JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
Cite This: J. Am. Chem. Soc. 2019, 141, 3875−3884
pubs.acs.org/JACS

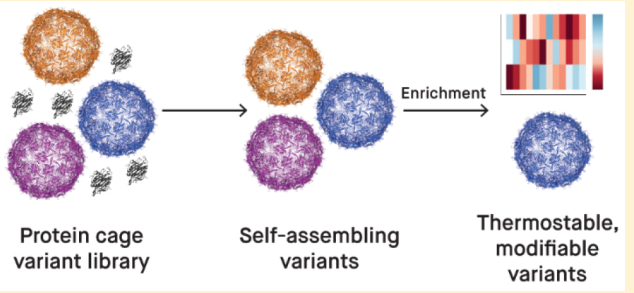
Systematic Engineering of a Protein Nanocage for High-Yield, Site-Specific Modification

Daniel D. Brauer,[†,⊥] Emily C. Hartman,[†,⊥] Daniel L. V. Bader,[†] Zoe N. Merz,[†] Danielle Tullman-Ercek,[*,‡] and Matthew B. Francis[*,†,§]

[†]Department of Chemistry, University of California, Berkeley, California 94720-1460, United States
[‡]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Technological Institute E136, Evanston, Illinois 60208-3120, United States
[§]Materials Sciences Division, Lawrence Berkeley National Laboratories, Berkeley, California 94720-1460, United States

Supporting Information

ABSTRACT: Site-specific protein modification is a widely used strategy to attach drugs, imaging agents, or other useful small molecules to protein carriers. N-terminal modification is particularly useful as a high-yielding, site-selective modification strategy that can be compatible with a wide array of proteins. However, this modification strategy is incompatible with proteins with buried or sterically hindered N termini, such as virus-like particles (VLPs) composed of the well-studied MS2 bacteriophage coat protein. To assess VLPs with improved compatibility with these techniques, we generated a targeted library based on the MS2-derived protein cage with N-terminal proline residues followed by three variable positions. We subjected the library to assembly, heat, and chemical selections, and we identified variants that were modified in high yield with no reduction in thermostability. Positive charge adjacent to the native N terminus is surprisingly beneficial for successful extension, and over 50% of the highest performing variants contained positive charge at this position. Taken together, these studies described nonintuitive design rules governing N-terminal extensions and identified successful extensions with high modification potential.



INTRODUCTION

Site-specific bioconjugation techniques are widely used to produce useful conjugate biomaterials. Many recently developed N-terminal modification strategies are of particular interest, as these reactions are high-yielding, can proceed under mild reaction conditions, and have the capacity to be site-selective.[1−9] Because nearly all proteins contain a single instance of an N terminus, these reactions are useful in a wide variety of contexts,[10] including the loading of cargo onto protein carriers[11] or the development of new biomaterials.[12,13] However, such reactions require free N-terminal residues that are uninvolved in secondary structure, limiting their usefulness on proteins with sterically hindered N termini. One such case is the MS2 bacteriophage, a well-studied protein nanocage that is being actively explored for applications in drug delivery,[14−16] disease imaging,[17] vaccines,[18,19] and biomaterials.[20−22] Limited genetic manipulations can be made to the MS2 coat protein (CP) without disrupting the assembly state,[23] and many inter- and intrasubunit contacts make mutability challenging to predict.[24] Additionally, the native N terminus is sterically hindered, and efforts to extend the N terminus have had limited success.[25] As such, currently developed N-terminal modification strategies are not compatible with the MS2 CP. Instead, the attachment of targeting groups to the exterior of the MS2 CP either relies on nonspecific chemistry, such as lysine modification, or requires the incorporation of nonstandard amino acids, lowering expression yields and complicating protocols.[26,27] The usefulness of the MS2 scaffold would be expanded substantially by enabling N-terminal modification of the CP in a manner that yields stable, easy-to-produce, and modifiable virus-like particles (VLPs).

Here, we combine a systematically generated library with direct functional selections to identify N-terminally extended variants of the MS2 CP that are well-assembled, thermostable, and amenable to chemical modification (Figure 1). In addition to identifying highly useful extensions that can be modified to >99% by oxidative couplings between the N terminus and oxidized catechols, we also uncovered surprising design rules governing which extensions are compatible with particle assembly. Of 8000 possible combinations of N-terminal extensions, merely 3% of the library remained assembled through stringent chemical and thermal selections. In addition to identifying useful VLP variants for biomedical applications, this study represents the first time that chemical modification conditions have been used as a selection for protein fitness. This

Received: October 4, 2018
Published: February 7, 2019

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

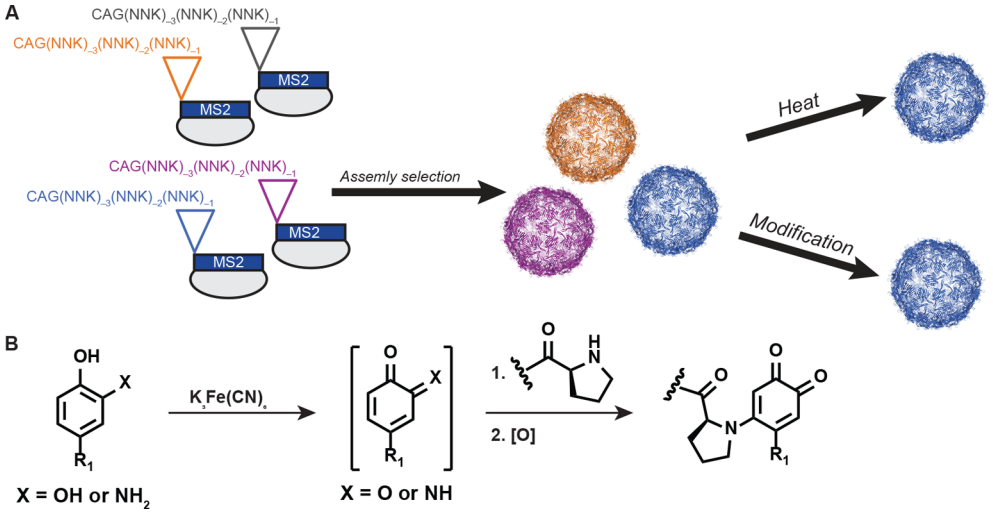


Figure 1. Scheme to isolate N-terminally extended VLPs with desired properties. (A) Three-codon NNK extensions at the N terminus generated a library of 8000 variants. Assembly, thennostability, and chemical modification challenges were used to identify HiPerX variants or high-performing N-terminal extensions with desirable properties, indicated in blue. (B) Oxidative coupling reactions can be used to modify N-terminal proline residues.

approach could be adapted to study the modification efficiency for other reactions or protein substrates and could provide rich information about the effects of amino acid sequence on reactivity.

RESULTS AND DISCUSSION

Characterization of a Comprehensive N-Terminally Extended MS2 Bacteriophage Library. The MS2 VLP is a 27 nm icosahedral particle that is composed of 180 copies of a protein monomer. Three N termini of these quasi-equivalent proteins are clustered together, forming a triangle with lengths of 11.7, 12.8, and 7.9 Å (Figure 2a).[28] This sterically confined local environment suggests that few N-terminal extensions would be compatible with particle assembly. As such, we sought to use Systematic Mutagenesis and Assembled Particle Selection (SyMAPS), a technique developed previously in our laboratories,[23] to evaluate all possible proline-terminated extensions of the MS2 CP with the pattern P-X-X-X-MS2, where X represents all amino acids. When expressed in *E. coli*, assembly-competent variants of the MS2 CP encapsulate available negative charge, including mRNA. SyMAPS capitalizes on this property, using the encapsulated nucleic acid as a convenient genotype-to-phenotype link. Well-assembled VLPs copurify with a snapshot of cellular nucleic acid, including variant mRNA, while mRNA from poorly assembled VLPs is lost.

