JMB

A Minimized M13 Coat Protein Defines the Requirements for Assembly into the Bacteriophage Particle

Tomer A. Roth, Gregory A. Weiss, Charles Eigenbrot and Sachdev S. Sidhu*

Department of Protein Engineering, Genentech Inc. 1 DNA Way, Mailstop 27 South San Francisco, CA 94080 USA

The M13 filamentous bacteriophage coat is a symmetric array of several thousand α-helical major coat proteins (P8) that surround the DNA core. P8 molecules initially reside in the host membrane and subsequently transition into their role as coat proteins during the phage assembly process. A comprehensive mutational analysis of the 50-residue P8 sequence revealed that only a small subset of the side-chains were necessary for efficient incorporation into a wild-type (wt) coat. In the three-dimensional structure of P8, these side-chains cluster into three functional epitopes: a hydrophobic epitope located near the N terminus and two epitopes (one hydrophobic and the other basic) located near the C terminus on opposite faces of the helix. The results support a model for assembly in which the incorporation of P8 is mediated by intermolecular interactions involving these functional epitopes. In this model, the N-terminal hydrophobic epitope docks with P8 molecules already assembled into the phage particle in the periplasm, and the basic epitope interacts with the acidic DNA backbone in the cytoplasm. These interactions could facilitate the transition of P8 from the membrane into the assembling phage, and the incorporation of a single P8 would be completed by the docking of additional P8 molecules with the second hydrophobic epitope at the C terminus. We constructed a minimized P8 that contained only nine non-Ala side-chains yet retained all three functional epitopes. The minimized P8 assembled into the wt coat almost as efficiently as wt P8, thus defining the minimum requirements for protein incorporation into the filamentous phage coat. The results suggest possible mechanisms of natural viral evolution and establish guidelines for the artificial evolution of improved coat proteins for phage display technology.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: phage display; protein engineering; combinatorial mutagenesis; M13 bacteriophage

*Corresponding author

Introduction

M13 bacteriophage is a filamentous, *Escherichia coli*-specific virus approximately 1 μm in length but less than 10 nm in diameter. The viral filament consists of a circular, single-stranded DNA (ssDNA) core surrounded by a protein coat.[1] The ends of the particle are capped by about five copies each of four minor coat proteins (two different proteins at either end), while the filament length is covered by approximately 2700 copies of the gene-8 major coat protein (protein-8, P8). Filamentous phage have proven useful as model systems for the study of membrane-associated
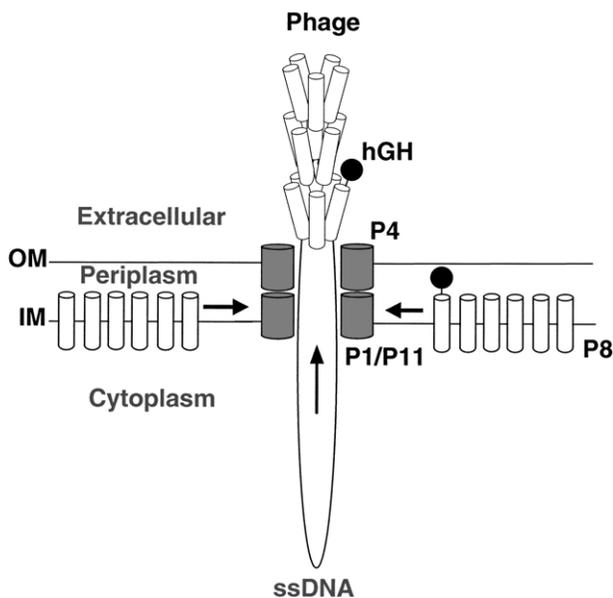
Figure 1. Filamentous bacteriophage assembly and phage display. Newly synthesized major coat protein molecules (P8, white cylinders) are imbedded in the IM with their N termini in the periplasm and their C termini in the cytoplasm. Single-stranded viral DNA (ssDNA) is extruded from the host at assembly sites (grey cylinders) composed of phage-encoded proteins (P1/P11 and P4) that together span the IM and the OM. Coat proteins interact with the assembly site, where they surround the ssDNA and are thus transferred from the bacterial membrane into the assembling phage particle, which is extruded without lysis of the host. Phage display of hGH (filled circles) was achieved by using an hGH-P8 fusion gene carried on a phagemid. *E. coli* harboring the phagemid were co-infected with a helper phage containing all the genes necessary for phage production, including a wt P8 gene. The resulting phage particles contained phagemid DNA surrounded by a coat that was composed mainly of wt P8 but contained a small amount of the phagemid-encoded hGH-P8 fusion. Only the elongation phase of assembly is depicted; the minor coat proteins involved in assembly initiation and termination are not shown. For clarity, the P5 molecules that bind to the ssDNA prior to assembly are omitted.

filament assembly and secretion.[2–4] More recently, the study of M13 assembly and structure has also enabled improvements in phage display technology.[5,6]

Filamentous phage assembly is a non-lytic, membrane-associated process (Figure 1);[7] phage particles are extruded from the infected host, which continues to grow and divide, albeit at a reduced rate. Prior to assembly, newly synthesized coat proteins are anchored in the inner membrane (IM) by single transmembrane domains. The phage genome encodes for proteins that are not found in the phage particle but are necessary for phage assembly and export. Multimers of protein-4 (P4) in the outer membrane (OM) interact with multimers of protein-1 and protein-11 (P1 and P11) in the IM[8] to form an assembly site that may

function as a gated pore through which assembling phage are extruded.[2,9] In the cytoplasm, another phage-encoded protein (protein-5, P5) binds to newly synthesized, viral ssDNA. The P5–ssDNA complex interacts with P1/P11 at the assembly site, and thus initiates the phage assembly process. At the assembly site, P5 is stripped from the ssDNA and replaced by coat proteins. Assembly is initiated by addition of the minor coat proteins P7 and P9, and the first few copies of P8.[10] In the elongation phase, the ssDNA is extruded from the host and concomitantly surrounded by P8 molecules. When the end of the ssDNA is reached, the process is terminated by the addition of the minor coat proteins P6 and P3, and the assembled phage particle is released into the extracellular milieu.[11]

