Cofactor binding and enzymatic activity in an unevolved superfamily of *de novo* designed 4-helix bundle proteins

Shona C. Patel,[1] Luke H. Bradley,[2] Sayuri P. Jinadasa,[2] and Michael H. Hecht[2]*

[1]Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544
[2]Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Received 1 March 2009; Revised 12 April 2009; Accepted 13 April 2009
DOI: 10.1002/pro.147
Published online 29 April 2009 proteinscience.org

Abstract: To probe the potential for enzymatic activity in unevolved amino acid sequence space, we created a combinatorial library of *de novo* 4-helix bundle proteins. This collection of novel proteins can be considered an "artificial superfamily" of helical bundles. The superfamily of 102-residue proteins was designed using binary patterning of polar and nonpolar residues, and expressed in *Escherichia coli* from a library of synthetic genes. Sequences from the library were screened for a range of biological functions including heme binding and peroxidase, esterase, and lipase activities. Proteins exhibiting these functions were purified and characterized biochemically. The majority of *de novo* proteins from this superfamily bound the heme cofactor, and a sizable fraction of the proteins showed activity significantly above background for at least one of the tested enzymatic activities. Moreover, several of the designed 4-helix bundles proteins showed activity in all of the assays, thereby demonstrating the functional promiscuity of unevolved proteins. These studies reveal that *de novo* proteins—which have neither been designed for function, nor subjected to evolutionary pressure (either *in vivo* or *in vitro*)—can provide rudimentary activities and serve as a "feedstock" for evolution.

Keywords: binary code; protein design; biomolecular evolution; 4-helix bundle; synthetic biology

Introduction

What were the functional capabilities of primitive proteins before they evolved into the active and specific biocatalysts that exist today? In 1976, Jensen suggested that "primitive enzymes possessed a very broad specificity, permitting them to react with a wide range of related substrates." Further, he hypothesized that broad specificity would have facilitated life at the early stages of evolution because it would "maximize the catalytic versatility of an ancestral cell that functioned with limited enzyme resources."[1] Jensen's hypothesis is difficult to test because unevolved proteins no longer exist, and the proteins in modern organisms have already been subjected to eons of selective pressure for specific biological functions.

An alternative approach toward understanding the properties of unevolved sequence space is to create a collection of amino acid sequences *de novo*, thereby ensuring that the proteins are not biased by billions of years of evolutionary history. Assessing the properties

Abbreviations: ABTS, 2,2′-azino-di(3-ethyl-benzthiazoline-6-sulfonic acid); PLE, porcine liver esterase.

Additional Supporting Information may be found in the online version of this article.

The authors dedicate this article to the memory of Walter Kauzmann, a long time member of the Princeton Department of Chemistry. His early insights into the importance of the hydrophobic effect in protein folding were decades ahead of their time, and laid the foundation for the protein design strategy described in this article.

Shona C. Patel's current address is Merck & Co., Inc., West Point, PA 19486.

Luke H. Bradley's current address is Departments of Anatomy & Neurobiology, Molecular & Cellular Biochemistry, University of Kentucky, Lexington, KY 40536.

Sayuri P. Jinadasa's current address is Columbia University Medical Center, New York, NY 10032.

Grant sponsor: NSF grant; Grant number: MCB-0817651.

**Correspondence to:* Michael H. Hecht, Department of Chemistry, Princeton University, Princeton, NJ, 08544.
E-mail: hecht@princeton.edu

of such collections would facilitate an understanding of the structural and functional capabilities of unevolved sequence space.

In principle, an ideal bias-free collection of proteins would be a combinatorial library of amino acid sequences constructed at random. However, the vast majority of random sequences do not fold into protein-like structures,[2–6] and thus would not be expected to exhibit biologically relevant activities. A more appropriate collection of sequences would still be combinatorially diverse, but would focus on a region of sequence space that favors folding into a stable three-dimensional structure.[7] To provide such a collection, we have designed and produced a library of sequences encoding an "artificial superfamily" of *de novo* 4-helix bundles.[8] [The term "superfamily" has been used in the literature to denote either a group of evolutionarily related proteins, or a group of structurally related proteins that are not necessarily of common ancestry (see http://pir.georgetown.edu/pirwww/otherinfo/sfdef.pdf). Since our proteins were designed *de novo*, and not evolved in nature, we use the term "superfamily" to describe a collection of proteins with similar structures.]

We previously reported the development of a "binary code" strategy for designing libraries of well-folded *de novo* proteins.[9,10] According to this strategy, the overall structure of a protein can be specified by designing the sequence periodicity of polar and nonpolar amino acids to match the structural periodicity of the desired secondary structure. Since only the type of residue—polar versus nonpolar—is designed explicitly, the strategy is inherently binary. Moreover, since the identities of the polar and nonpolar side chains are not specified explicitly, the strategy is well-suited for constructing vast combinatorial libraries of novel sequences.[8]

We have used the binary code strategy to construct several libraries of α-helical or β-sheet proteins.[8–12] Because the designed β-sheet proteins often form insoluble aggregates, our current work on artificial superfamilies focuses on α-helical structures.[11]

The binary patterned design of α-helical sequences places a hydrophobic amino acid every three or four residues, generating the following pattern: O●OO ●●OO●OO●●O, where O and ● represent polar and nonpolar residues, respectively.[9,10] Consistent with the 3.6 residues/turn periodicity of a canonical α-helix, this pattern encodes α-helices that are amphiphilic. When four such helices are linked, the hydrophobic effect[13] drives them to pack against one another, thereby forming a 4-helix bundle with nonpolar residues pointing towards the protein core, and polar residues exposed to solvent.

The construction of binary patterned libraries is facilitated by the organization of the genetic code: Six polar residues (Lys, His, Glu, Gln, Asp, and Asn) can be encoded by the degenerate DNA codon VAN, and five nonpolar residues (Met, Leu, Ile, Val, and Phe) can be encoded by the degenerate codon NTN (V = A, G, or C; N = A, G, C, or T). Using these degenerate codons, we have constructed large libraries of synthetic genes. The encoded binary patterned proteins are then expressed in *Escherichia coli* and readily purified.[14]

To date, we have constructed three libraries of binary patterned 4-helix bundles. The first library encoded 74-residue sequences, and produced structures that were mostly dynamic and only moderately stable.[9] Subsequently, a 2nd generation library was designed using an elongated template encoding 102 residues. Five proteins from the 2nd generation library were characterized in detail.[12] Of these, four were shown to form stable well-ordered structures; and two of these were solved by NMR.[7,15] Both proteins—S824 and S836—formed well-folded 4-helix bundles, and as designed, the hydrophobic residues were sequestered in the core and the hydrophilic residues were exposed on the exterior. Although the overall topologies of the two structures were similar, the sequences, local packing interactions, and dynamic properties were different.[7]

In addition to the structural studies of S824 and S836, the spectroscopic and thermodynamic properties of several proteins from the 1st and 2nd generation libraries were characterized.[9,12] Taken together, these studies confirmed that the binary code for protein design can produce vast collections of folded α-helical proteins. Among these collections, some proteins are extremely stable and well ordered, while others are less stable and more dynamic.[7,12,16–18] Moreover, although not specifically designed for function, several proteins from the 1st and 2nd generation libraries were found to bind cofactor and/or catalyze reactions.[19–21] In the current study, we demonstrate that proteins from a 3rd generation library of binary patterned 4-helix bundles display a range of activities, thereby establishing the functional potential of an unevolved artificial superfamily of proteins.

