Title

Systematic Mutagenesis of MS2 Virus-Like Particles for Engineered Properties and Computational Models of Assembly

Author

Brauer, Daniel

Publication Date

2021

Peer reviewed|Thesis/dissertation

Systematic Mutagenesis of MS2 Virus-Like Particles for Engineered Properties and Computational Models of Assembly

by

Daniel Brauer

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Matthew B. Francis, Chair
Professor Evan Miller
Professor Michi Taga

Fall 2021

Systematic Mutagenesis of MS2 Virus-Like Particles for Engineered Properties and Computational Models of Assembly

Abstract

Systematic Mutagenesis of MS2 Virus-Like Particles for Engineered Properties and Computational Models of Assembly

by

Daniel Brauer

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Matthew B. Francis, Chair

Virus-like particles (VLPs) are nanoscale proteinaceous materials that show promise as scaffolds for a plethora of applications, including vaccine development, targeted drug delivery, and nanoreactor production. These self-assembling structures are based on viral architectures, but lack genetic material to cause infection. Though theoretically particle features may be finely tailored via genetic manipulation, the phenotypic consequences of mutations to self-assembling proteins remains hard to predict. Throughout this work, advancements in both the functional engineering of VLPs and the fundamental understanding of VLP design constraints are described. A method for selection of chemically modifiable capsids was developed and applied to a systematic library of N-terminally extended bacteriophage MS2 VLPs, resulting in highly reactive capsids for site-specific bioconjugation. The one-dimensional fitness landscape of a non-native MS2 assembly configuration was also constructed and used to develop machine learning models for assembly state prediction. This work highlights the utility of fitness landscaping in producing materials with applications-driven properties and explores the possibilities of *in silico* modeling for targeted engineering of VLPs. Furthermore, efforts to assess and improve the equity of the Chemistry academic community for marginalized populations are discussed. An annual survey framework was developed for generating quantitative understanding of departmental climate over time, and longitudinal improvements in departmental inclusivity were observed. This method was also used to critically assess how academic values of underrepresented graduate students are not well represented in faculty modes of evaluation. The developed framework is readily adaptable to other institutions aiming to improve the inclusion of historically minoritized groups in STEM.

To my parents, Suzanne and David, who always encouraged me to be curious.

Contents

List of Figures

List of Tables

Acknowledgments

There are more people that supported me in some way throughout my Ph. D. than I can possibly list, but I'll give it a shot regardless. Although it was a challenging road, I can truly say I enjoyed my time in graduate school thanks to all of the friends I've made along the way.

First, I would like to thank Annette and Daniel, my grandparents who taught me the value of an education and brought me home from school almost every day. I would like to thank Matt Francis, for being an absolutely incredible graduate school mentor. Your commitment to your students as people first and your boundless scientific enthusiasm fostered an lab environment that was a joy to be a part of. To Danielle and the Northwestern group, thank you for being brilliant and supportive collaborators even from 2000 miles away. To my first-year cohort, thank you for struggling through Saturday review sessions with me after late Friday nights. To the Francis lab, thank you for building such a supportive lab environment, pushing me on the ski slopes, and keeping the best softball sportsmanship streak going. To Emily, thank you for being such an inspirational mentor and role model. To all of the iterations of 733, you've made work a great place for antics, over-caffeination, and taco/boba runs. To Kristin, I think I still owe you a poke bowl. To Zoe, thank you for making mentorship easy, keeping me excited about our work, and being such a mensch. To Bader, for being the number 2 Dan, and staying a great friend around the world. To Zaneta and Daria, thank you for pulling me away from work to relive the early 2000's, you're always welcome to Kramer (now it's in public record). To Chrissy, Hikaru, Jamie and Sonja on the climate survey team, thank you for giving me a place to vent and believing we can help the community do better. To Terrace House Thursdays, for coward-rice, *vibes* and too much snacking, To the DND crew, thank you for giving me an outlet for creativity, drama, and some accents of questionable quality. Thank you to Tucci, for making it impossible to leave the couch and being probably the best dog who has ever lived?

Most of all thank you to Alina for all of your love and support this entire time. I'll be forever grateful that you found me just about the second I arrived in Berkeley. For being a dance partner, a couch potato, a travel buddy, a sous-chef, a hiking pal, and the spooky silly person you are. I can't wait to see what's next.

Chapter 1

Introduction

1.1 Self-assembling protein nanomaterials

Proteins not only perform a wide array of functions as individual folded structures, but also self-assemble into a myriad of complex superstructures. These assemblies range from complicated molecular machinery such as the ribosome to key structural complexes of the cytoskeleton. Closed-shell protein assemblies, such as those that form the capsids of viruses, have become increasingly attractive as candidates for biotechnological applications. These nanoscale particles are well-suited for use as drug delivery vehicles, imaging agents, bioreactors, and vaccine scaffolds. While nature has provided valuable initial templates for such purposes, these materials must be tailored to fit each particular use case. In principle, genetic manipulation of the constituent proteins of a protein assembly enables fine tuning of the properties and molecular features of an assembled material. In practice, however, the phenotypic consequences of amino acid mutations, insertions, and deletions in self-assembled materials are difficult to predict.

This work aimed to build up understanding of the design constraints of protein nanocages using new high-throughput methodology for assessing self-assembly. We not only employ these tools to improve fundamental understanding of a self-assembled system, but also construct and implement new functional selections for self-assembled materials with various improved properties, such as thermostability and chemical reactivity. We further extend our exploration beyond the scope of the natural assembly geometry of a protein nanocage to evaluate the comparative mutational landscape of a non-native assembly phenotype. Lastly, we utilize the generated empirical assembly fitness information to benchmark *in silico* tools for modeling the energetic consequences of missense mutations in proteins. This information in tandem with our

experimental mutational information then serves as the basis for training of machine learning models for predicting the assembly state of mutated protein capsids.

1.1.1 Virus-like particles

Derived from viral structures, virus-like particles (VLPs) represent a class of protein-based nanomaterials that have attracted considerable attention as biotechnological platforms. First isolated from hepatitis patient sera in 1968, these structures self-assemble from components of viral structural proteins, but lack viral genetic material and are thus non-infectious.[1] To date, over 110 VLPs from 35 viral families have been identified, with sizes ranging from 10-200 nm.[2] VLPs encompass many levels of structural complexity: human papillomavirus (HPV) VLPs form readily from a single capsid protein, VP1,[3] while poliovirus VLPs require three structural proteins to form a closed-shell.[4] The incorporation of non-proteinaceous elements such as an outer lipid envelope present in Zika virus VLPs adds another layer of complexity.[5] Further, the structural diversity of viruses are reflected in VLPs, with spherical, polyhedral, and rod-like geometries observed (Figure 1.1)[6]

1.1.2 Biomedical Applications of VLPs

VLPs have many attractive features that lend themselves to emerging biotechnologies as well. As protein-based materials, VLPs are naturally biodegradable. They are relatively inexpensive to produce, and their homogenous size distribution allows for consistent quality control.[7] Many VLPs are patterned with nanoscale surface pores, enabling diffusion of small molecules to the interior lumen of the particles. In some cases, VLPs can be disassembled under particular pH or ionic strength and reassembled around proteinaceous, genetic, or even polymeric cargoes.

1.1.2.1 VLPs as vaccine scaffolds

The structural similarity of VLPs to active virions naturally sparked the pursuit of their use as improved vaccines over classical inactivated or live attenuated pathogens.[8] As they maintain a similar antigenic conformation to their parent viruses, VLPs represent a vaccine scaffold that may be produced safely in heterologous hosts. VLPs have been successfully expressed and purified from bacteria, yeast, insect, plant, and mammalian systems at high titers. They have been shown to engage dendritic cells through binding to surface pattern recognition receptors, resulting in high immunogenicity.[9] The first approved VLP vaccine, Recombivax HB for hepatitis B virus, was licensed for human use in 1986.[10] Recombivax HB is produced recombinantly via expression

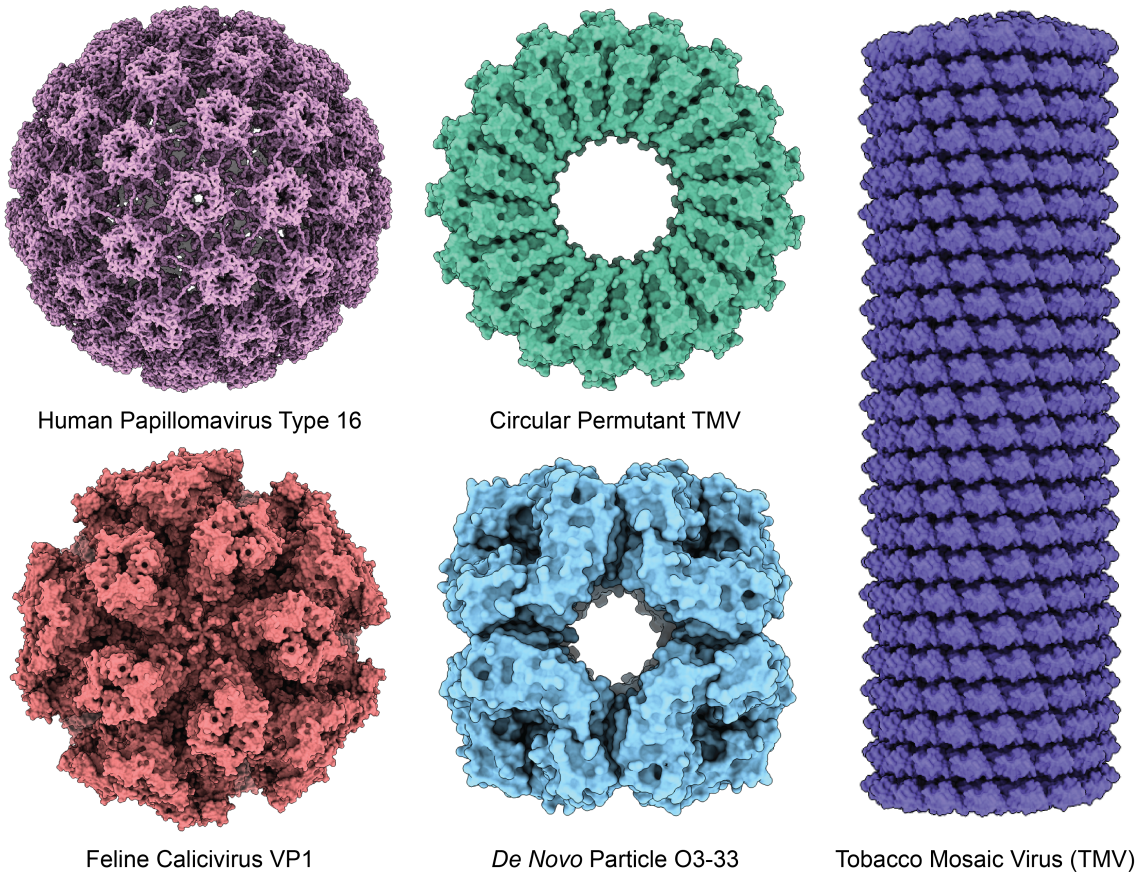

Figure 1.1: A sampling of structural geometries virus-like particles can adopt. HPV16 (PDB: 3J6R), cpTMV (PDB: 6X0R), calicivirus VP1 (PDB: 4PB6), O3-33 (PDB: 6FDB), and TMV (PDB: 3J06) are shown.

of the hepatitis B surface antigen (HBsAg) in *Saccharomyces cerevisiae*, drastically improving production over previous methods that relied on purification of HBsAg particles from human plasma. Several VLP vaccines have since been approved for clinical use, including Gardasil-9 for HPV and Hecolin for HEV.[11] These more recent examples offer protection against multiple pathogenic genotypes simultaneously by utilizing multiple antigenic subunits in a single vaccine formulation.

Several VLP platforms that aim to immunize against non-infectious vaccination targets are in clinical trial phases. These technologies utilize either genetic fusion or chemical bioconjugation to install antigens on the exterior surface of a core VLP. Using a bifunctional succinimate linker to covalently modify exposed lysine residues, chimeric VLPs based on the bacteriophage Qβ VLPs have been decorated with antigens for

Alzheimer's disease,[12] type 2 diabetes,[13] allergies,[14] and nicotine.[15] These modified VLPs present roughly 300-600 copies of their small molecule or peptide epitopes on their surface and exhibit robust humoral immune responses to their respective targets.

1.1.2.2 VLPs as nanoreactors

The porous, well-defined structure of VLPs has led to their exploration as nanoscale reactors for spatially confined catalysis. Handa and coworkers first demonstrated the encapsulation of a heterologous enzyme in 45 nm SV40 VLPs.[16] They fused EGFP to VP2, the minor capsid protein of SV40 VLPs and co-assembled the fusions with VP1, the major capsid protein. They determined via anti-GFP antibody immunoblotting that the GFP fusion was oriented to the interior of the capsid. They employed this strategy to encapsulate yeast cytosine deaminase (yCD), which converts cytosine to uracil, in SV40 capsids. yCD can also act as a potential prodrug-converting enzyme in gene therapy by converting 5-fluorocytosine to 5-fluorouracil, which leads to cell death by thymidylate synthase inhibition. SV40-yCD VLPs not only retained enzymatic activity, but also enabled delivery of functional yCD to CV-1 by protecting the encapsulated enzyme from protease degradation.

Douglas and coworkers recently have extended this concept to form a two-step biosynthetic pathway within a single VLP.[17] Using large (58 nm diameter) icosahedral P22 VLPs as a scaffold, they encapsulated the components to produce the cellular antioxidant tripeptide glutathione from its amino acid precursors. Glutamate cysteine ligase and glutathione synthetase were fused to the scaffolding proteins of P22 and co-assembled with the P22 coat protein (CP). The encapsulated enzymes maintained catalytic activity and kinetics comparable to free enzymes, but had an increased thermostability. Although no rate enhancement was observed due to the confined intermolecular proximity of the enzymes, such a phenomenon has been reported for P22 encapsulated EcHyd-1 hydrogenase.[18]

1.1.2.3 VLPs as drug delivery vehicles

The morphology and repeat subunit patterning of VLPs also makes them ideal candidates as drug-carrier systems. Therapeutic payloads may diffuse through the open pores of VLPs and be covalently attached to residues on the interior surface space of the capsid. This can both concentrate and protect a therapeutic from degradation. The multivalent exterior surface of VLPs can also be used for attachment of targeting ligands to enhance specificity of delivery and reduce off-target effects. Although no VLP delivery systems have been clinically approved thus far, many have entered preclinical development stages.

Steinmetz and coworkers have utilized the plant-derived Cowpea chlorotic mottle virus (CCMV) VLP as a platform for delivery.[19] CCMV VLPs disassemble at physiological pH or high ionic strength, releasing their encapsidated RNA. Buffer exchange and addition of a negatively-charged substrate triggers capsid reassembly. Using this method, they demonstrated the encapsulation of CpG oligodeoxynucleotides (ODNs), which have been investigated as immunoadjuvants for cancer therapy. Roughly 50 CpG-ODNs could be packaged per CCMV particle, and CCMV-ODN VLPs showed improved macrophage activation and antitumor activity *in vitro*. Evaluation of efficacy *in vivo* using CT26 colon cancer mice models showed decreased tumor size and prolonged survival. With Fast Track designation being recently granted to vidutolimod, a VLP vehicle for the treatment of metastatic melanoma, the advent of VLP-based therapeutics appears to be on the horizon.[20]

1.1.3 Bacteriophage MS2 VLPs

Bacteriophage MS2 is one of the most well-known and well-studied viral particles over the past sixty years. The MS2 bacteriophage was first characterized as an RNA phage in 1961, natively infecting male specific *Escherichia coli* (F+, F' or HFr).[21] The MS2 genome was the first complete genome to be sequenced and remains one of the smallest described genomes to date at 3,569 nucleotides.[22] Its genome encodes four proteins in total: coat, maturation, replicase, and lysis. The infectious virion is only composed of two of these proteins: 178 copies of the CP and a single maturation protein that binds to bacterial pili during infection.[23]

Recombinant expression of the MS2 CP results in the spontaneous self-assembly of 180 CPs into a non-infectious VLP devoid of the maturation protein. MS2 VLP production is robust, and has been demonstrated in bacterial,[24] yeast,[25] and cell-free expression systems.[26] The assembled MS2 VLP can tolerate harsh conditions – capsids do not denature upon exposure to temperatures up to 68 ^{o}C and pH values ranging from 2-12.

The structures of both the complete MS2 virion and the MS2 VLP have been characterized via x-ray crystallography as well as cryogenic electron microscopy.[27–29] MS2 particles assemble into 27 nm closed shell icosahedra. As dictated by Casper and Klug's seminal work describing the geometric rules for viral structural organization, MS2 capsids must undergo minor distortions to their symmetry in order to accommodate more than 60 CP subunits.[30,31]

The resultant MS2 VLPs may be described as T=3 icosahedra with three CP monomer conformations (termed A, B, and C) that exhibit quasi-equivalence.

