Supplementary Information

Over the last decade, the Baker Lab has developed an array of *de novo* designed helical repeat proteins (DHRs) and designed helical bundles (HBs)[1,2]. Due to their exceptional stability and accessible termini, DHRs and HBs can be combined to generate larger molecular assemblies, which can then be sequence-optimized using the RosettaDesign[3] platform. A major goal of these efforts is the design of self-assembling oligomers that could serve as starting points for novel drug delivery particles, vaccine platforms, and other nanomaterials. In addition to advancing progress toward these specific applications, structural characterization of *de novo* designed proteins allows evaluation of the effectiveness of various design strategies and is a critical part of the iterative computational-experimental workflow. A very large set of *de novo* designed proteins also gives us the unique opportunity to systematically and thoroughly optimize the cryo-EM workflow for these kinds of proteins, from sample preparation to model building.

The Bhabha / Ekiert Labs have focused on high-resolution structural characterization of these *de novo* designed proteins. Initial characterization was done using X-ray crystallography to solve structures of small DHRs[2]. These smaller building blocks are now being combined into larger oligomeric designs that are of sufficient size for structure determination by cryo-EM. Designed proteins are expressed and purified at the Institute of Protein Design at UW and analyzed by SDS-PAGE, SEC, and SEC-MALS, SAXS or nMS to assess homogeneity and intact mass of oligomeric assemblies. Proteins that exhibit a homogeneous population near the expected mass for the oligomeric design are shipped to the Bhabha/Ekiert group at NYU for structural characterization.

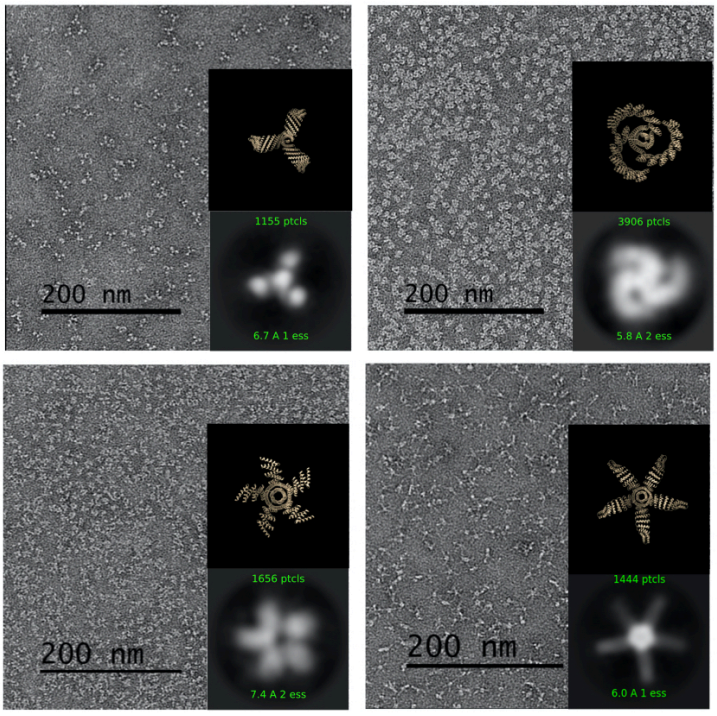

Figure 1. Negative stain EM of *de novo* designed trimers and pentamers. 2D class averages are shown below design models as insets on representative micrographs.

Upon receipt at NYU, we assess sample homogeneity and overall shape by negative stain EM (Fig. 1). Designs that appear monodisperse and have 2D averages that match the expected design are identified as candidates for cryo-EM. We have found that many of these *de novo* designed proteins exhibit one or more challenges that must be addressed to enable high-resolution cryo-EM, including: particles that are absent from the vitreous ice layer or aggregated, heterogeneity caused by particle damage or oligomer dissociation upon freezing, and preferred orientation of particles. In the latter case, we have improved the quality of 3D reconstruction by collecting tilted data at NYSBC to improve angular coverage of particles. We have also found grids with continuous coatings of graphene oxide or thin carbon to yield dramatic improvements in particle quality and distribution (Fig. 2). In the case of one design, C4-79, graphene oxide grids

