De Novo-Designed α-Helical Barrels as Receptors for Small Molecules

Franziska Thomas,†‡§ William M. Dawson,†‡ Eric J. M. Lang,‡§ Antony J. Burton,†,¶ Gail J. Bartlett,† Guto G. Rhys,† Adrian J. Mulholland,†‡§⊥ and Derek N. Woolfson†‡§⊥

†School of Chemistry, University of Bristol, Cantock’s Close, Bristol BS8 1TS, U.K.
‡Institute of Organic and Biomolecular Chemistry, Georg-August-Universität Göttingen, Tammannstrasse 2, 37077 Göttingen, Germany
§BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, U.K.
¶Frick Chemistry Laboratory, Princeton, New Jersey 084544, United States

Supporting Information

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-diphenylhexatriene (DPH) in the micromolar range and fluoresce. Displacement of the dye is used to report the binding of nonfluorescent molecules: palmitic acid and retinol bind to all three αHBs with submicromolar inhibitor constants; farnesol binds the hexamer and heptamer; but β-carotene binds only the heptamer. A co-crystal structure of the hexamer with farnesol reveals oriented binding in the center 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 diphasphate.

KEYWORDS: α-helical barrel, coiled coil, molecular dynamics, rational peptide design, small-molecule binding

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 (2) making the design methods open and accessible to others, particularly to nonspecialist users;7 and (3) building on these methods to deliver functional de novo proteins.5,10 Desired functions include biomolecular recognition and sensing and catalysis or enzyme-like activities.5,11 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 millimolar to femtomolar.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 ~400 G-protein-coupled receptors can distinguish up to 1 trillion different stimuli through low-affinity and relatively nonselective recognition of odorants.15

The design and engineering of new proteins that recognize and bind peptides and folded proteins have been considerably successful, 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 have 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. 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 two to four helices wrapped around a central superhelical axis leading to solid hydrophobic cores. These structures are directed by underlying sequence repeats of hydrophobic (h) and polar (p) residues, hpphppp, often denoted abcedfg. 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 homodimers, trimers, and tetramers and reliable sets of de novo heterodimeric and trimeric systems.22

Apart from a small number of examples,23,24 however, these traditional coiled-coil dimers—tetr Tomers do not have suitable internal cavities or central channels to provide a basis for ligand 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., hpphphp and others.26 Both natural and engineered αHBs are known but 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 With respect 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 helices33 with a central cavity that also binds hydrophobic dyes.34 There are other examples of coiled-coil oligomers above pentamer, but to the best of our knowledge, these do not have central channels or their binding and transport properties have not been fully characterized.35−37 Indeed, adding specificity of binding for small molecules has not been addressed formally by any of the reports cited above.

The coiled coil is a parametrizable protein fold;38 i.e., there are equations that allow coiled-coil structures to be built ab initio making them accessible to computational design.23,39−44 We have combined parametric design and the use of hpphphp-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 of >95 °C,25 and by analogy to de novo coiled-coil dimers—tetr Tomers, we estimate the dissociation
constants of the αHBs to be in the subnanomolar regime. Furthermore, these αHBs are robust to mutation, allowing the engineering of coiled-coil-based nanotubes, and the generation of variants with engineered lumens that perform rudimentary catalysis.

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 \( \approx 5 \) to 7 Å, provides a first level of discrimination: the different barrels bind different targets with a range of affinities from submicromolar upward. 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 modeling 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 designed structures, namely, CC-Pent, CC-Hex2, and CC-Hept (Figure 1A). The sequences of these hyperstable assemblies are similar (Supplementary Table 2 and Supplementary Figure 1), and the Ile/Leu-lined channels differ only in diameter, which increases through the series CC-Pent to CC-Hept. To quantify this, the three X-ray crystal structures were analyzed by Pore Walker (Supplementary Figure 2 and Supplementary Table 3), which revealed variations in the diameters of the channels in CC-Pent, CC-Hex2, and CC-Hept of 3.0–7.4, 4.7–7.7, and 5.4–10.1 Å, respectively.

