Biosynthetic phage display: a novel protein engineering tool combining chemical and genetic diversity

Mary A Dwyer, Wuyuan Lu, John J Dwyer and Anthony A Kossiakoff

Background: Molecular diversity in nature is developed through a combination of genetic and chemical elements. We have developed a method that permits selective manipulation of both these elements in one protein engineering tool. It combines the ability to introduce non-natural amino acids into a protein using native chemical ligation with exhaustive targeted mutagenesis of the protein via phage-display mutagenesis.

Results: A fully functional biosynthetic version of the protease inhibitor eglin c was constructed. The amino-terminal fragment (residues 8–40) was chemically synthesized with a non-natural amino acid at position 25. The remaining carboxy-terminal fragment was expressed as a 30-residue peptide extension of gIIIp or gVIIIp on filamentous phage in a phage-display mutagenesis format. Native chemical ligation was used to couple the two fragments and produced a protein that refolded to its active form. To facilitate the packing of the introduced non-natural amino acid, residues 52 and 54 in the carboxy-terminal fragment were fully randomized by phage-display mutagenesis. Although the majority of the observed solutions for residues 52 and 54 were hydrophobic – complementing the stereochemistry of the introduced non-natural amino acid – a significant number of residues (unexpected because of stereochemical and charge criteria) were observed in these positions.

Conclusions: Peptide synthesis and phage-display mutagenesis can be combined to produce a very powerful protein engineering tool. The physical properties of the environment surrounding the introduced non-natural residue can be selected for by evaluating all possible combinations of amino acid types at a targeted set of sites using phage-display mutagenesis.

Address: Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA.

Correspondence: Anthony A Kossiakoff
E-mail: koss@cummings.uchicago.edu

Key words: hydrophobic packing, molecular diversity, native chemical ligation, non-natural amino acids, phage display

Received: **7 December 1999**
Revisions requested: **7 January 2000**
Revisions received: **25 January 2000**
Accepted: **26 January 2000**

Published: **20 March 2000**

Chemistry & Biology 2000, **7**:263–274

1074-5521/00/$ – see front matter

Introduction

Structure and function in the natural protein world has been primarily developed through evolution, based on chemistries and structures using 20 left-handed amino acids. One of the grand challenges in biology is to relate how a pattern of 20 amino acids can be organized along a polypeptide chain to give a uniquely functional form. A very fruitful approach to investigate this principal issue has been protein engineering coupled with structural analysis, which has allowed functional elements to be systematically and thoroughly probed. With an increase of the knowledge base, the scope of some of the studies has expanded to include the creation of new functionalities and the design of new structural entities based on existing polypeptide scaffolds [1–4].

An important and powerful advance in protein engineering technology has been the development of phage-display mutagenesis [5]. Phage display is an efficient methodology used to produce and evaluate an extremely large number of proteins with a selectable physical property. Its underlying principle is that 'trying all possibilities' is a much more pragmatic and inclusive approach to developing optimized solutions than fully intellectual approaches using site-directed mutagenesis. Using this technique proteins with a number of novel and unexpected properties have been produced [6–12].

Although this method is very powerful in developing diversity using the 20 amino acids, it still cannot duplicate the versatility of nature. Nature has worked out systems to greatly expand functional characteristics using post-translational modification and the incorporation of specialized non-natural amino acids. Strategically placed, these chemical moieties form the basis for many essential regulating events, such as signaling and targeting functions [10,13–15]. It follows that protein engineering could be made even more powerful by expanding the genetic code in a similar way; that is, including the ability to incorporate non-natural amino acids with selected properties or the capability to make post-translational modifications.

Efforts to introduce non-natural amino acids into a protein sequence using *in vitro* translation methods have had

some success, but are technically daunting [16,17]. There is also considerable potential in using an intein–extein ligation strategy [18]. Native chemical ligation, however, has advanced rapidly as a technology for making large polypeptides (up to 200 amino acids) [19–23], and its success has been demonstrated through structural analysis that has shown that the ligated fragments form a stereochemically true peptide bond [24]. This technology has the potential to greatly expand the repertoire of protein function by strategic placement of highly functionalized non-natural amino acids.

Combining the versatility of synthetic proteins with the power of targeted random mutagenesis and screening using phage-display libraries offers extraordinary potential to both the understanding and the rational design of protein functionality. The question is how to do it? The obvious drawback is a technical one: phage display is genetics based and thus does not naturally partner with synthetic approaches. To overcome this apparent incompatibility, we have adopted a straightforward approach. In its simplest terms, it involves expressing a carboxy-terminal fragment of the protein of interest on the phage and coupling it to the chemically synthesized amino-terminal fragment using the native chemical ligation technique. The 'full-length' protein thus incorporates the polypeptide code to form a folded entity that contains synthetic and genetically derived elements.

The segregation of the synthetic and genetic diversities allows them to be manipulated separately; however, having them separated in the amino- and carboxy-terminal fragments, respectively, does not mean that the individual diversities are necessarily separated in terms of their influence on protein function. This situation reflects the inherent feature of protein structure: spatial arrangement of amino acid residues in a folded protein is generally independent of their position in the sequence. Thus, systems can be designed such that in the folded form a group introduced in the amino-terminal synthetic fragment can lie adjacent to groups expressed in the phage-originated peptide.

We report here the development and proof of concept of a facile methodology to couple synthesized peptide fragments to protein fragments expressed either monovalently or polyvalently on filamentous phage. We have named this method 'biosynthetic phage display', and discuss here two applications that explore packing complementarity in a folded protein. In these applications, non-natural amino acids with very different stereochemical characteristics were introduced into the hydrophobic core of eglin, a 70 amino acid residue serine protease inhibitor. The core of the protein is made up of residues derived from both the amino- and carboxy-terminal halves of the protein. In each experiment, a synthetic fragment containing the first 33 amino acids was synthesized with a non-natural amino acid at position 25, which is a phenylalanine residue in the wild-type protein. The sidechain of Phe25 makes van der Waals contacts with those of Val52 and Val54 [25–27]. Thus, although position 25 can be manipulated as part of the synthesized amino-terminal peptide, eglin carboxy-terminal fragments that include combinations for residues 52 and 54 are expressed as an extension on the amino terminus of either the gIIIp or gVIIIp polypeptides of the phage coat protein. Taken together these expressed fragments include all of the 400 possible combinations required for each of the amino acid types to be fully represented at positions 52 and 54.

The two non-natural amino acids introduced at position 25 were kynurenine (Kyn), an analog of tryptophan, and norvaline (Nva), a straight chain analog of valine. Kyn is noteworthy in that it has fluorescent characteristics that are very sensitive to dielectric environment, allowing us to assess the local dielectric in the core of eglin [28,29]. Because it is significantly larger than phenylalanine, it was assumed that adjustments of the packing environment, specifically involving the sidechains of residues 52 and 54, might be required to facilitate folding. Although common wisdom would suggest that sidechains smaller than valine would be required at these two positions to accommodate Kyn, larger sidechains might be preferred to maximize packing around the smaller and more flexible Nva.

The findings from the phage-display solutions for amino acid sidechain types for positions 52 and 54 support some of the above predictions; however, the variety of sizes and types of amino acids that were found at positions 52 and 54, producing viable folded proteins, was unanticipated. Consequently, we have developed a powerful technology to couple synthetic and genetic diversity that has produced a rich source of proteins to study with a striking broad set of packing characteristics.