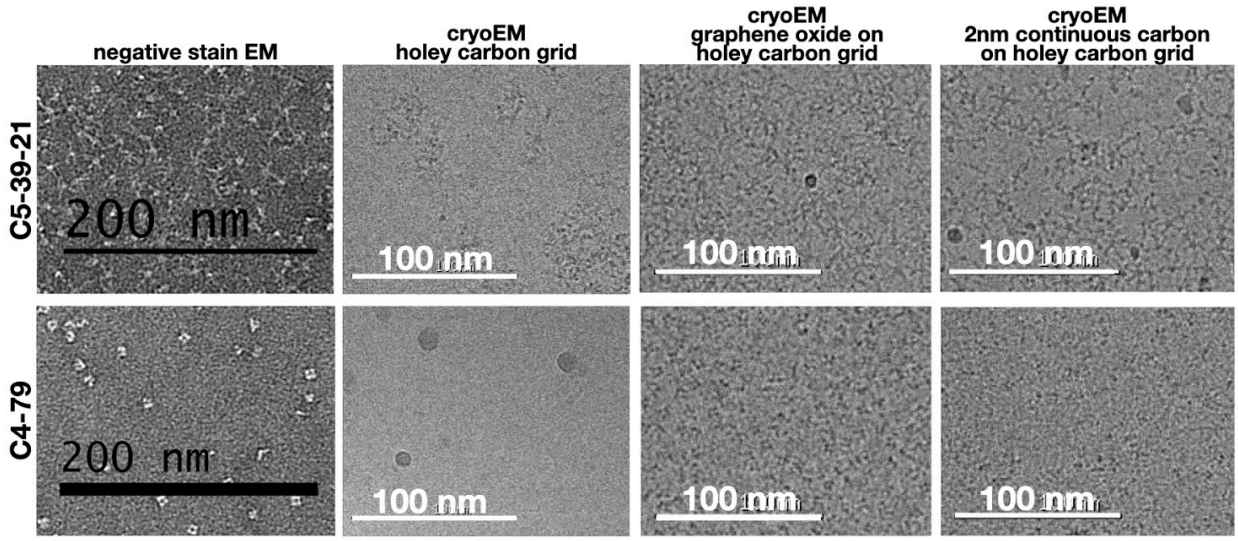

Figure 2. CryoEM sample prep optimization for pentameric and tetrameric designs C5-39-21 and C4-79.

enabled a cryo-EM map to be generated at 3.7 Å global resolution, whereas on uncoated grids, particles were nearly absent from the ice layer (Figs. 2, 3).

For each *de novo* oligomeric design, we will prepare samples for cryo-EM on holey carbon grids (uncoated as well as coated with graphene oxide and 2 nm carbon) using a Vitrobot Mark IV at the NYU Cryo-EM Facility. At NYU we also have access to Talos Arctica and Titan Krios microscopes for screening grids and collecting data. The high volume of designs generated by the Baker lab, combined with the high number of users of the NYU Cryo-EM facility, motivate our request for Krios time at NCCAT. We would also like to use the BAG to access the Chameleon vitrification system to optimize cryo-EM grid preparation for high-priority designs where none of the above grid preparation strategies yield acceptable particle distribution.

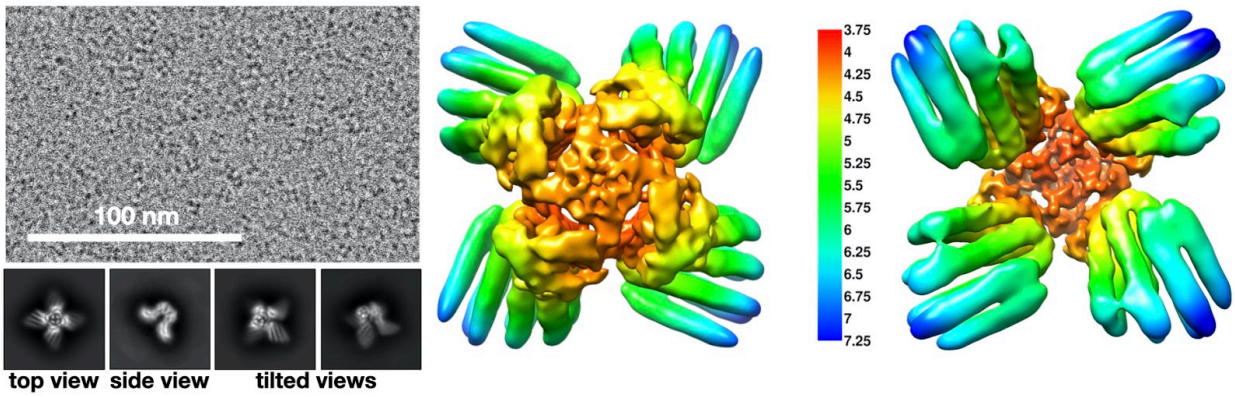

Figure 3. Cryo-EM data and reconstruction for C4-79. Left, representative motion-corrected micrograph with representative 2D classes shown below. Right, locally-filtered cryo-EM map colored by local resolution.

1. Lu, P. *et al.* Accurate computational design of multipass transmembrane proteins. *Science* **359**, 1042–1046 (2018).
2. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
3. Liu, Y. & Kuhlman, B. RosettaDesign server for protein design. *Nucleic Acids Research* vol. 34 W235–W238 (2006).