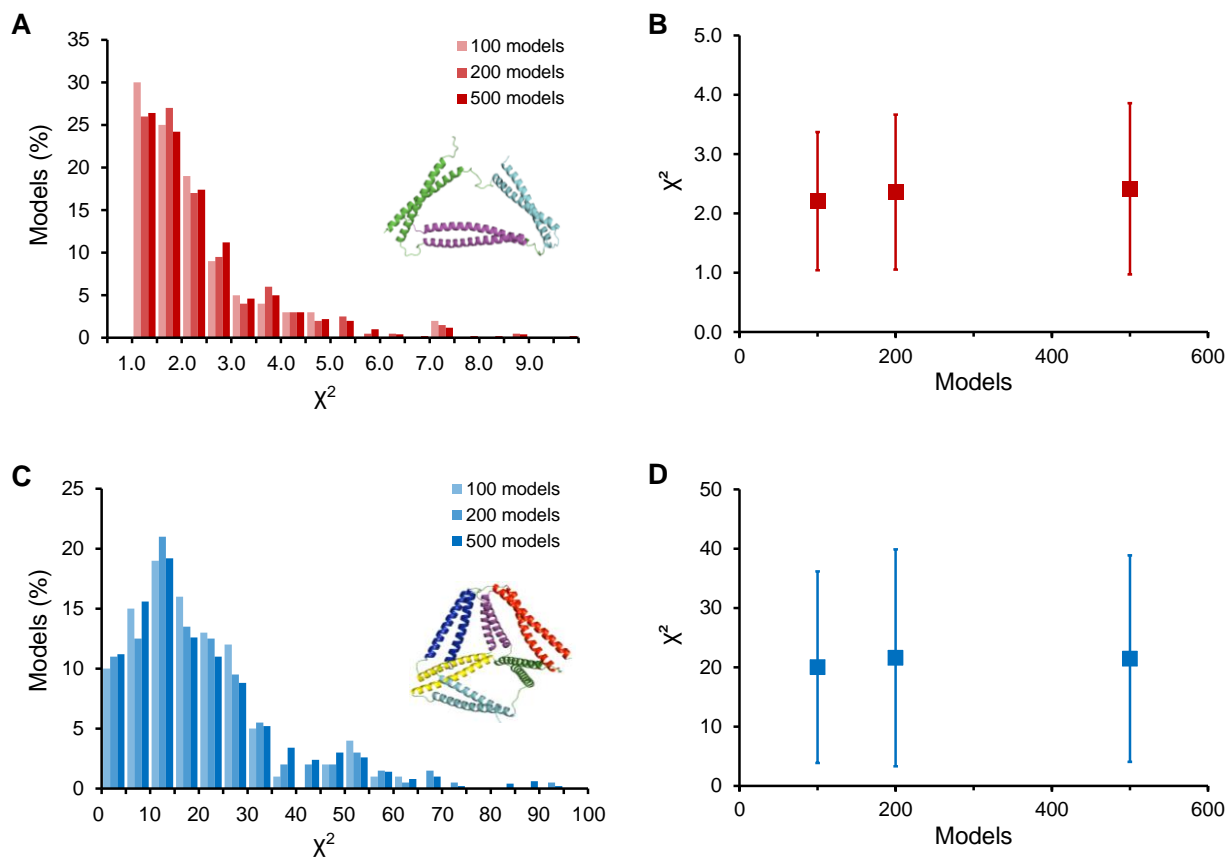
201     210     220     230     240     250     260     270     280     290     300
TET12SN          EDEIQQLEEKNSQLKQETISQLEEKQELKYGGPGQLEDKVEELLKSNYHLENEVERLKKLVGSGPGSPEDKISQLKEIKQQLKQENQQLKEEENSQLEYG
APH:APH          -----SPEDKISQLKEIKQQLKQENQQLKEEENSQLEY-----
P3:P4            -----
BCR:BCR          -----QLEDKVEELLKSNYHLENEVERLKKLV-----
GCNh:GCNh       -----
P7:P8            EDEIQQLEEKNSQLKQETISQLEEKQELKY-----
P5:P6            -----

301     310     320     330     340     350     360     370     380     390     400
TET12SN          SGPGSPEDENSQLEEKISQLKQKNSLKEEIQLEYGSGPGSPEDKISELKEENQQLKQIQQLEENSQLEYGSGPGDIEQELERAKESIRRLEQEVNQ