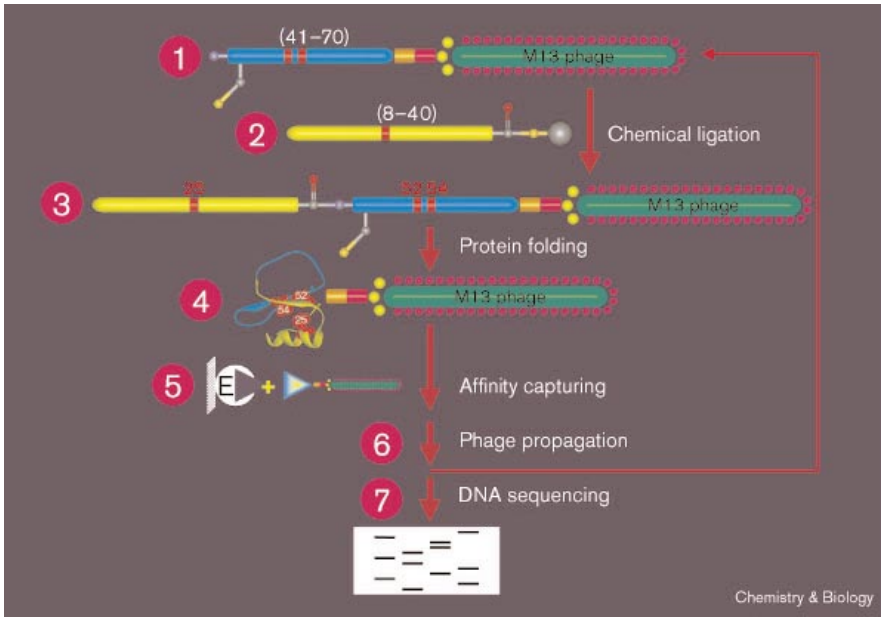
Results

Validating chemical ligation to phage peptides – technical considerations

Figure 1 outlines the method that has been developed and tested for chemically ligating a carboxy-terminal thioester-containing peptide to an amino-terminal cysteine-containing protein fragment, and expressed as a fusion on the surface of a coat protein of filamentous phage. The method involves a set of sequential steps. Consequently, the efficiency of each step affects the yields and ultimately the success of those steps that follow. It was therefore critical to establish protocols that maximized competent products at each step.

Successful application of chemical ligation to a phage-expressed peptide involves addressing several technical issues. The first challenge was in the expression of the

Figure 1



Biosynthetic phage ligation. Steps 1 and 2: the starting reagents are a carboxy-terminal (amino acids 41–70) eglin peptide fragment expressed as a fusion to either gVIIIp or gIIIp (the amino-terminal residue is required to be cysteine); and an amino-terminal synthetic (amino acids 8–40) eglin peptide fragment with a requisite carboxy-terminal thioester (the red stripes designate altered groups, and position 25 is the introduced non-natural amino acid; 52 and 54 are residues mutated in the phage libraries). Step 3: ligation reaction catalyzed by thiophenol yielding a full-length eglin (amino acids 8–70) fused to either gVIIIp or gIIIp on the surface of the filamentous phage. Step 4: the constructed eglin molecule folds into an active conformation. Positioning of the non-natural amino acid at 25 and the mutated residues at positions 52 and 54 are shown. Step 5: a selection assay using a capturing agent is performed to select for the folded eglin ligation products. Step 6: captured phage are propagated and the ligation reaction followed by the selection assay are repeated four times. Step 7: selected clones are isolated and sequenced, identifying the preferred eglin sequences selected from the phage sorting.

viable peptide fragment on the phage coat protein. Expression of amino-terminal cysteine-containing proteins in *Escherichia coli* is generally problematic [30]. As a consequence, we expected the expression level on the filamentous phage coat of this particular fusion product to be lower than for typical peptides. Expression of viable protein fragments can also be difficult because unfolded polypeptides are more susceptible to proteolytic cleavage and aggregation. In addition, we found that the presence of 0.5% thiophenol and 6 M guanidinium HCl used for the ligation reaction reduces the phage titer by ~tenfold (data not shown).

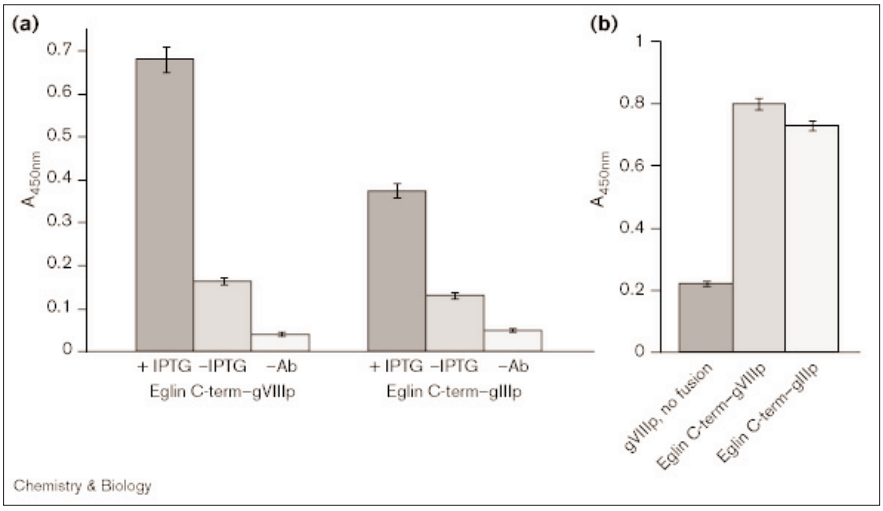
The ligation protocol is designed to couple exclusively through amino-terminal cysteine sidechains, but as the coat protein of the phage contains a large number of free cysteines from the multiple copies of gVIIIp, it was important to ascertain whether any competing reaction products were being produced. A method to ligate a synthetic peptide to an expressed protein using the intein–extein approach has been reported [31]. To take advantage of the expanded level of diversity of protein products available from the phage-display approach, however, we needed to perform this coupling on a phage substrate, which adds another level of complexity. Described below are the methodologies for the central steps of the procedure and the sets of controls used to validate their success and measure their efficiency.

Expression of the carboxyl-terminal eglin fragment on gIIIp and gVIIIp proteins

Expression of the carboxy-terminal eglin peptide fragment by both monovalent (gIIIp) and polyvalent (gVIIIp) phage display was performed by displaying this peptide as a fusion with the respective coat protein [32]. This carboxy-terminal fragment contains residues 41–70 with a Ser41→Cys mutation to provide the free amino-terminal cysteine for the subsequent ligation. This fragment is referred to as C-term. The expression levels of eglin C-term–gIIIp and eglin C-term–gVIIIp fusion protein were compared. Theoretically, a significantly higher expression level would be achieved for a protein fragment–gVIIIp fusion, as there are about 10^3 more copies of gVIIIp than of gIIIp present on the phage coat [33].

Two types of phage-binding ELISA (enzyme-linked immunoabsorbant assay) were used to quantify the expression of eglin C-term–gIIIp and eglin C-term–gVIIIp proteins (Figure 2). The first method used an incorporated amino-terminal His_6 tag as a means to capture the recombinant phage with a Ni-nitrilotriacetic acid (NTA) coated plate. Absorbance values (450 nm) shown in Figure 2a illustrate a 450% (gVIIIp) and a 300% (gIIIp) isopropylthiol galactosidase (IPTG) induction of expression as compared with no IPTG treatment. A saturating level of recombinant phage was applied to the ELISA plate each with a titer of 10^{13} pfu/ml. Interestingly,

Figure 2



Evaluation of eglin C-term–gIIIp and eglin C-term–gVIIIp protein expression with two types of phage-binding ELISAs (average of three ELISAs). **(a)** Absorbance (450 nm) values from a phage-binding ELISA using the incorporated His_6 tag as a means to capture the recombinant phage. IPTG induction of expression is shown by comparing +IPTG (10 μM) and –IPTG treatments. **(b)** Absorbance (450 nm) values from a phage-binding ELISA using the incorporated amino-terminal cysteine residue as a capturing agent for the recombinant phage. Both ELISAs used an anti-VCS-M13-horseradish peroxidase (HRP) conjugated antibody as a means of detection and used a saturating level of recombinant phage each with a titer of 10^{13} pfu/ml.

even though a higher level of expression was observed for the gVIIIp fusion compared with the gIIIp fusion, this difference is not as significant as might be expected given the greater number of potential fusion sites on gVIIIp than on gIIIp.

The second phage-binding ELISA detects the presence of the amino-terminal cysteine by capturing the recombinant phage using a maleimide-coated plate. Figure 2b illustrates roughly a 400% (gVIIIp and gIIIp constructs) absorbance increase as compared with the phage not expressing a fusion protein (VCS-M13). These absorbance 450 nm values represent the full, dynamic range for this assay. These data confirm the presence of the amino-terminal cysteine on the eglin C-term–gVIIIp and eglin C-term–gIIIp fusion proteins.