A consequence of this geometry is the patterning of pentavalent and hexavalent positions across the icosahedral lattice, which form the basis for 32 pores of 2 nm diameter patterned along the surface of MS2. Structurally, the monomeric CP of MS2 is notable because it does not adopt the jelly roll fold that is found in 16 distinct families of RNA and DNA viruses. The MS2 CP consists primarily of β-sheets, with an α-helix and the C-terminal region of its 129 amino acid length. Evidence suggests that MS2 exists as an obligate dimer that is locked into place via interdigitation of the α-helices of adjacent CP monomers.[32] In the assembly process, the MS2 CP dimer takes on two quasi-equivalent conformations, known as the A/B and C/C dimers. Transition from the symmetric C/C dimer to the asymmetric A/B dimer is mediated by binding to the 19 nucleotide RNA translational repressor (TR) of the MS2 genome.[33] TR-RNA binding results in a cis/trans isomerization of the proline residue at position 78, which in turn folds back the pore-forming FG loop region of the capsid. The high-affinity of MS2 to this TR-RNA stem-loop continues to be used as an purification tag for RNA-protein complexes and as a tool for studying RNA localization *in cellulo*.[34] Binding to non-specific genetic material or other negatively charged molecules appears to be sufficient to mediate this switch, as MS2 VLPs have been shown to assemble around DNA cargo, anionic polymers, and proteins with negatively charged tags.[35]

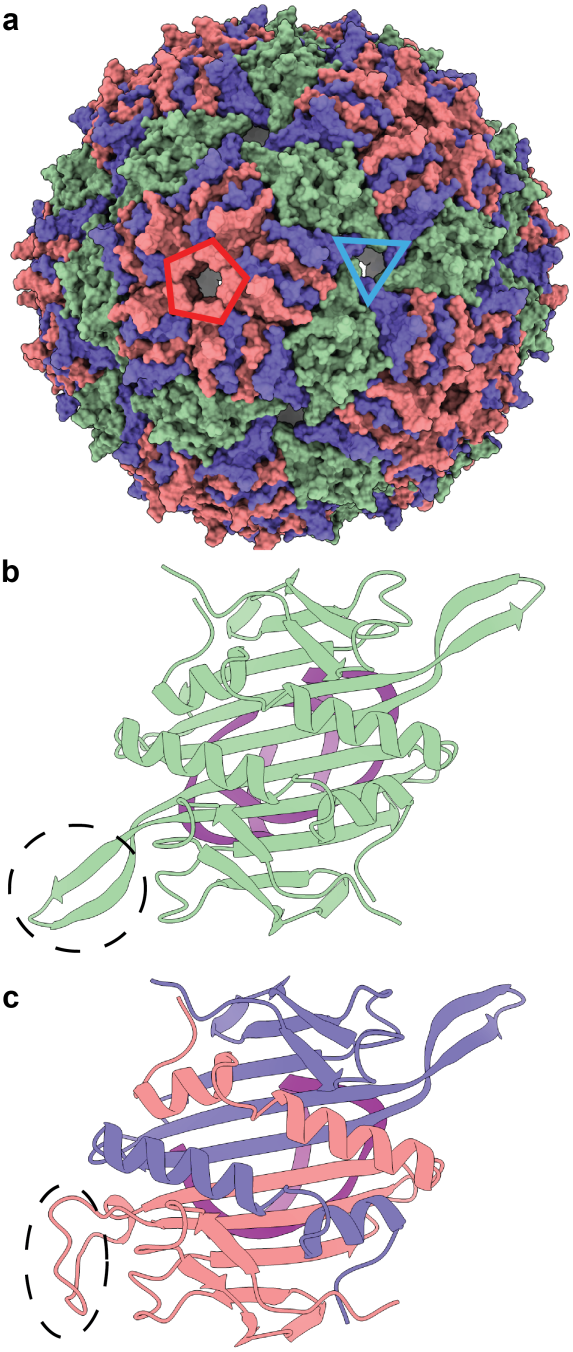

Figure 1.2: Bacteriophage MS2 VLP structure. (a) The MS2 VLP is depicted with its A, B, and C conformers shown in blue, red, and green, respectively. 5- and quasi-6-fold symmetry axes are indicated by a red pentagon and blue triangle, respectively. (b) C/C and (c) A/B dimers of the MS2 (CP) are shown bound to the TR-RNA depicted in purple. Conformational shift of the FG loop is circled.

1.2 Protein engineering

The broad range of protein functions, particularly the ability of enzymes to catalyze chemical reactions with exquisite selectivity, has motivated scientists to pursue the use of proteins as laboratory and industrial tools. Proteins however are highly evolved for their natural substrates, thus the scope of their applications are limited. As a result, the field of protein engineering has emerged as to attempt to remodel native proteins to fit desired uses. Discussed here are the historical successes of protein engineering and advances in methodology to modern techniques.

1.2.1 Site-directed mutagenesis

Advances in molecular biology in the 1980s, particularly methods for oligonucleotide synthesis, polymerase chain reaction (PCR), and site-specific molecular cloning enabled the deliberate substitution of amino acid residues in recombinantly expressed proteins.[36] Thus, it became possible to interrogate the role of individual amino acids in the structure and function of proteins, leading to hypothesis-driven attempts to alter native enzyme activity.[37] The first major breakthrough in this arena was published by the Wells lab in 1985.[38] They used site-directed mutagenesis to substitute Met222, a residue implicated as the primary site for oxidative inactivation of the enzyme, with all 19 other natural amino acids. They found that although enzyme activity was diminished, mutants bearing nonoxidizable amino acids were resistant to inactivation by 1 M H_2O_2. This landmark work established the foundation for modulating the stability and activity of proteins rationally guided by crystallographic information.

1.2.2 Directed evolution

While site-directed mutagenesis strategies can effectively produce improved proteins in cases where key residues are known from structural information, oftentimes the pathway to a desired protein function is incredibly complex and can involve many mutations. Researchers have used Darwinian evolution as inspiration as for "directed evolution" by which evolution for a chosen feature is accelerated in a laboratory setting. Directed evolution evolution hinges upon two steps that can be repeatedly cycled: 1) diversification of the coding sequence for a parent protein and 2) selection of protein variants with improved fitness. Iteration of these processes acts as an accelerated search algorithm for protein evolution.[39]

1.2.2.1 Mutagenic library generation

Widespread diversification of a target gene was first accomplished by Caldwell and Joyce in 1992 using a modified PCR protocol, now known as error-prone PCR (ep-PCR).[40] This method employs a low fidelity DNA polymerase to introduce variations in a protein coding sequence during DNA amplification. This method produces millions of unique variants of a given protein an average of 2-3 base pair substitutions per mutagenized gene. While this method generates a vast diversity of sequences, the probability that beneficial amino acid substitutions are discovered can be low, as improvements to protein function can be obscured by simultaneous neutral or deleterious amino acid changes. The introduction of DNA shuffling by Stemmer and coworkers overcomes some of the drawbacks of epPCR.[41,42] This method involves fragmentation of target genes after an initial round of epPCR and selection into oligonucleotides. Fragments are then reassembled via homologous recombination allowing lateral sharing of beneficial mutations and elimination of deleterious mutations. DNA shuffling and selection of TEM-1 β-lactamase resulted in a 32,000-fold increase in antibiotic resistance of TEM-1 containing *E. coli* while epPCR alone resulted in a 16-fold increase.

1.2.2.2 Functional protein selection

Effective acquisition of a new or optimized protein trait depends upon careful analysis of protein variant phenotypes. Two fundamental rules apply to effective analysis of a protein library: 1) there must be a link between a library member's genotype and its phenotype and 2) "you get what you select for".[43] The most common and straightforward method of establishing a genotype-to-phenotype link is via transformation of a mutated DNA library into an expression host. Each host cell maintains a single library variant, and screening of individual colonies or cellular sorting methods enables evaluation of a given property. Using this method requires a functional assay that can be performed in the host cell without interference from endogenous proteins.

Cellular surface-display systems levy host membrane proteins as fusion partners to transport protein cargo to the cell surfaces, broadening the scope of functional assays to include ligand affinity. Another significant method for enzyme evaluation pioneered by Griffiths and Tawfik is *in vitro* compartmentalization. This method involves dispersion of a substrate-linked gene library in water-in-oil emulsion droplets with interior volumes close to that of *E. coli*. Transcription and translation factors are added to the suspension to produce proteins of interest, which convert their linked substrate if they are active. Enrichment for the desired product linked to the coding

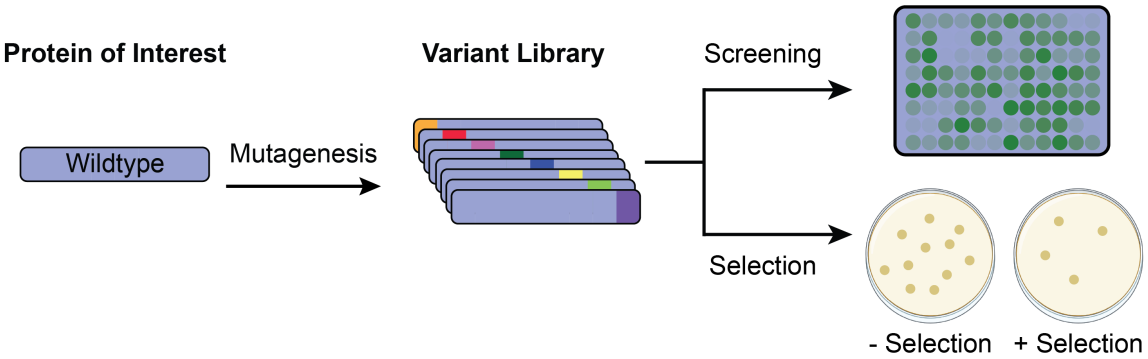

Figure 1.3: Protein engineering methodology. Protein gene of interest is diversified, then fitness of library variants is assessed. Screening methods evaluate variants individually through a designed assay (i.e. colorimetric product of enzyme activity). Selection methods isolate fit variants from unfit variants in bulk (i.e. protein function is tied to bacterial antibiotic resistance.)

gene gives a readout of functional variants.

Careful experimental evaluation of protein function is key to selecting variants that perform in the desired manner. Functional evaluation of protein variants fall into two categories: 1) selections, in which more fit variants out-compete less fit variants in bulk, and 2) screens, which assess the performance of a each protein individually. Common engineering targets include improved thermal stability, higher tolerance to organic solvents, modified pH sensitivity, and altered substrate specificity or enantioselectivity.

1.2.2.3 Protein selection strategies

Selection strategies either hinge upon linkage of a protein's activity with survival of a host organism or on physical separation of library members by their fitness. While careful planning is required to design assays that effectively link a protein function to survival, extremely large (10^{10}) variant libraries can be evaluated simultaneously via selections. The most common method for linking fitness to survival is the use of antibiotic resistance as a selection marker. Many studies have evolved enzymes that neutralize antibiotics in this way; however elegant design can produce assays for evolution of proteins typically unrelated to antibiotic resistance.[44] Work by Shultz and co-workers developed a selection for aminoacyl tRNA synthetases that aminoacylate suppressor tRNAs with non-canonical amino acids.[45] By inserting a stop codon with the chloramphenical efflux pump gene, they linked successful aminoacylation with chloramphenicol resistance.

1.2.2.4 Protein screening strategies

Screening offers more precise control over assay conditions and readout at the cost of decreased throughput. Single-cell screening methods such as fluorescence-activated cell sorting (FACS) can expand the number of screened variants up to 10^6 if a fluorescent indicator can be used. Non-fluorescent substrates meanwhile may be readily screened with moderate throughput as individual colonies if they present a readily visible phenotype. Chromatographic or mass spectrometry readouts may be used as screening methods, albeit with considerably reduced throughput.

1.2.3 Comprehensive Protein fitness landscapes

Advances in DNA synthesis technologies and multiplexed DNA sequencing have enabled high-throughput evaluation of comprehensive sets of mutated proteins.[46] While targeted protein mutations or random mutagenesis explore small or sporadic pockets of the possible mutational space of a given protein, deep mutational scanning generates exhaustive understanding of the mutability of the target protein. Deep mutational scanning introduces all possible amino acid substitutions of a protein at all possible positions using cassette-based cloning methods.[47] The comprehensive knowledge generated by this methodology illuminates what is known as the fitness landscape of a protein.

The concept of a protein fitness landscape is founded in John Maynard Smith's conception of protein evolution as a walk from one functional protein to another in the space of all possible protein sequences.[48] By understanding the fitness of each protein sequence, one can visualize the space as a landscape where mountainous peaks represent fit-

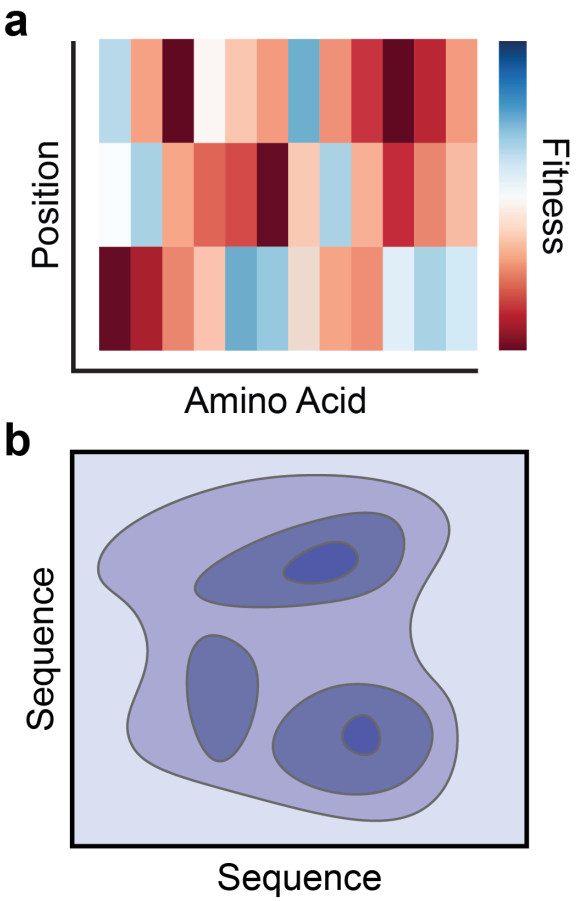

Figure 1.4: Visualization of protein fitness landscapes. The mutable landscape of a protein may be visualized as (a) a heatmap in which every cell represents the quantitative fitness score of a protein variant or (b) a 2D topological map where all possible sequences are spread across a surface and the fitness peaks are mapped to color saturation.

ness maxima and non-functional sequences lie along the valley floor. Understanding the fitness landscape of a protein not only identify protein variants with improved properties, but also can shed light on the evolutionary trajectory of proteins. Fitness landscapes generated via deep mutational scanning simultaneously access all amino acid substitutions of a given position—a feat that cannot be achieved by natural evolution, as the genetic code constrains a single nucleotide mutagenic step to 6 of the 20 amino acids on average.

In recent years, the experimental fitness landscapes of various proteins have been generated through deep mutational scanning. The first demonstration of a complete protein fitness landscape was generated by Bolon and coworkers in 2011.[49] They generated all possible point mutations of the entire yeast ubiquitin gene. Using engineered yeast strains that contained a copy of ubiquitin under tight regulation of a galactose promoter. This allowed yeast to grow normally in galactose media, but depend upon the function of the ubiquitin library variant when exchanged to dextrose media. Thus ubiquitin function was effectively linked to yeast growth. The full fitness landscape yielded information about the faces of regions of ubiquitin highly tolerant and highly sensitive to amino acid substitution, and confirmed the conservation of residues critical for proteasome recognition in a single experiment.

Fitness landscaping of viruses has also attracted significant attention over the past decade as understanding the mutability of a virus can generate information about potential evolutionary trajectories to immune evasion. Bloom and coworkers generated the fitness landscape of hemagglutinin (HA), an influenza surface protein that aids in viral binding and infectivity.[50] They diversified the the HA gene, produced mutant viruses by reverse genetics, and passaged viruses in tissue culture to select for viability. Their results shed light on the high frequency of antigenic mutations that necessitate frequent reformulation of flu vaccines. While host-binding site mutations were poorly tolerated in HA, regions of antigenic recognition were extremely mutationally flexible. This flexibility is unique to HA; other influenza proteins such as the influenza nucleoprotein do not permit mutation to their epitope sites without a loss of fitness.

While the wealth of information to be gleaned from protein fitness landscapes is apparent for a variety of protein archetypes, performing a deep mutational scan (all amino acid substitutions) only captures a single dimension of a protein's mutational space. The expansive nature of the protein landscape becomes apparent when simultaneous mutations are sampled. A 100 amino acid protein has 2000 possible variants. The double mutant space of the same protein meanwhile contains 2 million possibilities. Although fully elucidating the effects of simultaneous mutations rapidly becomes infeasible (a 300 amino acid protein has 20^{300} possible combinations, more than the number of atoms in the universe),[51] complex libraries have demonstrated the surprising effects multiple mutations can have on protein fitness.

1.2.4 Epistasis in fitness landscapes

In natural evolutionary patterns, accessible mutations are heavily constrained by the mutational path between variants. As beneficial mutations are propagated with high frequency, neutral and deleterious mutations drop out of the variant pool rapidly. Thus, the effect of combined mutations along the path of an evolutionary walk, also known as epistasis, has significant impact on the possibility of protein evolution.

The impact of two mutations to a given protein on resultant fitness can manifest in a number of ways. In an additive case in which no epistasis occurs, the fitness of a protein double-mutant is merely the simple sum of both single mutants' fitness. In cases known as magnitude epistasis, a double mutant produces a larger fitness effect than the sum of its corresponding single mutations (e.g. single mutants with positive effects on fitness results in a greater-than-additive increase in fitness when combined). The most drastic effects on a fitness landscape occur due to sign and reciprocal sign epistasis. Sign epistasis occurs when one mutation has the opposite effect in the presence of a second mutation (e.g. a deleterious mutation on its own enhances the positive effect of a beneficial second mutation). Most confounding is reciprocal sign epistasis, where two mutations with the same type of fitness effect alone result in the opposite fitness effect in tandem (e.g. two deleterious mutations together result in a fitness benefit).

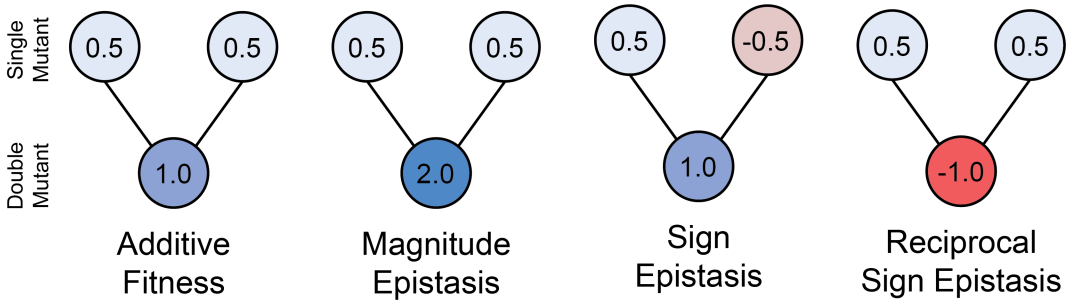

Figure 1.5: Epistatic outcomes of protein mutation. Mutations with positive and negative effects on protein fitness are depicted in green and red, respectively.

A number of studies have used deep mutational scanning methods to elucidate the epistatic effects in protein fitness landscapes. These works tend to focus on exploring the multi-residue mutational space of functional domains within a protein of interest. Groundbreaking work by Weinreich and coworkers demonstrated how sign epistasis constrains the evolutionary trajectory of TEM β-lactamase in its accumulation of antibiotic resistance activity. Only five mutations are required to increase the antibiotic resistance of bacteria to cefotaxime by a factor of 100,000.

In order to acquire these five mutations, TEM must traverse one of 120 possible mutational trajectories. By generating individual combinations of mutants along the possible mutational trajectories, they found that only 18 of the trajectories are accessible as the majority of possible pathways result in unfit mutants. However, sign epistasis along allowed trajectories enables fixing of mutations with deleterious effects. For example, the G238S mutation enhances hydrolysis but increases aggregation, while the M182T mutation increases stability at the cost of hydrolysis. M182T alone is selected against, but if the G238S mutation is first fixed, then the M182T/G238S mutant shows a net fitness improvement.