APH:APH          -----DIEQELERAKESIRRLEQEVNQ-----
P3:P4            -----
BCR:BCR          -----SPEDKISELKEENQQLKQIQQLEENSQLEY-----
GCNh:GCNh       -----
P7:P8            ---SPEDENSQLEEKISQLKQKNSLKEEIQLEY-----
P5:P6            -----

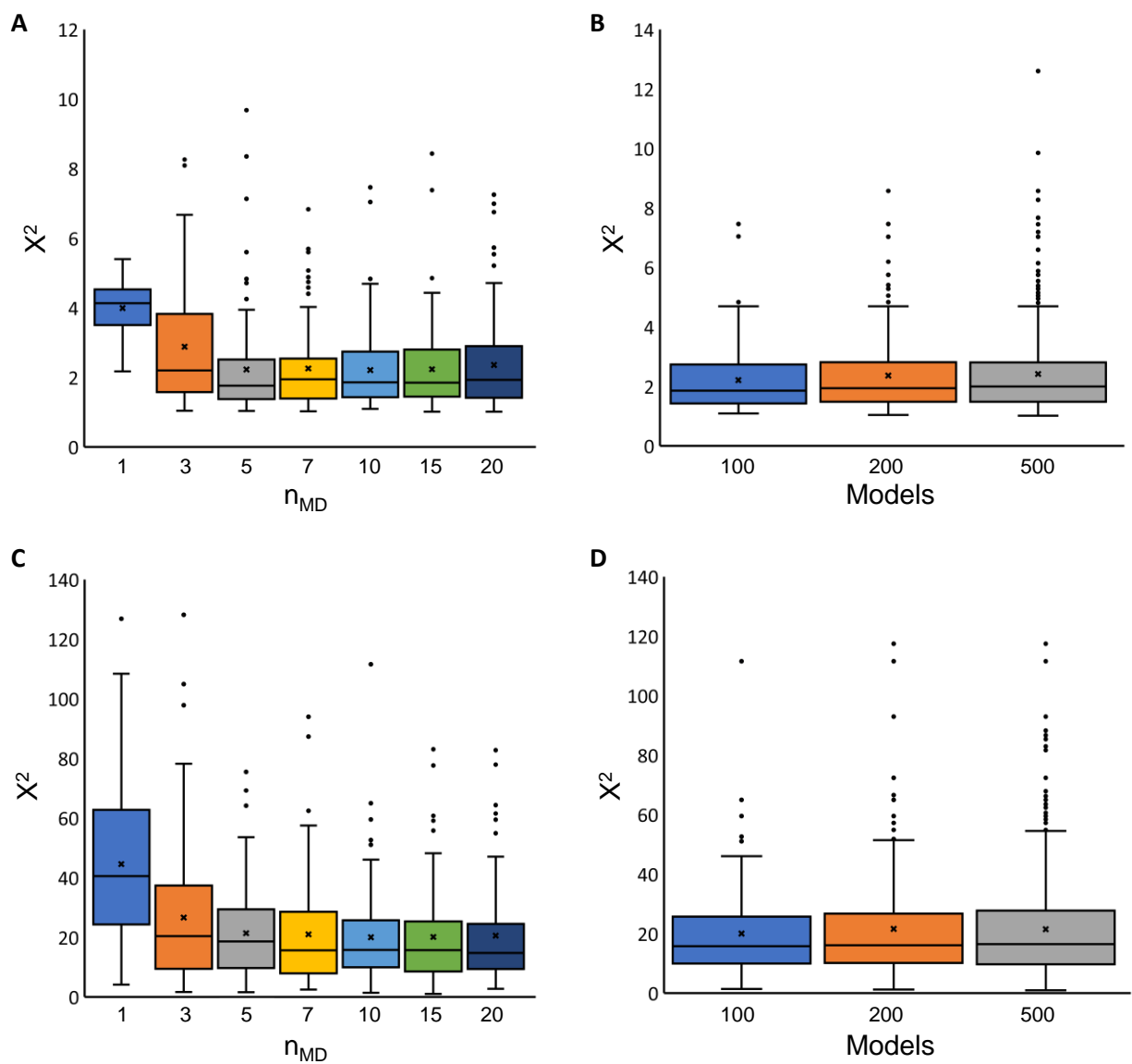
401     410     420     430     440     450     460
TET12SN          ERSRMQYLQTLLEKSGPGSPEDKNSLKEEIQLEEEENQQLKEEKISELKYGLEHHHHHHHH
APH:APH          -----
P3:P4            -----
BCR:BCR          ERSRMQYLQTLLEK-----
GCNh:GCNh       -----
P7:P8            -----
P5:P6            -----SPEDKNSLKEEIQLEEEENQQLKEEKISELKY-----

```

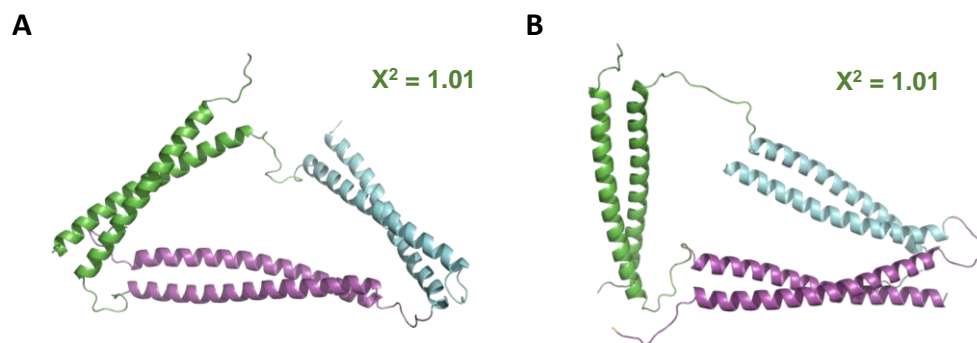
