Thomas, F., Dawson, W., Lang, E., Burton, A., Bartlett, G., Rhys, G., ... Woolfson, D. (2018). De Novo-Designed -Helical Barrels as Receptors for Small Molecules. *ACS Synthetic Biology*, *7*(7), 1808-1816. https://doi.org/10.1021/acssynbio.8b00225

De novo designed α-helical barrels as receptors for small molecules

Franziska Thomas,[a,b‡] William M. Dawson,[a‡] Eric J.M. Lang,[a,c] Antony J. Burton,[a,d] Gail J. Bartlett,[a] Guto G. Rhys,[a] Adrian J. Mulholland,[a,c,e] and Derek N. Woolfson[a,c,f]*.

[a]School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK.

[b]Institute of Organic and Biomolecular Chemistry, Georg-August-Universität Göttingen, Tammannstr. 2, 37077 Göttingen, Germany.

[c]BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK.

[d]Frick Chemistry Laboratory, Princeton, NJ 084544, USA.

[e]Centre for Computational Chemistry, School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK.

[f]School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Brisol, BS8 1TD, UK

ABSTRACT: We describe *de novo* designed α-helical barrels (αHBs) that bind and discriminate between lipophilic biologically active molecules. αHBs have five or more α helices arranged around central hydrophobic channels the diameters of which scale with oligomer state. We show that pentameric, hexameric and heptameric αHBs bind the environmentally sensitive dye, 1,6-diphenyl-hexatriene (DPH) in the μM range and fluoresce. Displacement of the dye is used to report the binding of non-fluorescent molecules: palmitic acid and retinol bind to all three αHBs with sub-μM inhibitor constants; farnesol binds the hexamer and heptamer; but β-carotene only binds to the heptamer. A co-crystal structure of the hexamer with farnesol reveals oriented binding in the centre of the hydrophobic channel. Charged side chains engineered into the lumen of the heptamer facilitate binding of polar ligands: a glutamate variant binds a cationic variant of DPH; and introducing lysine allows binding of the biosynthetically important farnesol diphosphate.

INTRODUCTION

Protein design has advanced sufficiently that it is now possible to generate successfully a variety of stable protein structures from first principles.[1-2] This can be done using rules of thumb that relate protein sequence to structure or by employing computational methods.[3-5] New challenges for this field of *de novo* protein design include: (1) taking forays into the so-called dark matter of protein space, *i.e.*, designing entirely new protein structures not observed in nature;[6-7] (2) making the design methods open and accessible to others, particularly to non-specialist users;[8] and (3) building on these scaffolds and methods to deliver functional *de novo* proteins.[9-10] Desired functions include biomolecular recognition and sensing, and catalysis or enzyme-like activities.[11-12] A requirement for many such functions is that the designed protein scaffolds bind and discriminate between various molecules, *e.g.*, bioactive small molecules and biological macromolecules.

In biological systems, protein-ligand dissociation constants range from mM to fM.[13] Moreover, many receptors respond to multiple ligands across a range of binding affinities. One advantage of the one receptor-multiple ligand model is that it removes the need to synthesize a specific receptor for every possible protein-ligand interaction. The olfactory network provides a natural example of such a system.[14] In humans, genes encoding *circa* 400 G-protein coupled receptors are able to distinguish up to 1 trillion different stimuli through low affinity and relatively non-selective recognition of odorants.[15]

The design and engineering of new proteins that recognize and bind peptides and folded proteins has achieved considerable success, particularly using directed evolution to modify either natural protein folds or *de novo* scaffolds generated by consensus design.[4, 12] However, the recognition and binding of small molecules has proven more challenging.[16-18] Two related issues here are the generation of scaffolds that act as good receptors for small molecules, and then the embellishment of these to discriminate between what are often very similar molecules. The small sizes and limited functional groups of the target molecules often confound attempts to recognize, bind and distinguish them. One solution is to adapt natural proteins with cavities already evolved to bind small-molecule ligands.[19-20] Here, we add *de novo* proteins with central accessible channels to this repertoire of small-molecule-binding proteins. These can be adapted for the recognition, binding and release of ligands.

Our designs are based on *de novo* α-helical coiled coils. The vast majority of coiled coils are bundles with 2 – 4 helices wrapped around a central superhelical axis leading to solid hydrophobic cores.[21] These structures are directed by underlying sequence repeats of hydrophobic (***h***) and polar (***p***) residues, ***hpphppp***, often denoted ***abcdefg***. The 3,4-hydrophobic repeats drive the folding of amphipathic α helices, which assemble into bundles to bury their hydrophobic, ***a***+***d*** faces. The helices supercoil around each other because the 3.5-residue sequence repeat and the 3.6-residue structural repeat do not match precisely. These relatively straightforward structural principles, and the sequence-to-structure relationships that have followed, have made coiled coils attractive targets for protein design.[3] This has resulted in myriad *de novo* coiled-coil homo-dimers, trimers and tetramers and reliable sets of *de novo* hetero-dimeric and trimeric systems.[22]

Apart from a small number of examples,[23-24] however, these traditional coiled-coil dimers – tetramers do not have suitable internal cavities or central channels to provide a basis for ligand