The assembly and structure of filamentous phage is dominated by protein–protein and protein–DNA interactions involving P8, since this protein is by far the major component of the viral coat. Fibre X-ray diffraction studies of the bacteriophage coat have revealed that P8 is α-helical; the P8 moities surround the ssDNA in an helical array with a 5-fold rotation axis and a 2-fold screw axis with a 16 Å pitch.[1,12,13] Thus, the coat consists of interlocking layers, with each layer containing five symmetrically identical P8 moities. Each P8 molecule makes extensive contacts with P8s in the three layers above and three layers below, but only minor contacts with P8s within its own layer.[14] While X-ray diffraction analyses have revealed the intermolecular contacts that define the major structure of the phage coat, they do not address the energetic contributions of these contacts to phage assembly and stability. Studies of protein–protein interfaces have shown that, in most cases, only a subset of the side-chains that make intermolecular contacts make energetic contributions to binding and, often, these side-chains cluster together in "functional epitopes".[15,16] Accurate mapping of functional epitopes requires a three-dimensional structure, and systematic site-directed mutagenesis combined with functional assays to assess the contributions of individual side-chains to protein function.[17,18]

We wanted to determine the functional epitopes that are necessary for the stable incorporation of P8 into the M13 phage coat, but this goal was complicated by the fact that P8 must perform multiple functions during its multi-stage lifecycle,[7] and each of these functions depends on particular P8 side-chains. In stage 1, nascent P8 inserts into the IM in a process that is independent of host proteins, but instead, depends on an N-terminal pro-peptide and on certain residues within mature P8.[4] Following membrane insertion, the pro-peptide is removed proteolytically by a host protease, leaving the 50-residue mature P8 anchored in the membrane. In stage 2, P8 interacts with the assembly site and transitions from the membrane into the viral coat (Figure 1). In the third and final

stage, P8 molecules interact with each other and with the viral DNA to form a structurally stable, viable phage particle.[1]

We established a system in which the contributions of P8 side-chains to phage assembly could be studied independently of the contributions to membrane insertion or phage viability. For this purpose, we used a phagemid vector designed for the heterologous display of human growth hormone (hGH) on the surface of M13 phage.[19] The phagemid contained an open reading frame that produced an hGH-P8 fusion protein with an N-terminal secretion signal that targeted the polypeptide to the bacterial secretion pathway and, thus, made the membrane insertion process independent of the P8 sequence. Furthermore, the phagemid display system resulted in the production of hybrid phage particles with coats containing predominantly wild-type (wt) P8 moieties from a helper phage and only a few hGH-P8 fusions. In this system, the level of hGH display could be used as an indicator of the incorporation efficiency of the fused P8 into a phage coat composed of wt P8 (Figure 1).

To assess the contributions of individual P8 side-chains to the assembly process, we measured the effects of truncating mutations that selectively removed wt side-chains. We used Ala mutations because this substitution removes all side-chain atoms past the β-carbon atom, but is not likely to cause major conformational changes.[18] At positions where Ala was the wt residue, we used Gly substitutions to remove even the β-carbon atoms. A comprehensive analysis of the entire mature P8 was conducted using shotgun scanning mutagenesis, a rapid combinatorial strategy.[20] Libraries of P8 mutants were constructed using degenerate codons designed to preferentially encode the wt or the truncating mutation. Library pools were subjected to binding selections to enrich for clones displaying heterologously fused hGH, and selected clones were sequenced to determine the wt/mutant ratio at each varied position. This ratio provided a quantitative assessment of the contribution of each side-chain to the process of P8 incorporation into the phage coat.

The P8 shotgun scan revealed that only a small subset of side-chains are necessary for the stable incorporation of P8 into the wt phage coat. In fact, we engineered an extremely minimized version of P8 that contained only nine non-Ala side-chains and yet assembled into the wt coat with an efficiency almost equivalent to that of wt P8. When mapped onto the three-dimensional structure of filamentous phage, our mutagenesis results reveal that the viral assembly process depends on the recognition of only a few key P8 side-chains that cluster in distinct functional epitopes. Our findings provide insights into the processes of natural viral evolution and establish guidelines for the artificial evolution of novel coat proteins with improved performance in phage display technology.

Results

Shotgun scan of P8

To conduct a comprehensive mutational analysis of P8, we constructed three shotgun scanning libraries within a phagemid vector designed to display hGH fused to the N terminus of P8.[19] Each library covered a contiguous region of P8 sequence but did not overlap with the other libraries; together, the three libraries spanned the entire 50-residue primary sequence (see Materials and Methods). The shotgun scanning code[20] was used to replace wt codons with degenerate codons that ideally encoded only the wt amino acid residue or Ala (m1 in Table 1), although the nature of the genetic code necessitated two other substitutions at some positions (m2 and m3 in Table 1). In positions where Ala was the wt residue, we used Gly as the truncating substitution. The libraries were constructed using mutagenic oligonucleotides that had been synthesized with equimolar nucleotide mixtures at degenerate positions, thus providing equal representation of each designed amino acid residue at each mutated position. The number of unique members in each library was at least 100-fold greater than the number of possible amino acid combinations encoded by the mutagenic oligonucleotide used for library construction, thus ensuring complete representation of the theoretical diversity.

Phage pools from each library were subjected separately to selections for binding to immobilized growth hormone binding protein (hGHbp), a receptor that binds to hGH with high affinity (K_d 1.6 nM).[21] These selections enriched for clones displaying hGH, and thus selected for P8 variants that incorporated into the phage coat efficiently. Approximately 100 hGH-displaying clones were sequenced from each library selection, the sequences were aligned, and the occurrence of wt or each designed substitution was tabulated at each mutated position (see Materials and Methods). These data were used to calculate the wt/mutant ratio for each mutation at each position (Table 1). The wt/mutant ratio is the statistical preference for the wt relative to the mutant, and it correlates with the effect of each mutation on hGH display, and by extension, with the effect of each mutation on P8 incorporation. Ratios greater than or less than 1 indicate deleterious or beneficial mutations, respectively.

While Ala residues contain only a β-carbon side-chain, the side-chains of five out of ten wt alanine residues in P8 are important for coat incorporation, as evidenced by conservation of the wt sequence relative to Gly (wt/m1 > 5 in Table 1). Aside from these Ala residues, only ten other side-chains appear important for incorporation (wt/m1 > 5) and, with the exception of Tyr24, these residues are located in two distinct groups at either end of the sequence (Figure 2). The N-terminal group of functionally important side-chains consists of two