Quantification of ligation products

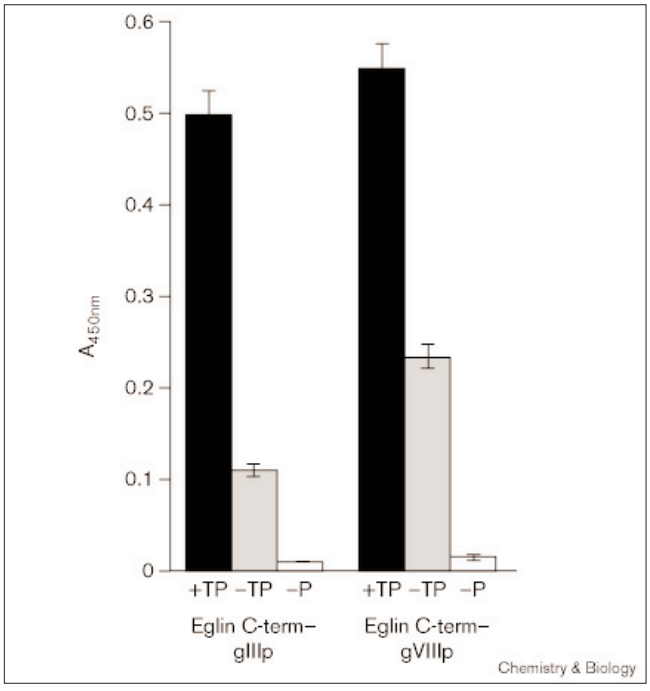
As outlined in Figure 1, chemical ligation of the synthetic eglin amino-terminal peptide to the phage-displayed eglin C-term peptide was conducted, yielding fully active eglin (8–70)–(gIIIp and gVIIIp) fusion proteins. The amino-terminal peptide, referred to as N-term, was synthesized with a carboxy-terminal thioester moiety. N-term had wild-type sequence and consisted of residues 8–40. It had previously been shown that the first eight residues are not a structural requirement and their exclusion has no effect on inhibitor potency [34]. Ligation products were detected with a phage-binding ELISA using the high affinity of eglin for Genenase ([35,36]; $K_d = 0.14$ nM), a variant of the serine protease subtilisin BPN′, as a capturing agent for properly folded eglin. As shown in Figure 3, thiophenol-catalyzed ligations yielded 670% (gIIIp) and 680% (gVIIIp) absorbance (450 nm) increases as compared with thiophenol-catalyzed ligations performed in the absence of peptide. Uncatalyzed ligation reactions yielded 200% (gIIIp) and 300% (gVIIIp) absorbance (450 nm) increases as compared with ligation reactions done without peptide. Reactions performed in the absence of peptide yielded an insignificant number of Genenase binders (0.015 $Å_{450\ nm}$) compared with ELISA background measurements (averaged 0.045 $Å_{450nm}$) that were subtracted from each treatment. These findings also indicate that, although ligation occurs without addition of thiophenol, this catalyst increases the formation of the product by about threefold. Noncovalent association of the eglin N-term and C-term fusion halves is highly unlikely because the sorting wash steps would most likely disrupt any such interactions.

Application to producing viable biosynthetic phage display libraries

The above studies demonstrated that the individual steps to ligate a carboxy-terminal thioester-containing peptide selectively to an amino-terminal cysteine-containing protein fragment of phage could be efficiently achieved. It was not clear, however, when these steps are linked together, whether enough competent biosynthetic phage could be produced to accommodate effective phage selections. To address this issue, we conducted four rounds of ligation of synthetic eglin N-term to eglin C-term–gVIIIp followed by Genenase-binding selection and phage propagation. Following each round of ligation/binding selection, the phage titer was ~10^{13} pfu/ml as determined using standard methods [33]. This finding indicates that ample phage titers are recovered to perform numerous rounds of ligation–binding selection.

To apply the ligation method in the context of the phage-display libraries, two different synthesized eglin N-term peptides were chemically ligated to separate polyvalent phage-display libraries of eglin C-term. These polyvalent

Figure 3



Detection of full-length eglin–(gVIIIp and gIIIp) chemical ligation products with a phage-binding ELISA using the high affinity of eglin for Genenase (K_d = 0.14 nM) to select for properly folded eglin. Thiophenol catalysis (final concentration = 0.5%) of the ligation reaction is indicated by comparison of the absorbance 450 nm values (average of three ELISAs) for the following treatments: thiophenol added (+TP); no thiophenol added (–TP); no peptide added (–P). Background absorbance 450 nm readings (0.045 A 450 nm) were subtracted for each treatment in Figures 3 and 4.

C-term libraries were designed to mutate simultaneously at positions Val52 and Val54. The two synthetically produced amino-terminal peptides contained different non-natural amino acid at position 25. (In the folded native protein, position 25 contains a phenylalaine residue that is buried in the hydrophobic core and packs against Val52 and Val54.) One of the non-natural amino acids was Kyn (N-term (Kyn)), which is a fluorescent analog of tryptophan, and the other was Nva (N-term (Nva)), a straight-chain analog of valine.

Variants of eglin C-term at positions Val52 and Val54 were generated by constructing two phagemid libraries. The first library (NNS) was designed to contain codons for all 20 amino acids at positions 52 and 54 (Table 1). This process is referred to as 'hard' randomization. The second library (VRS) was constructed to restrict the residues in these two positrons to histidine, glutamine, asparagine, lysine, aspartate, glutamate, arginine, serine or glycine, thus choosing these residues to be polar. The first library therefore contains all the potential solutions for substitutions at 52 and 54 to accommodate the larger Kyn and the flexible Nva. The polar-biased library is much more restricted, with glycine being the 'rescue' residue if no combination of polar residues is possible at these positions (Table 1).

Ligation to the phage fragment

Following construction and propagation, the expressed eglin C-term protein fragments produced in both these phage libraries were chemically ligated to the N-term (Kyn) and N-term (Nva), yielding two sets of binding-competent eglin variants. The ligation products from the hard-randomized (NNS) and polar-biased (VRS) libraries were detected using the Genenase-coated phage-binding ELISA described above (Figure 4). The N-term (Nva) selections did not include the VRS library. The eglin Val52 and Val54 variants from both NNS and VRS libraries showed a ~sixfold absorbance (450 nm) increase relative to ligations performed in the absence of peptide.

Screening involved four rounds of ligation followed by Genenase-binding selection were carried out for these eglin (with Phe25→Kyn and with Phe25→Nva) polyvalent phage libraries (as described in the Materials and methods

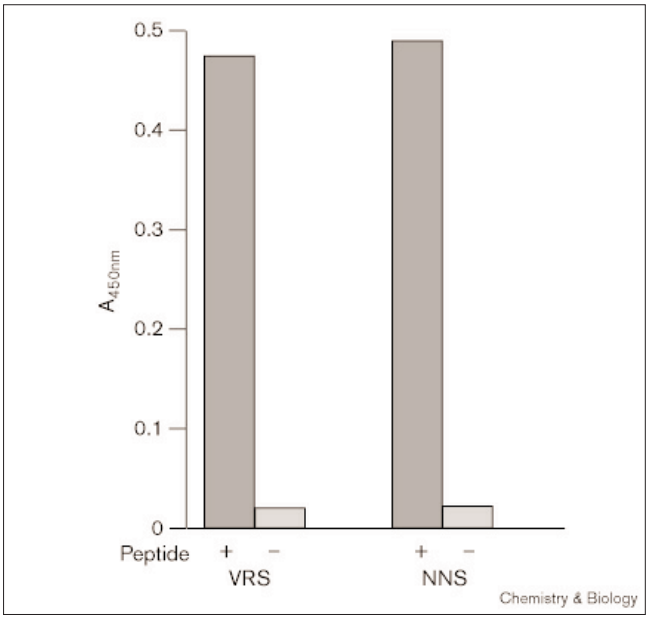
Table 1

Codons and relevant statistical information for choosing amino acids to be a specific type.