Results

An artificial superfamily of de novo 4-helix bundles

Although the thermodynamic and structural properties of the 2nd generation library were appropriate for the functional studies described below, it was a small library with limited sequence diversity. To facilitate a more meaningful probe of the functional potential of a designed superfamily of proteins, we constructed a 3rd generation library comprising $\sim 10^6$ binary patterned sequences designed to fold into 102-residue 4-helix bundles. The construction and biophysical characterization of this library are described elsewhere.[8,22] The design template for this library is illustrated in Figure S1 in the Supplementary Material. It must be emphasized that this library (and all previous binary patterned libraries) was designed for structure; no

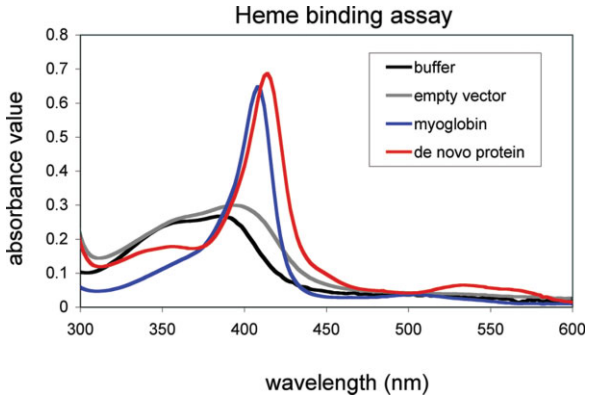

Figure 1. Absorption spectra for a representative heme binding assay. Heme incubated in a clarified lysate from cells expressing a 3rd generation *de novo* protein produces a sharp Soret peak, signifying heme binding. The negative controls are heme incubated in buffer, and heme incubated in a clarified lysate from *E. coli* cells harboring the empty vector. Both negative controls show broad absorption with no significant peaks. The positive control is purified myoglobin, which yields a sharp Soret peak.

attempts were made to explicitly design binding or activity. Therefore, this library is well-suited for studying the functional potential of an unevolved superfamily of proteins.

De novo proteins from an artificial superfamily bind cofactors and display catalytic activity

The Enzyme Classification (EC) system[23] and the CATH databases (Class, Architecture, Topology, and Homologous superfamily)[24] show that many α-helical proteins in nature function as hydrolases or oxidoreductases. Two common types of hydrolases are esterases and lipases, and a well-studied class of oxidoreductases is the heme-based peroxidases, which convert hydrogen peroxide to water. As an initial probe of the binding and catalytic potential of our unevolved superfamily of novel proteins, we screened our new (3rd generation) library of helical proteins for heme binding, and for peroxidase, esterase, and lipase activities.

Heme binding was detected by a bright red color and the presence of a Soret peak at 412 nm. The absorbance spectra in Figure 1 show typical results for the heme binding assay. Myoglobin, a natural heme protein, has a sharp Soret peak, as does a lysate from cells expressing a *de novo* 4-helix bundle protein from the 3rd generation library. In contrast, heme incubated with buffer shows only a broad absorbance from 300 to 450 nm. A similar featureless spectrum is observed for heme added to a clarified lysate prepared from cells harboring the empty vector (pET-3a). This demonstrates that background endogenous *E. coli* proteins do not exhibit significant heme binding, and assays for the ability of highly expressed *de novo* proteins to bind heme can be performed in the context of clarified *E. coli* lysates.

Representative results from 96-well plates are shown in Figure 2(A). Samples with heme in buffer or in clarified lysates from cells harboring the empty pET-3a vector produce a dim green/brown color. In contrast, heme incubated with lysates from cells expressing a *de novo* 4-helix bundle protein produces a bright red color. This colorimetric assay for heme binding was performed on clarified lysates from cells harboring expression plasmids encoding 360 different *de novo* sequences. Figure 3 shows the spectra for each well in a typical 96-well plate. The variations in peak intensities are easily detected. Spectra for the negative controls, buffer and empty vector, are shown in wells A1 and B1, respectively. Soret peaks are not seen in either control. In contrast, a sharp Soret peak is clearly visible in wells C1-H1, which contain six previously characterized *de novo* proteins from all three libraries. Each of the other wells contains the lysate from cells expressing arbitrarily chosen sequences from the 3rd generation library. The level of heme binding was estimated by the relative height of the peak at 412 nm. The distribution of heme binding for 360 randomly chosen clones from the 3rd generation library is shown in the top left panel of Figure 4. Each bar in the graph represents a single protein and the values are ranked from lowest to highest.

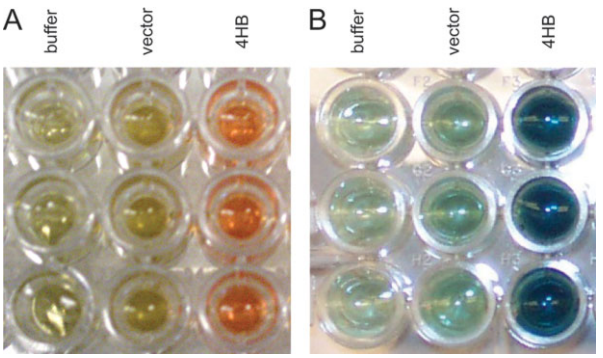

Figure 2. Colorimetric assays in 96-well plates. Each column is a different sample performed in triplicate. (A) Heme binding assay. Column 1 shows heme in buffer, yielding a light green/brown color. Column 2 shows heme incubated in lysates from cells harboring the empty vector. These samples also show a light green/brown color, indicating a low background from endogenous *E. coli* proteins. Column 3 shows heme incubated in lysates from cells expressing a *de novo* 4-helix bundle protein from the 3rd generation library. The pink/red color indicates heme binding. (B) Peroxidase activity using ABTS as a colorimetic assay. Column 1 shows the background peroxidase activity for heme in buffer. Column 2 shows the peroxidase activity for heme in lysates from cells containing the empty vector. This represents the peroxidase activity of background *E. coli* proteins. Column 3 shows the activity for heme added to lysates from cells expressing a 4-helix bundle protein from the 3rd generation library. Significant peroxidase activity produces the dark teal color.

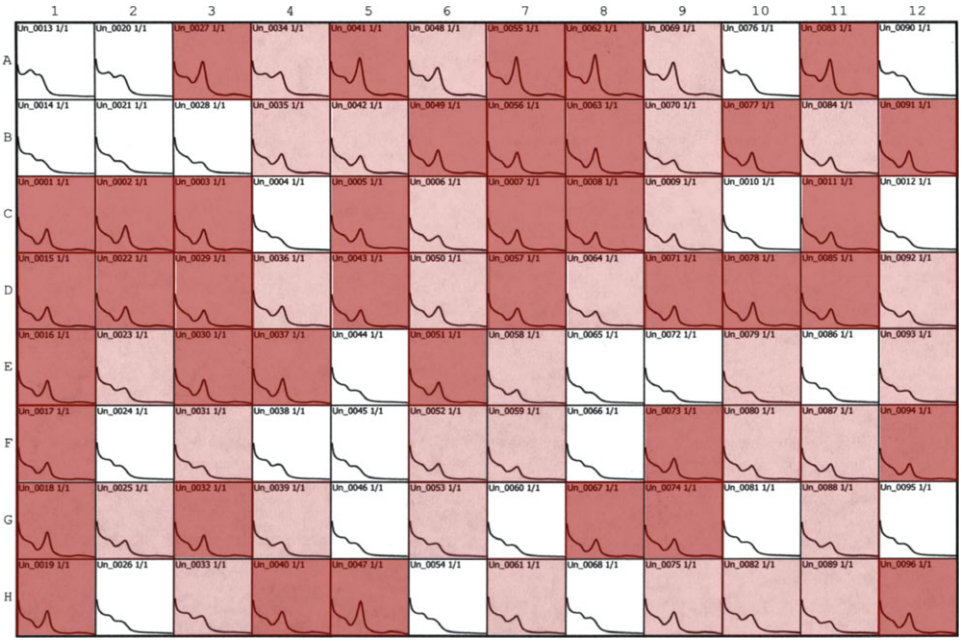

Figure 3. Heme binding screen. Spectra of wells in a 96-well plate containing heme added to *de novo* proteins in *E. coli* cell lysates. White boxes represent proteins that do not bind heme, pink boxes represent proteins showing moderate heme binding, and red boxes represent proteins showing high heme binding. Box A1 is heme in buffer alone, and Box B1 is heme in a lysate from cells harboring the empty vector. Boxes C1-H1 are for heme added to previously characterized proteins from the 2nd and 3rd generation libraries. The remaining boxes are randomly chosen protein samples from the 3rd generation library.

Approximately 66% of the library shows heme binding, and half of these (33%) show relatively high levels of heme binding.

A number of natural heme proteins have peroxidase activity. The ability of our novel heme proteins to catalyze the peroxidase reaction was assayed using 2,2′-azino-di(3-ethyl-benzthiazoline-6-sulfonic acid) (ABTS) as a reducing agent. ABTS is frequently used to monitor peroxidase activity because conversion of ABTS to the oxidized form yields a teal color with an absorbance at 650 nm. A typical peroxidase assay is shown in Figure 2(B), where the background activity shows very little color and the activity of the *de novo* 4-helix bundle protein is readily apparent. The *de novo* protein yields a dark teal color representing activity well above the background levels. This colorimetric assay was performed for the same 360 clones as the heme binding assay, and the absorbance levels were recorded at a single time point. As shown in the top right panel of Figure 4, ~50% of the sequences show some level of peroxidase activity, and several proteins show catalytic activity that is substantially above the background peroxidase activity in *E. coli* lysates.

