


## Design of functionalised circular tandem repeat proteins with longer repeat topologies and enhanced subunit contact surfaces

Jazmine P. Hallinan<sup>1,4</sup>, Lindsey A. Doyle <sup>1,4</sup>, Betty W. Shen<sup>1</sup>, Mesfin M. Gewe<sup>2</sup>, Brittany Takushi<sup>1</sup>, Madison A. Kennedy<sup>1</sup>, Della Friend<sup>1</sup>, James M. Roberts<sup>2</sup>, Philip Bradley<sup>1,3</sup>  & Barry L. Stoddard <sup>1</sup> 

Circular tandem repeat proteins ('cTRPs') are de novo designed protein scaffolds (in this and prior studies, based on antiparallel two-helix bundles) that contain repeated protein sequences and structural motifs and form closed circular structures. They can display significant stability and solubility, a wide range of sizes, and are useful as protein display particles for biotechnology applications. However, cTRPs also demonstrate inefficient self-assembly from smaller subunits. In this study, we describe a new generation of cTRPs, with longer repeats and increased interaction surfaces, which enhanced the self-assembly of two significantly different sizes of homotrimeric constructs. Finally, we demonstrated functionalization of these constructs with (1) a hexameric array of peptide-binding SH2 domains, and (2) a trimeric array of anti-SARS CoV-2 VHH domains. The latter proved capable of sub-nanomolar binding affinities towards the viral receptor binding domain and potent viral neutralization function.

<sup>1</sup>Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. North, Seattle, WA 98109, USA. <sup>2</sup>Lumen Bioscience Inc., 1441 North 34th Street, Seattle, WA 98103, USA. <sup>3</sup>Division of Public Health Sciences and Program in Computational Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98009, USA. <sup>4</sup>These authors contributed equally: Jazmine P. Hallinan, Lindsey A. Doyle. email: [pbradley@fredhutch.org](mailto:pbradley@fredhutch.org); [bstoddard@fredhutch.org](mailto:bstoddard@fredhutch.org)

Protein engineering has enabled the creation of novel protein folds and assemblages, as well as the modification of existing protein molecules<sup>1</sup>. The discipline of protein engineering has been greatly facilitated by at least three significant recent advances: the development of accurate computational algorithms for de novo structural design, an enhanced ability to construct and screen large protein libraries, and increasingly efficient and affordable deep sequencing approaches. Whether they are created via de novo computational design or laboratory selections (or a combination of those two approaches), engineered proteins are now being used for a wide variety of biotechnology applications, including those that can benefit from new ligand binding proteins<sup>2–4</sup>, novel immunological regulators<sup>5</sup> or targeted protein therapeutics<sup>6,7</sup>.

Tandem repeat proteins (‘TRPs’) contain modular units of repeated protein sequence and structure, that can be composed of a variety of structural motifs including  $\alpha$ -helical bundles,  $\beta$ -sheets, or mixed topologies<sup>8–11</sup>. They are particularly amenable to de novo design via purely computational approaches, due to their highly modular architectures<sup>8,9</sup>. Multiple types of computationally designed tandem repeat proteins (using either naturally existing or computationally designed motifs and topologies) have been described that form either extended linear shapes (in which the N- and C-termini are at opposite ends of the protein chain)<sup>12</sup> or closed circular shapes (in which the N- and C-termini contact with one another and thereby complete a closed protein ring)<sup>5,13,14</sup>. Engineered TRPs display many of the same properties as naturally evolved repeat proteins, including relatively few constraints on their length and size, significant flexibility that allows them to sample a wide variety of curvature and shapes, and a propensity to exhibit high thermostability and solubility.

We have previously described the creation of an array of circular tandem repeat proteins (‘cTRPs’) that are constructed from repeated two-helix bundles and that display a wide range of sizes and symmetries<sup>5,13</sup>. The largest of these constructs was used to create symmetric protein nanoparticles that can incorporate a variety of functional protein via covalent attachment around their periphery<sup>5</sup>. cTRPs that display multiple binding domains can display significant avidity effects (due to multimeric display of those domains around the molecular surface) and have been used to create novel formulations of immunological stimulatory and signaling molecules that can be used for biotechnology applications such as therapeutic T-cell manufacture.

While these cTRP constructs have many favorable properties, we have previously noted that their ability to self-assemble from smaller subunits is compromised by relatively limited contacts and small surface areas that are involved in packing between repeats and subunits. For example, a designed cTRP that was intended to form a homotrimer (corresponding to noncovalently assembled cTRP composed of three subunits that contain three repeats each) was instead found to inefficiently form a tetramer, indicating that the structural contacts between repeats and subunits was inadequate to enforce a desired stoichiometric assembly pattern<sup>13</sup>. Similarly, a series of cTRP subunits designed to self-assemble into a 24-repeat assemblage failed to assemble altogether, indicating that the structural contacts between repeats and subunits were inadequate to drive efficient assembly, a deficiency that was addressed through the incorporation of disulfide ‘staples’ across each subunit interface<sup>5</sup>.

Based on those results, we hypothesized that a new generation of computationally designed cTRP proteins with increased repeat size (corresponding to longer secondary structural elements that increase both particle thickness and the number of buried contacts between repeats and protein subunits) would improve both the energetics and the stoichiometric control of self-association. In this study, we describe the de novo design of thick cTRPs

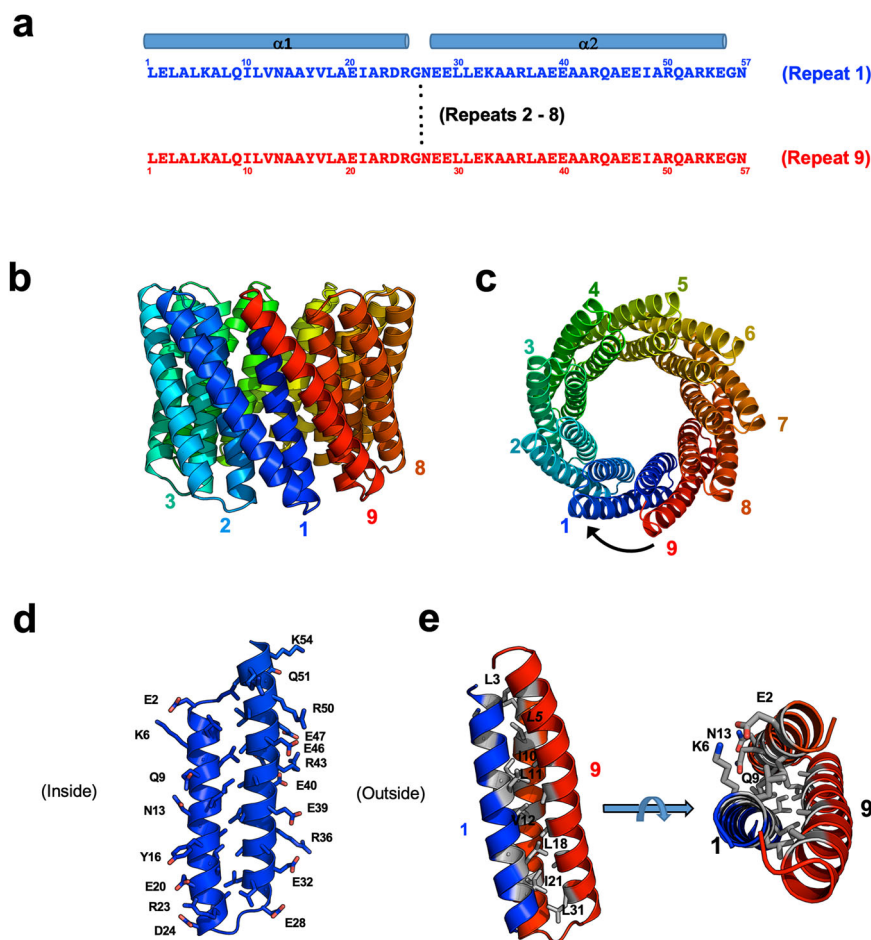
(‘tcTRPs’) with repeats composed of longer helical bundles that form a new underlying repeat topology and demonstrate their solution and assembly behaviors, their structural features via X-ray crystallography and CryoEM, and their ability to be functionalized with additional folded protein domains.

## Results

**Design, expression, and solution behavior of tcTRP9 and tcTRP24.** We previously developed an approach to geometry-guided repeat protein design that was described in<sup>13,15,16</sup> and is implemented in the Rosetta molecular modeling package<sup>17,18</sup>. Key features include symmetry of backbone and side-chain conformations extended across all repeats (allowing computational complexity to scale with repeat length rather than protein length); a pseudo-energy term that favors the desired inter-repeat geometry; clustering and resampling stages that allow intensified exploration of promising topologies; and an in silico validation step that assesses sequence-structure compatibility by attempting to re-predict the designed structure given only the designed sequence. When applying this approach to the design of closed tandem repeat proteins with increased thickness (‘tcTRPs’ or ‘thick(er) circular tandem repeat proteins’), we identified designs with repeats corresponding to right-handed helical bundles as displaying favorable predicted folding energetics for their designed backbone topologies, and well-formed energy funnels when subjecting their sequences to unbiased fold predictions in *Rosetta*. We ultimately selected a small number of individual designs corresponding to two different tcTRP sizes and symmetries (containing 9 repeats and 24 repeats, respectively) to examine and validate using biophysical and structural approaches.

The smaller of the two designed tcTRPs (‘tcTRP9’; Fig. 1) was initially expressed and purified both as a monomeric construct containing 513 residues (MW = 56.6 kDa) composed of nine equivalent repeats of 57 residues each, and as a smaller protein chain (‘tcTRP9<sub>3</sub>’) containing three repeats (172 residues; MW = 18.9 kDa) that was intended to assemble into a trimeric tcTRP particle with the same architecture and dimensions as its single-chain parent (with an exterior diameter of ~60 Å and an interior pore ~15 Å across). The larger of the designed tcTRPs (‘tcTRP24’; Fig. 2) was generated solely from smaller subunits containing either six repeats (‘tcTRP24<sub>6</sub>’) or eight repeats (tcTRP24<sub>8</sub>) that were intended to assemble into tetrameric or trimeric tcTRP particles, each with an overall architecture and dimensions corresponding to the original design of a single-chain tcTRP containing 24 repeats in total (with an exterior diameter of ~100 Å and an interior pore ~ 60 Å across).