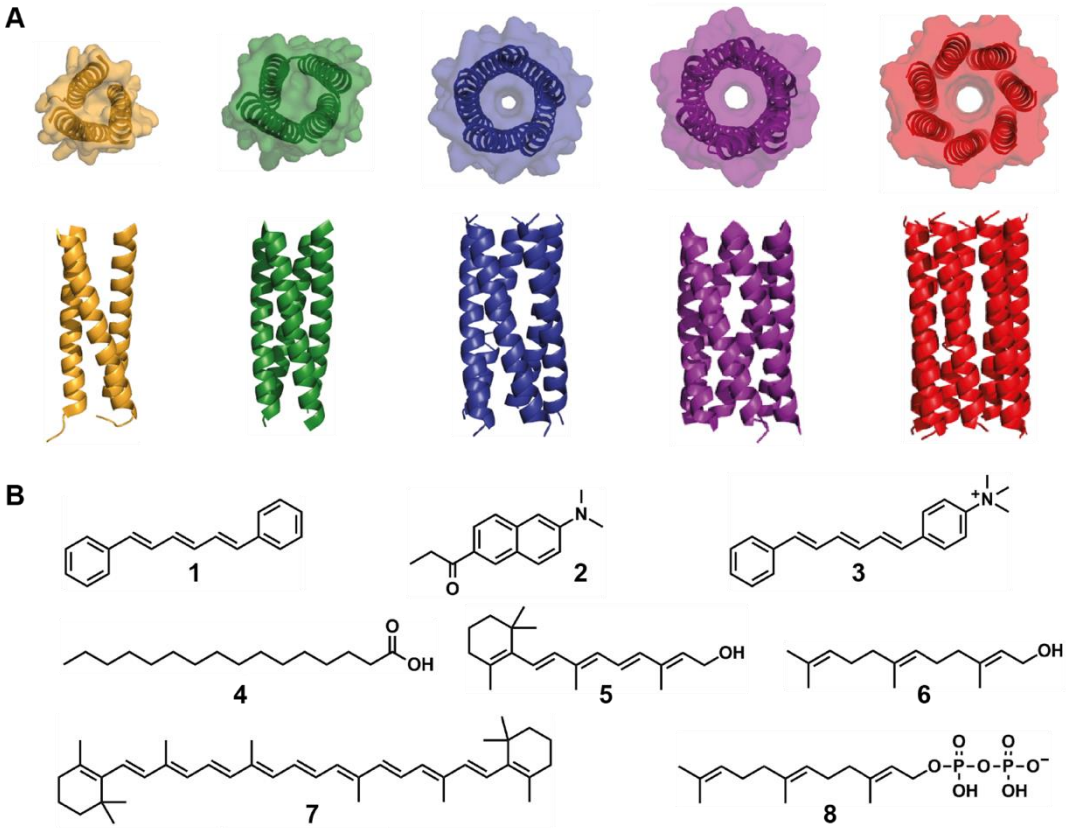

Figure 1. Structures of the coiled-coil assemblies and small molecules investigated herein. (A) Orthogonal views of the space-filled models (top) and ribbon diagrams (bottom) of the X-ray structures of CC-Tri (orange; PDB code 4dzl), CC-Tet (green; PDB code 3r4a), CC-Pent (blue; PDB code 4pn8), CC-Hex2 (purple; PDB code 4pn9), and CC-Hept (red; PDB code 4pna). (B) Hydrophobic molecules tested for encapsulation into the hydrophobic channels of the *de novo* designed αHBs: **1**, DPH; **2**, Prodan; **3**, TMA-DPH; **4**, palmitic acid; **5**, retinol; **6**, farnesol; **7**, β-carotene; **8**, farnesyl diphosphate (FPP).

binding. To make coiled coil-based receptors for small molecules requires oligomer states of five and above, because these are α-helical barrels (αHBs) with central, accessible channels.[25] Generally, these have sequence patterns that expand the traditional heptad repeat, *e.g.* ***hpphhph*** and others.[26] Both natural and engineered αHBs are known but they are rare.[27] For example, the oligomerization domain of cartilage oligomeric matrix protein (COMPcc), a pentamer with a 7 Å diameter central channel, sequesters hydrophobic molecules such as fatty acids, vitamins and cholesterol.[27-29] Turning to engineered and designed proteins: the pentameric Trp-zipper has a cavity that binds polyethylene glycol;[30] the serendipitously discovered hexamer, CC-Hex,[31] binds a small-molecule dye;[32] and a mutant of the GCN4 leucine zipper peptide, GCN4-pAA, forms a heptameric spiral of helices[33] with a central cavity that also binds hydrophobic dyes.[34] There are other examples of coiled-coil oligomers above pentamer but, to our knowledge, these either do not have central channels or their binding/transport properties have not been fully characterised.[35-37] Indeed, adding specificity of binding for small molecules has not been addressed formally by any of the above reports.

The coiled coil is a parameterizable protein fold;[38] *i.e.*, there are equations that allow coiled-coil structures to be built *ab initio* making them accessible to computational design.[25, 39-44] We have combined parametric design and the use of ***hpphhph***-type sequences to deliver a set of *de novo* αHBs from pentamer to heptamer and characterized these through to X-ray protein crystal structures.[25] These highly stable assemblies have melting temperatures above 95 °C,[25] and, by analogy to *de novo* coiled-coil dimers – tetramers, we estimate the dissociation constants of the αHBs to be in the sub-nM regime.[22, 45] Furthermore, these αHBs are robust to mutation allowing the engineering of coiled coil-based nanotubes,[32, 46] and the generation of variants with engineered lumens that perform rudimentary catalysis.[9]

Here we describe how these computationally designed αHBs can be adapted to render *de novo* protein receptors that bind and discriminate between a panel of biologically important lipophilic small molecules. The internal shape and diameter of the αHBs, which ranges from ≈ 5 – 7 Å, provides a first level of discrimination: the different barrels bind different targets with a range of affinities from sub-μM upwards. Further discrimination is achieved by adding charged side chains to the lumens to direct the binding of lipophilic molecules with formally charged headgroups. We combine experimental spectroscopic and high-resolution structural studies with molecular modelling and dynamics simulations to provide insight into the modes of binding and how these systems might be engineered for applications in biotechnology and synthetic biology.

RESULTS and DISCUSSION

Different αHBs have binding abilities. We chose three αHBs from the available computationally design structures,[25]

Table 1. Dissociation constants and inhibition constants of small molecules binding to the αHBs.

	CC-Pent		CC-Hex2		CC-Hept	
	K_D [μM]	K_I [μM]	K_D [μM]	K_I [μM]	K_D [μM]	K_I [μM]
DPH (**1**)	4.5 ± 0.9	-	1.6 ± 0.2	-	1.3 ± 0.3	-
Prodan (**2**)	401 ± 14	-	6.5 ± 1.1	-	5.8 ± 0.6	-
TMA-DPH (**3**)	15.5 ± 0.9	-	14.8 ± 1.4	-	13.3 ± 1.3	-
Palmitic acid (**4**)	-	0.8 ± 0.3	-	0.7 ± 0.2	-	0.5 ± 0.2
Retinol (**5**)	-	10.2 ± 2.8	-	2.9 ± 1.3	-	2.5 ± 0.4
Farnesol (**6**)	-	23.9 ± 2.4	-	0.6 ± 0.2	-	0.4 ± 0.1
β-Carotene (**7**)	-	nd[1]	-	74 ± 27	-	5.9 ± 3.0

[1]A K_I could not be determined because of weak binding of β-carotene to CC-Pent.

namely CC-Pent, CC-Hex2, and CC-Hept (Fig 1A). The sequences of these hyperstable assemblies are similar (Table S2, Supplementary Fig 1), and the Ile/Leu-lined channels differ only in diameter, which increases through the series CC-Pent to CC-Hept.[25] To quantify this, the three X-ray crystal structures were analyzed by Pore Walker (Supplementary Fig 2 and Supplementary Table 3),[47] which revealed variations in the diameters of the channels in CC-Pent, CC-Hex2 and CC-Hept of 3.0 Å – 7.4 Å, 4.7 Å – 7.7 Å, and 5.4 Å - 10.1 Å, respectively.

Channels of similar chemistry and dimensions bind hydrophobic dyes, including the rod-shaped 1,6-diphenylhexatriene (DPH; **1**, Fig 1B) and the more sterically demanding Prodan (**2**, Fig 1B).[32, 34, 46, 48] We reasoned that differences in channel diameter between CC-Pent, CC-Hex2 and CC-Hept should differentiate binding of the same molecule, which could lead to specific binding of particular hydrophobic molecules or molecular classes (Fig 1B). To explore this *in silico*, we built models with the dyes docked into the channels of the X-ray crystal structures of the three αHBs and minimized the energy of the complexes in Gromacs (Supplementary Fig 3). All three barrels accommodated DPH, whereas Prodan only fitted into the channels of CC-Hex2 and CC-Hept.