As shown in Figure 1A, an NNK-based strategy was used to encode all variants while minimizing biases due to genetic code redundancies.[23] Following expression, the N-terminal methionine of wild-type MS2 CP is cleaved, yielding an alanine in position 1. In the library, extensions were appended directly before alanine 1, starting with a −1 position. With this numbering, the N-terminal proline is located at the −4 position (Figure 2B). Proline also is compatible with efficient methionine cleavage, leading to a library with four total extended residues.[29] The invariant N-terminal proline was chosen because these residues were shown to modify to high conversion via an oxidative coupling bioconjugation reaction (Figure 1B).[1,30] While this modification strategy is mild, fast, and efficient, the wild-type MS2 CP was observed to modify in less than 5% yield.

Using SyMAPS, we characterized the assembly competency of each variant in the P-X-X-X-MS2 library, generating an apparent fitness landscape (AFL). We generated a quantitative assembly score for every mutant in the targeted library by comparing the relative log% abundance of each variant before and after an assembly selection with size exclusion chromatography (SEC), identifying the subset of P-X-X-X-MS2 extensions that were competent for VLP assembly (Figure 3, Supplementary Figure 1). In addition, we generated a non-proline-terminated library, X-X-X-MS2, to distinguish which assembly trends were general and which were specific to a proline at position −4 (Supplementary Figure 2).

Of the 8000 variants, around 92% were observed in the starting plasmid library, consistent with coverage of previous SyMAPS libraries.[23,24] Of these, 48% were absent in the VLP library after the assembly selection, indicating that these extensions likely did not permit assembly. These low-scoring variants could be a result of mutations that are assembly incompetent, poorly expressed, or unstable to protein expression.[23] Around 24% of the variants scored apparent fitness score (AFS) values greater than 0.2, indicating that assembly occurred readily.[23] Variants with a nonsense mutation had an average AFS value of −3.0 with a standard deviation of 1.5, indicating that these sequences were depleted from the population of selected VLPs by 1000-fold.

We observed striking trends in the AFL when the data were grouped by the identity of the −1 position (or the position nearest to the native N terminus) (Figure 3). We evaluated the number of variants with P-X-X-Z-MS2 that were compatible with assembly (Figure 4A), where Z is the amino acid at the −1 position. Positive charge was particularly well-tolerated at this position and enabled a wide variety of extensions with the pattern P-X-X-[R/K]-MS2 (Figure 3, Figure 4A). This was surprising given the sterically hindered environment of the N terminus in the MS2 CP. Nearly 80% of extensions with the pattern P-X-X-R-MS2 assembled (AFS value >0.2), and over 60% of extensions with P-X-X-K-MS2 assembled, compared to merely 12% of P-X-X-D-MS2 and 8% of P-X-X-E-MS2. These results suggest that the beneficial effect is due specifically to positive charges rather than any charge at all.

Glycine and alanine, both common choices for rational N-terminal extensions, performed worse than expected compared to other amino acids, with 23% and 18% extensions permitted,

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

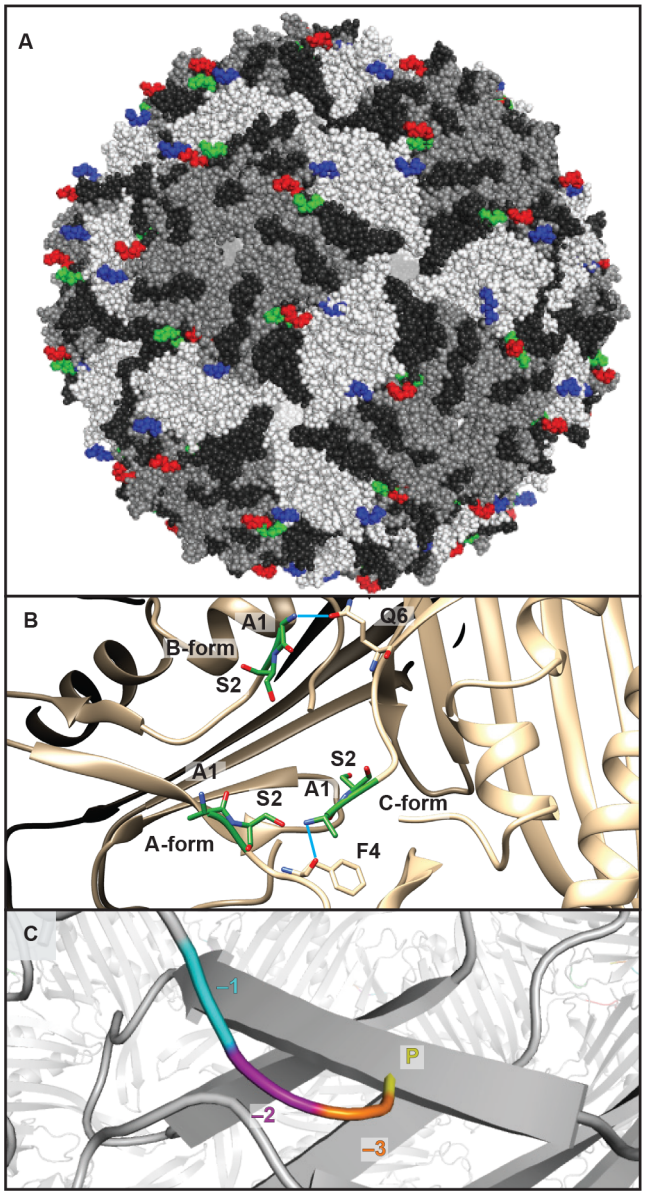


Figure 2. N termini of the MS2 capsid coat protein (MS2 CP) monomers. (A) Each quasi-equivalent form and N terminus is indicated with a shade of gray or color, respectively. The N terminus of the A form (dark gray) is shown in red; the N terminus of the B form (white) is shown in blue; and the N terminus of the C form (gray) is shown in green. (B) Hydrogen-bonding interactions are shown for the native N terminus. Hydrogen bonds are shown in blue. (C) The −1 (cyan), −2 (purple), −3 (orange), and −4 (proline, yellow) positions are indicated in relation to the native N terminus (alanine) of the MS2 CP.

respectively (Figure 4A). More intuitively, bulky residues such as tryptophan, phenylalanine, or leucine were poorly tolerated at the −1 position. In contrast, polar residues that can act as hydrogen donor and acceptors, such as serine, threonine, or asparagine, performed well. Asparagine was better tolerated than glutamine, indicating that side chain length may contribute to mutability of this position. Interestingly, histidine was also relatively well-tolerated and was the fifth most permitted amino acid at this position; however, only 40% of extensions with the pattern P-X-X-H-MS2 assembled, which is far lower than either arginine or lysine.

To visualize this effect, we plotted a histogram of AFS values with arginine or lysine at the −1 position compared to all other AFS values (Figure 4B). These residues in this position shift the average AFS values to be more positive, indicating that a higher percent of variants was compatible with self-assembly. Additionally, a histogram of arginine or lysine in the −1 position was compared with arginine or lysine at the −2 or −3 position to evaluate whether this effect was location specific. In this case, a notable shift to more positive AFS values was found with arginine or lysine only at the −1 position, suggesting the charge effect is indeed specific to this location (Figure 4C). A larger version of the data for arginine in position −1 appears in Figure 4D.

Finally, we confirmed that these trends were similar to N-terminal extensions in the absence of proline in the −4 position (Supplementary Figure 2). In this library, X-X-[K/R]-MS2 also resulted in a disproportionally high number of assembled particles compared to other amino acids at the −1 position, indicating that this trend is likely general for N-terminal extensions of the MS2 CP rather than specific to those starting with proline.