The differences in the design of repeat sequences, structures, and packing between the tcTRP9 and tcTRP24 models are illustrated in Fig. 3. The sequences of individual repeats from the two designs are similar to one another: each corresponds to a repeated 57 residue sequence that forms an anti-parallel two-helix bundle, with overall 37% sequence identity (50% similarity). The identities and similarities at individual residues between repeats in the two designed proteins are found both at surface-exposed positions and at interfacial positions that contribute to packing within individual repeats. The overall rmsd for all alpha carbons within comparable repeats in the two constructs is approximately 1.5 Å (dark and light blue, respectively in Fig. 3b–d), indicating that the backbone architecture of repeats from the two designs is quite similar. However, superposition of a single repeat from tcTRP9 and tcTRP24 and examination of the deviation in the positions of adjacent downstream repeats (red and salmon respectively in Fig. 3c, d) illustrates that the differences in the size and curvature between the two constructs are facilitated by



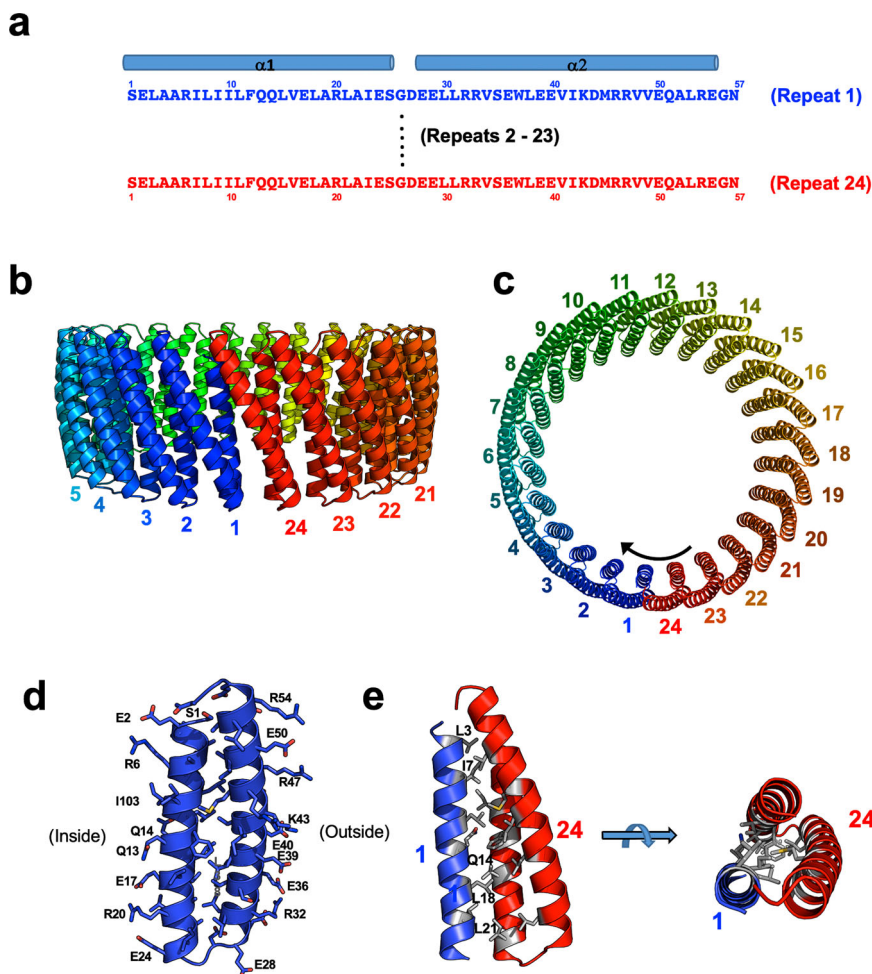
**Fig. 1 Design of tcTRP9.** **a** Sequence and secondary structure of individual repeats, which each consist of an anti-parallel two-helix bundle connected by short turns composed of ‘GN’ sequences. **b, c** Nine consecutive repeats (colored blue at the N-terminal repeat and progressing to red at the C-terminal repeat in all figure panels) form a closed cylindrical structure, with an exterior diameter of  $\sim 60$  Å and an interior pore diameter of  $\sim 15$  Å. **d** The structural composition of the two-helix bundle corresponding to each designed repeat places mostly charged side chains (largely glutamic acid and arginine residues) on the exterior of the particle, and a network of charged and hydrophilic residues (along with one aromatic tyrosine) on its interior surface. The interface between the two helices is largely composed of a series of leucine and alanine residues (not labeled). **e** The interface between individual repeats corresponds to a series of beta-branched (Ile and Val) and gamma-branched (Leu) hydrophobic residues that form a hydrophobic core between helices, flanked by hydrophilic residues extending towards the interior and exterior of the tcTRP particles.

altered packing angles between sequential repeats (resulting in differences in the positions of the immediately adjacent, downstream repeats of over 4 Å rmsd).

**Solution behavior and crystal structures of tcTRP9 and tcTRP9<sub>3</sub>.** To validate the accuracy of the tcTRP9 particle (harboring nine repeats), the recombinant protein was overexpressed and purified, its apparent molecular mass analyzed by size exclusion chromatography, its thermal stability verified by circular dichroism (CD) spectroscopy and its crystal structure determined. The protein was overexpressed and purified to homogeneity (Supplementary Fig. S1a) with a yield of approximately 4 mg per liter of LB culture media. The protein eluted from an SEC column in a single relatively uniform peak at a retention volume corresponding to the mass of the protein (56 kDa) (Supplementary Fig. S1b, left). CD spectra collected at 22° and 95° were nearly superimposable and indicated the presence and retention of the helical secondary structure after heating (Supplementary Fig. S1b, right). The thermal stability of these constructs was comparable to previously designed cTRP constructs<sup>5,13</sup>.

The crystal structure of the tcTRP9 particle (Table 1 and Fig. 4a–c), determined to 2.1 Å resolution, validated the accuracy of the design. Superposition of the design model and the experimental crystal structure (Fig. 4d–f) produced an rmsd value for all alpha-carbons of  $\sim 0.5$  Å, and close agreement between side-chain rotamers within the inter- and intrahelical repeats and interaction surfaces. The only positions where the side-chain rotameric conformation significantly differs between the design and the structure is for a pair of leucine residues (L11 and L37), which interact with one another through van der Waals interactions between their side chains in both models.

Because the original (thinner) ‘cTRP9’ construct (also harboring nine repeats; Supplementary Fig. S2a) was previously observed not to properly self-assemble from smaller subunits<sup>13</sup> (inefficiently forming an unexpected tetramer at high protein concentrations, instead of the intended trimer), we tested and compare the ability of the thicker ‘tcTRP9’ scaffold (Supplementary Fig. S2b) to self-assemble from smaller subunits. To accomplish this, we generated and purified a construct that again contained only three repeats per subunit (termed ‘tcTRP9<sub>3</sub>’). This construct was also expressed and purified at relatively high levels (Supplementary Fig. S1a; yield  $\sim 5$  mg per liter of culture) and



**Fig. 2 Design of tcTRP24.** **a** Sequence and secondary structure of individual repeats, which each consist of an anti-parallel two-helix bundle connected by short turns composed of ‘GD’ sequences. **b, c** Twenty-four consecutive repeats (colored blue at the N-terminal repeat and progressing to red at the C-terminal repeat in all figure panels) form a closed cylindrical structure, with an exterior diameter of  $\sim 100$  Å and an interior pore diameter of  $\sim 60$  Å. **d** The structural composition of the two-helix bundle corresponding to each designed repeat is chemically similar to that of the tcTRP9 design (Fig. 1), again placing mostly charged glutamic acid and arginine residues on the exterior of the particle, and a network of charged and hydrophilic residues on its interior surface. **e** The interface between individual repeats.

eluted from a final SEC column at a volume corresponding to the same approximate mass as the full-length protein (Supplementary Fig. S1b, left), indicating probable assembly into a higher-order trimeric species with 9 repeats total. The purified construct, also like the full-length protein, did not display significant unfolding behavior at  $95^\circ$  in a CD spectrum (Supplementary Fig. S1b, right). The purified construct from the final SEC column was crystallized and its structure was determined at  $3.2$  Å resolution (Table 1). This analysis confirmed a homotrimeric assembly that closely mirrored the dimensions, topology, and side-chain packing interactions found in the full-length tcTRP9 designed protein (Fig. 4g).

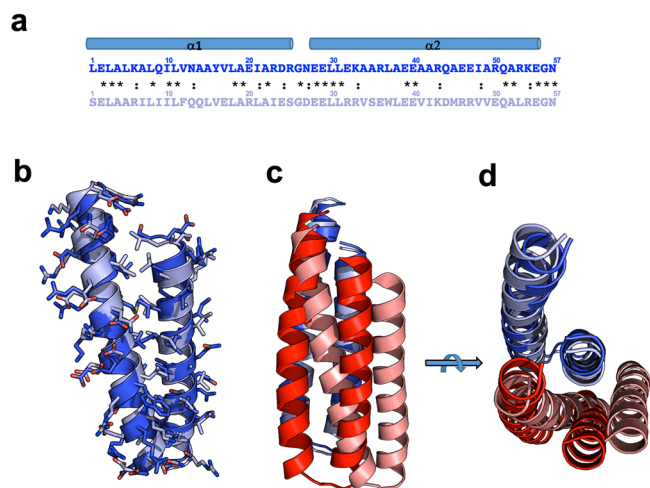
**Generation and analysis of tcTRP24<sub>8</sub>.** Having demonstrated the accuracy and behavior of a designed thick(er) cTRP (tcTRP) harboring nine repeats and a relatively acute radius of curvature between sequential repeats, we next decided to generate and examine a similar tcTRP harboring 24 repeats, corresponding to a considerably larger internal and external diameter and a shallower curvature between consecutive repeats. Because the large size of these constructs (and the corresponding large number of repeats within their designed structures) precluded generating monomeric, single-chain constructs of 24 linear repeats, we

decided to immediately test our designs using smaller subunits with fewer number of repeats that were again intended to self-assemble to form full-sized tcTRP toroidal particles.

Designs corresponding to subunits containing either 6 (tcTRP24<sub>6</sub>) or 8 (tcTRP24<sub>8</sub>) repeats, that were respectively intended to assemble into a full-sized particle with 24 repeats via tetramerization or trimerization (Supplementary Fig. S3a, left) were found to both express at high levels (Supplementary Fig. S3a, right). The tcTRP24<sub>8</sub> construct was slightly better behaved in subsequent purification attempts and was therefore used for further experiments. The protein was purified to relative homogeneity using a three-column protocol (“Methods” and Supplementary Fig. S3b).

The tcTRP24<sub>8</sub> construct eluted from the final SEC column over a wide range of elution volumes (Fig. 5a), corresponding to a heterogeneous population of sizes ranging from greater than 670 kDa to a final peak corresponding to a mass slightly greater than 158 kDa (near the expected mass of individual tcTRP particles composed of three subunits with a total of 24 repeats). Fractions from the SEC elution were subsequently subjected to TEM imaging of negative stained specimen (Fig. 5c, left panel). The resulting images contained mixtures of individual open (‘c’-shaped) and closed (circular) toroidal particles with the





**Fig. 3 Comparison of designed repeats from tcTRP9 and tcTRP24.**

Repeats from tcTRP9 and tcTRP24 are shown with darker shades and lighter shades, respectively. **a** Alignment of the sequences of an individual repeats from each of the two designs. Both designs are based on a 57-residue sequence that forms an anti-parallel two-helix bundle. The overall sequence identity (noted by asterisks) is 37%; sequence similarity (including D/E, N/Q, K/R, and I/V pairs) is 53%. Many of the identities between designs are observed at less-constrained surface-exposed positions, but also are found at interfacial positions that contribute to packing between helices and repeats. **b** Superposition of individual repeats from tcTRP9 and tcTRP24 gives an overall rmsd across all alpha carbons of approximately 1.5 Å, indicating that the backbone architecture of individual repeats from the two designs is similar. **c, d** Superposition of a single repeat from tcTRP9 and tcTRP24 (dark and light blue, respectively) and extension to the neighboring immediate downstream repeat (dark and light red, respectively) illustrates that the curvature and diameters between the two constructs are facilitated by altered packing angles between sequential repeats (resulting in an alpha-carbon rmsd value between the downstream repeats of over 4 Å).

diameter of the latter corresponding to the expected dimension of the designed protein. Those particles were interspersed with fibrous assemblages that appeared to correspond to elongated chains of protein subunits, forming spiral assemblages of variable lengths. The thickness of the fibers was similar to the diameter of the neighboring rings.

