Enzyme-like proteins from an unselected library of designed amino acid sequences

Yinan Wei and Michael H. Hecht

Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

To whom correspondence should be addressed.
E-mail: hecht@princeton.edu

Combinatorial libraries of de novo amino acid sequences can provide a rich source of diversity for the discovery of novel proteins with interesting and important activities. However, since arbitrary sequences rarely fold into well ordered protein-like structures, randomly generated libraries will yield functional proteins only very rarely. To enhance the likelihood of finding functional de novo proteins, we use binary patterning of polar and non-polar amino acids to design focused libraries of sequences that are predisposed to fold into ordered structures. Proteins isolated from binary patterned libraries have been shown previously to fold into well ordered and native-like three-dimensional structures. To probe the potential of such libraries to also yield proteins with enzyme-like activity, we measured the esterase activity of S-824, a de novo binary patterned protein whose $\alpha$-helical three-dimensional structure was reported recently. Protein S-824 displayed a rate enhancement ($k_{cat}/k_{uncat}$) of 8700. The observed activity is similar to, or better than, that observed for several esterases designed previously using rational design or automated computational methods. Moreover, the observed activity rivals those of the first catalytic antibodies. To assess whether the activity of S-824 is representative of other proteins in binary patterned libraries, we measured the esterase activity of six additional proteins from two libraries. These libraries were ‘naïve’ in that they were neither designed to bind substrate, nor subjected to high throughput screens for activity. All six of the additional proteins displayed esterase activity significantly above background. These findings demonstrate that novel proteins with enzyme-like properties are surprisingly common in focused libraries designed by binary patterning. Moreover, the activity of these unselected proteins provides a reference state for the levels of activity that have been obtained by selection and/or computational design.

Keywords: binary patterning/catalytic activity/combinatorial libraries/esterase activity/protein design

Introduction

Combinatorial libraries play increasingly important roles in the discovery of new molecules with desired functions. For libraries of amino acid sequences, powerful screening and selection methods have been developed to facilitate the isolation of rare ‘winners’ from vast collections of inactive candidates (Smith, 1985; Hanes and Plückthun, 1997; Chen et al., 2001; Keefe and Szostak, 2001). The likelihood that these methods will succeed in finding proteins with desired activities depends on the power of the screen (or selection), the diversity of the library, and the quality of the library.

For libraries of novel amino acid sequences, the potential for diversity is enormous: for example, a randomly generated combinatorial library of 100 residue sequences would contain $20^{100}$ possibilities. This number is so large ($20^{100} > 10^{30}$) that a collection containing one molecule of each sequence would fill a volume larger than Avogadro’s number of universes (Beasley and Hecht, 1997). Thus, the diversity of a random library is more than sufficient for any laboratory screen or selection. However, the quality of such libraries is likely to be quite low. Since most random sequences do not fold into stable protein-like structures (Mandecki, 1990; Davidson and Sauer, 1994; Davidson et al., 1995; Prijambada et al., 1996), and since well folded structures are a prerequisite for achieving enzymatic activity (Creighton, 1992; Persht, 1999), randomly generated libraries will yield functional proteins only very rarely (Keefe and Szostak, 2001).

To enhance the likelihood of finding enzyme-like proteins, combinatorial libraries must be confined into those regions of sequence space that are most likely to yield well folded structures. The enormous numerical power of the combinatorial approach must be tempered by elements of rational design.

Over the past decade we have blended combinatorial methods with rational design to produce focused libraries enriched in sequences that fold into stable protein-like structures (Kamtekar et al., 1993; West et al., 1999; Wang and Hecht, 2002; Wei et al., 2003a). Our approach is based on the premise that the binary pattern of polar (O) and non-polar (●) residues in the linear sequence of a protein can encode a substantial fraction of the information necessary to specify a particular three-dimensional structure. For example, the binary pattern O●O0●O●O0●O●O0O0, which specifies non-polar residues at every third or fourth position, has a sequence periodicity that matches the $\alpha$-helical structural periodicity of 3.6 residues per turn. Therefore, de novo sequences constrained by this pattern are predisposed to form amphipilic $\alpha$-helices. Conversely, the binary pattern O●O0●O●O favors amphipilic $\beta$-strands, which have an inherent structural periodicity of two residues per repeat. Designed sequences comprising several units of binary patterned amphipilic secondary structure ($\alpha$-helices or $\beta$-strands) are expected to fold into globular structures that bury hydrophobic side chains in the protein interior while exposing polar side chains to solvent.

Since a binary pattern stipulates whether a position is polar or non-polar, but does not specify the identities of the side chains, a designed pattern allows for enormous combinatorial diversity. Experimentally, this diversity is incorporated into libraries of de novo proteins, by expressing proteins from libraries of synthetic genes. These genes can be designed to encode libraries of binary patterned amino acid sequences because partitioning of DNA codons into polar and non-polar...
subgroups is an inherent feature of the genetic code. Thus, the
degenerate codon CNT encodes five non-polar amino acids
(Met, Leu, Ile, Val and Phe), while the degenerate codon VAN
codes six polar amino acids (Lys, His, Glu, Gln, Asp and
Asn) (N represents the DNA bases A, G, C or T, while V
represents A, G or C).

