Installing hydrolytic activity into a completely *de novo* protein framework

Antony J. Burton[1], Andrew R. Thomson[1], William M. Dawson[1], R. Leo Brady[2] and Derek N. Woolfson[1,2,3]*

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

[2]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol, BS8 1TD, UK

[3]BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK

*To whom correspondence should be addressed: d.n.woolfson@bristol.ac.uk; School of Chemistry, University of Bristol, Bristol BS8 1TS, UK. Tel: +44 (0)117 95 46347

Abstract

Designing enzyme-like catalysts tests our understanding of sequence-to-structure/function relationships in proteins. Here, we install hydrolytic activity predictably into a completely *de novo* and thermo-stable α-helical barrel, which comprises 7 helices arranged around an accessible channel. We show that the lumen of the barrel accepts 21 mutations to functional polar residues. The resulting variant, which has cysteine-histidine-glutamic acid triads on each helix, hydrolyses *p*-nitrophenyl acetate with catalytic efficiencies matching the most-efficient redesigned hydrolases based on natural protein scaffolds. This is the first report of a functional catalytic triad engineered into a *de novo* protein framework. The flexibility of our system also allows the facile incorporation of unnatural side chains to improve activity and probe the catalytic mechanism. Such predictable and robust construction of truly *de novo* biocatalysts holds promise for applications in chemical and biochemical synthesis.

(134 words)

Designing functional biomolecules *de novo* where both structure and function are built from first principles represents a considerable challenge. Enzyme-like catalysts constructed in this way would have wide-ranging uses for manipulating and synthesizing small molecules with potential applications in medicine, industrial biotechnology and basic science.[1-3] Such endeavors test our ability to generate protein folds that are not, or are rarely, observed in nature,[4] and also advance understanding of biological catalysis and how this might be emulated through design and engineering.

Current approaches to enzyme design largely center on the *redesign* of existing, natural protein scaffolds selected from the Protein Data Bank (PDB).[5,6] Arguably the most successful of these use theoretical enzyme substructures ("theozymes") that can be incorporated into such proteins for experimental validation. The designs are selected through computational modeling of transition-state analogues within side-chain constellations.[7,8] Iterative rounds of computational design, both in the active site and surrounding residues, place mutations to stabilize the transition state with the aim of accelerating the target reaction. Successes of this approach include redesigned proteins with activities for the Kemp elimination,[9] and the retro-aldol[10] and Diels-Alder[11] reactions. Often only moderate efficiencies are achieved, although these can be further improved by directed evolution.[12-16] Computational approaches have also been used to design esterases, again with moderate activities being achieved through the placement of a histidine (His) residue into the active site of thioredoxin,[17] the *C*-terminal domain of calmodulin[18] and cysteine-His (Cys-His) dyads installed into existing α/β protein folds.[19]

By contrast, much less has been demonstrated for the design of completely *de novo* catalysts; that is, where both the protein scaffold and the catalytic activity are built from scratch.[20-22] Although more challenging, this would carry advantages over redesign approaches: firstly, the designer could select and control all, or at least many of the residues, and so engineer the complete construct predictably; secondly, the resulting proteins could be made to function under conditions away from those required by natural proteins; and finally, success in this area would provide the acid test of our understanding of enzyme structure and function. Towards this fully *de novo* effort, successful designs of hydrolases have included: the decoration of small protein-folding motifs with His residues;[23,24] the employment of Zn^{2+} cations as Lewis-acidic cofactors;[25-28] and the identification of catalytically active proteins in combinatorial libraries of sequences patterned to form to all-α or all-β protein folds.[29]

One of the most productive areas in the *de novo* design of protein structure has been for coiled-coil assemblies.[30] Coiled coils are bundles of two or more α-helices, which wrap around one another to form rope-like structures.[31] Oligomer states above tetramer have central

cavities of defined size and side-chain chemistry. As a result, they are best described as α-helical barrels.[32,33] Recently, we have reported both the serendipitous discovery,[34] and the rational, computational design of a series of α-helical barrels, including pentamers, hexamers and a heptamer.[35] These structures have contiguous, mutable and accessible central channels,[36] which provide a basis for the installation of functional residues with precision and within defined environments. Here, using the heptamer, CC-Hept, which is highly thermostable and has a channel 8 Å in diameter,[35] we report that multiple mutations are accommodated in the lumen to give an α-helical barrel with catalytic activity.

Results

Rational design of multiple active sites

We chose to mimic active sites found in the majority of natural cysteine/serine (Cys, C; Ser, S) hydrolases.[37] In these, the Cys or Ser side chain acts as the nucleophile; the imidazole group of a proximal histidine (His, H) deprotonates the thiol or hydroxyl moiety; and an aspartic acid (Asp, D) or glutamic acid (Glu, E) residue polarizes the imidazole to activate the system further.

As with most coiled coils, α-helical barrels are founded on sequence patterns of hydrophobic and polar residues in so-called heptad repeats, ***abcdefg***.[30,31] The ***a*** and ***d*** positions are hydrophobic leucine (Leu, L) and isoleucine (Ile, I) residues, respectively, and project into the central pore. We targeted these sites for mutation. The peptide sequence of CC-Hept has four heptad repeats, **Table 1**. Thus, there were 12 possible contiguous ***d-a-d*** or ***a-d-a*** sites where C-H-E or E-H-C triads could have been installed. Because of concerns over helix fraying, we discounted mutations near the termini and focused on those in the third heptad repeat. Using CCBuilder[38] and structural parameters from the crystal structures of a range of *de novo* coiled coils,[35,39] we generated *in silico* all-atom models for single, double and triple mutants as tetramers to heptamers (**SI Table 1.1–1**). Encouragingly, for 3 of the 5 sequences heptamer was predicted as the most-stable state. With all 7 helices carrying the complete triad, a total of 21 mutations would be made to CC-Hept, corresponding to 10% of its residues. This places a high demand on the system. However, we posited that, in contrast to most natural protein structures, the high stability of *de novo* α-helical barrels and bundles might tolerate this level of change.[35,40] To test this, we adopted an iterative design strategy, fully characterizing point and double mutants *en route*.

Table 1. Peptide sequences and heptad register for CC-Hept and the designed variants.