**Fig. S1** The sequence alignment for the comparative modeling of the triangle (a heterotrimer of SYNZIP1-SYNZIP6, SYNZIP5-SYNZIP4, and SYNZIP3-SYNZIP2, top) and the tetrahedron (TET12SN, bottom).



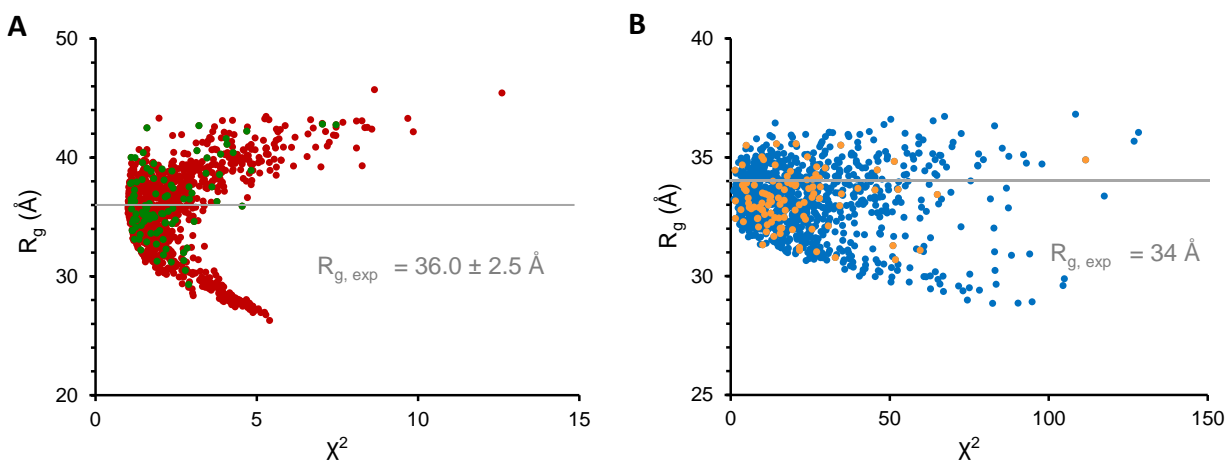
**Fig. S2** The quality of SAXS fitting of 100, 200, and 500 triangle (**A**, **B**) and tetrahedron models (**C**, **D**). **A**. Distribution of  $\chi^2$  values of the triangle models. **B**. Average of  $\chi^2$  values of the triangle models. **C**. Distribution of  $\chi^2$  values of the tetrahedron models. **D**. Average of  $\chi^2$  values of the tetrahedron models. All models were refined through 10 cycles of MD optimization. The error bars represent standard deviations.