The observed peroxidase activity of our *de novo* proteins, like that of natural peroxidases, depends on the binding of the redox-active heme cofactor. To assess whether catalytic activity could be found in our unevolved proteins even in the absence of cofactors, we assayed two hydrolytic activities: esterase and lipase. These activities were measured by monitoring the hydrolysis of *p*-nitrophenyl acetate and *p*-nitrophenyl palmitate, respectively. Both hydrolysis reactions yield *p*-nitrophenol, which is yellow and can be monitored at 405 nm. As shown in the bottom panels of Figure 4, 30% of the library shows esterase activity and 20% of the library shows lipase activity above background.

The screens described above for heme binding and catalytic activity were performed on clarified cell lysates, rather than purified proteins. Therefore, it was important to demonstrate (i) that endogenous proteins in the *E. coli* lysates do not produce a significant signal, and (ii) that the *de novo* proteins express at levels far above the endogenous *E. coli* proteins. This first issue was addressed for heme binding and peroxidase activity as described earlier, and shown in Figures 1–3. Similar controls showed that clarified lysates from cells harboring the empty vector did not display significant activity in the esterase and lipase assays (data not shown).

The second issue—expression level—was also assayed explicitly. Ninety-six clones were checked for expression, and as shown in Figure 5(A), ~70% of the *de novo* proteins express at high levels. Figure 5(B) shows the correlation between protein expression, heme binding, and the three catalytic activities. A surprising result is the finding that nearly all of the *de novo* proteins that expressed were capable of binding the heme cofactor. Moreover, among the proteins that bind heme, nearly 80% show some level of peroxidase activity. For the hydrolase activities, ~60% of

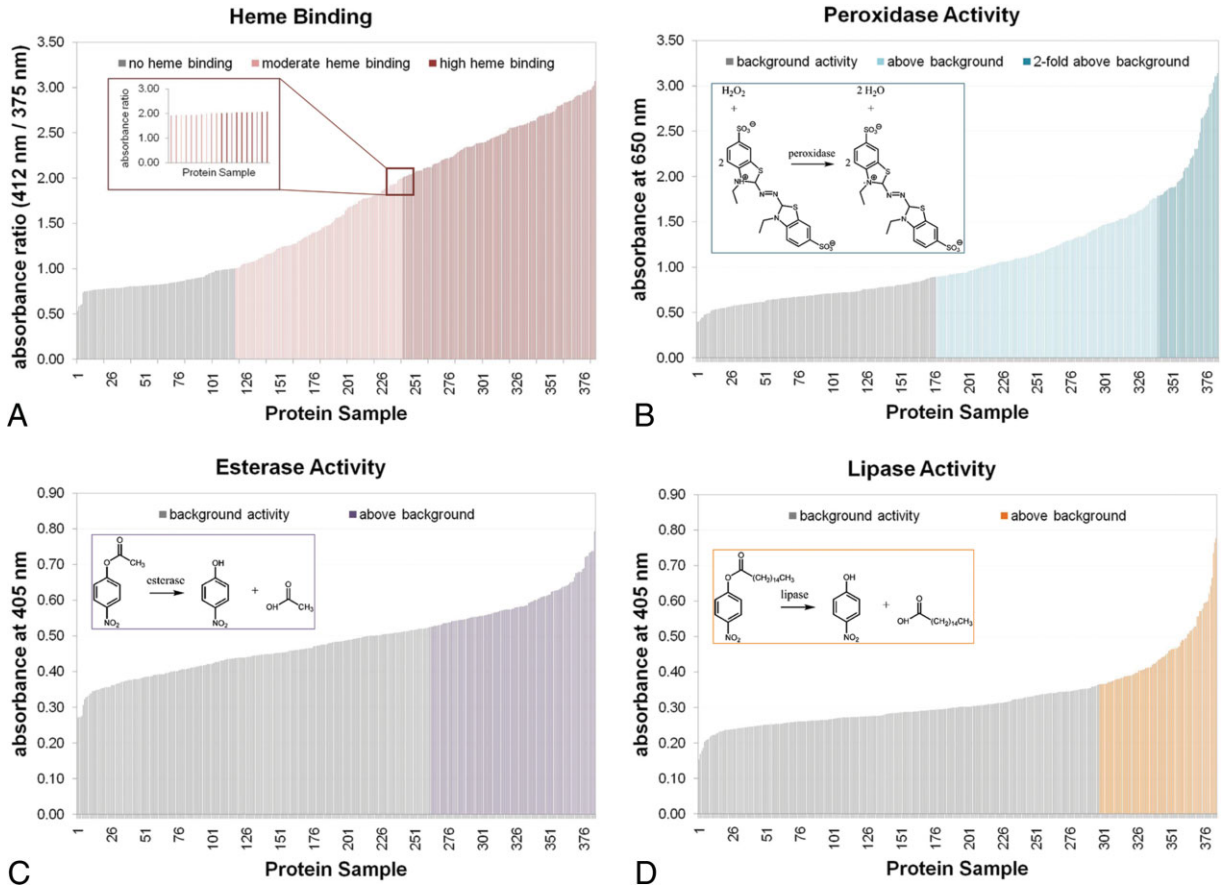

Figure 4. Frequency of active proteins in a superfamily of *de novo* 4-helix bundles. Each bar represents a different protein. Activity levels are ranked from lowest to highest. (A) Frequency of heme binding: The y-axis is the ratio of the absorbance at 412 nm (Soret peak) to the reading at 375 nm (local minimum). Gray bars represent proteins with a ratio $\leq$1. These samples do not show heme binding. Pink bars represent proteins that show heme binding with a ratio $>$1, and red bars represent proteins that show high heme binding with a ratio $>$2. Inset shows a magnification to illustrate individual bars. (B) Frequency of peroxidase activity at a single time point. Gray bars represent proteins that do not show peroxidase activity, light teal bars represent proteins with activity above background, and dark teal bars represent proteins with activity two-fold above background. (C) Frequency of esterase activity at a single time point. Gray bars represent proteins that do not show esterase activity, and purple bars represent proteins with activity above background. (D) Frequency of lipase activity at a single time point. Gray bars represent proteins that do not show activity, and orange bars represent proteins that show activity above background. For panels B–D, the background activity is calculated as the average value of the endogenous *E. coli* proteins (control cells harboring empty vector) across each 96-well plate plus three standard deviations. Insets in panels B–D show the reactions catalyzed by the *de novo* proteins.

the expressed proteins show esterase activity and 36% show lipase activity. An interesting observation is that ~30% of the highly expressed proteins exhibit some level of activity for all of the functions (binding and catalysis) that were tested.

Kinetic profiles of purified proteins

The studies described earlier and presented in Figures 2–5 were performed on clarified cell lysates, thereby enabling rapid screening of nearly 400 samples for four different activities. Although these relatively high throughput screens allowed us to assess the frequencies of cofactor binding and catalytic activity in our unselected library, cell lysates are not suitable for quantifying the levels of activity of the *de novo* proteins. To enable quantitative studies, we purified several proteins from the collections. Six proteins were chosen for purification: S824 and S836 from the 2nd generation library were chosen because their high-resolution structures are known.[7,15] WA20 and WA32 from the 3rd generation library were chosen because they had been characterized biophysically when this library was constructed.[22] Finally, T-C8 and T-D10 from the 3rd generation library were chosen solely based on their activity in the screens shown in Figure 4.