Subsequent fitness landscaping work has described epistasis in a variety of systems including GFP, RNA binding proteins, and various viruses. Full accounting of all double mutants of protein G domain 1 found that roughly 30% of mutant pairs experience weak epistasis, while roughly 5% display strong deviations from additivity. Further research is needed to establish the prevalence of epistasis in protein systems as well as determine how epistatic effects can be used to access mutational outcomes of proteins that may be disallowed in nature.

1.3 Conclusion and Outlook

Deep mutational scanning and fitness landscaping are powerful tools for studying the fundamental mutability of proteins, as well as for engineering new features based on selective pressures or screening conditions. This work utilized a recently-developed tool for generating fitness landscapes of virus-like particles, called Systematic Mutation and Assembled Particle Selection (SyMAPS) to engineer new properties in bacteriophage MS2 VLPs, as well as understand fundamental constraints on self-assembly of proteinaceous systems. The method uses particle assembly as a measure for fitness, allowing than previous landscaping studies of viruses, which rely on infectivity, as a measure of fitness. This allows for functional assessment of assembly in isolation when compared to the features necessary for viral viability (i.e. assembly, attachment, and replication).

The comprehensive landscape of N-terminal MS2 extensions was first generated with the intent to install site-specific reactive handles on the surface of MS2 VLPs. It was found that capsid CP extension is largely deleterious to assembly, but identify sequences that enable VLP formation. We further design selection strategies that use site-selective bioconjugation as a fitness parameter, ultimately producing engineered capsids with high reactivity.

Next, the epistatic landscape of MS2 VLPs with a point mutation that fixes an alternative assembly geometry to the wildtype T=3 icosahedra was constructed. This

point mutation, S37P, results in a homogenous geometry shift to T=1 icosahedral particles without drastically altering the secondary structure of the MS2 CP or the dimeric subunit of MS2. Thus, the S37P landscape uniquely enables separation of folding constraints on assembly from those that arise due to intersubunit interactions at MS2's symmetry axes. Surprisingly, this non-native assembly phenotype permits access to more mutations than wildtype MS2 despite being an inaccessible mutation in viral MS2. Further, these comprehensive landscape datasets are used to benchmark the performance of *in silico* mutational tools that predict the energetic effects of protein mutation. Finding that predicted shifts in folding free energy poorly correlate to VLP assembly competency, results are used to train machine learning models that predict assembly state of mutated VLPs.

Lastly, this work reports on grassroots efforts to improve the equity and inclusion of the UC Berkeley Department of Chemistry. Longitudinal outcomes from three years of departmental climate surveys and cooperative meetings between graduate students and faculty are discussed. Happily, respondents report improvements in sense of value and inclusion within the department of the course of survey administration. The survey framework is also used to generate primary data on how metrics of success are applied to chemistry graduate students. We report that the success metrics prioritized by advisors align less with the metrics valued by students from historically marginalized groups than those from well-represented backgrounds. Methods to ameliorate such issues are discussed, such as explicit goal-setting with graduate mentees. Prescriptions as to how academic practices can be shifted to improve equitable access to chemistry for people from all backgrounds are also offered.

1.4 References

(1) Bayer, M. E.; Blumberg, B. S.; Werner, B. *Nature* **1968**, *218*, 1057–1059.

(2) Zeltins, A. *Molecular Biotechnology* **2013**, *53*, 92–107.

(3) Salunke, D. M.; Caspar, D. L. D.; Garcea, R. L. *Cell* **1986**, *46*, Publisher: Elsevier, 895–904.

(4) Bräutigam, S.; Snezhkov, E.; Bishop, D. H. *Virology* **1993**, *192*, 512–524.

(5) Boigard, H.; Alimova, A.; Martin, G. R.; Katz, A.; Gottlieb, P.; Galarza, J. M. *PLOS Neglected Tropical Diseases* **2017**, *11*, Publisher: Public Library of Science, e0005608.

(6) Miller, R. A.; Presley, A. D.; Francis, M. B. *Journal of the American Chemical Society* **2007**, *129*, Publisher: American Chemical Society, 3104–3109.

(7) Nooraei, S.; Bahrulolum, H.; Hoseini, Z. S.; Katalani, C.; Hajizade, A.; Easton, A. J.; Ahmadian, G. *Journal of Nanobiotechnology* **2021**, *19*, 59.

(8) Plotkin, S. A. *Nature Medicine* **2005**, *11*, S5–S11.

(9) Zepeda-Cervantes, J.; Ramírez-Jarquín, J. O.; Vaca, L. *Frontiers in Immunology* **2020**, *11*, 1100.

(10) McAleer, W. J.; Buynak, E. B.; Maigetter, R. Z.; Wampler, D. E.; Miller, W. J.; Hilleman, M. R. *Nature* **1984**, *307*, 178–180.

(11) Zhao, Q.; Zhang, J.; Wu, T.; Li, S.-W.; Ng, M.-H.; Xia, N.-S.; Shih, J. W.-K. *Journal of Gastroenterology* **2013**, *48*, 159–168.

(12) Wiessner, C.; Wiederhold, K.-H.; Tissot, A. C.; Frey, P.; Danner, S.; Jacobson, L. H.; Jennings, G. T.; Lüönd, R.; Ortmann, R.; Reichwald, J.; Zurini, M.; Mir, A.; Bachmann, M. F.; Staufenbiel, M. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **2011**, *31*, 9323–9331.

(13) Cavelti-Weder, C.; Timper, K.; Seelig, E.; Keller, C.; Osranek, M.; Lässing, U.; Spohn, G.; Maurer, P.; Müller, P.; Jennings, G. T.; Willers, J.; Saudan, P.; Donath, M. Y.; Bachmann, M. F. *Molecular Therapy: The Journal of the American Society of Gene Therapy* **2016**, *24*, 1003–1012.

(14) Klimek, L.; Willers, J.; Hammann-Haenni, A.; Pfaar, O.; Stocker, H.; Mueller, P.; Renner, W. A.; Bachmann, M. F. *Clinical & Experimental Allergy* **2011**, *41*, 1305–1312.

(15) Maurer, P.; Jennings, G. T.; Willers, J.; Rohner, F.; Lindman, Y.; Roubicek, K.; Renner, W. A.; Müller, P.; Bachmann, M. F. *European Journal of Immunology* **2005**, *35*, 2031–2040.

(16) Inoue, T.; Kawano, M.-a.; Takahashi, R.-u.; Tsukamoto, H.; Enomoto, T.; Imai, T.; Kataoka, K.; Handa, H. *Journal of Biotechnology* **2008**, *134*, 181–192.

(17) Wang, Y.; Uchida, M.; Waghwani, H. K.; Douglas, T. *ACS Synthetic Biology* **2020**, *9*, Publisher: American Chemical Society, 3298–3310.

(18) Jordan, P. C.; Patterson, D. P.; Saboda, K. N.; Edwards, E. J.; Miettinen, H. M.; Basu, G.; Thielges, M. C.; Douglas, T. *Nature Chemistry* **2016**, *8*, 179–185.

(19) Cai, H.; Shukla, S.; Steinmetz, N. F. *Advanced Functional Materials* **2020**, *30*, 1908743.

(20) Ribas, A. et al. *Cancer Discovery* **2021**, Publisher: American Association for Cancer Research Section: Research Brief.

(21) Davis, J.; Sinsheimer, R.; Strauss, J. In *Science*; Issue: 348, Amer Assoc Advancement Science: 1961; Vol. 134, p 1427.

(22) Fiers, W.; Contreras, R.; Duerinck, F.; Haegeman, G.; Iserentant, D.; Merregaert, J.; Min Jou, W.; Molemans, F.; Raeymaekers, A.; Van den Berghe, A.; Volckaert, G.; Ysebaert, M. *Nature* **1976**, *260*, 500–507.

(23) Stockley, P. G.; Stonehouse, N. J.; Valegård, K. *The International Journal of Biochemistry* **1994**, *26*, 1249–1260.

(24) Stonehouse, N.; Stockley, P. *FEBS Letters* **1993**, *334*, 355–359.

(25) Legendre, D.; Fastrez, J. *Journal of Biotechnology* **2005**, *117*, 183–194.

(26) Bundy, B. C.; Swartz, J. R. *Journal of Biotechnology* **2011**, *154*, 230–239.

(27) Golmohammadi, R.; Valegård, K.; Fridborg, K.; Liljas, L. *Journal of Molecular Biology* **1993**, *234*, 620–639.

(28) Koning, R. I.; Gomez-Blanco, J.; Akopjana, I.; Vargas, J.; Kazaks, A.; Tars, K.; Carazo, J. M.; Koster, A. J. *Nature Communications* **2016**, *7*, 12524.

(29) Dai, X.; Li, Z.; Lai, M.; Shu, S.; Du, Y.; Zhou, Z. H.; Sun, R. *Nature* **2017**, *541*, 112–116.

(30) Caspar, D. L. D.; Klug, A. *Cold Spring Harbor Symposia on Quantitative Biology* **1962**, *27*, 1–24.

(31) Prasad, B. V. V.; Schmid, M. F. *Viral Molecular Machines* **2011**, *726*, 17–47.

(32) Lalwani Prakash, D.; Gosavi, S. *The Journal of Physical Chemistry B* **2021**, *125*, Publisher: American Chemical Society, 8722–8732.

(33) Dykeman, E. C.; Stockley, P. G.; Twarock, R. *Journal of Molecular Biology* **2010**, *395*, 916–923.

(34) George, L.; Indig, F. E.; Abdelmohsen, K.; Gorospe, M. *Open Biology*, *8*, Publisher: Royal Society, 180104.

(35) Glasgow, J. E.; Capehart, S. L.; Francis, M. B.; Tullman-Ercek, D. *ACS Nano* **2012**, *6*, 8658–8664.

(36) Wilkinson, A. J.; Fersht, A. R.; Blow, D. M.; Winter, G. *Biochemistry* **1983**, *22*, 3581–3586.

(37) Winter, G.; Fersht, A. R.; Wilkinson, A. J.; Zoller, M.; Smith, M. *Nature* **1982**, *299*, 756–758.

(38) Estell, D. A.; Graycar, T. P.; Wells, J. A. *Journal of Biological Chemistry* **1985**, *260*, 6518–6521.

(39) Arnold, F. H.; Volkov, A. A., 6.

(40) Cadwell, R. C.; Joyce, G. F. *Genome Research* **1992**, *2*, 28–33.

(41) Stemmer, W. P. *Proceedings of the National Academy of Sciences* **1994**, *91*, 10747–10751.

(42) Stemmer, W. P. C. *Nature* **1994**, *370*, 389–391.

(43) Olsen, M. J.; Stephens, D.; Griffiths, D.; Daugherty, P.; Georgiou, G.; Iverson, B. L. *Nature Biotechnology* **2000**, *18*, 1071–1074.

(44) Orencia, M. C.; Yoon, J. S.; Ness, J. E.; Stemmer, W. P. C.; Stevens, R. C. *Nature Structural Biology* **2001**, *8*, 238–242.

(45) Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. *Proceedings of the National Academy of Sciences* **1997**, *94*, 10092–10097.

(46) Packer, M. S.; Liu, D. R. *Nature Reviews Genetics* **2015**, *16*, 379–394.

(47) Fowler, D. M.; Fields, S. *Nature methods* **2014**, *11*, 801–807.

(48) Maynard Smith, J. *Nature* **1970**, *225*, 563–564.

(49) Roscoe, B. P.; Thayer, K. M.; Zeldovich, K. B.; Fushman, D.; Bolon, D. N. A. *Journal of Molecular Biology* **2013**, *425*, 1363–1377.

(50) Thyagarajan, B.; Bloom, J. D. *eLife* **2014**, *3*, ed. by Pascual, M.

(51) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell. 4th edition* **2002**.

Chapter 2

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

The following is adapted from Brauer, Hartman, Bader, Merz, Tullman-Ercek, and Francis; J. Am. Chem. Soc., 2019 with permission.

2.1 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.

2.2 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 2.1). 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 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.

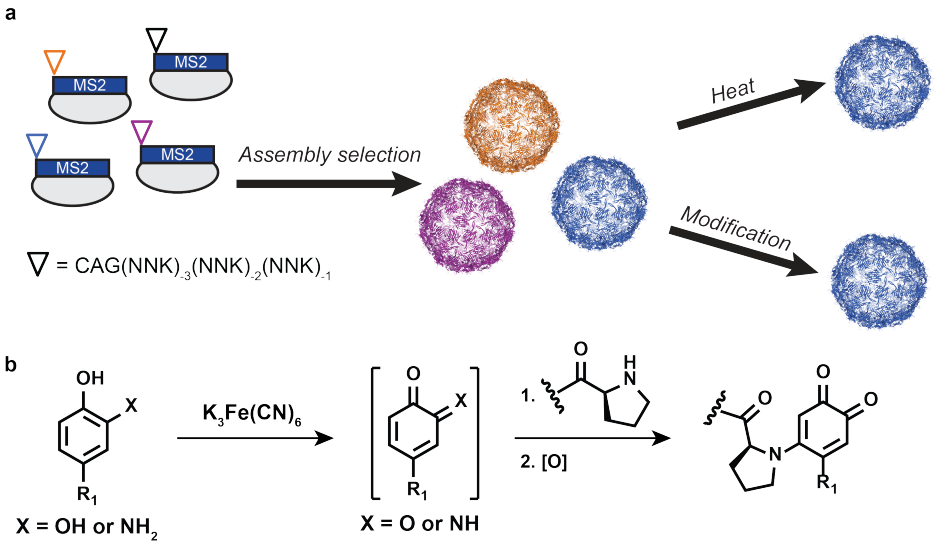

Figure 2.1: Scheme to isolate N-terminally extended VLPs with desired properties. (a) Three-codon NNK extensions at the N terminus generated a library of 8,000 variants. Assembly, thermostability, and chemical modification challenges were used to identify HiPerX variants, or high-performing N-terminal extensions with desirable properties, indicated in blue. (b) Oxidative coupling reactions can be used to modify N-terminal proline residues.

2.3 Results and Discussion

2.3.1 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 2.2a).[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 2.1a, 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 2.2b). 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 2.1b).[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 calculated the 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 extensions competent for VLP assembly (Figure 2.3, Additional Figure 2.9). 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 (Additional Figure 2.10).

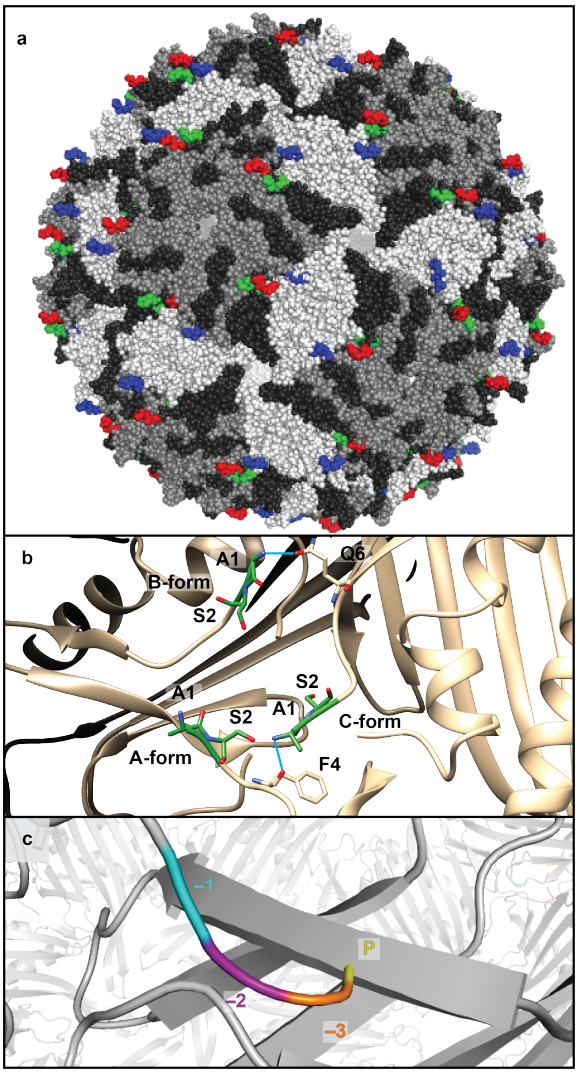

Figure 2.2: N termini of the MS2 capsid coat protein (MS2 CP) monomers. (a) Each quasi-equivalent form and N terminus is indicated with a shade of gray or color, respectively. The N terminus of the A form (dark gray) is shown in red; the N terminus of the B form (white) is shown in blue; and the N terminus 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.