We previously reported the design and construction of
binary patterned libraries of both α-helical and β-sheet proteins
(Kamtekar et al., 1993; West et al., 1999; Wang and Hecht,
2002). Recently, we described a second generation library of
sequences designed to fold into four-helix bundles of 102
residues (Wei et al., 2003a). Biophysical characterization of
five proteins from this new library demonstrated that all five
were α-helical and stable; and four of the five proteins formed
structures that were well ordered and/or native-like (Wei et al.,
2003a). The three-dimensional structure of one of these
proteins was determined recently by NMR spectroscopy (Wei
et al., 2003b). As specified by the design, the experimentally
determined structure is a four-helix bundle with non-polar side
chains buried in the protein interior and polar side chains
exposed to solvent. These biophysical and structural studies
were performed on proteins chosen arbitrarily from a ‘naïve’
library that had not been subjected to directed evolution or high
throughput screens. Hence, these findings demonstrate that in
appropriately designed binary patterned libraries, well folded
structures are not rare and may indeed occur frequently.

The finding that binary patterned libraries are rich in well
folded structures suggests that such libraries are likely to be far
more productive than randomly generated libraries for
the discovery of novel proteins with enzyme-like activities.
Here we demonstrate that the de novo protein S-824, which has a
well ordered four-helix bundle structure (Wei et al., 2003b),
has significant esterase activity. The level of activity is similar
to, or better than, those of several artificial proteins that were
deigned explicitly for esterase activity through the use of
rational design and computational methods (Broo et al., 1997,
1998; Bolon and Mayo, 2001; Andersson et al., 2002).
Moreover, the level of activity for protein S-824 rivals
that observed for some of the first catalytic antibodies, which
were isolated by immunological selection (Pollack et al., 1986;
Tramontano et al., 1986; Tawfik et al., 1990). To assess
whether protein S-824 is a rare fluke or a fairly representative
protein in our binary patterned libraries, we also studied the
esterase activity of six other proteins from two naive
(unselected) libraries. All six had activities significantly
above background, thereby suggesting that appropriately
designed binary patterned libraries can provide a rich source of
de novo proteins with enzyme-like activities.

Materials and methods

Proteins and substrates

The amino acid sequences of the binary patterned proteins are
shown in Figure 1. Proteins were expressed in Escherichia coli
and purified by methods described previously (Johnson and
Hecht, 1994; Roy and Hecht, 2000; Wei et al., 2003a). Stock
solutions of protein (1–2 mM) were made in 50 mM phosphate
buffer (pH 7.0). Because p-nitrophenyl esters are poorly
soluble in water, stock solutions of substrate were dissolved in
acetonitrile, and then diluted into the aqueous reactions. The
maximal concentration of acetonitrile in the final reaction
mixture was always <0.5%.

Reaction rates

The desired concentrations of substrate and protein were
brought to a final volume of 600 µl by adding buffer. The
buffers were 50 mM acetate buffer for pH 4–5.5, 50 mM
phosphate for pH 5.8–7.8, and 50 mM Tris–HCl for pH above
8.0. Protein concentration was typically in the range of 20–150
µM, and substrate concentration varied from 120 µM to 2 mM.
The hydrolysis reaction was followed using a HP 8452A diode
array spectrophotometer to monitor absorbance of the product,
p-nitrophenol, at 320 nm (at pH <6) or at 400 nm (at pH >6).
Absorbance was converted into product concentration by the
Beer–Lambert law. Because the extinction coefficient of the
product varies with pH, it was measured in control experiments
at each buffer condition.

Data analysis

We wished to compare our results with those reported
previously in other systems. Since some of the previous
studies analyzed their enzyme-like proteins as simple second-
order reactions, while others used the Michaelis–Menten
model, we calculated rate constants using both models.
(Under ideal conditions, where the concentration of substrate
is well below the value of $K_M$, $k_{cat}/K_M$ would be the same as $k_2$.
However, because the concentration of substrate in some of our
experiments was close to $K_M$, we calculated rate constants
using both models.)

Determination of the second-order rate constant. The observed
rate of the reaction (v) results from the sum of three intrinsic
rates: the protein catalyzed reaction, buffer effects and the background uncatalyzed reaction:

\[
v = k_{\text{obs}}[S] \\
v = k_2[E][S] + k_{\text{bf}}[bf][S] + k_{\text{uncat}}[S] \\
k_{\text{obs}} = k_2[E] + k_{\text{bf}}[bf] + k_{\text{uncat}}
\]

where \( k_2 \) is the apparent second-order rate constant for catalysis by the enzyme-like protein. The values of \( k_{\text{bf}} \) and \( k_{\text{uncat}} \) are readily determined by running the reaction in the absence of protein. Under such conditions:

\[
k_{\text{obs}} = k_{\text{bf}}[bf] + k_{\text{uncat}}
\]

and \( k_{\text{bf}} \) and \( k_{\text{uncat}} \) are the slope and y-intercept of a plot of \( k_{\text{obs}} \) versus [bf]. Once the values of \( k_{\text{bf}} \) and \( k_{\text{uncat}} \) are known, then for a given concentration of buffer, the value of \((k_{\text{bf}}[bf] + k_{\text{uncat}})\) is a constant. Therefore, in the presence of protein:

\[
k_{\text{obs}} = k_2[E] + \text{constant}
\]

Measurements were made at various concentrations of protein, and the value of \( k_2 \) was determined as the slope of the line in a plot of \( k_{\text{obs}} \) versus [E].