Name	Codon*	Amino acid	No. codons	Total†	Library size needed‡	Max no. residues§
Polar	VRS	H,Q,N,K,D,E,R,S,G	12	$4(12)^8$	2×10^9	8
Nonpolar	NYS	K,L,S,P,M,I,T,V,A	16	$4(16)^7$	1×10^9	7
Nonaromatic	VYS	L,P,M,I,T,V,A	12	$4(12)^8$	2×10^9	8
Aromatic	TNS	F,L,S,Y,stp,C,W	8	$4(8)^9$	5×10^8	9
Hard randomization	NNS	All	32	$4(32)^5$	1×10^9	5

Characteristics of phage libraries choosing polar, nonpolar, aromatic, nonaromatic and nonbasic residues. For each type of biased residue, the represented codon, amino acids (using single-letter code) and number of codons are listed in the above table. stp. stop codon (TAG). *The nomenclature for these codons is as follows: N = G + A + T + C; V = A + C + G; Y = C + T; R = A + G; S = C + G. Codon base wobbles are synthesized using deoxynucleotide mixtures as follows: four way, 25% each; three way, 33% each; two way, 50% each. †For each biased library type, the maximum number of randomized residues is determined based on statistical calculations for total representation ($4(\text{number codons})^n$), where n is number of randomized residues [33] and current electroporation limits (1×10^{10} independent transformants/library). ‡Transformants required for total representation of biased libraries. §Maximum number of residues that can be simultaneously mutated preserving total representation.

Figure 4



Chemical ligation to produce phage libraries consisting of eglin (amino acids 8–70)–gVIIIp variants containing the non-natural amino acid Kyn at position 25. Absorbance values (450 nm) from a phage-binding ELISA (described in Figure 3) performed following the final (fourth) round of phage library sorting are shown. Eglin C-term–gVIIIp served as the positive control for the ligation reaction (data not shown) and ligation in the absence of peptide was the negative control.

Table 2

Amino acids for eglin residues 52 and 54 after sorting eglin (F25Kyn)–gVIIIp and eglin (F25Nva)–gVIIIp phagemid libraries.

Kyn at position 25						Nva at position 25		
NNS			VRS			NNS		
Freq.	52	54	Freq.	52	54	Freq.	52	54
	G	V	6	G	G		K	L
	G	E		K	D	3	A	V
	G	A	2	D	G	3	L	V
	V	C		E	E		I	L
4	V	V		G	N		P	L
	K	H		G	K	2	N	V
2	V	G		N	K	2	T	V
	L	G		G	E		V	T
	V	E		E	G	2	P	V
2	L	V		Q	G	4	V	L
	G	D		H	G	3	V	V
	K	V		N	G	2	V	A
	M	F		K	R	2	I	V
	G	Y				2	L	D
	V	L				2	L	A
	W	G					I	A
	V	D					V	C
	P	C					D	A
	R	W				2	T	L
	R	G					V	I
	E	R				2	N	L
	G	G				2	L	I
							D	V
							T	A
						3	A	L

Frequency indicates the number of times the pattern of sequence was found.

section). After the fourth round of ligation/selection, enrichment of the eglin (with Phe25→Kyn) phage libraries was ~tenfold for the hard-randomized library (NNS) and 15-fold for the polar-biased library (VRS) compared with the background binding to a phosphate buffered saline (PBS)-coated control well. Similarly, four rounds of ligation/selection of the eglin (Phe25→Nva) phage libraries yielded an eightfold enrichment for the hard-randomized library (NNS) and 14-fold enrichment for the polar-biased library (VRS) as compared with background binding to a PBS-coated well.

Kynurenine at position 25

Table 2 contains the resulting amino acids obtained from the hard randomization (NNS) of positions 52 and 54 with Kyn at position 25. The phage-derived clones were isolated following four rounds of selection. The composite of the 27 sequenced clones of Phe25→Kyn showed a well-defined pattern at both positions 52 and 54. On the basis of codon usage, the hard-randomized library indicated a preference for either glycine, with a standard deviation (σ_n) of 3.4 σ_n for amino acid 52 and 3.3 σ_n for amino acid 54 or valine, 9.9 σ_n for amino acid 52 and 7.7 σ_n for amino acid 54 (Table 3). The σ_n and the significance scores were calculated as shown in Table 3. Only residues for which σ_n was greater than 1 were considered significant. Also noteworthy, clones were isolated that had a wide variety of amino acid types at these positions. These involved, among others, residues with large sidechains such as tryptophan, as well as those with both basic and acidic moieties. Although it would be expected that proteins with large or charged sidechain types at either position mutated would be less stable than those with more complementary groups, this latter category still represents a group of viable proteins.

In the polar-biased libraries, both positions showed a preference for glycine [3.1 σ_n (amino acid 52); 4.75 σ_n (amino acid 54)] (Table 3). In 6 of the 19 sequences, residues 52 and 54 were both glycines. In eight other sequences, glycine was found at one of the positions. The polar residues found at one or the other position had a fairly broad spectrum of stereochemical traits. Only serine was not represented among the possible sidechain types contained in the polar library. The extent of the variability is characterized by having sequences that were fully acidic, Glu:Glu; fully basic, Lys:Arg; or a mix, Lys:Asp.

Norvaline at position 25

Eglin (Phe25→Nva) library phage clones were isolated following four rounds of selection followed by DNA

Table 3

Consensus residues identified after sorting eglin (Phe25→Kyn)–gVIIIp phagemid libraries.

Position	Residue	P_e	σ_n	P_f	$P_f–P_e/\sigma_n$
NNS library					
Val52	G	0.062	0.046	0.259	3.4
	V	0.031	0.033	0.370	9.9
	K	0.031	0.033	0.074	1.2
Val54	G	0.062	0.046	0.222	3.3
	V	0.031	0.033	0.296	7.7
	D	0.031	0.033	0.074	1.2
	E	0.031	0.033	0.074	1.2
	C	0.031	0.033	0.074	1.2
	Y	0.031	0.033	0.074	1.2
VRS Library					
Val52	G	0.167	0.080	0.409	3.1
Val54	G	0.083	0.059	0.545	4.8

The most frequently occurring residues from the NNS and VRS phage libraries are shown, based on fractional representation (P_f) among all clones isolated after four rounds of binding selection. Expected frequencies (P_e) were calculated from the number of VRS and NNS codons for each amino acid theoretically in the starting library. Standard deviations (σ_n) were calculated as $\sigma_n = [P_e(1–P_e)/n]^{1/2}$ and termed a 'significance score'. Only residues for which the fraction found exceeded the fraction expected by at least $1\sigma_n$ (i.e. $[(P_f–P_e)/\sigma_n] > 2$) are shown. For the NNS library, n = 27 sequences; for the VRS library, n = 22 sequences.

isolation and sequencing. The sequence analysis of the 46 clones indicated a different preferred sequence for the two positions (Table 2). Position 54 showed a definite preference for valine (13.8 σ_n) and other nonaromatic hydrophobic residues (Leu, 4.8 σ_n; Ala, 2.5 σ_n; Ile, 1.2 σ_n). Position 52 showed a more variable set of sequences (Leu, 2.4 σ_n; Ala, 1.9 σ_n; Ile, 2.2 σ_n; Asn, 2.2 σ_n; Tyr, 1.3 σ_n), but retained a strong preference for the wild-type valine (8.8 σ_n).

Protein diversity versus DNA diversity

In the hard randomization of residues 52 and 54 of the Phe25→Nva mutant, several sets of identical solutions were isolated (Table 2). At issue was whether these represented true independent solutions or whether their repeated presence was a function of stability or expression factors. Table 4 shows the codon usage for three sets of these solutions: valine, leucine (frequency [n] = 4); alanine, leucine (n = 3); alanine, valine (n = 3). The DNA sequence data indicate that each of the solutions in the set are independent at the DNA level. Thus, it is apparent that the phage selection is a robust process that uses a diverse set of DNA sequences to derive a more narrow set of protein solutions.

Discussion

The coupling of synthetic chemistry to phage-display mutagenesis significantly expands the capabilities to build new functionality into proteins by combining chemical and genetic diversity into a single method. It has long been recognized that non-natural amino acids, if they could be substituted at specific sites in a protein, could alter protein function in quite unique and important ways. The question has centered on how to incorporate them.

Table 4

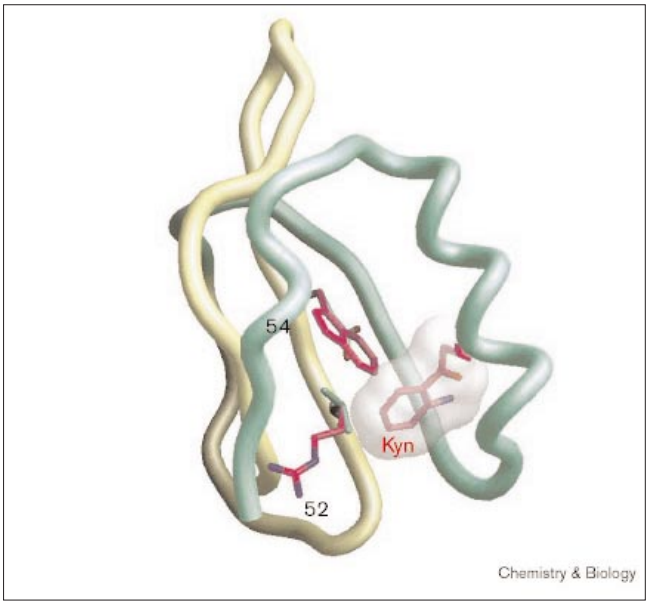
Comparison of position 52 and 54 eglin library amino acid and DNA sequences.

