5-1-2010

Engineering RNA phage MS2 virus-like particles for peptide display

Sheldon Keith Jordan

Sheldon Keith Jordan
Candidate

Biomedical Sciences
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

, Chairperson

Engineering RNA Phage MS2 Virus-Like Particles for Peptide Display

BY

SHELDON KEITH JORDAN

B.S., Biology, The University of New Mexico, 1993
B.S., Chemistry, The University of New Mexico, 1993

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Biomedical Sciences

The University of New Mexico
Albuquerque, New Mexico

May, 2010

Engineering RNA Phage MS2 Virus-Like Particles for Peptide Display

BY

SHELDON KEITH JORDAN

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Biomedical Sciences

The University of New Mexico
Albuquerque, New Mexico

May, 2010

Engineering RNA Phage MS2 Virus-Like Particles for Peptide Display

BY

SHELDON KEITH JORDAN

B.S., Biology, The University of New Mexico, 1993
B.S., Chemistry, The University of New Mexico, 1993
Ph.D., Biomedical Sciences, The University of New Mexico, 2010

ABSTRACT

Phage display is a powerful and versatile technology that enables the selection of novel binding functions from large populations of randomly generated peptide sequences. Random sequences are genetically fused to a viral structural protein to produce complex peptide libraries. From a sufficiently complex library, phage bearing peptides with practically any desired binding activity can be physically isolated by affinity selection, and, since each particle carries in its genome the genetic information for its own replication, the selectants can be amplified by infection of bacteria. For certain applications however, existing phage display platforms have limitations. One such area is in the field of vaccine development, where the goal is to identify relevant epitopes by affinity-selection against an antibody target, and then to utilize them as immunogens to elicit a desired antibody response. Today, affinity selection is usually conducted using display on filamentous phages like M13. This technology provides an efficient means for epitope identification, but, because filamentous phages do not display peptides in the high-density, multivalent arrays the immune system prefers to recognize, they generally

make poor immunogens and are typically useless as vaccines. This makes it necessary to confer immunogenicity by conjugating synthetic versions of the peptides to more immunogenic carriers. Unfortunately, when introduced into these new structural environments, the epitopes often fail to elicit relevant antibody responses. Thus, it would be advantageous to combine the epitope selection and immunogen functions into a single platform where the structural constraints present during affinity selection can be preserved during immunization.

This dissertation describes efforts to develop a peptide display system based on the virus-like particles (VLPs) of bacteriophage MS2. Phage display technologies rely on (1) the identification of a site in a viral structural protein that is present on the surface of the virus particle and can accept foreign sequence insertions without disruption of protein folding and viral particle assembly, and (2) on the encapsidation of nucleic acid sequences encoding both the VLP and the peptide it displays. The experiments described here are aimed at satisfying the first of these two requirements by engineering efficient peptide display at two different sites in MS2 coat protein. First, we evaluated the suitability of the N-terminus of MS2 coat for peptide insertions. It was observed that random N-terminal 10-mer fusions generally disrupted protein folding and VLP assembly, but by bracketing the foreign sequences with certain specific dipeptides, these defects could be suppressed. Next, the suitability of a coat protein surface loop for foreign sequence insertion was tested. Specifically, random sequence peptides were inserted into the N-terminal-most AB-loop of a coat protein single-chain dimer. Again we found that efficient display required the presence of appropriate dipeptides bracketing the peptide insertion. Finally, it was shown that an N-terminal fusion that tended to interfere

specifically with capsid assembly could be efficiently incorporated into mosaic particles when co-expressed with wild-type coat protein.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

History

Initially described by George Smith in 1985, phage display represents a method for the presentation of peptides or proteins on the surfaces of virus particles. Essentially, peptides are "displayed" by fusing nucleotides encoding the genomic information of the target peptide to genomic regions coding for coat protein of phage. Using recombinant DNA techniques, foreign sequences are joined to a gene encoding a coat protein, thereby creating a fusion protein [1]. These recombinant proteins are then incorporated into newly forming phage particles preparing for final release from the host cell. Libraries of phage particles are created when a mixture of recombinant sequences are utilized to produce phage clones each carrying a unique recombinant sequence.

Presently, numerous libraries of phage displaying peptides, antibodies, or other bio-molecules have been constructed and many are now being screened for the ability to bind to specific ligands. Each recombinant phage genome contains the DNA sequence encoding the peptide or protein displayed on its surface. This means that phages displaying peptides with a particular binding function (e.g. affinity for a monoclonal antibody) can be physically purified from a complex random sequence library and then amplified by simple infection of E. coli. It should be noted that filamentous phage display, and other related technologies, rely on two essential features: (i) the ability to display a foreign peptide in a biochemically accessible form, and (ii) a physical linkage of the displayed peptide or protein to the nucleic acid that encodes it.

Filamentous Phage Display

Although several different bacteriophages have been utilized for this purpose, display systems based on filamentous phage dominate the scientific landscape. The most commonly used are based on f1, fd, and M13 phages. As figure 1 depicts, several coat proteins are present in the final phage particle. Coat protein pVIII predominates and is frequently a target for display. However, other coat proteins including pIII and pVI are also utilized.

The Life Cycle of Filamentous Phage

The life cycle of filamentous phage is schematically presented in Figure 2 [2]. Infection starts as pIII attaches at the F pilus of *Escherichia coli*. The pilus then pulls in towards the cell until the pIII protein of phage particle is able to interface directly with the cellular boundary. Next, in a poorly understood process, the phage particle then begins to disassemble and transfer of the viral genome occurs. Once the viral genome is in the cell, it is converted into double stranded DNA (dsDNA). The genome is then amplified in preparation for mature viral particle assembly.

As genomic amplification occurs, viral proteins also begin to appear in the cell. Several of these viral proteins are targeted for export into the periplasm in preparation for final assembly of the mature particle. The viral particle is assembled at the cell membrane and is extruded from it while the cell itself remains intact. Of course, it is precisely these viral structural porteins that are the targets for foreign peptide insertion in phage display. Some peptides possess amino acid sequences incapable of traversing the host cell wall. This inability to traverse the cellular membrane relates to the characteristics of some amino acids (i.e hydrophobicity). As a result, some recombinant phage particles never

exit the host cell and are not represented in the final recombinant phage library population. This resulting loss of library sequence complexity is sometimes called library censorship.

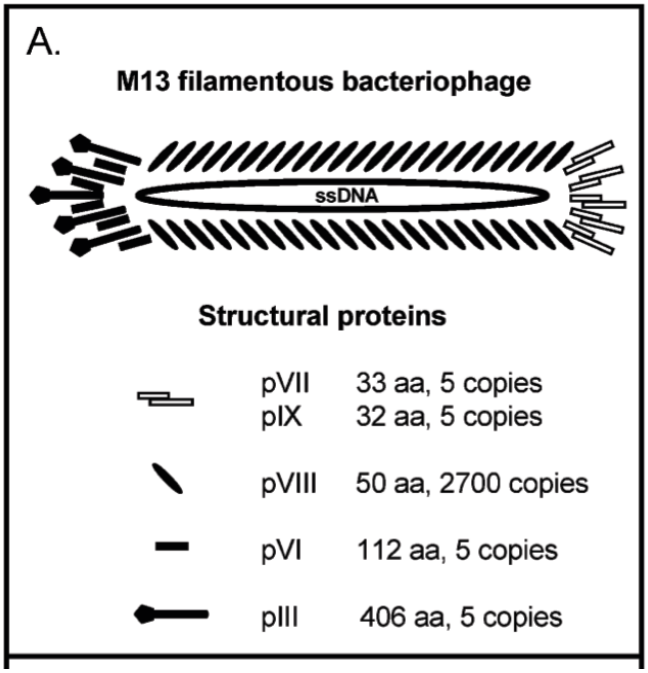


Figure 1. A view of M13 filamentous phage
The structure of M13 is similar to the structure of other filamentous phage particles. The particle is approximately 1 □m in length and about 6 nm in diameter. The major coat protein, pVIII, encircles the genome encoding the particle and has been a target for phage display. At the temini of the phage particle, pIII, represents another frequent target for phage display applications. Reprinted with permission [3].

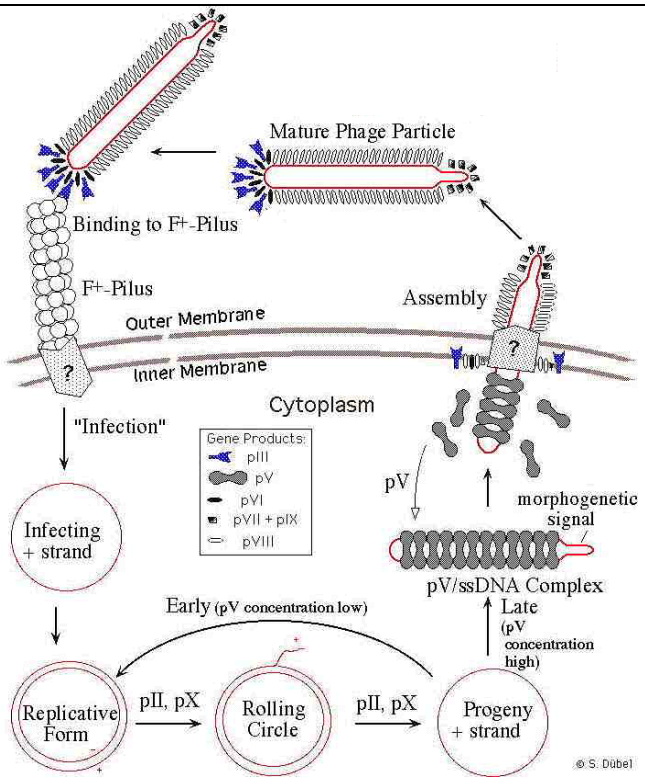


Figure 2. A view of M13 filamentous phage assembly
A critical aspect to M13 phage assembly is that it occurs in the periplasmic space in between the inner and outer membrane of the host cell. As the diagram depicts, however, some questions remains on how M13 is able to complete the highly complex and orchestrated process of cellular exit. Viral particle exit occurs without killing the host cell and is often described as a parasitic process. Reprinted with permission [2].

Alternative Phage Display Systems

As a result of library censorship and other limitations of the filamentous phage display system, efforts have focused on other bacteriophages for display applications. Of note, have been recent intensive efforts to bring lytic phages online for phage display. Lytic phage systems offer an advantage in that proteins associated with capsids are assembled and folded in the cytoplasm, thus avoiding avoiding constraints imposed by the secretory process filamentous phages utilize to generate phage particles [1]. Presently three prominent lytic phage systems have been identified for display applications. These

platforms include T4, T7, and λ. However, these new display systems have limitations of their own [4]. For example, T4 does not yet have an efficient *in-vitro* packaging system, and thus limits its usefulness in larger peptide display libraries [4]. In T7, it has already been demonstrated that some small proteins are refractory to high copy number display [4]. In the case of λ, capsid assembly involves two prominent proteins and takes several steps to complete the phage construction process – which may hinder efficient packaging and display of foreign peptides. In any case, all these systems are based on complex viruses whose replication and morphogenesis require infection of intact cells. This point is made because other display activities, not based on bacteriophages, have already moved away from intact cells to cell-free systems in an effort to increase library diversity and to potentially address concerns related to library censorship, cellular toxicity, etc [5].

Peptide Presentation Systems Not Based on Phages

Ribosome Display

Ribosome display offers a unique approach to peptide display [6-8]. Libraries of random peptide or protein sequences, usually linked by recombinant DNA methods to a carrier protein, are expressed in cell-free protein synthesis systems derived from an appropriate organism (e.g. E. coli, wheat germ, rabbit reticulocytes, etc). Thus, even proteins toxic to the hosts in which they are normally expressed can be collected and subjected to analysis. Ribosome display extends cell-free expression by linking newly expressed proteins to the mRNA template from which they were expressed (Figure 3). And these complexes of protein, ribosome, and mRNA are subjected to affinity selection. The linkage of foreign sequences to the mRNA that encodes them ensures that selected sequences can be recovered and amplified [7, 8].

Peptide Display using mRNA

An offshoot of ribosome display is mRNA display (Figure 3, bottom panel). In this system, peptides in the library become covalently fused to the mRNA strand that encodes them. Briefly, mRNA display starts with a library of DNA. The library is then transcribed into mRNA -with these transcripts becoming fused to a DNA linker. This DNA linker also contains a puromycin moiety. As the mRNA is translated *in-vitro*, the ribosome proceeds until it encounters the RNA-DNA boundary, where it stalls. It is at this point that the puromycin moiety binds at the A site of the ribosome. This action then allows for the generated peptide to become transferred to the puromycin moiety and released from the ribosome. This new complex of mRNA and protein is then used for subsequent peptide display processing.

Peptide Display on Hepadnavirus-Like Particles

The approaches described above all have the ability exploit affinity selection as a means of identifying peptides with desired binding activities. Sometimes, however, it is sufficient simply to display a known peptide sequence previously identified by other means. For example, peptides are sometimes displayed on virus-like particles (or VLPs) as a means of conferring to them a high degree of immunogenicity, thus making them potentially useful as vaccines. Billaud et al. have inserted foreign peptides into various surface-exposed sites on hepadnavirus core protein VLPs [9]. However, it is frequently observed that such insertions disrupt the ability of the viral protein to properly fold. To accommodate a wide variety of different peptide inserts, this group found it necessary to identify an "optimal combination" of insertion site coupled with other stabilizing modifications to allow for maximum incorporation of foreign sequences. They showed

that flanking acidic amino acid brackets around a recombinant sequence stabilized insertions and allowed for virus-like particle assembly. However, optimal insertion position, insert sequence, and C-terminal modification were also required. This approach has served well for generating virus-like particles displaying a variety of peptides, but the method is laborious, since each peptide can require an unpredictable combination of insertion site and C-terminal modification. Furthermore, the lack of an affinity selection capability makes this system unsuited to epitope identification by affinity selection against antibody targets. The work described below was motivated by a desire to develop a phage display platform combining the potent immunogenicity of the VLP with the affinity selection capabilities of existing phage display systems.

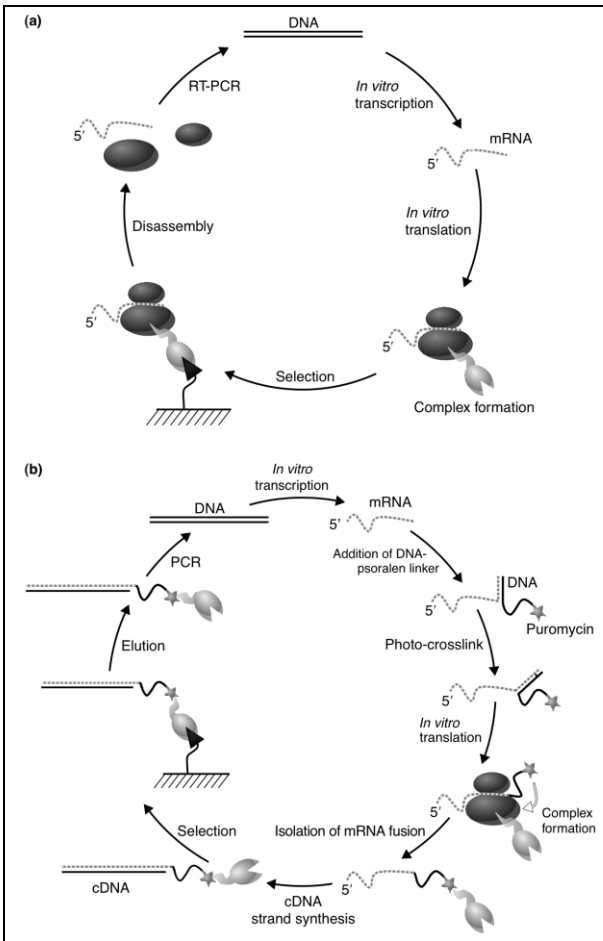
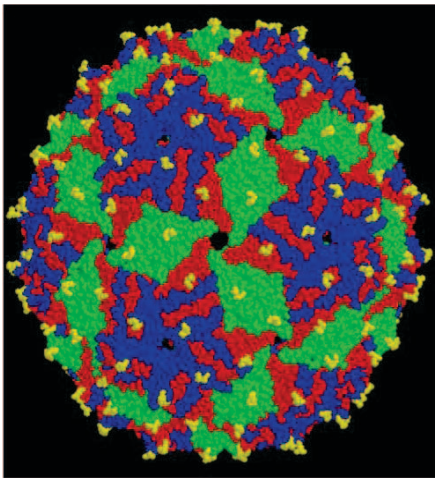


Figure 3. A view of ribosome and mRNA display

The top panel highlights ribosome display. In this system, a DNA library encompassing proteins of interest is first linked to a spacer sequence that has no stop codon. This new DNA hybrid is then transcribed into an mRNA complex ready for ribosomal docking and translation. As the ribosome translates the resultant mRNA, it encounters the spacer sequence and stalls. At this point, translation is terminated and the resulting complex of bound ribosome, mRNA, and protein become available for subsequent peptide display applications. In the bottom panel, a system is described for mRNA display. In this system the ribosome is ultimately released. In this system, the library of DNA is transcribed *in-vitro*, and the resulting mRNAs becomes available for further processing. The library of mRNA transcripts is fused to a DNA linker that carries puromycin. As the ribosome proceeds with translation, it reaches the RNA-DNA boundary and stalls. The puromycin moiety is then able to complex with the ribosome at the A site, where the peptide is transferred to the puromycin moiety. This final step causes ribosomal release resulting in a complex containing the mRNA (genotype) covalently linked to resulting protein (phenotype) via a puromycin linker. Reproduced with permission [10].