In order to evaluate the potential interactions responsible for this favorable effect, we performed a conformational search of a hexameric unit of P-A-A-R-MS2 (Supplementary Figure 3A). Most notably, a new salt bridge is formed in the *in silico* study between the N-terminal proline of B chain monomer and Asp17 of the C chain (Supplementary Figure 3B). In addition, hydrogen bonding is observed between arginine at the −1 position and Gln6 of the C chain in the minimized structure. We hypothesize that these hydrogen bonds are beneficial for assembly, as many extension combinations with a hydrogen bond donor residue at the −1 position are permitted.

In a conformational search of P-A-A-A-MS2 and P-A-R-A-MS2, both the hydrogen bond and salt bridge were not observed in either extension. These variants have lower AFL scores and lack a hydrogen bond donor side chain at the −1 position.

Finally, we analyzed P-A-R-R-MS2, a relatively poorly performing variant, via structure minimization. We found that while the −1 arginine formed the presumably beneficial salt bridge, multiple van der Waals clashing interactions were also found between arginine residues of the A and C chains (Supplementary Figure 3C). The stringent positional specificity of these interactions highlights the remarkable level of detail offered by a comprehensive mutational strategy such as SyMAPS.

Interpreting the Apparent Fitness Landscape. We evaluated the consistency of the data by plotting the two biological replicates of the P-X-X-X-MS2 data set as a scatterplot (Supplementary Figure 4A). In addition, we plotted the three biological replicates of the X-X-X-MS2 data set (Supplementary Figure 4B–D). In general, we find that the data sets do correlate, though the R^2 values are relatively low (0.42–0.59). We hypothesize that this variability may arise from a number of sources, including technical differences between assembly selections: for example, bacterial growth rates or expression levels are both variables that are not controlled that may affect the selections beyond assembly competency. Correlations between the two chemical modification selections are even lower (0.37, Supplementary Figure 4E), suggesting that significant variability may exist between replicates of the same challenge.

Interestingly, correlations within a replicate are somewhat higher, even when comparing chemical modification and

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

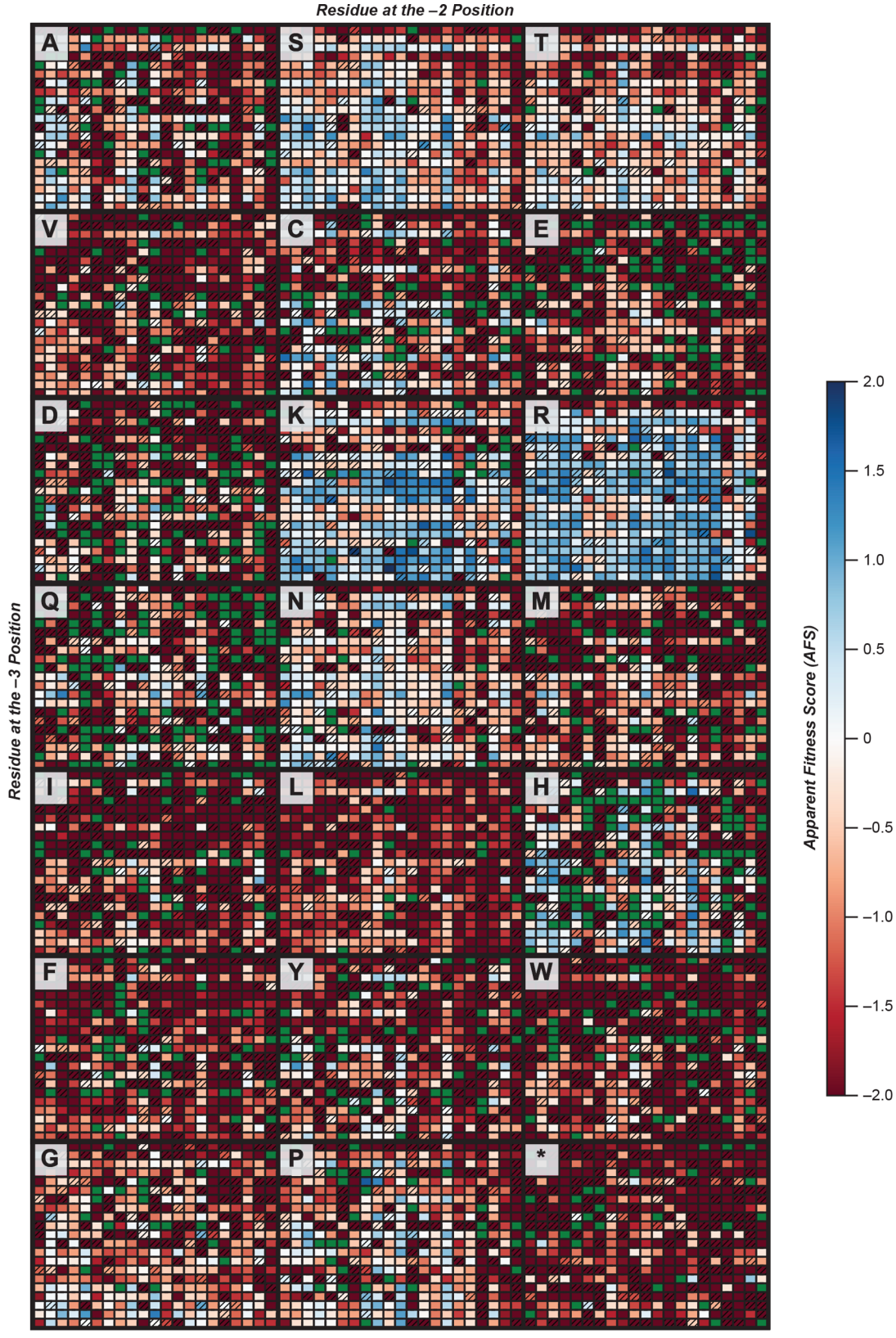


Figure 3. Apparent fitness landscape of P-X-X-X-MS2 N-terminal extensions. Extensions are labeled as the distance from the native N terminus (alanine), and the −1 position is indicated in the upper left corner of each quadrant. Blue indicates enriched variants and red indicates enriched combinations. Dark red indicates variants that were present in the plasmid library but absent in the VLP library. Missing values are shown in green. The unaveraged AFS is reported for variants with a missing value in a single replicate and is indicated by hatching. The nonsense mutations are marked with asterisks.