	Position 52		Position 54	
Pair at 52,54	Amino acid	CODON	Amino acid	CODON
V, L	V	GTC	L	CTC
	V	GTG	L	CTG
	V	GTC	L	TTG
A, L	A	GCC	L	CTG
	A	GCG	L	CTC
	A	GCC	L	CTC
A, V	A	GCC	V	GTG
	A	GCG	V	GTG
	A	GCC	V	GTC

The challenge of introducing non-natural amino acids into a protein sequence has been approached in several different ways. The more traditional methods involve chemical modification of existing amino acids [37] or *in vitro* translation using misacylated tRNAs [16,17]. However, using the first approach has inherent difficulties in achieving specificity and the latter approach often fails to produce a reasonable amount of product. In an alternative approach, synthetic peptides containing non-natural amino acids have been ligated to protein fragments by enzymatic means using a peptide ligase [38]. Many other approaches have been used for enzymatic semisynthesis [39]; however, each of these techniques have limitations related to enzymatic specificity, solvent constraints or solubility issues. Recently, native chemical ligation has emerged as a general method to synthesize full-length proteins and allows specific incorporation of a wide variety of non-natural amino acids [19]. This approach has been extended to allow ligation of synthesized peptides with expressed protein fragments using an intein–extein strategy (recently reviewed [40]).

As potentially powerful as the native ligation method is, it has one important drawback. Unless the introduced non-natural amino acid is isosteric with the one it replaces, forming a stable protein might require additional changes in the protein core to reoptimize the packing environment of the introduced group. For instance, we describe here a study where a large fluorescent reporter group is introduced into the hydrophobic core of eglin. It replaces a somewhat smaller group suggesting that the additional volume has to be accommodated by a set of packing adjustments.

Figure 5



Structure of eglin c. The green chain is the synthesized fragment 8–40 with Kyn at position 25. The yellow chain is an expressed phage fragment (41–70) with altered groups at 52 and 54. Proposed packing fit is shown with Kyn at residue 25 and arginine or valine at position 52 and tryptophan or aspartate at position 54.

Although rational design solutions to such problems are intellectually pleasing, lessons learned from many protein engineering studies suggest that genetic approaches are often better providers of the optimum solutions. It is at this interface, between introducing a nonisoteric amino acid into a folded protein and being able to accommodate it, that the biosynthetic phage-display approach is very powerful. We have shown that the method will select those solutions that are most highly preferred, as well as indicating the range of amino acid types that can produce viable protein products.

Introduction of Kyn at position 25 of eglin

Kyn is a derivative of tryptophan that can be used as a fluorescent reporter group for assessing local dielectric environments [28,29]. Our objective was to introduce Kyn into the interior of eglin thereby enabling us to obtain direct experimental data relating to apparent dielectric values within the protein core. Such information has been elusive and represents a critical step in improving modeling of electrostatic effects in protein interiors and active sites [41,42].

Previous attempts to introduce Kyn and other similar reporter groups into protein systems have relied on chemical modification or passive diffusion approaches [37]. Each of these methods suffers from specificity and efficiency issues. Using native chemical ligation, the eglin molecule can be readily synthesized [24]. As Kyn can be purchased in amino acid form, there are no technical issues with inserting Kyn in place of phenylalanine at position 25. However, Kyn is significantly larger (~ 60 Å^3) than phenylalanine and has a polar component, which together could perturb folding and stability of the protein. In an attempt to circumvent these problems and to ascertain the nature of possible environments around the Kyn sidechain, we used the phage display arm of the method to screen for compensating packing mutations. We note that this method samples all possibilities because the size of the library produces up to 10^9 protein sequences, and thus is highly over-determined to produce all 400 combinations possible in the simultaneous randomization of two positions.

Screening phage-display-derived clones of eglin (Phe25→Kyn) for binding to Genenase yielded a strong preference for nonpolar groups at positions 52 and 54. Only about 30% of the solutions had a polar group at either position 52 or 54. Based on expected probabilities, the presence of glycine and valine is disproportionally high. Although the occurrence of these small nonpolar amino acids might be predicted, the presence of other solutions with larger amino acids was unexpected.

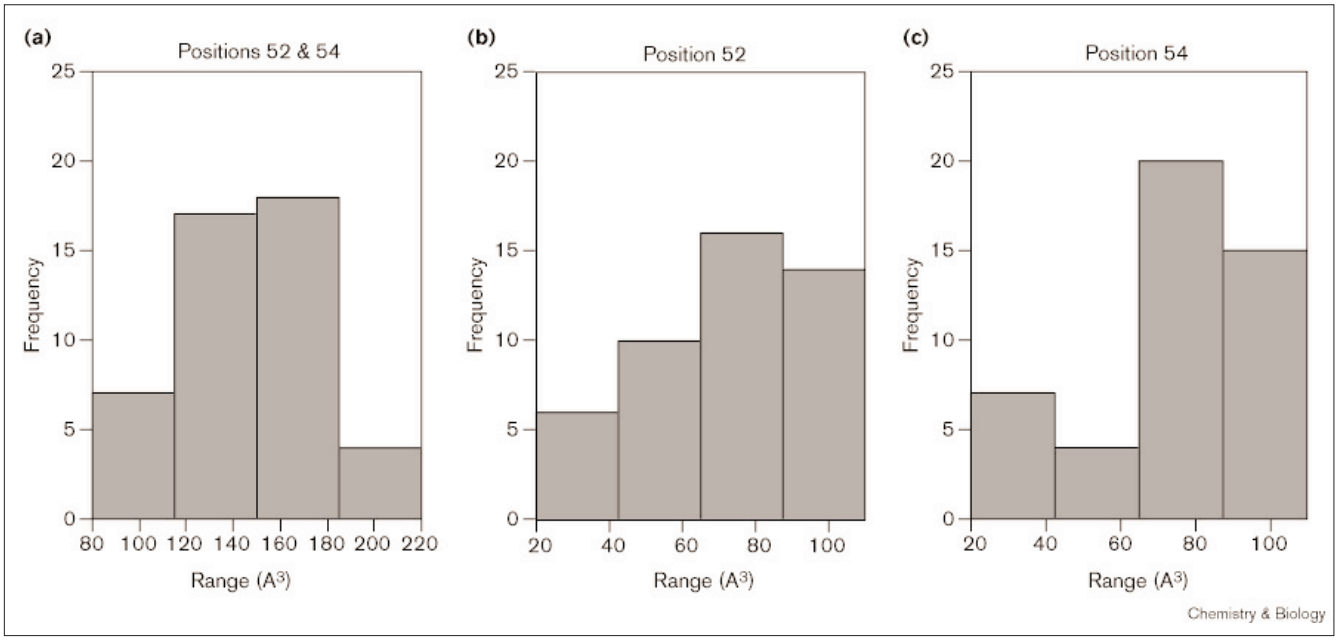
In the polar-biased library, six of the 19 sequenced clones had Gly52–Gly54 solutions. Nine clones had Gly52–X54 or visa versa and only four clones were found with polar residues at both positions. These results suggest that the rescue solution, Gly–Gly, predominates in the selections.

In modeling studies of the eglin (Phe25→Kyn), Gly and Val could be accommodated at positions 52 and 54. Figure 5 pictures the stereochemical relationships between a putative Kyn25 orientation and two different solutions for observed 52 and 54 sidechain types. These solutions were picked to be modeled because the combination of residue types found were not expected based on size and charge criteria. The model of Phe25→Kyn, Val52→Arg and Val54→Trp shows that the Kyn can stack perpendicular to the Trp54 sidechain with the arginine sidechain rotated such that some of it aliphatic carbons are packed against the tryptophan sidechain. In the modeled orientation, the guanidinium moiety has exposure at the surface. In the Val52–Val54→Glu mutant, the Kyn sidechain can be flipped (data not shown) to accommodate the energetically most favorable conformations of the 52 and 54 sidechains, while additionally forming an hydrogen bond between the carboxylate oxygen of Val52→Glu and the phenolic nitrogen of Kyn (~3.2 Å). We note that the accuracy of these assignments will have to be verified through a direct structural analysis.