\[\text{Michaelis–Menten treatment.}\] The Michaelis–Menten model is typically used under conditions where the initial concentration of substrate is much larger than the initial concentration of enzyme (*S_0* >> *E_0*). Under such conditions the concentration of the enzyme–substrate complex [ES] is small and *S_0* can be used to approximate [S]. However, for de novo proteins with only modest activity, high concentrations of protein must be used, so *E_0* cannot be small. Also, as a result of the substrate’s limited solubility, *S_0* cannot be large. Therefore, reactions cannot be run with *S_0* >> *E_0* so the Michaelis–Menten treatment must be used without this simplifying assumption. As described by Broo et al. (1998), this leads to the following relationship:

\[
v = k_{\text{cat}}[ES] \\
v = k_{\text{cat}} \frac{(E_0 + S_0 + K_M) - \sqrt{(E_0 + S_0 + K_M)^2 - 4E_0S_0}}{2}
\]

The rate of the reaction (v) is measured, and \( E_0 \) and \( S_0 \) are known. \( K_M \) and \( k_{\text{cat}} \) were determined by plotting v as a function of \( S_0 \) at a constant value of \( E_0 \).

**Results**

**Proteins from designed binary patterned libraries**

The first binary patterned library was designed to encode 74 residue four-helix bundles (Kamtekar et al., 1993). Proteins isolated from this initial library were indeed α-helical. While a few proteins from the original library displayed biophysical properties consistent with native-like structures, most formed fluctuating structures resembling molten globules (Roy et al., 1997; Rosenbaum et al., 1999; Roy and Hecht, 2000). A second generation library was designed recently to test whether elongation of the α-helices would produce a library containing an abundance of well ordered and native-like structures. Helices in proteins from the second generation library were designed to be 50% longer, and the overall length of the sequences was increased from 74 to 102 residues. The second generation sequences were constructed by using binary patterning to elongate the four α-helices of a molten globule-like protein (n86) from the first generation library. Five proteins from this second generation library were subjected to biophysical characterizations, and four of them (all except S-23) were shown to be well ordered and/or native-like (Wei et al., 2003a). The high-resolution structure of one of these proteins, S-824, was determined recently by NMR spectroscopy, and shown to be a well ordered four-helix bundle (Wei et al., 2003b). This protein was chosen for the studies described in the current paper.

To assess whether the enzyme-like properties of protein S-824 are unusual or typical of other sequences in our libraries, we chose several other proteins for comparison. These included two proteins from the first generation library, and four from the second generation library. The two proteins from the first generation library are n86, which was shown previously to resemble a molten globule, and M60, which was shown previously to be more native-like (Roy et al., 1997; Rosenbaum et al., 1999; Roy and Hecht, 2000). From the second generation library, we chose all five of the proteins (including S-824) that had been characterized previously (Wei et al., 2003a). These are S-23, S-213, S-285, S-824 and S-836. The sequences of all proteins used in this study are shown in Figure 1.

Neither the first nor the second generation libraries have been subjected to genetic selections or high throughput screens. Thus, all proteins used in this study are from ‘naïve’ libraries.

**Hydrolysis of p-nitrophenyl esters**

The reaction monitored in these studies is the hydrolysis of p-nitrophenyl esters, shown below:

\[
\text{Scheme 1}
\]

This reaction has four distinct advantages: (i) the activation barrier is relatively modest, thereby increasing the likelihood of finding active proteins in a naïve library; (ii) the mechanism is simple and well characterized; (iii) the reaction generates a colored product (p-nitrophenol), thereby facilitating kinetic studies; and (iv) hydrolysis of p-nitrophenyl esters has been used as a model reaction in several earlier studies using very different approaches to search for novel protein-based catalysts (Pollack et al., 1986; Tramontano et al., 1986; Tawfik et al., 1990; Broo et al., 1997, 1998; Bolon and Mayo, 2001; Andersson et al., 2002). This previous history enables direct comparison between our results and those obtained using approaches ranging from computer-aided rational design to immunological selection.

**Catalytic activity**

For the hydrolysis of p-nitrophenyl acetate, the catalytic activity of protein S-824 is substantially above background (Figure 2). Activity was compared with two controls: 4-methylimidazole is a control for the histidine side chain in isolation without the surrounding protein structure, and sample 48 is a control for residual esterase activity from *E. coli* that might co-purify with our proteins. [Sequence 48 was part of the original binary code library (Kamtekar et al., 1993), but contains a stop codon early in the gene. A ‘mock’ protein purification from cells harboring this sequence serves as a
control for any background activity from endogenous *E. coli* proteins. As shown in Figure 2, the activity of protein S-824 is substantially above these controls.

Figure 3 shows the pH dependence of the second-order rate constant, \(k_2\), for the hydrolysis of p-nitrophenyl acetate catalyzed by protein S-824. Experiments were typically carried out at a substrate concentration of 120 µM.

**Rate enhancement by protein S-824**

For spontaneous (uncatalyzed) ester hydrolysis, the rate of the reaction decreases dramatically with decreased pH. In contrast, for the catalyzed reaction, the effect of pH is relatively modest (<2-fold between pH 7 and pH 8.5; see Figure 3). Hence, the rate enhancement for the catalyzed reaction relative to the uncatalyzed reaction is most striking at lower pH. At pH 7.3, the rate constant for the uncatalyzed hydrolysis of p-nitrophenyl acetate (\(k_{\text{uncat}}\)) is \(3.8 \times 10^{-5}\) min\(^{-1}\). At this pH the rate enhancement (\(k_{\text{cat}}/k_{\text{uncat}}\)) by protein S-824 is 8700.