Name	Sequence and register *cdef gabcdef gabc def gabcdef gabcdef*
CC-Hept	Ac–G EIAQ ALKEIAK ALKE IAW ALKEIAQ ALK G–NH_2
CC-Hept-I-C-I	Ac–G EIAQ ALKEIAK ALKE IAW A**C**KEIAQ ALK G–NH_2
CC-Hept-I-H-I	Ac–G EIAK ALREIAK ALRE IAW A**H**REIAK ALR G–NH_2
CC-Hept-C-H-I	Ac–G EIAK ALREIAK ALRE **C**AW A**H**REIAK ALR G–NH_2
CC-Hept-H-C-I	Ac–G EIAK ALREIAK ALRE **H**AW A**C**REIAK ALR G–NH_2
CC-Hept-C-H-E	Ac–G EIAK ALREIAK ALRE **C**AW A**H**RE**E**AK ALR G–NH_2
C-H-E-short	Ac–G RE **C**AW A**H**RE**E**AK G–NH_2
CC-Hept-C-H-E-outside	Ac–G EIAK ALREIA**C** ALR**H** IA**E** ALREIAW ALR G–NH_2
CC-Di-C-H-E	Ac–G EIAALKK EIAA LK**C** ENA**H**LK**E** EIAALKW G–NH_2
CC-Hept-C-L-E	Ac–G EIAK ALREIAK ALRE **C**AW ALRE**E**AK ALR G–NH_2
CC-Hept-I-H-E	Ac–G EIAK ALREIAK ALRE IAW A**H**RE**E**AK ALR G–NH_2
CC-Hept-hC-H-E*	Ac–G EIAK ALREIAK ALRE**hC**AW A**H**RE**E**AK ALR G–NH_2
CC-Hept-βMC-H-E*	Ac–G EIAK ALREIAK ALRE**βC**AW A**H**RE**E**AK ALR G–NH_2

* hC denotes homocysteine; βC denotes β-methyl cysteine.

CC-Hept tolerates multiple, polar mutations in its lumen

First, to test if CC-Hept could accommodate the key catalytic residues individually, singly mutated peptides incorporating Cys or His at the third ***a*** position were synthesized and characterized in solution. These were named systematically as CC-Hept-I-C-I and CC-Hept-I-H-I (**Table 1**). Circular dichroism (CD) spectroscopy showed that both maintained high α-helicity and thermal stability, with no loss of structure detected up to 90 °C (**Fig. 1a**&**b**). Sedimentation-velocity and sedimentation-equilibrium experiments conducted by analytical ultracentrifugation (AUC) returned molecular weights consistent with heptamers for both peptides (**SI Fig 1.4–1**&**2**). Finally, the design models were confirmed by determining X-ray crystal structures to atomic resolution. 1.95 and 1.80 Å resolution structures for CC-Hept-I-C-I and CC-Hept-I-H-I, respectively, were obtained by molecular replacement using a poly-alanine search model derived from the CC-Hept structure. The structure of CC-Hept-I-C-I revealed a corona of free-thiol Cys side chains with Sγ-Sγ distances of 4.2–4.7 Å, *i.e.* not-disulfide linked helices; and that for CC-Hept-I-H-I showed a H-bonded ring of imidazole side chains (**Fig. 2a**&**b**).

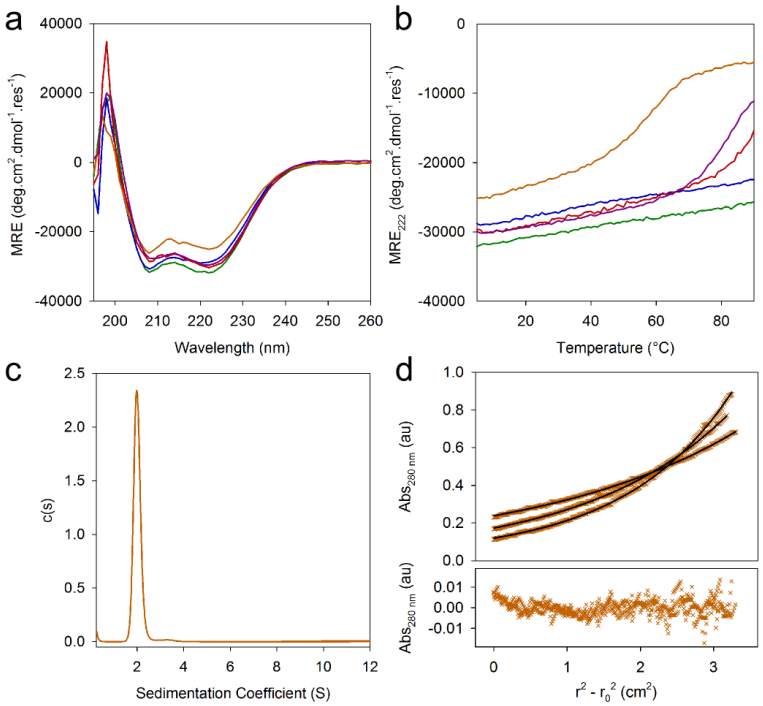

Figure 1: Solution-phase biophysical characterisation of the designed peptides. a: CD spectra at 20 °C for CC-Hept-I-C-I (green), -I-H-I (blue), -C-H-I (red), -H-C-I (purple) and -C-H-E (orange). MRE, mean residue ellipticity. **b:** Temperature-dependence of the CD signal monitored at 222 nm. Same colour scheme as for panel **a**. **c:** Sedimentation-velocity AUC c(s) distribution fit at 20 °C for CC-Hept-C-H-E, which returned a molecular weight of 22,800 Da, equivalent to 6.8 × monomer mass. **d:** Sedimentation-equilibium AUC for CC-Hept-C-H-E, with experimental data (top, crosses), fits (lines) to single-ideal species and residuals (bottom) between the experimental and fitted data points. These data were recorded at 20 °C, 280 nm and at 16, 19 and 22 krpm. The analysis returned a molecular weight of 24,060 Da (95% confidence limits 23,490–24,670 Da), equivalent to 7.1 × monomer mass. au, arbitrary units. Conditions: peptide concentrations of 10 μM (**a** and **b**), 150 μM (**c**), and 70 μM (**d**); phosphate-buffered saline (PBS, pH 7.4, **a** and **b**), HEPES-buffered saline (HBS, pH 7.0, **c** and **d**); 10-fold excess, relative to peptide, of *tris*-carboxyethylphosphine hydrochloride (TCEP) for Cys-containing sequences.

Next, to establish the best spacing and order of these key residues, two double mutants were made, CC-Hept-C-H-I and CC-Hept-H-C-I (**Table 1**), with a Cys-His or His-Cys dyad at consecutive ***d*** and ***a*** sites. This is important as side chains at these positions project differently into the lumen of the barrel. Indeed, computational scoring with CCBuilder[38] predicted CC-Hept-C-H-I to be heptameric, but CC-Hept-H-C-I as a hexamer (**SI Table 1.1–1**). Both synthetic peptides were highly α-helical and thermally stable (**Fig. 1a**). Moreover, the predicted oligomer states were verified by AUC experiments (**SI Fig 1.4–3**&**4**). Only an X-ray crystal structure of CC-Hept-C-H-I could be obtained. The 1.70 Å resolution structure confirmed a heptamer, and revealed the His side chain in an alternative rotamer to that observed in CC-Hept-I-H-I. This new conformer places the imidazole within 4 Å of the Cys thiol, with potential for a "catalytic" Cys-His interaction, although the geometry is not ideal as the thiol points away from the imidazole (**Fig. 2c**). Ordered water molecules were also observed H-bonded to each nitrogen atom of the His side chains. Thus, in this crystal structure

at least, the installed residues are accessible to water (**Fig. 2d**). For these reasons, we continued the design process with CC-Hept-C-H-I.