Based on those images, we reasoned that the tcTRP24<sub>8</sub> design might be forming a mixture of desired circular particles, along with partially closed assemblages with an exposed interface surface at their termini that might promote the incorporation of additional protein subunits (leading to a mixture of unassembled subunits, closed toroidal particles, and the growth of fibers until free protein was exhausted). We therefore attempted to promote the closed toroidal form of this construct by incorporating a pair of cysteine residues at neighboring positions across each subunit interface so that a disulfide staple might lock the protein into the desired closed topology, as described previously<sup>5,19</sup>. Based on the tcTRP24<sub>8</sub> design, an alanine at position 5 and an isoleucine at position 7 (respectively located within the N- and C-terminal repeats of each subunit) were mutated to cysteine residues, with the intention of installing potential disulfide bonds within each interface of the intended tcTRP24<sub>8</sub> trimeric particle ('tcTRP24<sub>8</sub>SS'). This construct was again well-expressed in *E. coli* and purified using the same three-column protocol including a terminal size exclusion chromatography (SEC) step. Like the original tcTRP24<sub>8</sub> construct, the protein again eluted over a range of volume and corresponding masses extending from approximately 150 to > 670 kDa. However, the peak

**Table 1 Crystallographic data collection and refinement statistics.**

	tcTRP9 (monomer)	tcTRP9 <sub>3</sub> (homotrimer)
<i>Data collection</i>		
Space group	P2 <sub>1</sub>	P2 <sub>1</sub>
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	52.5, 85.5, 56.1	57.6, 78.9, 112.7
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 103.6, 90	90, 93.8, 90
Resolution (Å)	2.1(2.18–2.1) <sup>a</sup>	3.2 (3.31–3.2) <sup>a</sup>
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.036 (0.109)	0.221 (0.698)
<i>I</i> / $\sigma$ <i>I</i>	40.4 (14.0)	8.5 (3.4)
Completeness (%)	99.5 (95.7)	88.0 (90.0)
Redundancy	7.3 (5.9)	5.5 (5.5)
<i>Refinement</i>		
Resolution (Å)	2.1	3.2
No. reflections	28299	16744
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.1701/0.2277	0.2480/0.3302
No. atoms		
Protein	3870	7053
Ligand/ion	0	0
Water	308	12
<i>B</i> -factors		
Protein	28.65	30.27
Ligand/ion	34.01	20.20
Water		
R.m.s. deviations		
Bond lengths (Å)	0.007	0.003
Bond angles (°)	0.75	0.72

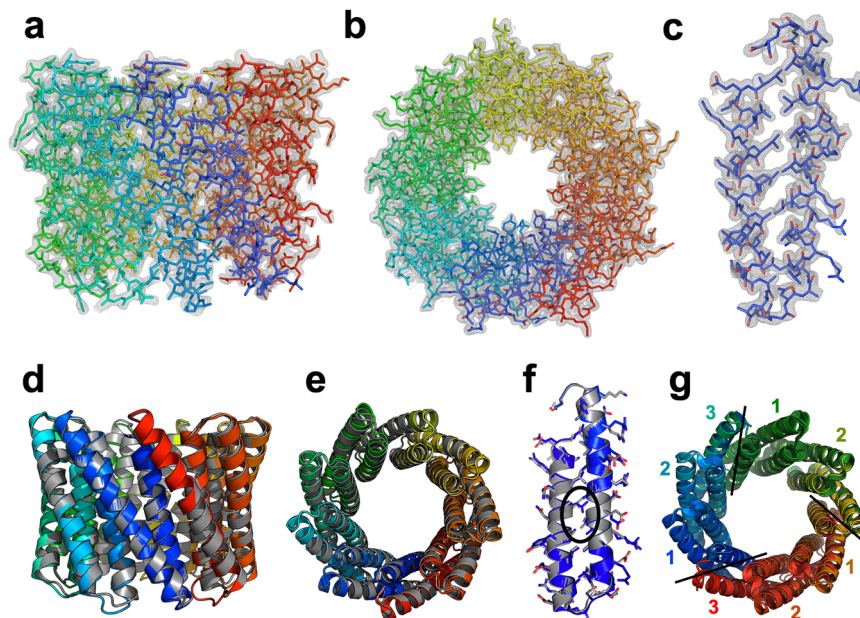
<sup>a</sup>1 crystal for each data set.

corresponding to the mass of the desired trimer appeared to be greater in relative height (Fig. 5b, 'tcTRP24<sub>8</sub>SS Run #1').

Reasoning that the designed disulfide bond might only form after cell lysis and exposure to a non-reducing environment (and during that time, would compete with misfolding and fiber formation), we isolated the slowest eluting half of the final peak (corresponding to the expected elution of a correctly formed trimeric particle), concentrated the protein, and subjected it to a second size exclusion step and analysis (Fig. 5b, 'tcTRP24<sub>8</sub>SS Run #2'). In that second run, the protein now ran as a single uniform peak at a volume and predicted mass corresponding to the desired trimeric assemblage. To further test that conclusion, fractions of tcTRP24<sub>8</sub>SS from this second SEC elution were again subjected to transmission electron microscopy using negative stained specimens (Fig. 5c, right panel) and found to correspond to individual toroidal particles with dimensions close to the computational design of a 24-repeat toroid.

A negative control construct ('tcTRP24<sub>8</sub>Cap'), containing bulky side-chain substitutions in the protein interface that were intended to block self-assembly through steric clashes, was observed to elute at a later volume and smaller corresponding mass, further validating the assembly of the tcTRP24<sub>8</sub> and tcTRP24<sub>8</sub>SS constructs (Fig. 5a).

The structure of the isolated tcTRP24<sub>8</sub>SS construct from the final SEC run was then further examined by CryoEM single particle analysis (Fig. 6, Table 2 and Supplementary Fig. S4). The resulting density map at a resolution of approximately 6.9 Å, at a Gold Standard Fourier Shell Correlation (GSFSC) of 0.143 between the two half maps, clearly indicated the presence of a close circular-shaped particle with two concentric layers of helices lining its periphery. The topology of the map is consistent with the trimeric tcTRP24<sub>8</sub>SS toroidal assembly and is closely superimposed on the original design model (Fig. 6). In the final electron density map, three strong features extending from equivalent positions around the tcTRP particle correspond to N-terminal poly-histidine affinity tags and linkers extending from the N-terminus of each protein



**Fig. 4** Crystal structures of tcTRP9 and tcTRP9<sub>3</sub>. **a–c** Refined models and corresponding 2Fo-Fc electron density map for full-length monomeric tcTRP9. **c** Illustrates the map and refined model for a single repeat within the construct. The structure was determined at 2.1 Å resolution. **d–f** Superposition of the designed model of tcTRP9 (gray) and the experimental crystallographic structure of the same construct (rainbow spectrum; N-terminal repeat = blue, C-terminal repeat = red). The alpha-carbon rmsd between the design and the refined crystallographic coordinates is 0.5 Å. The circled residues in **f** (which shows the side chains for repeat #1 from the design and from the crystal structure) are L11 and L37, which pack against one another in both models, but display opposing rotameric conformations that each promote their interaction in the interhelical interface of the repeat. **g** Superposition of the crystal structure of the monomeric tcTRP9 construct (from the earlier panels) and the crystal structure tcTRP9<sub>3</sub> homotrimer (a self-assembling particle-containing three subunits, each of which contains 3 repeats of the same sequence as shown in Fig. 1; the structure was solved at 3.2 Å resolution). The structure of the latter construct confirms its homotrimeric assembly (as previously indicated by size exclusion chromatography analysis) and alpha-carbon rmsd between superimposed monomer and homotrimer of 0.5 Å. The repeat numbering (1–3, repeated three times) and lines illustrate the subunit boundaries and locations for the homotrimeric tcTRP9<sub>3</sub> construct.

subunit (an unbiased feature that further validated the trimeric assemblage of the particle). Those atoms are clearly disordered and are not modeled into the density.

Inspection of the images from the selected 2D class averages of particles used in the reconstruction indicated that the protein assemblage, even with designed disulfide staples in place at the interfaces between protein subunits, displayed significant conformational flexibility or ‘breathing’ that are manifested in slight deviations from circularity and small but significant variations in the diameter of the particles, which severely limited the resolution of the 3D reconstruction.