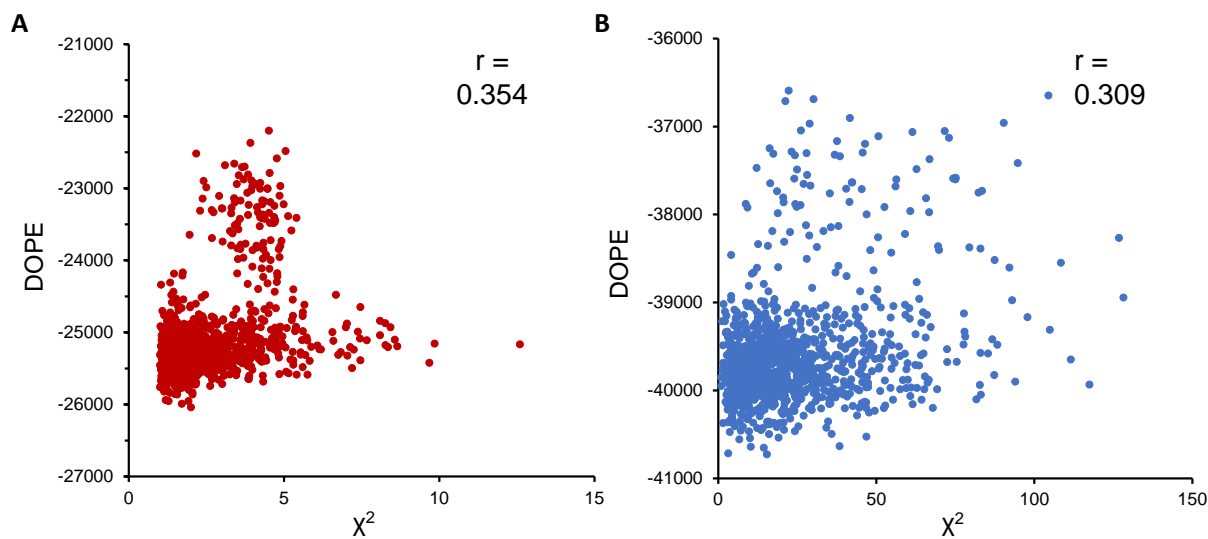
**Fig. S3** The distribution of the SAXS fitting data. **A, B.** Distribution of  $\chi^2$  values of the triangle models refined at  $n_{MD} = 1, 3, 5, 7, 10, 15,$  and  $20$  (**A**) and of  $100, 200,$  and  $500$  models refined at  $n_{MD} = 10$  (**B**). **C, D.** Distribution of  $\chi^2$  values of the tetrahedron models refined at  $n_{MD} = 1, 3, 5, 7, 10, 15,$  and  $20$  (**C**) and of  $100, 200,$  and  $500$  models refined at  $n_{MD} = 10$  (**D**). The whisker represents the highest and lowest values, and the box represents the upper and lower quartiles and the median. The average (x) and outliers (•) values are also indicated.