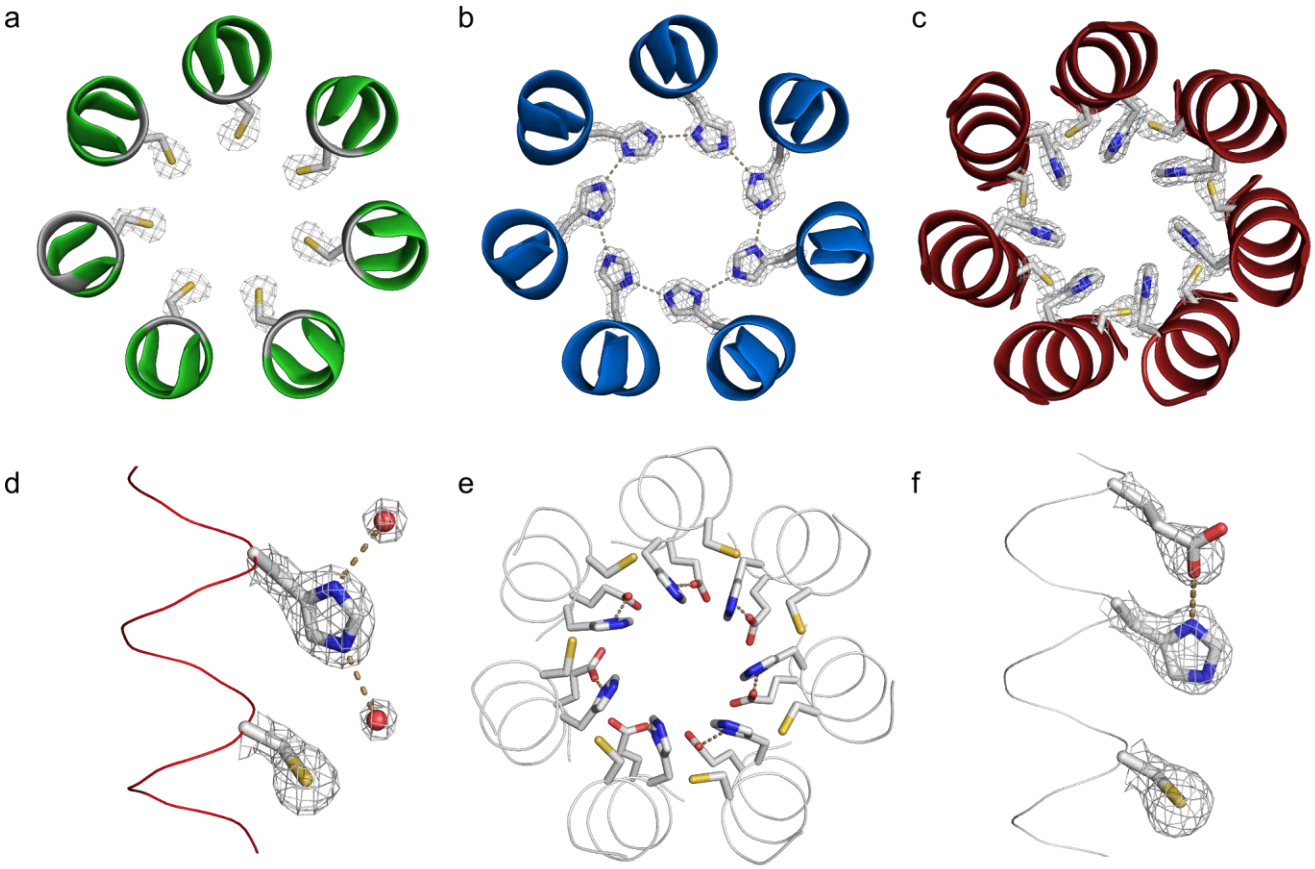

Figure 2. X-ray crystal structures of CC-Hept mutants. a-c: Slices through the structures showing the installed residues in CC-Hept-I-C-I (**a**, PDB 5ez8, 1.95 Å), CC-Hept-I-H-I (**b**, 5ez9, 1.80 Å) and CC-Hept-C-H-I (**c**, 5eza, 1.70 Å). Electron density ($2F_{obs}$-F_{calc}) for the mutated side chains is shown contoured at 2 σ. **d:** Electron density, $2F_{obs}$-F_{calc} contoured at 1.5 σ, for the Cys-His dyad in CC-Hept-C-H-I, revealing water molecules hydrogen bonded to the His side chain (Nδ/Nε–OH_2 = 2.9 Å). **e:** CC-Hept-C-H-E (5ezc, 1.80 Å), which harbors 21 mutations within its lumen. **f:** Electron density, $2F_{obs}$-F_{calc} contoured at 1.2 σ, for the C-H-E triad of CC-Hept-C-H-E. All images created with PyMOL (https://www.pymol.org).

Building further into the iterative design process, CC-Hept-C-H-E was modeled and synthesized. In this variant, Glu was placed at the *d* position immediately following the His of CC-Hept-C-H-I (**Table 1**). In total, this puts 21 polar mutations in the lumen of CC-Hept, which is otherwise lined by hydrophobic residues. Although approximately 25% less helical than the parent heptamer, CC-Hept-C-H-E was largely helical in solution, and had a cooperative and reversible unfolding transition with a T_M of 57 °C (**Fig. 1a**&**b**). Importantly, AUC showed that the construct was heptameric in solution (**Fig. 1c**&**d**), which was confirmed by a 1.80 Å resolution X-ray crystal structure (**Fig. 2e**&**f**). In the latter, electron density was clearly visible for the Cys, His and Glu side chains and the pore size was maintained at ≈8 Å.[41] The C-terminal regions of the helices had higher temperature factors (*B*-factors), suggesting fraying consistent with the decreased helicity in solution. The structure showed the intended His-Glu

H-bond (Oε-Nδ distance = 2.8 Å); and the Cys side chain adopted the same rotamer observed in CC-Hept-C-H-I. Whether the latter could adopt, or sample, different rotamers in solution was not clear from the X-ray crystal structure.[19] That 21 mutations, which corresponds to more than one third of the hydrophobic Ile or Leu residues, can be introduced into the lumen of CC-Hept is remarkable. These data highlight the hyperstability and mutability of these completely *de novo* α-helical barrels.

Hydrolytic activity is observed with the dyad and triad variants

For direct comparison to previous hydrolase designs,[17-19] we used a colorimetric *p*-nitrophenyl acetate (*p*NPA)-based assay to test for activity in the CC-Hept series. The single-mutant peptide assemblies, and hexameric CC-Hept-H-C-I showed no or little hydrolysis of *p*NPA above baseline at pH 7.0 (**Fig. 3a**). The heptameric CC-Hept-C-H-I did give product release over baseline, with an initial rate >10x background hydrolysis. However, considerably more activity was observed for CC-Hept-C-H-E, with initial rates >250x baseline (**Fig. 3a**).