We used fluorescence spectroscopy to probe binding of the two dyes to the αHBs (Fig 2) as the fluorescence of both dyes is environmentally sensitive:[49-50] DPH only fluoresces (λ_{max} 455 nm) when in hydrophobic environments; and the emission fluorescence spectrum of Prodan is enhanced and blue-shifted (from 520 nm to 424 nm) when transferred to apolar surroundings. As negative controls for any background binding, we used a *de novo* coiled-coil trimer (CC-Tri) and a *de novo* tetramer (CC-Tet), which do not have channels or pores (Fig 1A).[22] Neither of the dyes fluoresced with either control over peptide concentrations of 10 μM – 300 μM (Figs 2A-D). By contrast, 50 μM concentrations of all three αHBs fluoresced with 1 μM DPH (Fig 2A). Prodan bound to CC-Hex2 and CC-Hept, but not to the smaller CC-Pent (Fig 2B).

Binding was quantified by titrating increasing amounts of the αHB peptides into fixed, 1 μM concentrations of DPH or Prodan (Figs 2C-D). The resulting saturation-binding curves fitted to single-site binding models to return dissociation constants (K_D; Table 1): all three αHBs bound the rod-shaped DPH with low μM K_Ds; Prodan bound similarly to CC-Hex2 and CC-Hept, but weakly to CC-Pent.

In addition, sedimentation velocity analytical ultracentrifugation experiments for CC-Pent, CC-Hex2 and CC-Hept in the presence of DPH could be followed directly by monitoring the absorbance of DPH at 350 nm. Thus, the peptides and dye co-sediment indicating that the two are bound together, which we interpret as the dye being encapsulated by the αHBs. Moreover, analysis of the sedimentation data confirmed that the oligomeric states of the assemblies were unaffected by the presence of the small molecule (Supplementary Fig 4).

αHB receptors induce chirality in the encapsulated dyes. DPH and Prodan are flat, achiral molecules. Nonetheless, we reasoned that the chiral environment of the protein channels might induce a circular dichroism (CD) effect.[51] In free solution, although both molecules absorbed strongly in the near-UV neither had associated CD spectra (Figs 2E&F). However, when bound to DPH all three αHBs showed strong CD bands centered on ≈350 nm with vibrational fine structure (Fig 2E). Interestingly, these data showed a clear inverse correlation with αHBs pore size: the intensity of the CD signal decreased through the series CC-Pent to CC-Hex2 to CC-Hept (Fig 2E). However, CD spectra were only observed for Prodan with CC-Hex2 and CC-Hept (Fig 2F) consistent with the binding data from fluorescence spectroscopy.

Computational modelling sheds light on the modes of DPH binding to αHBs. MD simulations can identify binding sites and binding modes of small molecules in natural[52] and *de novo* proteins.[53] DPH was docked into the channels of the X-ray crystal structures for the three αHBs using AutoDock Vina (Fig 3A, Supplementary Fig 5A-C).[54] Three binding sites for each αHB were found: site 1, near the *C* terminus; a central site 2; and site 3 near the *N* terminus. Each site consists of two widened voids separated by a narrow section formed by the ring of Leu residues at the ***a*** position; and adjacent sites share a void (Supplementary Fig 5D). For each identified binding pose, the bulky phenyl rings are accommodated in the cavities whereas the thinner hexatriene moiety fits in the narrowed sections. For CC-Hept, the different poses were found to have similar binding scores (≤ 0.5 kcal/mol, as calculated by the Autodock vina scoring function), irrespective of the position of DPH in binding site 1, 2 or 3 (Supplementary Table 4). For CC-Hex2, similar scores were obtained for DPH in sites 1 and 2, but site 3 was predicted to be less favorable. For CC-Pent the three sites had markedly different binding scores with site 1 predicted to be the most favorable binding site followed by site 2 and then site 3. The lower score for site 3 can be rationalized as it is too small to fully accommodate the phenyl moiety of DPH, leaving it partially solvent exposed.

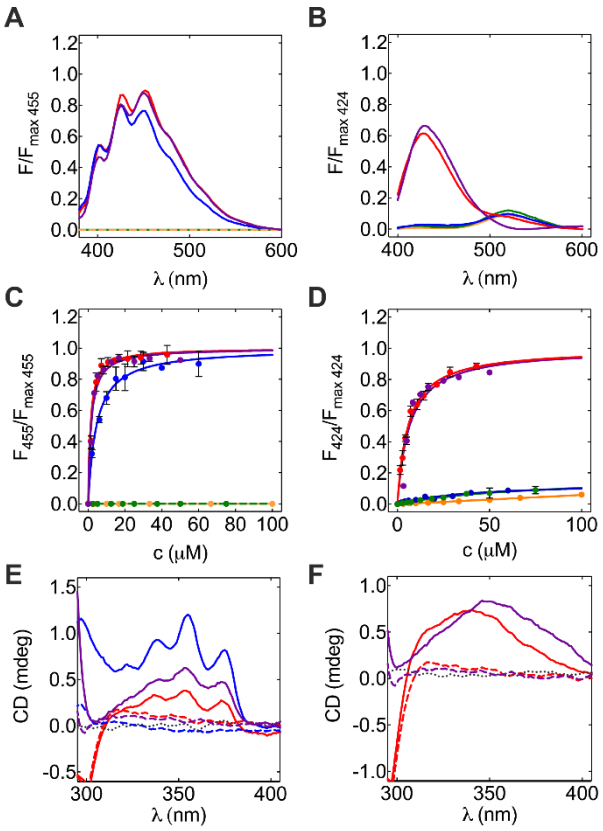

Figure 2. Fluorescence and CD spectroscopy of DPH (left) and Prodan (right) binding to αHBs. Emission fluorescence spectra for DPH (A) and Prodan (B) with coiled-coil peptides. Saturation binding curves of the peptides to DPH (C), and to Prodan (D). CD spectra of dye in the absence (black dotted line) and presence of αHBs (solid lines) for DPH (E), and Prodan (F), and of αHBs without dye (colored dotted lines). Color key: CC-Tri (orange), CC-Tet (green), CC-Pent (blue), CC-Hex2 (purple), CC-Hept (red). Conditions: (A&B) 50 μM peptide, 1 μM dye, HEPES buffered saline (HBS; 25 mM HEPES, 100 mM NaCl, pH 7.0), 5% v/v DMSO; (C&D) 1.4 μM – 100 μM of the coiled-coil assembly, 1 μM dye, HBS, 5% v/v DMSO; (E) 200 μM peptide, 5 μM DPH, HBS, 5% v/v DMSO; and (F) 200 μM peptide, 20 μM Prodan, HBS, 5% v/v DMSO.