Table 1. P8 shotgun scan

Residue[b]	Wt/mutant ratios[a]		
	wt/m1	wt/m2	wt/m3
A1	0.91		
E2	0.76		
G3	1.9		
D4	1.3		
D5	2.5		
P6	0.85		
A7	**7.1**		
K8	1.1	1.1(E)	**58(T)**
A9	**6.0**		
A10	**56**		
F11	**> 170**	**170(S)**	**> 170(V)**
N12	0.82	1.1(D)	1.9(T)
S13	0.28		
L14	**150**	**150(P)**	**8.3(V)**
Q15	0.40	1.3(E)	**20(P)**
A16	1.7		
S17	0.25		
A18	**6.1**		
T19	0.64		
E20	2.9		
Y21	1.5	**> 31(D)**	**31(S)**
I22	0.46	>11(T)	0.52(V)
G23	3.4		
Y24	**7.0**	**> 49(D)**	**> 49(S)**
A25	**18**		
W26	1.5	**10(G)**	**30(S)**
A27	0.55		
M28	1.1	1.6(T)	1.8(V)
V29	0.26		
V30	1.9		
V31	0.71		
I32	0.27	0.67(T)	0.19(V)
V33	0.48		
G34	1.6		
A35	4.6		
T36	1.2		
I37	1.0	0.64(T)	2.3(V)
G38	0.83		
I39	**100**	**> 103(T)**	0.62(V)
K40	**54**	**> 163(E)**	**82(T)**
L41	**6.8**	**136(P)**	**11(V)**
F42	**13**	**> 107(S)**	2.0(V)
K43	**81**	**54(E)**	**81(T)**
K44	**20**	**156(E)**	**31(T)**
F45	**80**	**160(S)**	**27(V)**
T46	1.4		
S47	4.6		
K48	0.84	1.8(E)	**7.4(T)**
A49	3.5		
S50	5.0		

The wt/mutant ratio provides an estimate of the effect of each mutation on the incorporation of P8 into the phage coat. Deleterious effects are indicated by wt/mutant ratios greater than 1, and mutations that have large deleterious effects (wt/mutant > 5) are shown in bold text. In cases where a particular mutation was not observed, only a lower limit could be defined for the wt/mutant ratio (indicated by >).
  [a] The mutation m1 was an Ala substitution in all cases, except in the case of wt Ala, where a Gly substitution was used. The identities of the mutations m2 and m3 depend on the particular shotgun codon used and are shown in parantheses to the right of each wt/mutant ratio.
  [b] Each wt residue is represented by the single-letter amino acid code followed by the position in the mature P8 sequence.

large hydrophobic residues (Phe11 and Leu14) and four alanine residues (Ala7, Ala9, Ala10, and Ala18), while the C-terminal group is a continuous stretch of seven residues consisting of four hydro-
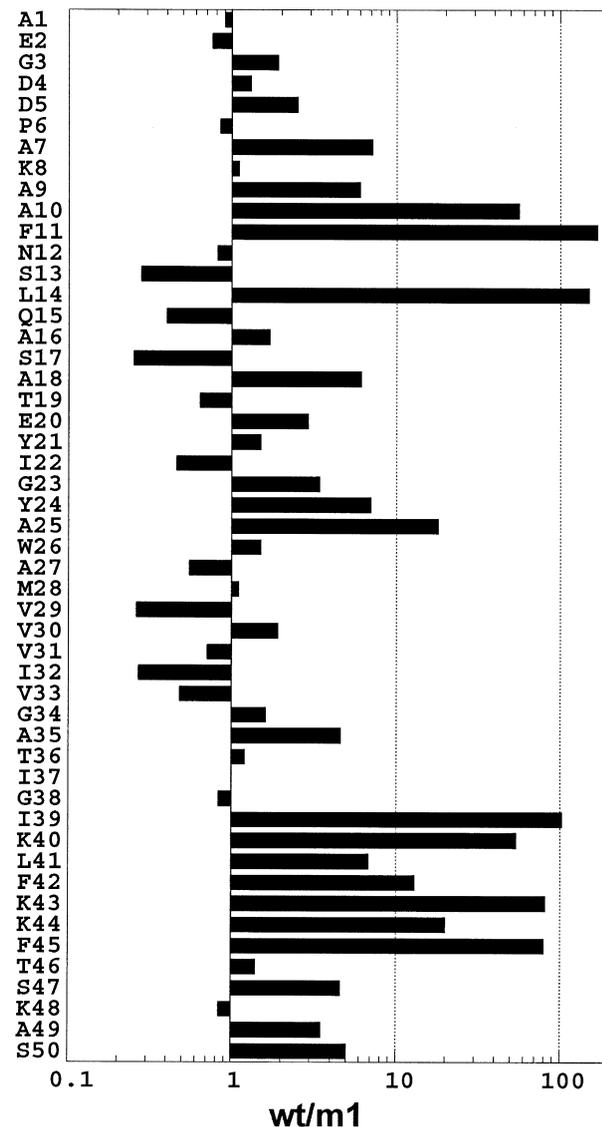
Figure 2. The wt/mutant ratios measuring the effects of P8 mutations on the efficiency of incorporation into the wt phage coat. For each P8 residue (vertical axis), the wt/m1 ratio from Table 1 is plotted (horizontal axis). The m1 substitution was Ala, except when the wt residue itself was Ala, in which case a Gly substitution was used.

phobes (Ile39, Leu41, Phe42, and Phe45) and three lysine residues (Lys40, Lys43, and Lys44). Several residues exhibit wt/mutant ratios of less than 1, indicating that Ala substitutions at these positions may actually increase incorporation efficiency and, while the individual ratios favor Ala only slightly, it is noteworthy that these residues also form two distinct groups. One group consists of hydrophilic residues (Asn12, Ser13, Gln15, Ser17, and Thr19) in the same region as the N-terminal cluster of functionally important hydrophobic residues, while the second group (Ile22, Val29, Val31, Ile32, and Val33) consists of β-branched hydrophobes near the center of the protein sequence.

(a)

```
                 *  ***   *    *      **        *******
      P8    AEGDDPAKAAFNSLQASATEYIGYAWAMVVVIVGATIGIKLFKKFTSKAS
 Mini-P8    AAGAAAAAAAFAALAAAAAAAAGAAAAAAAAAAGAAAGIKLFKKFAAAAA
            +-+---+-++  --  -+-+----+-+-+------++--+      ---+-
```

(b)