Since hydrolysis of p-nitrophenyl esters by protein S-824 is presumed to involve a histidine nucleophile (see Discussion), it is important to establish that the rate catalyzed by protein S-824 is above that catalyzed by free imidazole. As shown in Figure 2, at pH 7 protein S-824 hydrolyzes p-nitrophenyl acetate ~100-fold faster than does 4-methylimidazole.

**Activity following multiple turnovers**

A defining characteristic of natural enzymes is their ability to catalyze multiple turnovers. To assess whether our *de novo* protein possesses this enzyme-like property, protein S-824 was incubated overnight with a 10-fold excess of substrate. The protein was then separated from small molecules by filtration and re-assayed. As shown in Figure 4, the multiple turnovers associated with overnight incubation did not reduce the activity of protein S-824.

**Covalent modification of the protein**

Hydrolysis of an ester requires transfer of an acyl group. For enzymes that use a nucleophilic side chain, the mechanism typically occurs in two steps. First, the acyl group is transferred from the substrate to the nucleophilic side chain. Subsequently, the nucleophile is regenerated by transfer of the acyl group to another side chain or to water. To assess whether an acyl group is covalently attached to our *de novo* protein, S-824 was incubated overnight with a 20-fold excess of the substrate, p-nitrophenyl acetate, and the resulting protein was analyzed by electrospray mass spectrometry. As shown in Figure 5, prolonged incubation with excess substrate caused the attach-
ment of four to seven acetyl groups. Despite multiple acetylations (Figure 5), the protein remains active (Figure 4). This indicates that the active site nucleophile can be regenerated by transfer of the acetyl group to another site on the protein. Moreover, since the protein was incubated with a 20-fold excess of substrate and only four to seven acetyl groups remain attached to the protein, we conclude that acetyl groups are ultimately transferred to water.

**Substrate specificity**

The ability to discriminate between closely related molecules is an essential feature of biological processes. Consequently, many natural enzymes have evolved to possess exquisite substrate specificity. In contrast, our binary patterned proteins, which have neither been designed nor selected to bind particular substrates, would not be expected to possess such specificity. To assess whether protein S-824 can distinguish between closely related substrates, we compared the catalytic rates for the hydrolysis of the \( p \)-nitrophenyl esters of acetate, propionate and butyrate. As shown in Table I, the values of \( k_{\text{cat}}/k_{\text{uncat}} \) indicate that protein S-824 displays a very slight preference for the propionate ester relative to acetate or butyrate.

**Catalytic activity of proteins from unselected libraries**

Protein S-824 was chosen for the studies described in the previous sections because it is the first binary patterned protein for which a high-resolution structure is known (Wei et al., 2003b). However, since protein S-824 was not isolated by genetic selections or high throughput screens for activity, our choice to focus on S-824 was somewhat arbitrary. To ascertain whether protein S-824 was simply a ‘lucky’ choice, or whether its level of activity is representative of activity that might be found frequently in unselected binary patterned libraries, we purified six additional proteins from two binary patterned libraries and measured their esterase activities. The sequences of these proteins are shown in Figure 1.

![Mass spectrum of protein S-824 following overnight incubation with a 20-fold excess of \( p \)-nitrophenyl acetate. (Protein concentration was ~100 \( \mu \)M and the \( p \)-nitrophenyl acetate concentration was ~2 mM.) Following incubation, protein was filtered away from small molecules and transferred to distilled water prior to analysis. Identification of peaks in the mass spectrum is based on an increase of 42 AMU for each replacement of a hydrogen (1 AMU) by an acetyl group (43 AMU). The mass of the unmodified protein is 11 928 AMU.](image)

**Table I. Substrate specificity**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>( k_{\text{cat}} ) (min(^{-1}))</th>
<th>( K_M ) (mM)</th>
<th>( k_{\text{cat}}/k_{\text{uncat}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetate</td>
<td>1.5</td>
<td>4</td>
<td>940</td>
</tr>
<tr>
<td>Propionate</td>
<td>1.9</td>
<td>7</td>
<td>2000</td>
</tr>
<tr>
<td>Butyrate</td>
<td>0.6</td>
<td>4</td>
<td>1000</td>
</tr>
</tbody>
</table>

Comparison of catalytic activity of protein S-824 towards different \( p \)-nitrophenyl esters. Experiments were performed at pH 8.5.

**Table II. Kinetic data for the designed proteins measured at pH 8.5**

<table>
<thead>
<tr>
<th>Protein</th>
<th>( k_2 ) (min(^{-1} ) M(^{-1} ))</th>
<th>( k_{\text{cat}} ) (min(^{-1} ))</th>
<th>( K_M ) (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First generation proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n86</td>
<td>35</td>
<td>0.2</td>
<td>3</td>
</tr>
<tr>
<td>M60</td>
<td>300</td>
<td>1.0</td>
<td>3</td>
</tr>
<tr>
<td>Second generation proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-824</td>
<td>300</td>
<td>1.5</td>
<td>4</td>
</tr>
<tr>
<td>S-23</td>
<td>170</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>S-836</td>
<td>260</td>
<td>1.7</td>
<td>6</td>
</tr>
<tr>
<td>S-213</td>
<td>300</td>
<td>1.6</td>
<td>5</td>
</tr>
<tr>
<td>S-285</td>
<td>370</td>
<td>1.3</td>
<td>5</td>
</tr>
</tbody>
</table>

**Substrate specificity**

The ability to discriminate between closely related molecules is an essential feature of biological processes. Consequently, many natural enzymes have evolved to possess exquisite substrate specificity. In contrast, our binary patterned proteins, which have neither been designed nor selected to bind particular substrates, would not be expected to possess such specificity. To assess whether protein S-824 can distinguish between closely related substrates, we compared the catalytic rates for the hydrolysis of the \( p \)-nitrophenyl esters of acetate, propionate and butyrate. As shown in Table I, the values of \( k_{\text{cat}}/k_{\text{uncat}} \) indicate that protein S-824 displays a very slight preference for the propionate ester relative to acetate or butyrate.