The proteins were purified using an acid precipitation step (to remove a significant fraction of *E. coli* contaminants) followed by cation exchange chromatography. Chromatograms for each protein and SDS-PAGE analysis of each step in the purification process are shown in Figure S2 in the Supplementary Material. These gels show that the purified samples are free of

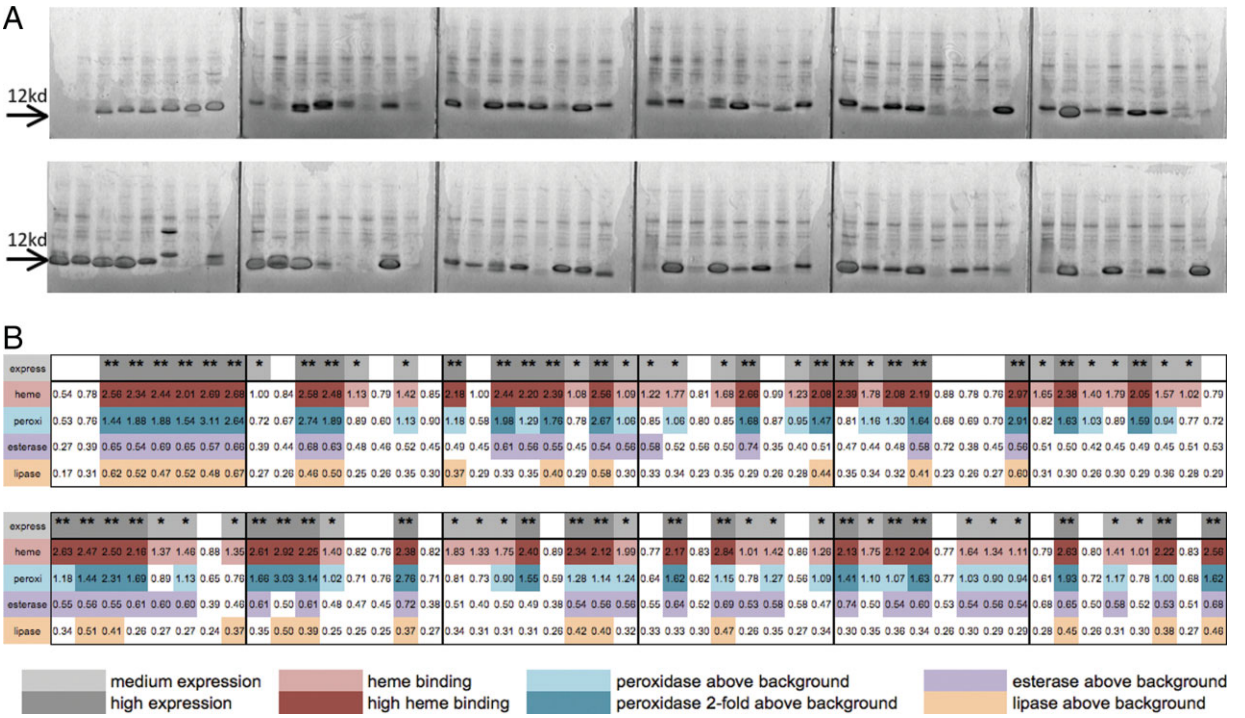

Figure 5. Expression levels of 96 arbitrarily chosen proteins, and the correlation of expression with activity. (A) Proteins were expressed using auto-induction media, and cell lysates were analyzed by SDS-PAGE. Expression is detected by the presence of a strong band at the bottom of each gel (MW 12 kD), as indicated by the arrow. Each lane shows a different protein sample. The first lane in the top row is buffer, and the second lane shows cells harboring the empty vector showing background expression of endogenous *E. coli* proteins. (B) The correlation of expression and activity. The top row in each chart shows a graphic representation of the protein gels. A single asterisk (*) represents proteins that express at moderate levels and two asterisks (**) represent proteins that express at high levels. The second row shows the heme binding. Pink represents proteins that bind heme and red represents proteins with high heme binding. The third row shows the peroxidase data with light teal boxes representing proteins that have activity above background and dark teal boxes representing proteins having activity two-fold above background. The fourth row and fifth row shows esterase and lipase data, respectively, where shaded boxes indicate proteins that have activity above background. Vertical columns of colored boxes indicate a given protein exhibits several different activities.

visible contaminants. The sequences of the purified proteins are shown in Figure 6. All sequences follow the designed binary pattern, yet the specific side chains at each polar and nonpolar position differ from one protein to another. The kinetic data for these six purified proteins for peroxidase, esterase, and lipase activities are shown in Figure 7 and summarized in Table I.

Peroxidase activity. All six of the *de novo* proteins show peroxidase activity at levels 10^5–10^6-fold above the uncatalyzed reaction [k_{cat}/k_{uncat}, Fig. 7(A), Table IA]. Nonetheless, their activities are well below that of horseradish peroxidase.[25] This is not surprising, since HRP evolved to bind and transform peroxide, whereas the binary patterned *de novo* proteins

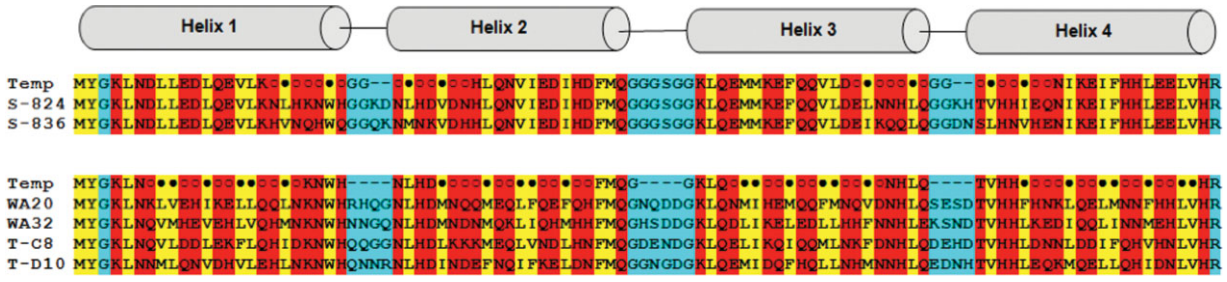

Figure 6. Sequences of proteins chosen for purification and biochemical characterization. Top: Design template for 2nd generation proteins and the amino acid sequences of S824 and S836. Bottom: Design template for 3rd generation proteins and the amino acid sequences of WA20, WA32, T-C8 and T-D10. The sequences follow the binary pattern design with red indicating polar residues and yellow indicating nonpolar residues. Turn sequences and charged residues or glycine residues are highlighted in blue.

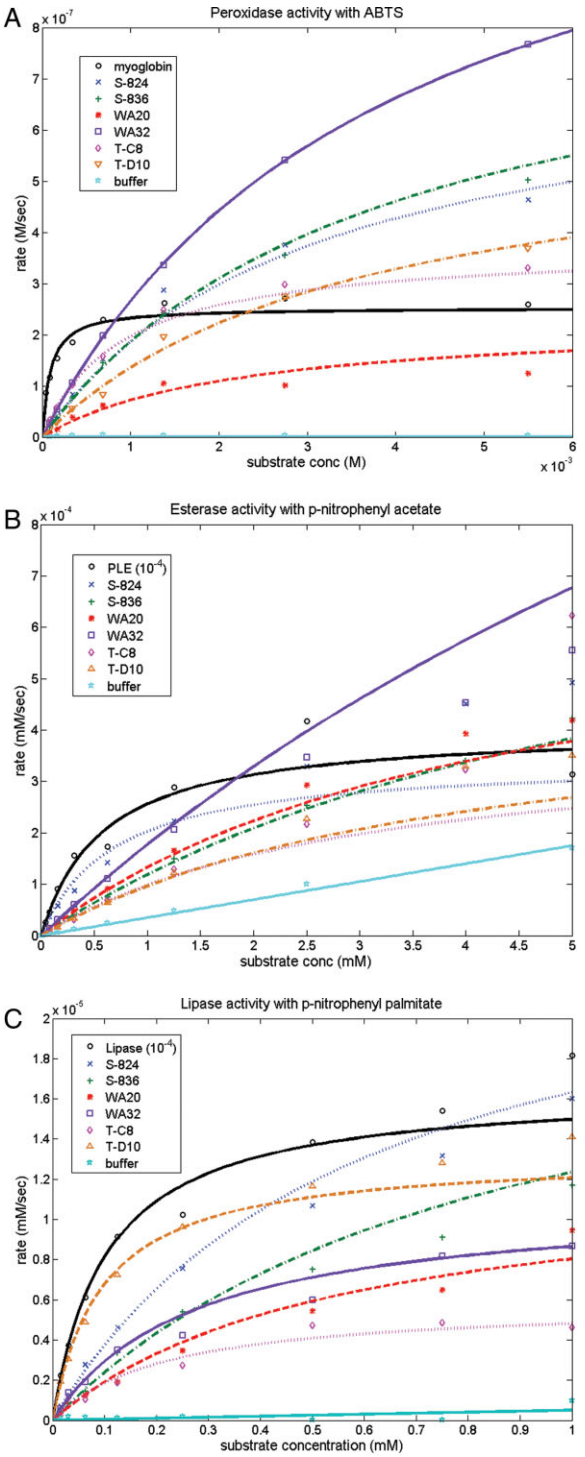

Figure 7. Kinetic profiles and rate constants calculated by Michaelis-Menten kinetics: (A) Peroxidase activity showing rate of product formation (oxidized ABTS) versus substrate (H_2O_2) concentration. The negative control includes buffer + heme + ABTS. (B) Esterase activity showing rate of *p*-nitrophenol formation versus substrate (*p*-nitrophenyl acetate) concentration. (C) Lipase activity showing rate of *p*-nitrophenol formation versus substrate (*p*-nitrophenyl palmitate) concentration.