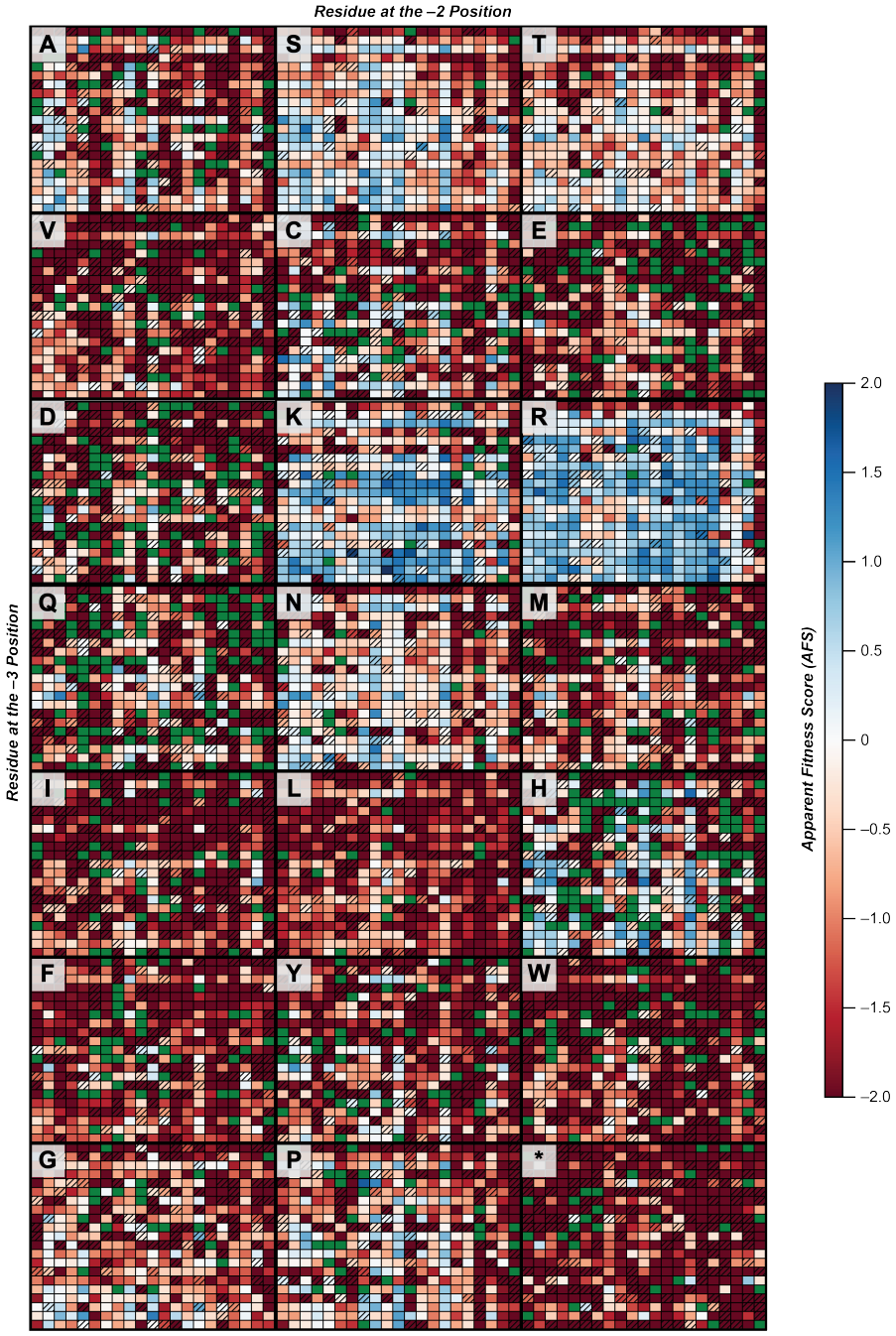

Figure 2.3: Apparent fitness landscape of P-X-X-X-MS2 N-terminal extensions. Extensions are labeled as the distance from the native N terminus (alanine), and the –1 position is indicated in the upper left corner of each quadrant. Blue indicates enriched variants and red indicates enriched combinations. 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.

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 2.3). We evaluated the number of variants with P-X-X-Z-MS2 that were compatible with assembly (Figure 2.4a), 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 2.3, Figure 2.4a). 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

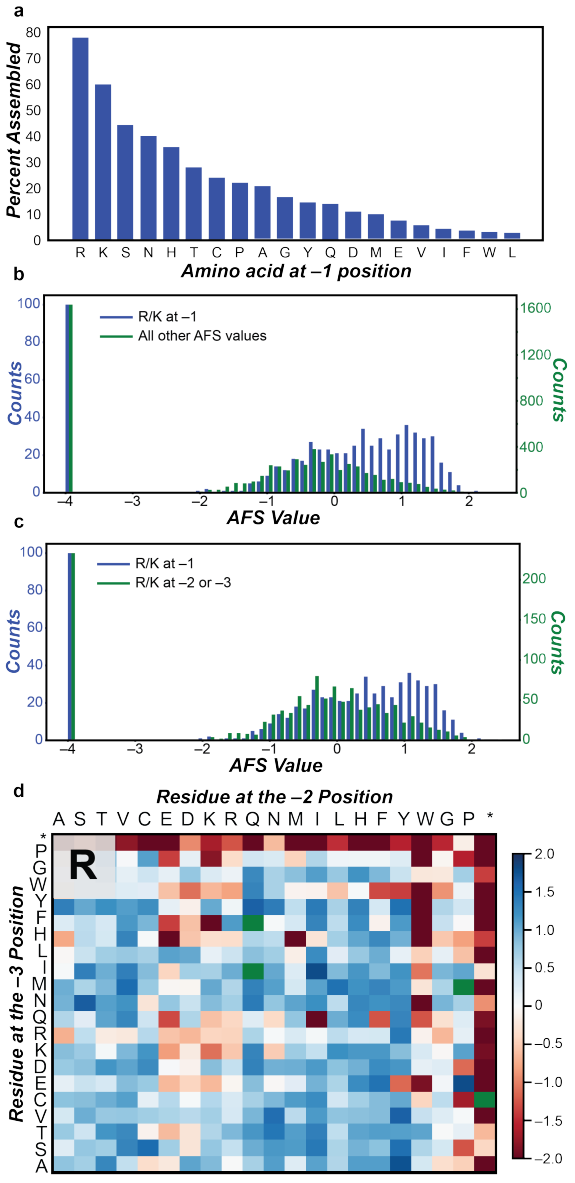

Figure 2.4: Effect of positive charge at the –1 position in N-terminal extensions. (a) Assembly competency differs by –1 amino acid. Arginine and lysine at the –1 position permit 77% and 61% of extensions. Arginine and lysine at position –1 result in a higher percent of positive AFS values (b) compared 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 when arginine is at the –1 position.

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, respectively (Figure 2.4a). 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 2.4b). 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 2.4c). A larger version of the data for arginine in position –1 appears in Figure 2.4d.

Finally, we confirmed that these trends were similar to N-terminal extensions in the absence of proline in the –4 position (Additional Figure 2.10). In this library, X-X-[K/R]-MS2 also resulted in a disproportionately 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 (Additional Figure 2.3a). 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 (Additional Figure 2.11b). 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 (Additional Figure 2.11c). The stringent positional specificity of these interactions highlights the remarkable level of detail offered by a comprehensive mutational strategy such as SyMAPS.

2.3.2 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 (Additional Figure 2.12a). In addition, we plotted the three biological replicates of the X-X-X-MS2 data set (Additional Figure 2.12b–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, Additional Figure 2.12e), 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 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 (Additional Figure 2.12f,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 (Additional Figure 2.12h). 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 (Additional Figure 2.12i), and further requiring at least 10 reads in both replicates increased the R2 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 (Additional Figures 2.13–2.15). 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 (Additional Figures 2.16–2.18). 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.

2.3.3 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 2.1b).

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 2.5a). 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 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.

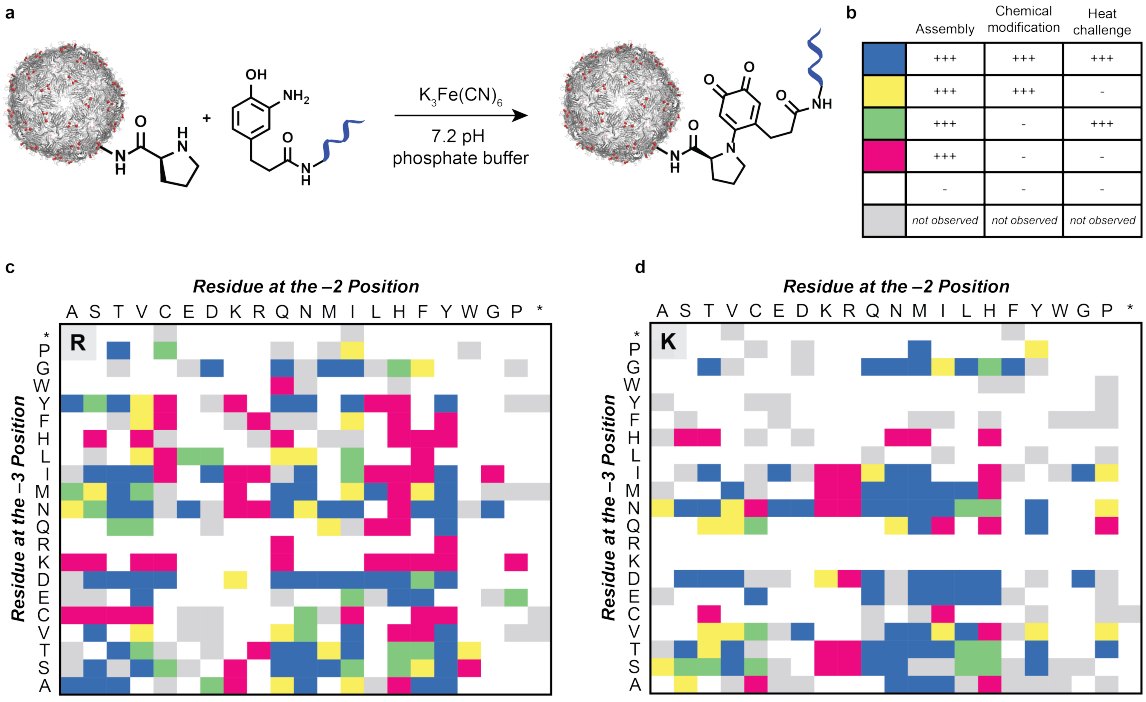

Figure 2.5: Combined fitness landscape of the P-X-X-X-MS2 N-terminal extensions. (a) The chemical modification-based selection of the variant library employs bioconjugation to a 25 bp DNA strand. (b) A color key is provided for the combined AFL data. (+++) indicates a score greater than 1.0 in the selection, 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 Additional Figure 7.

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 (Additional Figure 2.19) or chemical modification (Additional Figure 2.20) 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 2.5b, Additional Figure 2.21) 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 2.5c,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 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.

2.3.4 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 ^{o}C for 10 min, supporting the quality of the AFLs (Additional Figure 2.23).

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 2.6a). 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 2.6b). 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 vari-

ants 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.

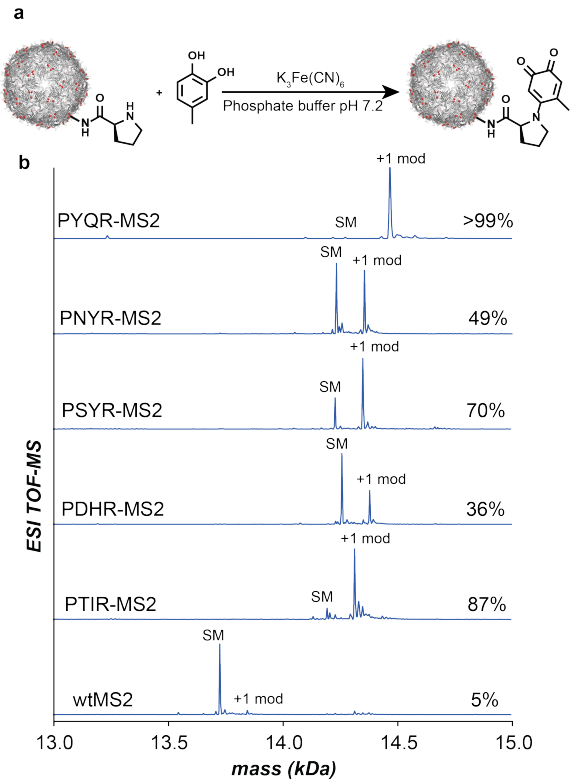

Figure 2.6: Chemical modification of HiPerX MS2 variants. (a) An oxidative coupling reaction was evaluated for proline-terminated MS2 variants. (b) Mass spectra of chemically modified HiPerX variants of the MS2 CP are shown. Percent modification is determined by integration of the unmodified (SM) vs modified (+1 mod) peaks.

HPLC SEC of modified samples confirmed that all of these VLPs remained assembled after modification (Additional Figure 2.25). 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 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 2.7a–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 2.7b).[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 2.7d). 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 2.7e).

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 2.7c).[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 (36-87%) to near-quantitative (>99%) in all cases (Figure 2.7e). 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 (Additional Figure 2.24). 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.

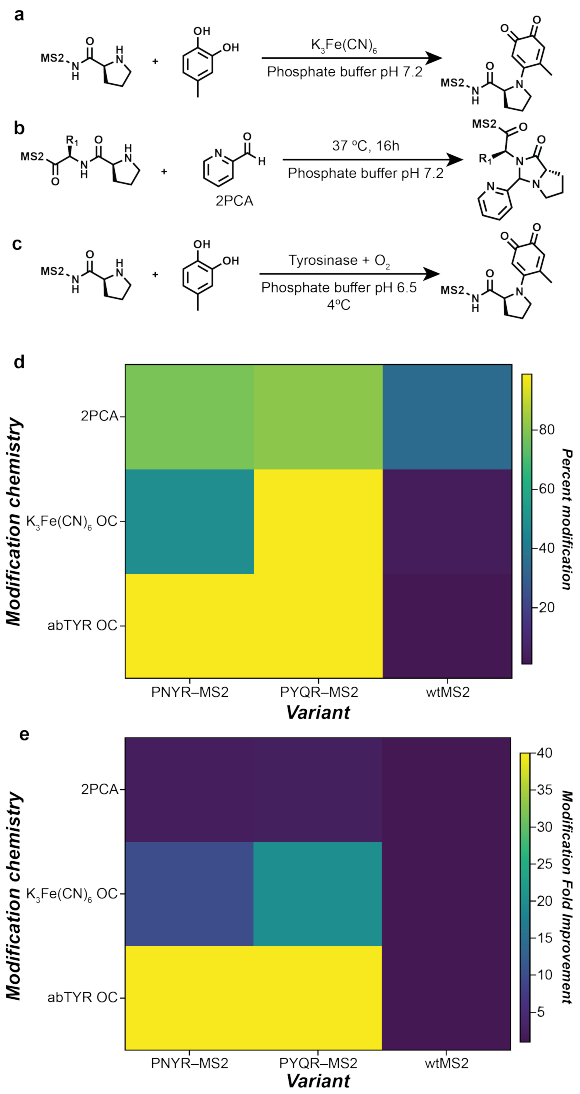

Figure 2.7: Conversion and fold improvement of N-terminal modification strategies of HiPerX MS2 variants. Reaction schemes are shown for (a) potassium ferricyanide-mediated oxidative coupling, (b) 2-pyridinecarboxaldehyde modification, and (c) tyrosinase-mediated oxidative coupling. (d) Percent modification and (e) fold improvement of two HiPerX MS2 variants is shown in contrast to wild-type MS2.

2.3.5 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 2.8a). This smaller-sized variant retains similar thermostability 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]

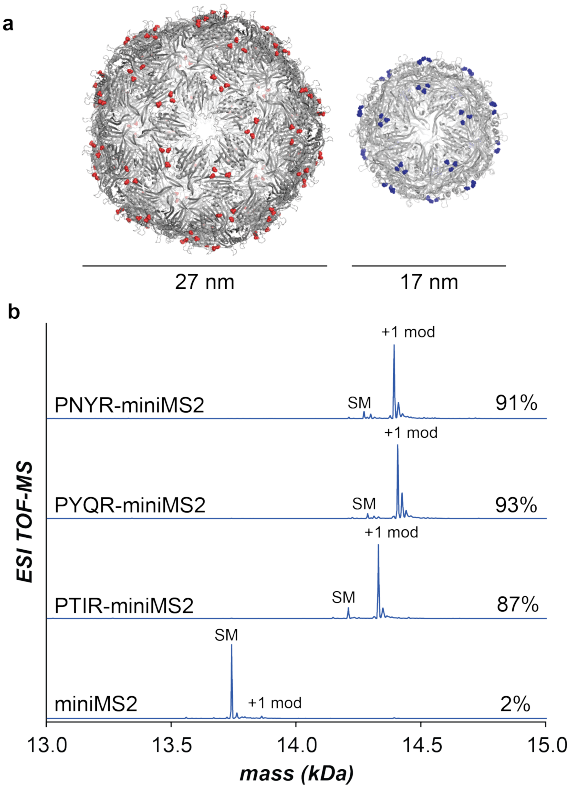

Figure 2.8: Chemical modification of HiPerX miniMS2 variants (CP[HiPerX-S37P]. (a) Crystal structures of CP[WT] and CP[S37P] are shown with N termini highlighted in red and blue, respectively. (b) Mass spectra of chemically modified HiPerX variants of the MS2 CP are shown.

We looked 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 (Additional Figure 2.26a). Additionally, variants tolerated 50 ºC for 10 min, indicating that thermostability was preserved in the new genetic background (Additional Figure 2.26b).

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 2.8b). As a comparison, CP[S37P] modified <5%, indicating that the extensions are

critical to achieve high modification rates. Despite changes to 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.

2.4 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.

2.5 Materials and Methods

2.5.1 Library generation

To generate libraries of N-terminal extensions, we modified a library generation and selection strategy used previously in our lab, called SyMAPS.[23] SyMAPS uses the EMPIRIC cloning developed in the Bolon lab.[43] In EMPIRIC cloning, a plasmid contains a self-encoded removable fragment (SERF) flanked by inverted BsaI recognition sites. Thus, BsaI digestion simultaneously removes the SERF and BsaI sites. This plasmid, referred to as an entry vector, was one of several previously used to generate all one amino acid mutations of the MS2 CP. Two single-stranded DNA primers with $(NNK)_3$ extensions, either with or without an N-terminal proline, were purchased, resuspended, and diluted to final concentrations of 50 ng/µL. The

reverse strand was completed by overlap extension PCR with a corresponding reverse primer. Subsequently, the double-stranded DNA was purified by PCR clean-up (Promega, Cat #A9282), diluted to 1-5 ng/µL, then cloned into the entry vector using Golden Gate cloning. Ligated plasmids were incubated on desalting membranes (Millipore Sigma, Cat #VSWP02500) for 20 min, followed by transformation into store-bought electrocompetent DH10B *E. coli* cells (Invitrogen, Cat #18290015) for the $P(NNK)_3$ extension, or homemade electrocompetent DH10B *E. coli* cells for the $(NNK)_3$ extension. Following electroporation and recovery, cells were plated onto two LB-A plates (VWR, Cat #82050-600) containing 32 µg/mL chloramphenicol. Cells were grown overnight at 37 °C. Colony number varied, but every transformation produced at least 3x the library size. The protocol was repeated in full for three (for $(NNK)_3$ extensions) or two (for $P(NNK)_3$ extensions) total biological replicates that are entirely independent from library generation through selection.

2.5.2 Assembly selection

Colonies were harvested by scraping plates into LB-M and growing the mixture for 2 h. Each library was subcultured 1:100 into 1 L of 2xYT (Teknova, Cat #Y0210) and grown to an OD of 0.6, then induced with 0.1% arabinose. Variant libraries were expressed overnight at 37 °C. Cells were harvested by centrifugation and lysed by sonication. Libraries were subjected to two rounds of ammonium sulfate precipitation (50% saturation), followed by FPLC size exclusion chromatography purification to select for well-assembled VLPs.

2.5.3 Heat selection

Assembly-selected libraries were subjected to 50 °C for 10 min. Precipitated VLPs were pelleted via centrifugation, and assembled VLPs were isolated via semi-preparative HPLC size exclusion chromatography. Fractions containing assembled VLPs (at the characteristic elution time of 11.2 min) were combined, subjected to RNA extraction, barcoded, and identified via high-throughput sequencing.