As a control, we made a short, unstructured peptide with Cys, His and Glu mutations at analogous sequence positions to CC-Hept-C-H-E (**Table 1**). This C-H-E-short peptide hydrolyzed *p*NPA at ≈1.5x the background rate (**Fig. 3b**). To probe the role of placing the triad within the lumen of the barrel over simply being in an α-helical configuration, an additional control peptide was made with the triad on the solvent-exposed exterior of a coiled coil. Initially, this was attempted with CC-Hept, but the peptide aggregated (**Table 1**; **SI Fig 1.3–7**&**8**). Therefore, we used the homodimer, CC-Di,[39] incorporating Cys, His and Glu at sequential ***f***, ***c*** and ***f*** positions (**Table 1**). This peptide was highly α-helical, soluble and dimeric (**SI Fig 1.3–9**&**10**; **SI Fig 1.4–5**). In the hydrolysis assay, CC-Di-C-H-E was inactive like the C-H-E-short control (**Fig. 3b**). Thus, placing the side-chain constellation in the lumen of an α-helical barrel facilitates hydrolytic catalysis.

Next, we knocked out these "active-site" residues. CC-Hept-C-H-I (*vide supra*) provided the Glu knockout, and the further Cys and His knockouts were synthesized. The His→Leu mutation, CC-Hept-C-L-E, eliminated all catalytic activity, and product generation returned to baseline levels (**Fig 3b**). Interestingly, the Cys→Ile mutation, CC-Hept-I-H-E, did generate product over the course of an hour (**Fig. 3b**), suggesting that the His residue can act as a nucleophile, though not as effectively as CC-Hept-C-H-E. Finally, to eliminate the possibility that divalent cations, such Zn^{2+}, were participating in the mechanism of action of CC-Hept-C-H-E, the assay was re-run in the presence of excess ethylenediaminetetraacetic acid (EDTA). The kinetic profile overlaid with the trace in the absence of EDTA (**SI Fig 1.5–1**).

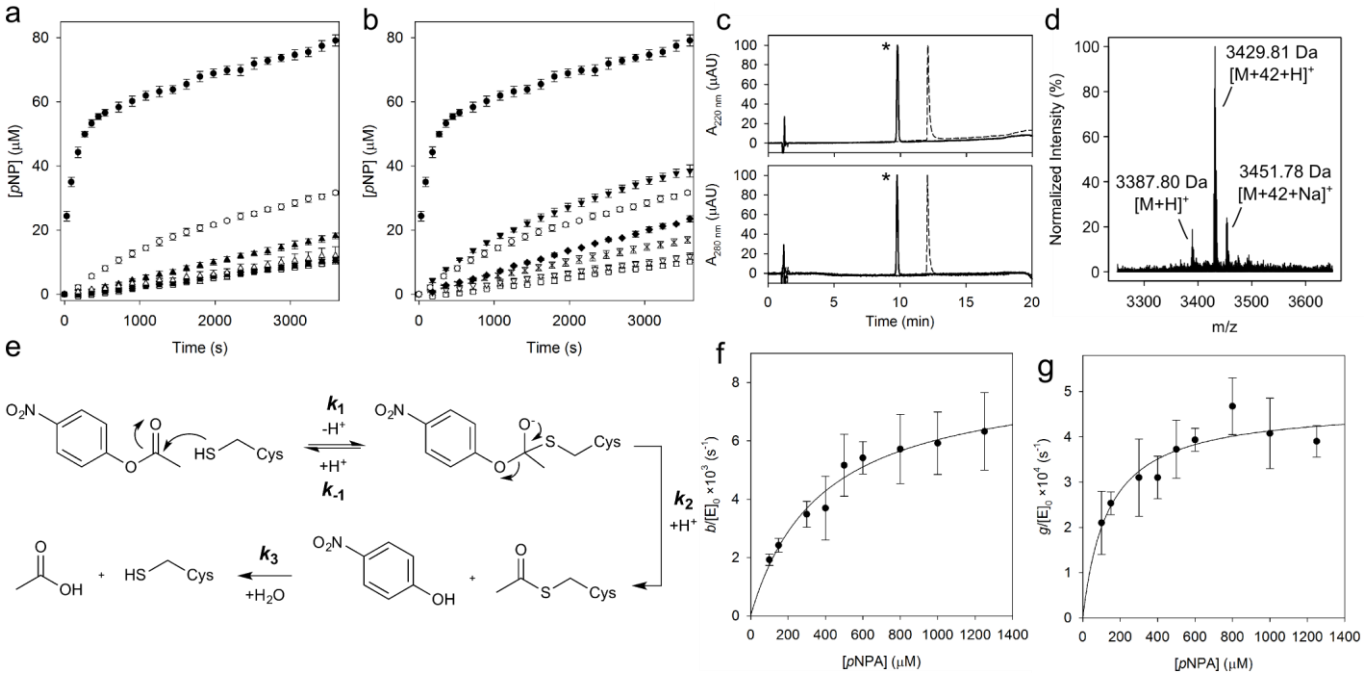

Figure 3. Mechanistic and kinetic analysis of the reaction of the designed peptides with *p*NPA. **a:** Assay traces in the presence of CC-Hept-I-H-I (■), CC-Hept-I-C-I (△), CC-Hept-H-C-I (▲), CC-Hept-C-H-I (**O**) and CC-Hept-C-H-E (●) against background hydrolysis (□). **b:** Assay traces for the control and "knock-out" peptides: C-H-E-short (◆), CC-Di-C-H-E (✕), CC-Hept-C-L-E (▽), CC-Hept-C-H-I (**O**) and CC-Hept-I-H-E (▼) against background hydrolysis (□) with, for comparison, CC-Hept-C-H-E (●). **c:** Normalized HPLC traces for CC-Hept-C-H-E (dashed) and acylated CC-Hept-C-H-E after reaction with *p*NPA for 10 min (*, solid line). **d:** MALDI-TOF MS of the new peak in **c** (labeled *). **e:** Proposed mechanism for the reaction of CC-Hept-C-H-E with *p*NPA via a thioester intermediate. **f:** Pre-steady state analysis of the first and second acylation events for the reaction of CC-Hept-C-H-E with *p*NPA; k_2/K_S = 22.0 ± 2.0 $M^{-1}s^{-1}$. **g:** Michaelis-Menten plot for the turnover step; k_{cat}/K_M = 3.7 ± 0.6 $M^{-1}s^{-1}$. Conditions: 50 μM (**a** and **b**), 100 μM (**c** and **d**) and 10 μM (**f** and **g**) peptide concentrations, 10-fold excess of TCEP relative to peptide, 300 μM *p*NPA (**a** and **b**), HBS (pH 7.0, the optimum pH (**SI Table 1.12–1)**), 5% v/v CH_3CN, 20 (**a** and **b**) and 22 ± 1 °C (**f** and **g**). Evolution of *p*NP was monitored at its pH-independent isosbestic point, 348 nm; ε = 5,400 $cm^{-1}M^{-1}$. For the fitted data in panels **f** and **g** $[E]_0$ = 10 μM, the total concentration of peptide chains in the assay.