Norvaline at position 25

Substitution of phenylalanine for Nva at position 25 may result in the formation of a cavity, because the volume of

Figure 6



A histogram plotting the frequency of occurrence versus the range of volumes for eglin (Phe25→Nva) positions 52 and 54. **(a)** The volume distribution for both positions 52 and 54 indicates a symmetric distribution around the volumes of the two wild-type residues. Comparison of the volume distributions for **(b)** position 52 and **(c)** position 54 suggests a slight skewing toward sidechain 54 being larger than 52.

this residue is ~50 Å^3 less than phenylalanine. After the phage sorting, hydrophobic residues were observed for position 54, but a broader distribution of residue types were observed for position 52. Of the 46 clones sequenced, 31 had hydrophobic groups at both positions. (valine or leucine appeared 21 times at position 52 and 32 times at position 54.) The remaining 14 clones had some combination of a nonpolar, polar mix. Note, however, that in 12 of these cases, it was position 52 that accommodated the polar sidechain. This suggests that the environment surrounding it is somewhat more amenable to polar groups than that around residue 54.

Figure 6a is a histogram showing the distribution of the summed volumes of the 52 and 54 sidechains with a Nva at position 25. Generally, there is a symmetric distribution around the volumes of two valine residues (2×75 Å^3 = 150 Å^3), which are the wild-type residue types. There are clearly viable proteins made that have significantly larger volume displacements, for instance Lys52–Leu54 (206 Å^3) or Leu52–Ile54 (204 Å^3). Although these occurrences are somewhat rare, modeling suggests that these can be accommodated with modest adjustments of surrounding groups. The distribution of size between positions 52 and 54 indicates a slight skewing towards the sidechain 54 being larger than the 52 sidechain (Figure 6b,c).

Significance

Protein engineering is powerful approach to developing a better understanding of the fundamental rules linking a protein's structure to its function. Two recently developed methodologies promise to make significant contributions. In the first technique, total chemical synthesis of proteins, using native chemical ligation, facilitates the introduction of non-natural amino acids and alteration of peptide bond stereochemistries. This synthetic access to chemical derivatives greatly expands the potential repertoire of function. The second technique, phage-display mutagenesis, is extraordinarily efficient in producing and evaluating an extremely large number of possible solutions with a desired physical property. It has been coined 'evolution in a test tube' because the library sizes allow up to 10^9 protein sequences to be evaluated in a single experiment.

Biosynthetic phage display synergizes the integration of synthetic chemistry and genetic diversity, providing a unique method to build new and uniquely functional attributes into proteins. We believe that application of this method to engineer protein cores, active sites and macromolecular interfaces will contribute greatly to our ability to both understand and rationally manipulate the physicochemical properties that drive protein function. Non-natural amino acids with a functional group can be

introduced via the chemical synthesis arm of the method, whereas the local packing environment of the incorporated group can be optimized for a desired trait by altering the steric and electrostatic features of neighboring residues using phage-display mutagenesis.

An important aspect of the biosynthetic phage display method is that it not only identifies the optimal solution, but it can also generate a broad spectrum of unexpected but viable solutions. This latter category provides a rich source for the biophysical analysis of proteins that have undergone structural compromise to accommodate non-ideal packing features. For instance, for the introduction of the fluorescent reporter group, kynurenine (Kyn), the resulting sequences based on phage-display mutagenesis suggested that the most preferred solutions contained small hydrophobic amino acids at positions 52 and 54. However, solutions with tryptophan or charged sidechains in these positions were also found. This was unexpected based on the stereochemistry of Kyn. Additionally, phage-display libraries were constructed to express a subset of residue types based on a particular stereochemistry or hydrophobicity. For example, only hydrophobic or charged amino acids would be allowed for determining a given packing solution. In the study presented here, phage-display libraries were created that required residues 52 and 54 to be polar or glycine. Although a significant fraction of the solutions involved selection of glycine, a number of combinations of charged sidechains also made viable proteins. These are obviously the ones that will be interesting to analyze further structurally and biophysically.

Materials and methods

Construction of gVIIIp and gIII phagemid vectors with eglin (nucleotides 120–210)

Standard recombinant DNA techniques were used to construct recombinant phagemid transfer vectors based on the vectors ps657(gVIIIp) and ps657(gIIIp) (provided by S. Sidhu, Genentech). The M13 gIIIp display vector ps657(gIIIp) was constructed by replacing the *Eco*RI/*Nsi*I fragment of pH0753 (a monovalent human growth hormone (hGH) display vector [33]) with a PCR amplified 1.6 kb fragment from pMalp2 (New England Biolabs). Beginning at nucleotide 1 for both the pMalp2 and ps657(gIIIp) vectors, this fragment contained the *lacIq* gene and a gene fragment encoding the signal peptide from the maltose-binding protein under control of the isopropylthiogalactoside (IPTG)-inducible Ptac promoter. As a result, production of the eglin C-term fusion protein was both IPTG-inducible and tightly regulated in the absence of IPTG by the *lacIq* gene.

The hGH gene was then removed and replaced by a glycine/serine linker (SGGGSGSG). The M13 gVIIIp display vector ps657(gVIIIp) was analogous to ps657(gIIIp) except that gIIIp was replaced by gVIIIp and the glycine/serine linker (GGGSGSSS) was slightly different. Both vectors contained linker regions with unique *Nsi*I and *Xba*I sites which allowed directional in-frame cloning of the eglin gene fragment using *Nsi*I (at the 5′ end) and *Xba*I (at the 3′ end). Specifically, both the ps657(gVIIIp and gIIIp) vectors were linearized with *Nsi*I and *Xba*I and treated with shrimp alkaline phosphatase (Promega). The carboxy-terminal portion of eglin (nucleotides 120–210) was PCR amplified to introduce two modifications: first, the first 5′-end codon (nucleotides 120–123) was mutated from serine to cysteine; secondly, a His_6 tag coding sequence was introduced adjacent to the 3′-end of the eglin sequence. The template for the polymerase chain reaction (PCR) amplification was the vector, pET27b (Novagen) with the eglin gene inserted at *Nde*I (5′) and *Bam*HI (3′) sites (provided by T. Komiyama at the University of Michigan).

The following oligonucleotides: forward, 5′-PAGCATC**ATGCAT**TGT-CCTGTTACCCTG-3′ *Nsi*I site in bold; reverse, 5′-PAGCATC-**TCT-AGA***GTGATGGTGATGGTGAT**GACCAACATGCGG***-3′ (*Xba*I site in bold, His_6 tag in italics and the eglin sequence is bold and italicized) were used as primers. The linearized vectors were then ligated with the eglin (nucleotides 120–210)/ His_6 tag cDNA insert. Following ligation, competent *E. coli* XL1-Blue were transformed, and bacteria were selected for the correct recombinant plasmids (termed ps657(gVIIIp)–eglin (nucleotides 120–210) and ps657(gIIIp)–eglin (nucleotides 120–210) by DNA sequence analysis using the dRhodamine dye-terminator method and an Applied Biosystems ABI Model 377 automated DNA sequencer. This construct yielded a fusion protein that consisted of a secretion signal, eglin C-term peptide, a His_6 tag, a glycine/serine linker and either protein gIIIp or protein gVIIIp. An amber stop (TAG) codon was present between the eglin (nucleotides 120–210)/ His_6 tag protein and the linker region (within the *Xba*I cut site). An amber suppression *E. coli* strain was therefore used for phage display (e.g. XL1-Blue).