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

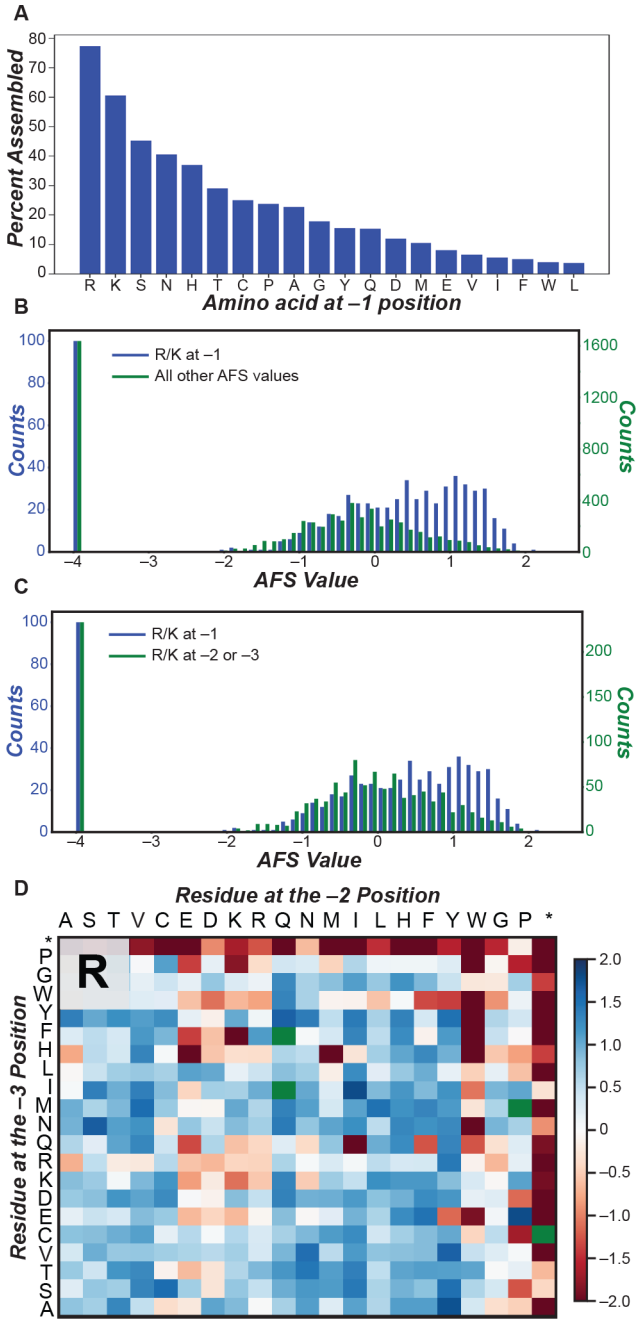


Figure 4. Effect of positive charge at the −1 position in N-terminal extensions. (A) The amino acid identity at the −1 position alters how many extensions are assembly competent. Arginine and lysine permit 77% and 61% of possible extensions with the pattern P-X-X4R/KFMS2. Arginine and lysine at position −1 result in a higher percent of positive AFS values (B) compared with all other AFS values or (C) compared to positive charge at −2 or −3. (D) The assembly scores of all P-X-X-R-MS2 extensions are shown as an example, in which arginine is at the −1 position.

assembly selections. While several extensions are positive in the assembly selection and negative in the chemical modification selection (as is to be expected for additional selective pressure), very few of the opposite are seen. Correlations for these are 0.52 and 0.67 for replicates 1 and 2, respectively (Supplementary Figure 4F,G). We find that the heat selection correlates well with the chemical selection for replicate one, yielding an R^2 of 0.75 and few off-axis data points (Supplementary Figure 4H). From these analyses, we hypothesize that replicate variability likely arises from growth or expression rather than the selections themselves.

Finally, we evaluated whether low abundances in the plasmid library contributed to the low correlation. Requiring at least two reads in both replicates for the P-X-X-X-MS2 data set did increase the correlation of the replicates to 0.52 from 0.42 (Supplementary Figure 4I), and further requiring at least 10 reads in both replicates increased the R^2 to 0.67. While increased stringency does improve correlation, our interpretation of this result is that factors beyond read abundances, such as biological noise, contribute to differences between the replicates.

To reduce the impact of stochastic variation on our data set and highlight variants with higher certainty in their determined fitness scores, we developed two additional methods for data processing. The first method filters out variants with low plasmid read counts (<4), as these are more prone to error. Though this reduces the coverage of our P-X-X-X library, we were pleased to find that many trends remain apparent in the filtered heatmaps (Supplementary Figures 5−7). For example, the stark favorable effect of positive charge at the −1 position on assembly competency was retained, as over half (54%) of all assembling variants bear a lysine or arginine at this position.

The second method of data processing aimed to remove all variants with ambiguous assembly competency and simplify the output to a binary "assembling" and "nonassembling" value. All variants with an AFS near 0 or an AFS that changed sign between replicants were removed from analysis. Variants with consistent high or low fitness scores were marked as "assembling" or "nonassembling" mutants, respectively (Supplementary Figures 8−10). This method of processing precludes detailed comparison of variant scores but allows for rapid selection of N-terminally extended MS2 variants with a clear assembly phenotype. While many of the trends discussed above are replicated in all methods of data analysis, we recommend using either of the high stringency methods to select individual extensions for further experiment. These supplementary processing methods serve as complementary approaches to interpreting SyMAPS data sets.

Direct Functional Selections for HiPerX Variants. The chemical modification of VLPs imposes a number of challenges to self-assembly, and any useful variant must tolerate reaction conditions as well as strain introduced by the covalent attachment of new functionality. As such, we designed a selection for tolerance to chemical modification conditions to identify variants that are well-suited for use as protein scaffolds. We used an N-terminal oxidative coupling reaction for this challenge.[1] The oxidative coupling uses a mild metal oxidant to convert methoxyphenols,[31] aminophenols,[1] and catechols[32] to ortho-quinone and ortho-iminoquinone intermediates that react selectively with anilines,[27] reduced cysteines,[33] and N-terminal amines of proteins or peptides.[1] In this study, we used aminophenols and catechols as ortho-quinone precursors, as both can be rapidly oxidized via $K_3Fe(CN)_6$ (Figure 1B).

The library was chemically coupled to DNA oligomers bearing *o*-aminophenol handles, simultaneously exposing the library to chemical modification conditions and to the strain of coupling large biomolecules to the VLP surfaces (Figure 5A). Variants that remained assembled under these conditions were enriched through HPLC SEC and sequenced. As with the assembly-selected AFL, we compared percent abundance of the library after the selection to the plasmid library to generate a quantitative score of chemical modification compatibility. As a

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

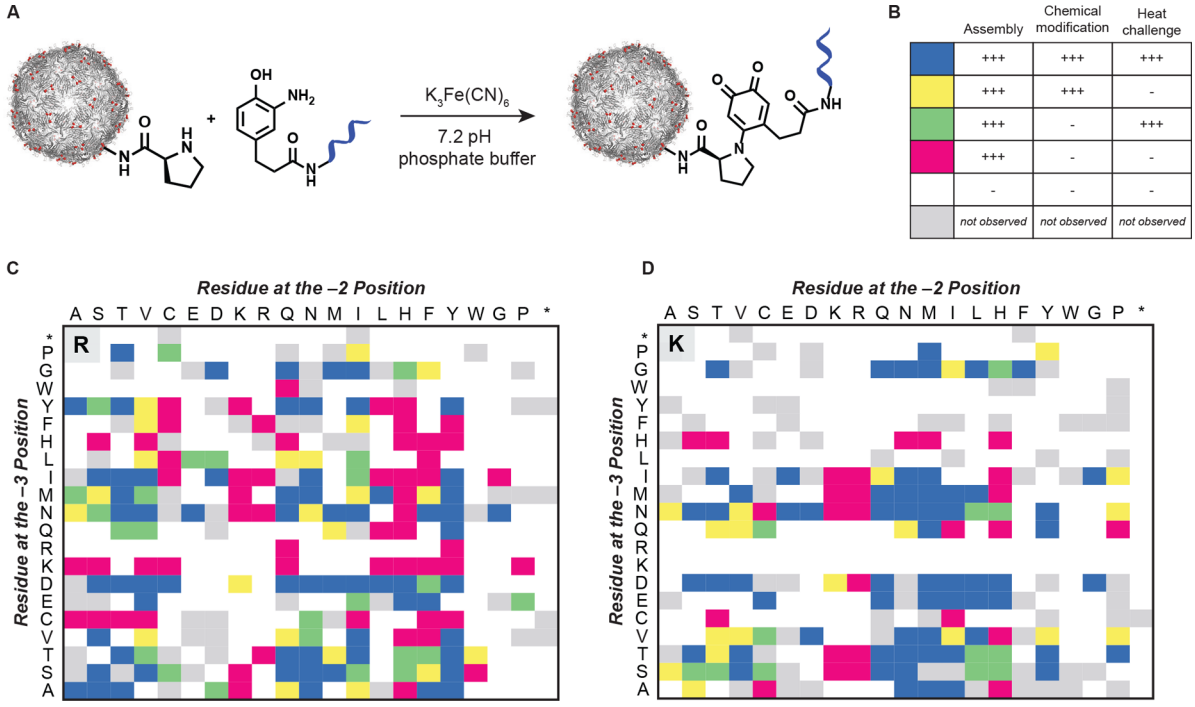


Figure 5. Combined fitness landscape of the P-X-X-X-MS2 N-terminal extensions. (A) The chemical modification-based selection of the variant library employs bioconjugation to a 25 bp DNA strand. (B) A color key is provided for the combined AFL data. (+++) indicates a score greater than 1.0 in the selection, and HiPerX, or high performing extensions, are indicated in blue. (−) indicates a score less than 1.0 in the selection. Combined AFLs are displayed for extensions (C) P-X-X-R-MS2 and (D) P-X-X-K-MS2. The full combined fitness landscape can be found in Supplemental Figure 7.

complement, we also evaluated the thermostability of all variants, subjecting the library to 50 °C for 10 min to differentiate between wild-type-like variants and those with reduced thermostability. As a comparison, the wild-type VLPs are stable up to 65 °C. Variants that remained assembled after this challenge were also purified by HPLC SEC, sequenced, and processed to generate a heat-selected AFL.