Bacteriophage MS2: A New Phage Display Platform

The efforts described in this dissertation focused on the development of a peptide display system based on VLPs of bacteriophage MS2 (Figure 4), a small (+)-sense, single-stranded RNA virus with a geonome consisting of only 3,569 nucleotides. The icosahedral shell of the viral particle is made up of a single coat protein, which, in addition to its structural role, acts as translational repressor of viral replicase synthesis. As we will see, it is □□ 7E □□ IC V H 7C HA I □□ CA 7 □□ Eressor activity that is of significant interest for a new phage display system. As manipulation of the coat protein for display applications requires the insertion of peptides, it is anticipated that non-functional coat proteins will easily be identified as non-functional repressors and quickly be excluded from analysis as the operational parameters of a display system based on MS2 are established.



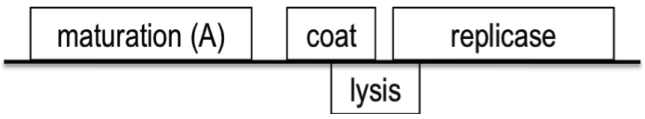


Figure 4. A View of MS2
A view of the capsid produced by MS2. The genome of bacteriophage MS2

Engineering Advantages of MS2

As bacteriophage MS2 VLPs were being evaluated for phage display applications, several advantages were quickly identified. First, MS2 consists only of a single coat protein. Filamentous phage, on the other hand, employs several coat proteins to achieve final particle assembly. In the case of MS2, it is expected that by having only one coat gene to manipulate that problems routinely encountered during peptide insertion will be easier to identify and correct. Second, MS2 coat protein readily self-assembles into capsids of icosahedral geometry. Coat protein alone is sufficient – i.e. no other viral components are required for assembly and coat protein expressed from a plasmid efficiently produces virus-like particles [11]. Importantly, MS2 VLPs package the mRNA that encodes them, thus providing the linkage between the phenotype and genotype required for amplification of affinity selected sequences.

Because of the structural simplicity of the MS2 VLP, it should be possible to produce the particles *in vitro* by coupled transcription/translation. Due to constraints imposed by the need to propagate other phage display vehicles in bacteria, the practical upper limit on peptide library complexity is thought to be around 10^{10}-10^{11}. However, if VLP synthesis and assembly could be made to occur *in vitro*, growth in *E. coli* would be unnecessary and the numerical advantages inherent in working with molecules instead of cells, would become manifest. The first requirement of a display system based on MS2 VLPs is the identification of sites in coat protein that both tolerate the insertion of foreign peptide sequences and present the peptide in accessible form at the VLP surface. In a nutshell, this is the purpose of the work described here.

Proposed Insertion Sites for MS2 Display

A first step in developing bacteriophage MS2 for display applications was to identify sites in the coat protein that accepted foreign peptides and presented them at the capsid surface. Inspection of the coat protein's structure revealed two obvious candidates, the AB-loop, a short sequence connecting two ß-strands of coat protein on the surface of the viral particle, and the N-terminus. The AB-loop has already been employed for the engineered presentation of a few specific peptides [11-13]. Additionally, limited work has been performed to suggest that the N-terminus may also be a viable target for display applications [11]. To determine the utility of the AB-loop and N-terminus for peptide presentation, we proposed to create a library of peptide insertions and determine the fraction of recombinants that retained function. It was anticipated that these peptide insertions, as in other systems, would lead to non-viable phage virus-like particles. These deficiencies have occurred either because of a failure of the capsid protein to properly fold, or because to folded recombinant coat protein failed to assemble into a VLP. This dissertation describes efforts to accomplish peptide display on MS2 VLPs and has the following aims:

Specific Aim #1: *To determine the tolerance of the AB-loop to peptide insertions generally, and assess the potential use of single-chain dimers to allow for greater peptide library diversity.*

Specific Aim #2: *To determine the tolerance of the N-terminus to peptide insertions generally, and assess the potential use of single-chain dimers to allow for greater peptide library diversity.*

Prior Work on MS2 VLP Display

A great deal of effort has been devoted to understanding the RNA recognition properties of coat protein. During those studies two technical tools were devised that have special relevance to the present work [14]. First, since coat protein is a translational repressor of replicase synthesis it was possible to construct a two-plasmid system in which coat protein expressed from one plasmid represses the translation of a replicase-ß-galactosidase fusion protein expressed from a second plasmid. Thus bacteria expressing a functional repressor give white colonies on plates containing the chromogenic substrate x-gal, while repressor defects yield blue colonies. In the present work this translational repression assay provided a convenient means of determining whether a peptide insertion inactivated coat protein by disrupting folding. Second, a convenient electrophoretic assay was devised that permits the facile determination of the presence of VLPs in crude lysates of coat protein expressing cells.

Preliminary efforts to display peptides on MS2 VLPs were described by Peabody, who showed that insertion of the so-called Flag peptide in the AB-loop of MS2 coat protein resulted in a failure of the protein to properly fold [11]. However, this folding failure could be suppressed in coat protein single-chain dimers, a modified form of coat protein in which the two identical chains of the coat dimer are fused into a single polypeptide chain. Subunit fusion increases the thermodynamic stability of the protein and greatly increases its tolerance of amino acid substitutions, peptide insertions and chemical denaturants. The presence of the Flag peptide in the AB-loop was tolerated as long as it was restricted to one of the two halves of the single-chain dimer. In the same work, Peabody described the display of the Flag peptide by fusion to the coat protein N-

terminus, and showed that although the presence of the peptide did not interfere with protein folding (as assessed by the translational repression assay), assembly of the VLP required the use of the single-chain dimer. Presumably the presence of Flag caused molecular crowding a local 3-fold symmetry axes. Of course the single-chain dimer has only half as many N-termini as the wild-type protein, so crowding was sufficiently reduced to allow VLP assembly. Finally, a more comprehensive study conducted by Peabody et al., showed that the vast majority of random sequence insertions into the AB-loop of the downstream half of the single-chain dimer was well tolerated as assessed by retention of translational repressor activity and the ability to produce a VLP [12]. This body of prior work provided a model for the experimental approaches taken in the work described in this dissertation.

Characterize Recombinant Folding via Translation Repression Assay

In order to characterize the functional behavior of recombinant proteins, a translational repression assay was first performed. This assay takes advantage of a unique functional feature associated with MS2 coat protein. Under normal conditions, the coat protein of MS2 acts to repress translation by inhibiting the synthesis of viral replicase. MS2 coat protein, as it dimerizes, is able to achieve this by binding to the translational operator region (i.e. ribosome-binding site) of replicase. As described in a study by Peabody, the region encompassing the translational operator of replicase can be fused to the *E.coli* lacZ gene to form a replicase-β-galactosidase fusion protein[14]. This fusion protein, as demonstrated in the same study by Peabody, is then capable of being modulated by the production and interaction with wild-type MS2 coat protein. The lacZ gene, which is part of the lac operon, allows for the synthesis of β-galactosidase. On

indicator plates that contain 5-bromo-4-chloro-indolyl-β-D-galactosidase (X-gal), β-galactosidase cleaves X-gal producing a characteristic blue color. In this same study by Peabody, the replicase-β-galactosidase fusion construct was placed into a separate plasmid and then stably transformed with a plasmid containing the MS2 coat sequence under study. The results demonstrated that by employing a two-plasmid system whereby coat protein generated on one plasmid repressed the production of β-galactosidase on another plasmid, MS2 coat protein function could readily be assessed. Thus, this study established a link between the ability of MS2 coat protein to repress translation and its inherent folding behavior. In the present study, a similar repression assay was utilized.

Characterize Recombinants by Restriction Digest

The next step in the experimental process was to verify whether or not recombinants were being generated in the newly synthesized libraries. From each library, a number of individual white and blue colonies were be selected and picked to two 1 ml cultures in LB medium and grown overnight with agitation at 37° C. One set of cultures was then used to isolate the recombinant plasmid via mini-prep methodology. In this case, once the plasmids were obtained, those same restriction endonucleases used to insert the original peptides fragments were used to generate new fragments from the library and compared against restriction fragments generated by parental plasmids that contain unmodified coat sequences. Digested fragments were then subjected to gel electrophoresis and characterized. Those fragments that were recombinant had a greater size than those generated from parental plasmids. From those colonies that were deemed to be recombinant, an additional step of DNA sequence analysis was performed.

Virus-Like Particle (VLP) Formation Assay

A key outcome of this experimental effort was the expectation that the introduction of peptides into the coat of MS2 would not prevent VLP assembly. To monitor VLP formation, individual white or blue clones were selected from each experimental variation and subjected to analysis. First, sonicated cell lysates from individual clones were collected and then subjected to agarose gel electrophoresis. Next, these lysates were visualized via an indirect method by which gels were stained with ethidium bromide. This step was critical in that ethidium bromide intercalates with the nucleic acid normally encapsidated by properly formed virus-like particles. Although not definitive, ethidium bromide staining provided the first indication that VLPs incorporating N-terminal extensions were being formed.

Once completed, ethidium bromide staining was then contrasted with Western blot analysis to firmly establish the presence or absence of MS2 VLPs. Western blot analysis was accomplished by staining a blot with anti-MS2 serum, followed by a resolution step involving the application of alkaline phosphatase-conjugated anti-rabbit IgG. As recombinant MS2 VLPs are inherently made up different amino acids, the mobilities associated with amino acids served to further establish the presence of unique recombinant VLPs. Thus, this assay provided insight into VLP formation.

To test the hypothesis that MS2 could act as a suitable platform to generate peptide libraries, we then initially investigated insertions at the N-terminus and in the AB-loop of the MS2 coat protein. Expecting that these insertions would be highly disruptive to VLP formation, we added an additional element in which we would introduce the use of a single chain fused dimer of MS2 coat protein to potentially

stabilize any aberrant protein recombinants[11]. Figure 5 illustrates the nature of the insertions made.

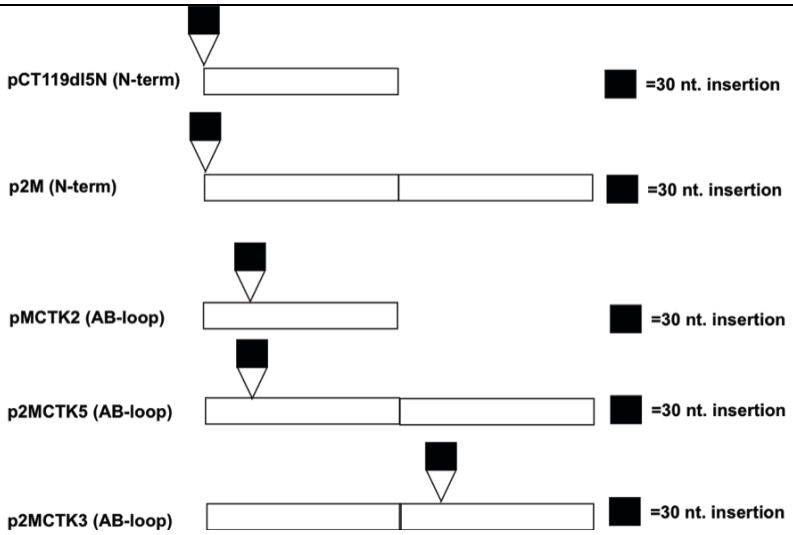


Figure 5. Templates for construction of MS2 libraries
Recombinant libraries were generated in the monomeric and dimeric forms of MS2 coat protein. Insertions were made at the N-terminus and AB-loop of the coat gene and recombinants were subsequently obtained. The nature of these insertions was randomly generated 30 nucleotide sequences. These sequences were generated using primers of the $(NNY)_{10}$ replicated motif respectively - where N = A,C,G, or T; Y= T or C. Use of these primers allowed for the generation of 15 amino acids without the inclusion of any stop codons. Amino acids not produced utilizing this primer motif were Lys, Glu, Gln, Trp, and Met. This approach was undertaken to allow for rapid initial assessment of recombinants.

Recombinant sequences were then generated based on the proposed scheme for peptide insertion and transformed into a laboratory strain of e-coli and plated on nutrient rich agar to generate coat protein. Utilizing the aforementioned repression assay, colonies were then segregated based on their implied ability to fold properly into dimers and then subjected to sequence analysis. After obtaining sequence information, only those recombinants with proper sequence structure (i.e no frameshifts, stop mutations, etc.)

were included in subsequent analysis. Table I highlights the results of these initial experiments.

Insertions into the monomer of MS2 produced very few white colonies – suggesting that these insertions were highly disruptive at both the N-terminus and AB-loop. However, we also generated recombinants into a single-chain fused dimer of the MS2 coat protein. The results of these dimer experiments are also shown in Table I. In the case of 30 nucleotide or 10 amino acid insertions, modifications to AB-loops benefitted from the implementation of the single-chain dimer. Insertions into the 3' downstream loop showed greater than 90 % fold competence. However, the results for the 5' upstream half of the dimer were still discouraging. With fold competence near 40% for the 5' AB-loop, it was assumed that the experimental design had introduced some anomaly into the effort and the construction of recombinants and subsequent analysis was repeated. However, at the end of this new effort, the initial results were confirmed and it appeared that only the 3' loop of the single-chain dimer would be viable in any future phage display efforts. In the case of N-terminus, insertions into the monomer and dimer of MS2 demonstrated that the n-terminus was almost entirely refractory to proper folding in both test conditions. This result was unexpected given that previous work had shown that peptides insertions were possible at the N-terminus of MS2, and that the single-chain dimer was a beneficial addition[11].

As part of this study by Peabody, the Flag peptide was inserted at the N-terminus of MS2. It was expected that the Flag peptide could be expressed in a VLP, and that a commercially available antibody would detect the presence of Flag. Experimental data showed that Flag incorporation did not impact the ability of MS2 coat to fold. However,

the Flag coat fusion did fail to assemble into a VLP. To correct for this assembly defect, two coat protein monomers were fused into a single-chain dimer and the Flag peptide was re-incorporated. In this new configuration, the Flag peptide was shown to properly fold and assemble into a VLP.

During preliminary experiments for this new effort, however, the n-terminus appeared refractory to random insertions even with the additional use of the single-chain dimer. It was at this point that a key observation was made. N-terminal insertions, in general, were highly destabilizing, and that future attempts would likely require extensive engineering of the recombinant insert to produce viable proteins.

Table I. Preliminary translation repression results

Translational repression as express by colony color provides insight into ability of recombinant protein to fold. Recombinant libraries were generated in the monomeric and dimeric forms of the MS2 coat protein. Insertions were made at the AB-loop or N-terminus of the coat gene and recombinants were subsequently obtained. The nature of these insertions was 30 nucleotide (10 amino acid) sequences.

			Percent White Colonies		
N-Terminus			AB-Loop		
	pCT119dl5n	p2M	pMCTK2	p2MCTK5	p2MCTK3
$(NNY)_{10}$	ND	16	ND	44	92

A NEW DIRECTION

Given the inability of bacteriophage MS2 to effectively tolerate insertions at the N-terminus and in the 5' AB-loop, the focus of this effort quickly shifted to explore the potential nature of the peptide insertions themselves. Length of insert, volume of insert, hydrophobicity, and relative isoelectric point were suggested as key determinants influencing peptide behavior in preliminary literature reviews. However, a deeper review of research extending back thirty years suggested that peptide behavior in the context of a phage display system may be a much more complex problem. After a more detailed review of relevant research on peptide insertion, *de novo* protein synthesis, and general aspects of protein folding, five novel approaches emerged in relation to the display of foreign peptides. These five ideas are listed as follows:

- Masking the nature of peptide insertions
- Amino acid propensities in the context of secondary structures
- Stereochemical initiation/termination of secondary structures
- Exploitation of cellular processes that enhance folding
- Helper proteins in mosaic capsid formation

Each of these approaches encompassed an expanded approach to allow for the accommodation of foreign peptides at the N-terminus and AB-loop of bacteriophage MS2. Background on each approach follows in the next sections. As a result of this revised, expanded effort, the final specific aims for this peptide display study were expanded to include on additional theme. The final theme centered on several targeted approaches was devised to specifically address N-terminal insertions. Results from all experimental efforts are detailed later in this study.

Masking the Effects of Peptide Insertions

Peptide insertions in the AB-loop and N-terminus of bacteriophage MS2 have proven to be highly destabilizing. Ballaud et al. were faced with a similar problem in their attepmtps to display foreign peptides on hepadnaviral VLPs. Basically, this group made the observation that the "optimal combination" of insertion site coupled with stabilizing modifications could allow for maximum incorporation of foreign sequences [9]. Along with this observation, was a key finding that the charge of the foreign sequence appeared to greatly influence whether or not peptide incorporation and successful virus-like particle assembly was achieved. As the group continued to refine their combinatorial approach, it was noted that the increased presence of highly basic amino acids correlated with the inability of hepadnavirus-like particles to assemble [9]. Additionally, as more data was accumulated, a theme emerged that suggested that the isoelectric point of the target foreign sequence also correlated with particle assembly [9]. These findings were significant in that the final experiments described in this study employed the use of flanking acidic amino acid brackets to revert assembly defects observed with peptides that had high net positive charge (high pI) [9]. However, optimal insertion position, insert sequence, and C-terminal modification were also required to effectively allow for peptide insertion in this hepadnavirus model. Yet, this idea of being able to influence charge could provide the impetus to try and "engineer" peptides prior to insertion at the N-terminus of MS2 to revert folding and assembly defects encountered in our preliminary experiments. Although the efforts of this study suggested a rational approach to increase successful peptide incorporation, other factors including hydrophobicity, high β-strand index, insertion length, etc., also required consideration

[9]. With this notion in mind, literature reviews identified other opportunities to potentially aid in the display of peptides.