To confirm that the binding sites identified were favorable and the docking poses stable, extensive molecular dynamics (MD) simulations were initiated starting from 5 poses for each site and running 200 ns of simulation per pose, leading to a total of 1 μs of simulation for each binding site. These indicated that ligand binding to each site was stable (Fig 3B-D, Supplementary Figs 6 & 7 and Supplementary Movie 1). For CC-Pent and CC-Hept the overall protein structures did not change significantly from the X-ray crystal structures with or without DPH. This was quantified by backbone root mean square deviations (RMSDs) through the trajectories (Supplementary Fig 8), which reached maxima ≤ 2 Å in both cases. By contrast, the average RMSDs for CC-Hex2 for both the apo and DPH-bound forms were >3 Å, suggesting that the solution-phase structure might deviate from the crystal structure. We attribute this to a change in superhelix of CC-Hex2, which relaxed to a higher pitch over the MD trajectories (Supplementary Fig 9) leading to a longer, narrower channel of diameter 3 Å – 6 Å calculated by HOLE (Supplementary Fig 10).[55] A similar analysis for CC-Pent revealed that the presence of DPH lead to an expansion of the channel around each binding site to accommodate the ligand when compared with the apo simulations. This is especially marked in the narrowed sections where the diameter of the channel increase by up to 1.5 Å. In addition, this expansion to accommodate DPH increases the overall curvature of the helices in the central region of the barrel (Supplementary Fig 9). In contrast, CC-Hex2 and CC-Hept showed a subtle tightening in response to the ligand (Supplementary Fig 10). Presumably, this improves van der Waals' contacts between protein and ligand.

Turning to the dynamics of the DPH molecule (Figs 3B-D, supplementary Figs 6 & 11): at site 1 of CC-Pent, CC-Hex2 and CC-Hept DPH was stable in its starting binding position; whereas, at sites 2 and 3 of CC-Hept we observed DPH moving between these sites in several trajectories. This suggests a relatively low energy barrier between these two sites of CC-Hept compared to the barrier between site 1 and 2, and between any adjacent sites in CC-Pent and CC-Hex2. Moreover, the central site 2 was preferred over site 3 in these transitions in CC-Hept indicating a lower binding energy of DPH in site 2. Detailed analysis of these events showed that the transition itself was fast (between 10 and 20 ps) and did not noticeably perturb the protein structure (Supplementary Fig 12). The preferred positions

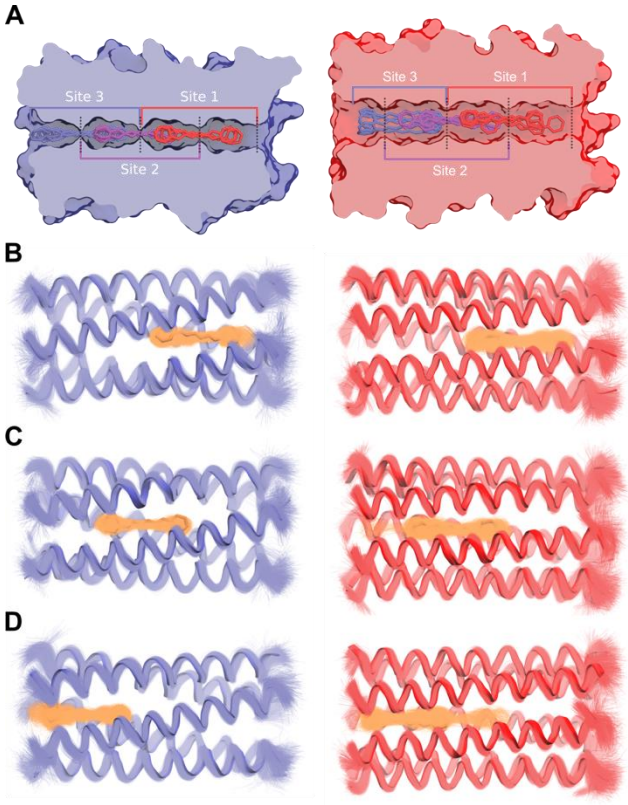

Figure 3. Docking and molecular dynamics simulations of DPH bound to CC-Pent (blue) and CC-Hept (red). (A) Binding poses of DPH obtained from docking for each identified binding site (5 poses per site). Superpositions of snapshots from the ensuing MD simulations starting with DPH docked in site 1 (B), site 2 (C) and site 3 (D) of each αHB. The snapshots (fine ribbons and sticks) are for every 10 ns of 1 μs trajectories. The thick ribbons and sticks correspond to a representative conformation of each of the trajectories. The proteins are oriented with the *C* termini of the peptides on the right.

for the phenyl groups were the widened voids, even for CC-Hept that has a larger channel than the other αHBs. This suggests that less-sterically-demanding linear chains are better accommodated in the narrowed sections of the channels, whereas bulkier groups reside more favorably in the wider voids, adding a new level of selection to the binding modes of small molecules inside αHBs.

For CC-Hept, the ligand also moved in the plane orthogonal to the long-axis of the channel axis and rotated freely, presumably due to the larger diameter of its channel (Supplementary Figs 13 & 14). Consistent with the tighter channels, these motions were restricted for CC-Pent and CC-Hex2, and changes in ligand position or orientation were discrete, infrequent events, especially in the case of CC-Hex2 site 2 and CC-Pent site 2 and 3. Together, these simulations reveal increased restrictions in ligand binding and movement as channel diameter shrinks in the series CC-Hept, CC-Hex2, CC-Pent. It is interesting that this correlates with in the increased intensity of induced CD signal of DPH.

αHB bind biological lipophilic molecules. To explore if αHBs are able to sequester other rod-shaped, lipophilic molecules, we tested the binding of palmitic acid (**4**), and the terpenoids retinol (**5**), farnesol (**6**) and β-carotene (**7**) (Fig 1). Palmitic acid plays roles in the synthesis and regulation of fatty acids, and is a common post-translational modification serving as a membrane anchor for protein trafficking.[56-57] The retinoids, retinol and retinoic acid, are involved in regulatory processes, including epithelial cell growth and cellular differentiation, and retinal (vitamin A) is found in the opsin protein class, *e.g.* rhodopsin.[58-59] Retinal cannot be synthesized by humans and is taken up as a retinylester or the provitamin β-carotene (**7**), which is cleaved enzymatically to give two retinal molecules.[60-61] Farnesol has anti-fungal properties and anti-tumor and chemopreventative effects in animal models, which is mainly due to it inhibiting cell growth and proliferation and inducing apoptosis.[62-63]

The binding propensities of these molecules to the αHBs were determined from a displacement assay as follows. In place of DPH, the more-water-soluble trimethylammonium-DPH (TMA-DPH; 3, Fig 1) was used as the fluorescent reporter. This bound all three αHBs with K_Ds of 13 μM - 16 μM (Table 1 and Supplementary Fig 15 & Table 5). αHBs were equilibrated for 2 hours with TMA-DPH in approximately 25-fold excess at 20 °C and titrated with the analytes **4** – **7**. Decreases in fluorescence with increased analyte concentration were fitted to a one-site competition model to obtain IC_{50} values, which were converted to the inhibition constants (K_I; Equation S3).