Surprisingly, the chemical modification selection was more stringent than the thermal selection: only 16% of the mutants were assembled following exposure to chemical modification conditions, while 22% of the mutants tolerated 50 °C for 10 min. In addition, chemical modification and thermostability scores showed stark differences in trends when compared to assembly-selected AFS values. While variants with multiple positive charges expressed and assembled far better than the library average (61% compared to 24%), these VLPs were almost universally sensitive to chemical and thermal challenges, suggesting that these types of extensions are unstable and therefore undesirable. Histidine behaved similarly in these challenges, and histidine at the −1 position when combined with positive charge at the −2 or −3 position was sensitive to thermal or chemical challenges. AFLs following thermal (Supplementary Figure 11) or chemical modification (Supplementary Figure 12) challenges present this phenomenon as distinct red bands within plots in which lysine and arginine are grouped by the −1 position. These data exhibit why functional challenges to variant libraries are crucial to disentangle subtle changes to VLP properties.

We next sought to generate insight into the variants performing well across all selections, which included assembly, thermal stability, and oxidative coupling selections. We generated an aggregate AFL that incorporated the results of each enrichment, in which a stringent threshold score for each parameter was used to isolate the most promising and useful variants. This aggregate AFL identified 238 thermally stable, chemically modifiable N-terminal extensions of the MS2 CP, indicated in blue (Figure 5B, Supplementary Figure 13) and termed high-performing extended (HiPerX) variants. Consistent with the findings above, 129 of these 238 variants possessed lysine or arginine at the −1 position, accounting for 54% of the HiPerX variants (Figure 5C,D). With a stringent score of 10-fold enrichment in all selections, most amino acids at the −1 position resulted in few or no HiPerX variants. Interestingly, unsuccessful sequences included glycine, which is commonly used in rational design to engineer extensions or linkers between protein domains.[34,35] Branched amino acids were also poorly tolerated at the −1 position: a comparison between serine and threonine at the −1 position revealed that threonine performed far worse than serine. Proline was better tolerated and outperformed glycine, even though these extensions have at least two proline residues in the first four amino acids.

We also found many nonintuitive results that diverged from common protein engineering assumptions. For example, tyrosine at the −2 or −3 position, when combined with arginine at the −1 position, was observed in many HiPerX variants. Combinations with multiple charges (P-D-H-R-MS2) or multiple large amino acids (P-S-Y-R-MS2) are also assembly competent, thermostable, and highly modifiable extensions. In particular, P-D-X-R-MS2 folded well across a broad range of X identities, such as when X was a small residue like serine, a hydrophobic residue such as isoleucine, or a polar, bulky residue like tyrosine. These results underscore the importance of experimental efforts to describe the mutability of large protein assemblies.

Bulky residues were tolerated at the −2 and −3 position in combination with arginine at the −1 position; however, by this metric, multiple positive charges were still detrimental to VLP stability. Even negative charge could not rescue stability in

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

nearly all of these cases. The only extensions with multiple positive charges with any increased stability are P-D-K-[K/R]-MS2, which are thermally stable but do not tolerate chemical modification. Additionally, glycine was only tolerated at the −3 position and, even then, only when there is a positively charged residue at the −1 position.

These trends, where multiply charged or bulky combinations of residues are permitted, are difficult to reconcile with the structure of the N terminus of the MS2 CP. For example, the close proximity of the monomer N termini means that an extension like P-S-Y-R-MS2 positions multiple large and/or charged residues within 9 to 12 Å. These results also contrast with most rational N-terminal extensions, which rely on small residues such as serine or glycine to disrupt the local protein folding environment minimally.[34,35]

We hypothesize that many of these mutations may enhance the critical charge interactions that make lysine and arginine desirable variants. For example, hydrophobic residues at the −2 position could create a more hydrophobic environment, reducing the local dielectric constant.[36,37] This in turn could strengthen the interactions involved in the proposed salt bridges. Alternatively, nonpolar residues in the −2 or −3 position could interact through hydrophobic effects. Regardless of the cause, in the absence of a systematic library approach and direct functional selections, these many nonintuitive yet critical findings would almost certainly have been missed. Ultimately, only 3% of the 8000 possible P-X-X-X-MS2 extensions were identified as HiPerX variants, enriched in assembly, thermal stability, and chemical modification.

Characterization and Modification of HiPerX Variants. Based on the stringent selection conditions, HiPerX variants were expected to have increased tolerance to chemical reaction conditions; however, it was not known whether the N termini of these variants would be modified at higher rates than CP[WT]. As such, we sought to validate trends identified in high-throughput sequencing and to characterize the usefulness of HiPerX variants as protein scaffolds. To do so, five randomly selected HiPerX variants with P-X-X-R-MS2 extensions were cloned and evaluated individually. These variants were selected because this population showed the largest enrichment across all challenges. All five variants expressed in high yield, formed assembled VLPs, and tolerated the thermal challenge of 50 °C for 10 min, supporting the quality of the AFLs (Supplementary Figure 14).

We next evaluated whether the engineered extensions indeed enhanced reactivity to the N-terminal oxidative coupling reaction. We performed a reactivity test to modify the VLP N termini with a small-molecule catechol derivative (Figure 6A). Gratifyingly, all tested variants showed a significant enhancement in reaction conversion compared to the wild-type MS2 CP. HiPerX variants showed 36−87% modification, compared to <5% modification in wild-type VLPs (Figure 6B). The dramatic increase in conversion under these conditions is notable in and of itself; additionally, as there are 180 MS2 CPs per VLP, these N-terminally extended variants are capable of displaying up to 65−160 copies of the new functionality per VLP, representing a substantial increase in targeting or drug-carrying capabilities. HPLC SEC of modified samples confirmed that all of these VLPs remained assembled after modification (Supplementary Figure 15). This result shows that, for the first time, SyMAPS can be combined with a chemical modification enrichment to identify highly modifiable variants. In addition, given that all five randomly selected variants modified at higher rates than

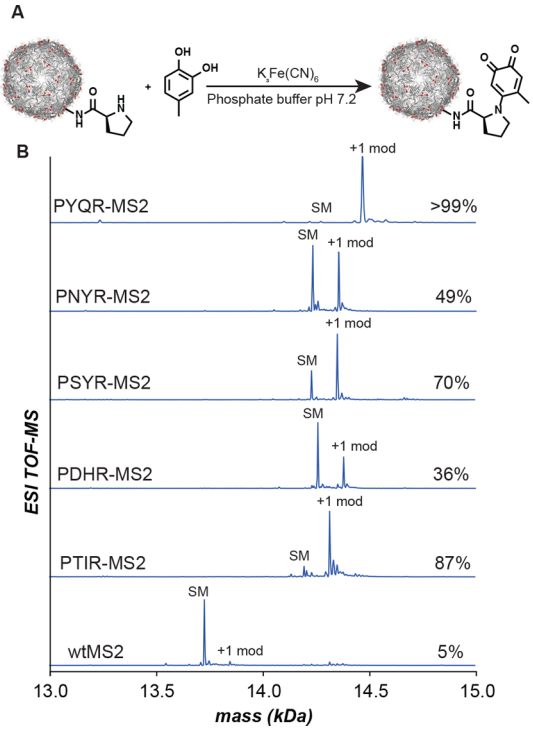


Figure 6. Chemical modification of HiPerX MS2 variants. (A) An oxidative coupling reaction was evaluated for proline-terminated MS2 variants. (B) Mass spectra of chemically modified HiPerX variants of the MS2 CP are shown. Percent modification is determined by integration of the unmodified (SM) vs modified (+1 mod) peaks.