Figure 3. A minimized version of P8 (mini-P8) incorporates efficiently into the wt phage coat. (a) The 50-residue mini-P8 contains only nine non-Ala side-chains. Ala/Gly residues that are present in both wt and mini-P8 are underscored with + symbols, while Ala mutations in mini-P8 are unders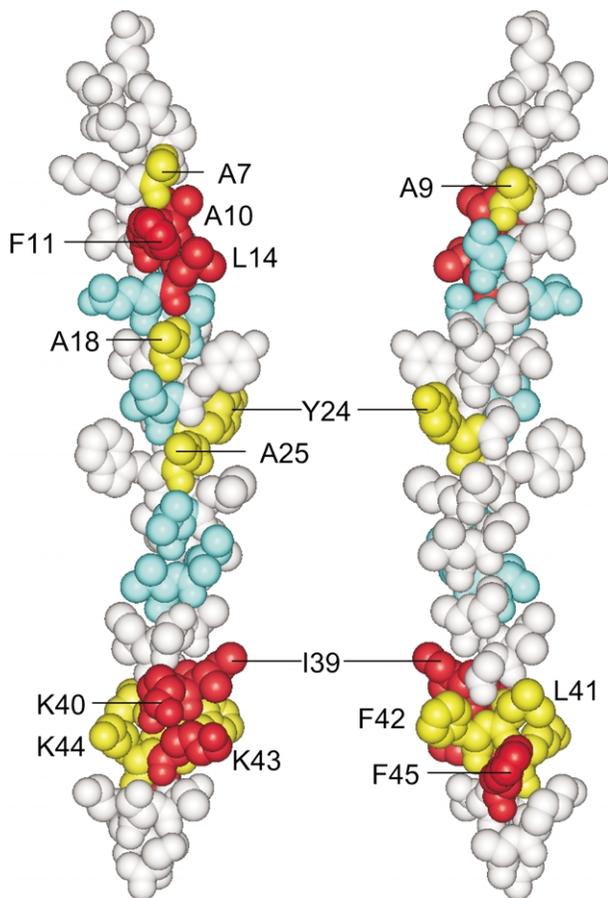cored with dashes. Mini-P8 retains identity with P8 at all of the positions (except Y24) that are required for efficient incorporation into the phage coat (wt/mutant > 5, Table 1; indicated by asterisks). (b) The incorporation efficiency of mini-P8 relative to wt P8 was determined using a phage ELISA that measured the display of hGH fused to the N terminus of either P8 (circles) or mini-P8 (squares). The amount of bound phage (*y*-axis) was directly proportional to the phage concentration (*x*-axis, log scale). For each set of data, the least-squares linear fit is shown, with the corresponding equation and Pearson's correlation coefficient (*R*) given at the top (see Results for further details).

Minimized P8

The P8 shotgun scan results suggested that, aside from a few residues located at either end of the protein, most side-chains could be replaced with Ala without affecting assembly into the wt phage coat significantly and, furthermore, some Ala mutations might actually increase incorporation efficiency. On the basis of these data, we constructed a minimized version of P8 (mini-P8) by simultaneously substituting with Ala for 27 side-chains predicted to be minor contributors to the assembly process. While our data indicate that the four Gly residues in P8 can be replaced by Ala

without significantly affecting incorporation, we chose not to mutate these residues, since our intention was to minimize the size of P8, and Gly is the one amino acid that is actually smaller than Ala. Mini-P8 is a 50-residue protein consisting of four Gly, 37 Ala, and only nine larger residues, yet it retains essentially all of the side-chains required for coat incorporation (Figure 3(a)). Assuming additivity for the mutational effects, we predicted that mini-P8 should assemble into the wt phage coat with essentially the same efficiency as wt P8 itself.

We used a phage ELISA to quantify the efficiency of hGH fusion display,[19] and with these data, we determined the incorporation efficiency of mini-P8 relative to that of wt P8. In this assay, serial dilutions of phage were incubated in wells containing hGHbp to capture hGH-displaying phage. Following washing, specifically bound phage could be detected spectrophotometrically (see Materials and Methods). Signal strengths depend on both phage concentration and the amount of hGH displayed on the phage particles. At subsaturating concentrations of phage, the signal strengths were found to be directly proportional to phage concentration, as evidenced by an almost perfect correlation in a least-squares linear fit (Figure 3(b)). For each phage sample, the slope of the line provided a measure of the increase in ELISA signal per unit of phage concentration, and by dividing the slope for wt P8 by the slope for mini-P8, we determined that the amount of phage required to generate an equivalent signal was only 14-fold greater when hGH was fused to mini-P8 rather than to wt P8. Thus, on the basis of the relative efficiency of hGH display, we estimated that mini-P8 incorporated into the wt phage coat only 14-fold less efficiently than wt P8.

Discussion

The P8 side-chains required for assembly constitute a subset of those required for phage viability

While our selections required only that P8 mutants retain the ability to incorporate into a wt coat supplied by a helper phage, Williams *et al.* have mutagenized the single P8 gene within the M13 genome and investigated the variations in P8 sequence that are compatible with phage viability.[22] Williams *et al.* observed conservation of the wt sequence at all of the positions that we found to be important for coat incorporation. In particular, they observed no mutation in the C-terminal stretch from Ile39 to Phe45, while in the N-terminal region, Ala10, Phe11 and Leu14 were conserved completely and Ala7, Ala9 and Ala18 exhibited only limited divergence from the wt. Furthermore, Hunter *et al.* have shown that the three lysine residues conserved in our assembly-competent P8 variants (Lys40, Lys43,

Figure 4. Mapping of the functional epitopes for coat incorporation onto the structure of P8. While the C-terminal region of the α-helical P8 is buried completely in the phage coat, only one face of the N-terminal portion is buried. The left panel shows the buried face, while in the right panel, rotation by 180° reveals the exposed face. The residues are color-coded according to the magnitudes of the wt/m1 ratios (Table 1), as follows: red, >25; yellow >5; blue <0.5. Thus, the red and yellow side-chains are predicted to be important for incorporation of P8 into the phage coat. The blue side-chains may actually hinder incorporation, since mutations are favored over wt at these positions (at least twofold). The grey side-chains are predicted to neither significantly enhance nor hinder incorporation efficiency (5 > wt/mutant > 0.5). The functionally important side-chains cluster into three distinct functional epitopes: epitope 1 (Ala7, Ala10, Phe11, Leu14, and Ala18), epitope 2 (Ile39, Leu41, Phe42, and Phe45), and epitope 3 (Lys40, Lys43, and Lys44). The wt P8 of the Ff phage coat (PDB entry 1IFJ)[13] is shown in CPK representation. M13 P8 differs from Ff P8 in only one position (Asn12 in M13 P8 is an Asp in Ff P8). The Figures were generated using Insight II (Accelrys, San Diego).

and Lys44) were essential for phage viability.[23] Interestingly, while Lys48 is located close to these essential lysine residues, this side-chain is dispensable for both assembly (Figure 2) and viability.[23] The more stringent restraints imposed by the requirements of phage viability resulted in the conservation of many residues that are not required

for coat incorporation, including several large hydrophobes (Tyr21, Ile22, Trp26, and Met28).[22] Some of these additional conserved residues may be involved in the membrane insertion of nascent wt P8. Alternatively, or in addition, they may be required to maintain the structural stability of the assembled phage particle, and this is particularly likely for large hydrophobes that make extensive intermolecular contacts, and for small side-chains that permit close helix–helix contacts, as described.[22]