There was a clear trend in the data across the panel of analytes (Fig 4A, Table 1). The most flexible n-alkyl chain of palmitic acid (**4**) was sequestered equally by all three αHBs with sub-μM K_I values. In contrast, αHBs bound the unsaturated ligands differentially as follows. Retinol was bound by CC-Hex2 and CC-Hept with $K_I \approx 3$ μM, but binding to CC-Pent was three times poorer (Fig 4B). The αHBs discriminated farnesol with sub-μM inhibition constants to CC-Hept and CC-Hex2, respectively, but no appreciable binding to CC-Pent (Fig 4C). Finally, for the largest ligand, β-carotene (**7**), only binding to CC-Hept could be quantitated (Table 1, Fig 4D).

We offer the following explanation for the differences in binding of the three terpenoids. First, compared with palmitic acid, which is not discriminated by the αHBs at all, the partially unsaturated and methylated chain of farnesol results in selectivity of binding for the larger barrels (CC-Hex2 and CC-Hept) and against CC-Pent. Second, the bulky cyclohexene group that terminates retinol lowers the binding efficiency to all αHBs. Third, the two cyclohexene head groups of β-carotene inhibit binding to the two smaller barrels (CC-Pent and CC-Hex2).

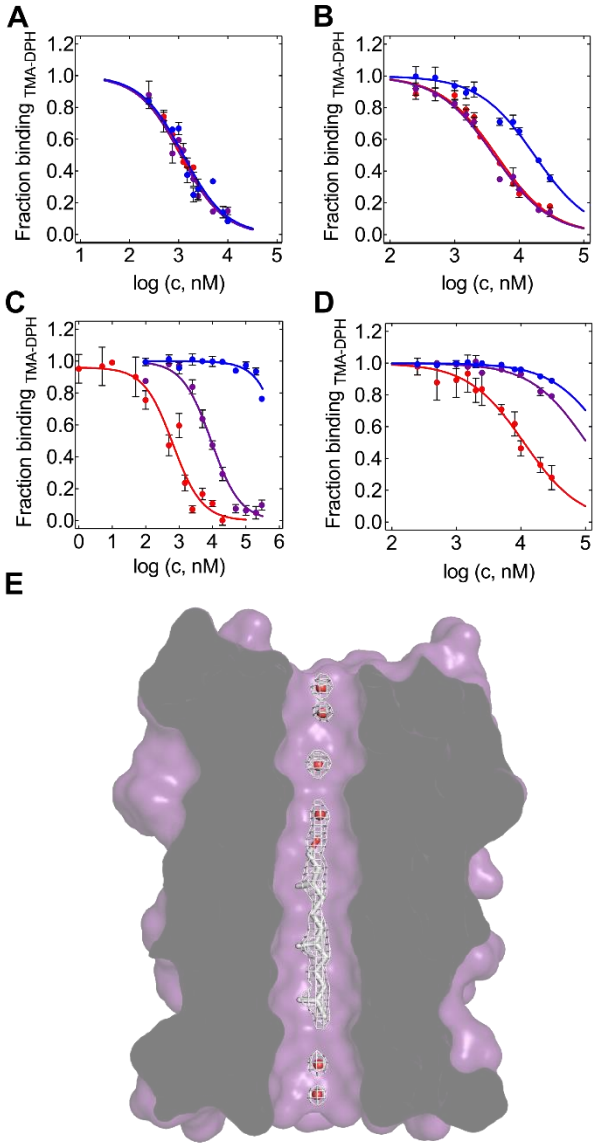

Figure 4. Binding of various lipophilic biomolecules to αHBs. Fluorescence-based displacement assays using TMA-DPH as the reporter in complex with CC-Pent (blue), CC-Hex2 (purple) or CC-Hept (red), and titrated against palmitic acid (**4**, A), retinol (**5**, B), farnesol (**6**, C), and β-carotene (**7**, D). (E) 1.85 Å X-ray crystal structure of farnesol bound in the channel of CC-Hex2 (PDB: 6EIZ). Electron density $2F_o$-F_{calc} for the ligand is shown at 1.4 σ. Conditions for A - C: 50 μM TMA-DPH, 1.4 - 2 μM coiled-coil assembly, HBS, 5% v/v DMSO, 20 ˚C, concentration of ligands varied between 0.1 and 500 μM. D: As A – C with addition of 4% THF.

A CC-Hex2: farnesol complex elucidated by X-ray crystallography. To define the binding in atomic detail, we screened crystal soaks and found that CC-Hex2 co-crystallized with far-

nesol. A 1.85 Å resolution X-ray crystal structure for the complex was solved by molecular replacement using apo-CC-Hex2 (PDB code 4pn9)[25] as the search model (Supplementary Table 6, Fig 4E, and Supplementary Fig 16). This confirmed the single-molecule binding mode assumed above, and revealed farnesol bound approximately in the middle of the channel. Moreover, and in contrast to the many other X-ray crystal structures and MD simulations of αHBs that we have solved and performed, there was density consistent with ordered hydrogen-bonded water molecules connecting the hydroxyl functionality of farnesol to bulk solvent (Fig 4E).

The lumens of αHBs can be engineered to accept charged molecules. Hydrophilic and charged side chains can be introduced into the otherwise exclusively hydrophobic lumens of *de novo* αHBs without major structural changes.[9, 31] For instance, a robustly folded CC-Hept variant with three mutations per chain, CC-Hept-Cys-His-Glu, has been designed and shows rudimentary catalytic activity in ester hydrolysis.[9] We reasoned that ionizable residues could be included to help sequester charged analytes into the channels. To test this, we chose the larger scaffold, CC-Hept as it should tolerate charged side chains and be receptive to larger, more-complex molecules. Specifically, we introduced lysine and glutamate residues to bind anions and cations, respectively. We modified positions Ile-17 and Ile-24 and found that Ile-24 of CC-Hept was the most robust towards substitution for Lys or Glu—to give CC-Hept-I24K and CCHept-I24E—without compromising the structure or stability of the αHB (Supplementary Table 2, and Supplementary Figs 17-22). We obtained an X-ray crystal structure for the latter (Supplementary Fig 17).