**Catalytic activity of proteins from unselected libraries**

Protein S-824 was chosen for the studies described in the previous sections because it is the first binary patterned protein for which a high-resolution structure is known (Wei et al., 2003b). However, since protein S-824 was not isolated by genetic selections or high throughput screens for activity, our choice to focus on S-824 was somewhat arbitrary. To ascertain whether protein S-824 was simply a ‘lucky’ choice, or whether its level of activity is representative of activity that might be found frequently in unselected binary patterned libraries, we purified six additional proteins from two binary patterned libraries and measured their esterase activities. The sequences of these proteins are shown in Figure 1.

![Mass spectrum of protein S-824 following overnight incubation with a 20-fold excess of \( p \)-nitrophenyl acetate. (Protein concentration was ~100 \( \mu \)M and the \( p \)-nitrophenyl acetate concentration was ~2 mM.) Following incubation, protein was filtered away from small molecules and transferred to distilled water prior to analysis. Identification of peaks in the mass spectrum is based on an increase of 42 AMU for each replacement of a hydrogen (1 AMU) by an acetyl group (43 AMU). The mass of the unmodified protein is 11 928 AMU.](image)

**Table I. Substrate specificity**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>( k_{\text{cat}} ) (min(^{-1}))</th>
<th>( K_M ) (mM)</th>
<th>( k_{\text{cat}}/k_{\text{uncat}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetate</td>
<td>1.5</td>
<td>4</td>
<td>940</td>
</tr>
<tr>
<td>Propionate</td>
<td>1.9</td>
<td>7</td>
<td>2000</td>
</tr>
<tr>
<td>Butyrate</td>
<td>0.6</td>
<td>4</td>
<td>1000</td>
</tr>
</tbody>
</table>

Comparison of catalytic activity of protein S-824 towards different \( p \)-nitrophenyl esters. Experiments were performed at pH 8.5.

**Table II. Kinetic data for the designed proteins measured at pH 8.5**

<table>
<thead>
<tr>
<th>Protein</th>
<th>( k_2 ) (min(^{-1} ) M(^{-1} ))</th>
<th>( k_{\text{cat}} ) (min(^{-1} ))</th>
<th>( K_M ) (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First generation proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n86</td>
<td>35</td>
<td>0.2</td>
<td>3</td>
</tr>
<tr>
<td>M60</td>
<td>300</td>
<td>1.0</td>
<td>3</td>
</tr>
<tr>
<td>Second generation proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-824</td>
<td>300</td>
<td>1.5</td>
<td>4</td>
</tr>
<tr>
<td>S-23</td>
<td>170</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>S-836</td>
<td>260</td>
<td>1.7</td>
<td>6</td>
</tr>
<tr>
<td>S-213</td>
<td>300</td>
<td>1.6</td>
<td>5</td>
</tr>
<tr>
<td>S-285</td>
<td>370</td>
<td>1.3</td>
<td>5</td>
</tr>
</tbody>
</table>

**Substrate specificity**

The ability to discriminate between closely related molecules is an essential feature of biological processes. Consequently, many natural enzymes have evolved to possess exquisite substrate specificity. In contrast, our binary patterned proteins, which have neither been designed nor selected to bind particular substrates, would not be expected to possess such specificity. To assess whether protein S-824 can distinguish between closely related substrates, we compared the catalytic rates for the hydrolysis of the \( p \)-nitrophenyl esters of acetate, propionate and butyrate. As shown in Table I, the values of \( k_{\text{cat}}/k_{\text{uncat}} \) indicate that protein S-824 displays a very slight preference for the propionate ester relative to acetate or butyrate.

**Catalytic activity of proteins from unselected libraries**

Protein S-824 was chosen for the studies described in the previous sections because it is the first binary patterned protein for which a high-resolution structure is known (Wei et al., 2003b). However, since protein S-824 was not isolated by genetic selections or high throughput screens for activity, our choice to focus on S-824 was somewhat arbitrary. To ascertain whether protein S-824 was simply a ‘lucky’ choice, or whether its level of activity is representative of activity that might be found frequently in unselected binary patterned libraries, we purified six additional proteins from two binary patterned libraries and measured their esterase activities. The sequences of these proteins are shown in Figure 1.

Surprisingly, all six proteins displayed esterase activity well above background levels. Table II compares the kinetic constants for protein S-824 with the other six binary patterned proteins. Data are presented at pH 8.5, the pH of maximal activity for protein S-824 (see Figure 3).

To facilitate comparison of our results with other systems reported in the literature, we analyzed the catalytic activities of our binary patterned proteins using two different models.

(i) If the reaction between enzyme (E) and substrate (S) is modeled as a simple second-order reaction:

\[
E + S \xrightarrow{k_2} E + P
\]
then $k_2$ is the apparent second-order rate constant, and the rate of the reaction is $k_2[E][S]$. Values of $k_2$ for the seven proteins are listed in Table II.