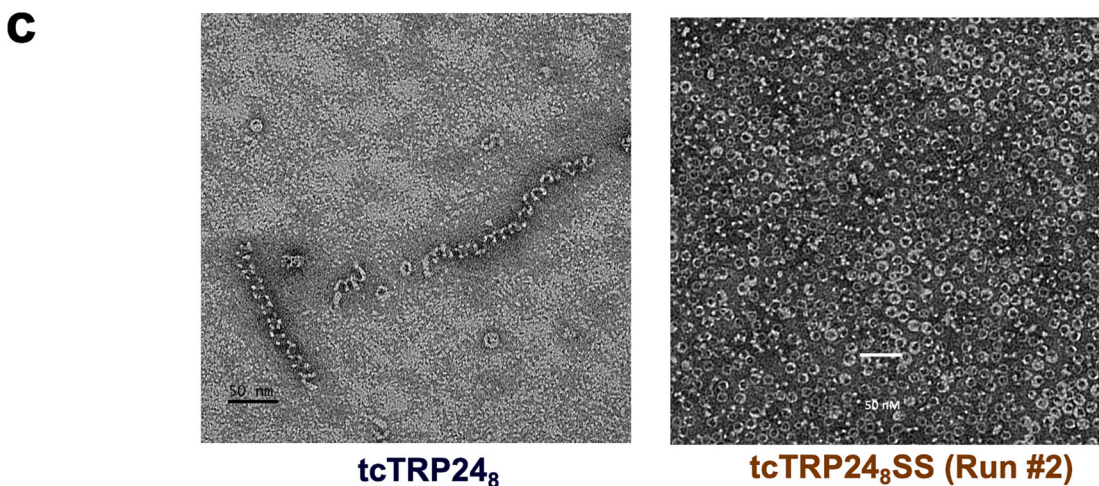
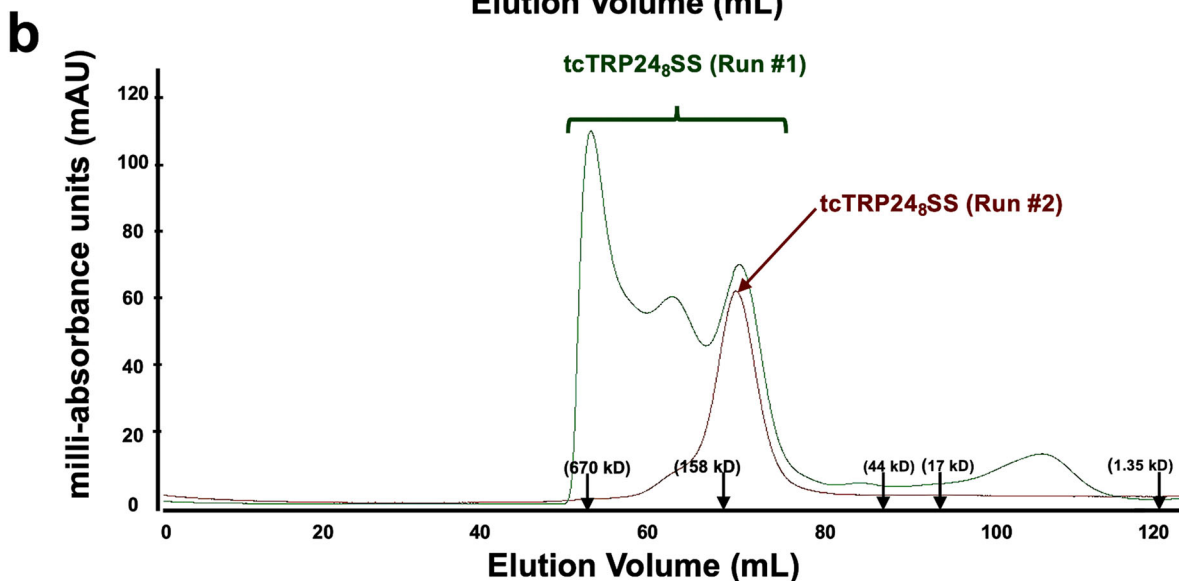
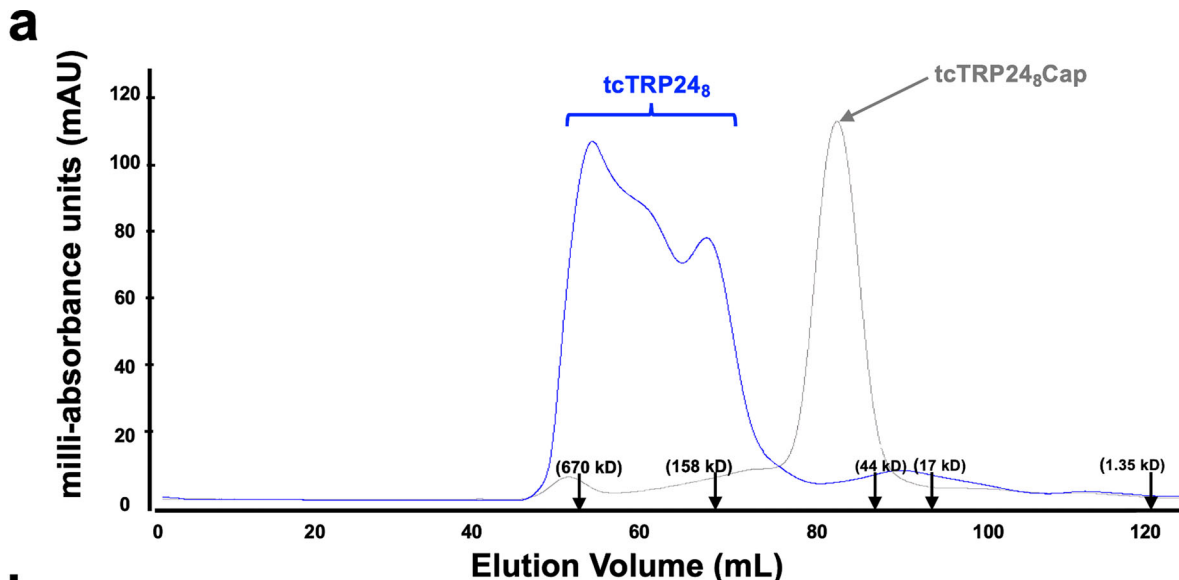
In the same CryoEM analyses, we also noted that while the vast majority of the 2D class averages of the closed circular tcTRP24<sub>8</sub>SS particles contained 24 repeats as designed, a small fraction of particles (less than 5%) contained more or less than 24 repeats (Supplementary Fig. S4), indicating the presence of ‘ $n \pm 1$ ’ and/or ‘ $n \pm 2$ ’ repeats in a small fraction of the underlying tcTRP subunits. One possible explanation for this observation could be occasional recombination events within the repetitious tcTRP24<sub>8</sub> coding sequence, leading to the insertion or deletion of individual DNA repeats from the expression vector. Using PCR, we compared the length of the coding sequence inserts in the original expression vector with plamid populations recovered from cells after extended culture growth and expression but did not see any indication of abnormal length inserts accumulating during cell growth and expression. Thus, the basis for the small percentage of tcTRP24<sub>8</sub> particles harboring more or less than 24 repeats is unclear.

**Generation and analysis of functionalized tcTRPs.** Finally, we tested the ability of the tcTRP24<sub>8</sub> and tcTRP9<sub>3</sub> constructs to

promote the display and function of multiple copies of a functional protein domains, as has been described previously for other engineered protein scaffolds<sup>5,6,20–22</sup>. For this purpose, we conducted experiments using two underlying tcTRP trimeric constructs (tcTRP24<sub>8</sub> and tcTRP9<sub>3</sub>).

In the first experiment, we used fluorescence polarization and surface plasmon resonance (SPR) to demonstrate (1) that the peptide-binding activity of an SH2 domain is maintained when fused to a tcTRP scaffold, and (2) that the association and dissociation behavior of the multimeric SH2-tcTRP construct differs from the free SH2 domain. The tcTRP24<sub>8</sub> construct was functionalized by inserting two copies of a sequence-specific SH2 domain derived from the Nck adapter protein (used previously for a similar purpose with the original, thinner cTRP24 construct)<sup>13</sup> into each protein subunit. The assembled particle (‘tcTRP24<sub>8</sub>-SH2<sub>6</sub>’) thereby displayed six total copies of the same SH2 domain arranged around the same face of the protein scaffold, with four repeats separating each domain (Fig. 7a). The construct was then assayed for binding to its cognate phosphotyrosyl-containing peptide (EHIpYDEVAAD) both free in solution (using fluorescence polarization; Fig. 7b) and on an immobilized surface (using surface plasmon resonance; Fig. 7c).

The solution-based binding function and affinity of free SH2 domains versus tcTRP displayed SH2 domains, measured via fluorescence polarization, were relatively comparable (approximate  $K_D$  values of 1–3 micromolar). The tcTRP24<sub>8</sub> scaffold alone did not display significant peptide binding. In this experiment, the labeled ligand is not immobilized on a surface, and individual ligand molecules are bound independently (in a completely non-cooperative manner) by each SH2 domain around the periphery of the particle. Therefore, there is no



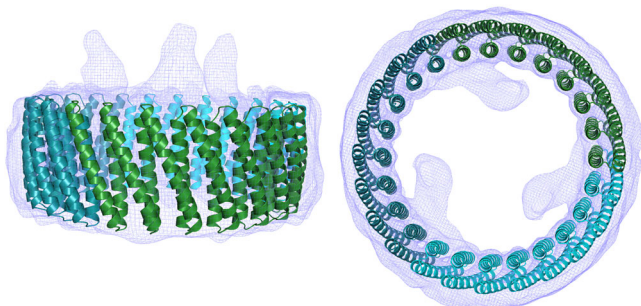
avidity effect to be expected in this experiment, nor is any effect observed.

In contrast, when the peptide binding experiment was carried out using surface plasmon resonance (with the peptide target presented to the protein on the surface of a chip), a significant avidity effect is expected and observed. We observed a greatly

reduced dissociation rate ( $k_{off}$ ) from the chip, offset by a reduced rate of association ( $k_{on}$ ). While we might have expected to see an increased on-rate coupled to a decreased off-rate, a decreased on-rate could instead be caused a combination of (1) greatly increased molecular mass (252 kDa for tcTRP24<sub>8</sub>-SH2<sub>6</sub>, versus 12 kDa for free SH2) and/or (2) reduced



**Fig. 5 Assembly behavior of tcTRP24<sub>8</sub> and tcTRP24<sub>8</sub>SS.** **a** Size-exclusion chromatographic (SEC) elution profiles of the original 'tcTRP24<sub>8</sub>' homotrimeric construct (blue), and the 'tcTRP24<sub>8</sub>Cap' construct (gray), which harbors a pair of disrupting mutations at the subunit interface that prevents self-association. The original construct displays a wide range of apparent masses, with the final peak corresponding to the approximately expected elution and mass for the desired homotrimeric species. **b** SEC elution profiles of the 'tcTRP24<sub>8</sub>SS' homotrimeric construct directly from the initial purification out of bacterial cells ('Run#1', green), and after collecting, concentrating, and re-running the tail of the original, properly sized fractions ('Run#2', brown). The original prep (Run#1) still forms a distribution of masses after purification, albeit with the earliest and latest eluting peaks increased in relative percentage of material. However, collection of the late-arriving fractions corresponding to the predicted mass of the intended homotrimer and a second run over the same column now results in a single uniform peak corresponding to the intended mass of the designed homotrimeric construct. **c** Negative stain EM image of tcTRP24<sub>8</sub> (left) and of tcTRP24<sub>8</sub>SS ('Run #2'; right). The original construct (lacking disulfide staples; left panel) displays a mixture of correctly sized trimeric particles (small rings on the micrograph) and long extended fibers. The resized stapled construct (right panel) displays a uniform distribution of tcTRP particles. Subsequent CryoEM analyses of this construct (Supplementary Figs. S4 and 6) verify the symmetry, size, and repeat copy number of the construct.



**Fig. 6 CryoEM single-particle reconstruction of tcTRP24<sub>8</sub>SS.** Two different views of the electron density map superposed with the original model. The three-strong features extending from equivalent positions around the tcTRP particle correspond to N-terminal poly-histidine affinity tags and linkers extending from the N-terminus of each protein subunit (an unbiased feature that further validated the trimeric assemblage of the particle). Those atoms are clearly disordered and are not modeled into the density.

accessibility of the binding domain after tethering it to the tcTRP scaffold.

In our second, more comprehensive experiment intended to more rigorously examine the effect of binding domain multi-merization, the tcTRP9<sub>3</sub> construct was functionalized with a previously described camelid 'nanobody' (a single chain VHH domain) that was previously raised against the SARS-CoV-1 spike receptor-binding domain (RBD) protein<sup>23</sup>. The VHH contained a point mutation (D61R) that is hypothesized to increase binding to SARS-CoV-2 spike RBD and was fused to the tcTRP domains in several ways, either to each of the three separate N-termini in the trimeric tcTRP or to each of the three separate C-termini (sequences and VHH insertion sites shown in Supplementary Table S1). In both cases, the assembled particles ('tcTRP9<sub>3</sub>-VHH<sub>3</sub>') display three separate copies of the antiviral VHH domain in a trimeric organization extending from one face of the construct. For each construct, four different linker sequences were tested in parallel (Fig. 8, Supplementary Tables S1 and S2 and Supplementary Fig. S5).