2.5.4 Chemical modification selection

$P(NNK)_3$ libraries (final concentration 200 µM) were added to a solution of aminophenol-DNA (final concentration 1 mM) in 10 mM phosphate buffer, pH 7.2. $K_3Fe(CN)_6$ in Milli-Q water was added (final concentration 5 mM) and the solution was incubated at room temperature for 30 min. The reaction was quenched with tris(2-carboxyethyl)phosphine (TCEP, final concentration 50 mM) and excess oxidant and

DNA were removed using a 100 kDa MWCO filter spin filter (EMD Millipore, Burlington, MA). Well-formed VLPs were isolated by semi-preparative HPLC SEC, sample preparation, and high-throughput sequencing as described.

2.5.5 Sample prep for high-throughput sequencing

Plasmid DNA was isolated prior to expressions using Zyppy Plasmid Miniprep Kit (Zymo, Cat #D4036). RNA was extracted from the assembly-selected libraries using previously published protocols.[23,44] Briefly, homogenization was carried out with TRIzol (Thermo Fisher Cat #15596026), followed by chloroform addition. The sample was centrifuged to separate into aqueous, interphase, and organic layers. The aqueous layer, containing RNA, was isolated, and the RNA was precipitated with isopropanol and washed with 70% ethanol. RNA was dried and resuspended in RNAse free water. cDNA was synthesized with the Superscript III first strand cDNA synthesis kit from Life (Cat #18080051, polyT primer). PolyT primers have historically been used with success for SyMAPS projects,[23,24] likely because a small percent of the *E. coli* transcriptome is known to be polyadenylated in small amounts[45,46]. A head-to-head comparison of random hexamer primers with polyT primers shows that both successfully produce cDNA that can be used for downstream high-throughput sequencing steps (Additional Figure 2.22). Though unusual, this low level of polyadenylation, coupled with high expression levels achieved in the library generation stage, has led to the successful use of polyT primers. Future work will further compare polyT vs. random hexamer primers in this system. cDNA and plasmids were both barcoded for high-throughput sequencing. Both types of samples were amplified with two rounds of PCR (10 cycles, followed by 8 cycles) to add barcodes and Illumina sequencing handles, following Illumina 16S Metagenomic Sequencing Library Preparation recommendations. Libraries were quantified with Qubit and combined in equal molar ratio. Libraries were analyzed by 150 PE MiSeq in collaboration with the UC Davis Sequencing Facilities. A total of 18 million reads passed filter, and had an overall Q30 $>$ 85%.

2.5.6 Individual variant cloning

Individual variants were cloned through a variation of the methods described above. Briefly, overlap extension PCR was used to produce double-stranded fragment that spanned the length of the missing 26-codon region in the entry vector. Each fragment was cloned into the entry vector using standard Golden Gate cloning techniques.[47] Plasmids were then transformed into chemically competent DH10B cells. Plasmid identities were confirmed by Sanger sequencing prior to expression. Variants with

multiple mutations (i.e. CP[HiPerX–S37P] and CP[HiPerX–N87C]) were cloned into a similar Entry Vector bearing the desired mutation at position 37 or 87, which was installed via site-directed mutagenesis.[48]

2.5.7 Individual variant expression

Selected variants were expressed individually in 5 or 50 mL cultures of 2xYT. Expressions were lysed by sonication and precipitated twice with 50% ammonium sulfate. Variants were evaluated for assembly by HPLC SEC and thermostability by native gel.

2.5.8 Oxidative coupling of HiPerX MS2 variants

To a solution of HiPerX MS2 (final concentration 10 µM, 1 eq.) in 10 mM sodium phosphate buffer at pH 7.2 was added 4-methylcatechol in DMF (final concentration 100 µM, 10 eq.). To initiate oxidation, $K_3Fe(CN)_6$ in Milli-Q water was added (final concentration 1 mM, 100 eq.) and the solution was incubated at room temperature for 30 min. The reaction was quenched with tris(2-carboxyethyl)phosphine (TCEP, final concentration 10 mM) and excess oxidant and DNA were removed using a 100 kDa MWCO filter spin filter. Percent conversion was analyzed by ESI-TOF-LC/MS analysis and VLP integrity was confirmed by HPLC SEC.

2.5.9 Enzyme-catalyzed oxidative coupling of HiPerX MS2 variants

A solution of HiPerX MS2 (final concentration 10 µM, 1 eq.) in 50 mM sodium phosphate buffer at pH 6.5 was supplied with 4-methylcatechol in DMF (final concentration 100 µM, 10 eq.). To initiate the oxidative coupling reaction, tyrosinase enzyme (final concentration 0.5 µM) was added to the reaction mixture. After 2 h of incubation at room temperature, the reaction was quenched with a solution of tropolone and TCEP (final concentration 2.1 mM tropolone and 2.1 mM TCEP). Excess 4-methylcatechol was removed using a 100 kDa MWCO filter spin filter. The percent conversion was determined using ESI-TOF-LC/MS analysis and VLP integrity was confirmed by HPLC SEC.

2.5.10 2PCA modification of HiPerX MS2 variants

A 10 mM solution of 2-pyridiencarboxaldehyde (2PCA) in water was added to a solution of HiPerX MS2 (final concentrations: 50 µM MS2 CP, 12.5 mM 2PCA) in 10

mM sodium phosphate buffer, pH 7.2 with 100 mM NaCl. The reaction was allowed to proceed for 18 h at room temperature. Excess 2-pyridinecarboxaldehyde was removed using a 100 kDa MWCO filter spin filter. Percent conversion was analyzed by ESI-TOF-LC/MS, and VLP assembly was assessed by HPLC SEC.

2.5.11 MS2-fluorophore labeling

P-X-X-R/N87C MS2 (final concentration 100 µM) was mixed with Alexa Fluor 488-Maleimide (stock solution 10 mM in DMSO, final concentration 100 µM, Invitrogen, Cat #A10254) in 50 mM phosphate buffer, pH 7.2. Solution was briefly vortexed then incubated at room temperature for 1 h. Excess dye was removed by Nap-5 size exclusion column (GE Healthcare, Cat #17-0853-01) equilibrated with 10 mM phosphate buffer, pH 7.2. Modification of assembled VLPs was verified by HPLC SEC and quantified by ESI-MS.

2.5.12 FPLC SEC

MS2 CP libraries and select individual variants were purified on an Akta Pure 25 L Fast Protein Liquid Chromatography (FPLC) system with a HiPrep Sephacryl S-500 HR column (GE Healthcare Life Sciences, Cat #28935607) Size Exclusion Chromatography (SEC) column via isocratic flow with 10 mM phosphate buffer at pH 7.2 with 200 mM sodium chloride and 2 mM sodium azide.

2.5.13 HPLC SEC

Variants or libraries were analyzed on an Agilent 1290 Infinity HPLC with an Agilent Bio SEC-5 column (5 um, 2000Å, 7.8x300mm) with isocratic flow of 10 mM phosphate buffer at pH 7.2 with 200 mM sodium chloride and 2 mM sodium azide. Fractions were harvested at 11.2 min, or the characteristic elution time for wild-type MS2. Harvested VLPs were then subjected to RNA extraction and high-throughput sequencing sample preparation.

2.5.14 Native gel electrophoresis

Variants were analyzed in a 0.8% agarose gel in 0.5X TBE buffer (45 mM Tris-borate, 1 mM EDTA) and 2X SYBR Safe DNA Gel Stain (ThermoFisher Scientific, Cat #S33102) for 120 min at 40 V. Agarose gels were imaged on a BioRad GelDoc EZ Imager. Densitometry with ImageJ was carried out to compare experimental conditions to pH 7.4 in each case.

2.5.15 ESI-TOF

Modified and unmodified variants were analyzed with an Agilent 1200 series liquid chromatograph (Agilent Technologies, USA) connected in-line with an Agilent 6224 Time-of Flight (TOF) LC/MS system with a Turbospray ion source.

2.5.16 Strains

All library experiments were conducted with MegaX DH10B *E. coli* electrocompetent cells (ThermoFisher Scientific, Cat #C640003). Chemically competent *E. coli* DH10B cells were used for expression of individual variants of interest. Overnight growths from a single colony were incubated for 16-20 h at 37 °C shaking at 200 RPM in LB-Miller media (Fisher Scientific, Cat #BP1426-2) with chloramphenicol at 32 mg/L. Expressions were subcultured 1:100 into 2xYT media (Teknova, Cat #Y0210) with 32 mg/L chloramphenicol, allowed to grow to an OD600 of 0.6, then induced with 0.1% arabinose. Expressions continued overnight at 37 °C shaking at 200 RPM.

2.5.17 High-throughput sequencing data analysis

Data were trimmed and processed described previously,[23] with minor variation. Briefly, data were trimmed with Trimmomatic[49] with a 4-unit sliding quality window of 20 and a minimum length of 32. Reads were merged with FLASh (fast length adjustment of short reads) with a maximum overlap of 150 base pairs. Reads were sorted and indexed with SAMtools[50] and unmapped reads were filtered with the Picard function CleanSam. Reads shorter than 106 base pairs were removed. Reads were processed with in-house code to produce various AFLs.

2.5.18 AFL calculations

Cleaned and filtered high-throughput sequencing reads were analyzed using Python programs written in-house. Briefly, the N-terminal region of the MS2 CP was located, and the three codons following the proline at the new N terminus were recorded. Codons were then translated into amino acids to generate counts for each tripeptide combination. These calculations were repeated for all experiments to generate abundances before and after each selective pressure. Relative abundances were calculated similarly to the previously described protocol.[23] Briefly, the relative percent abundance of a selection was calculated by dividing the percent abundance generated from an assembled VLP, heat-selected, or chemically-selected library by the plasmid percent abundance from its respective biological replicate. Two biological

replicates were assessed for each challenge except for the heat-selected library, where one biological replicate was analyzed. All Nan (null) values, which indicate variants that were not identified in the plasmid library were ignored. Scores of zero, which indicate variants that were sequenced in the unselected library but absent in the VLP library, were replaced with an arbitrary score of 0.0001. We calculated the mean relative abundances across replicates. We then calculated the log10 of the Relative Abundance array to calculate the final array for each replicate. The log10 relative abundance value for an individual variant in a particular challenge is referred to as its Apparent Fitness Score (AFS). Arrays are displayed in Figure 2.3 and Additional Figures 2.9, 2.10, 2.19, and 2.20. Heatmaps for each of five individual biological replicates (two replicates of P-X-X-X-MS2 and three replicates of X-X-X-MS2) are shown in Additional Figures 19-23, and the correlation between all replicates are shown in Additional Figure 4, with plasmid abundance cutoffs indicated where relevant. For the correlation analyses, any extension with a –4 or nan score in either replicate was not included.

2.5.19 High-stringency heatmap calculation

Two additional data processing methods were applied to the P-X-X-X libraries in order reduce the impact of stochastic variation on the generated heatmaps show in Additional Figures 2.13-2.18. The first set of heatmaps were produced by removing low read count variants from the dataset. Variants with three or fewer reads in the plasmid library of any individual replicate were treated as null values in the average heatmap and are shown in green (Additional Figures 2.13-2.15). The second processing method removed ambiguous fitness scores and simplified scoring to a positive (blue) or negative (red) score. In this filter, any variant with a fitness score between -0.2 and 0.2 in any individual replicate was considered a null value and colored in grey. Variants with scores of opposite charge between replicates were also removed. The remaining variants were marked blue if their fitness scores were >0.2 in all replicates or marked red if their fitness scores were $<$-0.2 in all replicates (Additional Figures 2.16-2.18).

2.5.20 Making the aggregate AFL

The aggregate AFL was produced using in-house Python code. In brief, variants were sorted by their AFS values in assembly, heat, and chemical selection. Variants with an AFS value greater than or equal to 1.0 in all three criteria were designated Highest Performing Extensions (HiPerX). Variants with AFS values of less that 1.0 in one or

more challenge(s) were sorted as shown in Figure 2.4b. If a variant returned a null value for any of the three challenges it was sorted into the not observed category.

2.5.21 Random selection of HiPerX variants

All HiPerX variants of the form P-X-X-R-MS2 were assigned a unique value from 1-65. Five random integers between 1-65 were produced using Python's *randint()* function and the corresponding variants were subsequently cloned and expressed.

2.5.22 Computational modeling of extended MS2 coat proteins

A hexameric subunit of the MS2 capsid was prepared in based off of the crystal structure of wildtype MS2 (PDB ID:2MS2). The structure was imported into Schrodinger's Maestro suite and various N-terminal extensions were constructed with the build tool. Preprocessing of the extended structures was performed with Maestro's protein preparation wizard. In brief, hydrogen bonds were assigned with the H-bond optimization tool at a PROPKA pH of 7. Subsequently, a restrained minimization of the structure using an OPLS3e forcefield was performed. MacroModel was then used to carry out a Large-scale Low Mode conformational search of the minimized structures. All residues within 10 Å of the N-terminal extensions were restrained with a force constant of 200 kJ/mol. Atoms beyond this subshell were frozen in place. Sampling used 1000 maximum steps with 100 steps per rotatable bond.

2.6 Additional Figures

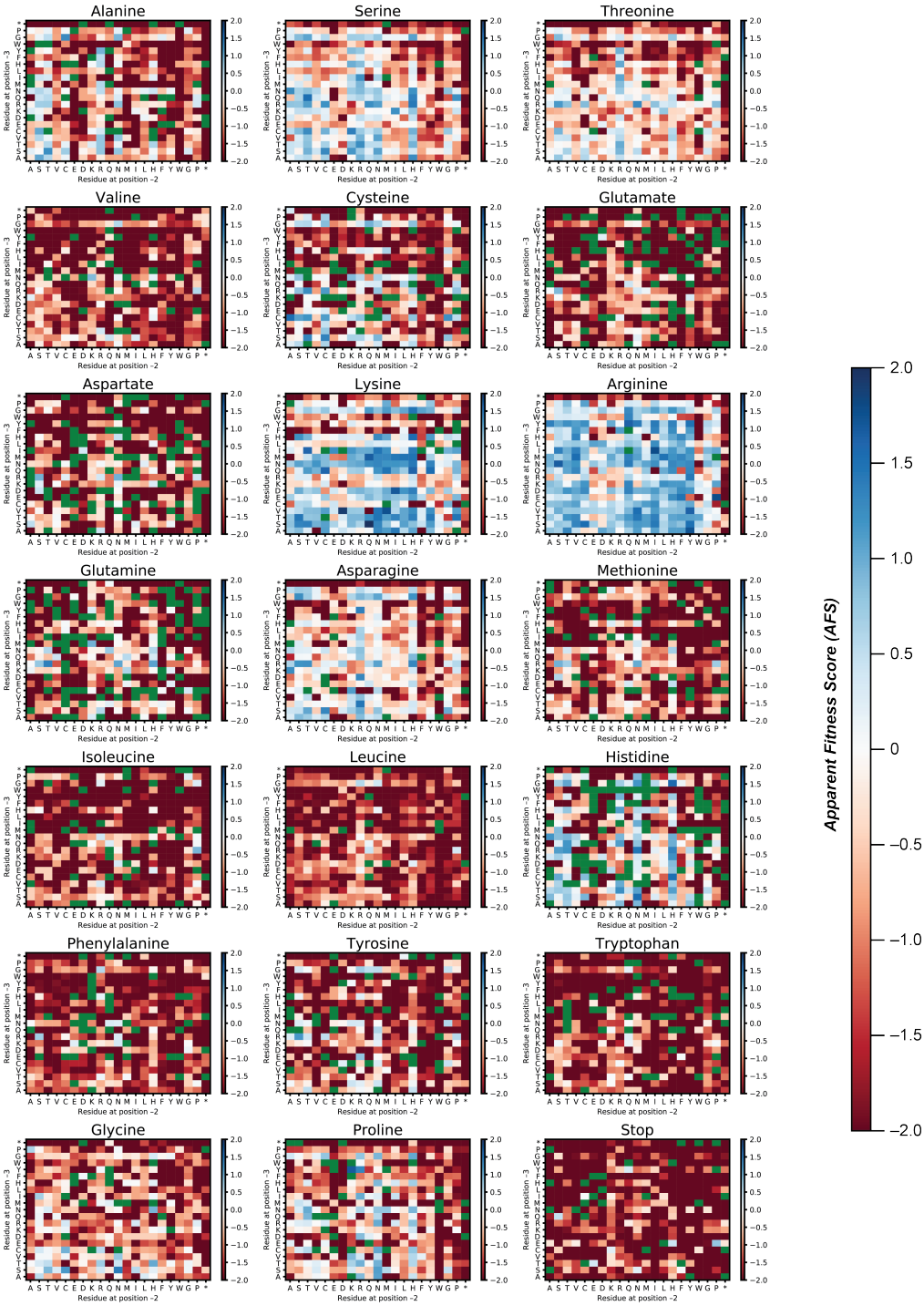

Figure 2.9: Labeled full assembly-selected AFL of N-terminal extensions with the pattern P-X-X-X-MS2, as shown in Figure 2.3

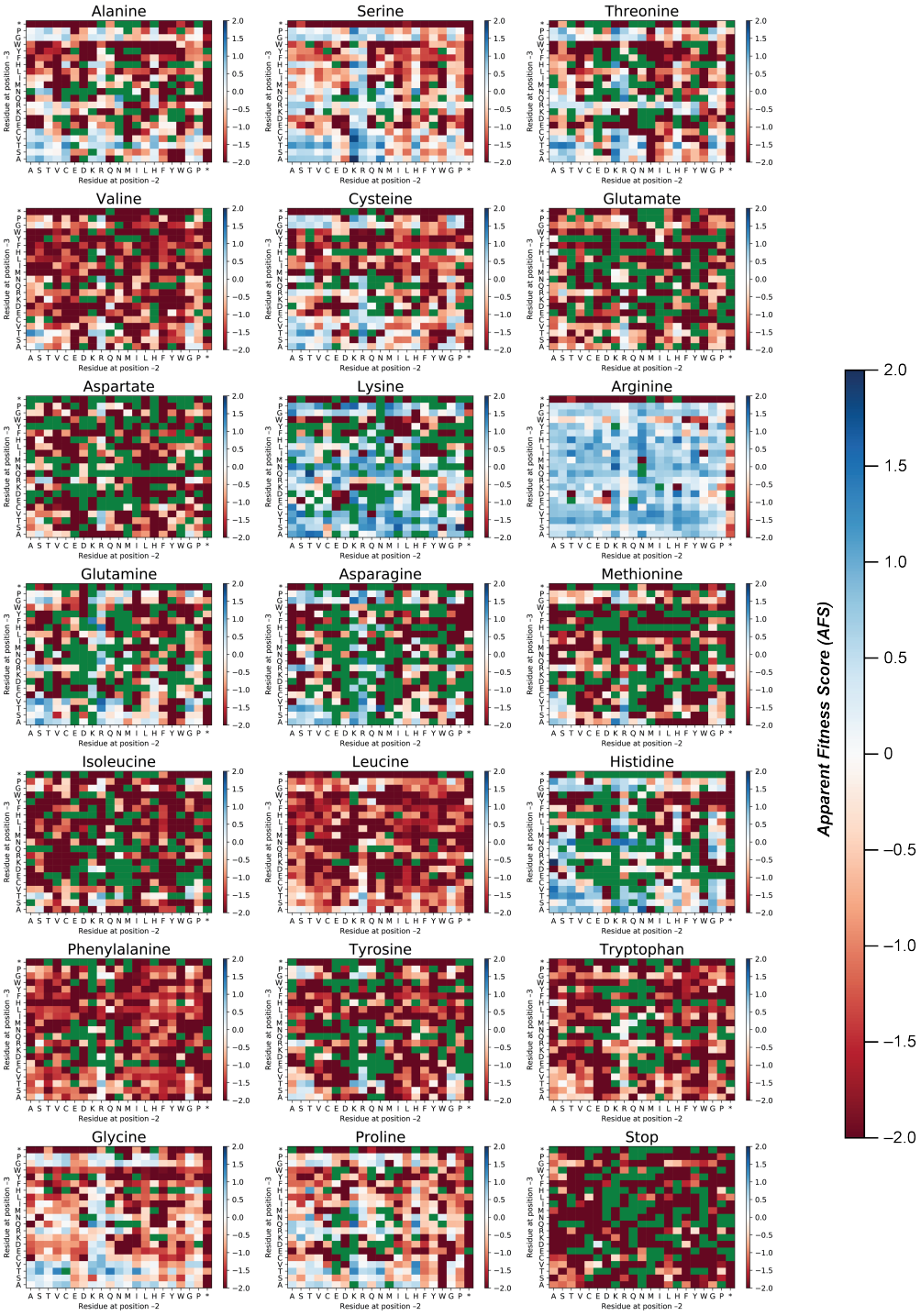

Figure 2.10: Assembly-selected AFL of N-terminal extensions with the pattern X-X-X-MS2. Blue indicates enriched amino acids, red indicates combinations that are not enriched, and green indicates missing values.