Phagemid particles displaying the eglin C-term fragment (either monovalently or polyvalently) were prepared using the following method. ps657 (gIIIp and gVIIIp)–eglin (nucleotides 120–210) constructs were transformed into *E. coli* XL1-Blue cells (Stratagene). Then, mid-log cultures (250 ml 2× YT with 50 μg/ml ampicillin and 10 μM IPTG, in a 2 l beveled flask at 225 rpm, 37°C) were infected with VCS-M13 helper phage (Stratagene) at a multiplicity of infection equal to three. Infected cultures were grown for 16–18 h and resultant phagemid particles were purified by two polyethyleneglycol (PEG; MW = 8000, 20% with 2.5 M sodium chloride) precipitations from 250 ml of clarified culture supernatants. Phagemid particles were resuspended in peptide ligation buffer (0.2 N potassium phosphate pH 7.5/6 M guanidine hydrochloride) Phage titers were determined by infecting tenfold serial dilutions with mid-log phase *E. coli* XL1-Blue and observing resultant plaques. Typical phage titers ranged from 1×10^{14} to 1×10^{15} plaque forming units/ml.

Peptide synthesis

Experimental details for the total synthesis of eglin C-term by native chemical ligation have been described by Lu *et al.* [24]. For this study, the two thioester peptides, (8–40)Kyn25–SR and (8–40)Nva25–SR, were synthesized in stepwise fashion on custom-made thioester-generating resin [21] using the published *in situ* neutralization/HBTU activation protocol for Boc chemistry solid-phase peptide synthesis [43]. The following sidechain protection was used: arginine(Tosyl); asparagine(Xanthyl); aspartate(OcHxl); glutamine(OcHxl); histidine(Bom); lysine(2ClZ); serine(Bzl); threonine(Bzl); tyrosine(BrZ). After HF deprotection and cleavage from the resin, the crude peptides were purified by preparative reversed phase high-performance liquid chromatography. Molecular weights of the peptides were verified by electrospray ionization mass spectrometry (data not shown).

Phage-binding ELISA for detection of eglin C-term–gVIIIp and eglin C-term–gIIIp fusion proteins

Eglin C-term–gIIIp and eglin C-term–gVIIIp recombinant phage expression was evaluated by ELISA using the incorporated His_6 tag as a means to capture the fusion phage–proteins. Thus, eight fivefold serial dilutions of the phage (100 μl/well; phage resuspended and diluted in Buffer #1 (PBS with 0.2% bovine serum albumin (BSA; Calbiochem) and 0.1% Tween 20) were applied to Ni-NTA Hisorb 96-well microtiter plates (Qiagen; 1.5 h, room temperature, moderate agitation). Non-induced phage and no application of detecting antibody served as negative controls. The plates were washed 3× with 400 μl/well of Buffer #1 and blotted dry. Anti-VCS-M13-horse-radish peroxidase (HRP; Pharmacia)

conjugated antibody was diluted 1:5000 in Buffer #1 and 100 µl/well was applied (1 h, room temperature, moderate agitation). The wash step was then repeated as described. Colour was developed by adding 100 µl/well of 1-step Turbo 3, 3′, 5, 5′-tetramethylbenzidine (TMB) substrate (Pierce) for 5–10 min and stopped by adding 100 µl/well 4.5 N sulfuric acid. The absorbance 450 nm ($A_{450\ nm}$) was measured and plotted in histograms as shown (Figures 2–4).

Amino-terminal cysteine determination ELISA

The presence of the eglin C-term (gVIIIp and gIIIp) amino-terminal cysteine residue was detected by a phage-binding ELISA, using a maleimide-coated plate (Pierce; 100 µl/well; phage resuspended in Buffer #1 for 1.5 h at room temperature with moderate agitation). This maleimide-coated plate covalently captured recombinant phage expressing the amino-terminal cysteine. Phage expressing gVIIIp without a fusion protein were included as a negative control. All subsequent washing and detection steps were performed as described in the phage-binding ELISA section.

Chemical ligation of eglin N-term to eglin C-term/gIIIp and eglin C-term/gVIIIp

Figure 1 depicts the chemical ligation process. Specifically, the ligation was conducted using recombinant phage resuspended in 1 ml reaction buffer (0.2 N potassium phosphate, pH 7.5 with 6 M guanidinium hydrochloride). The concentration of recombinant phage fusion proteins was estimated to be 0.2 nM, assuming a titer of 10^{13} pfu/ml with a viable fusion protein expression rate of 1%. The eglin (N-term) peptide with a thioester at the carboxyl terminus was resuspended in the reaction buffer at a final concentration of 100 µM, yielding a large ratio of peptide to recombinant phage fusion protein. Thiophenol (0.5% final concentration) was added to catalyze the ligation reaction performed at room temperature with gentle rotary mixing for 24 h. Addition of 10 mM (final concentration) dithiothreitol quenched the reaction. Then, eglin (gVIIIp and gIIIp) ligation products were detected using the phage-binding ELISA described below. Two controls were included for each treatment. As the ligation reaction will proceed at a slower rate in the absence of thiophenol, a reaction omitting thiophenol was included to show thiophenol catalysis of the reaction. In addition, a reaction without addition of peptide was included for each treatment as a negative control.

Phage-binding ELISA detection of eglin (gVIIIp and gIIIp) ligation products

Eglin has a high affinity for Genenase ($K_d = 0.14$ nM [36]) allowing this protease to be used as a selection agent for properly folded eglin following the ligation reaction described above. Genenase is a subtilisin variant generated by mutating the catalytic histidine (His64) to alanine [35]. Although this enzyme is inactive, it still retains its strong binding capacity for most substrates and inhibitors.

Maxisorb 96-well plates (Nunc) were coated with Genenase (10 µg/ml in 50 mM Carbonate buffer, pH 9.6; 100 µl/well, o/n, 4°C). After washing the plates as described above, the plates were blocked with 5% BSA, 0.01% Tween-20 in PBS (200 µl/well, 2 h, room temperature with moderate agitation). After further plate washing, the phage from the ligation reaction were applied to the blocked, Genenase-coated plate (100 µl/well, 2 h, room temperature with moderate agitation). All subsequent wash and detection steps were performed as described above.

Construction of eglin C-term–gVIIIp phagemid libraries

Two eglin C-term–gVIIIp-phagemid libraries were constructed varying the residues at positions Val52 and Val54. The hard-randomized library (NNS) randomly mutated Val52 and Val54 using the following phosphorylated oligonucleotide: 5′-TAC AAC CGT NNS CGC NNS TTC TAC AAC-3′, where N represents any of the four nucleotides and S represents either a G or a C. The polar-biased library (VRS) allowed Val52 and Val54 to be histidine, glutamine, asparagine, lysine, aspartate, glutamate, arginine, serine or glycine, using the following phosphorylated oligonucleotide: 5′-TAC AAC CGT VRS CGC VRS TTC TAC AAC-3′, where V represents a C, A, or G; R represents either an A or a G; and S represents a G or a C (J.J.D., unpublished observations). Oligonucleotide-directed mutagenesis was used to construct these libraries using ps657(gVIIIp)–eglin (nucleotides 120–210) as a template. The library mutagenesis products were electrotransformed into *E. coli* XL1-Blue (Stratagene) cells for phagemid preparation yielding 1×10^9 independent transformants as described previously [44].

Sorting of NNS and VRS (positions V52, V54) eglin(C-term)–phagemid libraries:

Four rounds of ligation followed by binding selection using Genenase were performed for these libraries. First, propagated phage generated from each library was chemically ligated to the eglin peptide. Then, ligation products corresponding to eglin with either Kyn or Nva at position 25 were selected on the basis of their ability to properly fold and bind to Genenase. Specifically, Genenase was coated onto 96-well Maxisorb plates (Nunc) (10 µg/ml in 50 mM Carbonate buffer, pH 9.6; 100 µl/well, overnight, 4°C) and blocked as described above. Both libraries were sorted by adding ~1×10^{13} phage particles to the Genenase-coated wells as well as duplicate PBS mock coated wells for a negative control. All wells were washed 10× as described above. Phage particles were eluted with 0.2 M glycine, pH 2.0, neutralized with 1 M Tris base. The phage were titered and propagated as previously described. Following propagation, phage from each library were harvested (PEG precipitated similar to described above) and chemically ligated to the eglin peptide to begin a second round of selection. Following four rounds of sorting, DNA from individual clones was isolated and sequenced (described above).

Acknowledgements

We would like to thank Dr Sidhu and Dr Komiyama for providing plasmid vectors. Additionally, we would like to thank Dr Pal and Dr Bernat for providing useful suggestions during the manuscript preparation.