Channels of similar chemistry and dimensions bind hydrophobic dyes, including the rod-shaped 1,6-diphenylhexatriene [DPH (1) (Figure 1B)] and the more sterically demanding Prodan [2 (Figure 1B)]. We reasoned that differences in channel diameter among 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 (Figure 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 Figure 3). All three barrels accommodated DPH, whereas Prodan fitted into the channels of only CC-Hex2 and CC-Hept.

We used fluorescence spectroscopy to probe binding of the two dyes to the αHBs (Figure 2) as the fluorescence of both dyes is environmentally sensitive: DPH fluoresces \( (\lambda_{\text{max}} = 455 \text{ nm}) \) only when in hydrophobic environments, and the emission fluorescence spectrum of Prodan is enhanced and blue-shifted (from 520 to 424 nm) when it is 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 (Figure 1A). Neither of the dyes fluoresced with either control over peptide concentrations of 10–300 \( \mu \text{M} \) (Figure 2A–D). By contrast, 50 \( \mu \text{M} \) concentrations of all three αHBs fluoresced with 1 \( \mu \text{M} \) DPH (Figure 2A). Prodan bound to CC-Hex2 and CC-Hept, but not to the smaller CC-Pent (Figure 2B).

Figure 2. Fluorescence and CD spectroscopy of DPH (left) and Prodan (right) binding to αHBs. Emission fluorescence spectra for (A) DPH and (B) Prodan with coiled-coil peptides. Saturation binding curves for binding of the peptides to (C) DPH and (D) Prodan. CD spectra of the dye in the absence (black dotted line) and presence of αHBs (solid lines) for (E) DPH and (F) Prodan 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 and B) 50 \( \mu \text{M} \) peptide, 1 \( \mu \text{M} \) dye, HEPES-buffered saline (HBS; 25 mM HEPES and 100 mM NaCl (pH 7.0)), and 5% (v/v) dimethyl sulfoxide (DMSO); (C and D) 1.4–100 \( \mu \text{M} \) the coiled-coil assembly, 1 \( \mu \text{M} \) dye, HBS, and 5% (v/v) DMSO; (E) 200 \( \mu \text{M} \) peptide, 5 \( \mu \text{M} \) DPH, HBS, and 5% (v/v) DMSO; and (F) 200 \( \mu \text{M} \) peptide, 20 \( \mu \text{M} \) Prodan, HBS, and 5% (v/v) DMSO.

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

In addition, sedimentation velocity analytical ultracentrifugation experiments with 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 Figure 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.
Table 1. Dissociation constants and inhibition constants of small molecules binding to the αHBs

<table>
<thead>
<tr>
<th></th>
<th>CC-Pent</th>
<th></th>
<th>CC-Hex2</th>
<th></th>
<th>CC-Hept</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_d$ (μM)</td>
<td>$K_i$ (μM)</td>
<td>$K_d$ (μM)</td>
<td>$K_i$ (μM)</td>
<td>$K_d$ (μM)</td>
</tr>
<tr>
<td>DPH (1)</td>
<td>4.5 ± 0.9</td>
<td>–</td>
<td>1.6 ± 0.2</td>
<td>–</td>
<td>1.3 ± 0.3</td>
</tr>
<tr>
<td>Prodan (2)</td>
<td>401 ± 14</td>
<td>–</td>
<td>6.5 ± 1.1</td>
<td>–</td>
<td>5.8 ± 0.6</td>
</tr>
<tr>
<td>TMA-DPH (3)</td>
<td>15.5 ± 0.9</td>
<td>–</td>
<td>14.8 ± 1.4</td>
<td>–</td>
<td>13.3 ± 1.3</td>
</tr>
<tr>
<td>palmitic acid (4)</td>
<td>–</td>
<td>0.8 ± 0.3</td>
<td>–</td>
<td>0.7 ± 0.2</td>
<td>–</td>
</tr>
<tr>
<td>retinol (5)</td>
<td>–</td>
<td>10.2 ± 2.8</td>
<td>–</td>
<td>2.9 ± 1.3</td>
<td>–</td>
</tr>
<tr>
<td>farnesol (6)</td>
<td>–</td>
<td>23.9 ± 2.4</td>
<td>–</td>
<td>0.6 ± 0.2</td>
<td>–</td>
</tr>
<tr>
<td>β-carotene (7)</td>
<td>–</td>
<td>nd*</td>
<td>–</td>
<td>74 ± 27</td>
<td>–</td>
</tr>
</tbody>
</table>