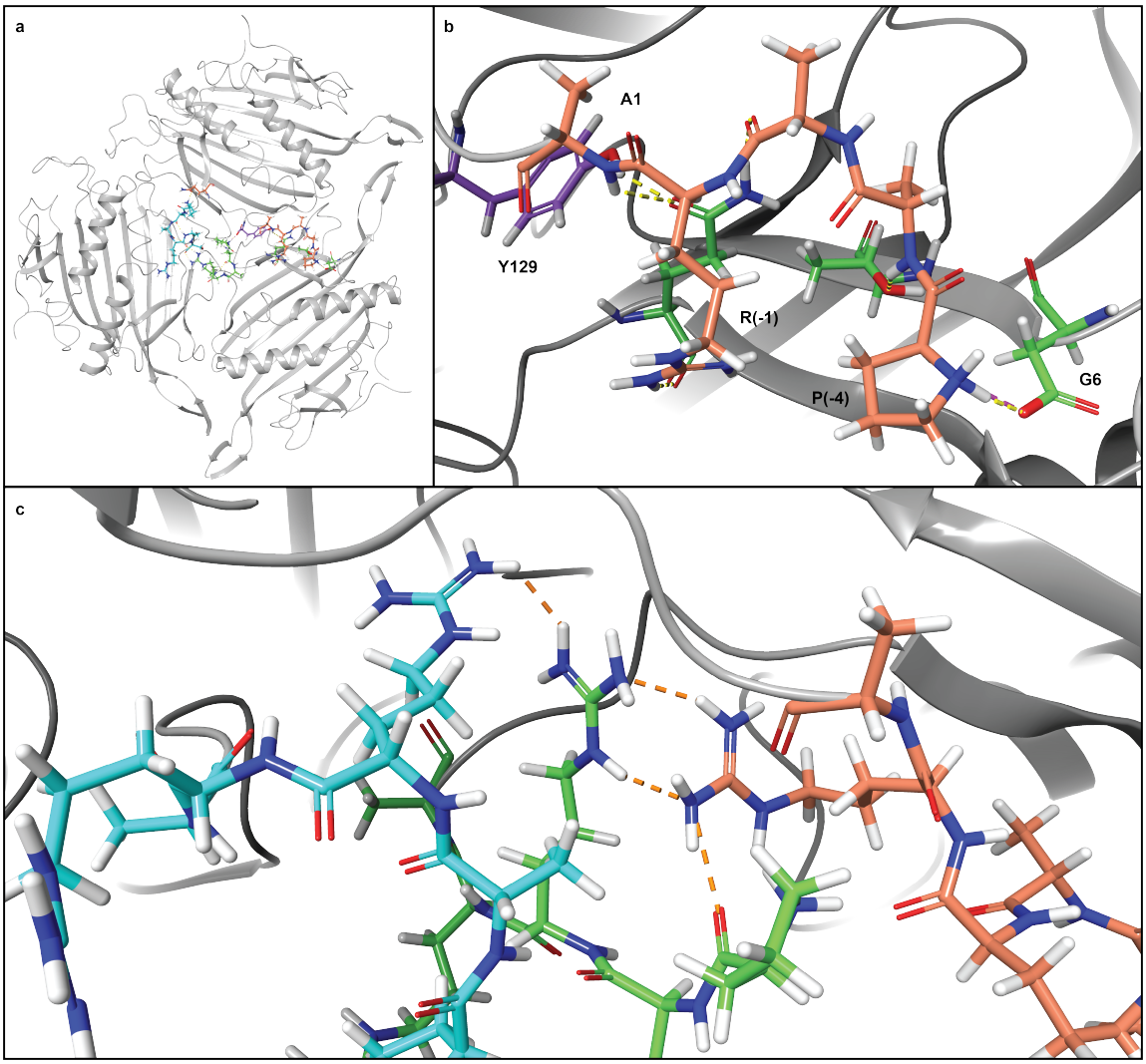

Figure 2.11: Modeling of N-terminally extended MS2 coat protein variants. a) Modeling of the extended variants uses a hexameric subunit of the MS2 capsid. The A, B, and C forms of the coat protein monomer are labeled in blue, orange, and green respectively. b) A view of P-A-A-R MS2 is shown with hydrogen bonds in yellow and salt bridges in pink. c) A close up of P-A-R-R MS2 is presented with unfavorable van der Waals interactions highlighted in orange.

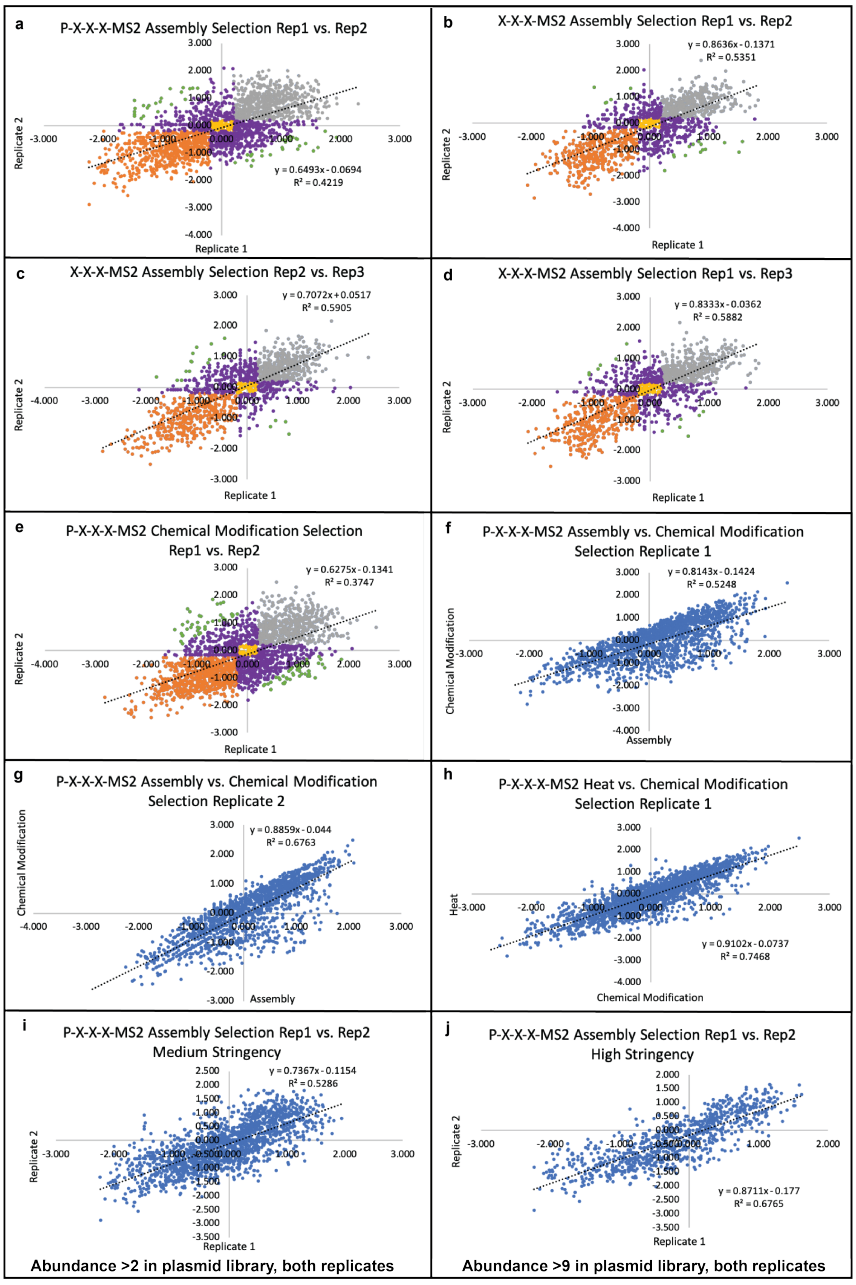

Figure 2.12: Correlation analyses of libraries described in this work. In all cases, variants with –4 or NaN values in either replicate were excluded. The correlations for a) P-X-X-X-MS2 Assembly Replicates 1-2 and (b-d) X-X-X-MS2 Assembly replicates 1-3 are shown. e) Correlation of modification conditions for P-X-X-X-MS2 is shown. In each case, extensions with two positive scores (> 0.2) are shown in gray, and replicates with two negative scores (<-0.2) are shown in orange. Replicates with scores close from –0.2 to 0.2 in both replicates are shown in yellow, while green indicates extensions that are both inverted and greater than 1.5 apart. The remaining scores are in purple. In addition, (f-h) correlation between various selections within a given replicate are shown. Finally, medium and high stringency correlations for the P-X-X-X-MS2 library, determined by abundances (>2 and >9, respectively), are shown.

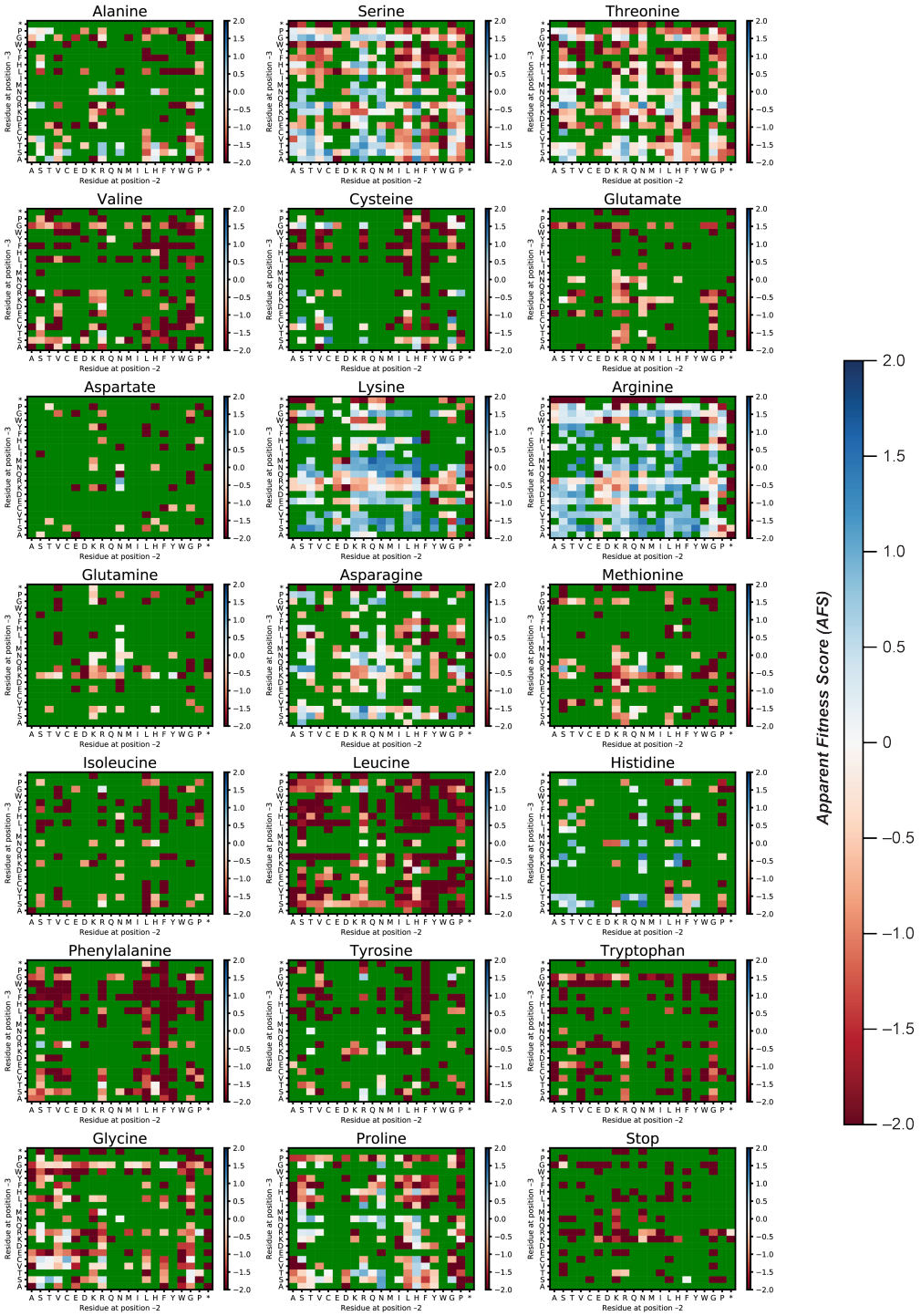

Figure 2.13: Apparent Fitness Landscape of P-X-X-X-MS2 N-terminal extensions with low read filter. Variants with three or fewer reads in the plasmid library of either replicate were treated as a missing value and are shown in green.

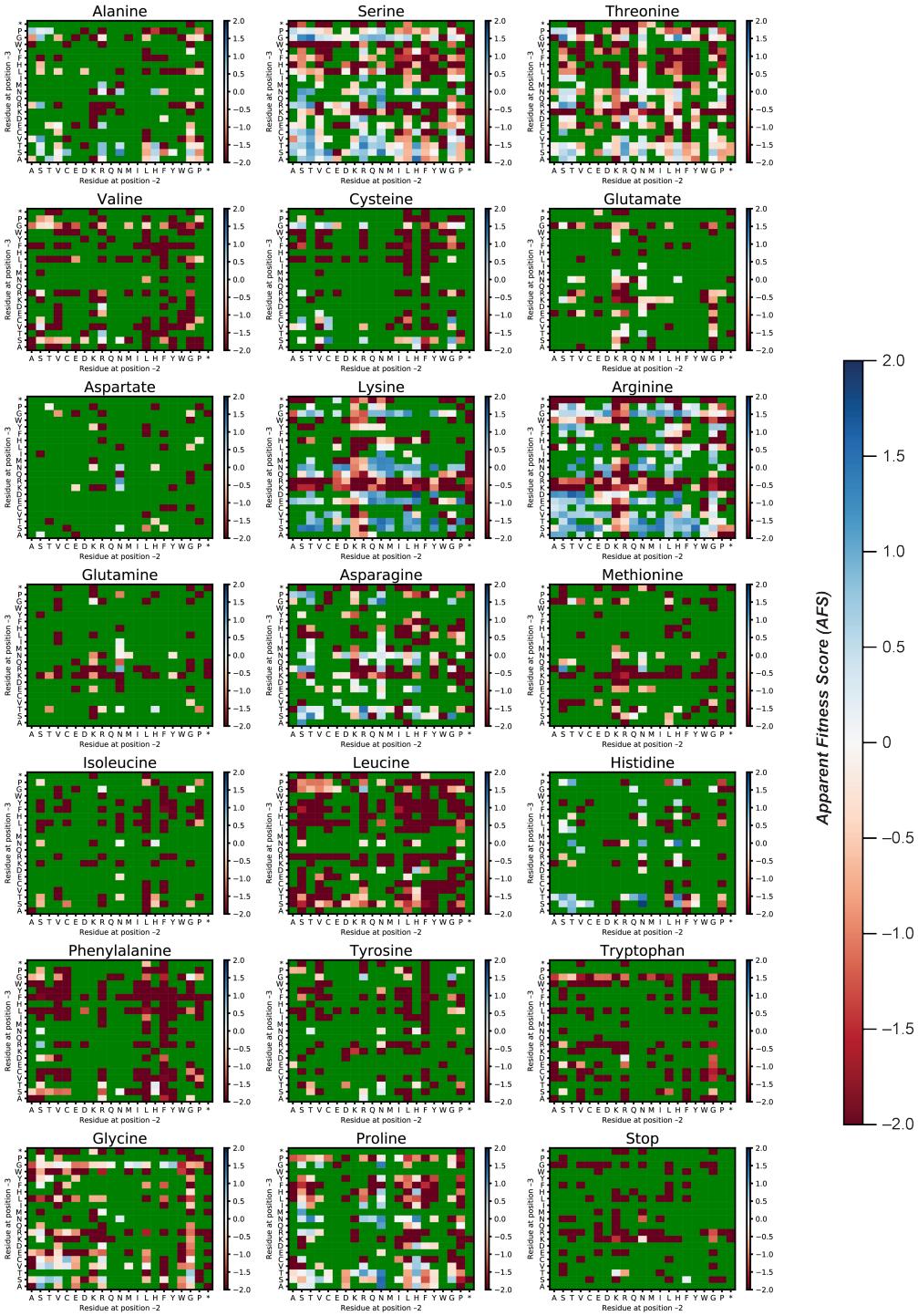

Figure 2.14: Chemical challenge of P-X-X-X-MS2 N-terminal extensions with low read filter. Variants with three or fewer reads in the plasmid library of either replicate were treated as a missing value and are shown in green.