(ii) Alternatively, the reaction kinetics can be analyzed according to the Michaelis–Menten model:

$$E + S \quad \overset{K_M}{\rightleftharpoons} \quad ES \quad \xrightarrow{k_{cat}} \quad E + P$$

where $ES$ is the transient enzyme–substrate complex. According to this model, the rate of the overall reaction is $k_{cat}[ES]$. $K_M$ and $k_{cat}$ for the de novo proteins are also listed in Table II.

Discussion

Esterase activity of protein S-824

Because the structure of S-824 is known, this protein was chosen for initial studies of esterase activity in an unselected library. As shown in Figure 2, protein S-824 catalyzes ester hydrolysis at rates significantly above background. For the hydrolysis of p-nitrophenyl acetate at pH 7.3 the rate enhancement ($k_{cat}/k_{uncat}$) by S-824 is 8700.

The maximal activity of S-824 occurs at pH 8.5. The bell-shaped curve of the pH dependence (Figure 3) is consistent with a deprotonated residue acting as a nucleophile and a protonated residue stabilizing a negatively charged intermediate. Because activity drops as pH is lowered from 8.5 to 5.5, and also by analogy with earlier work on novel protein-based esterases (Pollack et al., 1986; Tramontano et al., 1986; Tawfik et al., 1990; Broo et al., 1997, 1998; Bolon and Mayo, 2001; Andersson et al., 2002), we presume the active site nucleophile is histidine.

Natural enzymes, as true catalysts, are not degraded through the course of a reaction, and therefore can catalyze multiple turnovers. As shown in Figure 4, protein S-824 also remains active after many rounds of catalysis. Yet, as shown in Figure 5, the protein is modified by attachment of acyl groups. Acylation was also observed for catalytic antibodies and for rationally designed esterases (Stewart et al., 1994; Broo et al., 1997; Bolon and Mayo, 2001). Andersson et al. (2002) did a systematic study of the acylation of a model protein following reaction with p-nitrophenyl esters. They note that ‘in the first, and rate-limiting step of the reaction, the unprotonated form of the histidine attacks the ester to form an acyl intermediate. The acyl group is then transferred to the flanking lysine in a fast intramolecular reaction and an amide is formed at the lysine side chain... If there is more than one lysine in close proximity to the His residue, the site of modification is determined by intramolecular competition between the flanking lysines’ (Andersson et al., 2002). We presume a similar mechanism is responsible for acylation in our system. The sequence of protein S-824 contains 12 histidines and eight lysines. While we do not yet know which of these are responsible for activity, it is clear from the NMR structure (Figure 6) that there are several sites where histidines and lysines occur in close proximity.

As shown in Figures 4 and 5, when protein S-824 was incubated overnight with a 20-fold excess of substrate, four to seven acyl groups became attached, yet the protein remained active. How does the protein retain activity despite acylation? Apparently, the lysines that are modified are not essential for activity. Our finding that after incubation with a 20-fold excess of substrate only four to seven lysines are modified indicates that once these reactive lysines have become acylated, further transfer of acyl groups is to water, as must be the case for a true hydrolysis reaction.

Esterase activity in unselected libraries

As shown in Table II, all seven of the proteins tested thus far catalyze ester hydrolysis at rates significantly above background. Because these proteins were chosen arbitrarily—without screens or selections—these findings suggest that the requirements for this level of activity are relatively promiscuous.

The goal of these initial studies was simply to assess whether esterase activity could be found in focused, but unselected libraries. Although significant activity was observed for all the proteins, these initial experiments are not sufficiently detailed to ascribe a mechanism to the observed activity. Nonetheless,
by analogy with earlier work on novel esterases (Pollack et al., 1986; Tramontano et al., 1986; Tawfik et al., 1990; Stewart et al., 1994; Broo et al., 1997, 1998; Bolon and Mayo, 2001; Andersson et al., 2002), it seems reasonable that at least for some of the proteins the active site nucleophile is histidine. (As mentioned above, the pH profile for the activity of protein S-824 is also consistent with an active site histidine.) Of the seven proteins assayed for activity thus far (Table II), the two proteins from the first generation library (n86 and M60) have five histidines each, while the second generation proteins contain 7–12 histidines each.

In comparing the activities of the binary patterned proteins to one another (Table II), we note that activity correlates with structural integrity: among the first generation proteins, M60 is more active than n86 (Table II). Earlier studies showed that structural integrity: among the first generation proteins, M60 is the most native-like proteins in the first generation library, whereas n86 is a molten globule (Roy et al., 1997; Rosenbaum et al., 1999; Roy and Hecht, 2000). Likewise, for the second generation library, S-23 has the lowest activity (Table II), and is also the only protein from this library that does not possess native-like features (Wei et al., 1997; Rosenbaum et al., 1999; Roy and Hecht, 2000). The observed correlation between enzyme-like activity and structural integrity suggests that for the binary code proteins, as for natural enzymes, formation of an active site is favored by formation of a well folded structure (Creighton, 1992; Matthews et al., 1994; Fersht, 1999).

**Comparison with other studies**

Hydrolysis of p-nitrophenyl esters has been studied by several laboratories as a model reaction for the isolation of novel enzyme-like catalysts. As explained by Menger and Ladika (1987), catalysis of p-nitrophenyl esters is rather easy to achieve. Therefore, the choice of this reaction as a model system has both advantages and disadvantages. The advantage is that the likelihood of finding some level of activity is reasonably good. The associated disadvantage is that although observing activity may be easy, assessing the significance of this activity can be difficult.