*A $K_i$ could not be determined because of weak binding of β-carotene to CC-Pent.

free solution, although both molecules absorbed strongly in the near ultraviolet, neither had associated CD spectra (Figure 2E,F). However, when bound to DPH, all three αHBs showed strong CD bands centered on ≈350 nm with vibrational fine structure (Figure 2E). Interestingly, these data showed a clear inverse correlation with αHB pore size: the intensity of the CD signal decreased through the series CC-Pent to CC-Hex2 to CC-Hept (Figure 2E). However, CD spectra were observed for Prodan with only CC-Hex2 and CC-Hept (Figure 2F), consistent with the binding data from fluorescence spectroscopy.

Computational Modeling Sheds Light on the Modes of Binding of DPH to αHBs. Molecular dynamics (MD) simulations can identify binding sites and binding modes of small molecules in natural and de novo proteins. DPH was docked into the channels of the X-ray crystal structures for the three αHBs using AutoDock Vina (Figure 3A and Supplementary Figure 5A–C). 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 α position, and adjacent sites share a void (Supplementary Figure 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.

To confirm that the binding sites identified were favorable and the docking poses stable, extensive MD simulations were initiated starting from five poses for each site and running simulation for 200 ns 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 (Figure 3B–D, Supplementary Figures 6 and 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 Figure 8), which reached maxima of ≤2 Å in both cases. By contrast, the average RMSDs for CC-Hex2 for 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 the superhelix of CC-Hex2, which relaxed to a higher pitch over the MD trajectories (Supplementary Figure 9) leading to a longer, narrower channel with a diameter of 3–6 Å calculated by HOLE (Supplementary Figure 10). A similar analysis for CC-Pent revealed that the presence of DPH led 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 increases by ≤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 Figure 9). In contrast, CC-Hex2 and CC-Hept showed a subtle tightening in response to the ligand (Supplementary Figure 10). Presumably, this improves van der Waals contacts between protein and ligand.

With respect to the dynamics of the DPH molecule (Figure 3B–D and Supplementary Figures 6 and 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 sites 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 Figure 12). The preferred positions for the phenyl groups were the widened voids, even for CC-Hept that has a channel larger than those of 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 because of the larger diameter of its channel (Supplementary Figures 13 and 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 sites 2 and 3. Together, these simulations reveal increased restrictions in ligand binding and movement as the channel diameter shrinks in the series CC-Hept, CC-Hex2, and CC-Pent. It is interesting that this correlates with the increased intensity of the induced CD signal of DPH.

αHBs 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 (S), farnesol (6), and β-carotene (7) (Figure 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 opsins protein class, e.g., rhodopsin.58,59 Retinal cannot be synthesized by humans and is taken up as a retinyl ester or the provitamin β-carotene (7), which is cleared enzymatically to give two retinol molecules.60,61 Farnesol has antifungal properties and antitumor and chemopreventative effects in animal models, which is mainly due to it inhibiting cell growth and proliferation and inducing apoptosis.62,63

The propensities of binding of these molecules to αHBs were determined from a displacement assay as follows. In place of DPH, the more water-soluble trimethylammonium-DPH [TMA-DPH (3) (Figure 1)] was used as the fluorescent reporter. This bound all three αHBs with K_{D} values of 13–16 μM (Table 1 and Supplementary Figure 15 and Table S1). αHBs were equilibrated for 2 h with TMA-DPH in approximately 25-fold excess at 20 °C and titrated with the analytes 4–7. Decreases in fluorescence with an 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} (eq S3)].

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

We offer the following explanation for the differences in the binding of the three terpenoids. First, compared with palmitic acid, which is not discriminated by the $\alpha$HBs 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 $\alpha$HBs. Third, the two cyclohexene headgroups of $\beta$-carotene inhibit binding to the two smaller barrels (CC-Pent and CC-Hex2).