Amino Acid Propensities in the Context of Secondary Structures

Amino acid sequences carry key structural information for native proteins as they fold into their final conformations. Although this concept is not new, it offers a fundamental premise that must be considered when attempting to insert foreign sequences into virus-like particles. As the literature suggests, considerable effort has been expended to link an amino acid sequence to a propensity to form a defined secondary structure [15-20]. In a hallmark paper, an innovative predictive model was established in which amino acid sequences prompted the development of secondary structure [18]. In this study, crystallographic data was employed to establish amino acid preferences for β-sheet, random coil, and helical regions for 15 proteins. Amino acids were classified as residues in helix, residues in inner helix, residues in β region, or residues in coil region. Concurrent to this effort, a far-reaching mechanism was proposed that posited helix and beta nucleation originated centrally in their respective structures and spread bi-directionally until strong terminal residues were encountered [18, 19]. As a result, rules established from this study allowed for the generation of an early model to predict secondary structure from an amino acid sequence [19]. With this in mind, Chou and Fasman became one the first groups to establish conformational parameters or what others call propensities for secondary structures.

In another foundational study, preferences were noted for amino acids throughout the span of a α-helix and also at positions close to its termini. This study conducted by Richardson and Richardson differed from the one presented by Chou and Fasman in that

amino acid propensities were identified specifically for positions throughout a helix as opposed to "general" secondary structure regions (i.e. inner, terminal region, etc.). Essentially, Richardson and Richardson identified for the first time position-specific preferences for amino acids in helices – which could provide benefit in reverting MS2 fold defects encountered in the N-terminus. In this study, 215 helices were compiled into a database so that statistically significant single-position helix preferences were established for helices [15]. Additionally, a new concept was defined as part of this effort in which N- and C-helix terminal residues became defined as capping residues. This concept was significant in that the N-cap or C-cap residues, which by definition are residues half in and half out of a helix, represented true terminal boundaries in secondary structures [15]. This ability to identify an amino acid at the helix interface was an outgrowth of a prevalent theory that helix boundaries could be defined by α-carbon positions. More concisely, cap residues in a helix could be defined as the first (N-cap) or last (C-cap) amino acid whose α-carbons reside in a cylinder formed by the spiral structural backbone of a helix [15]. Prior to this assertion, three theories dominated formation of helix boundaries. One defined these boundaries by ϕ and ψ angles, while the other two looked at hydrogen bonding and α-carbon positions [15]. In this study by Richardson and Richardson, each of these theories was tested – with the α-carbon theory emerging as being more likely. Once this definition for a helical boundary was established, Richardson and Richardson identified Gly as having a strong C-cap preference and Ser as having a strong N-cap preference. In both cap instances, other researchers reported similar findings, so the theory and inferences made from α-carbon position appeared to be reinforced. Additionally, Asn, Gly, Thr and Asp were identified

as also having strong N-cap preference. Ser, Asn, Lys, Ala, and Leu emerged as having the highest propensity for the C-cap. However, Gly was clearly most often found at the C-cap. Based on the results of this study, Richardson and Richardson speculated if the substitution of strongly preferred amino acids in correct locations would stabilize helices or whole proteins. However, for the purposes of this study on MS2 as a platform for peptide display, other structural considerations would have to be evaluated as well.

If secondary structure propensities represented a new opportunity to manipulate peptides in a phage display context, then β-structures must also be considered. As a review of current literature reflects, β-sheet propensities have proven to be more difficult to determine as opposed to those from α-helices [21]. The fact that β-sheets do not appear to fold in isolation and conflicting experimental data from several model systems appears to be the reason for this problem [21, 22]. Additionally, the fact that β-sheets can exist as parallel or anti-parallel strands further complicates the elucidation of β-sheet propensities. Simply put, amino acid residues in a parallel conformation could express different folding effects when positioned in an anti-parallel orientation. Thus, even the orientations of amino acid residues in a particular secondary structure are a consideration when trying to revert folding defects.

As in the case of helices, several models have emerged that support the initiation and termination of β-sheets. One model suggests that side chain interactions can sterically prevent the establishment of hydrogen bonds between adjacent amino acid residues and solvent [23]. Another, more recent model, suggests that β-sheet propensities are guided by the tendency to minimize the surface area of β-strand backbones [24].

However, neither model fully clarifies the interactions of amino acid residues in β-sheets. One model, albeit equally incomplete, appears to better clarify this relationship between amino acids and the formation of β-sheet secondary structure. In this model, sheet propensity is characterized as simply the avoidance of steric clashes between sheet backbone and residue side chains [21]. However, because this model does not delineate position specific β-sheet propensities, extracting useful *de novo* protein design cues would be limited.

A recent study extended the work of Street, Mayo, and others to actually identify capping propensities for β-sheets. In this study, Asp, Asn, Pro, and Ser were identified as β-sheet terminators, while Lys and Arg emerged as β-sheet initiators in both parallel and anti-parallel strands [22]. This effort re-emphasized the importance side-chain sheet backbone interactions and could, if successful provide some insight into a mechanism for stabilizing N-terminal insertions in MS2. One other key finding of this effort related to the under-representation of Asp and Asn in β-sheet interiors. FarzadFard et al. theorized that the interactions of these two amino acids in the interiors of β-sheet could limit backbone-backbone interactions typically seen between partner strands. Thus, the limiting interactions of Asp and Asn could prove to be detrimental to β-sheet formation. Taken in context, this additional outcome could prove to be useful for *de novo* protein design considerations.

After review of these early experiments on helices and sheets, a key research question emerged - could amino acids be strategically employed to influence peptide insertions? In order to do so effectively, amino acids cannot confer any significant

constraints on the random peptide targeted for insertion. In fact, if peptides do become constrained, it has been suggested that a library constructed from these sequences would represent less conformational complexity than peptides in an unconstrained library [1]. Additionally, as it has been noted in other efforts, random peptides will not generally fold into defined secondary structures [1]. Therefore, successful strategic placement of amino acids around a peptide must have subtle impact, and most importantly, preserve maximum conformational complexity. In this same review by Smith and Petrenko, it was also noted that peptides could become constrained within the framework of a protein scaffold. More precisely, β-structures or α-helices have the ability to impart structural constraints on inserted peptides- especially in the context of phage display. However, this particular literature review was trying to infer that the constraints imparted by β-sheets, α-helices, or other secondary structure motifs would be beneficial to peptides given that they mostly lack structures of their own [1]. With this in mind, it should also be noted that peptides do not natively exist in one rigid structural conformation. Instead, even in the case where a peptide is 'helical," peptides readily move between non-helical, helical, and partially helical arrangements [25]. Therefore, it is conceivable that employing selected amino acids to influence the presentation of peptides within the context of phage display could have detrimental, unintended consequences.

Whether the context is a α-helix or a β-structure, amino acids have been shown to influence the propagation of secondary structures [15, 16, 22, 25, 26]. Most research efforts have focused specifically on understanding the role of amino acids at the boundaries of β-sheets and α-helices. However, amino acid propensities have also been established for positions that span across a given secondary structure [18, 21, 26, 27]. In

the case of α-helices, propensities have even been shown for amino acids 15 residues in from the n-terminus [27]. The implication here was that amino acid residues throughout the length of a defined secondary structure all work in concert to govern the overall nature of the protein in which they reside. Therefore, in the case where we seek to revert n-terminal folding defects in bacteriophage MS2, flanking residues may not exert as much structural influence as once thought. Instead, these flanking amino acids may simply interact favorably with the backbone of the recombinant peptide- thereby allowing the inherent structural nature of the inserted peptide to be conveyed.

Stereochemical Initiation/Termination of Secondary Structures

The AB-loop of bacteriophage MS2 has recently shown initial promise in being able to present viral epitopes and stimulate B-cells [12]. During preliminary experiments for this effort, it was quickly recognized that the use of the single-chain dimer would be required to allow for effective peptide presentation in the AB-loop. As illustrated in Fig. 5, once the coat sequence of MS2 was effectively duplicated, two AB-loop sites become available for potential peptide presentation. For point of clarity, the upstream (blue) AB-loop region has been labeled as the 5' insertion site -with the downstream AB-loop region (red) labeled as the 3' insertion site.

As preliminary efforts began to progress on the viability of the MS2 platform for vaccine applications, it was observed that the 5' AB-loop insertion site, even in the context of the single-chain dimer, remained mostly refractory to random peptide insertions (data not shown). This was surprising given the observation that peptide insertions in the 3' AB-loop are well tolerated in the single-chain dimer [12]. We assumed that 3' insertions were clearly benefiting from being in the context of a single-

chain dimer. Conversely, we also speculated that 5' insertions could potentially destabilize the overall single-chain dimer. If, as the single-chain dimer begins to form, the upstream coat sequence remains improperly folded, then even a viable downstream wild-type coat sequence cannot provide sufficient stability to overcome this defect. With this in mind, initial attention turned to opportunities to stabilize 5' random insertions.

A group of researchers working with hepadnavirus employed a combinatorial approach to maximize peptide incorporation. In this system, the research group observed that the "optimal combination" of insertion site coupled with flanking acidic amino acids could allow for maximum incorporation of foreign sequences [9]. Thus, it was postulated that insertions into the 5' AB-loop of MS2 could benefit from the efforts of Billaud et al. Indeed, incorporating charged amino acids to stabilize random insertions seemed innovative. However, Billaud et al., correctly highlight other factors including hydrophobicity, high β-strand index, insertion length, etc., for consideration in reverting folding defects. The ideas presented in this study represented a starting point for reverting AB-loop failures. However, it was clear that other approaches would need to be considered.

The concept that amino acids confer intrinsic structural information for native proteins as they fold into their final conformations is not new. In a hallmark paper, a mechanism was proposed that suggested beta strand and helix formation originated centrally within a polypeptide and spread bi-directionally until "terminating" residues were encountered [18, 19]. Given that two beta strands bound the AB-loop of MS2, attention was turned to opportunities to manipulate beta structures. A quick review of current literature revealed that β-sheet propensities have proven to be much more difficult

to resolve [21]. However, several models have been proposed to describe the initiation and termination of β-sheets. One model proposes that side chain interactions work to prevent the hydrogen bond interaction that occurs between adjacent amino acid residues and solvent [23]. Another model posits that β-sheet propensities are guided by an inherent need to minimize the surface area of β-strand backbones [24]. However, one model simply describes sheet propensity as the minimization of unfavorable interactions between sheet backbone and residue side chains [21]. In 2008, a research team extended this simple approach and developed a model for the capping propensities of β-sheets. Asp, Asn, Pro, and Ser emerged as β-sheet terminators, while Lys and Arg were identified as β-sheet initiators [22]. With this data now in hard, we speculated if the propagation of secondary structures could be terminated by the strategic placement of amino acids? Additionally, could the effects of randomized sequences inserted into the 5' AB-loop be minimized with the same approach?

As emphasis has been made on employing amino acids to positively influence random sequence incorporation into the AB-loop of MS2, other cellular activities may also provide direct opportunities to revert folding defects. As we will explore further in detail, we suggest the possible recruitment and interaction of molecular chaperones with arginines as a novel folding mechanism for MS2 display. Briefly, the premise behind this approach was that the incorporation flanking arginines around a recombinant peptide region would recruit DnaK and encourage coat folding. In the case of the AB-loop, we expect that arginines will ultimately allow for recombinant coat sequences to fold.

Exploitation of Cellular Processes that Enhance Folding

As consideration has been given to utilizing amino acids to directly modulate peptide insertions into MS2, other cellular processes may also offer equally innovative opportunities to revert folding defects. Of particular significance is the potential use of molecular chaperones. Scientists that employ filamentous phage in protein engineering applications readily acknowledge limitations in the types of peptides incorporated into this display system. With these limitations in mind, numerous attempts have been made to identify other cellular mechanisms to improve peptide incorporation into filamentous phage. Recently, a group employed the twin-arginine translocation (Tat) pathway to transport and integrate a particular protein into filamentous phage [28]. Central to this effort was the understanding that bacteria employ several pathways to generate and deliver proteins to designated cellular locations. Of these pathways, the secretory (Sec) pathway represents a general means by which proteins are exported to the periplasm [29]. In this pathway, proteins are exported to the periplasm in an unfolded state utilizing a signal peptide sequence. This signal sequence, generally composed of hydrophobic amino acids, allows for efficient targeting of Sec cellular machinery involved in peptide export [29].

As in the case of proteins designated for the Sec pathway, many proteins must interact with specific cellular machinery to achieve suitable final fold geometries. Proteins can be targeted exclusively for folding in the cytoplasm as well as for the extra-cytoplasmic space. To achieve this specific level of protein targeting, signal sequences are often encoded as part of the nascent peptide. Chaperones or other cellular machinery, in turn, recognize these signal sequences and direct protein folding and export. In the

case of the Tat pathway, proteins are delivered for translocation pre-folded. Thus, there exists a tight coupling between signal sequence and compartment-specific chaperones to ensure appropriate folding [29-33]. With this in mind, researchers speculated that some refractory proteins could be incorporated into filamentous phage via the Tat pathway given that proteins in this pathway would be pre-folded [28, 34]. In order to accomplish Tat pathway activation, a cleavable N-terminal signal peptide derived from proteins having Tat interaction was integrated into a reference protein and tracked [28, 33, 34]. The results of these experiments demonstrated that Tat targeting ensured proper peptide incorporation and display in filamentous phage.

Bacteriophage MS2, as a new platform for phage display, does not require periplasmic assembly to allow for virus-like particle formation. Instead, MS2 assembly occurs entirely in the cytoplasm. Preliminary recombinant experiments at the n-terminus of MS2 coat also uncovered refractory peptide sequences. Although MS2 coat, because of its cytoplasmic assembly, initially appears to be unsuitable for Tat-mediated peptide incorporation, Tat pathway activation could actually prove to be useful for peptide integration and display. Central to the Tat pathway is the requirement that proteins be delivered to this translocation system pre-folded. It is, therefore, not unreasonable to assume that other cellular machinery, including molecular chaperones, are involved in this process. A recent study looked at the role of molecular chaperones in the Tat pathway and identified a potentially significant role for DNAK. In this study, it was found that DNAK played an intimate role in protein folding prior to protein translocation via the Tat pathway [33]. Specifically, DnaK maintained the stability of tat-mediated proteins prior to translocation – even under conditions where Tat-specific machinery

were saturated [33]. This ability to promote efficient folding and to maintain proteins as "export-competent," represented a process with implications for rescuing misfolded proteins.

In other studies, DnaK has been shown to have binding specificity for arginine and lysine at its boundaries [35-37]. As suggested by its name, the twin arginine translocation pathway is distinguished by its hallmark twin arginine motif in the signal sequence that modulates it activation. Thus, by extension, it was proposed as part of this effort that sequences flanked by arginines would allow for efficient incorporation of refractory peptides into a MS2 VLP. Its was anticipated that by flanking a recombinant region with arginines, that DnaK would be recruited along with other co-factors to aid in nascent peptide folding and stabilization. With this mind, some mention must be made of DnaK peptide binding in general. We proposed to flank a recombinant region attached to MS2 coat protein with two arginines on each side because we believed that DnaK binding to this protein would be enhanced by their inclusion. Similar findings have been confirmed by several studies [35, 38]. Thus, we sought to demonstrate that the incorporation of randomized insertions at the N-terminus of RNA bacteriophage MS2 coat protein could be greatly enhanced through the selective, engineered addition of amino acids.

Approaches to the Suppression of Defects imparted by N-Terminal Peptide Fusion.

The N-terminus of bacteriophage MS2 coat protein has been our primary focus in developing a new location to display random peptide sequences. As experimental activities commenced for this effort, N-terminal attachment of peptides appeared to inhibit folding or generate assembly-incompetent clones. However, the strategic

placement of amino acids and the potential recruitment of molecular chaperones emerged as two new approaches to revert improperly folded recombinant proteins. As these two key technical approaches were being developed, other opportunities were also identified. One such opportunity related to efforts to display random insertions in filamentous phage.

Mosaic VLPs

Filamentous phage display systems can be classified based on the coat gene being exploited for presentation. For example type 88 vectors denote that a phage genome encompasses two types of pVIII coat molecules [1, 39, 40]. In the context of phage display, type 88 systems signify that one coat gene is recombinant and the other coat gene is wild-type. The virion that emerges from this system is a mosaic in that both coat molecules are present in the final particle. A variation on a type 88 phage display system is the type 8+8 system. This system produces a mosaic phage particle from VIII genes encoded on separate genomes.

In this case, the genome that encodes the wild-type coat sequence is called a helper phage[1]. The recombinant coat sequence, on the other hand, represents a unique type of plasmid called a phagemid [41, 42]. A key feature of this phagemid system is that this type of plasmid encodes a replication origin amenable to *E. coli* hosts and antibiotic resistance to aid in selection. Additionally, phagemids also encode the filamentous phage origin of replication. This origin is dormant until cell is infected with helper phage. Once both genomes are in the cell, both coat sequences are expressed in progeny virus particles. It was this dual genome approach that prompted laboratory efforts to design a system in which n-terminal recombinants would be co-expressed with

wild-type MS2 coat proteins. The expectation was that mosaic capsids would be produced and that this novel assembly pathway would allow poorly folded or assembly-deficient n-terminal MS2 coat recombinants to be rescued.

Genetic Selection for Stabilizing Bracket Sequences

As part of our first approach, we considered combinatorial approaches proposed in the literature to influence the incorporation of peptides. However, many of these approaches were specific to the phage systems from which they were derived. Additionally, their applicability to the challenges encountered in this MS2 coat effort may be limited. Masking the nature of peptide insertions emerged as a leading technical option. Again, this approach was inspired by the combinatorial approach applied in the generation of woodchuck hepadnavirus-like particles [9]. We extended the effort described by Billaud et al., and eventually identified flanking amino acid combinations capable of reverting insertion failures.

As a follow-up, a second, rather simplistic question emerged to offer a different perspective on the use of flanking amino acid residues. Could nature offer a better solution than simply employing propensity models or acidic sequences to stabilize peptide insertions? If so, how could nature, or more appropriately, natural selection yield a similar, or even better outcome? With this in mind, a new experimental approach was established in which a recombinant sequence known to induce folding failures was flanked by two randomized amino acid positions. The choice for randomizing two flanking positions was based on experimental data suggesting that this was the minimum configuration required to maintain insertion stability at the N-terminus. Taken in context,

this additional outcome could prove to be useful for *de novo* protein design considerations.