CC-Hept-C-H-E has enzyme-like behaviour

The CC-Hept-C-H-E assembly had a distinct two-state kinetic profile (**Fig. 3a**), reminiscent of a "ping-pong" mechanism involving a covalently bound acyl intermediate as observed for the serine protease α-chymotrypsin with similar substrates.[42] Furthermore, the y-intercept for the second, steady state curve of this profile matched the concentration of peptide in the assay. This indicates that the intermediate is present on each peptide chain of the heptamer. An acyl intermediate could be bound to the Cys (as a thioester) or the His (as an acyl-imidazolium cation) side chains. We probed this by isolating a new species that appeared in HPLC of the reaction mixture (**Fig. 3c**). When analyzed by MALDI-TOF mass spectrometry (MS), the isolate gave a peak of the mass expected for CC-Hept-C-H-E plus 42 Da (**Fig. 3d**), consistent with single acylation of each of 7 peptide chains in the coiled-coil assembly. Moreover, MS/MS analysis of this peak confirmed acylation at the Cys residue only (**Fig. 3e**; **SI Fig 1.13–1**).

Finally, CC-Hept-C-H-E effected robust catalysis with >9 turnovers in an extended 16-hour assay (**SI Fig. 1.6–1**).

As a first step to determining complete kinetic parameters for the reaction of CC-Hept-C-H-E with *p*NPA, we fitted the kinetic profiles recorded for a range of substrate concentrations to a model with (1) an exponential pre-steady state parameter, followed by (2) a linear steady state parameter. This has been used for the reaction of α-chymotrypsin with *p*NP esters to deliver k_2/K_S and k_{cat}/K_M, where K_S is the pseudo K_M for the acylation step.[42] In our case, this model did not fit the experimental data (**SI Fig. 1.11–1**). However, fitting to a double exponential followed by a linear phase returned good fits, and was the simplest solution found (**SI Fig. 1.11–1**). This suggests that the acylation events of the burst phase do not occur at the same rate. This stands to reason given the demands on substrate entry, binding and product release in an 8 Å lumen with multiple active sites. The two pre-exponential terms indicate that the first and second acylation events occur faster than the remaining five (**SI Table 1.11–1**). The k_2/K_S value for the first two acylation events was 22.0 ± 2.0 $M^{-1}s^{-1}$ (**Fig. 3f**), which allows comparisons with foregoing hydrolase designs (*vide infra*). The overall catalytic efficiency of CC-Hept-C-H-E (k_{cat}/K_M) was 3.7 ± 0.6 $M^{-1}s^{-1}$ (**Fig. 3g**).

Non-proteinogenic amino acids provide insight into the catalytic mechanism

In an attempt to improve catalytic efficiency, we replaced Cys in CC-Hept-C-H-E with homocysteine (hCys), giving CC-Hept-hC-H-E (**Table 1**). The rationale was that the greater conformational flexibility of hCys should increase productive nucleophile-His encounters, and give better access of water to effect hydrolysis of the thioester intermediate. Balanced against this, however, the thiol of hCys is approximately 100-fold less acidic than that of Cys. CC-Hept-hC-H-E was folded, thermally stable (**Fig. 4a**; **SI Fig. 1.3–15&16**) and heptameric in solution (**SI Fig. 1.4–8**), which was confirmed by a 2.10 Å X-ray crystal structure. The latter revealed the hCys side chains as free thiols (**Fig. 4e**). CC-Hept-hC-H-E had catalytic activity with improved substrate turnover, carrying out three turnovers in 3 hours (**Fig. 4b**), and >12 in the extended 16-hour assay (**SI Fig. 1.8–1**). Conversely, the burst phase with CC-Hept-hC-H-E was slower than for CC-Hept-C-H-E, presumably because of the decreased nucleophilicity of the homocysteine side chain compared with cysteine.

In the X-ray crystal structure of CC-Hept-C-H-E the Cys thiol group is further from the His side chain than is ideal for catalysis (**Fig. 2f**). Presumably, a more-active conformer of the Cys side chain can be accessed by rotation in solution. We reasoned that modification of the Cys side chain to β-methyl-Cys should hold the thiol away from the His side chain permanently, and so reduce or even abolish activity. A β-methyl-Cys building block, Fmoc-L-β-Me-Cys-(*p*-methoxybenzyl)-OH, suitable for peptide synthesis was made from L-Thr in 6 steps and a 13%

overall yield (**SI Scheme 1.15–1**). A peptide incorporating this, CC-Hept-βMC-H-E (**Table 1**), was folded and heptameric in solution (**Fig. 4a**; **SI Fig. 1.4–9**), and with *p*NPA was effectively inactive like the control peptide, C-H-E-short. This hinted that the nucleophile-His interaction had indeed been disrupted (**Fig. 4b**). A 1.80 Å resolution crystal structure for CC-Hept-βMC-H-E confirmed this, revealing the thiol groups of the βMC side chains forced away from the His side chains (**Fig. 4f**). This highlights the precision and predictability with which chemical groups can be installed into CC-Hept.