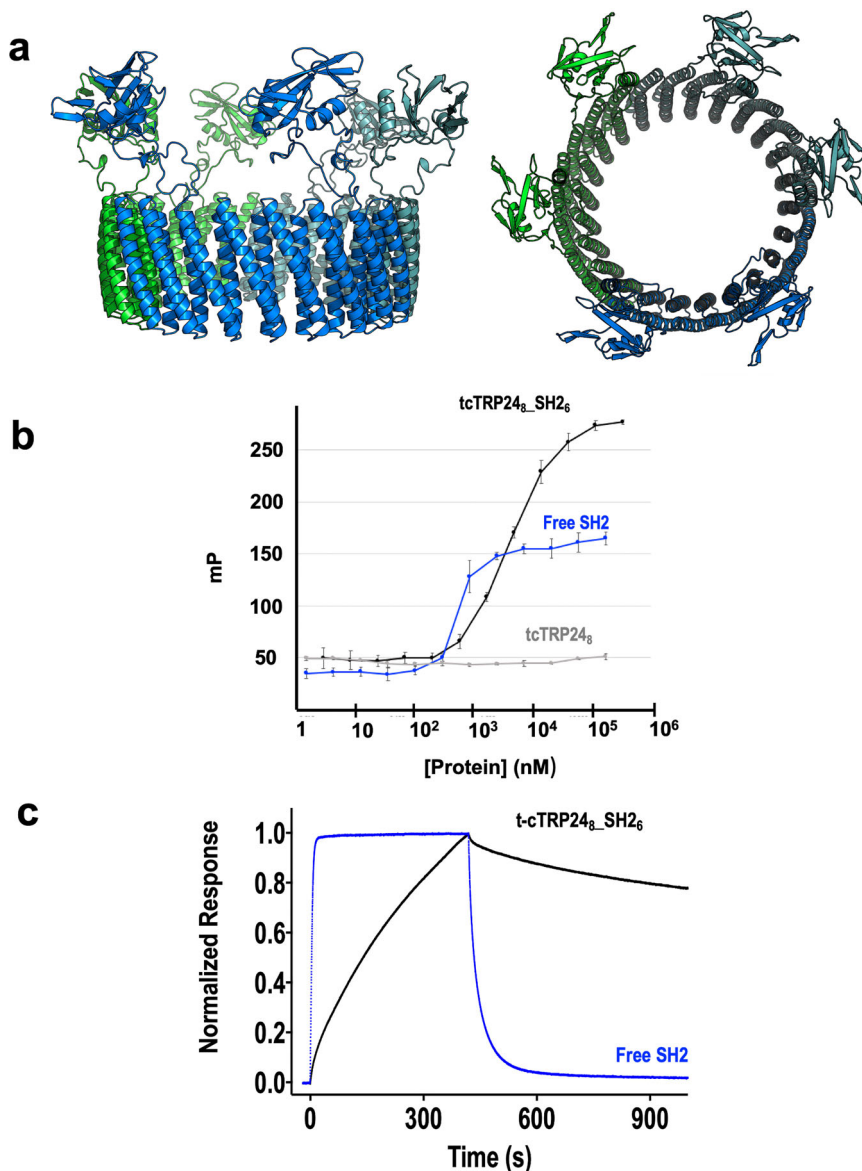
The tcTRP9<sub>3</sub>-VHH<sub>3</sub> constructs (Fig. 8a) were expressed and purified from *E. coli* (Supplementary Table S2 and Fig. 8b) and then assayed for binding to the SARS-CoV-2 spike protein RBD (Supplementary Table S2, Fig. 8c, d; Supplementary Fig. S5) using BLI and ELISA methods. The same constructs were then assayed for their ability to block infection of an HEK293 cell line expressing the ACE2 receptor by SARS-CoV-2 pseudoviral neutralization assay (Supplementary Table S2 and Fig. 8e). Protein expression was somewhat variable, depending upon the position (N- or C-terminal) of the VHH domain relative to the tcTRP subunit to which it was fused, and the identity of the linker

**Table 2 Cryo-EM data collection, refinement, and validation statistics.**

	tcTRP24 <sub>8</sub> SS (EMD-24425) (PDB 7RDR)
<i>Data collection and processing</i>	
Magnification	38 Kx
Voltage (kV)	200
Electron exposure (e-/Å <sup>2</sup> )	40
Defocus range (μm)	-0.8 to -2.3
Pixel size (Å)	1.16
Symmetry imposed	C3
Initial particle images (no.)	627763
Final particle images (no.)	121426
Map resolution (Å)	6.9
FSC threshold	0.143
Map resolution range (Å)	30-6.9
<i>Refinement</i>	
Initial model used (PDB code)	tcTRP24 <sub>8</sub> SS
Model resolution (Å)	6.9
FSC threshold	0.143
Model resolution range (Å)	30-6.9
Map sharpening B factor (Å <sup>2</sup> )	-418.6
<i>Model composition</i>	
Non-hydrogen atoms	None
Protein residues	1368
Ligands	None
<i>B factors (Å<sup>2</sup>)</i>	
Protein	20.0
Ligand	N/A
<i>R.m.s. deviations</i>	
Bond lengths (Å)	0.01
Bond angles (°)	1.2
<i>Validation</i>	
MolProbity score	1.6
Clashscore	1.05
Poor rotamers (%)	0
<i>Ramachandran plot</i>	
Favored (%)	98.32
Allowed (%)	1.68
Disallowed (%)	0

used to fuse the VHH to the tcTRP repeats. Most of the constructs expressed at high levels corresponding to multiple milligrams of recovered protein per liter of culture, with the constructs corresponding to a C-terminal fusion of the VHH (tcTRP9<sub>3</sub>-linker-VHH-6xHis) displaying more uniform expression and recovery. The trimeric constructs show binding to the SARS-CoV2 RBD and viral neutralization at concentrations lower than or comparable to the reported binding ( $K_D \sim 39$  nM) and virus neutralization ( $IC_{50} \sim 4$  nM) by the same VHH organized in a dimeric arrangement via fusion to an





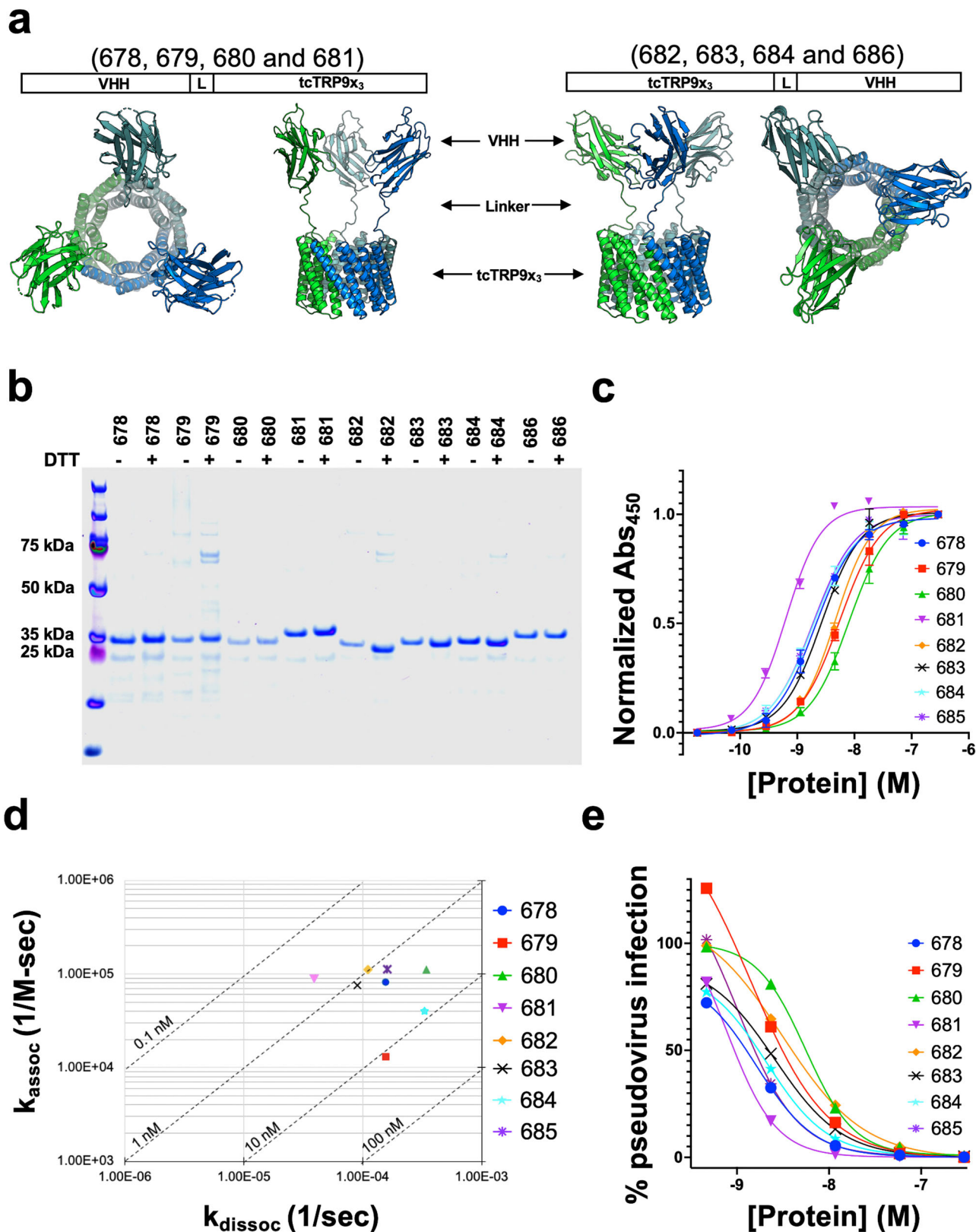
**Fig. 7 Peptide binding function of tcTRP24<sub>8</sub>SS\_SH2<sub>6</sub> in solution and on a surface.** **a** Schematic model of the construct shown in two orientations. Two copies of the SH2 domain are fused into a loop (between repeats 2 and 3, and between repeats 6 and 7) in each of three subunits, resulting in the presence of six domains distributed around the periphery of the construct in a hexagonal arrangement. **b** Binding of fluorescently labeled Tir10 phosphotyrosyl peptide (corresponding to the physiological binding target for the Nck SH2 domain) was measured in solution using fluorescence polarization. Binding activity was measured using isolated SH2 domain (blue), and tcTRP24<sub>8</sub>-SH2<sub>2</sub> (a tcTRP trimer containing two copies of fused SH2 domains per subunit, i.e., six copies total; black). A control experiment was also conducted using the free ‘naked’ tcTRP24<sub>8</sub> (gray) that contained no functional protein domains. The protein concentration is normalized to account for the ratio of SH2 domains per molecule (6 SH2 domains per tcTRP<sub>8</sub>-SH2<sub>2</sub> versus 1 per free SH2). Both constructs display saturable binding and an approximate K<sub>D</sub> of 1–3 micromolar. All binding experiments were conducted in triplicate using independent aliquots of each protein. Data shown as mean and standard deviation for  $n = 3$  measurements. **c** Binding and dissociation of free SH2 (blue) and tcTRP24<sub>8</sub>-SH2<sub>2</sub> (black) to surface-bound peptide. Surface Plasmon Resonance (‘Biacore’) was used to measure the binding to captured biotinylated phosphotyrosyl peptide as described in Methods.

immunoglobulin Fc domain<sup>23</sup>. The most active construct in our experiments (tcTRP9<sub>3</sub>-VHH681<sub>3</sub>) displayed the lowest, sub-nanomolar values for both binding assays (K<sub>D</sub> and EC<sub>50</sub> ~ 0.43 and 0.64 nM, respectively, and viral neutralization IC<sub>50</sub> ~ 0.78 nM).

## Discussion

The study described above expands upon previously published descriptions of de novo designed circular tandem repeat proteins (cTRPs)<sup>5,13</sup> that were originally inspired by naturally occurring tandem repeat proteins such as TAL (transcription activator-like

effectors<sup>24</sup>, PUF (Pumilio and FBP) proteins<sup>25</sup>, and TPR (tetra-tricopeptide repeat) proteins<sup>26</sup>. Each of those protein families, which have respectively evolved to bind double-stranded DNA, single-stranded RNA, and peptide targets in a sequence and/or structure-specific manner, are composed of underlying repeats of approximately 34 residues. In carrying out their functions, they display significant plasticity corresponding to interdomain dynamic flexibility, differing amounts of sequence divergence between repeats within individual protein factors, and different ligand-binding mechanisms that generally depend on the identity of residues at canonical positions within each repeat<sup>27</sup>.