In the present study, we have identified and characterized amino acid sequences that promote the effective incorporation and display of randomized sequences in bacteriophage MS2 VLPs. We demonstrate that peptide design cues from nature can be systematically employed to achieve peptide incorporation. We also suggest a methodology to employ mosaic virus-like particles for presentation of refractory sequences. Finally, we suggest a methodology to recruit cellular molecular machinery to aid in the folding of recombinant peptide sequences.

MATERIALS AND METHODS

Plasmid Constructions

N-Terminal Insertions

Single-chain dimer versions of all N-terminal recombinant coat sequences were produced by duplication of the recombinant and wild-type coat sequences. The upstream half that included the recombinant region was produced via PCR amplification. In this case, 5'-primers annealing to pCT119dl5N upstream of the coat sequence, and a downstream primer creating a *BglI* restriction site at the 3' end of the coat coding sequence were employed. The generated PCR products were subsequently digested with *HindIII* and *BglI* to produce a fragment that contained the recombinant coat sequence with a new *BglI* site. The downstream half was synthesized using a primer that created a *BglI* site at the 5'-end of the coat coding sequence and a 3'-primer that annealed to plasmid sequences downstream of coat. This downstream fragment was then digested with *BglI* and *BamHI* -which is a downstream restriction site encoded in the plasmid. The upstream and downstream coat fragments were then joined by ligation to a vector fragment generated via *HindIII-BamHI* cleavage of pCT119dl5N, thereby creating the required single-chain dimer construction. This process was repeated for all 12 primer constructions to generate recombinant single-chain dimers. This arrangement is identical to p2CT-dl13, but encoding two amino acid substitutions to create the required *BglI* site [43].

The plasmid pACTdl5n, a pACYC-derived plasmid, incorporates the wild-type MS2 coat sequence and has expression directed from the T_7 promoter. pACET is a control plasmid that does not generate coat protein. pACTdl5N and pACET were

subsequently utilized to characterize the generation of mosaic MS2 coat capsids. pFCT (which has previously been described) was employed to generate Flag fusions under the control of the *lac* promoter [11]. pAFCT, which was constructed in a fashion similar to pFCT, expresses Flag fusions under the control of the t_7 promoter. To generate the randomized bracket configuration, the PCR primer 5-GGCCATGGCT$(NNY)_2$–(LNLVFILLSG)-$(NNY)_2$-GGCGGCGCTTCTAACTTTACTCAGTTC was employed. NNY (where N is A, C, G, or T and Y is T or C) was utilized to generate the randomized flanking sequences.

Briefly, the randomized N-terminal insertion was produced by PCR amplification using above mentioned 5'-primer that annealed at the *NcoI* site, and a 3'-primer downstream of a plasmid *BamHI* restriction site. This PCR product was then digested with *NcoI* and *BamHI*, gel-purified and ligated to a similarly digested vector fragment from pCT119dl5N. This PCR fragment was carefully recovered in ligated into a new vector, pP7K. To generate this new randomizing bracket configuration, we first digested plasmid pP7K with *SapI* and *BamHI*. Then, in a three-part ligation, we digested the recovered PCR fragments (derived from pCT119dl5N) with *HindIII* and *BamHIII*. As we still required a *SapI-HindIII* fragment to complete the ligation, we digested pCT119dl5N with *SapI* and *HindIII* to generate the final required fragment. The two insert fragments were ligated with the pP7K vector and transformed into strain CSH41F-containing plasmid pRZ5 and plated on LB medium containing X-Gal [14].

AB-Loop Insertions

Single-chain dimer versions of all AB-loop recombinant coat sequences were produced by duplication of the recombinant and wild-type coat sequences. The upstream

half that included the recombinant region was produced via PCR amplification. In this case, 5'-primers annealing to plasmid sequences upstream of the coat sequence in pMCTK2, and a downstream primer creating a *BglI* restriction site at the 3' end of the coat coding sequence were employed. The generated PCR products were subsequently digested with *HindIII* and *BglI* to produce a fragment that contained the recombinant coat sequence with a new *BglI* site. The downstream half was synthesized using a primer that created a *BglI* site at the 5'-end of the wild-type coat coding sequence and a 3'-primer that annealed to plasmid sequences downstream of coat. This downstream fragment was then digested with *BglI* and *BamHI* -which is a downstream restriction site encoded in the plasmid. The upstream and downstream coat fragments were then joined by ligation to a vector fragment generated via *HindIII-BamHI* cleavage of pMCT [12], thereby creating the required single-chain dimer construction. This arrangement is identical to p2CT-dl13, but encodes two amino acid substitutions to create the required *BglI* site [43]. The new construct is called p2MCTK5. It provides a single-chain dimer with a unique *KpnI* restriction site in the AB-loop of the upstream half.

Libraries of Random Sequence Peptides

N-Terminal Insertions

To insert random DNA sequences into the N-terminus, the primers described below were used to amplify a coat fragment from pCT119dl5N in eight different reactions. The 5'-primers that attach at an *NcoI* site were 5-ATGGCTAGCAGC-$(NNY)_{10}$AGCAGCGGCGGC-3, 5- ATGGCTACCACC-$(NNY)_{10}$ACCACCGGCGGC-3, 5- ATGGCTGCGGCG-$(NNY)_{10}$GCGGCGGGCGGC-3, 5- ATGGCTGGCGGC-$(NNY)_{10}$GGCGGCGGCGGC-3, 5- ATGGCTGACGAC-$(NNY)_{10}$GACGACGGCGGC-3,

5- ATGGCTGAGGAG-$(NNY)_{10}$GAGGAGGGCGGC-3, 5- ATGGCT-$(NNY)_{10}$GAGGAGGAGGAGGGCGGC-3, and 5- ATGGCTAGGAGG-$(NNY)_{10}$AGGAGGGGCGGC-3. NNY (where N is A, C, G, or T and Y is T or C) is utilized to generate the random intervening sequences. For modifications to serine brackets, the following primers were used: 5- ATGGCTAGCAGC-$(NNY)_{10}$AGCAGCGGCGGC-3, 5- ATGGCTAGCAGC$(NNY)_{10}$GGCGGC-3, 5-ATGGCT-$(NNY)_{10}$AGCAGCGGCGGC-3, and 5- ATGGCT$(NNY)_{10}$AGCGGCGGC-3. Each reaction described above also employed a single 3'-primer that annealed downstream of a *BamHI* site in the plasmid vector. The generated PCR products were subsequently digested with *NcoI* and *BamHI*, gel-purified and ligated to similarly digested vector fragments from pCT119dl5N.

AB-Loop Insertions

To insert random DNA sequences into the AB-loop, the primers described below were used to amplify a coat fragment from pMCT into six different recombinant reactions. Each reaction utilized a single 5'-primer that annealed upstream of a *HindIII* site in the plasmid vector. Six different 3'-primers (5-CCGGTACCGCTGCT$(RNN)_{10}$AGTTCC-GCCATT-3, 5- CCGGTACCATCATC-$(RNN)_{10}$AGTTCCGCCATT-3, 5-CCGGTACC-GCC$(RNN)_{10}$AGTTCCGCCATT-3, 5-CCGGTACCCGG$(RNN)_{10}$AGTTCCGCCATT-3, 5- CCGGTACCGCGGCG-$(RNN)_{10}$GCGGCGAGTTCCGCCATT-3, and 5- CCGGTACCGCGGCGGCGGCG-$(RNN)_{10}$GCGGCGGCGGCGAGTTCCGCCATT-3) attach at a *KpnI* site. NNY (where N is A, C, G, or T and Y is T or C) is utilized to generate the random intervening sequences. The resulting PCR products were digested with *KpnI* and *HindIII*, gel-

purified and ligated to similarly digested fragments of pMCTK2 or P2MCTK5. The resulting plasmids were then introduced by transformation into strain CSH41F- containing plasmid pRZ5, and plated on LB medium containing X-Gal [14].

Protein Expression, Purification, and Functional Assays

To test the recombinant proteins for translational repressor function, each plasmid was introduced into E. coli strain CSH41F– containing the translational repression reporter plasmid pRZ5 and plated on LB medium containing 5- bromo-4-chloro-3-indolyl-β-d-galactoside (X-Gal) [14]. Control ligations containing only vector DNA produced at least 1000-fold fewer colonies than ones with an inserted sequence. After overnight incubation, blue and white colony counts were assessed. Again, properly folded recombinant proteins repression translation of β-galactosidase and yield white colonies. For each plasmid construction, 24 white colonies and 12 blue colonies were picked to two different 1 mL cultures in LB medium and grown overnight with shaking at 37 °C. One set of cultures was employed for plasmid isolation and DNA sequence analysis.

The second set was then lysed by sonication and subjected to agarose gel electrophoresis. Note that after sequence analysis, a few white clones were dropped from the analysis. A few clones either lacked a peptide insertion or contained mutations at secondary sites that might complicate interpretation of the results. Assessment of a recombinant protein's ability to assemble into a VLP was performed by electrophoresis of sonicated cell lysates (from 1 mL overnight cultures) in gels of 1% agarose in 50 mM potassium phosphate, pH 7.5 [44]. Gels were then stained with ethidium bromide to intercalate with host RNAs to indicate the presence of VLPs. Verification of VLPs was

then confirmed by transferring gel contents to nitrocellulose. This was then probed with rabbit anti-MS2 serum and finally resolved with an alkaline phosphatase-labeled second antibody.

RESULTS: N-TERMINAL FUSIONS

Insertion of Randomized Peptides in the N-Terminus of MS2 Coat Protein

As illustrated in Figure 6, the N-terminus of RNA bacteriophage MS2 is located at the surface of the viral capsid. It was due to the potential to exploit this surface accessibility in immunological applications that previous research focused on this area for peptide insertion [11]. In a foundational study, the Flag peptide, which has been utilized to immunologically tag and track proteins, was successfully incorporated at the N-terminus of MS2 [11]. To facilitate N-terminal incorporation of the Flag peptide, a plasmid, pCT119dl5N, was constructed from pCT119 as previously described [11, 14]. It possesses an Nco I site at the translational start codon, and thus facilitates fusion of peptides to the N-terminus. As described in the Introduction, a previous study inserted the Flag peptide and this site. The protein seemed to fold properly, but was unable to assemble into VLPs. Utilization of the single-chain dimer largely corrected the assembly defect.

In the present study, pCT119dl5N was again utilized to generate fusion proteins at the N-terminus of MS2. Primers containing either a random ten amino acid sequence or a random ten amino acid sequence flanked by other amino acids were synthesized and subsequently used to generate recombinant proteins (Figure 7a,b). The resulting plasmids, pCT119dl5n-$(NNY)_{10}$,pCT119dl5n-XX$(NNY)_{10}$XX, and pCT119dl5n-$(NNY)_{10}$XXXX were employed to express all recombinant proteins for this study. X, in all cases, was represented by a specific amino acid suggested in the literature as providing potential benefit in this recombinant effort and are referenced accordingly (Fig. 3b). The NNY triplet (N= any nucleotide, Y= T or C) was utilized to generate the

intervening random peptide region because it produces 15 out of the 20 standard amino acids. Although Met, Gln, Trp, Lys, or Glu are not produced in NNY libraries, this approach avoided the incorporation of stop codons, which obviously would have resulted in failure to produce the peptide-coat fusion.

Tests of translational repression demonstrated that the vast majority of $(NNY)_{10}$ random insertions at the N-terminus were essentially repressor-defective, presumably indicating interference with proper coat protein folding. However, the selective incorporation of amino acids in the form of $XX(NNY)_{10}XX$, and $(NNY)_{10}XXXX$ generally produced recombinants that were effective in repressing β-galactosidase production. These results were expressed as a percentage of clones repressing translation (Table II). The beneficial effect of the incorporation of either glutamate or aspartate brackets (Table II) is consistent with the results of Ballaud et al. obtained earlier with hepadnavirus VLPs (see ref. [9] and the Introduction). The presence of glutamate or aspartate brackets increased the tolerance of a 10-mer insertion from a baseline of 2% functional repressors to a high of almost 80%.

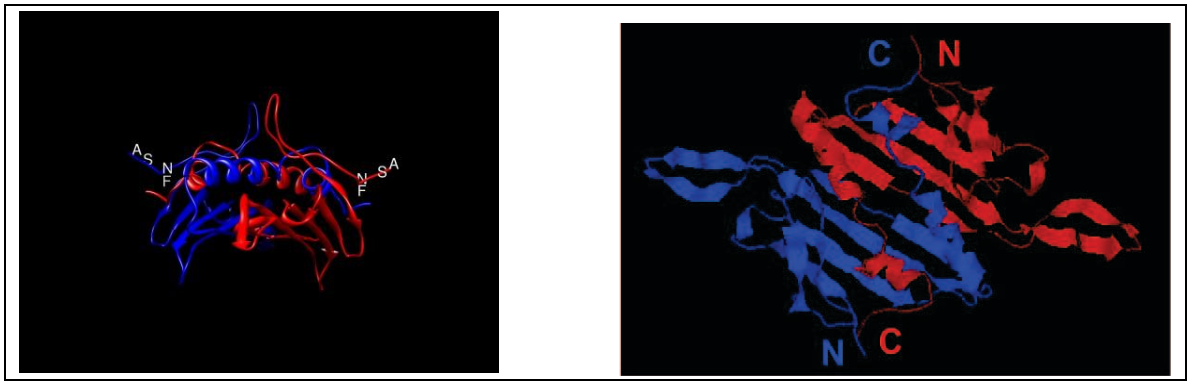


Figure 6. Alternative view of MS2
(a) The structure of the Bacteriophage MS2 coat protein dimer with polypeptide chains highlighted in blue and red. The N-terminus for each polypeptide is designated by the first four amino acids in its chain. (b) Structure of the MS2 coat protein dimer. The proximity of the N-terminus of one coat monomer to the C-terminus of the other coat monomer allows the single-chain dimer to be constructed.

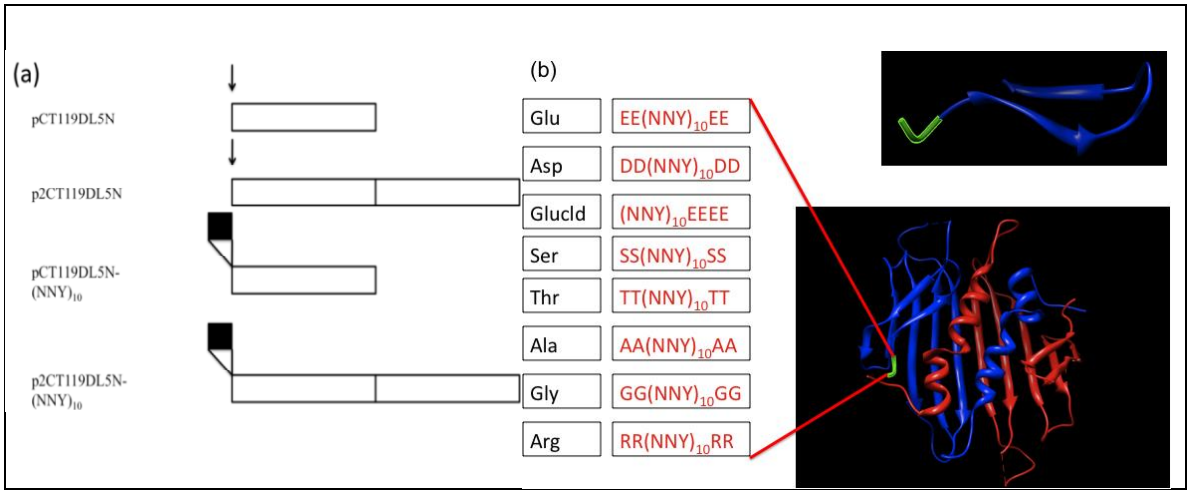


Figure 7. Summary of N-terminal efforts
(a) Arrangements of the coat protein reading frames on the plasmids used in this experiment. All express coat protein from the *lac* promoter. Plasmid p2CT119DL5N expresses a single-chain dimer version of the protein. Black boxes represent the peptide insertions made in each plasmid configuration. (b) Oligonucleotides utilized to make insertions in each plasmid configuration.

Table II. N-terminal monomer functional repressors.

The percentages of 10-mer N-terminal insertions that result in functional translational repressors. Results are shown for monomer configuration (white colonies on X-Gal plates).

Monomer Platform		Percent White Colonies
pUC119	Control	0
pCT119dl5n (pC5n)	Control	100
pC5n- $(NNY)_{10}$	Random Insertion	2
pC5n- $EE(NNY)_{10}EE$	+Glutamate Brackets	65
pC5n- $DD(NNY)_{10}DD$	+Aspartate Brackets	79
pC5n- $(NNY)_{10}EEEE$	+Glutamate Cloud	65
pC5n- $SS(NNY)_{10}SS$	+Serine Brackets	70
pC5n- $TT(NNY)_{10}TT$	+Threonine Brackets	18
pC5n- $AA(NNY)_{10}AA$	+Alanine Brackets	2
pC5n- $GG(NNY)_{10}GG$	+Glycine Brackets	5
pC5n- $RR(NNY)_{10}RR$	+Arginine Brackets	77

Controls are pUC119, which produces no coat protein. Plasmid pCT119dl5n (pC5n) produces wild-type coat protein.

Table III. N-terminal dimer functional repressors.

The percentages of 10-mer N-terminal insertions that result in functional translational repressors. Results are shown for single-chain dimer configuration (white colonies on X-Gal plates).