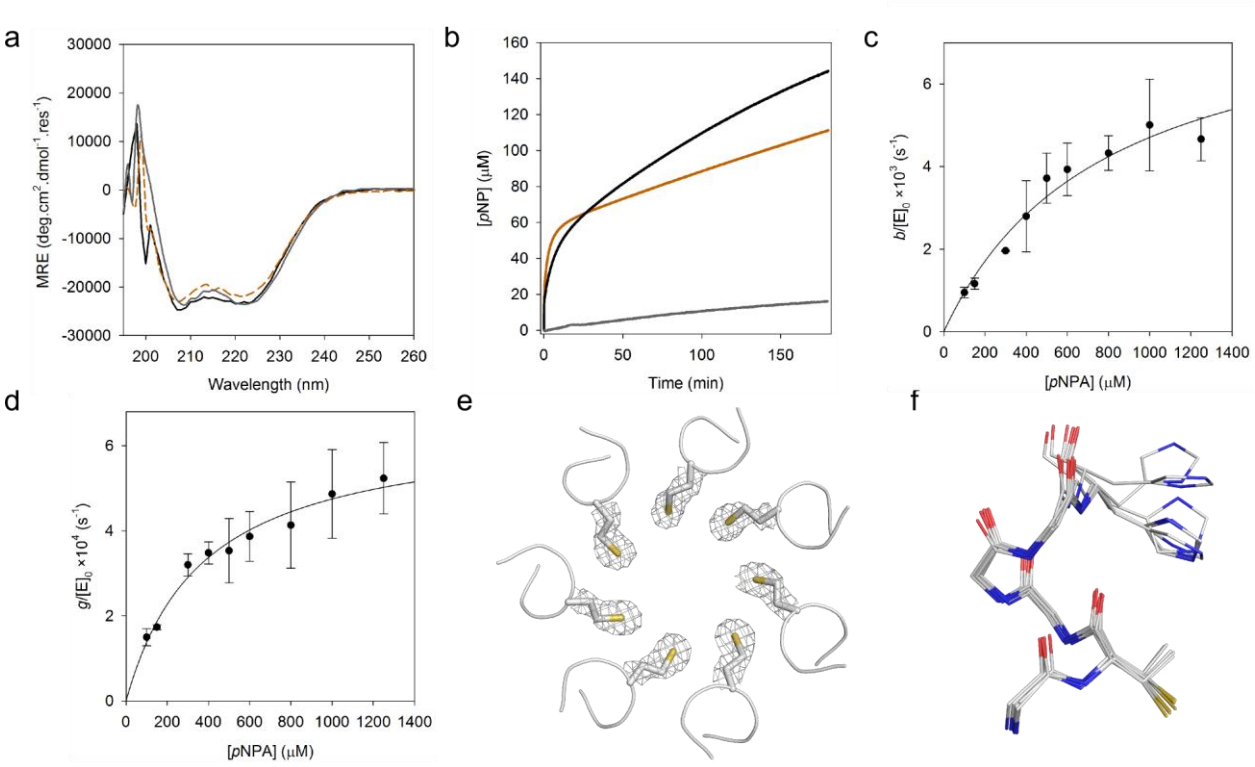

Figure 4. Characterization and hydrolytic evaluation of CC-Hept-hC-H-E and CC-Hept-βMC-H-E. a: CD spectra at 20 °C for CC-Hept-hC-H-E (black), CC-Hept-βMC-H-E (grey) and, for comparison, CC-Hept-C-H-E (orange). **b:** Baseline-corrected assay traces for the hydrolysis of 300 μM *p*NPA by CC-Hept-C-H-E (orange), CC-Hept-hC-H-E (black) and CC-Hept-βMC-H-E (gray). **c:** Pre-steady state analysis of the first and second acylation events for the reaction of CC-Hept-hC-H-E with *p*NPA at 22 ± 1 °C; k_2/K_S = 11.9 ± 1.2 $M^{-1}s^{-1}$. **d:** Michaelis-Menten plot for the turnover step; k_{cat}/K_M = 1.9 ± 0.2 $M^{-1}s^{-1}$. **e:** Electron density, $2F_{obs}$-F_{calc} contoured at 2 σ, for the hCys mutations in CC-Hept-□□-H-E (5f2y, 2.10 Å). **f**: Overlay of the seven chains of CC-Hept-βMC-H-E (5eze, 1.80 Å) displaying the βMC and His mutations, with the thiol forced from the His side chain. Conditions: 10 μM (**a**, **c** and **d**) and 50 μM (**b**) peptide concentrations; PBS (**a**), HBS (**b**, **c** and **d**); 10-fold excess of TCEP relative to peptide; 5% v/v CH_3CN (**b**, **c** and **d**). For the fitted data in panels **c** and **d**, $[E]_0$ = 10 μM, the concentration of peptide chains in the assay. Images **e** and **f** were created in PyMOL (https://www.pymol.org).

Discussion

We have described the successful installation of multiple catalytic triads into a completely *de novo* protein framework. Hydrolytic activity is observed when Cys-His-Glu catalytic triads are installed into the lumen of a *de novo* designed heptameric α-helical barrel, CC-Hept, with

improved turnover achieved when hCys is used as the nucleophile. Forcing the thiol from the His side chain, by adding a β-methyl group, abolishes the pre-steady state kinetic profiles.

The kinetic profile for ester hydrolysis by CC-Hept-C-H-E is clearly two-state (**Fig. 3a**). We interpret this as follows: the initial burst phase corresponds to the formation of a thioester intermediate with release of *p*NP (**Fig. 3e**; k_2/K_S); this is then followed by a rate-limiting slower phase, corresponding to the hydrolysis of this adduct and release of acetate (**Fig. 3e**; k_3). This "ping-pong" mechanism is well established for α-chymotrypsin, which displays similar profiles with *p*NP esters and proceeds *via* an ester intermediate;[42] and it is similar to that proposed for previously designed hydrolase catalysts.[19] Rapid and stoichiometric acylation within the heptameric barrel leads to seven acylated Cys side chains, followed by rate-limiting hydrolysis of the thioester intermediate. In support of this mechanism, we could not find any evidence for acylation at His. That said, acylation was observed in assays with the Cys-knockout peptide CC-Hept-I-H-E (**SI Fig. 1.10–1**). This suggests that His may act as the nucleophile, but not in the presence of adjacent Cys or hCys, which are expected to be better nucleophiles under the experimental conditions that we use.

Kinetic analysis of the burst phase of the reaction of CC-Hept-C-H-E with *p*NPA gives k_2/K_S = 22.0 ± 2.0 $M^{-1}s^{-1}$ for the first two acylation events, which compares favorably with previously reported *redesign* studies using natural protein scaffolds (**Table 2** entries 6–9), and is ≈30x more efficient than the *de novo* designed KO-42 peptide catalyst.[23] The overall catalytic efficiency (k_{cat}/K_M = 3.7 ± 0.6 $M^{-1}s^{-1}$) is approximately equal to previous redesign attempts (**Table 2** entries 3, 8&9). In our case, the reaction is turned over expelling acetate, and it is not driven by other nucleophilic side chains near the active site, which can become irreversibly acylated as the reaction proceeds contributing to observed kinetic profiles.[18,29] Indeed, even over a long time course and >6 turnovers, we observed only one small additional peak by MALDI-TOF MS at [M+84] Da (**SI Fig. 1.7–1**), indicating limited secondary acylation. For CC-Hept-hC-H-E, the burst phase is slowed and k_{cat} improved compared with CC-Hept-C-H-E (**Fig. 4c**). This is consistent with the mechanism and designs: the lower nucleophilicity of homocysteine affecting the first step is compensated by the increased conformational flexibility of the acylated side chain, which improves turnover (**Fig. 4d**).

Our best design is ≈10x less efficient than recently reported Zn^{2+}-dependent hydrolase designs (**Table 2** entries 10, 12&13), where the reaction proceeds *via* direct hydrolysis of the substrate without a thioester intermediate, thus removing the requirement for His to activate water in the hydrolytic step; and it is still ≈10^3x less efficient than the reaction of the natural enzyme α-chymotrypsin with the unnatural substrate *p*NPA.[43,44] We envisage improving the efficiency and selectivity of our designs in the future through directed evolution of residues close to the

active site,[12] and/or via the precise installation of a surrogate for the oxyanion hole to stabilize the tetrahedral intermediates.[19]

In conclusion, this is the first report of catalytically active Cys-His-Glu triads being installed into a completely *de novo* protein framework. In total, seven such triads have been engineered into a heptameric α-helical barrel, resulting in hydrolytic activities at least equal to those achieved in previous enzyme redesigns using natural scaffolds. Moreover, the peptide-based architecture that we describe allows the facile introduction of standard and non-canonical amino acids predictably to test and control aspects of the mechanism of action of these *de novo* assemblies. Therefore, this study highlights the possibilities for using highly stable, mutable and completely *de novo* architectures as scaffolds for the precise and predictable introduction of catalytic function.

Table 2. Kinetic parameters for the hydrolysis of *p*NPA by *de novo* and redesigned catalysts at 25 °C.