The length and helical content of the repeats that were incorporated in our first generation of cTRPs<sup>5,13</sup> were similar to those found in the naturally occurring repeat proteins listed above, while otherwise exploring geometric constraints and folding topologies (including the requirement of a closed circular architecture) that were unique from those observed in nature or

found in previously solved structures. The additional structural requirement for a closed circular architecture imposed the need for heightened accuracy in the design of contacts that lead to precise rotational symmetry, such that the toroidal organization of the folded particle is realized in the correct packing and register between N- and C-terminal repeats at each protein interface.

**Fig. 8 Generation and functional characterization of homotrimeric tcTRP<sub>9x3</sub>-VHH<sub>3</sub> constructs.** **a** Each construct (sequences and locations of inserted VHH domains provided in Supplementary Table S1) corresponds to a fusion at their N- or C-termini with a VHH domain previously shown to bind to the SAR—CoV2 spike protein receptor binding domain<sup>23</sup>. Eight constructs were tested, corresponding to various linkers that fuse the VHH domain to the N- or C-terminal ends of each subunit, in either case generating a trimeric array of VHH domains arranged on one side of the tcTRP<sub>9x3</sub> scaffold. The ability of each construct to bind the SARS CoV2 spike receptor-binding domain (RBD) was measured using two approaches (Biolayer interferometry and ELISA); the ability of the same constructs to inhibit cell infection was measured using a pseudoviral neutralization assay. **b** Purification of individual constructs from *E. coli*. Yields ranged from ~15 to ~80 mg/L. **c** The relative ability of each construct to bind to the SARS-Cov2 receptor-binding domain (RBD) was measured using an ELISA analysis, leading to EC<sub>50</sub> values (Supplementary Table S2) ranging from a low of 0.4 nM (construct 681) to 12 nM (construct 680). All binding experiments were conducted in triplicate using independent aliquots of each protein. Data shown as mean and s.d. for *n* = 3 measurements. **d** The relative ability of each construct to bind to the SARS-Cov2 receptor-binding domain (RBD) was measured using Biolayer interferometry (BLI). The rate of association and dissociation are calculated from concentration-dependent responses. The rate of association (*k*<sub>a</sub>) is plotted on the Y-axis as inverse molar second (1/MS), and the rate of dissociation (*k*<sub>dis</sub>) is plotted on the X-axis as inverse second. Corresponding equilibrium binding constants (*K*<sub>D</sub>) are plotted on a diagonal axis as nano molar concentrations. See also Supplementary Fig. S5. **e** The relative ability of each construct to achieve 50% inhibition of infection of 293T-Ace2 cells using a pseudoviral neutralization assay with SARS-CoV-2 spike pseudotyped lentiviruses was measured. The IC<sub>50</sub> values (Supplementary Table S2) ranged from 0.78 nM (construct 681) to 5.5 nM (construct 680). All experiments were conducted in triplicate using independent aliquots of each protein. Data shown as the mean and standard deviation for *n* = 3 measurements.

While the final designed scaffolds were well-behaved in many ways (including their expression, solution behavior, thermal stability, and ability to accommodate additional functional protein domains), their ability to assemble from smaller subunits was occasionally observed to be sub-optimal, in terms of how efficiently they assembled<sup>5</sup> and whether they incorporated the desired number of subunits into the full-sized construct<sup>13</sup>.

In this study, we tested a hypothesis that the expansion of cTRP repeat size and a corresponding increase in the area and residues within the packing interfaces between repeats and subunits might be a route by which the energetics and control of the self-assembly of such constructs could be improved. This appears to have been true, at least in part: the assembly of a trimer containing nine repeats (from subunits containing three repeats each) is enhanced both energetically and sterically relative to the original, comparable design and analysis of a corresponding 9-repeat trimer<sup>13</sup>. Additionally, whereas prior assembly a 24-repeat cTRP construct from smaller subunits was inefficient<sup>5</sup>, the newly designed tcTRP<sub>24</sub> construct displayed enhanced assembly (as indicated by size exclusion chromatography analyses that indicated all subunits were incorporated into larger particles). However, assembly of that construct was accompanied by the formation of undesired, higher-order fibrils (Fig. 5c, left) that we believe are likely promoted by ‘off-register’ packing between subunits, leading to a translational component (i.e., a spiral screw-axis) that confounds designed strict rotational symmetry and provides a partially exposed surface at one interface that can act as a nucleation site for subunit addition and continued extension of the protein complex. The introduction of a disulfide staple at the designed interface (which was previously required simply to promote assembly of otherwise disassociated subunits)<sup>5</sup> was helpful in promoting and trapping desired circularized constructs, which could then be isolated as stable particles for further analyses (Fig. 5c, right). Future studies will examine this behavior, and the computational improvements necessary to prevent this behavior, further.

A wide variety of applications would appear possible for de novo designed protein scaffolds (either tcTRPs, or alternative designed protein particles) that exhibit efficient self-assembly in defined copy number, size, and symmetry. Prior studies with such designed protein scaffolds have demonstrated the creation of high-avidity binding particles<sup>5</sup>, highly active cell stimulatory factors<sup>5</sup>, and epitope display particles for vaccine development<sup>20,22</sup>. In this study, we examine the possibility of functionalizing such designed protein scaffolds for the disruption of biological particles and assemblages, specifically examining how the trimeric assemblage of a previously described camelid

nanobody that was selected to bind a viral spike protein receptor-binding domain might improve on such a construct’s viral neutralization properties, relative to a commonly used dimerization motif<sup>23</sup>. It should be possible to expand on this concept for other purposes, such as the degradation of bacterial biofilms (themselves a high copy-number, high avidity target)<sup>28</sup> or other bioremediation applications. An additional area for investigation might be the use of such self-assembling protein scaffolds for the organization of enzyme domains within a defined pathway, potentially leading to kinetic gains through proximity effects between sequential activities<sup>29,30</sup>.

The comparative examination between various tcTRP trimers augmented with a three-fold arrangement of an anti-SARS CoV-2 VHH domain<sup>23</sup> demonstrates the overall amenability of the underlying protein scaffold towards variable fusion strategies at the N- or C-terminus of each subunit, as well as when various linker compositions and lengths are employed. At the same time, the fact that one construct (tcTRP<sub>9x3</sub>-VHH681<sub>3</sub>) demonstrates consistently superior target binding affinity and biological activity across multiple assays indicates the importance of systematically examining multiple functionalization strategies and architectures for potential applications such as direct viral neutralization.

This study points towards several steps that should be further addressed as the cTRP platform (and other self-assembling protein scaffolds) are further optimized. The mixed results from the assembly of the larger of the tcTRP constructs (harboring 24 repeats in various assemblages) appear to indicate that small design inaccuracies may be magnified as construct size increases and indicates that further refinement of the computational approach may yield additional improvement in behavior and performance. At the same time, the successful creation of self-assembling tcTRPs with increased surfaces and interface areas (and corresponding larger numbers of residues involved in the contacts within those interfaces) may facilitate yet another generation of self-assembling symmetric protein particles, in which homo-oligomeric assemblages are further diversified through the design and incorporation of unique contacts between each subunit (perhaps in the form of directional hydrogen-bond networks) that facilitate the creation of hetero-oligomeric particles and enable their functionalization with multiple forms of functional protein domains and cargo in precisely defined stoichiometric ratios and organizational patterns.

## Methods

**Constructs and nomenclature.** The construct names, sequences, and figures for all constructs described in this paper are provided in Supplementary Table S1. The constructs described in this article (and listed in that Table) are referred to as



'tcTRPs' ('thick circular Tandem Repeat Proteins'). Following the nomenclature for previous cTRP constructs described and used in a prior study<sup>5</sup>, the exact size and assemblies are further annotated as 'tcTRP9' and 'tcTRP24', respectively, where the underlying tcTRP9 and tcTRP24 scaffolds contain a total of 9 or 24 repeats. For scaffolds that are assembled from smaller identical protein subunits, the constructs are annotated as tcTRP9<sub>x</sub> and tcTRP24<sub>x</sub>, where each subunit contains 'x' repeats. For example, 'tcTRP9<sub>3</sub>' (read as 'tcTRP9 sub3') refers to a particle containing a total of nine repeats, assembled from the trimerization of protein subunits containing 3 repeats each. Similarly, 'tcTRP24<sub>8</sub>' corresponds to a particle containing a total of 24 repeats, assembled from the trimerization of protein subunits containing eight repeats each. Some tcTRP24 constructs contain disulfide staples between protein subunits and are named 'tcTRP24<sub>8</sub>SS'.

Finally, constructs harboring additional protein cargo fused at defined positions around the tcTRP periphery are further denoted in the form 'tcTRP24SS-Cargo<sub>y</sub>', where 'y' is the total number of cargo domains per construct. One example in this paper ('tcTRP24<sub>8</sub>-SH2<sub>6</sub>') corresponds to a trimeric tcTRP containing a total of 24 repeats ('tcTRP24...'), assembled from three identical subunits that contain 8 repeats each ('tcTRP24<sub>8</sub>...'), and with each subunit further displaying two SH2 peptide-binding domains, for a total of six such domains displayed ('tcTRP24<sub>8</sub>-SH2<sub>6</sub>').

**Computational protein design.** Protein design simulations were conducted exactly as described previously<sup>5</sup>. That approach corresponds to a geometry-guided repeat computational strategy implemented in the Rosetta package<sup>18</sup> with additional de novo design elements<sup>17</sup>. Key features include the application of parametric symmetrization of backbone and side-chain conformations applied across all repeats (such that computational complexity scales only with repeat length); a pseudo-energy term that optimizes the inter-repeat geometry; clustering and resampling protocols that allow intensified exploration of promising topologies; and an *in silico* validation step that assesses sequence-structure compatibility by attempting to re-predict the designed structure given only the designed sequence. Applying this design procedure produces a diverse array of toroidal structures.

In this work, two additional modifications of the previously described approach were implemented: the 'Ref2015' energy function<sup>31</sup> was used for all protein design and structure recapitulation calculations, and the range of allowed helix lengths was increased to 20–45 residues. Initial simulations explored helical linkers of length 1–5 residues with unconstrained backbone torsion angles. Clustering analysis of low-energy designs from these simulations revealed convergence on a 2-residue, antiparallel connection with backbone conformation 'GB' (one residue in a left-handed alpha-helical conformation and one residue in an extended conformation). A subsequent round of designs focused on 'GB' linkers was conducted to enhance sampling in this low-energy region of conformational space.

The identification of residue positions for the incorporation of disulfide staples into the tcTRP24<sub>8</sub> trimer was performed by utilizing the Rosetta 'Disulfidize Mover' routine<sup>32</sup>. Each edge helix involved in the trimerization was selected and corresponding residues scanned. The distance between adjoining beta-carbons was used to determine potential residues; once identified they were mutated to cysteine residues and tested through rotamer optimization and energy minimization.