Single-Chain Dimer Platform		Percent VLP Formation
pUC119	Control	0
p2M	Control	100
p2M- $(NNY)_{10}$	Random Insertion	8
p2M - $EE(NNY)_{10}EE$	+Glutamate Brackets	88
p2M - $DD(NNY)_{10}DD$	+Aspartate Brackets	79
p2M - $(NNY)_{10}EEEE$	+Glutamate Cloud	70
p2M - $SS(NNY)_{10}SS$	+Serine Brackets	88
p2M - $TT(NNY)_{10}TT$	+Threonine Brackets	42
p2M - $AA(NNY)_{10}AA$	+Alanine Brackets	7
p2M - $GG(NNY)_{10}GG$	+Glycine Brackets	39
p2M - $RR(NNY)_{10}RR$	+Arginine Brackets	94

Controls are pUC119, which produces no coat protein. Plasmid p2Mproduces a wild-type coat dimer protein.

The other amino acid choices listed in Table IV reflected the culmination of an extensive literature review on the incorporation of amino acids in nascent peptides [15, 16, 18, 19, 22, 25]. Serine and arginine emerged as two other amino acid partners that allowed randomly generated 10-mer sequences to repress translation. In the cases of serine and arginine, baseline repression increased from 2% in the random 10-mer to 70% and 77% respectively.

Incorporating Peptide Fusions into the Single-Chain Dimer

Although several bracket configurations allowed β-galatosidase repression, none of the recombinants constructed as fusions to wild-type coat protein were competent for VLP formation (Table II). We suspect that this failure of apparently properly folded coat proteins to assemble into VLPs is due to crowding at local 3-fold symmetry axes where N-termini of three adjacent coat proteins come into close proximity, a problem which is reduced when fusions are introduced into the single-chain dimer. As illustrated by the results, a striking improvement in repression was observed for N-terminal recombinants when the single-chain dimer was incorporated in conjunction with brackets (Table III). With the noted exceptions of the baseline random 10-mer insertion and alanine brackets, percent white colony counts improved across all recombinant variations.

MS2 VLP Formation

A properly assembled VLP stains with ethidium bromide and reacts with coat-specific antiserum. To monitor VLP formation, sonicated cell lysates from individual clones were subjected to agarose gel electrophoresis. VLPs were visualized by staining with ethidium bromide –which intercalates with and stains the RNA they encapsidate. Gels were also blotted to nitrocellulose and probed with anti-MS2 serum and alkaline

phosphatase conhugated anti-rabbit IgG [11, 45]. Note that VLP mobility varies with changes in surface charge associated with the insertion of peptides and flanking brackets (Figure 8).

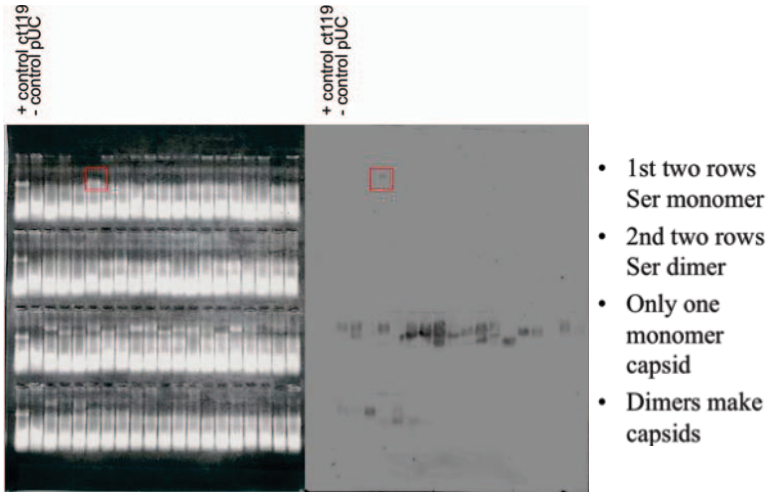


Figure 8. Serine bracketed VLP formation
Electrophoresis on agarose gel of VLPs identified in lysates of cells producing recombinant coat proteins. The first two lanes of each gel are controls: pCT119 produces wild-type coat, while pUC produces no coat protein. An Ethidium bromide-stained gel (left panel) and western blot probed with MS2 anti-serum (right panel) for serine-bracketed recombinants are shown.

Table IV. N-terminal monomer VLP formation

The percentages of 10-mer N-terminal insertions that suggest positive VLP formation in monomer (positive ethidium bromide staining in capside gel, western blot for MS2 coat).

Monomer Platform		Percent VLP Formation
pUC119	Control	0
pCT119dl5n (pC5n)	Control	100
pC5n- $(NNY)_{10}$	Random Insertion	0
pC5n- $EE(NNY)_{10}EE$	+Glutamate Brackets	0
pC5n- $DD(NNY)_{10}DD$	+Aspartate Brackets	0
pC5n- $(NNY)_{10}EEEE$	+Glutamate Cloud	0
pC5n- $SS(NNY)_{10}SS$	+Serine Brackets	4
pC5n- $TT(NNY)_{10}TT$	+Threonine Brackets	0
pC5n- $AA(NNY)_{10}AA$	+Alanine Brackets	0
pC5n- $GG(NNY)_{10}GG$	+Glycine Brackets	4
pC5n- $RR(NNY)_{10}RR$	+Arginine Brackets	0

Controls are pUC119, which produces no coat protein. Plasmid pCT119dl5n (pC5n) produces wild-type coat protein.

Table V. N-terminal dimer VLP formation.

The percentages of 10-mer N-terminal insertions that suggest positive VLP formation in dimer (positive ethidium bromide staining in capside gel, western blot for MS2 coat).

Single-Chain Dimer Platform		Percent VLP Formation
pUC119	Control	0
p2M	Control	100
p2M- $(NNY)_{10}$	Random Insertion	0
p2M - $EE(NNY)_{10}EE$	+Glutamate Brackets	4
p2M - $DD(NNY)_{10}DD$	+Aspartate Brackets	4
p2M - $(NNY)_{10}EEEE$	+Glutamate Cloud	4
p2M - $SS(NNY)_{10}SS$	+Serine Brackets	100*
p2M - $TT(NNY)_{10}TT$	+Threonine Brackets	4
p2M - $AA(NNY)_{10}AA$	+Alanine Brackets	0
p2M - $GG(NNY)_{10}GG$	+Glycine Brackets	4
p2M - $RR(NNY)_{10}RR$	+Arginine Brackets	0

Controls are pUC119, which produces no coat protein. Plasmid p2Mproduces a wild-type coat dimer protein. *Greatest impact on VLP formation

Characterizing the Effects of Amino Acid Brackets

Given that di-serine brackets produced the highest overall white colony count, and yielded the highest number of VLPs, we randomly selected 5 blue recombinants for sequence analysis to determine whether their failure to repress translation was indeed due to a protein folding failure or to some defect in their construction that might fail to yield the desired fusion protein. In a previous effort, we noticed that some reportedly repressor-defective clones in a phage library were the results of oligonucleotide synthesis errors that yielded frameshift mutations or of aberrant ligation events [12]. In fact, four of the five blue di-serine bracket clones proved to contain frameshift mutations (results not shown). Therefore, again, a noteworthy percentage of repressor-defective clones in this library were not the consequence insertion failures, but rather due to other defects.

Determining the Importance of Individual Serines in the Brackets.

To determine the importance of individual serines in the serine bracket configuration, serines were selectively removed and the functional consequences were determined (Figure 9). As suggested by the data, serine change 42 ($ATGGCT(NNY)_{10}AGCGGCGGC$ in dimer) appeared to have the greatest impact on the folding behavior of N-terminal recombinant proteins (see Table VI). Serine bracketed recombinants in the single-chain dimer started with a baseline of 88% repression effectiveness. The change imparted by the 42-configuration moved repression downward to a low of only 1%. Thus, the 42-configuration yielded a dramatic downward impact on recombinant folding behavior. With regard to assembly, serine change 12 ($ATGGCTAGCAGC(NNY)_{10}GGCGGC$ in dimer) had a dramatic impact on VLP formation for N-terminal extensions (Table VI). In this case, serine bracketed

recombinants started with a baseline of 100 % VLP formation. However, once serines were removed to reflect the 12-configuration, VLP formation fell from 100% to low of 4%.

Figure 9. Codons for serine modification study
Codons of the oligonucleotides utilized to make insertions in each plasmid configuration. S3 □T□□E□□H□C□H□□I□□H□7A□C□□□HC□□□□A□C□JA□□I□□□JA□O□□□I□7B□□CH□1□□□CH□□C□TS□ denotes the changes made to the original serine bracketed insertion.

Table VI. VLP formation for serine modifications.
The percentage of serine-bracketed N-terminal insertions suggesting proper folding (white clones) and proper assembly (VLP formation). SRX1 defines the monomer baseline configuration, while SRX2 identifies the single-chain dimer baseline. SR42 has the greatest impact on folding behavior, while SR12 indicates a dramatic impact on VLP formation.

Serine Changes	Percent White clones	Percent VLP Formation
SerX1 Monomer	70	4
SR11	46	0
SR31	61	0
SR41	49	4
SerX2 Dimer	88	100
SR12	92	4**
SR32	99	96
SR42	1**	83
**denotes greatest net change in baseline protein behavior		

RESULTS: PEPTIDE INSERTIONS IN THE UPSTREAM AB-LOOP OF THE SINGLE-CHAIN DIMER

Insertion of Engineered Randomized Peptides in the AB-loop of MS2 Coat Protein

As illustrated (Figure 10), the AB-loop of RNA bacteriophage MS2 is located at the surface of the viral capsid, an ideal site for peptide display. In fact, recent efforts to exploit this surface accessibility and incorporate peptides for immunological applications have shown considerable promise [12]. In this 2008 study by Peabody et al., it was demonstrated that a high frequency of peptide insertion tolerance (greater than 95%) was obtained when the insertions were confined to the AB-loop of the downstream half of a single-chain dimer. Insertions in the AB-loop of wild-type coat protein, or in both halves of the single-chain dimer nearly always resulted in a sever folding defect. To facilitate insertions a Kpn I site was produced by silent mutations in codons 14 and 15 [11, 14]. In this case insertion introduced a duplication of codons 14 and 15 within the AB-loop – resulting in 12, not 10 amino acid inserted.

PCR primers containing either a random ten amino acid sequence or a random ten amino acid sequence flanked by the selective incorporation of other amino acids were synthesized and subsequently used to generate recombinant proteins for the present effort (Figure 11a,b). Literature reviews had suggested the use of structure-terminating amino acids, i.e. “stereochemical punctuation marks” or “gatekeepers” to stop the propagation of secondary structures [22, 25, 46, 47]. Our premise for this effort was that the introduction of random sequences potentially extended or introduced nascent secondary structures. Therefore, it seemed logical to employ amino acids shown to cap or provide termination of structure as these insertions were made into the 5’ loop of MS2. The

resulting plasmids, pMCTK2-$(NNY)_{10}$,pMCTK2-$(NNY)_{10}$XX, pMCTK2-XX$(NNY)_{10}$XX and pMCTK2-XXXX$(NNY)_{10}$XXXX were employed to express all recombinant proteins for this study. X, in all cases, was represented by a specific amino acid suggested in the literature as providing potential benefit in this recombinant effort and are referenced accordingly (Figure 11b). The NNY triplet (N= any nucleotide, Y= T or C) was utilized to generate the intervening random peptide region because it produces 15 out of the 20 standard amino acids. Although Met, Gln, Trp, Lys, or Glu are not produced in NNY libraries, this approach avoided the incorporation of stop codons and consequently limited the inclusion of truncated sequences in downstream analyses.

The □-galactosidase repression assay was again utilized to assess the folding behavior of recombinants coat sequences. Functional tests demonstrated that $(NNY)_{10}$ random insertions in the AB-loop of the wild-type protein were essentially all repressor defective. Additionally, the selective incorporation of amino acids yielded no recombinants that were effective in repressing □-galactosidase production.

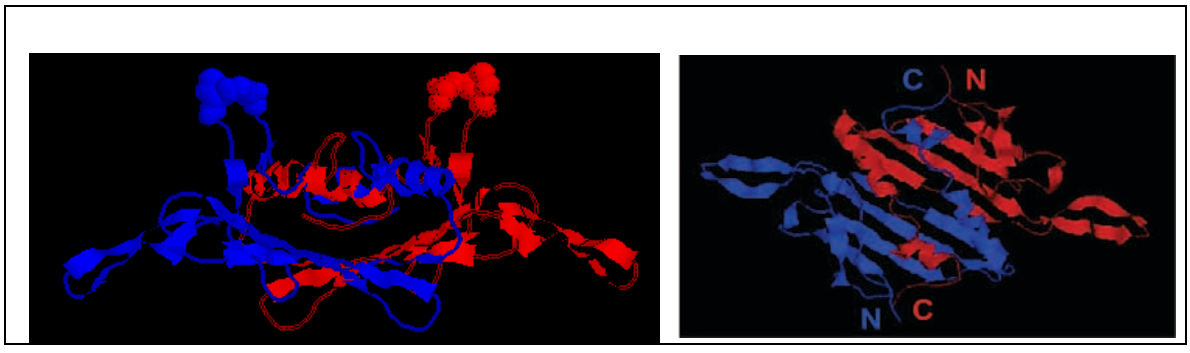


Figure 10. AB-loop of MS2
(a) The structure of the Bacteriophage MS2 coat protein dimer, with polypeptide chains highlighted in blue and red. The AB-loop for each polypeptide is designated by the enlarged, filled amino acids in its chain. (b) Structure of the MS2 coat protein dimer. The closeness of the N-terminus of one coat monomer to the C-terminus of the other coat monomer allows for the single-chain dimer to be constructed.

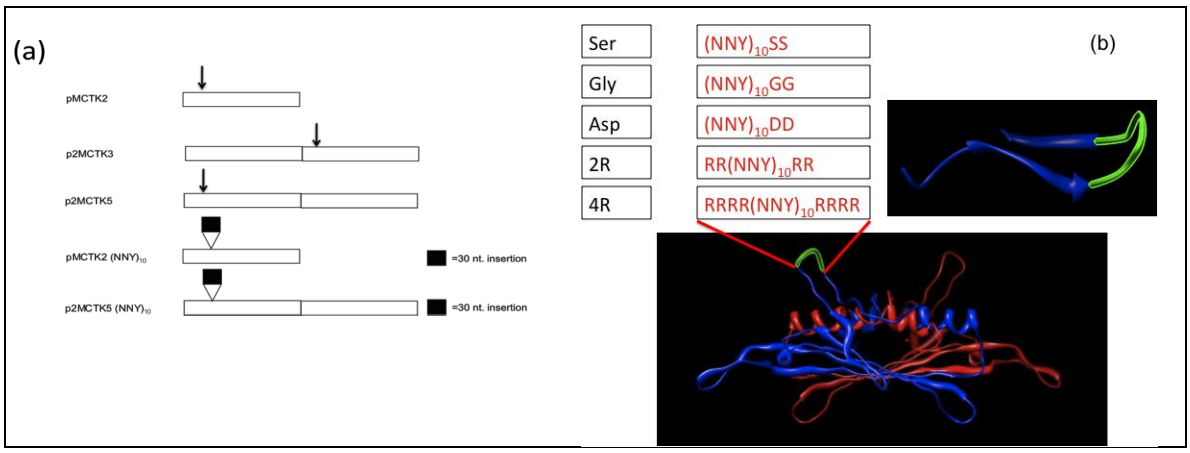


Figure 11. Recombinant insertions in AB-loop

(a) Arrangements of the coat protein reading frames on the plasmids used in this experiment. All express coat protein from the *lac* promoter. Plasmid pMCTK is similar to the previously described pCT119[11, 14], except for *KpnI* site that also incorporates the coat protein initiation codon (delineated by arrow). Plasmid p2MCTK5 expresses a single-chain dimer version of the protein. Black boxes represent the peptide insertions made in each plasmid configuration. (b) Oligonucleotides utilized to make insertions in each plasmid configuration. As AB-A□ □ E□H□I□□ C□H□ J□□□ 7V□ □□A C□JA□ □ I□ □□I □ generate recombinant inserts, the orientation of the nucleotides and random insert reflect an anti-sense orientation.

Table VII. AB-loop monomer functional repressors.

The percentages of 10-B □□□ □ V-loop insertions that result in functional translational repressors (white colonies on X-Gal plates).

Monomer Platform		Percent White Colonies
pUC119	Control	0
pMCTK2	Control	100
pMCTK2-$(NNY)_{10}$	Random Insertion	ND
pMCTK2-$(NNY)_{10}$SS	+Serine	0
pMCTK2-$(NNY)_{10}$DD	+Aspartate	0
pMCTK2-$(NNY)_{10}$GG	+Glycine	2
pMCTK2-RR$(NNY)_{10}$RR	+2Arginine	1
pMCTK2-RRRR$(NNY)_{10}$RRRR	+4Arginine	3

Controls are pUC119, which produces no coat protein and, as a result, yields no repression. Plasmid pMCTK2 produces wild-type coat protein and explicitly represses translation. ND= not detected

Folding Defects are Suppressed by Incorporation Into the Single-Chain Dimer

It was previously demonstrated that genetic fusion of two coat protein chains can efficinetly suppress folding/stability defects imparted by peptides inserted into the AB-loop when they are confined to the C-terminal half of the single-chain dimer [12]. Here we sought to determine whether the single-chain dimer similarly accommodates insertions in the N-terminal half of the single-chain dimer. We also wanted to determine whether bracketing amino acids affect the protein's tolerance of insertions. To facilitate peptide insertion, we constructed plasmid p2MCTK5. This plasmid is also similar to a previously described plasmid, p2CTdl13, and relates to pMCTK2 in its construction [11, 14]. However, in this case, a *KpnI* insertion site is generated in the 5' or upstream loop only. Implementation of the single-chain dimer increased repression analysis significantly (Table VII). Engineered serine and quad arginine recombinant generated the highest number of repression-competent clones.