**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 farnesol. A 1.85 Å resolution X-ray crystal structure for the complex was determined by molecular replacement using apo-CC-Hex2 (PDB entry 4pn9) as the search model (Supplementary Table 6, Figure 4E, and Supplementary Figure 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 $\alpha$HBs that we have designed and performed, respectively, there was density consistent with ordered hydrogen-bonded water molecules connecting the hydroxyl functionality of farnesol to bulk solvent (Figure 4E).

**The Lumens of $\alpha$HBs Can Be Engineered To Accept Charged Molecules.** Hydrophobic and charged side chains can be introduced into the otherwise exclusively hydrophobic lumens of de novo $\alpha$HBs without major structural changes. 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. 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 toward substitution with Lys or Glu, to give CC-Hept-I24K or CC-Hept-I24E, respectively, without compromising the structure or stability of the $\alpha$HB (Supplementary Table 2 and Supplementary Figures 17–22). We obtained an X-ray crystal structure for the latter (Supplementary Table 6 and Supplementary Figure 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 ($K_d$ of 13.3 ± 1.3 and 12.8 ± 0.6 $\mu M$, respectively (Figure 5A)), whereas binding to CC-Hept-I24E was almost 1 order of magnitude tighter at 1.6 ± 0.1 $\mu M$ (Figure 5A, purple trace). Using the TMA-DPH displacement assay, CC-Hept-I24K and CC-Hept-I24E accommodated neutral farnesol with $K_d$ of 1.8 ± 0.3 and 1.0 ± 0.04 $\mu M$, respectively (Figure 5B).

Finally, we tested the binding of the anionic and biologically important farnesyl diphosphate [FPP (8)]. FPP is the precursor of sesquiterpenoids, including the antimalarial drug artemisinin and squalene, the precursor of sterols, and it is a substrate of protein prenyltransferases involved in trafficking of proteins to membranes. 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 (Figure 5C). By contrast, FPP bound to CC-Hept-I24K >10 times tighter with a measurable $K_d$ of 1.5 ± 0.2 $\mu M$ (Figure 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 heptalysine. 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 (Figure 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 Figures 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, five of the Lys residues were protonated, and for the remainder, either six, three, or seven lysine residues (ranked by decreasing prevalence) were charged (Supplementary Figure 28). The charged FPP remained in a stable position throughout the course of the simulations (Supplementary Figure 26), with the hepta-Lys ring accommodating the three negative charges of the pyrophosphate through strong Coulombic interactions. However, the Lys side chains were mobile (Supplementary Figure 27), and not all interacted with FPP throughout the simulation (Supplementary Figure 24); however, the protein backbone remained stable (Supplementary Figure 28).

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 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 environmentally 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 nonfluorescent analytes to αHBs. The varying internal diameters of the structures, which are in the range of 5–7 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 with respect to the αHBs can be rationalized by the physical properties of the small molecules, and through molecular modeling and dynamics simulations. In particular, a second discrimination factor was identified as the internal shape of the channels that alternate with 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 sequestered 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. While 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 for the development of membrane-spanning transporters and sensors.

ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00225.

Experimental details and supplementary characterization, simulation, and binding data (PDF)
Supplementary Movie 1 (MPG)
Supplementary Movie 2 (MPG)

AUTHOR INFORMATION

Corresponding Author
E-mail: d.n.woolfson@bristol.ac.uk.

ORCID
Franziska Thomas: 0000-0002-1176-7018
William M. Dawson: 0000-0003-2710-6879
Derek N. Woolfson: 0000-0002-0394-3202

Author Contributions
F.T. and W.M.D. contributed equally to this work.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

F.T. and D.N.W. are grateful for funding by the Leverhulme Trust (RPG-2012-536). W.M.D., G.J.B., and D.N.W. are supported by a European Research Council Advanced Grant (340764). D.N.W., E.J.M.L., and A.J.M. are supported by the BBSRC and EPSRC through the BrisSynBio Synthetic Biology Research Centre (BB/L01386X1). A.J.M. is supported by EPSRC Grant EP/M022609/1. A.J.B. and G.G.R. thank the Bristol Chemical Synthesis Centre for Doctoral Training funded by the Engineering and Physical Sciences Research Council (EP/G036764/1). D.N.W. holds a Royal Society Wolfson Research Merit Award. The authors 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