CP[WT]—and because we expect the N-terminal prolines to be solvent accessible—we anticipate that many other HiPerX variants will also be amenable to modification.

We next sought to evaluate whether these extensions were compatible with other bioconjugation strategies (Figure 7A−C). One such N-terminal modification strategy using 2-pyridinecarboxaldehyde (2PCA) modifies most N-terminal residues to high yield in a single step through a mechanism that is distinct from oxidative coupling reactions (Figure 7B).[2] These two chemistries do not share common intermediates and proceed under different reaction conditions. To evaluate HiPerX variant performance with 2PCA, extended variants and CP[WT] were incubated with excess reagent overnight at room temperature, according to the published protocol.[2] We observed that HiPerX variants resulted in around 80% modification with 2PCA, compared to around 30% modification with CP[WT], even though these extensions were optimized for the $K_3Fe(CN)_6$ oxidative coupling reaction (Figure 7D). While the fold improvement was lower for this reaction, an increase from 30% modification (CP[WT]) to 80% (CP[HiPerX]) modification represents a useful increase in the number of functional groups installed on the exterior, from 50 modifications to 140 modifications (Figure 7E).

We also investigated a new tyrosinase-mediated variant of the oxidative coupling reaction (abTYR, tyrosinase from *Agariucus bisporus*) that proceeds through a similar mechanism to $K_3Fe(CN)_6$ oxidative coupling after the ortho-quinone intermediate is produced (Figure 7C).[30] The enzymatic oxidation is compatible with phenols as well as catechols; thus, compatibility with abTYR would widen the scope of potential small-molecule partners to include many shelf-stable phenols. We found that modification yields with catechols increased from good (36−

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

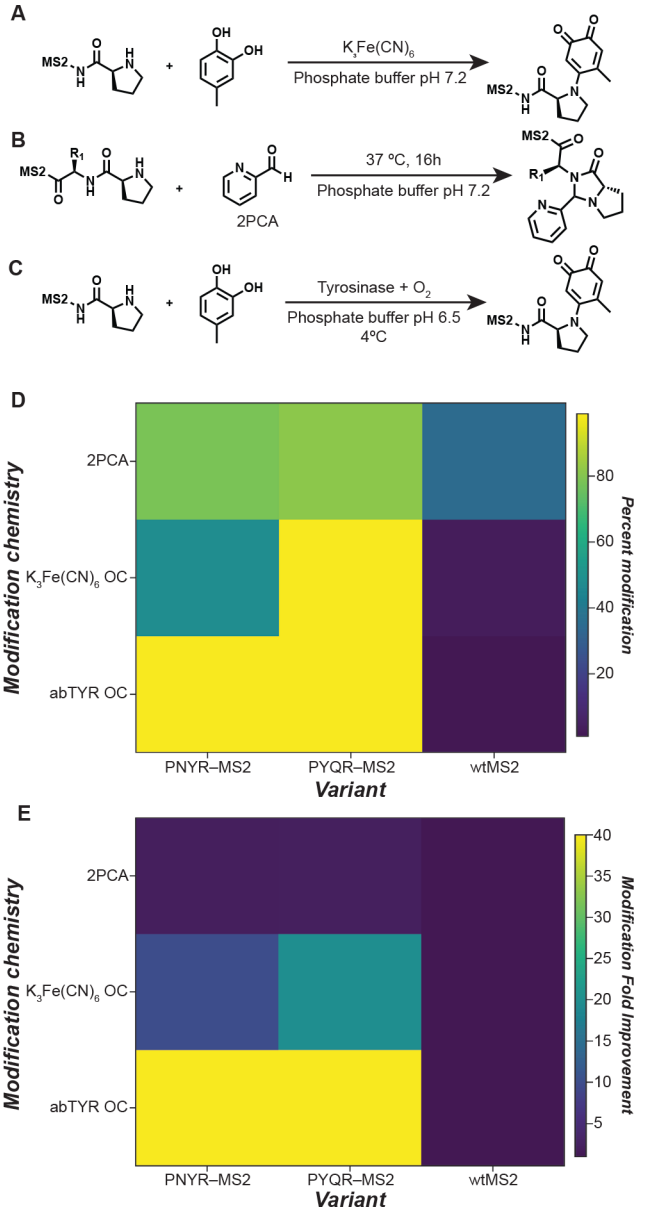


Figure 7. Conversion and fold improvement of N-terminal modification strategies of HiPerX MS2 variants. Reaction schemes are shown for (A) potassium ferricyanide-mediated oxidative coupling, (B) 2-pyridinecarboxaldehyde (2PCA) modification, and (C) tyrosinase-mediated (abTYR) oxidative coupling reactions. (D) Modification of two HiPerX MS2 variants is shown in contrast to wild-type MS2 across these modification strategies. (E) Fold improvement compared to wild-type MS2 is shown.

87%) to near-quantitative (>99%) in all cases (Figure 7E). In addition, CP[PQYR] was found to be compatible with installation and modification of a reactive cysteine in the interior cavity.[38] Interior labeling was performed with an AlexaFluor-488 maleimide dye, and modification efficiency with this strategy was high (>99%), as previously reported.[16,38−40] More importantly, subsequent exterior modification via abTYR-mediated oxidative coupling also proceeded to over 99% conversion, resulting in doubly modified VLPs with 180 copies of both functionalities (Supplementary Figure 16). Altogether, these extensions are thermally stable and highly modifiable and can carry cargo, making them promising carriers with highly desirable properties for a number of biomedical applications.

Extensions Are Well-Assembled and Modified in Combination with CP[S37P]. Previous work in our lab identified a variant of the MS2 VLP with altered quaternary geometry.[41] This CP[S37P] mutation alters the global structure from a 27 nm wild-type-sized VLP to a smaller, 17 nm VLP (Figure 8A). This smaller-sized variant retains similar thermo-

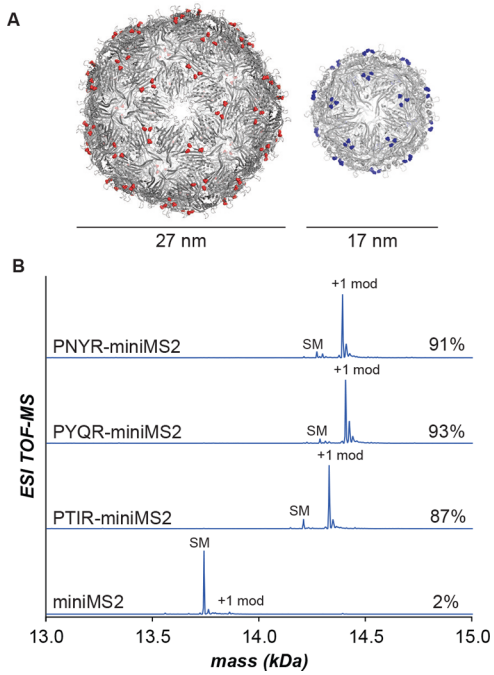


Figure 8. Chemical modification of HiPerX miniMS2 variants (CP[HiPerX-S37P]. (A) Crystal structures of CP[WT] and CP[S37P] are shown with N termini highlighted in red and blue, respectively. (B) Mass spectra of chemically modified HiPerX variants of the MS2 CP are shown.

stability and is a useful tool to probe the effect of carrier size directly in applications such as drug delivery or imaging.[42] However, the N terminus of the CP[S37P] is distinct from CP[WT], both in minor structural differences and spatial positioning. To date, the exterior of CP[S37P] has not been modified, and its N terminus is sterically unavailable, similar to the parent CP[WT].[24]

We sought to determine whether HiPerX sequences could be appended to the CP[S37P] structure, enabling facile modification without repeating the library generation and functional selections. Despite the differences in geometry and secondary structure, all three N-terminally extended CP[HiPerX-S37P] variants assembled into well-formed VLPs. Each variant retained the T = 1 geometry and smaller size, as confirmed by dynamic light scattering (Supplementary Figure 17A). Additionally, variants tolerated 50 °C for 10 min, indicating that thermostability was preserved in the new genetic background (Supplementary Figure 17B).