Binding experiments with cationic TMA-DPH demonstrated that the two variants discriminated this ligand as expected: the ligand was relatively weakly bound by both CC-Hept and CC-Hept-I24K (K_D's of 13.3 ± 1.3 μM and 12.8 ± 0.6 μM, respectively, Fig 5A); whereas, binding to CC-Hept-I24E was almost an order of magnitude tighter at 1.6 ± 0.1 μM (Fig 5A, purple trace). Using the TMA-DPH displacement assay, CC-Hept-I24K and CC-Hept-I24E accommodated neutral farnesol with K_Is of 1.8 ± 0.3 μM and 1.0 ± 0.04 μM, respectively (Fig 5B).

Finally, we tested the binding of the anionic and biologically important farnesyl diphosphate (**8**, FPP). FPP is the precursor of sesquiterpenoids, including the antimalarial drug artemisinin and squalene the precursor of sterols;[64] and it is a substrate of protein prenyltransferases involved in trafficking of proteins to membranes.[65] Using the displacement assay, we found that the exclusively hydrophobic channel of CC-Hept and the negatively charged channel of CC-Hept-I24E bound FPP only weakly (Fig 5C). By contrast, FPP bound to CC-Hept-I24K greater than ten times tighter with a measurable K_I of 1.5 ± 0.2μM (Fig 5C).

These data for FPP are most probably explained by binding being driven by its hydrophobic tail displacing TMA-DPH from the channels of CC-Hept and CC-Hept-I24E, and leaving the charged diphosphate headgroup outside of the channel. With CC-Hept-I24K the latter could also be accommodated within the channel by the positively charged hepta-lysine. We probed this *in silico* by docking charged FPP in the channel of a model for CC-Hept-I24K. Three binding poses were selected and a 100 ns constant-pH MD (CpHMD) simulation at pH 7.0 was run for each system (Figs 5D & E and Supplementary Movie 2). In the lowest-energy binding poses the pyrophosphate group interacted with the ring of lysine residues and the hydrophobic tail was encapsulated by the larger hydrophobic section of the channel (Supplementary Figs 23 and 24). For the CpHMD simulations, FPP was kept fully deprotonated and the protonation states of the Lys-24 residues were allowed to vary. For approximately half of the simulation, 5 of the Lys residues were protonated, and for the remainder either 6, 3 or 7 lysine residues (ranked by decreasing prevalence) were charged (Supplementary Fig 25). The charged FPP remained in a stable position throughout the course of the simulations (Supplementary Fig 26), with the hepta-Lys ring accommodating the 3 negative charges of the pyrophosphate through strong Coulombic interactions. However, the Lys side chains were mobile (Supplementary Fig 27) and not all interacted with FPP throughout the simulation (Supplementary Fig 24); but the protein backbone remained stable (Supplementary Fig 28).

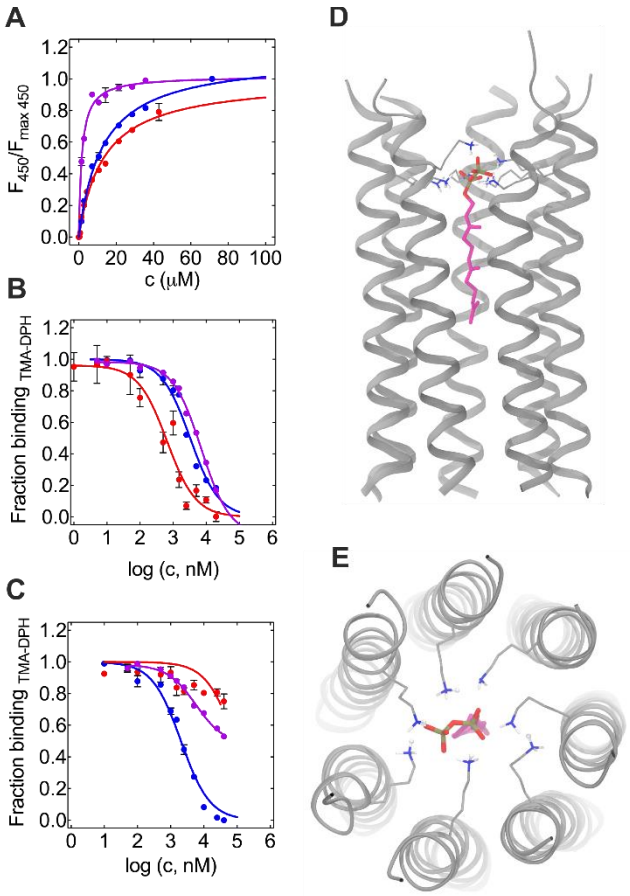

Figure 5. Binding of bioactive hydrophobic molecules to engineered CC-Hept channels. (A) Saturation binding of TMA-DPH to CC-Hept (red), CC-Hept-I24K (blue) and CC-Hept-I24E (purple). Conditions: 1.4 – 100 μM coiled-coil assembly, 1 μM dye, HBS, 5% v/v DMSO. (B&C) Fluorescence-based competition assays with TMA-DPH bound to CC-Hept, CC-Hept-I24K, and CC-Hept-I24E (colour key as above) being displaced by farnesol (B) or farnesol diphosphate (C). Conditions: 50 μM TMA-DPH, 1.4 μM coiled-coil assembly, HBS, 5% v/v DMSO, 20 ˚C. Concentration of ligands varied between 0.1 and 500 μM. (D&E) Orthogonal views of a pose from the MD simulation of FPP bound to CC-Hept-I24K. FPP is shown in sticks with the carbon atoms colored magenta; Lys-24 residues are represented by sticks with spheres for the titratable protons.

CONCLUSION

We have demonstrated the ability to sequester lipophilic and charged molecules with varying affinities into the central lumens of a series of *de novo* designed α-helical barrel (αHB) scaffolds including a pentamer, a hexamer and a heptamer. For the first-generation αHBs, the lumens are lined exclusively by hydrophobic residues. Consistent with this, all of the barrels reversibly bind the environment-sensitive hydrophobic dye diphenylhexatriene (DPH) to give signals in both fluorescence and CD spectroscopy. The former provides a reporter in a displacement assay for the binding of non-fluorescent analytes to αHBs. The varying internal diameters of the structures, which are in the range 5 Å – 7 Å, provide a first-order discriminating factor for differential binding of a small panel of lipophilic small biomolecules, which includes palmitic acid, retinol, farnesol and β-carotene.

The different binding modes and affinities of the various molecules to the αHBs can be rationalized by the physical properties of the small molecules, and through molecular modelling and dynamics simulations. In particular, a second discrimination factor was identified as the internal shape of the channels which alternate widened voids separated by narrower sections. Although bulky groups can be accommodated within the narrowed sections, binding there is less favorable and transient. Linear unbranched chains are however stable in these tight sections. Consequently, the length of the cavities and the distance between narrowed sections may influence the type of small molecule that can be sequestrated by the αHBs in the future.