Table VIII. AB-loop dimer functional repressors.
The percentages of 10-mer 5' AB-loop insertions that result in functional translational repressors (white colonies on X-Gal plates).

Single-Chain Dimer Platform		Percent White Colonies
pUC119	Control	0
p2MCTK5	Control	100
p2MCTK5-$(NNY)_{10}$	Random Insertion	41
p2MCTK5-$(NNY)_{10}$SS	+Serine	80
p2MCTK5-$(NNY)_{10}$DD	+Aspartate	74
p2MCTK5-$(NNY)_{10}$GG	+Glycine	51
p2MCTK5-RR$(NNY)_{10}$RR	+2Arginine	71
p2MCTK5-RRRR$(NNY)_{10}$RRRR	+4Arginine	86

Controls are pUC119, which produces no coat protein and, as a result, yields no repression. Plasmid p2MCTK5 produces a dimerized wild-type coat protein and explicitly represses translation.

Engineered Single-Chain Dimer Insertions Generate Recombinant MS2 VLPs

A key outcome of this experimental effort was the expectation that the introduction of AB-loop insertions would not prevent VLP assembly. To monitor VLP formation, individual white or blue clones were obtained and subjected to analysis. First, sonicated cell lysates from individual clones were subjected to agarose gel electrophoresis. Next, these lysates were visualized with ethidium bromide staining. This step was critical in that ethidium bromide intercalates with the nucleic acid normally encapsidated by properly assembled virus-like particles. Although not definitive, ethidium bromide staining established early evidence of VLPs that incorporated insertions at the 5' upstream loop of MS2.

Once ethidium staining was completed, we then assessed VLP formation by Western blot. Western blot analysis involved staining a blot with anti-MS2 serum, followed by a resolution step with the application of alkaline phosphatase-conjugated anti-rabbit IgG. As expected, AB-loop insertions in the dimer configuration did yield VLPs (Table VIII). The results were expressed as a percentage of white clones that yielded VLPs over the total number of white clones. In the single-chain dimer, the serine termination cap yielded the highest percentage of VLP formation (Table VIII). Other engineered sequences failed to yield VLP formation higher than that of the baseline insert only condition. With recombinant MS2 VLPs being made up different amino acids, their electrophoretic mobilities can be employed to further ascertain the presence of unique recombinant VLPs (Figure 12). The VLPs generated by serine terminal capped recombinants demonstrated the anticipated variable pattern in mobility.

Table IX. AB-loop VLP formation

The percentages of 10-mer AB-loop insertions that suggest positive VLP formation in dimer (positive ethidium bromide staining in capside gel, western blot for MS2 coat).

Dimer Platform		Percent VLP Formation
pUC119	Control	0
p2MCTK5	Control	100
p2MCTK5-$(NNY)_{10}$	Random Insertion	42
p2MCTK5-$(NNY)_{10}$SS	+Serine	75
p2MCTK5-$(NNY)_{10}$DD	+Aspartate	54
p2MCTK5-$(NNY)_{10}$GG	+Glycine	4
p2MCTK5-RR$(NNY)_{10}$RR	+2Arginine	30
p2MCTK5-RRRR$(NNY)_{10}$RRRR	+4Arginine	8

Controls are pUC119, which produces no coat protein. Plasmid pCT119dl5n (pC5n) produces wild-type coat protein.

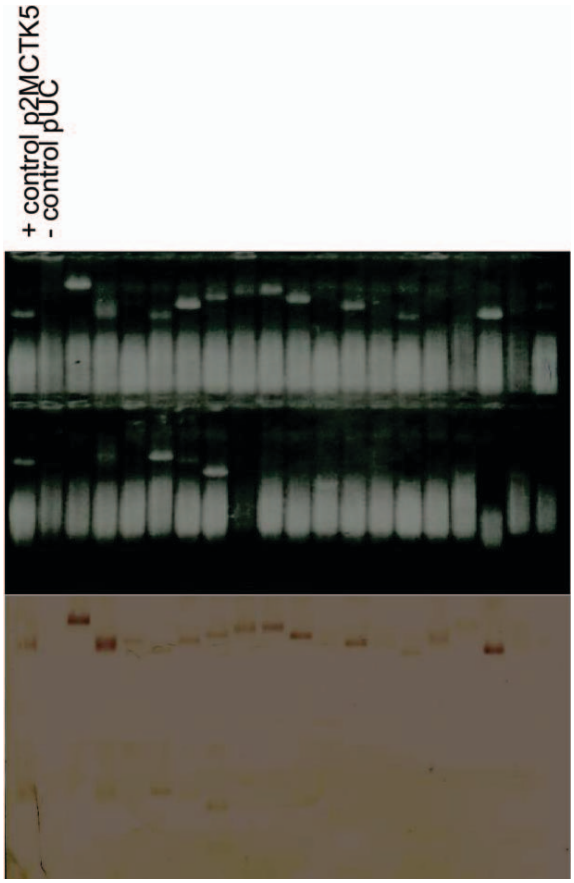


Figure 12. Termine serines in AB-loop produce VLPs
Electrophoresis on agarose gel of VLPs identified in lysates of cells producing recombinant coat proteins. The first two lanes of each gel are controls: p2MCTK5 produces wild-type coat, while pUC produces no coat protein. An ethidium bromide-stained gel (left panel) and western blot probed with MS2 anti-serum (right panel) for serine-bracketed recombinants are shown.

RESULTS: TARGETED APPROACHES TO N-TERMINAL PEPTIDE DISPLAY

Co-Expression of Flag-Coat Fusions with Wild-Type Coat Produce Mosaic VLPs

Co-expression of proteins to produce a mosaic virus-like particle represents a common approach for filamentous phage peptide display. However, in the case of bacteriophage MS2, such an effort has not been attempted. In anticipation of fold and assembly problem with the introduction of a foreign sequence into the coat of MS2, we proposed that co-expression could offer an alternative approach to address these challenges. In a previous study, the Flag peptide was successfully incorporated into the N-terminus of MS2 [11]. However, attempts to replicate the study yielded conflicting results. In the present effort, two plasmids, pFCT and pAFCT, were employed to generate Flag-coat fusions (Figure 13). Results of the repression assay in this effort again demonstrated that the incorporation of the Flag peptide in both plasmid configurations [illegible]-galactoside synthesis [11].

However, as VLP assembly was assessed, N-terminal constructions incorporating the Flag peptide failed to produce viable capsid-like particles. As an alternative technical approach, we then sought to generate mosaic capsids in which Flag fusion monomeric proteins could potentially be displayed. The plasmid pCT119dl5n was again utilized to produce fusion proteins at the N-terminus of MS2. Next, a second, inducible plasmid (pACTdl5n) was constructed to support co-expression and eventually VLP assembly. pACTdl5n, a pACYC-derived plasmid, incorporates the wild-type MS2 coat sequence and has expression directed from the T_7 promoter. To induce protein synthesis on this plasmid, isopropyl-[illegible]D -thiogalactopyranoside, IPTG was employed indirectly. IPTG,

which is an analogue of lactose, has the ability to displace the *lac* repressor from the *lac* operator. In our *E. coli* host, T_7 polymerase is under the control of a lac promoter and operator on the host chromosome.

Under normal conditions, this polymerase is effectively shut off via *lac* repressor interaction. Once induced, however, host-derived T_7 polymerase is synthesized and its interaction with the T_7 promoter on our inducible plasmid then directs coat synthesis. In short, this inducible system allowed for very strong promoter to be employed without sacrificing tight gene regulatory control. The resulting plasmid constructs, pCT119dl5n-Flag (pFCT) and pACT119dl5n-Flag (pAFCT), were employed to express all recombinant proteins for this study (Figure 13). As in the original study by Peabody, neither plasmid incorporating the Flag peptide produced VLPs. To potentially remedy these apparent assembly failures, a co-expression protocol was then employed to potentially generate mosaic capsids.

In order to generate these mosaic capsids, two co-expression scenarios were created. In the first scenario, wild-type MS2 coat was co-expressed with an inducible Flag fusion coat sequence. For the second scenario, a Flag fusion coat construct was co-expressed with an inducible wild-type coat construction. Next, to verify that mosaic VLP particles were actually being generated, a few control conditions were also established (Table X). First, a wild-type coat sequence on a non-inducible plasmid was co-expressed with an inducible plasmid that carried no coat sequence. Second, an inducible plasmid that carried wild-type coat protein was co-expressed with a non-inducible plasmid that carried no coat information. In each case, it was anticipated that VLPs would be formed and that a baseline for their electrophoretic mobility in an agarose gel assay would be

established. These two control conditions would be key as we expected that the net charge associated with Flag would impart a different electrophoretic mobility to mosaic VLPs as compared to either of these control conditions. Thus, we would know immediately by a mobility shift if mosaic capsids were being generated. Table IX highlights all co-expression experiments and associated control conditions.

To test for electrophoretic mobility, agarose gel electrophoresis was performed. Individual white clones were selected from each experimental test condition and subjected to analysis. Five co-expression test cases yielded VLPs (Figure 14). Three of these cases were representative of a control condition in which wild-type coat protein was generated. The other two test cases represented the heart of this co-expression experiment. Co-expression of a wild-type coat with an inducible Flag fusion (CT119dl5n/pAFCT) generated a VLP as per capsid gel analysis. However, its mobility in the gel did not differ from that of wild-type coat protein essentially generated via three different means (Figure 14). Co-expression of a non-inducible Flag fusion coat with an inducible wild-type coat protein generated a capsid as well (pFCT/pACT119dl5n). However, its mobility in agarose differed significantly from that of baseline wild-type coat protein (Figure 14). In the other remaining test cases where Flag fusions were expressed in the absence of wild-type coat or expressed via both plasmids, capsid gels failed to demonstrate the presence of VLPs.

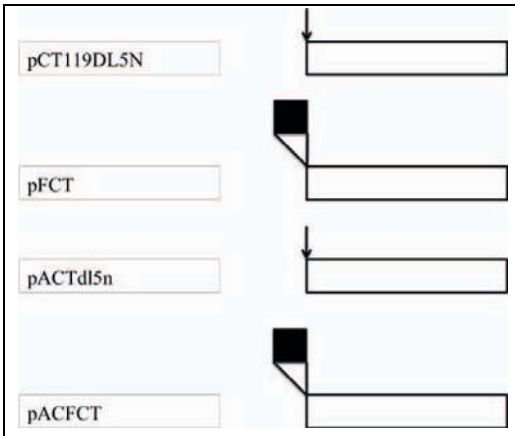


Figure 13. N-terminal Flag insertions
Plasmid pCT119DL5N is similar to the previously described pCT119[11, 14], except for *NcoI* site that also incorporates the coat protein initiation codon (delineated by arrow). This plasmid expresses coat protein from the *lac* promoter. Plasmid pACTdl5n, a pACYC-derived plasmid, expresses coat from the T_7 promoter. Black boxes represent the Flag insertions made in each plasmid configuration.

Table X. Co-expression partners for mosaic VLPs
Non-induced plasmids are under the control of the *lac* promoter. Inducible plasmids are under the control of the t_7 promoter. In each case, a non-inducible plasmid and inducible plasmid are co-transformed into *E. coli* and grown. Colonies were picked and cultured in LB media with chloramphenicol and ampicillin to mid-log phase. IPTG was then added and cultures were grown for an additional 3 hours. Cultures were then pelleted and subjected to a capsid gel protocol. pUC and pACET are control plasmids that do not generate coat protein.

Table X. Matrix of co-expression partners generated in this experiment.

	Inducible	pACET	pACTdl5n	pAFCT
Non-inducible				
pUC		X	X	X
pCT119dl5n		X	X	X
pFCT		X	X	X

Once ethidium bromide staining was completed, the gels were then employed via nitrocellulose transfer to generate Western blots. Again, Western blots detect the presence of coat protein and this information can be correlated with ethidium staining to

establish the presence or absence of VLPs. Western blot analysis confirmed the presence of MS2 protein in the same 5 co-expression test cases. In the co-expression test case pFCT/pACTdl5n, we see the blot reflects the mobility shift of a VLP that apparently incorporates both coat proteins (Figure 26, right panel). In a final step, a low-melt agarose gel was run for all generated proteins (Figure 26). In this gel, bands demonstrating positive ethidium bromide staining and correlating to the original bands in the agarose gel were excised. These recovered bands were then slowly melted and subjected to SDS-polyacrylamide electrophoresis. Results indicated the production of wild-type coat protein in all lanes. In the lanes with two bands, the Western blot reflects the incorporation of two proteins, presumably the ones of interest.

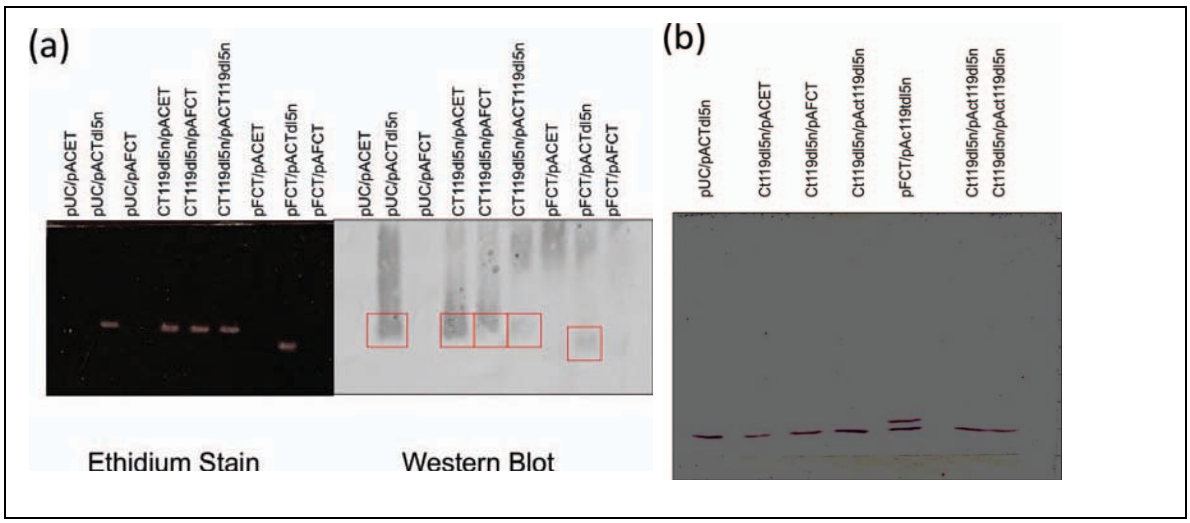


Figure 14. Characterizing mosaic virus-like particles
(a) Agarose gel electrophoresis assay (capsid gel) and Western blot of co-expression partners in approach to generate mosaic capsids. Given that VLPs contain RNA, they are readily visualized with ethidium bromide (left panel). In Western blot, coat protein was visualized using anti-MS2 anti-serum (right panel). (b) SDS-polyacrylamide gel electrophoresis and Western blot analysis of proteins produced by co-expression experiments. Proteins were visualized using anti-MS2 anti-serum. pUC and pACET are control plasmids that do not generate coat protein. The bottom bands reflect expression of wild-type coat protein. In the lane with two bands, the top band reflects the production of Flag fused to coat protein.

Attempts to Maximize Tolerance of N-terminal Fusions By Randomizing Bracket Sequences

Random inserts at the N-terminus of MS2 are highly destabilizing. In the body of this research effort, several attempts have been employed to stabilize these insertions and allow for eventual VLP formation. One approach that included the use of flanking serines has, thus far, offered the greatest opportunity for VLP assembly. As this approach was a direct attempt to resolve folding failures, a simplified, subtler hypothesis emerged to address the same protein-folding problem. In this new approach, a sequence known to result in a folding failure at the N-terminus was flanked by two randomized amino acid positions. By doing so it was expected that Nature (natural selection) would provide a more optimized solution for generating N-terminal MS2 recombinants. The choice of having two randomized positions was based on the same serine study in which flanking serines significantly improved n-terminal insertions. As an additional part of this serine study, the configuration of serines that flanked a randomized sequence was incrementally changed. In this extended effort, it was shown that by simply removing serines either upstream or downstream of a randomized insert resulted in changes to overall folding or assembly behavior. With this information in hand, it became clear that the two serines located on either side of a randomized sequence offered significant benefits. Thus, in the present effort, it was hypothesized that two randomized positions offered the greatest potential benefit in yielding new N-terminal recombinants.

In order to facilitate N-terminal incorporation of this specialized peptide, a plasmid, pCT119dl5n, was constructed from pCT119 as previously described [11, 14]. We then used a PCR primer encoding the known ten amino acid sequence flanked by two

randomizing amino acid locations to generate a recombinant coat sequence. This produced the resulting plasmid, pCT119dl5n- $(NNY)_2$ –(LNLVFILLSG)-$(NNY)_2$. The NNY triplet (N= any nucleotide, Y= T or C) was utilized to generate the two randomized amino acid locations that flanked the intervening known peptide region because it produces 15 out of the 20 standard amino acids. Although Met, Gln, Trp, Lys, or Glu are not produced in NNY libraries, this approach avoided the incorporation of stop codons and consequently limited the inclusion of truncated sequences in downstream analyses.

As initial efforts began to progress, it became clear that this plasmid configuration was generating an unacceptably high background of non-recombinant clones. When creating a vector for ligation of oligonucleotide products, there inevitably remains some uncut plasmid even after restriction digest. We suspected that this was the source of the high background of non-recombinant clones. Under normal conditions, the relatively low background would be experimentally acceptable. However, even under increased restriction endonuclease conditions, non-recombinants clones dominated experimental recovery. We then decided that an new approach would be required in which the PCR products generated with pCT119dl5n would be carefully re-separated via gel electrophoresis. Next, the resulting product would be transferred into a new plasmid that limited the potential to propagate non-recombinant clones forward.