**Protein expression and purification.** All constructs encoding tcTRPs described in this study were designed and ligated into an in-house pET15HE expression vector<sup>33</sup> or a commercially available pET28b expression vector and sequence verified. The coding sequence and the corresponding translated protein sequences, including the N-terminal poly-histidine affinity tag and thrombin cleavage site preceding the first tcTRP repeat, are provided in Supplementary Table S1. The free SH2 domain was subcloned and purified as previously described<sup>5</sup>.

Plasmids were transfected into BL21(DE3)-RIL *Escherichia coli* cells (Agilent Technologies) and plated on LB medium augmented with 100 µg mL<sup>-1</sup> ampicillin. Protein was expressed via a previously described autoinduction protocol<sup>34</sup>. Briefly, 1 L of ZYP-5052 media containing 100 µg mL<sup>-1</sup> ampicillin was inoculated with individual transformants, shaken at 37 °C for 8 h followed by 16 °C for 24 h. Expression cultures were pelleted by centrifugation and stored at -20 °C until purification.

Frozen cell pellets were thawed at room temperature and resuspended in 100 mL of 1× phosphate-buffered saline (PBS; 137 mM NaCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.7 mM KCl, pH 7.4). PMSF was added to a final concentration of 0.5 µM. Cells were lysed via sonication and centrifuged in an SS34 rotor at 16,000 rpm for 20 min at 4 °C to remove cell debris. The supernatant was passed through a 5 µm filter, added to 2 mL of nickel-NTA metal affinity resin (Invitrogen) equilibrated with 1× PBS, and then incubated on a rocker platform at 4 °C for 1 h. After loading onto a gravity-fed column, the resin was washed twice with 25 mL of PBS containing 25 mM Imidazole. The protein was then eluted from the column by three additions of 5 mL 1× PBS containing 300 mM Imidazole. Fractions containing the eluted protein were pooled, concentrated, and buffer exchanged into 1× PBS. The sample was then filtered through a 0.2 µm filter and run over a size exclusion column (Cytiva HiLoad 16/60 Superdex 200) equilibrated in either 1× PBS or 20 mM Tris pH 7.5 + 150 mM NaCl.

Those constructs encoding tcTRPs trimers functionalized with three identical copies of VHH 72D61R (an anti-SARS-CoV-2 camelid nanobody targeting the viral receptor-binding domain or 'RBD')<sup>23</sup> were subcloned into a modified pET28

b(+) vector where the Kanamycin bacterial resistance gene is replaced with Ampicillin bacterial resistance gene. The constructs were sequence verified by Sanger sequencing. Protein expression was carried out using the method described above for tcTRPs expression, with the additional step of buffer exchange via dialysis from Ni-NTA elution buffer to 1×PBS, prior to use in assays.

**Circular dichroism.** Purified proteins were dialyzed overnight into 10 mM potassium phosphate buffer at pH 7.0, then diluted to 20.8 µM, as determined using the trimeric molecular weight of 166077 Daltons and extinction coefficient on a NanoDrop spectrophotometer (Thermo Fisher). Thermal denaturation experiments were performed on a JASCO J-815 spectrometer with a Peltier temperature controller. Wavelength scans from 190–250 nm were performed at 25 °C, 95 °C, and cooling back to 25 °C.

**Crystallization and structure determination of tcTRP9 and tcTRP9<sub>3</sub>.** Both purified proteins were crystallized at 22 mg/mL with 100 mM sodium acetate pH 4.5 and 25% polyethylene glycol 400 in a 24-well hanging drop tray. Crystals were cryocooled in the same buffer via direct plunge into liquid nitrogen. Data was collected under cryocooled conditions (-150 °C) on a Saturn 944+ CCD area detector (Rigaku Inc.) using X-rays produced at 1.54 Å wavelength by a Rigaku HF-007 rotating anode generator. Data were processed using program HKL2000<sup>35</sup>. Molecular replacement was performed using the de novo designed model of tcTRP9 using PHASER<sup>36</sup> in the PHENIX program suite<sup>37</sup>. Refinement was done utilizing programs COOT<sup>38</sup> and REFMAC<sup>39</sup>. Figures were generated using program PYMOL<sup>40</sup>. The final Ramachandran distribution for backbone angles (favored, allowed, outliers) were 98.6, 1.4, 0% for tcTRP9 and 96.3, 2.5, 1.2% for tcTRP9<sub>3</sub> (Table 1).

**Cryogenic electron microscopy (CryoEM) visualization of tcTRP24<sub>8</sub> and tcTRP24<sub>8</sub>SS.** Both purified proteins were screened with negative-stained transmission electron microscopy (TEM) using a 120 KV JOELI400 electron microscope equipped with a 16 megapixel (4k × 4k) GATAN RIOL CMOS detector. The samples were prepared by depositing 4 µL of purified proteins at approximately 40 nM to the surface of a glow-discharged uniform carbon-coated grid. The particles were allowed to adsorb to the carbon film for ~ 1 min and washed three times with 20 µL of water and once with a drop of 0.7% uranyl formate followed by staining for 25 s with a 40 µL droplet of uranyl formate solution. Excess stains were wicked away with filter paper and the grids were air-dried overnight prior to analysis.

The tcTRP24<sub>8</sub>SS particles were further analyzed by CryoEM. Samples were prepared by applying an aliquot of 3 µL protein sample of tcTRP24<sub>8</sub>SS to a glow-discharged Quantifoil 1.2/1.3 holey carbon grid, blotted with filter paper for 5 s and plunge-cooled in liquid ethane using an FEI Vitrobot Mark IV. Cryo-EM micrographs were collected on a 200 kV Glacios microscope (FEI) equipped with a Gatan K2 Summit direct detection camera. The microscope was operated at a calibrated magnification of 37,000×, yielding a pixel size of 1.16 Å on micrographs with an accumulated dosage of 60 e<sup>-</sup>/Å<sup>2</sup>S. In total, 627 movies were collected from two screening sessions, including 82 at a tilt angle of 45°.

All data preprocessing, 2D classification, and 3D model generation and refinement, as well as post refinement polishing, were performed using the software package CryoSPARC<sup>41</sup>. For each movie stack, the frames were aligned for beam-induced motion correction using Patch-motion-correction. Patch-CTF was used to estimate the contrast transfer function (CTF) parameter. A new ring-shape algorithm with inner/outer diameters of 100/120 was used for automated blob picking. After inspection and local motion correction, 627763 particles were accepted for reference-free 2D classification. Two consecutive runs of 2D classification/selection were used to root out false positive and bad (overlapping) particles. A total of 121426 particles in 20 classes were used for ab initio 3D reconstruction.

It is obvious from the selected classes that there were at least two populations of particles with different diameters. Three models were requested for ab initio 3D reconstruction. Results from 3D reconstruction showed multiple circular-disk particles with different diameters. The proportion of the three 3D classes varied with the number of consecutive 2D classification/select and images selected. Multiple trials were performed with different particle picking protocols and particle diameters. All approaches yielded similar results.

**Peptide binding assays in solution via fluorescence polarization.** A 10-residue peptide, Tir10, containing a phosphorylated tyrosine ('pY'), was chemically synthesized with a FITC tag at the 5'-end linked to the peptide with a 7 atom aminohexanoyl spacer, Ahx (GenScript).

Tir10 : FITC – Ahx – EHI(pY)DEVAAD

Tir10 stock was re-suspended to 5.7 mM in DMSO, then diluted to 0.5 µM in fluorescence polarization ('FP') Buffer (20 mM HEPES, 150 mM KCl, pH 7.4). Proteins were exchanged into FP Buffer then two-fold serially diluted from 23 to 0.01 µM (tcTRP24<sub>8</sub>SS and free SH2) or 34 to 0.02 µM (tcTRP24<sub>8</sub>SS-SH2<sub>2</sub>). Diluted proteins were mixed with Tir10 at a ratio of 9:1 for final concentrations of 21–0.01 µM or 31–0.015 µM protein, respectively, and 0.05 µM Tir10, then

incubated, shielded from light, at room temperature for 20 min. FP values were read at excitation of 485 nm and emission of 525 nm (SpectraMax M5). After subtracting FP buffer only background from the raw perpendicular (S) and parallel (P) measurements, polarization (mP) and anisotropy (r) were calculated with the following equations:

$$mP = \left( \frac{P - S}{P + S} \right) \times 1000$$

$$r = \frac{P - S}{P + S}$$

#### Peptide binding assays on a surface via surface plasmon resonance (SPR).

SPR experiments were performed at 25 °C on a Biacore T100 instrument (Cytiva) with a Series S SA chip using a running buffer of 10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05% surfactant P20 with 0.1 mg/mL bovine serum albumin. Biotinylated Tir-10\_v2 (Biotin-Ahx-EHI-pY-DEVAAD) at 10 ng/mL was injected over one flow cell at a flow rate of 10 µL/minute to capture ~15 RUs of peptide. A flow cell with streptavidin alone was used as a reference surface. Analytes were repurified by SEC just prior to use. Buffer blanks and analytes (10 nM tcTRP24<sub>8</sub>SS, 60 nM free SH2, and 10 nM tcTRP24<sub>8</sub>SS-SH2<sub>2</sub>) were injected at a flow rate 50 µL/minute with 7 min of association and 10 min of dissociation. An overlay plot of double-referenced data was generated, then normalized for off-rate comparison by dividing each curve by its maximum response in Scrubber 2.0b software (BioLogic Software). Maximum binding responses observed were 125 and 123 RUs for free SH2 and tcTRP24<sub>8</sub>SS-SH2<sub>2</sub>, respectively. The tcTRP24<sub>8</sub>SS control did not bind. The figure was made in Prism 7 (GraphPad) for Mac OS X version 7.0d.

#### SARS-CoV-2 pseudoviral neutralization assays.

The SARS-CoV-2 pseudotyped lentiviral particle generation, titrating and neutralization assays were performed as previously described protocol<sup>42</sup> with minor modifications, as described here. Poly-L-lysine coated clear bottom 96-well black plates (Thermo Scientific, 12-566-70) were seeded with 293-ACE2 cells (provided by the Jessie Bloom lab at the Fred Hutchinson Cancer Research Center) at a density of 1.25 × 10<sup>4</sup> cells per well in 50 µL volume. Twelve hours after seeding, five-fold antibody dilutions, starting at 50 µg/mL, were prepared. Control, virus only and cell only samples were prepared as previously reported. 60 µL of the titered virus was added and mixed with antibody dilutions and virus only wells. The mix was incubated at 37 °C for 1 h. 100 µL of the mix from each well was added to the corresponding well on 293-ACE2 seeded plates. Polybrene (Sigma Aldrich, P4707) was added to each well at a final concentration of 5 µg/mL. Plates were incubated at 37 °C for 60 h post infection. Virus neutralization was assessed by measuring luminescence. While incubating, Bright-Glo Luciferase reagents (Promega, E2610) were thawed, equilibrated at room temperature, and prepared following the manufacturer's recommendation. 100 µL of growth media was removed from each well and 30 µL per well of luciferase reagent added. Plates were incubated for 2 min at room temperature in the dark and luminescence was measured using an M2 plate reader (Molecular Devices). Luminescence RLU from virus only wells were normalized as 100% infectivity and RLUs from cells only were normalized as 0% infectivity. Infectivity and IC50 were calculated using Four Parameter Logistic Regression on GraphPad Prism (GraphPad Software).