Overall, the restrictions on P8 sequence imposed by the requirements of phage viability are considerably more strict than those imposed by the requirements of assembly alone. To our knowledge, reports of P8 mutants that can support phage viability independently have been limited to single, double, or triple mutations,[22,24] although it has been noted that single Ala mutants are viable at 23 out of 36 N-terminal positions.[22] In contrast, the mini-P8 reported here contains a total of 27 Ala mutations that have only a negligible effect on incorporation efficiency (Figure 3). In a previous study, we engineered a P8 variant that contained mutations at 25 out of 30 N-terminal positions, yet actually incorporated into a wt coat more efficiently than wt P8, thus demonstrating that a P8 variant optimized for coat incorporation could tolerate an extremely large number of mutations.[25] The previous study supports the findings presented here, because, despite the extreme divergence from the wt sequence, the engineered P8 variant retained the side-chains necessary for efficient assembly; Ala7, Ala10, Leu14, and Ala18 were among the five N-terminal residues conserved as wt, Phe11 was mutated conservatively to Tyr, and the 20 C-terminal residues were retained as wt.

The functional epitopes for P8 assembly

We mapped our mutagenesis results onto the structure of P8,[13] thus revealing three functional epitopes required for the assembly of P8 into the wt phage coat (Figure 4). Epitope 1 consists of N-terminal side-chains (Ala7, Ala10, Phe11, Leu14, and Ala18) that form a hydrophobic patch that is buried against other P8 molecules in the phage coat (see below). Near the C terminus, four side-chains (Ile39, Leu41, Phe42, and Phe 45) form a second hydrophobic patch (epitope 2) that is buried completely within the phage coat. Epitope 3 is located directly opposite to epitope 2 on the other side of the α-helix, and it consists of three positively-charged lysine side-chains (Lys40, Lys43, and Lys44) that interact with the negatively charged DNA in the phage particle core. Minor contributions to assembly are provided by a small patch (Tyr24 and Ala25) at the center of the α-helix. It is notable that all the functionally important residues (except Ala9) are buried in the phage coat. Furthermore, the hydrophilic residues that surround epitope 1 can be substituted readily

with Ala (Figures 2 and 4), suggesting that decreased hydrophilicity around epitope 1 may be advantageous for incorporation. Indeed, a previous study demonstrated that hydrophobic substitutions in this region increased the display of N-terminal fusions.[19,25]

The functional epitopes for P8 assembly are arranged in a dumbell fashion at either end of a central spacer region, as exemplified by the central poly(Ala/Gly) stretch in mini-P8 (Figure 3(a)). This central stretch contains the entire trans-membrane domain of wt P8, and it is surprising that all the residues within this hydrophobic stretch can be replaced simultaneously with alanine without affecting incorporation efficiency significantly (Figure 3). While the subcellular localization of mini-P8 remains to be investigated, we assume that it is still an integral membrane protein similar to wt P8, since the assembly process is associated intimately with the membrane and perturbations in subcellular localization would presumably have major effects on the incorporation efficiency. This supposition is further supported by the observation that hydrophilic substitutions are disfavored strongly relative to the wt or Ala at positions 21, 22, 24, and 26 (see wt/m2 and wt/m3 in Table 1), suggesting that a general hydrophobic character is still required in this transmembrane region. It has been noted that the Phe and Lys side-chains that are essential components of the C-terminal functional epitopes for P8 assembly also serve as membrane anchors.[14] Our results suggest that such membrane anchors may be sufficient to allow even a poly(Ala/Gly) stretch to function effectively as a transmembrane domain and, while hydrophilic side-chains are disfavored specifically, large hydrophobic side-chains are not necessarily required in a transmembrane domain that is followed by a strong membrane anchor sequence.

The efficient assembly of mini-P8 suggests that the transmembrane domain does not contribute functionally important side-chain interactions during phage assembly. Instead, it serves to present the functional epitopes in the correct subcellular locations for recognition by the assembly machinery; specifically, the transmembrane domain acts as a connector between epitope 1 in the periplasm and epitopes 2 and 3 in the cytoplasm. Thus, while filamentous phage assembly is an obligatory membrane-associated process, the key molecular recognition events for P8 assembly involve epitopes located in the aqueous periplasm and cytoplasm.

Protein–protein interactions in phage assembly

Fibre X-ray diffraction studies have provided significant insights into the protein–protein interactions that contribute to the structural stability of filamentous phage particles.[1,13] These studies, together with NMR studies of P8 in detergent micelles,[14] have led to the development of models

for phage assembly. Our mutagenesis data and the simplified structure of mini-P8 support the assembly model proposed by Papavoine *et al*.[14] and further illuminate the protein–protein interactions required for phage assembly, independent of the requirements for phage structural stability.

The filamentous phage structure reveals that each P8 makes intermolecular contacts with ten other P8 molecules within the coat array, with contact defined as at least one distance between heavy atoms being less than 5 Å. Each P8 makes 172 intermolecular contacts that involve 44 out of 50 P8 residues.[14] Phage assembly depends on the subset of these interactions that involve the functional epitopes delineated by our mutagenesis data (Figure 4). While the positively charged epitope 3 interacts mainly with the negatively charged ssDNA core, the hydrophobic epitopes 1 and 2 are involved in extensive protein–protein interactions with neighboring P8 molecules.

Epitope 1 (excluding Ala18, a minor contributor on the periphery of the epitope) of an arbitrary "P8(0)" within the phage coat contacts nine residues of three neighboring P8 molecules in the three layers above P8(0). Epitope 2 of P8(0) contacts 15 residues of three neighboring P8 molecules in the three layers below P8(0) (Figure 5). These interactions bury epitopes 1 and 2 in hydrophobic clusters involving several aromatic residues. While it has been noted that these aromatic clusters are important for phage stability,[14,26] our results show that they are crucial for phage assembly. Due to the symmetry of the phage particle, the cluster involving the N terminus of P8(0) is identical with that involving the C terminus (Figure 5(b)); in comparing the C-terminal cluster to the N-terminal cluster, the C terminus of P8(0) adapts the position of P8(17) and the N terminus of P8(−17) adapts the position of P8(0). These clusters involve contacts between epitopes 1 and 2 from different P8 molecules, and this results in the intercalation of Phe11 between Phe42 and Phe45 (Figure 5(c)), an interaction that likely contributes significant binding energy to both phage assembly and stability. In the case of P8(0), eptitope 1 interacts with epitope 2 of P(17), and epitope 2 interacts with epitope 1 of P(−17).