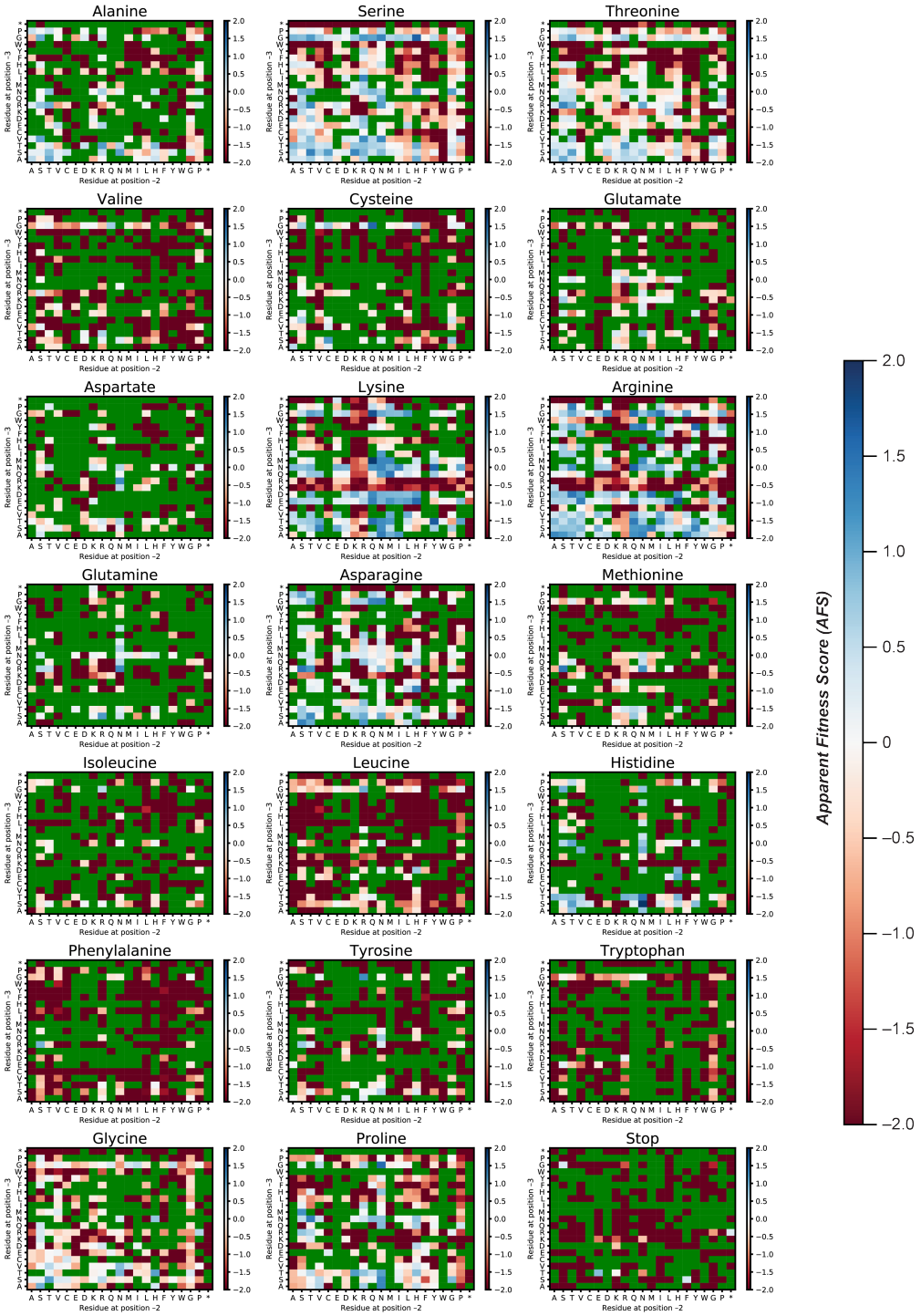

Figure 2.15: Heat challenge of P-X-X-X-MS2 N-terminal extensions with low read filter. Variants with three or fewer reads in the plasmid library were treated as a missing value and are shown in green.

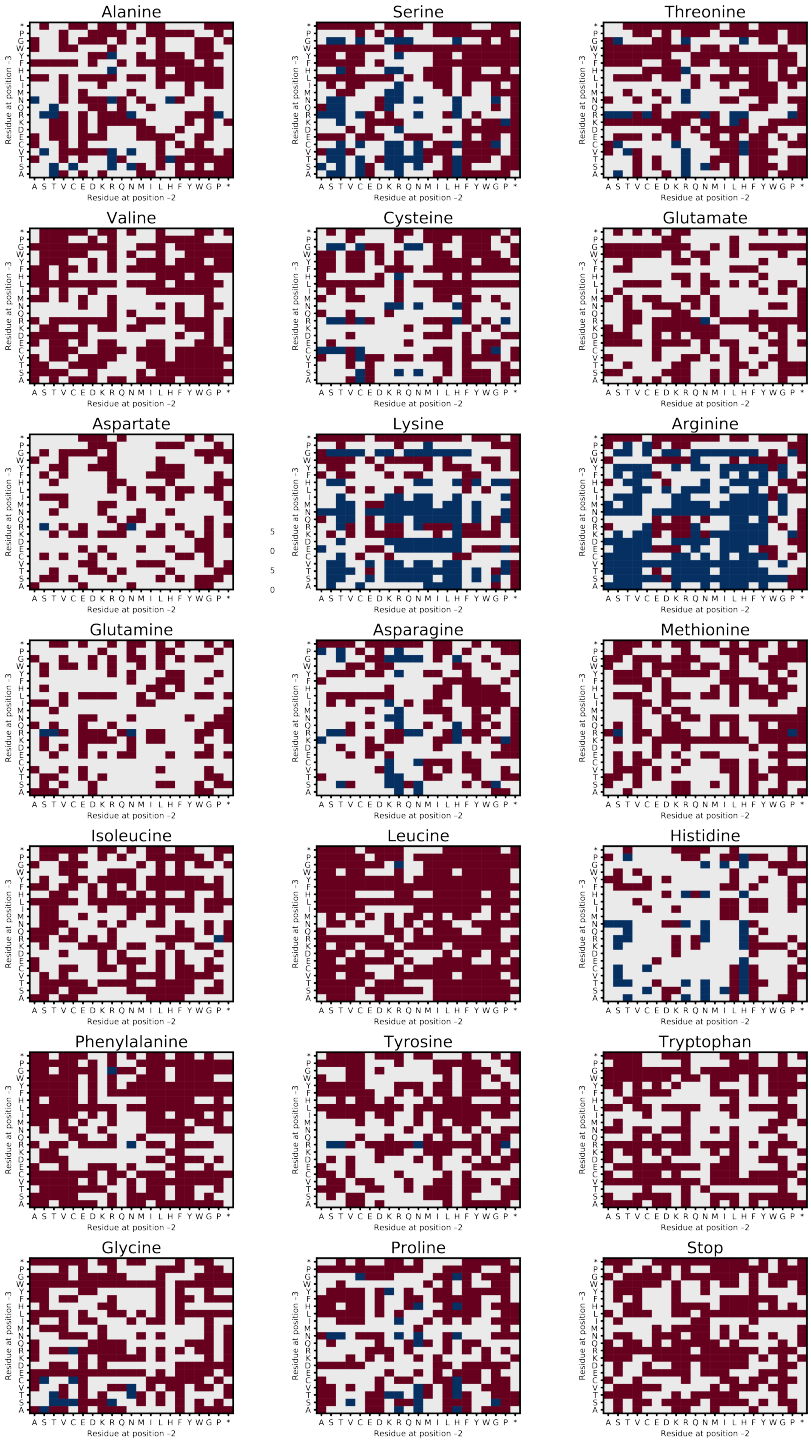

Figure 2.16: Apparent Fitness Landscape of P-X-X-X-MS2 N-terminal extensions with alternative data processing. Variants with an AFL score between -0.2 and 0.2 in either replicate or with opposite signs across replicates are shown in grey. Variants in which both replicate scores are >0.2 are shown in blue and variants with both replicate scores <-0.2 are shown in red.

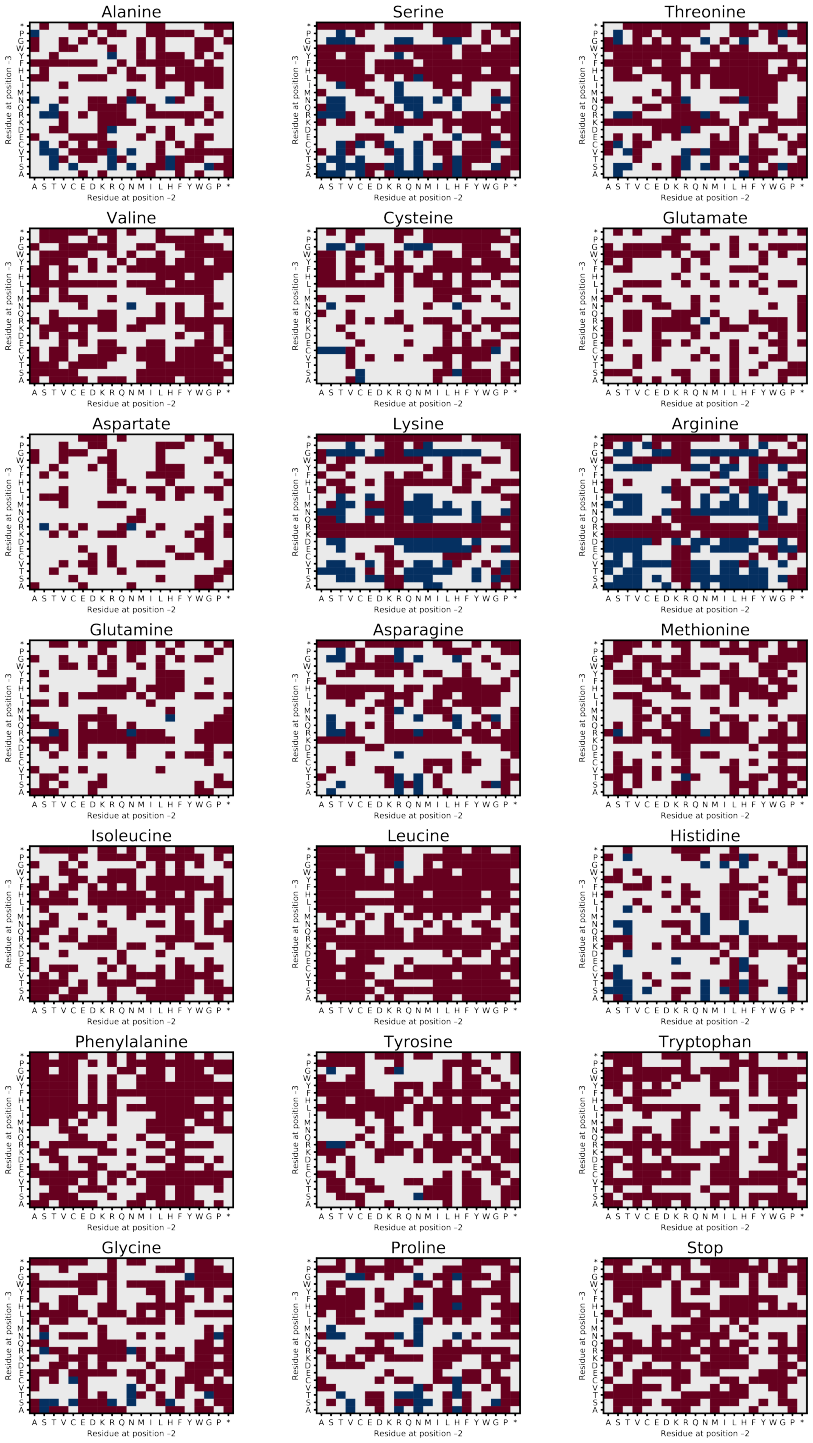

Figure 2.17: Chemical challenge of P-X-X-X-MS2 N-terminal extensions with alternative data processing. Variants with a chemical challenge score between -0.2 and 0.2 in either replicate or with opposite signs across replicates are shown in grey. Variants in which both replicate scores are >0.2 are shown in blue and variants with both replicate scores <-0.2 are shown in red.

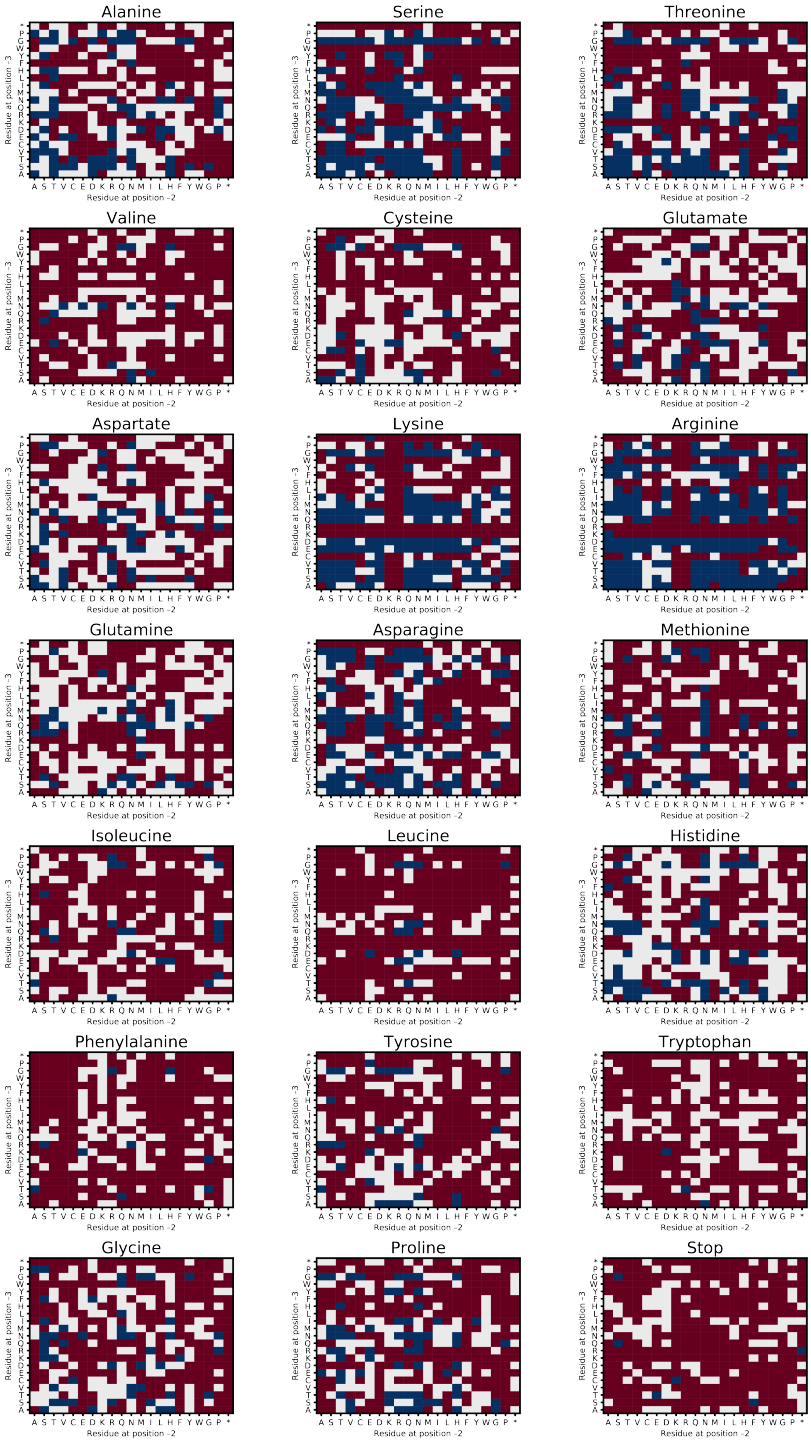

Figure 2.18: Heat challenge of P-X-X-X-MS2 N-terminal extensions with alternative data processing. Variants with a score between -0.2 and 0.2 in were removed and are shown in grey. Variants with heat challenge scores >0.2 are shown in blue and variants with a score <-0.2 are shown in red.

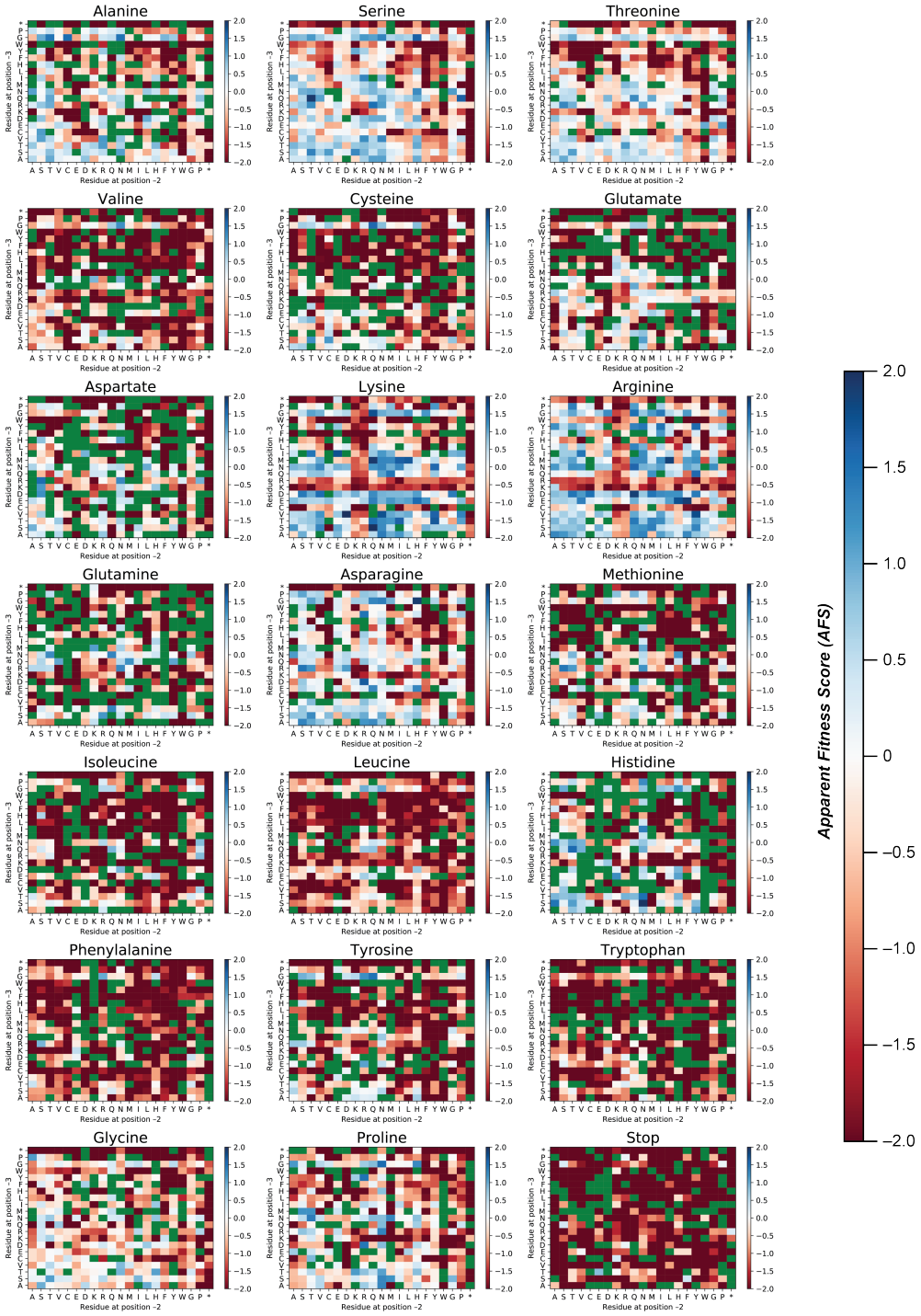

Figure 2.19: Heat-selected AFL for the P-X-X-X-MS2 library. The library was subjected to 50 °C for ten minutes, and assembled VLPs were enriched with semi-preparative HPLC size exclusion chromatography. Enriched amino acids are blue, variants that are not enriched are indicated in red, and missing values are shown in green.