We next modified the exterior of the N-terminally extended CP[S37P] variants with the $K_3Fe(CN)_6$ oxidative coupling reaction, appending a catechol small molecule to the N-terminus. We found that CP[HiPerX-S37P] variants modified equally as well as the parent HiPerX variants, achieving >85% modification in all cases (Figure 8B). As a comparison, CP[S37P] modified <5%, indicating that the extensions are critical to achieve high modification rates. Despite changes to

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

surface curvature and quaternary structure geometry, the selected HiPerX variants performed remarkably well as useful N-terminal extensions with CP[S37P]. Furthermore, this presents the first successful exterior modification of MS2 CP[S37P], enabling future study of a 17 nm VLP variant as a targeted protein scaffold.

CONCLUSION

The site-specific modification of proteins is of fundamental importance for many applications, including drug delivery, vaccines, and protein biomaterials. Here, we combined a systematically generated library with a functional selection under chemical modification conditions to identify variants of the MS2 CP that are highly compatible with N-terminal modification. The fact that only 3% of the library were enriched after the full set of challenges underscores the fact that the introduction of non-native amino acids into proteins remains a nonintuitive process a priori. This is particularly true in the case of self-assembling proteins, as single-point mutations lead to amplified effects when propagated throughout the quaternary structure. In this study, an unexpected charge interaction was uncovered that counters these effects and, in some cases, was bolstered by additional hydrophobic interactions. The selection procedure for bioconjugation conditions could be used with many future libraries to identify new reactive sequences. Finally, the MS2 CP variants identified in this study can be doubly modified to >99% yield on both the interior and exterior surfaces, providing homogeneous carrier materials in two different sizes for a variety of drug delivery applications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b10734.

Excel file of sequences (XLSX)
Jmol files (ZIP)
Experimental procedures and additional figures (PDF)
Raw data (ZIP)

AUTHOR INFORMATION

Corresponding Authors

*ercek@northwestern.edu
*mbfrancis@berkeley.edu

ORCID

Danielle Tullman-Ercek: 0000-0001-6734-480X
Matthew B. Francis: 0000-0003-2837-2538

Author Contributions

⊥D. D. Brauer and E. C. Hartman contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Army Research Office (W911NF-15-1-0144 and W911NF-16-1-0169), the BASF CARA program, and the Chemical Biology Graduate Program at UC Berkeley (NIH T32-GM066698). E.C.H. was supported by the DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. The sequencing was carried by the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant S10OD010786.

REFERENCES

(1) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. N-Terminal Modification of Proteins with *O* -Aminophenols. *J. Am. Chem. Soc.* **2014**, *136* (27), 9572−9579.

(2) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. One-Step Site-Specific Modification of Native Proteins with 2-Pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **2015**, *11* (5), 326−331.

(3) Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M.; Weng, C.-H.; Iavarone, A. T.; Francis, M. B. Site-Specific Protein Transamination Using *N* -Methylpyridinium-4-Carboxaldehyde. *J. Am. Chem. Soc.* **2013**, *135* (45), 17223−17229.

(4) Sur, S.; Qiao, Y.; Fries, A.; O'Meally, R. N.; Cole, R. N.; Kinzler, K. W.; Vogelstein, B.; Zhou, S. PRINT: A Protein Bioconjugation Method with Exquisite N-Terminal Specificity. *Sci. Rep.* **2016**, *5* (1), DOI: 10.1038/srep18363.

(5) Spicer, C. D.; Pashuck, E. T.; Stevens, M. M. Achieving Controlled Biomolecule−Biomaterial Conjugation. *Chem. Rev.* **2018**, *118* (16), 7702−7743.

(6) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. A New Ligation Method for N-Terminal Tryptophan-Containing Peptides Using the Pictet−Spengler Reaction. *Tetrahedron Lett.* **2000**, *41* (21), 4069−4073.

(7) Geoghegan, K. F.; Stroh, J. G. Site-Directed Conjugation of Nonpeptide Groups to Peptides and Proteins via Periodate Oxidation of a 2-Amino Alcohol. Application to Modification at N-Terminal Serine. *Bioconjugate Chem.* **1992**, *3* (2), 138−146.

(8) Casi, G.; Huguenin-Dezot, N.; Zuberbühler, K.; Scheuermann, J.; Neri, D. Site-Specific Traceless Coupling of Potent Cytotoxic Drugs to Recombinant Antibodies for Pharmacodelivery. *J. Am. Chem. Soc.* **2012**, *134* (13), 5887−5892.

(9) Palla, K. S.; Witus, L. S.; Mackenzie, K. J.; Netirojjanakul, C.; Francis, M. B. Optimization and Expansion of a Site-Selective *N* -Methylpyridinium-4-Carboxaldehyde-Mediated Transamination for Bacterially Expressed Proteins. *J. Am. Chem. Soc.* **2015**, *137* (3), 1123−1129.

(10) Rosen, C. B.; Francis, M. B. Targeting the N Terminus for Site-Selective Protein Modification. *Nat. Chem. Biol.* **2017**, *13* (7), 697−705.

(11) Li, D.; Han, B.; Wei, R.; Yao, G.; Chen, Z.; Liu, J.; Poon, T. C. W.; Su, W.; Zhu, Z.; Dimitrov, D. S.; et al. N-Terminal α-Amino Group Modification of Antibodies Using a Site-Selective Click Chemistry Method. *mAbs* **2018**, *10* (5), 712−719.

(12) Lee, J. P.; Kassianidou, E.; Macdonald, J. I.; Francis, M. B.; Kumar, S. N-Terminal Specific Conjugation of Extracellular Matrix Proteins to 2-Pyridinecarboxaldehyde Functionalized Polyacrylamide Hydrogels. *Biomaterials* **2016**, *102*, 268−276.

(13) Esser-Kahn, A. P.; Iavarone, A. T.; Francis, M. B. Metallothionein-Cross-Linked Hydrogels for the Selective Removal of Heavy Metals from Water. *J. Am. Chem. Soc.* **2008**, *130* (47), 15820−15822.

(14) ElSohly, A. M.; Netirojjanakul, C.; Aanei, I. L.; Jager, A.; Bendall, S. C.; Farkas, M. E.; Nolan, G. P.; Francis, M. B. Synthetically Modified Viral Capsids as Versatile Carriers for Use in Antibody-Based Cell Targeting. *Bioconjugate Chem.* **2015**, *26* (8), 1590−1596.

(15) Ashley, C. E.; Carnes, E. C.; Phillips, G. K.; Durfee, P. N.; Buley, M. D.; Lino, C. A.; Padilla, D. P.; Phillips, B.; Carter, M. B.; Willman, C. L.; et al. Cell-Specific Delivery of Diverse Cargos by Bacteriophage MS2 Virus-like Particles. *ACS Nano* **2011**, *5* (7), 5729−5745.

(16) Aanei, I. L.; Huynh, T.; Seo, Y.; Francis, M. B. Vascular Cell Adhesion Molecule-Targeted MS2 Viral Capsids for the Detection of Early-Stage Atherosclerotic Plaques. *Bioconjugate Chem.* **2018**, *29* (8), 2526−2530.