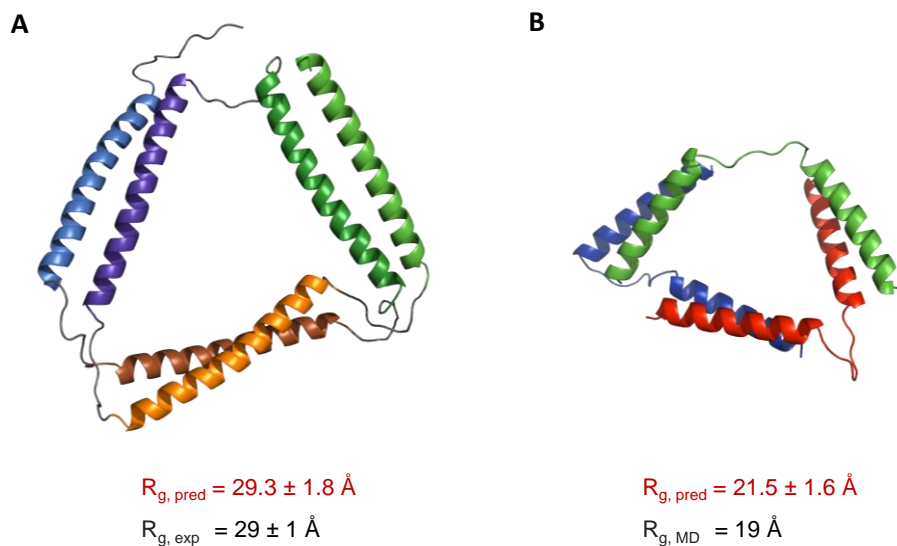
**Fig. S4** The triangle models with a  $\chi^2$  value of 1.01. **A.** The best-quality model refined at  $n_{MD} = 15$ . **B.** The best-quality model refined at  $n_{MD} = 20$ .



**Fig. S5** The  $R_g$  values plotted against  $\chi^2$  of comparative models. **A.** Total 1100 triangle models refined through MD optimization ( $n_{MD}$ ) ranging from 1 to 15 cycles (red) and the 100 models refined at  $n_{MD} = 10$  (green). **B.** Total 1100 tetrahedron models refined at  $n_{MD}$  ranging 1 to 15 (blue) and the 100 models refined at  $n_{MD} = 10$  (yellow). The grey lines indicate the experimentally determined values from the SAXS data ( $R_{g, \text{exp}}$ ).



**Fig. S6** The unnormalized DOPE score plotted against the  $\chi^2$  values of the triangle (A) and tetrahedron models (B). Total 1100 models refined at  $n_{MD}$  ranging from 1 to 15 were plotted in each panel and the Pearson correlation coefficient,  $r$ , is indicated.



**Fig. S7** The structure models of CCPO triangles with different linkers. (A) A model of the TRI6 triangle<sup>1</sup> with a CORGE score of 0.00017. (B) A model of the DoNA-GN<sub>4</sub> triangle<sup>2</sup> with a CORGE score of 0.0137. The predicted  $R_g$  values ( $R_{g, pred}$ ) are compared with the experimentally determined value ( $R_{g, exp}$ <sup>1</sup>) or the value from the all-atom MD simulation ( $R_{g, MD}$ <sup>2</sup>).

## Supplementary References

- 1 S. Božič Abram, H. Gradišar, J. Aupič, A. R. Round and R. Jerala, *ACS Chem. Biol.*, 2021, **16**, 310–315.
- 2 A. L. Boyle, E. H. C. Bromley, G. J. Bartlett, R. B. Sessions, T. H. Sharp, C. L. Williams, P. M. G. Curmi, N. R. Forde, H. Linke and D. N. Woolfson, *J. Am. Chem. Soc.*, 2012, **134**, 15457–15467.