No.	Catalyst	$k_2 \times 10^2$ (s^{-1})	K_S (μM)	k_2/K_S ($M^{-1}s^{-1}$)	$k_{cat} \times 10^2$ (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
1	CC-Hept-C-H-E[a,d]	0.83 ± 0.1	377 ± 67	22.0 ± 2.0	0.05 ± 0.0	134 ± 36	3.7 ± 0.6
2	CC-Hept-hC-H-E[a,d]	0.72 ± 0.1	604 ± 134	11.9 ± 1.2	0.07 ± 0.0	373 ± 60	1.9 ± 0.2
3	PZD2[d 17]	-	-	-	0.05 ± 0.0	170 ± 20	2.70
4	KO-42[d 23]	-	-	≈ 0.7	-	-	-
5	S824[e 29]	-	-	-	0.55	3000	1.83
6	FR29[b,f 19]	0.41 ± 0.1	120 ± 20	34	-	-	-
7	FR29 A44S/T112L/V151L[b,f 19]	1.54 ± 1.1	38 ± 5	405	-	-	-
8	AlleyCatE[f 18]	1.3 ± 0.2	400 ± 70	32.5 ± 2.8	0.4 ± 0.2	700 ± 300	5.5 ± 0.8
9	AlleyCatE2[f 18]	1.5 ± 0.2	200 ± 24	76.7 ± 8.3	0.12 ± 0.0	200 ± 15	6 ± 0.5
10	Ac-IHIHIQI-$CONH_2$ [c,g 25]	-	-	-	2.6 ± 0.0	400 ± 100	62 ± 2
11	$(TRIL9CL23H)_3^+$ [c,f 26]	-	-	-	0.02 ± 0.01	1600 ± 400	1.38 ± 0.04
12	MID1-zinc[c 27]	-	-	-	4.2 ± 0.1	1180 ± 100	35
13	^{A104}AB3[c,h 28]	-	-	-	2.7 ± 0.8	850	32 ± 8

[a] 22 ± 1 °C. [b] 29 °C and with *p*-nitrophenyl-2-phenylpropanoate. [c] In the presence of Zn^{2+}. [d] pH 7.0. [e] pH 7.3. [f] pH 7.5. [g] pH 8.0. [h] pH 9.0.

Methods

Peptide synthesis and purification

Peptides were synthesised by standard Fmoc solid-phase peptide synthesis (SPPS) methods on a CEM (Buckingham, UK) Liberty Blue automated synthesizer with inline UV monitoring on a 0.1 mmol scale. Peptide coupling was achieved using DIC/HOBt activation and standard coupling protocols on Rink amide ChemMatrix resin. All peptides were cleaved from the resin as C-terminal amides. Iterative Fmoc deprotections were carried out with 20% v/v piperidine or morpholine in DMF. All peptides were N-terminally acetylated on addition of acetic anhydride (0.25 ml) and pyridine (0.30 ml) in DMF (5 ml) for 20 min at RT. Cleavage from the solid support and side-chain deprotection (with the exception of CC-Hept-βMC-H-E; **Supplementary methods**) was effected by addition of 15 ml of a 95:2.5:2.5 v/v TFA:H_2O:TIPS solution followed by shaking for 3 h at rt. Following cleavage, the solution was reduced to ca. 5 ml in volume under a flow of nitrogen. The crude peptides were precipitated using cold diethyl ether (45 ml), recovered via refrigerated centrifugation at 4 krpm, re-dissolved in 1:1 CH_3CN:H_2O (10 ml) and lyophilized to yield a white solid.

Crude peptides were purified by reversed-phase HPLC running a gradient of 20 to 80% CH_3CN in H_2O over 30 min on a Phenomenex (Macclesfield, UK) Luna® 5 μm particle size, 100 Å pore size, C18 stationary-phase column (150 × 10 mm), with UV detection at 220 and 280 nm. More-hydrophobic and shortened peptides were purified with gradients of 40 – 100% or 0 – 60% CH_3CN in H_2O over 30 min, respectively, on the same column. All buffers contained 0.1% TFA. Pure fractions were identified by analytical HPLC and MALDI-TOF MS (**Supplementary methods**), and were pooled and lyophilized.

UV/vis activity assays

Activity assays were carried out in a 1 or 10 mm path length quartz cuvette on a Perkin-Elmer (Coventry, UK) Lambda 25 spectrophotometer, recording at 348 nm (the pH-independent isosbestic point for *p*-nitrophenol (*p*NP; ε = 5,400 $cm^{-1}M^{-1}$)) at 20 °C. Solutions were made up in HEPES-buffered saline (HBS) to a final peptide concentration of 50 μM with a 10-fold excess of TCEP for all peptides. A freshly prepared solution of *p*NPA in CH_3CN was added (300 μM final concentration; the final CH_3CN content was 5%) and measurements were taken in triplicate at 30 s intervals to a total time of 1 or 3 h. Data were background corrected (measured in the presence of TCEP), and errors obtained from the standard deviation of the repeats.

Kinetic assays

Kinetic assays were carried out in a 10 mm path length quartz cuvette on an Agilent (Stockport, UK) Cary-100 UV/vis spectrophotometer recording at 348 nm at 22 ± 1 °C. Solutions were made up in HBS to a final peptide concentration of 10 μM with a 10-fold excess of TCEP. A freshly prepared solution of *p*NPA in CH_3CN was added (100–1250 μM final concentration; the final CH_3CN content was 5%) and measurements were taken in triplicate at 1 s intervals at a 0.1 s averaging time to a total time of 0.5, 1 or 16 h (for extended time course measurements). Data were background corrected (measured in the presence of TCEP), and errors obtained from the standard deviation of the repeats. Fitting of the obtained curves was carried out following the method of Bender *et al.*[42] Where k_2, K_S, k_{cat} and K_M can

be obtained by fitting of pre-steady state kinetic curves to an equation with an exponential and a linear term. In the case of CC-Hept-C-H-E and CC-Hept-hC-H-E fitting to a double exponential and a linear term returned a good fit to the data (**SI Fig. 1.11–1**). The data were fitted to **Equation 1**, where the sum of the pre-exponentials *a* and *c* approximately equaled the concentration of peptide in the assay (10 μM), *b* and *d* provide rate constants k_{2a} and k_{2b}, and *g* provides rate constant k_3.

$$f = a\left(1 - e^{(-bx)}\right) + c\left(1 - e^{(-dx)}\right) + gx \qquad \textbf{Equation 1}$$

The rate constants were evaluated by fitting of the individual blank-subtracted (measured in the presence of TCEP) progress curves to **Equation 1** in SigmaPlot 13.0. The kinetic parameters (k_2, K_S, k_{cat} and K_M) were determined *via* fitting of the obtained rate constants across the range of substrate concentrations to a single-site saturation binding equation in SigmaPlot 13.0. $[E]_0$ = 10 μM, the concentration of peptide chains in the assay.

Keywords

De novo protein design, coiled coil, α-helical barrels, enzyme design, bioorganic catalysis.

Acknowledgements

AJB thanks the EPSRC-funded Bristol Chemical Synthesis Centre for Doctoral Training (EP/G036764/1) and the University of Bristol for the provision of a PhD studentship. AJB, ART, WMD and DNW are supported by the European Research Council (340764). DNW holds a Royal Society Wolfson Research Merit Award. We thank the Diamond Light Source (Didcot, UK) for access to beamlines I03, I04 and I24 (award MX-8922), and Dr. Franziska Thomas and members of the Woolfson group for helpful discussions.