were neither designed nor selected to catalyze this reaction. In the current studies, we compared the activities of the *de novo* proteins to that of myoglobin, which has moderate peroxidase activity.[26–28] Like our *de novo* proteins, myoglobin is an α-helical heme protein. Moreover, like our *de novo* proteins—and in contrast to natural peroxidases—myoglobin was not evolutionarily selected for peroxidase activity. As shown in Figure 7(A) and Table IA, all six of the *de novo* proteins have both higher k_{cat} and higher K_M values than myoglobin. Most of the *de novo* proteins have k_{cat}/K_M values ~10-fold lower than myoglobin. Their rate enhancements (k_{cat}/k_{uncat}), like myoglobin, are typically 10^5–10^6 above the uncatalyzed reaction.

Esterase activity. All six of the *de novo* proteins display rate enhancements (k_{cat}/k_{uncat}) for esterase activity that are 100–1000-fold higher than background (Table IB). Yet, not surprisingly, they are substantially less active than the naturally evolved porcine liver esterase (PLE). As shown in Figure 7(B), the rate profiles appear similar to PLE when this natural esterase is used at 10,000-fold lower concentration. The higher activity of the natural enzyme is due primarily to increased turnover rather than enhanced binding; thus the *de novo* proteins have K_M values similar to PLE, but k_{cat} values that are ~10,000-fold lower.

Lipase activity. As shown in Table IC, the *de novo* proteins produce rate enhancements (k_{cat}/k_{uncat}) for lipase activity 100–1000-fold above background. These activities were compared to the natural lipase from *Candida cylindracea*. As shown in Figure 7(C) and Table IC, and like the esterase activity, the *de novo* proteins are ~10,000-fold less active than the naturally evolved enzyme. As was observed for the esterase activity, the lower activity is mostly due to effects on k_{cat}, rather than K_M.

In summary, biochemical characterization of the purified *de novo* proteins confirms the results of the high throughput screens shown in Figure 4: All six of the purified proteins display enzymatic activities that are substantially above background. Furthermore, individual proteins catalyze several different reactions. Although the rate enhancements provided by the *de novo* proteins are substantially below those of natural enzymes, this is exactly what would be expected when comparing proteins from an unevolved collection (a "feedstock" for evolution) to natural proteins that have been selected for function by eons of biological evolution.

The observed activities are not due to endogenous E. coli proteins

When assaying the activity of a *de novo* protein expressed in *E. coli*, it is essential to consider the possibility that the observed activity might be due to an endogenous bacterial protein.[29] We address this

Table I. *(A) Peroxidase Rate Constants of De Novo Proteins Compared to Myoglobin. (B) Esterase Rate Constants of De Novo Proteins Compared to Porcine Liver Esterase. (C) Lipase Rate Constants of De Novo Proteins Compared to Lipase from Candida cylindracea*

Protein	k_{cat} (s^{-1})	K_M (M)	k_{cat}/K_M (s^{-1} M^{-1})	k_{cat}/k_{uncat}
A				
Myoglobin	0.051	8.8×10^{-5}	574	4.7×10^{5}
S-824	0.15	3.1×10^{-3}	49	1.4×10^{6}
S-836	0.18	3.8×10^{-3}	47	1.7×10^{6}
WA20	0.046	2.2×10^{-3}	21	4.3×10^{5}
WA32	0.26	4.0×10^{-3}	67	2.4×10^{6}
T-C8	0.074	8.9×10^{-4}	83	6.9×10^{5}
T-D10	0.13	3.6×10^{-3}	35	1.2×10^{6}
B				
PLE	81	5.8×10^{-4}	1.4×10^{5}	2.3×10^{6}
S-824	6.8×10^{-3}	6.9×10^{-4}	9.9	1.9×10^{2}
S-836	1.7×10^{-2}	6.3×10^{-3}	2.8	5.0×10^{2}
WA20	1.4×10^{-2}	4.3×10^{-3}	3.3	4.0×10^{2}
WA32	4.6×10^{-2}	1.2×10^{-2}	3.8	1.3×10^{3}
T-C8	8.0×10^{-3}	3.0×10^{-3}	2.6	2.3×10^{2}
T-D10	9.7×10^{-3}	4.0×10^{-3}	2.4	2.8×10^{2}
C				
Lipase	3.3	1.7×10^{-4}	1.9×10^{4}	6.6×10^{6}
S-824	5.2×10^{-4}	6.0×10^{-4}	0.87	1.0×10^{3}
S-836	4.6×10^{-4}	8.7×10^{-4}	0.53	9.2×10^{2}
WA20	2.5×10^{-4}	5.5×10^{-4}	0.45	5.0×10^{2}
WA32	2.2×10^{-4}	2.8×10^{-4}	0.80	4.4×10^{2}
T-C8	1.2×10^{-4}	2.2×10^{-4}	0.54	2.3×10^{2}
T-D10	2.6×10^{-4}	9.5×10^{-5}	2.8	5.2×10^{2}

possibility both for our high throughput screens on cell lysates, and for our kinetic studies on purified proteins.

To demonstrate that endogenous *E. coli* proteins are not responsible for the activities observed in cell lysates we consider the following two results: (i) Lysates from cells harboring the empty vector do not bind heme and do not catalyze the reactions. These controls for heme binding and peroxidase activity are shown in Figure 2; similar controls for the esterase and lipase reactions were also negative (data not shown). (ii) As shown in Figure 5(B), lysates from cells that contain a *de novo* DNA sequence, but which fail to express a *de novo* protein do not display any of the measured activities. Some of these non-expressing clones contain intact sequences, while others contain frameshifts or stop codons. Either way, if there is no expression, there is no activity. Thus, the presence of a novel DNA sequence and/or the induction of the T7 expression system *per se* are not sufficient to yield activity. The observed activities require that the *de novo* protein must be expressed.

To further rule out contaminating *E. coli* proteins as the source of the observed activities, quantitative measurements of enzyme activity (Fig. 7 and Table I) were performed on proteins that had been purified to a point where they were free of visible contaminants on SDS-PAGE gels (Fig. S2). Nonetheless, one must consider the possibility that an *E. coli* protein that is too dilute to see on a gel might have co-purified with the *de novo* protein, and even at this dilution might have sufficient activity to account for the observed catalysis. This concern is addressed by the following two considerations: (i) In our initial work probing esterase activity in *de novo* proteins, we used a binary patterned sequence containing a stop codon at the third codon to control for the possibility that an *E. coli* protein had co-purified with our 4-helix bundles. Cells expressing this truncated sequence were subjected to the identical purification protocol used for an intact 4-helix bundle. Although the purified protein showed esterase activity, the same chromatographic fraction from the mock purification was not active.[21] (ii) The six proteins purified for the current study eluted from the ion exchange column at different positions in the salt gradient (Fig. S2 and Table SI in Supplementary Material). An endogenous protein would not co-purify with all these fractions.

In summary, we have carried out an extensive series of control experiments to convince ourselves that despite the surprising nature of our results, the observed activities are due to the *de novo* proteins themselves, and do not result from contamination by natural proteins from *E. coli*.

Discussion

A combinatorial library of *de novo* α-helical proteins was used to investigate the functional potential of an unevolved superfamily. The proteins were designed using the binary code strategy, which partitions polar and nonpolar side chains into the core and exterior of a structure, respectively. The proteins were not

explicitly designed for any specific type of binding or activity; thus this library can be viewed at a model for the "feedstock" of evolution. Here, we have begun to assess the functional potential of this pre-evolved feedstock.