Secondary binding features can be added to the basic αHB scaffolds to tailor binding and selection. Specifically, we have incorporated rings of both negatively and positively charged side chains within the lumen of the heptamer to allow the selective binding of cationic and anionic ligands, respectively. Whilst by no means a fully comprehensive set of mutations, this highlights the versatility and potential of this system for the design of ligand binding sites. The αHBs and the principles that we advance provide a strong basis for the design and construction of specific *de novo* receptor proteins for the recognition, binding, and release of bioactive small molecules using water-soluble αHBs and αHB-based materials, and to develop membrane-spanning transporters and sensors.

In respect of applications, naturally occurring and engineered αHB coiled-coil scaffolds—*e.g.*, COMPcc, GCN4-pAA—bind and encapsulate small molecules.[27-29, 48] However, the *de novo* designed systems described herein have important differences and potential advantages: 1) By changing the lengths and oligomer states of the peptides, the lengths and diameters of the hydrophobic channels can be tuned to accommodate specific target molecules or classes of molecule. 2) As the parent designs are thermostable and have low-complexity repeat sequences, the *de novo* αHBs tolerate a large number of mutations in their lumens, enabling polar and charged side chains to be incorporated and, in turn, functionalized molecules to be sequestered.

The demonstration of designed binding also adds to their emerging potential as *de novo* biocatalysts for specific substrates.[9] This variety, modularity and flexibility of the *de novo* αHB systems bode well for their application in biotechnology and synthetic biology as vehicles for small-molecule recognition, capture, storage and release and as the active components of new biosensor and diagnostics devices.

ASSOCIATED CONTENT

Supporting Information.

Experimental details and supplementary characterization, simulation and binding data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* d.n.woolfson@bristol.ac.uk

Author Contributions

‡These authors contributed equally.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

FT and DNW are grateful for funding by the Leverhulme Trust (RPG-2012-536). WMD, GJB and DNW are supported by a European Research Council Advanced Grant (340764). DNW, EJML and AJM are supported by the BBSRC and EPSRC through the BrisSynBio Synthetic Biology Research Centre (BB/L01386X1). AJM is supported by EPSRC grant number EP/M022609/1. AJB and GGR thank the Bristol Chemical Synthesis Centre for Doctoral Training funded by the Engineering and Physical Sciences Research Council (EP/G036764/1). DNW holds a Royal Society Wolfson Research Merit Award. We thank the University of Bristol School of Chemistry Mass Spectrometry Facility for access to the EPSRC-funded Bruker Ultraflex MALDI TOF/TOF instrument (EP/K03927X/1) and BrisSynBio for access to the BBSRC-funded BMG Labtech Clariostar Plate Reader (BB/L01386X/1).

REFERENCES

(1) Regan, L.; Caballero, D.; Hinrichsen, M. R.; Virrueta, A.; Williams, D. M.; O'Hern, C. S., *Biopolymers* **2015,** *104*, 334.

(2) Huang, P. S.; Boyken, S. E.; Baker, D., *Nature* **2016,** *537*, 320.

(3) Woolfson, D. N., *Subcell. Biochem.* **2017,** *82*, 35.

(4) Porebski, B. T.; Buckle, A. M., *Protein Eng. Des. Sel.* **2016,** *29*, 245.

(5) Coluzza, I., *J. Phys. Condens. Matter* **2017,** *29*.

(6) Taylor, W. R.; Chelliah, V.; Hollup, S. M.; MacDonald, J. T.; Jonassen, I., *Structure* **2009,** *17*, 1244.

(7) Woolfson, D. N.; Bartlett, G. J.; Burton, A. J.; Heal, J. W.; Niitsu, A.; Thomson, A. R.; Wood, C. W., *Curr. Opin. Struct. Biol.* **2015,** *33*, 16.

(8) Cooper, S.; Khatib, F.; Treuille, A.; Barbero, J.; Lee, J.; Beenen, M.; Leaver-Fay, A.; Baker, D.; Popovic, Z.; Players, F., *Nature* **2010,** *466*, 756.

(9) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N., *Nat. Chem.* **2016,** *8*, 837.

(10) Chevalier, A.; Silva, D.-A.; Rocklin, G. J.; Hicks, D. R.; Vergara, R.; Murapa, P.; Bernard, S. M.; Zhang, L.; Lam, K.-H.; Yao, G., *Nature* **2017,** *550*, 74.

(11) Zanghellini, A., *Curr. Opin. Biotechnol.* **2014,** *29*, 132.

(12) Pluckthun, A., *Annu. Rev. Pharmacol. Toxicol.* **2015,** *55*, 489.

(13) Kuriyan, J.; Konforti, B.; Wemmer, D., *The Molecules of Life: Physical and Chemical Principles*. Garland Science, Taylor & Francis Group: New York, 2013.

(14) Gaillard, I.; Rouquier, S.; Giorgi, D., *Cell. Mol. Life Sci.* **2004,** *61*, 456.

(15) Bushdid, C.; Magnasco, M. O.; Vosshall, L. B.; Keller, A., *Science* **2014,** *343*, 1370.

(16) Malisi, C.; Schumann, M.; Toussaint, N. C.; Kageyama, J.; Kohlbacher, O.; Hocker, B., *PLoS One* **2012,** *7*, e52505.

(17) Allison, B.; Combs, S.; DeLuca, S.; Lemmon, G.; Mizoue, L.; Meiler, J., *J. Struct. Biol.* **2014,** *185*, 193.

(18) Polizzi, N. F.; Wu, Y.; Lemmin, T.; Maxwell, A. M.; Zhang, S.-Q.; Rawson, J.; Beratan, D. N.; Therien, M. J.; DeGrado, W. F., *Nat. Chem.* **2017,** *9*, 1157.

(19) Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; Baker, D., *Nature* **2013,** *501,* 212.

(20) Bick, M. J.; Greisen, P. J.; Morey, K. J.; Antunes, M. S.; La, D.; Sankaran, B.; Reymond, L.; Johnsson, K.; Medford, J. I.; Baker, D., *eLife* **2017,** *6,* e28909.

(21) Lupas, A. N.; Bassler, J.; Dunin-Horkawicz, S., The Structure and Topology of α-Helical Coiled Coils. In *Fibrous Proteins: Structures and Mechanisms*, Parry, D. A. D.; Squire, J. M., Eds. Springer International Publishing: Cham, 2017; pp 95.

(22) Fletcher, J. M.; Boyle, A. L.; Bruning, M.; Bartlett, G. J.; Vincent, T. L.; Zaccai, N. R.; Armstrong, C. T.; Bromley, E. H. C.; Booth, P. J.; Brady, R. L.; Thomson, A. R.; Woolfson, D. N., *ACS Synth. Biol.* **2012,** *1,* 240.

(23) Gonzalez, L.; Plecs, J. J.; Alber, T., *Nat. Struct. Biol.* **1996,** *3,* 510.

(24) Yadav, M. K.; Redman, J. E.; Leman, L. J.; Alvarez-Gutierrez, J. M.; Zhang, Y. M.; Stout, C. D.; Ghadiri, M. R., *Biochemistry* **2005,** *44,* 9723.