To aid in this specialized peptide insertion, we utilized another previously described plasmid pP7K [48]. This plasmid generates the coat for *Pseudomonas* phage pP7. As we were trying to isolate very rare natural selection events, it was important to efficiently identify colonies that contained our insertion of interest. pP7K under normal conditional generates blue colonies in a coat repression assay designed for MS2. We,

therefore, digested pP7K with *SapI* and *BamHI*. In a three-part ligation, we then digested the recovered PCR fragments with *HindIII*. As we still required a *SapI-HindIII* fragment, we digested pCT119dl5N with *SapI* and *HindIII* to generate the final required fragment. The two insert fragments were ligated with the pP7K vector and transformed into strain CSH41F- containing plasmid pRZ5 and plated on LB medium. Although this step seemed trivial, we felt the step necessary as we were trying to isolate what we believed to be statistically hyper-infrequent randomization events. The resulting plasmid, pP7K-$(NNY)_2$–(LNLVFILLSG)-$(NNY)_2$, was then utilized to generate additional n-terminal recombinant proteins. Repression analysis yielded 6 potential clones capable of generating fold-competent recombinant coat proteins. These six clones were then subjected to sequence analysis to verify their coat constructions. The resulting data indicated that all clones either contained frameshift mutations or had other ligation errors. This entire experiment was repeated with no functional clones ever being recovered. Thus, no recombinants clones with randomized brackets were generated in the monomer configuration.

Insertion of Fusion Sequence into the Single-Chain Dimer

Attempts to generate recombinant proteins with randomized flanking sequences failed in the monomer configuration. We then constructed a new plasmid incorporating the single-chain dimer to express our randomizing specialized peptide. As illustrated by the results, recombinant proteins with the correct genetic sequence were generated (Table XI).

MS2 Recombinant VLP Formation

A key outcome of this experimental effort was the expectation that the introduction of N-terminal extensions would not prevent VLP assembly. When a capsid gel was performed for the recovered sequences, no VLPs were visualized via ethidium bromide staining. The control, p2CT119dl5N (p2M), generates a single-chain dimer of the wild-type MS2 coat sequence. The presence of p2M was verified in the positive control lane via ethidium bromide staining. To complete the experiment, a Western blot was performed. However, as demonstrated by capsid gel, no recombinant VLPs were generated.

Figure 15. Randomized brackets in N-terminus of MS2
(a) Arrangements of the coat protein reading frames on the plasmids used in this experiment. Plasmid pCT119DL5N is similar to the previously described pCT119 [11, 14], except for *NcoI* site that also incorporates the coat protein initiation codon (delineated by arrow). Plasmid p2CT119DL5N expresses a single-chain dimer version of the protein. Black boxes represent the peptide insertions made in each plasmid configuration. (b) Codons of the oligonucleotides utilized to make insertions in each plasmid configuration.

Table XI. Summary of randomized bracket configurations

White colonies, as they repress translation were the primary intent of this effort. Blue colonies, were included to expand the data set on randomized brackets and the functional behavior they imparted. Blue colonies were recovered from the original experiment in which a high background of non-recombinants was generated.

Table XI. Summary of recombinant proteins expressed in this effort					
Sample	NNY	NNY2	Know Seq	NNY3	NNY4
WHITE					
F1	Ser	Pro	LNLVFILLSG	Ser	Leu
F6	Asn	Tyr	LNLVFILLSG	His	Ser
F7	His	Ser	LNLVFILLSG	Ile	Thr
F8	Arg	Asn	LNLVFILLSG	Asn	His
F9	His	Asp	LNLVFILLSG	Ser	Arg
F11	Tyr	Tyr	LNLVFILLSG	Asn	Phe
BLUE					
F2	Asp	Asp	LNLVFILLSG	His	Asn
F4	Asp	Asn	LNLVFILLSG	Asp	His
F10	Ile	Asp	LNLVFILLSG	Asp	Pro
F13	Asn	Asp	LNLVFILLSG	Asp	Ser
F14	Asp	Asn	LNLVFILLSG	Ser	Ser

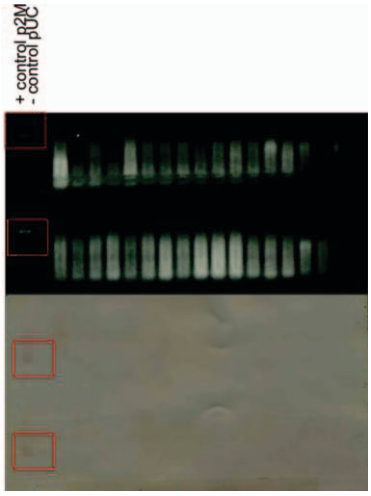


Figure 16. VLP formation in randomized bracket configuration

Electrophoresis on agarose gel of VLPs identified in lysates of cells producing recombinant coat proteins. The first two lanes of each gel are controls: pCT119 produces wild-type coat, while pUC productes no coat protein. An Ethidium bromide-stained gel (top panel) and western blot probed with MS2 anti-serum (bottom panel) for

recombinants are shown. Red boxes highlight control coat protein (p2M). Plasmid p2M generates a single-chain wild-type coat protein.

DISCUSSION

N-Terminal Fusions

Since the N-terminus resides near the surface of the MS2 capsid, it represented a potential location for peptide display. Initial efforts suggested that 10-amino acid N-terminal fusions were highly destabilizing. Even in the single-chain dimer random N-terminal fusions generated an unacceptably low number of functional VLPs. A review of the literature identified several opportunities to further stabilize peptide stretches as they were incorporated. The first approach suggested that the incorporation of glutamate or aspartate sequences would introduce localized acidity around a peptide insertion, thereby changing overall insertion hydrophobicity [9]. In the monomer configuration, only 2% of random insertions yielded functional translational repressors. After the addition of glutamate and aspartate brackets, the numbers improved to 65% and 79%. With these acidic additions, however, VLP formation was still essentially zero.

As the data showed, incorporation of the single-chain dimer produced only 2% functional repressors. Once acidic brackets were incorporated along with the single-chain dimer, functional repressors again increased to 79% for aspartate and 88% for glutamate. However, even with the incorporation of the single-chain dimer and acidic brackets, VLP formation improved to only 4%. What is clear, however, is that acidic brackets do have the ability to minimize disruption to coat function upon peptide insertion. Clearly, proper coat folding is the critical first step in VLP formation. However, as implied by Billaud et al., acidic brackets alone will not fully stabilize protein insertions. In the same study by Billaud, C-terminal modifications were also required to allow for recombinant peptides to be displayed in a VLP. Perhaps acidic brackets, in

concert with other modifications to MS2 coat, could allow for a high degree of VLP display. But as it stands now, acidic brackets do not represent an ideal configuration for N-terminal VLP display in MS2.

A second approach identified to potentially provide coat protein stability related to the inclusion of flanking arginines. The premise behind this approach was that flanking arginines would allow for the recruitment of chaperones and allow for refractory peptide incorporation. As a technical approach, similar efforts have been executed with some success in filamentous phage [28, 34]. In the monomer configuration, random insertions flanked by arginines yielded 58% functional repressors. VLP formation for this configuration was zero. As efforts then transitioned into the single-chain dimer, flanking arginine sequences yielded 95% functional recombinant repressors – an experimental high for all engineering modifications. VLP formation, however, remained unchanged at zero. Again, as in the case of acidic brackets, coat function was significantly restored with the addition of arginines. But, VLP formation, a critical goal for this effort showed no improvement.

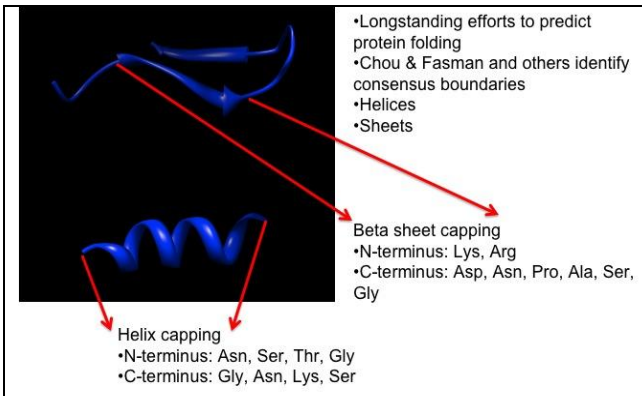


Figure 17. Amino acid propensities
Summary of amino acid propensities at the boundaries of a β-Sheet and α-helix. Amino acids listed here formed the basis for flanking sequences employed to stabilize foreign peptide insertions.

The final approach offered in this experimental effort to correct folding failures in MS2 coat was the implementation of amino acid propensities to predict optimum capping residues for recombinant sequences. Although several amino acid combinations were tried, flanking serines emerged as the best approach to stabilize random N-terminal fusions. In the context of the monomer, flanking serines improved the fraction of recombinants producing functional repressors from 2% to 70%. VLP formation, however, improved only marginally to 4%. Once the single-chain dimer and flanking serines were incorporated with random sequences, repressor function improved to 90-98% (corrected for frameshift errors). Of these functional repressors, 100% also produced VLPs. Thus, we have established a potentially significant role for serines in stabilizing recombinant insertions at the N-terminus of MS2.

The increased stability that serines provide may offer another general approach to stabilizing recombinant proteins. The importance of the single-chain dimer for the success of VLP display applications is now well documented [11-13, 49]. In the several

cases presented here, the incorporation of certain amino acid bracket sequences stabilized the folded structure of MS2 coat. As serine emerged as being able to stabilize MS2 coat and allow for VLP formation, we speculated if another, simpler mechanism could account for the stabilizing effects observed for glutamate, aspartate, arginine, and serine. Protein folding has, until relatively recently, relied heavily on statistical models that describe secondary structures only in the context of helices and sheets. However, other research efforts have extended secondary structure prediction to include intra-residue, inter-residue, side-chain conformations, and related protein contacts in their models [50-54].

Of particular note, are the hydrogen bonding interactions that can occur between amino acid residues in polypeptides. In the case of arginine, it has been suggested that because of the guanidinium group at the end of its chain, this amino acid, in particular, has significant H-bonding potential [55]. In this same review by Borders et al., it was highlighted that the guanidinium group of arginine has the ability to donate up to 5 hydrogen bonds to interact with peptide carbonyl atoms. This interaction of arginine with the peptide backbone was thought to stabilize and maintain correct protein folding [55, 56]. In another article, it was suggested that small polar residues like serine posses the ability to compensate for peptide backbone distortions because of their ability to form compensating hydrogen bond interactions [51]. The implication here was that as random insertions perturb the polypeptide backbone of MS2 coat, serine might minimize these distortions through offsetting, non-standard H-bonds.

In the final case of charged amino acids like glutamate and aspartate, structural polypeptide stabilization may be achieved by similar means. Glutamate has been shown

to interact with its own backbone amide via the ε-carbonyl oxygen in its side chain as well as interacting with main-chain atoms upstream and downstream to its relative polypeptide position [53]. Aspartate also interacts with main-chain atoms upstream and downstream to its relative position. However, aspartate and its ability to h-bond to main-chain atoms 2 positions downstream of its location offers a unique side-chain/main-chain conformation. This type of conformation mimics the type II' β-turn and has been shown to stabilize other β-turns [57]. Additionally, this unique conformation for aspartate has been identified as critical in the propagation of some α-helices [15, 57]. Thus, for arginine, serine, glutamate, and aspartate, it becomes clear that an alternate hypothesis of simply side-chain/main-chain stabilization could be offered to explain the positive effects uncovered in this study.

Although hydrophobicity, amino acid propensities, and molecular chaperone recruitment represented the initial approaches identified to resolve folding failures in MS2, a new, more simplified approach has also emerged. This modified approach simply relates to protein stabilization. The increased stabilization offered from each technical approach presented in this study does stand on its own merits. But, when you consider protein folding in its most generic context, it becomes clear that the path to protein stability can take many routes. In the case of MS2 coat protein, the use of the single-chain dimer represented but one path to coat protein stability.

In fact, it has been suggested that another equally viable approach to revert folding defects could simply be to identify mutations at other sites in MS2 coat that suppress effects introduced with the incorporation of recombinant sequences [13]. Thus, protein contacts, side-chain interactions, hydrophobicity, backbone mimics, amino acid

position, conformational geometries, etc., all represent technical approaches to be considered in evaluating protein folding. Thus, as long as any technical approach to protein folding considers its relative contribution to protein stability, it can only serve to further expand our understanding of protein folding.

In the context of the present study, our primary goal was to identify opportunities that allowed for the effective insertion of small peptides at the N-terminus of MS2 for VLP display. With the identification of serines and the use of the single-chain dimer, we have identified a system of modifications that correct most folding and assembly errors at the N-terminus of MS2. Additionally, having a second insertion location in MS2 VLPs for peptide display could prove to be very beneficial in targeted immunological applications. First, one recombinant site on MS2 (n-terminus) could be employed for cellular targeting, while the other site (AB-loop) could be allowed to locally induce the immune response [12, 58]. This could potentially offer a highly specific immunological approach to deal with disease. Second, the N-terminus represents an opportunity where peptides can be presented in an un-constrained fashion – which has been suggested as another, but less effective means of generating biologically relevant three-dimensional targets for antibodies [59]. Thus, VLPs generated by randomized insertions at the N-terminus should expand libraries constructed to identify epitopes in immunological applications and increase the utility of the MS2 platform for other biotechnology applications.

AB-Loop Insertions

The AB-loop, as it resides near the surface of the MS2 capsid, has recently become a target for peptide display. Initial efforts suggested that peptide insertions at the

AB-loop were highly destabilizing. In our original technical approach, we expected that the stabilizing effects of the single-chain dimer would restore folding in recombinant AB-loop coat sequences. As the single-chain dimer is essentially comprised of two coat protein monomers, two AB-loop targets become available for insertion in this configuration. In the research activities that followed, it was shown that peptides could be displayed in downstream (3') AB-loop via this specialized coat construction [12]. However, it was also clear that the upstream or 5' AB-loop was refractory to insertion under most conditions. A review of the literature outlined several opportunities to potentially stabilize peptide stretches as they were incorporated in the 5' loop. We identified two approaches that follow: one in which amino acid sequences were shown to cap or essentially provide termination of secondary structures; and two, the potential recruitment of molecular chaperones via specialized amino acid sequences.

Our first approach related to the inclusion of flanking arginines. The thought behind this methodology was that flanking arginines would recruit chaperones to the AB-loop capable of assisting in refractory peptide insertion. Similar approaches have been attempted with some success in filamentous phage [28, 34]. In the monomer configuration, random insertions flanked by arginines produced only 2% functional repressors. However, once focus shifted to the single-chain dimer, flanking arginine sequences yielded 71% for twin flanking arginines and 86% functional recombinant repressors for quad flanking arginines – an experimental high for all engineering modifications. VLP formation, however, only improved to 17% for twin arginines and 8% for quad flanking arginines. What is clear, though, was that coat repressor function was significantly restored with the addition of arginines.

As the data outlined, incorporation of the single-chain dimer alone produced 41% functional repressors in the 5' AB-loop. VLP formation in the insert-only configuration was at 42%. For immunological activities planned in the future, having low effective insert frequencies and low particle formation would be technically unacceptable. Once capping sequences, our second approach, were incorporated along with the single-chain dimer, functional repressors increased to a high of 80% for serines. Additionally, VLP formation for serine capped sequences, improved to 75% VLP formation.

Although sequences capped by aspartate showed a 40% improvement in VLP formation, this result was not above that of the baseline, single-chain dimer configuration. However, given that aspartate produced 74% functional repressors, it was clear that acidic brackets do have the ability to minimize disruption to coat function upon peptide insertion. As suggested by Billaud et al., acidic brackets alone will not fully stabilize protein insertions. Other polypeptide changes including C-terminal modifications were needed to allow for recombinant peptides to be displayed in a VLP. Perhaps other offsetting modifications to MS2 coat, as suggested by other research efforts will allow for a high degree of VLP display [13]. But as it stands now, acidic brackets do not represent an ideal engineering modification to allow for 5' AB-loop peptide insertions and eventual VLP display.

The increased stability that serines provide may again offer the best technical solution to stabilizing recombinant proteins in this application. In this effort, we have shown that the use of certain di-peptide brackets can stabilize the folded structure of MS2 coat. However, di-serine brackets proved the most efficient at suppressing both coat protein folding and VLP assembly defects. As we considered how terminal serines

benefitted AB-loop insertions, we tried to identify an underlying mechanism. To get at this problem, we reviewed other technical literature that described efforts to expand secondary structure prediction. As a result, we identified several technical approaches that put a new light on protein folding in general. These other approaches included intra-residue, inter-residue, side-chain conformations, and related protein contacts in their models [50-54]. As noted with N-terminal insertions, hydrogen bonding emerged as the consensus approach to best expand our understanding of how amino acids interact in the context of a polypeptide chain.

Arginine, in and of itself, offers a very complex network of H-bonding interactions. The guanidinium group at the end of its chain has the potential to provide up to 5 hydrogen bonds as its side chain interacts with the main polypeptide chain and other amino acids [55]. Serine, which was found to offer the greatest ability to stabilize AB-loop insertions in the 5' dimer and promote VLP formation, also possesses a great ability to promote hydrogen interactions during protein formation. In fact, serine was identified it as having a unique capacity to compensate for peptide backbone distortions [51]. Mechanistically, serine could work to stabilize AB-loop insertions through intra- or inter-strand h-bonds. In the final case of a charged amino acid like aspartate, H-bonding appeared once again to be able to account for the stabilization offered by this amino acid residue. Aspartate possesses the ability to cooperate with main-chain atoms upstream and downstream to its relative position. Aspartate makes one particular conformation which is though to mimic the type II' β-turn, stabilize other β-turns, and even potentially nucleate to formation of some α-helices [15, 57]. Hence, the effects of arginine, serine,

and aspartate demonstrated in this study can all otherwise be defined by the stabilizing interactions between side-chains and main-chains.