When considered in the context of the membrane-associated assembly process, our results further simplify the model proposed by Papavoine *et al*.[14] The initial interaction of membrane-bound P8(0) with the assembling phage likely involves two events: the periplasmic interaction of epitope 1 with P8 molecules already assembled into the particle, and the cytoplasmic interaction of epitope 3 with the ssDNA. The interaction between epitope 1 of P8(0) and epitope 2 of P8(17) would complete the assembly of P8(17) and, at the same time, would establish interactions between P8(0) and the assembling coat. Together, these interactions tether P8(0) to both the protein and DNA components of the phage particle, and this may be sufficient to induce transition of P8(0) from the
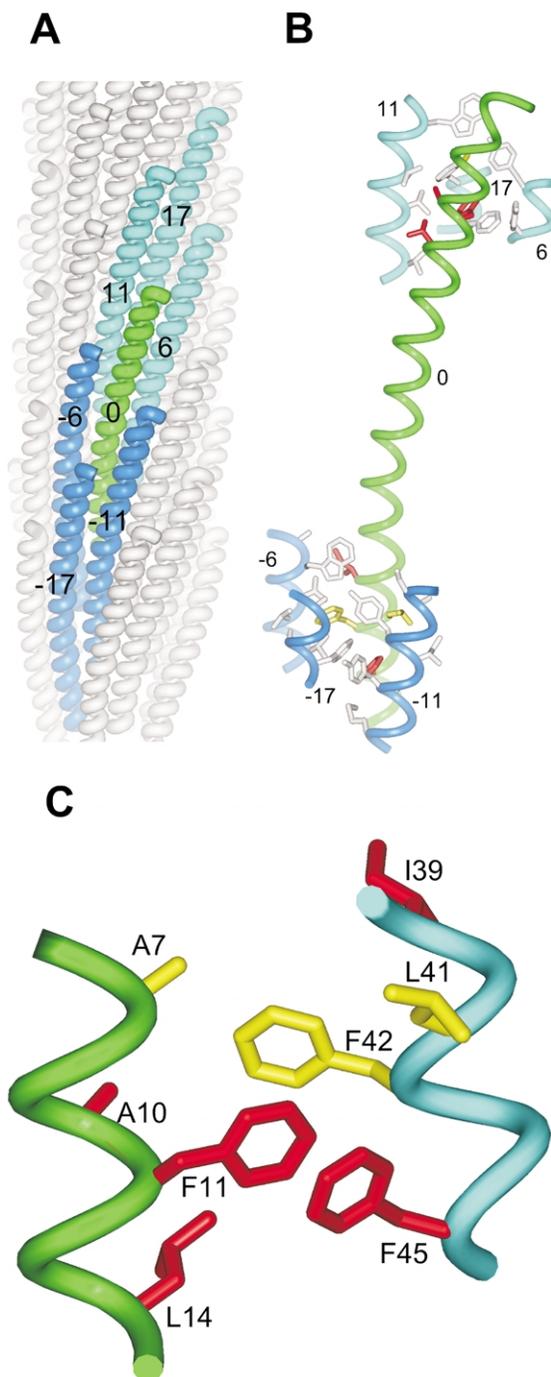
Figure 5. Protein–protein interactions involved in the incorporation of P8 into the wt coat. (a) The filamentous phage coat is a repeating array of P8 molecules,[1,13] shown as ribbons and labeled according to the numbering scheme used by Papavoine *et al.*[14] A single P8 molecule, P8(0), is shown in green. P8 molecules that make contact with epitope 1 or epitope 2 of P8(0) (see Figure 4) are shown in light blue or dark blue, respectively. (b) Intermolecular side-chain interactions involving epitopes 1 or 2 of P8(0). The orientation, numbering and backbone coloring are as in (a). Side-chains in the P8(0) functional epitopes are color-coded as in Figure 4, while side-chains in neighboring P8 molecules are colored grey. Epitope 1 (excluding Ala18) is surrounded by side-chains from P8(6) (Tyr21 and Tyr24), P8(11) (Trp26, Val30, Val33 and Ile37), and P8(17) (Phe42 and Phe45). Epitope 2 is surrounded by

membrane into the assembling phage. During the transition phase, the remaining neighbor P8s could assemble around P8(0), culminating with the docking of epitope 1 of P8(−17) with epitope 2 of P8(0), an event that would complete the assembly of P8(0) and initiate the assembly of P8(−17). Thus, the assembly process is dependent on interactions involving epitopes at either end of the P8 molecule, and the protein–protein interactions involving the N and C-terminal epitopes likely contribute to the initiation and completion of P8 incorporation, respectively. Intermolecular interactions with the central region of P8 contribute to phage viability,[22] but they do not play a significant role in the dynamics of phage assembly. Finally, it should be noted that genetic evidence suggests that P8 interacts with P1 in the assembly site,[8] and it is likely that the functional epitopes for this interaction are contained within the minimized sequence of mini-P8.

Implications for natural and artificial viral evolution

In summary, we have demonstrated that the assembly of proteins into the filamentous phage coat requires the recognition of only a few key side-chains displayed on either end of a trans-membrane domain. With such lax requirement for incorporation, it seems likely that host membrane proteins could be readily co-opted as new viral coat proteins, especially since many natural membrane proteins possess positively charged cytoplasmic regions similar to epitope 3 of P8. The facile acquisition of host-derived coat proteins could, in turn, facilitate the rapid evolution of new viral functions, and it is tempting to speculate that at least some of the minor coat proteins may have arisen through such a process. It would be interesting to map and compare the functional epitopes required for the assembly of other bacteriophage and eukaryotic viruses.

Our results support and explain artificial evolution experiments that have demonstrated the promiscuous nature of filamentous phage assembly[5] and enabled the development of novel phage display scaffolds in the form of highly divergent P8 variants[19,25] or even completely artificial

side-chains from P8(−6) (Ala25, Trp26, Val29, Val30 and Val33), P8(−11) (Ser17, Ala18, Tyr21, Ile22, Tyr24, Ala25 and Met28), and P8(−17) (Ala7, Ala10 and Phe11). (c) Intermolecular contacts between epitope 1 of P8(0) (green ribbon) and epitope 2 of P8(17) (blue ribbon). The structure is presented as viewed from the right of (a) and (b). The side-chains are labeled and color-coded as in Figure 4. A symmetrically related interaction between epititope-1 of P8(−17) and eptitope 2 of P8(0) can be seen at the bottom of (b). The Figures were generated using Insight II with coordinates from PDB entry 1IFJ.[13]

coat proteins.[27] The mini-P8 sequence represents a minimal blueprint for coat incorporation that can be used to further improve phage display technology, and, to further investigate the structure and function of filamentous phage.