Therefore, to evaluate the significance of our results, we compared the activities of our unselected binary patterned proteins to those obtained by four other approaches: (i) computational design of an esterase active site into the natural protein, thioredoxin (Bolon and Mayo, 2001); (ii) iterative rational design of de novo peptides with esterase activity (Broo et al., 1997, 1998; Andersson et al., 2002); (iii) selection of catalytic antibodies from immunological repertoires raised against transition state analogs (Pollack et al., 1986; Tramontano et al., 1986; Tawfik et al., 1990); and (iv) selection of p-nitrophenyl esterases by phage display (Yamauchi et al., 2002).

Bolon and Mayo (2001) used computational methods to engineer an active site into E.coli thioredoxin. The target activity for their studies was hydrolysis of p-nitrophenyl acetate by a histidine nucleophile. Novel active sites were designed by modeling the transition state of the reaction into numerous loci on the protein, and then computationally searching for mutations and rotamers compatible with formation of an active site. Several active sites, containing two to four mutations, were designed computationally, and the two most promising sequences were synthesized. Both displayed esterase activity above background. At pH 7, the more active of the mutant thioredoxins (called PZD2) displayed a $K_M$ of 0.17 mM and a $k_{cat}$ of 0.028 min$^{-1}$ (Bolon and Mayo, 2001). At a

<table>
<thead>
<tr>
<th>Protein R</th>
<th>pH</th>
<th>$k_{cat}$ (min$^{-1}$M$^{-1}$)</th>
<th>$K_M$ (mM)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>S8-24 CH$_3$</td>
<td>7.3</td>
<td>0.17</td>
<td>3</td>
<td>This work</td>
</tr>
<tr>
<td>MNKR CH=CHCOOH</td>
<td>5.0</td>
<td>0.21</td>
<td>0.017</td>
<td>Broo et al. (1997)</td>
</tr>
<tr>
<td>MOPC67 CH$_2$CH$_2$N(CH$_3$)$_2$+</td>
<td>7.0</td>
<td>0.0006</td>
<td>0.21</td>
<td>Pollack et al. (1986)</td>
</tr>
<tr>
<td>6D4 NHCOCF$_3$PhCH$_2$</td>
<td>8.0</td>
<td>0.0006</td>
<td>0.2</td>
<td>Tramontano et al. (1986)</td>
</tr>
<tr>
<td>CNJ series</td>
<td>8.0</td>
<td>0.06 to 3.0</td>
<td>0.2</td>
<td>Tawfik et al. (1990)</td>
</tr>
<tr>
<td>YSLP6-1 CH$_3$</td>
<td>5.1</td>
<td>6.4</td>
<td>0.2</td>
<td>Broo et al. (1998)</td>
</tr>
<tr>
<td>S8-24 CH$_3$</td>
<td>8.5</td>
<td>300</td>
<td>This work</td>
<td></td>
</tr>
<tr>
<td>S213 CH$_3$</td>
<td>8.5</td>
<td>300</td>
<td>This work</td>
<td></td>
</tr>
<tr>
<td>S285 CH$_3$</td>
<td>8.5</td>
<td>370</td>
<td>This work</td>
<td></td>
</tr>
<tr>
<td>S836 CH$_3$</td>
<td>8.5</td>
<td>260</td>
<td>This work</td>
<td></td>
</tr>
<tr>
<td>KO-42 CH=CHCOOH</td>
<td>7 and above</td>
<td>45</td>
<td>KO-42</td>
<td></td>
</tr>
<tr>
<td>KO-42 CH$_3$</td>
<td>7 and above</td>
<td>45</td>
<td>KO-42</td>
<td></td>
</tr>
</tbody>
</table>

The identity of the ester is indicated by the R group. $K_M$ for S-824 at pH 7.3 is 3 x 10$^{-5}$ M. See Tramontano et al. (1986).

Uncat for S-824 at pH 7.3 is 3 x 10$^{-10}$ M. See Tramontano et al. (1986).
similar pH (7.3), protein S-824 from our binary patterned library had a $K_M$ of 3 mM and a $k_{cat}$ of 0.33 min$^{-1}$ (Table III). Thus, PZD2 binds substrate with an apparent affinity (measured by $K_M$) 18-fold better than protein S-824. This is not surprising since the site in PZD2 was computationally designed around a model of the transition state. Unexpectedly, however, the $k_{cat}$ of protein S-824 is 10-fold faster than that of PZD2 (Table III). This is surprising since PZD2 was produced by computational design of an active site into the scaffold of a well folded natural protein, whereas S-824 was chosen arbitrarily from a library of de novo sequences that had neither been designed nor selected for activity.

Baltzer and coworkers used a different approach to devise novel esterases (Broo et al., 1997, 1998; Andersson et al., 2002). They rationally designed a 42 residue $\alpha$-helical hairpin (called KO-42) capable of dimerizing into a four-helix bundle (Broo et al., 1997). At pH 5.1, the second-order rate constant, $k_2$, for hydrolysis of $p$-nitrophenol acetate was 18 M$^{-1}$ min$^{-1}$ (Broo et al., 1997). Under similar conditions, protein S-824 was >-2-fold faster, with a $k_2$ of 45 M$^{-1}$ min$^{-1}$ (Table III). At pH values above 7, the $k_2$ of KO-42 for hydrolysis of the fumarate ester plateaus at 42 M$^{-1}$ min$^{-1}$ (Broo et al., 1997). By comparison, for the hydrolysis of the acetate ester at pH 8.5, four of the binary patterned proteins display $k_2$ values between 260 and 370 M$^{-1}$ min$^{-1}$ (Table III). Thus, although the binary patterned proteins were not explicitly designed as catalysts, many proteins from this unselected library catalyze ester hydrolysis at rates faster than the rationally designed peptides.