(17) Farkas, M. E.; Aanei, I. L.; Behrens, C. R.; Tong, G. J.; Murphy, S. T.; O'Neil, J. P.; Francis, M. B. PET Imaging and Biodistribution of Chemically Modified Bacteriophage MS2. *Mol. Pharmaceutics* **2013**, *10* (1), 69−76.

(18) Crossey, E.; Frietze, K.; Narum, D. L.; Peabody, D. S.; Chackerian, B. Identification of an Immunogenic Mimic of a Conserved

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884

Epitope on the Plasmodium Falciparum Blood Stage Antigen AMA1 Using Virus-like Particle (VLP) Peptide Display. *PLoS One* **2015**, *10* (7), No. e0132560.
(19) Zhai, L.; Peabody, J.; Pang, Y.-Y. S.; Schiller, J.; Chackerian, B.; Tumban, E. A Novel Candidate HPV Vaccine: MS2 Phage VLP Displaying a Tandem HPV L2 Peptide Offers Similar Protection in Mice to Gardasil-9. *Antiviral Res.* **2017**, *147*, 116−123.
(20) Capehart, S. L.; Coyle, M. P.; Glasgow, J. E.; Francis, M. B. Controlled Integration of Gold Nanoparticles and Organic Fluorophores Using Synthetically Modified MS2 Viral Capsids. *J. Am. Chem. Soc.* **2013**, *135* (8), 3011−3016.
(21) Glasgow, J. E.; Asensio, M. A.; Jakobson, C. M.; Francis, M. B.; Tullman-Ercek, D. Influence of Electrostatics on Small Molecule Flux through a Protein Nanoreactor. *ACS Synth. Biol.* **2015**, *4* (9), 1011−1019.
(22) Glasgow, J. E.; Capehart, S. L.; Francis, M. B.; Tullman-Ercek, D. Osmolyte-Mediated Encapsulation of Proteins inside MS2 Viral Capsids. *ACS Nano* **2012**, *6* (10), 8658−8664.
(23) Hartman, E. C.; Jakobson, C. M.; Favor, A. H.; Lobba, M. J.; Álvarez-Benedicto, E.; Francis, M. B.; Tullman-Ercek, D. Quantitative Characterization of All Single Amino Acid Variants of a Viral Capsid-Based Drug Delivery Vehicle. *Nat. Commun.* **2018**, *9* (1), 1385.
(24) Hartman, E. C.; Lobba, M. J.; Favor, A. H.; Robinson, S. A.; Tullman-Ercek, D.; Francis, M. B. Experimental Evaluation of Coevolution in a Self-Assembling Particle. *Biochemistry* **2018**, DOI: 10.1021/acs.biochem.8b00948.
(25) Peabody, D. S. Subunit Fusion Confers Tolerance to Peptide Insertions in a Virus Coat Protein. *Arch. Biochem. Biophys.* **1997**, *347* (1), 85−92.
(26) Hooker, J. M.; Esser-Kahn, A. P.; Francis, M. B. Modification of Aniline Containing Proteins Using an Oxidative Coupling Strategy. *J. Am. Chem. Soc.* **2006**, *128* (49), 15558−15559.
(27) Behrens, C. R.; Hooker, J. M.; Obermeyer, A. C.; Romanini, D. W.; Katz, E. M.; Francis, M. B. Rapid Chemoselective Bioconjugation through Oxidative Coupling of Anilines and Aminophenols. *J. Am. Chem. Soc.* **2011**, *133* (41), 16398−16401.
(28) Ni, C.-Z.; Syed, R.; Kodandapani, R.; Wickersham, J.; Peabody, D. S.; Ely, K. R. Crystal Structure of the MS2 Coat Protein Dimer: Implications for RNA Binding and Virus Assembly. *Structure* **1995**, *3* (3), 255−263.
(29) Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-Terminal Methionine Excision from Escherichia Coli Proteins Is Governed by the Side-Chain Length of the Penultimate Amino Acid. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (21), 8247−8251.
(30) Maza, J.; Bader, D. L. V.; Xiao, L.; Marmelstein, A. M.; Brauer, D. D.; ElSohly, A. M.; Smith, M. J.; Francis, M. B. Enzymatic Modification of N-Terminal Proline Residues Using Simple Phenol Derivatives. *J. Am. Chem. Soc.***2019**, DOI: 10.1021/jacs.8b10845.
(31) ElSohly, A. M.; MacDonald, J. I.; Hentzen, N. B.; Aanei, I. L.; El Muslemany, K. M.; Francis, M. B. *Ortho* -Methoxyphenols as Convenient Oxidative Bioconjugation Reagents with Application to Site-Selective Heterobifunctional Cross-Linkers. *J. Am. Chem. Soc.* **2017**, *139* (10), 3767−3773.
(32) Furst, A. L.; Smith, M. J.; Lee, M. C.; Francis, M. B. DNA Hybridization To Interface Current-Producing Cells with Electrode Surfaces. *ACS Cent. Sci.* **2018**, *4* (7), 880−884.
(33) Obermeyer, A. C.; Jarman, J. B.; Netirojjanakul, C.; El Muslemany, K.; Francis, M. B. Mild Bioconjugation Through the Oxidative Coupling of *Ortho* -Aminophenols and Anilines with Ferricyanide. *Angew. Chem., Int. Ed.* **2014**, *53* (4), 1057−1061.
(34) Reddy Chichili, V. P.; Kumar, V.; Sivaraman, J. Linkers in the Structural Biology of Protein-Protein Interactions. *Protein Sci.* **2013**, *22* (2), 153−167.
(35) Klein, J. S.; Jiang, S.; Galimidi, R. P.; Keeffe, J. R.; Bjorkman, P. J. Design and Characterization of Structured Protein Linkers with Differing Flexibilities. *Protein Eng., Des. Sel.* **2014**, *27* (10), 325−330.
(36) Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Velu, P. D.; Garcia-Moreno, E. B. Charges in the Hydrophobic Interior of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (37), 16096−16100.
(37) Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; García-Moreno, E. B. High Apparent Dielectric Constants in the Interior of a Protein Reflect Water Penetration. *Biophys. J.* **2000**, *79* (3), 1610−1620.
(38) Tong, G. J.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B. Viral Capsid DNA Aptamer Conjugates as Multivalent Cell-Targeting Vehicles. *J. Am. Chem. Soc.* **2009**, *131* (31), 11174−11178.
(39) Stephanopoulos, N.; Tong, G. J.; Hsiao, S. C.; Francis, M. B. Dual-Surface Modified Virus Capsids for Targeted Delivery of Photodynamic Agents to Cancer Cells. *ACS Nano* **2010**, *4* (10), 6014−6020.
(40) Aanei, I. L.; Francis, M. B. Dual Surface Modification of Genome-Free MS2 Capsids for Delivery Applications. In *Virus-Derived Nanoparticles for Advanced Technologies*; Wege, C., Lomonossoff, G. P., Eds.; Springer New York: New York, NY, 2018; Vol. *1776*, pp 629−642.
(41) Asensio, M. A.; Morella, N. M.; Jakobson, C. M.; Hartman, E. C.; Glasgow, J. E.; Sankaran, B.; Zwart, P. H.; Tullman-Ercek, D. A Selection for Assembly Reveals That a Single Amino Acid Mutant of the Bacteriophage MS2 Coat Protein Forms a Smaller Virus-like Particle. *Nano Lett.* **2016**, *16* (9), 5944−5950.
(42) Gaumet, M.; Vargas, A.; Gurny, R.; Delie, F. Nanoparticles for Drug Delivery: The Need for Precision in Reporting Particle Size Parameters. *Eur. J. Pharm. Biopharm.* **2008**, *69* (1), 1−9.

DOI: 10.1021/jacs.8b10734
J. Am. Chem. Soc. 2019, 141, 3875−3884