Materials and Methods

Materials

Enzymes were from New England Biolabs. Maxisorp immunoplates were from NUNC (Roskilde, Denmark). *E. coli* XL1-Blue and M13-VCS were from Stratagene. Bovine serum albumin (BSA) and Tween 20 were from Sigma. Horse radish peroxidase/anti-M13 antibody conjugate was from Pharmacia Biotech. 3,3′,5,5′-Tetra-methyl-benzidine/H_2O_2 (TMB) peroxidase was from Kirkegaard & Perry Laboratories Inc.

Oligonucleotides

Equimolar DNA degeneracies are represented in the IUB code (K = G/T, M = A/C, R = A/G, S = G/C, Y = C/T). Degenerate codons are shown in bold text. The following mutagenic oligonucleotides were used:

g8sg1: TCCGGGAGCTCCAGC**GSTGMAGSTGM-TGMTSCAGSTRMAGSTGSTKYTRMCKCCSYTS-MAGSTKCCGSTRCT**GAATATATCGGTTATGCGTGG

g8sg2: CTGCAAGCCTCAGCGACC**GMAKMTRY-TGSTKMTGSTKSGGSTRYTGYTGYTGYTRYTGY-TGSTGSTRCT**ATCGGTATCAAGCTGTTT

g8sg3: ATTGTCGGCGCAACT**RYTGSTRYTRMAS-YTKYTRMARMAKYTRCTKCCRMAGSTKCC**TGA-TAAACCGATACAATT

g8A1: GGGAGCTCCAGCGCCGCTGGTGCAGCT-GCAGCAGCTGCGGCCTTTGCAGCTCTGGCAGCC-GCTGCGGCAGCTGCAGCTGGTGCAGCGTGGGCG-ATGGTT

g8A2: GCTGGTGCAGCGGCTGCGGCAGCTGCA-GCTGCAGCTGGCGCAGCAGCTGGTATCAAGCTG-TTTAAGAAATTCGCAGCTGCAGCAGCTTGATAA-ACCGATACA

Construction of P8 shotgun scanning libraries

Libraries were constructed as described[28] with two "stop template" versions of a previously described phagemid (pS1607) designed for the display of hGH fused to the N terminus of P8.[19] The stop templates contained TAA stop codons in place of P8 codons 19, 20, and 21 (pR212a), or in place of codons 44 and 45 (pR212b). For library construction, either pR212a or pR212b was used as the template for the Kunkel muta-genesis method,[29] with a mutagenic oligonucleotide designed to simultaneously repair the stop codons and introduce mutations at the desired sites. Three separate libraries were constructed, with each library designed to mutate a continuous stretch of P8 sequence. There was no overlap between the mutated regions of each library and together they scanned the entire length of P8. Libraries 1 and 2 were constructed using template pR212a, while the construction of library 3 used pR212b. Libraries 1, 2, or 3 were constructed using oligo-

nucleotide *g8sg1*, *g8sg2*, or *g8sg3* and mutated residues 1–19, 20–36, or 37–50, respectively. The library diversities were as follows: library 1, 1.2×10^{10}; library 2, 2.9×10^9; and library 3, 7.5×10^9. In each library, the number of unique members exceeded the theoretical diversity (the number of possible codon combinations encoded by the mutagenic oligonucleotide) by at least 100-fold.

Library sorting and binding assays

Maxisorp immunoplates (96-well) were coated over-night at 4 °C with hGHbp (5 μg ml^{-1}) and blocked for two hours with BSA. Phage from the libraries described above were propagated in *E. coli* XL1-Blue with the addition of M13-VCS helper phage. The hGH-P8 fusion gene of pS1607 is IPTG-inducible, but the cultures were grown under non-inducing conditions to ensure mono-valent display. After growth overnight at 37 °C, phage were concentrated by precipitation with PEG/NaCl and resuspended in PBS, 0.5% BSA, 0.1% Tween 20, as described.[28] Phage solutions (10^{12} phage ml^{-1}) were added to the coated immunoplates to capture hGH-displaying phage by binding to hGHbp. Following two hours incubation, the plates were washed 12 times with PBS, 0.05% Tween 20. Bound phage were eluted with 0.1 M HCl for ten minutes and the eluant was neutralized with 1.0 M Tris base. Eluted phage were amplified in *E. coli* XL1-Blue and used for further rounds of selection.

Individual clones from each round of selection were grown in a 96-well format in 500 μl of 2YT broth supplemented with carbenicillin and M13-VCS, and the culture supernatants were used directly in phage ELISAs with hGHbp-coated plates to detect clones displaying hGH. After two (libraries 1 and 2) or three (library 3) rounds of selection, greater than 50% of the clones exhibited positive phage ELISA signals at least twofold greater than signals on control plates coated with BSA. These positive clones were subjected to DNA sequence analysis.

DNA sequencing and analysis

Culture supernatant containing phage particles was used as the template for a PCR that amplified the P8 gene. The PCR primers were designed to add M13(-21) and M13R universal sequencing primers at either end of the amplified fragment, thus facilitating the use of these primers in sequencing reactions. Amplified DNA fragments were sequenced using Big-Dye terminator sequencing reactions, which were analyzed on an ABI Prism 3700 96-capillary DNA analyzer (PE Biosystems, Foster City, CA). All reactions were performed in a 96-well format.

The sequences were analyzed with the program SGCOUNT as described.[20] SGCOUNT aligned each DNA sequence against the wt DNA sequence by using a Needleman–Wunch pairwise alignment algorithm, translated each aligned sequence of acceptable quality, and tabulated the occurrence of each natural amino acid at each position. The following numbers of clones were analyzed: library 1, 170; library 2, 95; and library 3, 170.

Construction of the gene encoding mini-P8

Two rounds of Kunkel mutagenesis were used to convert the P8 gene in pS1607 to a gene encoding

mini-P8 (Figure 3(a)). In the first round, oligonucleotide *g8A1* was used to introduce Ala substitutions in the N-terminal half of P8, resulting in a gene encoding the sequence of mini-P8 from residues 1–25 and the sequence of wt P8 from codons 26–50. The resulting phagemid was used as the template for a second round of mutagenesis in which oligonucleotide *g8A2* was used to introduce the additional Ala substitutions necessary for the conversion of P8 to mini-P8. The resulting phagemid (pS1946e) contained a gene encoding hGH fused to mini-P8.