Most of the de novo designed proteins bind heme

Heme proteins are fairly abundant in nature, comprising 5% of proteins in the PDB.[30] Among natural heme proteins, the majority of structures are mainly α helical (77%).[30] Most of these form orthogonal α-helical structures, such as myoglobin and hemoglobin. However, heme proteins that form up–down 4-helix bundles (similar to proteins in our *de novo* superfamily) are not uncommon. Examples include cytochrome b562 and cytochrome c′.

In our designed superfamily of 4-helix bundles, the vast majority of sequences bind the heme co-factor. Indeed, for those proteins that expressed at levels sufficient for the assay, nearly all (99%) bind heme. Our finding that heme binding is so easily achieved by unevolved *de novo* α-helical proteins suggests that at early stages of biological evolution, promiscuous binding of the heme cofactor might have been fairly common among α-helical structures.

Cofactors, such as heme, can be thought of as "pre-organized activity modules" capable of endowing proteins with a range of functions that may be difficult or impossible to achieve using a polypeptide sequence alone.[20] Hence, cofactor binding would have enhanced "the catalytic versatility of an ancestral cell that functioned with limited enzyme resources,"[1] thereby providing early proteomes with a wider range of biochemical functions. With the passage of time, as proteins evolved towards specialized functions, some α-helical heme proteins became highly efficient as oxidoreductases (e.g., horseradish peroxidase), electron transfer agents (e.g., cytochromes), or oxygen carriers (e.g., hemoglobin). Other early α-helical heme proteins may have lost their ability to bind heme as they evolved towards functions seen in non-heme proteins today. For example, ROP is a 4-helix bundle that evolved to bind RNA and does not bind heme. Interestingly, it has been shown that by mutating just a few side chains, ROP can readily be converted into a heme binding protein.[31]

A majority of the de novo α-helical heme proteins possess peroxidase activity

The heme cofactor, with its pre-organized macrocycle, delocalized electrons in the porphyrin ring, and redox-active metal is well-suited for catalyzing oxidoreductase reactions. In our unevolved library, fully 80% of the proteins that bound heme were also capable of catalyzing the peroxidase reaction. Among these, several showed rate enhancements 10^5 to 10^6-fold above background (Table IA). Thus, binding of this pre-organized activity module readily imparts activity into polypeptide chains that were neither designed nor selected for enzymatic function.

De novo α-helical proteins also display activity in the absence of cofactors

It has been estimated that ~50% of natural enzymes harbor a metal and/or other cofactor.[32] The remainder—half of the known enzymes—are able to catalyze reactions using only those chemical moieties found in the 20 amino acids and the polypeptide backbone. To assess the capabilities for unassisted catalytic activity in an unevolved α-helical superfamily, we measured two hydrolytic activities: esterase and lipase. We found that 30% of the sequences in the library show esterase activity and 20% show lipase activity. The esterase and lipase rate constants for the purified *de novo* proteins were 100–1000-fold above background. While this is a significant enhancement relative to the uncatalyzed reactions, it is considerably lower than is observed for natural esterases and lipases. Thus, both the frequency of "hits" and the levels of activity measured for the individual hits were much lower for the hydrolase screens than for the peroxidase screen. Presumably, this is because the hydrolase activities rely on the protein alone, whereas the peroxidase activity has the benefit of the bound heme.

Functional proteins occur frequently in unselected libraries

Our results suggest that libraries of novel sequences, which were neither selected by evolution (*in vivo* or *in vitro*) nor explicitly designed for function, can nonetheless yield proteins that bind biological cofactors and catalyze reactions. This is a surprising and unexpected result. Therefore, it is important to compare our findings with those observed in other systems.

Heme binding. Promiscuous heme binding was highlighted by a recent study of antibodies, which demonstrated that immunoglobulins have an intrinsic propensity to bind heme.[33] The immune system takes advantage of this binding to acquire new antigenic specificities. Moreover, sequestering free heme by antibodies is advantageous in fighting various pathological states that are accompanied by release of free heme into the circulation. The field of *de novo* design has also shown that heme binding is relatively easy to achieve. Many peptides and proteins have been designed to bind heme.[34–40] The diversity of these designs—and the frequency of their success—suggests that a great variety of sequences containing histidine side chains can readily bind heme. Indeed, the inherent propensity of polypeptides to bind heme was observed in one of the first designs of a novel heme protein, when it was found that not only the designed sequence bound heme, but a control "retro" sequence synthesized backwards also bound the cofactor.[34]

Peroxidase activity. Peptides and proteins that bind heme often display peroxidase activity. For example, a proteolytic fragment of cytrochrome c in which the heme is covalently linked to an 11-residue peptide has long been known as microperoxidase.[41] Moreover, when antibodies bind heme promiscuously (see above), the resulting antibody/heme complexes are active as peroxidases. Numerous designed heme proteins also display peroxidase activity.[42]

Hydrolase activity. Although the peroxidase activity of our *de novo* proteins requires the metallo-porphyrin cofactor, the hydrolase activities (esterase and lipase) occur without bound heme. Our finding that many of our proteins show hydrolase activity 100–1000-fold above background leads one to question whether this level of activity is easily achieved by other non-natural polypeptides. Indeed, several novel esterases have been rationally designed[43,44] and/or isolated from combinatorial libraries.[45] These findings suggest that the presence of catalytic side chains, such as histidine, may suffice to yield rudimentary hydrolytic activity. This hypothesis is supported by findings that peptide dendrimers rich in histidine, and even non-peptide polymers containing imidazole moities, can function as esterases.[46,47]

The frequent occurrence of heme binding, peroxidase activity, and hydrolase activity in other systems based on the polypeptide backbone and natural amino acid side chains is consistent with our results with binary patterned *de novo* proteins, and suggests that rudimentary activity is relatively common in the feedstock of evolution.

Specificity versus promiscuity in evolved and unevolved proteins

In contrast with most natural proteins, and even with the novel proteins reported in the literature (and summarized in the previous section), the binary code proteins exhibit activity across a variety of functions. For example, our results for esterase and lipase activities show that many proteins in our unevolved collection are active in both types of hydrolytic reaction. Thus, these *de novo* proteins are promiscuous hydrolases. This contrasts with natural hydrolytic enzymes, which usually are specific for particular types of substrates. Two examples of natural hydrolases, an esterase (PLE) and a lipase (phospholipase A2), are illustrative: PLE catalyzes the hydrolysis of the ester substrate, *p*-nitrophenyl acetate, with a rate enhancement of 10^6 above background, but acts on the related lipase substrate, *p*-nitrophenyl palmitate, with a rate enhancement that is only 10^4 above background.[48] Thus, PLE catalyzes lipid hydrolysis—a reaction for which it presumably has not evolved—with a level of activity that is only 10-fold higher than our *de novo* proteins which are indeed unevolved, and act promiscuously on the lipase substrate (Table IC).

The second example of substrate-specific activity in natural hydrolases is the helical bundle, phospholipase A2. This lipase catalyzes the hydrolysis of phosphotidyl choline, but shows no activity towards the *p*-nitrophenyl palmitate substrate used in our studies.[48] Thus, the promiscuous binary patterned *de novo* proteins are better catalysts for this lipase substrate than is the highly specific and non-promiscuous natural lipase.

Some of the *de novo* proteins that show activity as esterases and lipases also bind heme and catalyze the peroxidase reaction [Fig. 5(B)]. These proteins are highly promiscuous, consistent with Jensen's hypothesis that primitive enzymes had broad specificities, enabling them to act on a range of substrates, thereby enhancing the catalytic versatility of ancestral cells functioning with limited number of enzymes.[1] As described by Matsumura and coworkers, such proteins can be viewed as "generalists" rather than the "specialists" that typically arise after eons of natural selection.[49]

Although promiscuity may have been advantageous in the early stages of evolution,[50] the eventual development of microbial cells adapted to specific niches, and of differentiated cells with particular functions in higher organisms required that proteins acquire specificity for particular substrates. Indeed, in modern organisms, promiscuous enzymes, such as hydrolases that cleave any variation of an ester bond, would be disadvantageous or even toxic.