In an effort to locate the active site and enhance the substrate selectivity of KO-42, Baltzer and coworkers replaced several of the histidines in the four-helix bundle (Broo et al., 1998). The new peptides had fewer histidines—hence fewer sites for activity—and saturation kinetics were observed. Michaelis-Menten kinetics were reported for a peptide called MNKR, which is similar to KO-42 except that three histidines per chain were mutated. For the hydrolysis of the fumarate ester of $p$-nitrophenol at pH 5.1, MNKR had a $K_M$ of 1 mM and a $k_{cat}$ of 0.01 min$^{-1}$ (Broo et al., 1998). By comparison, for the hydrolysis of the acetate ester of $p$-nitrophenol at pH 7.3, protein S-824 has a $K_M$ of 3 mM and a $k_{cat}$ of 0.33 min$^{-1}$ (Table III). Although the conditions of the experiments differ somewhat, it seems clear that although S-824 has a weaker affinity for substrate, it catalyzes hydrolysis at a rate similar to (or perhaps faster than) the rationally designed MNKR.

A third approach for isolating novel esterases involves selecting catalytic antibodies. The first catalytic antibodies were reported in 1986 (Pollack et al., 1986; Tramontano et al., 1986). Schultz and coworkers isolated MOPC167, an antibody that catalyzed hydrolysis of $p$-nitrophenyl carbonate with a $K_M$ of 0.21 mM and a $k_{cat}$ of 0.4 min$^{-1}$ (Pollack et al., 1986). Concurrently, Lerner and coworkers isolated 6D4, which catalyzed ester hydrolysis. Depending on the substrate, 6D4 had a $K_M$ of 1.9 or 0.62 $\mu$M, and a $k_{cat}$ of 1.6 or 0.5 min$^{-1}$ (Tramontano et al., 1986). In 1990, Tawfik et al. (1990) reported several antibodies that catalyzed the hydrolysis of $p$-nitrophenyl esters with $K_M$ values ranging from 0.06 to ~30 mM, $k_{cat}$ values ranging from 0.63 to 2.39 min$^{-1}$, and enhancement ratios ($k_{cat}/K_{uncat}$) ranging from 2600 to 9700. These early antibodies functioned by stabilizing the transition state corresponding to direct hydroxide attack on the scissile carbonyl (Hilvert, 2000). In contrast, our proteins appear to catalyze hydrolysis by promoting acyl transfer to histidine, followed by transfer to water. Despite these different mechanisms, it is instructive to compare the kinetic parameters reported for the early catalytic antibodies to those of the binary patterned protein S-824 (Table III). The catalytic antibodies typically bind substrate with higher affinities than protein S-824. It would be surprising if this were not the case, since the antibodies were selected by the immune system for binding to transition state analogs, whereas S-824 is from an unselected library. The catalytic antibodies also exhibit greater substrate specificity (Hilvert, 2000) than observed for protein S-824 (Table I). This is also not surprising since the antibodies were raised against transition state analogs. It is surprising, however, that the catalytic rate ($k_{cat}$) and rate enhancement ($k_{cat}/K_{uncat}$) of protein S-824 are in the same range as many of the early catalytic antibodies (Table III).

As an alternative to the catalytic antibodies generated by the mammalian immune system, one can also use phage display to select among diverse sequences for those that bind transition state analogs. Yamauchi et al. (2002) displayed random polypeptides of ~140 residues on the surface of phage, and selected for binding to a transition state analog for the $p$-nitrophenyl esterase reaction. In contrast to typical phage display experiments, Yamauchi et al. (2002) used very small libraries containing only 10 clones per generation. In each generation, the clone showing the highest affinity for the target was subjected to random mutagenesis, and advanced to the next round. After six generations, the sequence with the highest esterase activity, YSLP6-1, had a $k_{cat}/K_M$ at pH 5 of 0.064 M$^{-1}$ min$^{-1}$. (Individual $k_{cat}$ and $K_M$ values were not reported.) By comparison, after no rounds of selection, protein S-824 showed a $k_{cat}/K_M$ at pH 7.3 of 0.11 M$^{-1}$ min$^{-1}$ (Table III).

In summary, when comparing the binary patterned proteins to the artificial protein-based esterases described in the previous four examples, it is important to note that the binary patterned proteins (i) were constructed entirely de novo and not engineered onto a pre-existing natural protein scaffold, (ii) were not designed explicitly to bind substrate, and (iii) were chosen at random and not by selections or high throughput screens. Nonetheless, these proteins possess enzyme-like activity, which although lower than those of natural enzymes, rivals those of earlier artificial esterases devised using natural protein scaffolds, rational design, immunological selection or phage display.

These findings suggest that although achieving the requisite levels of activity and specificity typical of natural enzymes may have required eons of evolutionary selection, moderately active macromolecular catalysts (Hollfelder et al., 1996; Suh and Oh, 2000) may occur frequently in focused combinatorial libraries that have undergone neither evolutionary selection nor active site design.

**Acknowledgements**

We thank Don Hilvert for many helpful suggestions. Supported by NIH grant RO1-GM062869.

**References**


Enzyme-like de novo proteins


Received September 8, 2003; accepted October 20, 2003

Edited by Greg Winter