Although β-sheet capping interactions and molecular chaperone recruitment were the initial approaches taken to address AB-loop failures in MS2, a new, more simplified approach has emerged. Proteins clearly can be stabilized by novel protein interactions introduced with the strategic placement of amino acids. The increased stabilization offered from each technical approach presented in this study highlights how H-bonding can be applied to stabilize a recombinant protein. In the case of MS2 coat protein, the use of the single-chain dimer represents but one path to coat protein stability. Protein contacts, side-chain interactions, hydrophobicity, backbone mimics, amino acid position, conformational geometries, etc., also offer the potential of equal validity in addressing folding failures identified in this effort.

In the context of the present study, our primary goal was to identify opportunities that allowed for the effective insertion of small peptides in the 5' AB-loop of MS2 for VLP display. With the identification of terminal serines and the use of the single-chain dimer, we have identified a system of modifications that corrects a significant percentage of folding and assembly errors in the 5' AB-loop. Additionally, we have made available a second loop insertion location in MS2 VLPs. We believe that this additional site could prove to be very beneficial in targeted immunological applications. First, having two identical sites to display an epitope has the potential to induce maximum immunogenicity – as it increases the density of antigens available to activate the immune system [12, 60]. Second, having an additional site could allow for multiple epitope presentation on the

same particle. As the flu vaccine is designed to provide a spectrum of viral epitopes for immunization, having multiple sites for antigen presentation could offer a similar benefit.

As part of this effort we also considered the type of peptides not tolerated in the 5' loop and sought to identify any obvious trends that emerged from their analysis. For example, it had already been demonstrated that peptide sequences that were refractory in the 3' loop appeared to be highly enriched in hydrophobic sequences [12]. In the case of the 5' loop, we do not see the same phenomena. On a related concern, we also investigated the potential of amino acids to impart localized constraints through the introduction of novel secondary structures. Therefore, we chose to look at all sequences that were accommodated or were refractory in the 5' loop and modeled their propensity to form sheets or helices. In conclusion, we identified engineering modifications that allowed for increased display of recombinant peptides in the 5' AB-loop of the single-chain dimer. With this new display site in hand, MS2 VLPs should provide the means to identify peptides with novel immunological properties and allow for their display as part of an expanded vaccine platform.

Other Approaches to N-terminal Peptide Display

The N-terminus, as it resides near the surface of the MS2 capsid, offers the potential to be utilized in the display of peptides. Initial efforts suggested that, without extensive engineering, viable insertions at the N-terminus would be extremely rare. To correct for potential folding failures, we planned to implement the incorporation of the single-chain dimer. Our expectation was, as in other phage display efforts, that the single chain dimer would offer additional stability necessary to revert the folding defects that we had encountered [11, 12, 61]. However, the addition of the single-chain dimer proved

to not be the optimum solution that we required. Thus, we went back to the literature and identified two potential enhancements that we felt, in concert with the single-chain dimer, would provide the solution we needed. The first approach involved the co-expression of a wild-type coat sequence with a recombinant coat sequence in the hopes of generating a mosaic capsid.

To execute this study, we drew from an earlier effort in which attempts had been made to attach the Flag peptide to the N-terminus of MS2. In this study, the Flag peptide could only be expressed in a virus-like particle with the help of the single-chain dimer. In our new approach, we recognized that Flag incorporation into the monomer configuration actually produced a properly folded, repressor-competent protein. This idea of a properly folded protein was key. Although Flag recombinant proteins failed to assemble into a VLP, its properly folded configuration could allow for its incorporation with wild-type coat sequences.

As anticipated, the monomer configuration with Flag was successfully incorporated into a mosaic VLP. We believe this effort succeeded because the flexibility of the wild-type coat sequence offered the greatest opportunity to incorporate a recombinant protein that may have lost some of its own conformational flexibility. We base our conclusion on the fact that we knew that the wild-type coat protein in the assembly of the native MS2 particle assumed three different conformations to generate a capsid. Thus, native coat protein's flexibility could offer an assembly advantage.

The requirement for a properly folded recombinant protein came from another effort to incorporate improperly folded recombinant proteins into mosaic capsids. In those efforts, we failed to identify any improperly folded proteins capable of

incorporation into a mosaic particle (unpublished results). However, in the case of the Flag fusion construct, this peptide was capable of acting as a functional repressor –which suggests it being properly folded. As a result, Flag fusions with the addition of a wild-type coat sequences were efficiently incorporated into a VLP. The only challenge that remains with this approach for future efforts is to ensure that the RNA sequence that encodes for the recombinant protein is efficiently incorporated into the mosaic capsid. But, this approach may prove to be viable in instances where a know sequence remains refractory to VLP display and needs to be incorporated into a mosaic capsid to determine its interaction with other proteins.

In our second approach to incorporate sequences at the N-terminus, we sought to randomize two amino acid locations around a known refractory sequence. We believed that this would introduce sequences that stabilized the overall recombinant coat sequence. As the data showed, this approach was capable of producing functional repressors as long as the single-chain dimer was part of the technical approach. However, none of the identified pairs of amino acids yielded VLPs. As we were able to show in another study, two serines flanking a randomized peptide insert generated function repressors and VLPs. As we worked to identify a mechanism to explain the phenomena of serines, we found that this amino acid, in particular, offered incredible conformational flexibility. In fact, the increased stability that serines provided may offer another general approach to stabilizing recombinant proteins.

At this point, we believe that serine offers the ability to form intra- and inter-strand hydrogen bonds. As we identified in a second study on the beneficial aspects of serine, we noted that both positions that flanked a randomized sequence were required to

generate a VLP. In our case of randomized locations, we identified several amino acid combinations that allowed for dimer stabilization. However, none of the conformations identified were able to nucleate assembly. Protein folding continues to be one of the most complex issues faced by scientists. Mosaic particle formation, strategic incorporation of stabilizing amino acids, single-chain dimer construction – all represent new tools to aid in the development of an optimum platform for phage display. Both strategies presented herein provide additional opportunities to aid in the creation and identification of novel peptides. As we come to better understand those factors that influence the nature of protein folding, we will better be able to expand the *de novo* strategies available for protein engineering. For the MS2 platform, having a repository of technical approaches on which to generate VLPs will ensure platform complexity and allow for the flexibility to address the challenges that lay ahead.

REFERENCES

1. Smith, G.P. and V.A. Petrenko, *Phage Display.* Chemical Reviews, 1997. 97(2): p. 391-410.
2. Breitling, F., and Dübel, S., *Recombinant Antibodies*. 1999, New York: John Wiley and Sons.
3. Kehoe, J.W. and B.K. Kay, *Filamentous Phage Display in the New Millennium.* Chemical Reviews, 2005. 105(11): p. 4056-4072.
4. Castagnoli, L., et al., *Alternative Bacteriophage Display Systems.* Combinatorial Chemistry & High Throughput Screening, 2001. 4(2): p. 121.
5. Rothe, A., R.J. Hosse, and B.E. Power, *In vitro display technologies reveal novel biopharmaceutics.* FASEB J., 2006. 20(10): p. 1599-1610.
6. He, M. and M.J. Taussig, *Rapid discovery of protein interactions by cell-free protein technologies.* Biochemical Society Transactions, 2007. 035(5): p. 962-965.
7. He, M. and M. Taussig, *Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites.* Nucl. Acids Res., 1997. 25(24): p. 5132-5134.
8. Hanes, J. and A. Pl√°ckthun, *In vitro selection and evolution of functional proteins by using ribosome display.* Proceedings of the National Academy of Sciences of the United States of America, 1997. 94(10): p. 4937-4942.
9. Billaud, J.-N., et al., *Combinatorial Approach to Hepadnavirus-Like Particle Vaccine Design.* J. Virol., 2005. 79(21): p. 13656-13666.
10. Amstutz, P., et al., *In vitro display technologies: novel developments and applications.* Current Opinion in Biotechnology, 2001. 12(4): p. 400-405.
11. Peabody, D.S., *Subunit Fusion Confers Tolerance to Peptide Insertions in a Virus Coat Protein.* Archives of Biochemistry and Biophysics, 1997. 347(1): p. 85-92.
12. Peabody, D.S., et al., *Immunogenic Display of Diverse Peptides on Virus-like Particles of RNA Phage MS2.* Journal of Molecular Biology, 2008. 380(1): p. 252-263.
13. Peabody, D., *A Viral Platform for Chemical Modification and Multivalent Display.* Journal of Nanobiotechnology, 2003. 1(1): p. 5.
14. Peabody, D., *Translational Repression by Bacteriophage MS2 Coat Protein Expressed From a Plasmid: A System for Genetic Analysis of a Protein-RNA Interaction.* J Biol Chem, 1990. 265: p. 5684 - 5689.
15. Richardson, J. and D. Richardson, *Amino acid preferences for specific locations at the ends of alpha helices.* Science, 1988. 240(4859): p. 1648-1652.
16. Presta, L.G. and G.D. Rose, *Helix Signals in Proteins.* Science, 1988. 240(4859): p. 1632-1641.
17. Chakrabartty A et al., *Helix capping propensities in peptide parallel those in proteins.* Proceedings of the National Academy of Sciences of the United States of America, 1993. 90(23): p. 11332-11336.
18. Chou, P.Y. and G.D. Fasman, *Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins.* Biochemistry, 1974. 13(2): p. 211-222.

19. Chou, P.Y. and G.D. Fasman, *Prediction of protein conformation.* Biochemistry, 1974. 13(2): p. 222-245.
20. Guzzo, A., *The Influence of Amino Acid Sequence on Protein Structure.* Biophysical Journal, 1965. 5(6): p. 809-822.
21. Street, A.G. and S.L. Mayo, *Intrinsic Œ≤ -Sheet Propensities Result from van der Waals Interactions between Side Chains and the Local Backbone.* Proceedings of the National Academy of Sciences of the United States of America, 1999. 96(16): p. 9074-9076.
22. FarzadFard, F., et al., *[beta]-Sheet capping: Signals that initiate and terminate [beta]-sheet formation.* Journal of Structural Biology, 2008. 161(1): p. 101-110.
23. Yawen Bai, S.W.E., *Hydrogen bond strength and beta-sheet propensities: The role of a side chain blocking effect.* Proteins: Structure, Function, and Genetics, 1994. 18(3): p. 262-266.
24. Koh, E., T. Kim, and H.-s. Cho, *Mean curvature as a major determinant of {beta}-sheet propensity.* Bioinformatics, 2006. 22(3): p. 297-302.
25. Chakrabartty, A., A.J. Doig, and R.L. Baldwin, *Helix Capping Propensities in Peptides Parallel Those in Proteins.* Proceedings of the National Academy of Sciences of the United States of America, 1993. 90(23): p. 11332-11336.
26. Koehl, P. and M. Levitt, *Structure-Based Conformational Preferences of Amino Acids.* Proceedings of the National Academy of Sciences of the United States of America, 1999. 96(22): p. 12524-12529.
27. Engel, D.E. and W.F. DeGrado, *Amino Acid Propensities are Position-dependent Throughout the Length of [alpha]-Helices.* Journal of Molecular Biology, 2004. 337(5): p. 1195-1205.
28. Thammawong, P., et al., *Twin-arginine signal peptide attributes effective display of CD147 to filamentous phage.* Applied Microbiology and Biotechnology, 2006. 69(6): p. 697-703.
29. Pugsley, A.P., *The complete general secretory pathway in gram-negative bacteria.* Microbiol. Mol. Biol. Rev., 1993. 57(1): p. 50-108.
30. Schafer, U., K. Beck, and M. Muller, *Skp, a Molecular Chaperone of Gram-negative Bacteria, Is Required for the Formation of Soluble Periplasmic Intermediates of Outer Membrane Proteins.* J. Biol. Chem., 1999. 274(35): p. 24567-24574.
31. Kuehn, M.J., S. Normark, and S.J. Hultgren, *Immunoglobulin-like PapD chaperone caps and uncaps interactive surfaces of nascently translocated pilus subunits.* Proceedings of the National Academy of Sciences of the United States of America, 1991. 88(23): p. 10586-10590.
32. Saier, M., *Protein Secretion and Membrane Insertion Systems in Gram-Negative Bacteria.* Journal of Membrane Biology, 2006. 214(1): p. 75-90.
33. Perez-RodrÌguez, R., et al., *An Essential Role for the DnaK Molecular Chaperone in Stabilizing Over-expressed Substrate Proteins of the Bacterial Twin-arginine Translocation Pathway.* Journal of Molecular Biology, 2007. 367(3): p. 715-730.
34. Paschke, M. and W. Hˆhne, *A twin-arginine translocation (Tat)-mediated phage display system.* Gene, 2005. 350(1): p. 79-88.

35. Tapley, T.L., J.R. Cupp-Vickery, and L.E. Vickery, *Sequence-Dependent Peptide Binding Orientation by the Molecular Chaperone DnaK&#x2020.* Biochemistry, 2005. 44(37): p. 12307-12315.
36. Crouy-Chanel, A.d., M. Kohiyama, and G. Richarme, *Specificity of DnaK for Arginine/Lysine and Effect of DnaJ on the Amino Acid Specificity of DnaK.* J. Biol. Chem., 1996. 271(26): p. 15486-15490.
37. Rudiger, S., et al., *Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries.* The EMBO Journal, 1997. 16(7): p. 1501-1507.
38. Patrik Kasper, P.C., Heinz Gehring,, *Empirical calculation of the relative free energies of peptide binding to the molecular chaperone DnaK.* Proteins: Structure, Function, and Genetics, 2000. 40(2): p. 185-192.
39. Smith, G.P., *Preface.* Gene, 1993. 128: p. 1-2.
40. McConnell, S.J., M.L. Kendall, et al. , *Constrained peptide libraries as a tool for finding mimotopes.* Gene, 1994. 151: p. 115-8.
41. Mead, D.A., B. Kemper, *Vectors: A survey of molecular cloning vectors and their uses.* Biotechnology (Reading, Mass), 1988. 10: p. 1-578.
42. Cesareni, G., *Vectors: A survey of molecular cloning vectors and their uses.* Biotechnology (Reading, Mass), 1988. 10: p. 1-578.
43. Peabody, D. and F. Lim, *Complementation of RNA binding site mutations in MS2 coat protein heterodimers.* Nucleic Acids Res, 1996. 24: p. 2352 - 2359.
44. Peabody, D., *The RNA-binding site of bacteriophage MS2 coat protein.* EMBO J, 1993. 12: p. 595 - 600.
45. Burnette, W., *"Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A.* Analytical Biochemistry, 1981. 112(2): p. 195-203.
46. Gunasekaran, K., et al., *Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals.* Journal of Molecular Biology, 1998. 275(5): p. 917-932.
47. Stoycheva, A.D., C.L. Brooks Iii, and J.N. Onuchic, *Gatekeepers in the Ribosomal Protein S6: Thermodynamics, Kinetics, and Folding Pathways Revealed by a Minimalist Protein Model.* Journal of Molecular Biology, 2004. 340(3): p. 571-585.
48. Lim, F., T.P. Downey, and D.S. Peabody, *Translational Repression and Specific RNA Binding by the Coat Protein of the Pseudomonas Phage PP7.* Journal of Biological Chemistry, 2001. 276(25): p. 22507-22513.
49. Caldeira, J. and D. Peabody, *Stability and assembly in vitro of bacteriophage PP7 virus-like particles.* Journal of Nanobiotechnology, 2007. 5(1): p. 10.
50. Faure, G., A. Bornot, and A.G. de Brevern, *Protein contacts, inter-residue interactions and side-chain modelling.* Biochimie, 2008. 90(4): p. 626-639.
51. Gunasekaran, K., C. Ramakrishnan, and P. Balaram, *Disallowed Ramachandran Conformations of Amino Acid Residues in Protein Structures.* Journal of Molecular Biology, 1996. 264(1): p. 191-198.

52. Doig, A.J. and M.J.E. Sternberg, *Side-chain conformational entropy in protein folding.* Protein Science, 1995. 4(11): p. 2247-2251.
53. Eswar, N. and C. Ramakrishnan, *Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures.* Protein Eng., 2000. 13(4): p. 227-238.
54. Stapley, B.J. and A.J. Doig, *Free energies of amino acid side-chain rotamers in [alpha]-helices, [beta]-sheets and [alpha]-helix N-caps.* Journal of Molecular Biology, 1997. 272(3): p. 456-464.
55. Borders, C.L., et al., *A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens.* Protein Science, 1994. 3(4): p. 541-548.
56. White, S.H., *Amino acid preferences of small proteins : Implications for protein stability and evolution.* Journal of Molecular Biology, 1992. 227(4): p. 991-995.
57. Eswar, N. and C. Ramakrishnan, *Secondary structures without backbone: an analysis of backbone mimicry by polar side chains in protein structures.* Protein Eng., 1999. 12(6): p. 447-455.
58. Morris, M.C., et al., *Cell-penetrating peptides: from molecular mechanisms to therapeutics.* Biology of the Cell, 2008. 100(4): p. 201-217.
59. Herman, R.E., et al., *The Trp Cage Motif as a Scaffold for the Display of a Randomized Peptide Library on Bacteriophage T7.* Journal of Biological Chemistry, 2007. 282(13): p. 9813-9824.
60. Houten, N.E.V. and J.K. Scott, eds. *Phage Libraries for Developing Antibody-Targeted Diagnostics and Vaccines*. Phage Display in Biotechnology and Drug Discovery, ed. S. Sidhu. 2005, CRC Press: Boca Raton. 165-254.
61. Liang, H., W.S. Sandberg, and T.C. Terwilliger, *Genetic Fusion of Subunits of a Dimeric Protein Substantially Enhances its Stability and Rate of Folding.* Proceedings of the National Academy of Sciences of the United States of America, 1993. 90(15): p. 7010-7014.