Our results suggest that a simple folded structure may suffice to provide low levels of promiscuous activities in unevolved systems. Ultimately, however, evolution selects for sequences and structures that have high activity and exquisite specificity. As selection progresses, promiscuity is diminished and enzymes become specific for one function while loosing their ability to catalyze other types of reactions.[51]

The results reported here indicate that for sequences capable of folding into protein-like structures, achieving some level of binding and/or catalytic activity is not difficult,[52] and can occur frequently in unevolved collections. Thus, the challenge—both for the early stages of biological evolution and for modern efforts in protein design—is not simply to produce activity, but to rein in the promiscuity of unevolved proteins, and ultimately produce biocatalysts that are highly active towards particular substrates.[53]

The availability of our superfamily of unevolved proteins with low levels of promiscuous activity will enable further exploration into several aspects of molecular evolution. The collection can be combined with high-throughput screens such as phage display,[54] mRNA display,[55] and *in vitro* compartmentalization,[56] or with genetic selections *in vivo*[57] to isolate *de novo* proteins with high levels of unique (non-promiscuous) activity. Ultimately, the isolation of biologically active proteins

from a designed library of *de novo* sequences would represent a significant advance in synthetic biology.

Materials and Methods

Screening in 96-well plates

The library of genes encoding the *de novo* proteins was constructed as described previously.[8] The library was cloned into an IPTG-inducible pET3a vector with ampicillin resistance. BL21(DE3) *E. coli* cells were transformed by electroporation, and colonies were picked and placed in 96-well plates (2 mL) containing auto-induction media (900 μL), and grown overnight on a plate shaker at 37°C. Auto-induction media was prepared by combining glycerol (0.4% v/v), glucose (0.05% w/v), and α-lactose (0.2% w/v) in LB/amp.[58] Following protein expression, cells were harvested by centrifugation and disrupted using BugBuster lysis solution (Novagen). Following centrifugation, clarified lysates were used for subsequent assays. All assays were performed in 96-well plates using the Varioskan Flash Spectral Scanning Multimode Reader and SkanIt software (Thermo Fisher Scientific, Waltham, MA).

Heme binding assay. Samples were prepared by mixing protein supernatant (50 μL) with hemin chloride (10 μ*M*, Sigma) in activity buffer (50 m*M* Tris-HCl, pH 8) to a final volume of 200 μL. (Heme stock solutions were prepared by dissolving hemin chloride in DMSO, followed by dilution into activity buffer) Heme binding was detected by the presence of a Soret peak at 412 nm.

Peroxidase assay. Samples were prepared by mixing protein supernatant (50 μL), hemin chloride (10 μ*M*, Sigma), ABTS (1 mg mL^{-1}, Sigma), H_2O_2 (0.006%) in activity buffer (50 m*M* Tris-HCl, pH 8) to a final volume of 200 μL. Product formation was detected at 650 nm 30 min after the addition of H_2O_2.

Esterase assay. Samples were prepared by mixing protein supernatant (50 μL) and *p*-nitrophenyl acetate (0.5 m*M*, Sigma) in activity buffer (50 m*M* Tris-HCl, pH 8) to a final volume of 200 μL. Product formation was detected at 405 nm 30 min after the addition of *p*-nitrophenyl acetate.

Lipase assay. Samples were prepared by mixing protein supernatant (50 μL) and *p*-nitrophenyl palmitate (0.5 m*M*, Sigma) in activity buffer (50 m*M* Tris-HCl+0.5% Triton X-100, pH 8) to a final volume of 200 μL. Product formation was detected at 405 nm 2 h after the addition of *p*-nitrophenyl palmitate.

Protein purification

Cells were streaked on Petri dishes and grown overnight. A single colony was picked, cells were grown in liquid culture, and DNA was purified. The DNA was sent for sequencing to ensure that the gene insert encoded the binary pattern. To ensure clonal purity, purified plasmid DNA was retransformed into fresh *E. coli* cells (BL21-DE3) prior to protein purification. These newly transformed cells were struck out, a single colony was picked, cells were grown in liquid culture, and protein over-expression was induced with IPTG (200 μg mL^{-1}). (Note: An aliquot was also sent for re-sequencing to confirm that the appropriate protein sequence was purified).

The protein was extracted from cells using the freeze-thaw method, and resuspended in $MgCl_2$ (100 m*M*) to extract the protein from the lysed cells.[14] Cellular contaminants were removed by acid precipitation (50 m*M* sodium acetate buffer, pH 4). The resulting supernatant was loaded onto a Poros 20 (Perseptive Biosysems) cation exchange column for purification. Protein was concentrated and exchanged into activity buffer using Centricon Plus-20 filters (Millipore).

Kinetic profiles and rate constants

Purified proteins were used for assays of peroxidase, esterase, and lipase activity. All assays were performed in 96-well plates in a final volume of 100 mL. Kinetic data was recorded using the Varioskan Flash Spectral Scanning Multimode Reader and SkanIt software (Thermo Fisher Scientific, Waltham, MA).

Peroxidase rate constants. Samples were prepared by mixing protein (50 μ*M*) with hemin chloride (5 μ*M*) and ABTS (1 mg mL^{-1}) in activity buffer (50 m*M* Tris-HCl, pH 8). Since the heme concentration is much lower than protein concentration, we assume the heme-protein complex was 5 μ*M*. Myoglobin from horse heart (Sigma) was used as a positive control at a concentration of 5 μ*M*. The negative control included hemin chloride (5 μ*M*) and ABTS (1 mg mL^{-1}) in activity buffer (50 m*M* Tris-HCl, pH 8). Upon the addition of H_2O_2 (0.086 μ*M*–11 μ*M*), timepoints were recorded for 30 min. Peroxidase rate constants were analyzed using Michaelis-Menten kinetics and were determined by fitting kinetic data to the equation: $1/V = K_M/(k_{cat}[E_o][S]) + 1/(k_{cat}[E_o])$.

Esterase rate constants. Samples were prepared by diluting protein (50 μ*M*) in activity buffer (50 m*M* Tris-HCl, pH 8). Porcine liver esterase (Sigma) was used as a positive control at a concentration of 0.005 μ*M*. Upon the addition of *p*-nitrophenyl acetate (0.08 m*M*–10 m*M*), timepoints were recorded for 2 h. Esterase rate constants were analyzed using Michaelis-Menten kinetics and were determined by fitting kinetic data to the equation: $1/V = K_M/(k_{cat}[E_o][S]) + 1/(k_{cat}[E_o])$.

Lipase rate constants. Samples were prepared by diluting protein (50 μ*M*) in activity buffer (50 m*M*

Tris-HCl +0.5% Triton X-100, pH 8). Lipase from *Candida cylindracea* (Sigma) was used as a positive control at a concentration of 0.005 μ*M*. Upon the addition of *p*-nitrophenyl palmitate (0.016 m*M*–1 m*M*), timepoints were recorded for 6 h. Lipase rate constants were analyzed using Michaelis-Menten kinetics and were determined by fitting kinetic data to the equation: $1/V = K_M/(k_{cat}[E_o][S]) + 1/(k_{cat}[E_o])$. There is no significant difference between the fit with and without the assumption of $[S] \gg [E]$.

Electronic supplementary material

The supplementary material provides more detail regarding (1) the design template for binary patterned 4-helix bundle proteins and (2) protein purification.

Acknowledgments

The authors thank Ellen Duncan and Steve Sasson for their help in establishing the screening procedures. S.C.P. was supported by the National Science Foundation Graduate Fellowship. L.H.B. was supported by a postdoctoral fellowship from the Princeton University Council on Science and Technology.