Author contributions

AJB, ART and DNW designed the research. AJB synthesized the peptides and the β-Me-Cys amino acid, performed the solution-phase biophysical analyses and solved the X-ray crystal structures. AJB and WMD performed the kinetic experiments. AJB and DNW wrote the manuscript. All authors analysed the data, and reviewed and contributed to the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Accession codes

Coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 5ez8, 5ez9, 5eza, 5ezc, 5f2y and 5eze for CC-Hept-I-C-I, CC-Hept-I-H-I, CC-Hept-C-H-I, CC-Hept-C-H-E, CC-Hept-hC-H-E and CC-Hept-βMC-H-E, respectively.

References

1 Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409**, 258-268, (2001).

2 Nanda, V. & Koder, R. L. Designing Artificial Enzymes by Intuition and Computation. *Nat. Chem.* **2**, 15-24, (2010).

3 Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem* **82**, 447-470, (2013).

4 Woolfson, D. N. *et al.* De novo protein design: how do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* **33**, 16-26, (2015).

5 Kiss, G., Çelebi‐Ölçüm, N., Moretti, R., Baker, D. & Houk, K. Computational enzyme design. *Angew. Chem., Int. Ed.* **52**, 5700-5725, (2013).

6 Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817-1819, (2010).

7 Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS ONE* **6**, e19230, (2011).

8 Rajagopalan, S. *et al.* Design of activated serine–containing catalytic triads with atomic-level accuracy. *Nat. Chem. Biol.* **10**, 386-391, (2014).

9 Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195, (2008).

10 Jiang, L. *et al.* De Novo Computational Design of Retro-Aldol Enzymes. *Science* **319**, 1387-1391, (2008).

11 Siegel, J. B. *et al.* Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **329**, 309-313, (2010).

12 Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567-573, (2009).

13 Blomberg, R. *et al.* Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418-421, (2013).

14 Kries, H., Blomberg, R. & Hilvert, D. De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **17**, 221-228, (2013).

15 Giger, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494-498, (2013).

16 Preiswerk, N. *et al.* Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8013-8018, (2014).

17 Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14274-14279, (2001).

18 Moroz, Y. S. *et al.* New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *J. Am. Chem. Soc.* **137**, 14905-14911, (2015).

19 Richter, F. *et al.* Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **134**, 16197-16206, (2012).

20 Kaplan, J. & DeGrado, W. F. De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11566-11570, (2004).

21 Zastrow, M. L. & Pecoraro, V. L. Designing functional metalloproteins: From structural to catalytic metal sites. *Coord. Chem. Rev.* **257**, 2565-2588, (2013).

22 Armstrong, C. T., Watkins, D. W. & Anderson, J. L. R. Constructing manmade enzymes for oxygen activation. *J. Chem. Soc., Dalton Trans.* **42**, 3136-3150, (2013).

23 Broo, K. S., Brive, L., Ahlberg, P. & Baltzer, L. Catalysis of Hydrolysis and Transesterification Reactions of p-Nitrophenyl Esters by a Designed Helix−Loop−Helix Dimer. *J. Am. Chem. Soc.* **119**, 11362-11372, (1997).

24 Nilsson, J. & Baltzer, L. Reactive-Site Design in Folded-Polypeptide Catalysts-The Leaving Group pKa of Reactive Esters Sets the Stage for Cooperativity in Nucleophilic and General-Acid Catalysis. *Chem. Eur. J.* **6**, 2214-2220, (2000).

25 Rufo, C. M. *et al.* Short peptides self-assemble to produce catalytic amyloids. *Nat. Chem.* **6**, 303-309, (2014).

26 Zastrow, M. L., Peacock, A., F. A., Stuckey, J. A. & Pecoraro, V. L. Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **4**, 118-123, (2012).

27 Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a De Novo Zinc-Mediated Protein Interface: Implications for Natural Enzyme Evolution and Rational Enzyme Engineering. *Biochemistry* **51**, 3933-3940, (2012).

28 Song, W. J. & Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **346**, 1525-1528, (2014).

29 Wei, Y. & Hecht, M. H. Enzyme‐like proteins from an unselected library of designed amino acid sequences. *Prot. Eng. Des. Sel.* **17**, 67-75, (2004).

30 Woolfson, D. N. The Design of Coiled-Coil Structures and Assemblies. *Adv. Prot. Chem.* **70**, 79-112, (2005).

31 Lupas, A. Coiled coils: new structures and new functions. *Trends Biochem. Sci* **21**, 375-382, (1996).

32 Calladine, C. R., Sharff, A. & Luisi, B. How to untwist an α-helix: structural principles of an α-helical barrel. *J. Mol. Biol.* **305**, 603-618, (2001).

33 Woolfson, D. N., Bartlett, G. J., Bruning, M. & Thomson, A. R. New currency for old rope: from coiled-coil assemblies to α-helical barrels. *Curr. Opin. Struct. Biol.* **22**, 432-441, (2012).

34 Zaccai, N. R. *et al.* A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.* **7**, 935-941, (2011).

35 Thomson, A. R. *et al.* Computational design of water-soluble α-helical barrels. *Science* **346**, 485-488, (2014).

36 Burton, A. J. *et al.* Accessibility, Reactivity, and Selectivity of Side Chains within a Channel of de Novo Peptide Assembly. *J. Am. Chem. Soc.* **135**, 12524-12527, (2013).

37 Smith, A. J. T. *et al.* Structural Reorganization and Preorganization in Enzyme Active Sites: Comparisons of Experimental and Theoretically Ideal Active Site Geometries in the Multistep Serine Esterase Reaction Cycle. *J. Am. Chem. Soc.* **130**, 15361-15373, (2008).

38 Wood, C. W. *et al.* CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics* **30**, 3029-3035, (2014).

39 Fletcher, J. M. *et al.* A Basis Set of de Novo Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. *ACS Synth. Biol.* **1**, 240-250, (2012).

40 Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481-485, (2014).

41 Pellegrini-Calace, M., Maiwald, T. & Thornton, J. M. PoreWalker: A Novel Tool for the Identification and Characterization of Channels in Transmembrane Proteins from Their Three-Dimensional Structure. *PLoS Comput. Biol.* **5**, e1000440, (2009).

42 Bender, M. L., Kezdy, F. J. & Wedler, F. C. alpha-Chymotrypsin: Enzyme concentration and kinetics. *J. Chem. Educ.* **44**, 84-87, (1967).

43 Kezdy, F. J. & Bender, M. L. The Kinetics of the α-Chymotrypsin-Catalyzed Hydrolysis of p-Nitrophenyl Acetate. *Biochemistry* **1**, 1097-1106, (1962).

44 Faller, L. & Sturtevant, J. M. The Kinetics of the α-Chymotrypsin-catalyzed Hydrolysis of p-Nitrophenyl Acetate in Organic Solvent-Water Mixtures. *J. Biol. Chem.* **241**, 4825-4834, (1966).