Phage ELISA for measuring hGH display

Phage ELISA protocols were adapted from previous work.[30] Cultures of *E. coli* XL1-Blue harboring phagemids were grown for eight hours at 37 °C in 1 ml of 2YT, 50 µg ml^{-1} of carbenicillin, 10 µg ml^{-1} of tetracycline. The cultures were transferred to 30 ml of the same medium, supplemented with M13-VCS helper phage (10^{10} phage ml^{-1}) and grown overnight at 37 °C. Phage were harvested from the culture supernatant by precipitation twice with PEG/NaCl[28] and resuspended in 1 ml of PBS, 0.2% BSA, 0.1% Tween 20 (blocking buffer). Phage concentrations were determined spectrophotometrically using $\varepsilon_{268} = 1.2 \times 10^8$ M^{-1} cm^{-1}.

Maxisorp immunoplates (96-well) were coated with hGHbp overnight at 4 °C, blocked for one hour at room temperature with 0.2% BSA in PBS and washed eight times with PBS, 0.05% Tween 20. Phage particles were diluted serially into blocking buffer and 100 µl portions were transferred to coated wells. After one hour, plates were washed six times with PBS, 0.05% Tween 20, incubated with 100 µl of 1:5000 (v/v) horse radish peroxidase/anti-M13 antibody conjugate in blocking buffer for 30 minutes, and then washed six times with PBS, 0.05% Tween 20 and twice with PBS. Plates were developed using a TMB peroxidase substrate system (100 µl), stopped with 1.0 M H$_3$PO$_4$ (100 µl) and absorption at 450 nm was measured.

References

1. Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **8**, 150–158.
2. Marciano, D. K., Russel, M. & Simon, S. M. (1999). An aqueous channel for filamentous phage export. *Science*, **284**, 1516–1519.
3. Russel, M., Linderoth, N. A. & Sali, A. (1997). Filamentous phage assembly: variation on a protein export theme. *Gene*, **192**, 23–32.
4. Wickner, W. (1988). Mechanisms of membrane assembly: general lessons from the study of M13 coat protein and *Escherichia coli* leader peptidase. *Biochemistry*, **27**, 1081–1086.
5. Sidhu, S. S. (2001). Engineering M13 for phage display. *Biomol. Eng.* **18**, 57–63.
6. Sidhu, S. S. (2000). Phage display in pharmaceutical biotechnology. *Curr. Opin. Biotechnol.* **11**, 610–616.
7. Webster, R. E. (2001). Filamentous phage biology. In *Phage Display: A Laboratory Manual* (Barbas, C. F. III, Burton, D. R., Scott, J. K. & Silverman, G. J., eds), pp. 1.1–1.37, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
8. Russel, M. (1993). Protein–protein interactions during filamentous phage assembly. *J. Mol. Biol.* **231**, 689–697.
9. Marciano, D. K., Russel, M. & Simon, S. M. (2001). Assembling filamentous phage occlude pIV channels. *Proc. Natl Acad. Sci. USA*, **98**, 9359–9364.
10. Lopez, J. & Webster, R. E. (1983). Morphogenesis of filamentous bacteriophage f1: orientation of extrusion and production of polyphage. *Virology*, **127**, 177–193.
11. Rakonjac, J. & Model, P. (1998). Roles of pIII in filamentous phage assembly. *J. Mol. Biol.* **282**, 25–41.
12. Glucksman, M. J., Bhattacharjee, s. & Makowski, L. (1992). Three-dimensional structure of a cloning vector. X-ray diffraction studies of filamentous bacteriophage M13 at 7 Å resolution. *J. Mol. Biol.* **226**, 455–470.
13. Marvin, D. A., Hale, R. D., Nave, C. & Helmer, C. M. (1994). Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages. *J. Mol. Biol.* **235**, 260–286.
14. Papavoine, C. H. M., Christiaans, B. E. C., Folmer, R. H. A., Konings, R. N. H. & Hilbers, C. W. (1998). Solution structure of the M13 major coat protein in detergent micelles: a basis for a model of phage assembly involving specific residues. *J. Mol. Biol.* **282**, 401–419.
15. Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone–receptor interface. *Science*, **267**, 383–386.
16. Bogan, A. A. & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9.
17. Wells, J. A. (1991). Systematic mutational analyses of protein–protein interfaces. *Methods Enzymol.* **202**, 390–411.
18. Cunningham, B. C. & Wells, J. A. (1989). High-resolution epitope mapping of hGH–Receptor interactions by alanine-scanning mutagenesis. *Science*, **244**, 1081–1085.
19. Sidhu, S. S., Weiss, G. A. & Wells, J. A. (2000). High copy display of large proteins on phage for functional selections. *J. Mol. Biol.* **296**, 487–495.
20. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. (2000). Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl Acad. Sci. USA*, **97**, 8950–8954.
21. Pearce, K. H., Cunningham, B. C., Fuh, G., Teeri, T. & Wells, J. A. (1999). Growth hormone binding affinity for its receptor surpasses the requirements for cellular activity. *Biochemistry*, **38**, 81–89.
22. Williams, K. A., Glibowicka, M., Li, Z., Li, H., Khan, A. R., Chen, Y. M. Y. *et al.* (1995). Packing of coat protein amphipathic and transmembrane helices in filamentous bacteriophage M13: role of small residues in protein oligomerization. *J. Mol. Biol.* **252**, 6–14.
23. Hunter, G. J., Rowitch, D. H. & Perham, R. N. (1987). Interactions between DNA and coat protein in the structure and assembly of filamentous bacteriophage fd. *Nature*, **327**, 252–254.

24. Deber, C. M., Khan, A. R., Li, Z., Joensson, C., Glibowicka, M. & Wang, J. (1993). Val to Ala mutations selectively alter helix–helix packing in the transmembrane segment of phage M13 coat protein. *Proc. Natl Acad. Sci. USA*, **90**, 11648–11652.

25. Weiss, G. A., Wells, J. A. & Sidhu, S. S. (2000). Mutational analysis of the major coat protein of M13 identifies residues that control protein display. *Protein Sci.* **9**, 647–654.

26. Overman, S. & Thomas, G. (1995). Raman spectroscopy of the filamentous virus Ff (fd, f1 M13): structural interpretation for coat protein aromatics. *Biochemistry*, **34**, 5440–5451.

27. Weiss, G. A. & Sidhu, S. S. (2000). Design and evolution of artificial M13 coat proteins. *J. Mol. Biol.* **300**, 213–219.

28. Sidhu, S. S., Lowman, H. B., Cunningham, B. C. & Wells, J. A. (2000). Phage display for selection of novel binding peptides. *Methods Enzymol.* **328**, 333–363.

29. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367–382.

30. Pearce, K. H., Potts, B. J., Presta, L. G., Bald, L. N., Fendly, B. M. & Wells, J. A. (1997). Mutational analysis of thrombopoietin for identification of receptor and neutralizing antibody sites. *J. Biol. Chem.* **272**, 20595–20602.