References

1. Jensen RA (1976) Enzyme recruitment in evolution of new function. Ann Rev Microbiol 30:409–425.
2. Mandecki W (1990) A method for construction of long randomized open reading frames and polypeptides. Protein Eng 3:221–226.
3. Davidson AR, Sauer RT (1994) Folded proteins occur frequently in libraries of random amino acid sequences. Proc Natl Acad Sci USA 91:2146–2150.
4. Keefe AD, Szostak JW (2001) Functional proteins from a random sequence library. Nature 410:715–718.
5. Watters AL, Baker D (2004) Searching for folded proteins in vitro and in silico. Eur J Biochem 271: 1615–1622.
6. Chiarabelli C, Vrijbloed JW, De Lucrezia D, Thomas RM, Stano P, Polticelli F, Ottone T, Papa E, Luisi PL (2006) Investigation of de novo totally random biosequences. Part II: On the folding frequency in a totally random library of de novo proteins obtained by phage display. Chem Biodivers 3:840–859.
7. Go A, Kim S, Baum J, Hecht MH (2008) Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. Protein Sci 17:821–832.
8. Bradley LH, Kleiner RE, Wang AF, Hecht MH, Wood DW (2005) An intein-based genetic selection enables construction of a high-quality library of binary patterned de novo sequences. Protein Eng Des Sel 18:201–207.
9. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262:1680–1685.
10. Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004) De novo proteins from designed combinatorial libraries. Protein Sci 13:1711–1723.
11. West MW, Wang W, Patterson J, Mancias JD, Beasley JR, Hecht MH (1999) De novo amyloid proteins from designed combinatorial libraries. Proc Natl Acad Sci USA 96:11211–11216.
12. Wei Y, Liu T, Sazinsky SL, Moffet DA, Pelczer I, Hecht MH (2003) Stably folded de novo proteins from a designed combinatorial library. Protein Sci 12:92–102.
13. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. Adv Protein Chem 14:1–63.
14. Johnson BH, Hecht MH (1994) Recombinant proteins can be isolated from E. coli cells by repeated cycles of freezing and thawing. Biotechnology 12:1357–1360.
15. Wei Y, Kim S, Fela D, Baum J, Hecht MH (2003) Solution structure of a de novo protein from a designed combinatorial library. Proc Natl Acad Sci USA 100: 13270–13273.
16. Roy S, Ratnaswamy G, Boice JA, Fairman R, McLendon G, Hecht MH (1997) A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. J Am Chem Soc 119:5302–5306.
17. Roy S, Helmer KJ, Hecht MH (1997) Detecting native-like properties in combinatorial libraries of de novo proteins. Fold Des 2:89–92.
18. Roy S, Hecht MH (2000) Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. Biochemistry 39:4603–4607.
19. Rojas NR, Kamtekar S, Simons CT, McLean JE, Vogel KM, Spiro TG, Farid RS, Hecht MH (1997) De novo heme proteins from designed combinatorial libraries. Protein Sci 6:2512–2524.
20. Moffet DA, Certain LK, Smith AJ, Kessel AJ, Beckwith KA, Hecht MH (2000) Peroxidase activity in heme proteins derived from a designed combinatorial library. J Am Chem Soc 122:7612–7613.
21. Wei Y, Hecht MH (2004) Enzyme-like proteins from an unselected library of designed amino acid sequences. Protein Eng Des Sel 17:67–75.
22. Platt J (2007) Senior Thesis. NJ: Princeton University.
23. Webb EC (1992) Recommendations of the nomenclature committee of the International union of biochemistry and molecular biology. Enzyme nomenclature. San Diego: Academic Press, p 23.
24. Orengo CA, Michie AD, Jones DT, Swindells MB, Thornton JM (1997) CATH: a hierarchic classification of protein domain structures. Structure 5:1093–1108.
25. Hiner ANP, Hernandez-Ruiz J, Arnao MB, Garcia-Canovas F, Acosta M (1996) A comparative study of the purity, enzyme activity and inactivation by hydrogen peroxide of commercially available horseradish peroxidase isoenyzmes A and C. Biotechnol Bioeng 50:655–662.
26. Wan L, Twitchett MB, Eltis LD, Mauk AG, Smith M (1998) In vitro evolution of horse heart myoglobin to increase peroxidase activity. Proc Natl Acad Sci USA 95: 12825–12831.
27. Witting PK, Mauk AG, Lay PA (2002) Role of tyrosine-103 in myoglobin peroxidase activity: kinetic and steady-state studies on the reaction of wild-type and variant recombinant human myoglobins with H2O2. Biochemistry 41:11495–11503.
28. Carlsen CU, Skovgaard IM, Skibsted LH (2003) Pseudoperoxidase activity of myoglobin: kinetics and mechanism of the peroxidase cycle of myoglobin and H2O2 and 2,2-Azino-bis(3-ethylbenzthiazoline-6-sulfonate) as substrates. J Agric Food Chem 51:5815–5823.
29. Dwyer MA, Looger LL, Hellinga HW (2008) Retraction of "computational design of a biologically active enzyme." Science 319:569.
30. Reedy CJ, Gibney BR (2004) Heme protein assemblies. Chem Rev 104:617–649.
31. Wilson JR, Caruana DJ, Gilardi G (2003) Engineering redox functions in a nucleic acid binding protein. Chem Commun 356–357.

32. Webb EC (1992) Recommendations of the nomenclature committee of the International union of biochemistry and molecular biology. Enzyme nomenclature. San Diego: Academic Press, p 862.
33. Dimitrov JD, Roumenina LT, Doltchinkova VR, Mihaylova NM, Lacroix-Desmazes S, Kaveri SV, Vassilev TL (2007) Antibodies use heme as a cofactor to extend their pathogen elimination activity and to acquire new effector functions. J Biol Chem 282:26696–26706.
34. Choma CT, Lear JD, Nelson MJ, Dutton PL, Robertson DE, DeGrado WF (1994) Design of a heme-binding four-helix bundle. J Am Chem Soc 116:856–865.
35. Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL (1994) Design and synthesis of multi-haem proteins. Nature 368:425–432.
36. Arnold PA, Shelton WR, Benson DR (1997) Peptide helix induction in a self-assembling hemoprotein model. J Am Chem Soc 119:3181–3182.
37. Huffman DL, Rosenblatt MM, Suslick KS (1998) Synthetic heme-peptide complexes. J Am Chem Soc 120:6183–6184.
38. Gibney BR, Dutton PL (1999) Histidine placement in de novo-designed heme proteins. Protein Sci 8:1888–1898.
39. Rau HK, DeJonge N, Haehnel W (2000) Combinatorial synthesis of four-helix bundle hemoproteins for tuning of cofactor properties. Angew Chem Int Ed 39:250–253.
40. Koder RL, Valentine KG, Cerda J, Noy D, Wand AJ, Dutton PL (2006) Nativelike structure in designed four α-helix bundles driven by buried polar interactions. J Am Chem Soc 128:14450–14451.
41. Adams PA (1990) The peroxidasic activity of the haem octapeptide microperoxidase-8 (MP-8): the kinetic mechanism of the catalytic reduction of H2O2 by MP-8 using 2,2′-azinobis-(3-ethylbenzothiazoline-6-sulphonate)(ABTS) as reducing substrate. J Chem Soc Perkin Trans 2:1407–1414.
42. Tsuruzono M, Obataya I, Ueno A, Mihara H (2003) Design and peroxidase-like activity of 3 alpha-helix hemoproteins. Peptide Sci 2002:367–368.
43. Broo KS, Nilsson H, Nilsson J, Baltzer L (1998) Substrate recognition and saturation kinetics in de novo designed histidine-based four-helix bundle catalysts. J Am Chem Soc 120:10287–10295.
44. Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. Proc Natl Acad Sci USA 98: 14274–14279.
45. Yamauchi A, Nakashima T, Tokuriki N, Hosokawa M, Nogami H, Arioka S, Urabe I, Yomo T (2002) Evolvability of random polypeptides through functional selection within a small library. Protein Eng 15:619–626.
46. Kunitake T, Okahata Y (1976) Catalytic hydrolysis by synthetic polymers. Adv Polym Sci 20:159–221.
47. Clouet A, Darbre T, Reymond JL (2006) Combinatorial synthesis, selection, and properties of esterase peptide dendrimers. Biopolymers 84:114–123.
48. Patel SC (2008) PhD Thesis. NJ: Princeton University.
49. O'Loughlin TL, Patrick WM, Matsumura I (2006) Natural history as a predictor of protein evolvability. Protein Eng Des Sel 19:439–442.
50. O'Brien PJ, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. Chem Biol 6: R91–R105.
51. Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, Tawfik DS (2005) The "evolvability" of promiscuous protein functions. Nat Genet 37:73–76.
52. Axe DD (2004) Estimating the prevalence of protein sequences adopting functional enzyme folds. J Mol Biol 341:1295–1315.
53. Petrounia PP, Arnold FH (2000) Designed evolution of enzymatic properties. Curr Opin Biotechnol 11:325–330.
54. Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228:1315–1317.
55. Roberts RW, Szostak JW (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci USA 94:12297–12302.
56. Tawfik DS, Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. Nat Biotechnol 16: 652–656.
57. Arnold FH, Georgiou G (2003) Methods in molecular biology—directed enzyme evolution—screening and selection methods. Totowa, NJ: Humana Press, pp 3–84.
58. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41: 207–234.