References

1. Dalal, S., Balasubramanian, S. & Regan, L. (1997). Transmuting α helices and β sheets. *Fold. Des.* **2**, R71-R79.
2. Hall, T.M., Porter, J.A., Young, K.E., Koonin, E.V., Beachy, P.A. & Leahy, D.J. (1997). Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* **91**, 85-97.
3. Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T. & Kim, P.S. (1998). High-resolution protein design with backbone freedom. *Science* **282**, 1462-1467.
4. Porter, J.A., *et al.,* & Beachy, P.A. (1996). Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell* **86**, 21-34.
5. Scott, J.K. & Smith, G.P. (1990). Searching for peptide ligands with an epitope library. *Science* **249**, 386-390.
6. Adams, G.P., *et al.,* & Weiner, L.M. (1998). Increased affinity leads to improved selective tumor delivery of single-chain Fv antibodies. *Cancer Res.* **58**, 485-490.
7. Atwell, S., Ultsch, M., De Vos, A.M. & Wells, J.A. (1997). Structural plasticity in a remodeled protein–protein interface. *Science* **278**, 1125 1128.
8. Demartis, S., *et al.,* & Neri, D. (1999). A strategy for the isolation of catalytic activities from repertoires of enzymes displayed on phage. *J. Mol. Biol.* **286**, 617-633.
9. Lowman, H.B. & Wells, J.A. (1993). Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol.* **234**, 564-578.
10. Oliff, A. (1999). Farnesyltransferase inhibitors: targeting the molecular basis of cancer. *Biochim. Biophys. Acta* **1423**, C19-C30.
11. Spada, S. & Pluckthun, A. (1997). Selectively infective phage (SIP) technology: a novel method for in vivo selection of interacting protein–ligand pairs. *Nat. Med.* **3**, 694-696.
12. Wolfe, S.A., Greisman, H.A., Ramm, E.I. & Pabo, C.O. (1999). Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.* **285**, 1917-1934.
13. Dennis, J.W., Granovsky, M. & Warren, C.E. (1999). Protein glycosylation in development and disease. *Bioessays* **21**, 412-421.
14. Magnaghi-Jaulin, L., Ait-Si-Ali, S. & Harel-Bellan, A. (1999). Histone acetylation in signal transduction by growth regulatory signals. *Semin. Cell Dev. Biol.* **10**, 197-203.

15. Swope, S.L., Moss, S.I., Raymond, L.A. & Huganir, R.L. (1999). Regulation of ligand-gated ion channels by protein phosphorylation. *Adv. Second Mess. Phosphoprot. Res.* **33**, 49-78.
16. Cload, S.T., Liu, D.R., Froland, W.A. & Schultz, P.G. (1996). Development of improved tRNAs for *in vitro* biosynthesis of proteins containing unnatural amino acids. *Chem. Biol.* **3**, 1033-1038.
17. Heckler, T.G., Zama, Y., Naka, T. & Hecht, S.M. (1983). Dipeptide formation with misacylated tRNAPhes. *J. Biol. Chem.* **258**, 4492-4495.
18. Paulus, H. (1998). The chemical basis of protein splicing. *Chem. Soc. Rev.* **27**, 375-386.
19. Dawson, P.E., Muir, T.W., Clark-Lewis, I. & Kent, S.B. (1994). Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779.
20. Hackeng, T.M., Mounier, C.M., Bon, C., Dawson, P.E., Griffin, J.H. & Kent, S.B. (1997). Total chemical synthesis of enzymatically active human type II secretory phospholipase A2. *Proc. Natl Acad. Sci. USA* **94**, 7845-7850.
21. Canne, L., Botti, P., Simon, R., Chen, Y., Dennis, E. & Kent, S.B.H. (1999). Chemical protein synthesis by solid phase ligation of unprotected peptide segments. *J. Am. Chem. Soc.* **121**, 8720-8727.
22. Tam, J.P., Lu, Y.A., Liu, C.F. & Shao, J. (1995). Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proc. Natl Acad. Sci. USA* **92**, 12485-12489.
23. Erlanson, D.A., Chytil, M. & Verdine, G.L. (1996). The leucine zipper domain controls the orientation of AP-1 in the NFAT AP-1 DNA complex. *Chem. Biol.* **3**, 981-991.
24. Lu, W., Randal, M., Kossiakoff, A. & Kent, S.B. (1999). Probing intermolecular backbone H-bonding in serine proteinase–protein inhibitor complexes. *Chem. Biol.* **6**, 419-427.
25. Bode, W., Papamokos, E. & Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. Structural analysis, subtilisin structure and interface geometry. *Eur. J. Biochem.* **166**, 673-692.
26. Hipler, K., Priestle, J.P., Rahuel, J. & Grutter, M.G. (1992). X-ray crystal structure of the serine proteinase inhibitor eglin c at 1.95 Å resolution. *FEBS Lett.* **309**, 139-145.
27. McPhalen, C.A. & James, M.N. (1988). Structural comparison of two serine proteinase–protein inhibitor complexes: eglin-c–subtilisin Carlsberg and CI-2–subtilisin Novo. *Biochemistry* **27**, 6582-6598.
28. Okajima, T., Kawata, Y. & Hamaguchi, K. (1990). Chemical modification of tryptophan residues and stability changes in proteins. *Biochemistry* **29**, 9168-9175.
29. Fukunaga, Y., Katsuragi, Y., Izumi, T. & Sakiyama, F. (1982). Fluorescence characteristics of kynurenine and N′-formylkynurenine. Their use as reporters of the environment of tryptophan 62 in hen egg-white lysozyme. *J. Biochem.* **92**, 129-141.
30. Pedersen, H., Holder, S., Sutherlin, D.P., Schwitter, U., King, D.S. & Schultz, P.G. (1998). A method for directed evolution and functional cloning of enzymes. *Proc. Natl Acad. Sci. USA* **95**, 10523-10528.
31. Muir, T.W., Sondhi, D. & Cole, P.A. (1998). Expressed protein ligation: a general method for protein engineering. *Proc. Natl Acad. Sci. USA* **95**, 6705-6710.
32. Lowman, H. B. (1997). Bacteriophage display and discovery of peptide leads for drug development. *Annu. Rev. Biophy. Biomol. Struct.* **26**, 401-24.
33. Lowman, H.B. & Wells, J.A. (1991). Monovalent phage display: a method for selecting variant proteins from random libraries. *Methods: A Companion to Methods in Enzymology* **3**, 205-216.
34. Okada, Y., Tsuboi, S., Tsuda, Y., Nagamatsu, Y. & Yamamoto, J. (1990). Synthesis of a trihexacontapeptide corresponding to the sequence 8–70 of eglin c and studies on the relationship between the structure and the inhibitory activity against human leukocyte elastase, cathepsin G and α-chymotrypsin. *FEBS Lett.* **272**, 113-116.
35. Carter, P., Nilsson, B., Burnier, J.P., Burdick, D. & Wells, J.A. (1989). Engineering subtilisin BPN′ for site-specific proteolysis. *Protein Struct. Func. Genet.* **6**, 240-248.
36. Lu, W.Y., Starovasnik, M., Dwyer, J., Kossiakoff, A., Kent, S.B.H., & Lu, W. Deciphering the role of the electrostatic interactions involving gly70 in eglin c by total chemical protein synthesis. *Biochemistry*, in press.
37. Harding, J.J. (1985). Nonenzymatic covalent posttranslational modification of proteins *in vivo*. *Adv. Prot. Chem.* **37**, 247-334.
38. Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J. & Wells, J.A. (1994). A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* **266**, 243-247.
39. Wallace, C.J. (1995). Peptide ligation and semisynthesis. *Curr. Opin. Biotechnol.* **6**, 403-410.
40. Cotton, G.J. & Muir, T.W. (1999). Peptide ligation and its application to protein engineering. *Chem. Biol.* **6**, R247-R256.
41. Warshel, A. & Papazyan, A. (1998). Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* **8**, 211-217.
42. Garcia-Moreno, B., Dwyer, J.J., Gittis, A.G., Lattman, E.E., Spencer, D.S. & Stites, W.E. (1997). Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **64**, 211-224.
43. Schnolzer, M., Alewood, P., Jones, A., Alewood, D. & Kent, S.B. (1992). *In situ* neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int. J. Pept. Protein Res.* **40**, 180-193.
44. Sachev, S.S., Weiss, G., Wells, J. (2000). High copy display of large proteins on phage for functional selections. *J. Mol. Biol.* **296**, 487-495.