#### ELISA based trimer VHH binding to SARS-CoV-2 RBD.

To assay trimeric VHH binding to RBD, ELISA-based binding assays were performed using the following protocol. High-affinity ELISA plates (Greiner Bio-one, Catalog Number 655084) were coated with SARS-CoV-2 RBD (Roland Strong, Fred Hutchinson Cancer Research Center) at 1 µg/mL in bicarbonate buffer (Sigma, Catalog Number C3041-100CAP) overnight at 4 °C. Coated plates were washed three times with ELISA wash buffer (1XPBS, (Fisher BioReagents, Catalog Number B399-4), supplemented with 0.05% Tween-20 (Thermo Fisher Scientific, Catalog Number B2337-500)). ELISA plates were washed on AquaMax 2000 Microplate Washer (Molecular Devices). Washed plates were blocked with ELISA blocking buffer (ELISA wash buffer supplemented with 5% Non-fat dried milk) for 2 h at room temperature. While blocking, VHH dilutions were prepared as follows. Trimeric VHH 72D61R were diluted with ELISA blocking buffer in 4-fold serial dilution starting at 10 µg/mL. ELISA blocking solutions were aspirated, and 100 µL of antibody dilutions were added into each corresponding well. Plates were incubated at room temperature for 60 min and washed three times with ELISA wash buffer.

The detection antibody, HRP conjugated MonoRab<sup>TM</sup> Rabbit Anti-Camelid VHH Cocktail (GenScript, A02016) was diluted 1:10,000 in ELISA blocking buffer. The detection antibody was added at 100 µL/well and plates were incubated for 30 min at room temperature then washed two times with ELISA wash buffer followed by a wash with 1XPBS. Peroxidase activity was measured using chemiluminescence by adding 100 µL Sera Care KPL TMB Microwell Peroxidase Substrate (Sera Care Life Sciences Inc., 5120-0047) following the manufacturer's recommendation. Peroxidase activity was quenched after 5 min incubation at room temperature by adding 50 µL per well 1 M Hydrochloric acid. Chemiluminescence was measured on M2 plate reader (Molecular Devices) at 450 nm wavelength using SoftMax software (Molecular Devices). Binding EC50 was calculated using averages

of replicates and Four Parameter Logistic Regression on GraphPad Prism (GraphPad Software).

#### Bio-Layer Interferometry (BLI) based Trimer VHH binding to SARS-CoV-2 RBD.

BLI measurements were performed on the Octet RED96 system (ForteBio) using High Precision Streptavidin (SAX) Biosensors (ForteBio). Biosensors were hydrated with phosphate-buffered saline (PBS) at pH 7.4 at room temperature for 10 min in 96-well flat-bottom microplate (Greiner, 655209). All kinetics experiments were performed at 30 °C with 1000 rpm agitation in the kinetics module. Biosensors were dipped into PBS containing wells for 60 s prior to antigen loading. Biosensors were loaded with enzymatically biotinylated RBD (provided by the Roland Strong Lab at the Fred Hutchinson Cancer Research Center) at 1 µg/mL in phosphate buffer, pH 7.4 for 300 s to achieve ~0.6–1 nm response. Loading was quenched by incubating biosensors in 50 µM Biotin (Sigma Aldrich, 576-19-2) for 60 s. Baseline were established by incubating antigen-loaded biosensors in kinetics buffer (PBS + 0.02% Tween 20, 0.1% BSA, 0.05% Sodium azide) for 120 s. Following baseline measurements to determine the rate of association, antigen-loaded biosensor tips were dipped for 50 s into three-fold dilution series of trimeric tcTRP9<sub>3</sub>-VHH<sub>3</sub> fusions starting at a protein concentration corresponding to approximately 100 nM. Analyte bound biosensors were dipped into kinetics buffer for 120–300 s to measure the rate of dissociation. Kinetic analyses were performed using the HT 11.1.1.39 Data Analysis module (ForteBio). Results were double referenced. The association and dissociation steps were both used in a 1:1 binding model with global fitting.

**Statistics and reproducibility.** Biochemical experiments reporting binding interactions (fluorescence polarization (FP) assays for the SH2-tcTRP fusions in Fig. 7; BLI and ELISA assays for VHH-tcTRP constructs in Fig. 8 and viral pseudoneutralization assays in Fig. 8) were conducted in triplicate using independent aliquots of each protein. Data are shown as mean and standard for  $n = 3$  measurements.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

#### Data availability

The structures reported in this study have been deposited in the protein data base ([www.rcsb.org](http://www.rcsb.org)) under PDB ID (entries 6XR1, 6XR2, and 7RDR) and the EMDB (entry 24425) and are available for immediate download. The original source data corresponding to all biochemical analyses have been added to the Harvard Dataverse depository (at <https://dataverse.harvard.edu/dataverse/tcTRP>) for public download and are also archived at our respective institutions and available upon request. Plasmids corresponding to tcTRP9 and tcTRP24 constructs described in this paper are available from the authors upon request and have also been provided to the Addgene public depository (IDs 175793 through 175800). Any remaining information can be obtained from the corresponding author upon reasonable request.

#### Code availability

The use of all computational algorithms in this study is described in Methods and the corresponding code and programs are all available to the public.

Received: 5 April 2021; Accepted: 7 October 2021;

Published online: 29 October 2021

#### References

- Sinha, R. & Shukla, P. Current trends in protein engineering: updates and progress. *Curr. Protein Pept. Sci.* **20**, 398–407 (2019).
- Dou, J. et al. De novo design of a fluorescence-activating b-barrel. *Nature* **561**, 485–491 (2018).
- Stoddard, B. L. in *Computational Design of Ligand Binding Proteins*, Vol. 1414 (ed. Stoddard, B. L.) v–ix (Humana Press, 2016).
- Tinberg, C. E. et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013).
- Correnti, C. E. et al. Engineering and functionalization of large circular tandem repeat protein nanoparticles. *Nat. Struct. Mol. Biol.* **27**, 342–350 (2020).
- Mohan, K. et al. Topological control of cytokine receptor signaling induces differential effects in hematopoiesis. *Science* **364**, eaav7532 (2019).
- Silva, D. A. et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
- Hannan, A. J. Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* **19**, 286–298 (2018).
- Jernigan, K. K. & Bordenstein, S. R. Tandem-repeat protein domains across the tree of life. *PeerJ* **3**, e732 (2015).

10. Kajava, A. V. Tandem repeats in proteins: from sequence to structure. *J. Struct. Biol.* **179**, 279–288 (2012).
11. Paladin, L. et al. RepeatsDB in 2021: improved data and extended classification for protein tandem repeat structures. *Nucleic Acids Res.* **49**, D452–D457 (2021).
12. Park, K. et al. Control of repeat-protein curvature by computational protein design. *Nat. Struct. Mol. Biol.* **22**, 167–174 (2015).
13. Doyle, L. et al. Rational design of alpha-helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 (2015).
14. Fallas, J. A. et al. Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).
15. Bradley, P. Structural modeling of TAL effector-DNA interactions. *Protein Sci.* **21**, 471–474 (2012).
16. Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716–719 (2012).
17. Koga, N. et al. Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
18. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
19. Mejias, S. H., Sot, B., Guantes, R. & Cortajarena, A. L. Controlled nanometric fibers of self-assembled designed protein scaffolds. *Nanoscale* **6**, 10982–10988 (2014).
20. Antanasijevic, A. et al. Structural and functional evaluation of de novo-designed, two-component nanoparticle carriers for HIV Env trimer immunogens. *PLoS Pathog.* **16**, e1008665 (2020).
21. Diamante, A. et al. Engineering mono- and multi-valent inhibitors on a modular scaffold. *Chem. Sci.* **12**, 880–895 (2021).
22. Ueda, G. et al. Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *Elife* **9**, e57659 (2020).
23. Wrapp, D. et al. Structural basis for potent neutralization of betacoronaviruses by single-domain camelid antibodies. *Cell* **181**, 1004–1015.e15 (2020).
24. Moore, R., Chandras, A. & Bleris, L. Transcription activator-like effectors: a toolkit for synthetic biology. *ACS Synth. Biol.* **3**, 708–716 (2014).
25. Nishanth, M. J. & Simon, B. Functions, mechanisms and regulation of Pumilio/Puf family RNA binding proteins: a comprehensive review. *Mol. Biol. Rep.* **47**, 785–807 (2020).
26. Perez-Riba, A. & Itzhaki, L. S. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Curr. Opin. Struct. Biol.* **54**, 43–49 (2019).
27. Rowling, P. J., Sivertsson, E. M., Perez-Riba, A., Main, E. R. & Itzhaki, L. S. Dissecting and reprogramming the folding and assembly of tandem-repeat proteins. *Biochem Soc. Trans.* **43**, 881–888 (2015).
28. Tran, H. M. et al. Nanomaterials for treating bacterial biofilms on implantable medical devices. *Nanomaterials* **10**, 2253 (2020).
29. Bauler, P., Huber, G., Leyh, T. & McCammon, J. A. Channeling by proximity: the catalytic advantages of active site colocalization using Brownian dynamics. *J. Phys. Chem. Lett.* **1**, 1332–1335 (2010).
30. Castellana, M. et al. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.* **32**, 1011–1018 (2014).
31. Park, H. et al. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).
32. Marcos, E. et al. De novo design of a non-local beta-sheet protein with high stability and accuracy. *Nat. Struct. Mol. Biol.* **25**, 1028–1034 (2018).
33. Mak, A. N., Lambert, A. R. & Stoddard, B. L. Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure* **18**, 1321–1331 (2010).
34. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
35. Otwinowski, Z. & Minor, W. in *Methods in Enzymology*, Vol. 276 (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic Press, 1997).
36. McCoy, A. J. et al. PHaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
37. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. Sect. D.* **66**, 213–221 (2010).
38. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr.* **D66**, 486–501 (2007).
39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240–255 (1997).
40. Schrödinger, L. The PyMol Molecular Graphics System. 1.2r3pre edn (PyMOLThe PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, 2020).
41. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
42. Crawford, K. H. D. et al. Protocol and reagents for pseudotyping lentiviral particles with SARS-CoV-2 spike protein for neutralization assays. *Viruses* **12**, 513 (2020).

## Acknowledgements

This work was supported by the National Institutes of Health (NIH R01 GM123378 and NIH R01 GM115545) and the Fred Hutchinson Cancer Research Center. We thank members of the Strong and Bloom labs at the Fred Hutchinson Cancer Center for research materials and advice.

## Author contributions

All computational design work was conducted by P.B. and M.A.K. All constructs described in this manuscript were expressed and purified by J.P.H., L.A.D., B.T., B.L.S., and M.G. Crystallographic analyses of tTRP9 and tTRP9<sub>3</sub> were conducted by J.P.H.; single-particle electron microscopy analyses of tTRP24<sub>8</sub> and tTRP24<sub>8</sub>SS were conducted by B.W.S. Surface plasmon resonance analyses were conducted by D.F.; viral neutralization and biolayer interferometry studies were conducted by J.M.R. and M.G. P.B. and B.L.S. wrote the initial draft of the manuscript, which was then rewritten and finalized through the joint efforts of all the coauthors.

## Competing interests

P.B. and B.L.S. are named inventors of the designed protein nanoparticles described in this work, which are described in IP filings by the Fred Hutchinson Cancer Research Center and are currently being licensed for commercialization. M.M.G. and J.M.R. are employees of Lumen Bioscience Inc., where studies of cTRP functionalization for potential commercial applications have been conducted. All other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02766-y>.

**Correspondence** and requests for materials should be addressed to Philip Bradley or Barry L. Stoddard.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Mark Collins and Anam Akhtar.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021