(25) Thomson, A. R.; Wood, C. W.; Burton, A. J.; Bartlett, G. J.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N., *Science* **2014,** *346,* 485.

(26) Walshaw, J.; Woolfson, D. N., *Protein Sci.* **2001,** *10,* 668.

(27) Guo, Y.; Bozic, D.; Malashkevich, V. N.; Kammerer, R. A.; Schulthess, T.; Enger, J., *EMBO J.* **1998,** *17,* 5265.

(28) Ozbek, S.; Engel, J.; Stetefeld, J., *EMBO J.* **2002,** *21,* 5960.

(29) MacFarlane, A. A.; Orriss, G.; Okun, N.; Meier, M.; Klonisch, T.; Khajehpour, M.; Stetefeld, J., *PLoS One* **2012,** *7*.

(30) Liu, J.; Yong, W.; Deng, Y. Q.; Kallenbach, N. R.; Lu, M., *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101,* 16156.

(31) Zaccai, N. R.; Chi, B.; Thomson, A. R.; Boyle, A. L.; Bartlett, G. J.; Bruning, M.; Linden, N.; Sessions, R. B.; Booth, P. J.; Brady, R. L.; Woolfson, D. N., *Nat. Chem. Biol.* **2011,** *7,* 935.

(32) Burgess, N. C.; Sharp, T. H.; Thomas, F.; Wood, C. W.; Thomson, A. R.; Zaccai, N. R.; Brady, R. L.; Serpell, L. C.; Woolfson, D. N., *J. Am. Chem. Soc.* **2015,** *137,* 10554.

(33) Liu, J.; Zheng, Q.; Deng, Y. Q.; Cheng, C. S.; Kallenbach, N. R.; Lu, M., *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103,* 15457.

(34) Xu, C. F.; Liu, R.; Mehta, A. K.; Guerrero-Ferreira, R. C.; Wright, E. R.; Dunin-Horkawicz, S.; Morris, K.; Serpell, L. C.; Zuo, X. B.; Wall, J. S.; Conticello, V. P., *J. Am. Chem. Soc.* **2013,** *135,* 15565.

(35) Spencer, R. K.; Hochbaum, A. I., *Biochemistry* **2016,** *55,* 3214.

(36) Spencer, R. K.; Hochbaum, A. I., *Biochemistry* **2017,** *56,* 5300.

(37) Lizatovic, R.; Aurelius, O.; Stenstrom, O.; Drakenberg, T.; Akke, M.; Logan, D. T.; Andre, I., *Structure* **2016,** *24,* 946.

(38) Crick, F. H. C., *Acta Crystallogr.* **1953,** *6,* 689.

(39) Offer, G.; Sessions, R., *J. Mol. Biol.* **1995,** *249,* 967.

(40) Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S., *Science* **1998,** *282,* 1462.

(41) Dunin-Horkawicz, S.; Lupas, A. N., *J. Struct. Biol.* **2010,** *170,* 226.

(42) Grigoryan, G.; DeGrado, W. F., *J. Mol. Biol.* **2011,** *405,* 1079.

(43) Huang, P. S.; Oberdorfer, G.; Xu, C. F.; Pei, X. Y.; Nannenga, B. L.; Rogers, J. M.; DiMaio, F.; Gonen, T.; Luisi, B.; Baker, D., *Science* **2014,** *346,* 481.

(44) Wood, C. W.; Bruning, M.; Ibarra, A. A.; Bartlett, G. J.; Thomson, A. R.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N., *Bioinformatics* **2014,** *30,* 3029.

(45) Thomas, F.; Boyle, A. L.; Burton, A. J.; Woolfson, D. N., *J. Am. Chem. Soc.* **2013,** *135,* 5161.

(46) Thomas, F.; Burgess, N. C.; Thomson, A. R.; Woolfson, D. N., *Angew. Chem. Int. Ed.* **2016,** *55,* 987.

(47) Steinbacher, S.; Bass, R.; Strop, P.; Rees, D. C., *Curr. Top. Membr.* **2007,** *58,* 1.

(48) Hume, J.; Sun, J.; Jacquet, R.; Renfrew, P. D.; Martin, J. A.; Bonneau, R.; Gilchrist, M. L.; Montclare, J. K., *Biomacromolecules* **2014,** *15,* 3503.

(49) Lentz, B. R., *Chem. Phys. Lipids* **1989,** *50,* 171.

(50) Parasassi, T.; Krasnowska, E. K.; Bagatolli, L.; Gratton, E., *J. Fluoresc.* **1998,** *8,* 365.

(51) Sreerama, N.; Woody, R. W., *Methods Enzymol.* **2004,** *383,* 318.

(52) Ge, Y.; van der Kamp, M.; Malaisree, M.; Liu, D.; Liu, Y.; Mulholland, A. J., *J. Comput.-Aided Mol. Des.* **2017,** 1.

(53) Watkins, D. W.; Jenkins, J. M. X.; Grayson, K. J.; Wood, N.; Steventon, J. W.; Le Vay, K. K.; Goodwin, M. I.; Mullen, A. S.; Bailey, H. J.; Crump, M. P.; MacMillan, F.; Mulholland, A. J.; Cameron, G.; Sessions, R. B.; Mann, S.; Anderson, J. L. R., *Nat. Commun.* **2017,** *8,* 358.

(54) Trott, O.; Olson, A. J., *J. Comput. Chem.* **2010,** *31,* 455.

(55) Smart, O. S.; Neduvelil, J. G.; Wang, X.; Wallace, B. A.; Sansom, M. S. P., *J. Mol. Graphics* **1996,** *14,* 354.

(56) Wakil, S. J.; Stoops, J. K.; Joshi, V. C., *Annu. Rev. Biochem.* **1983,** *52,* 537.

(57) Linder, M. E.; Deschenes, R. J., *Nat. Rev. Mol. Cell Biol.* **2007,** *8,* 74.

(58) Blomhoff, R.; Blomhoff, H. K., *J. Neurobiol.* **2006,** *66,* 606.

(59) Bushue, N.; Wan, Y. J. Y., *Adv. Drug Del. Rev.* **2010,** *62,* 1285.

(60) Kloer, D. P.; Schulz, G. E., *Cell. Mol. Life Sci.* **2006,** *63,* 2291.

(61) Britton, G., *FASEB J.* **1995,** *9,* 1551.

(62) Albuquerque, P.; Casadevall, A., *Med. Mycol.* **2012,** *50,* 337.

(63) Joo, J. H.; Jetten, A. M., *Cancer Lett.* **2010,** *287,* 123.

(64) Grunler, J.; Ericsson, J.; Dallner, G., *Biochim. Biophys. Acta* **1994,** *1212,* 259.

(65) McTaggart, S. J., *Cell. Mol. Life Sci.* **2006,** *63,* 255.

Insert Table of Contents artwork here