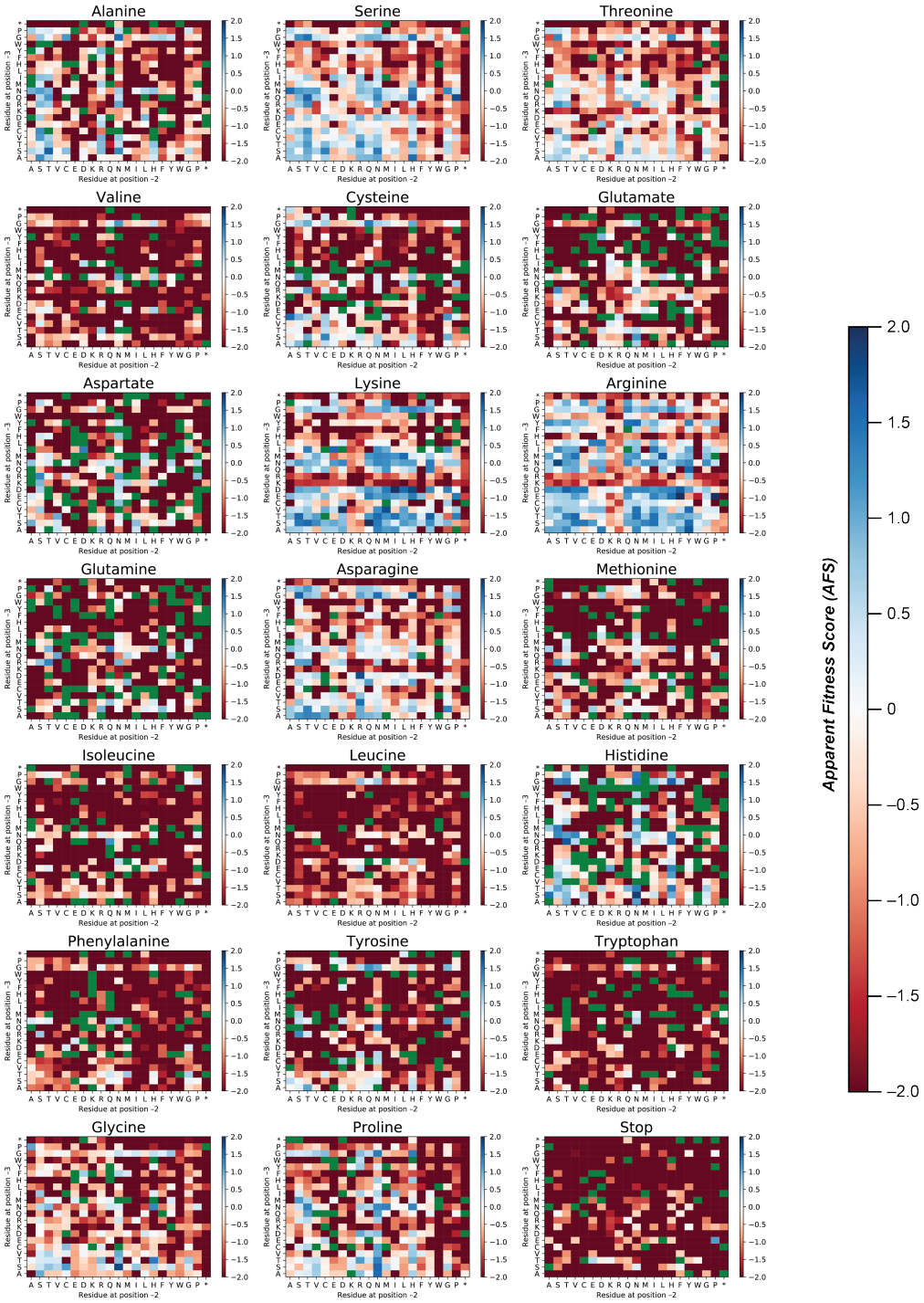

Figure 2.20: Chemical modification AFL of the P-X-X-X-MS2 library. The library was subjected to chemical modification conditions, and assembled VLPs were enriched with semi-preparative HPLC size exclusion chromatography. Enriched combinations are blue, variants that are not enriched are shown as red, and missing values are green.

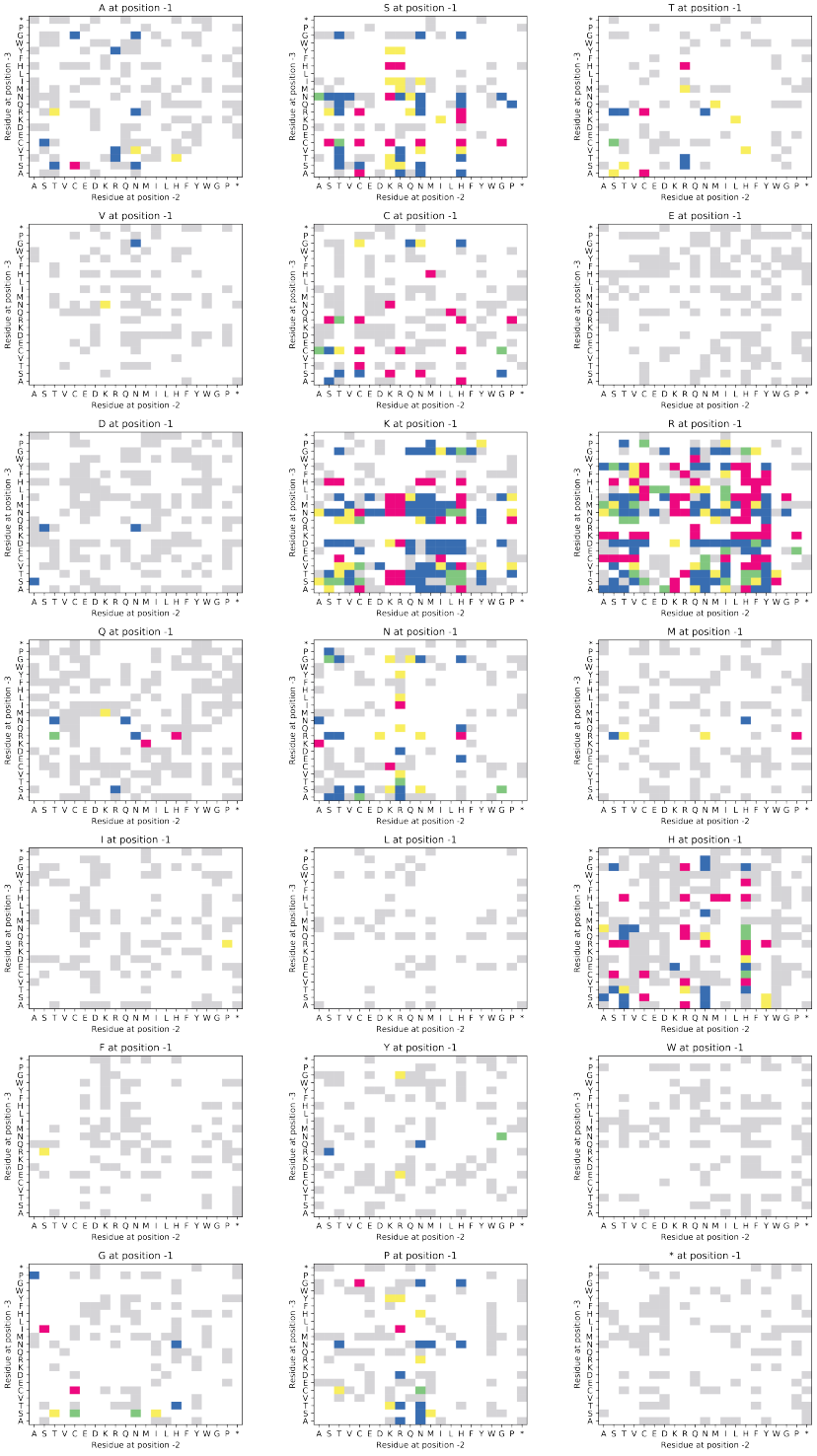

Figure 2.21: An aggregated heatmap combining results from the assembly, thermal, and chemical modification selections. a) All combinations of P-X-X-X-MS2 are given a color based on the key shown in Figure 2.5b

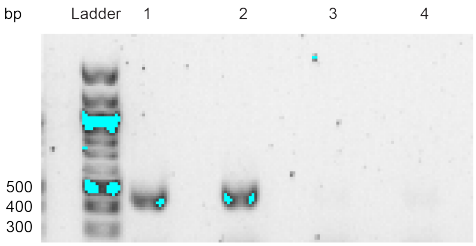

Figure 2.22: Comparison of polyT and random hexamer primers for cDNA synthesis. Agarose gel shows DNA following cDNA synthesis and the first PCR amplification of barcoding. Lanes 1 and 2 contain DNA synthesized from wtMS2-derived RNA using polyT and random hexamer primers, respectively. Lanes 3 and 4 contain DNA synthesized from an assembly-incompetent MS2 variant using polyT and random hexamer primers, respectively.

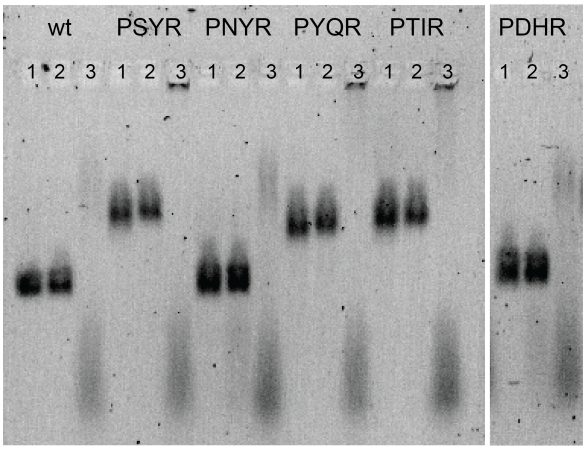

Figure 2.23: Native agarose gel of HiPerX variants following a thermal challenge. Lanes 1-3 for each variant were incubated at room temperature, 50° C, and 100° C respectively.

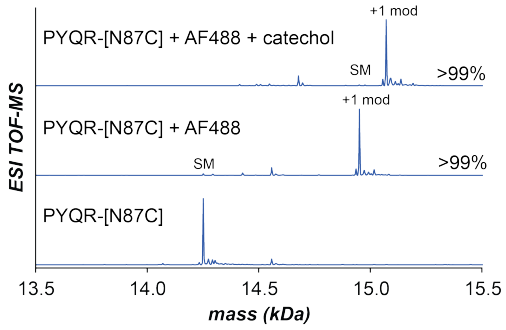

Figure 2.24: Dual chemical modification of CP[PYQR-N87C] MS2. CP[PYQR-N87C] can be modified with an AF-488 maleimide dye to quantitative conversion. Subsequent modification with a small molecule cathechol via oxidative coupling proceeds to full conversion as well.

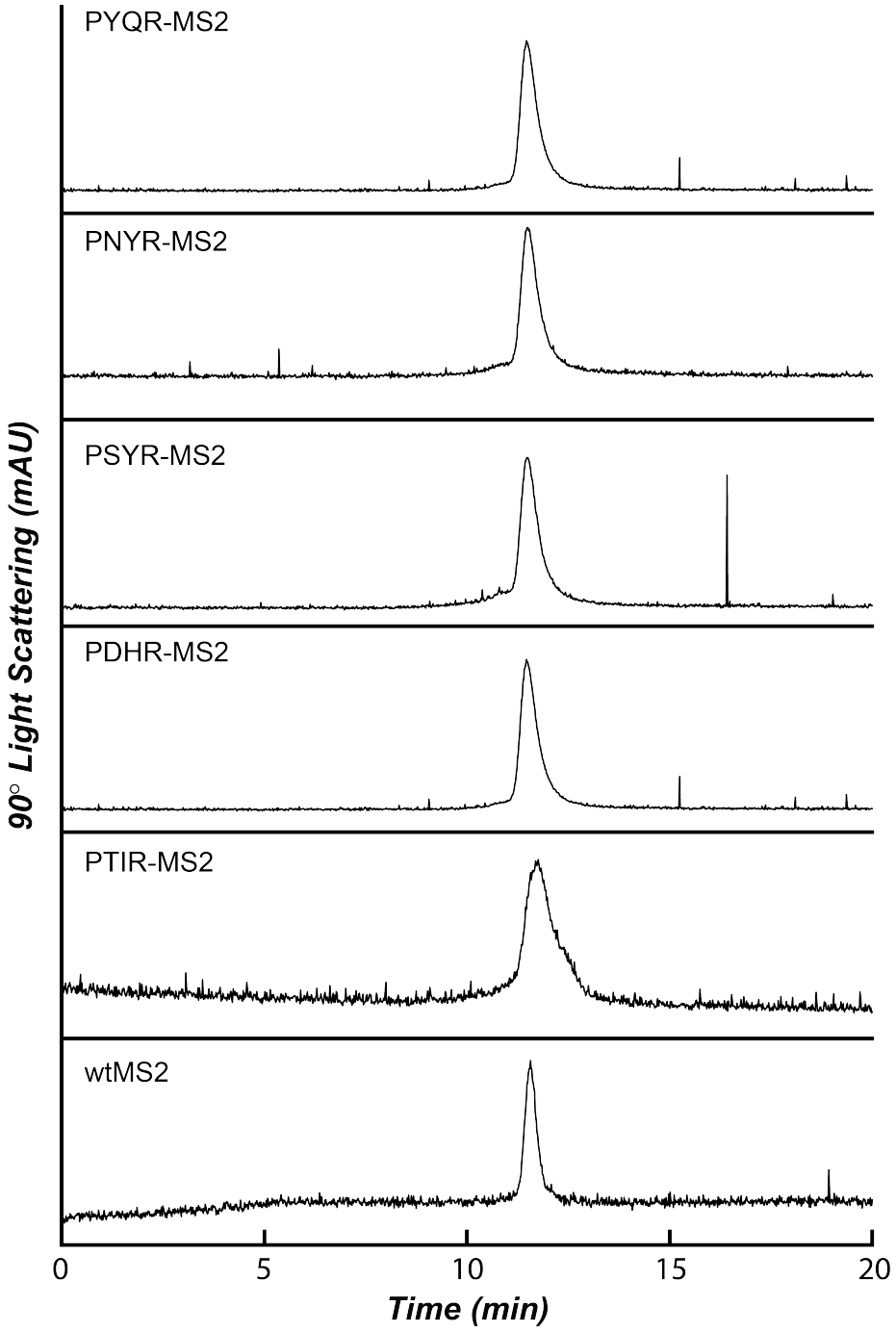

Figure 2.25: HPLC SEC traces of CP[HiPerX] variants following $K_3Fe(CN)_6$-mediated oxidative coupling. Presence of a light scattering peak at 11 min indicates that the capsids remain assembled.

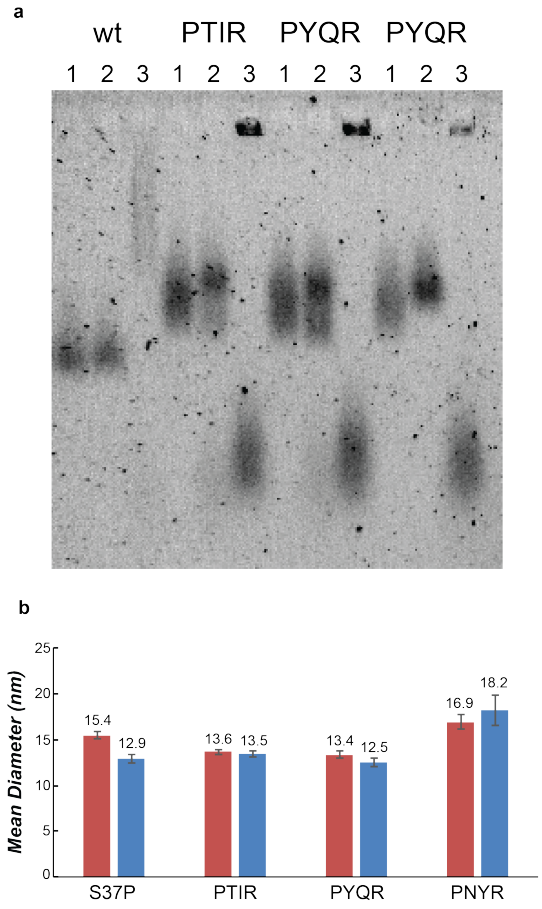

Figure 2.26: CP[HiPerX–S37P] variants following thermal and modification challenges. a) Native agarose gel of CP[HiPerX-S37P] variants following a thermal challenge. Lanes 1-3 for each variant represent incubation at room temperature, 50° C, and 100° C respectively. b) DLS of CP[HiPerX-S37P] prior to small molecule oxidative coupling (red) and following modification (blue).

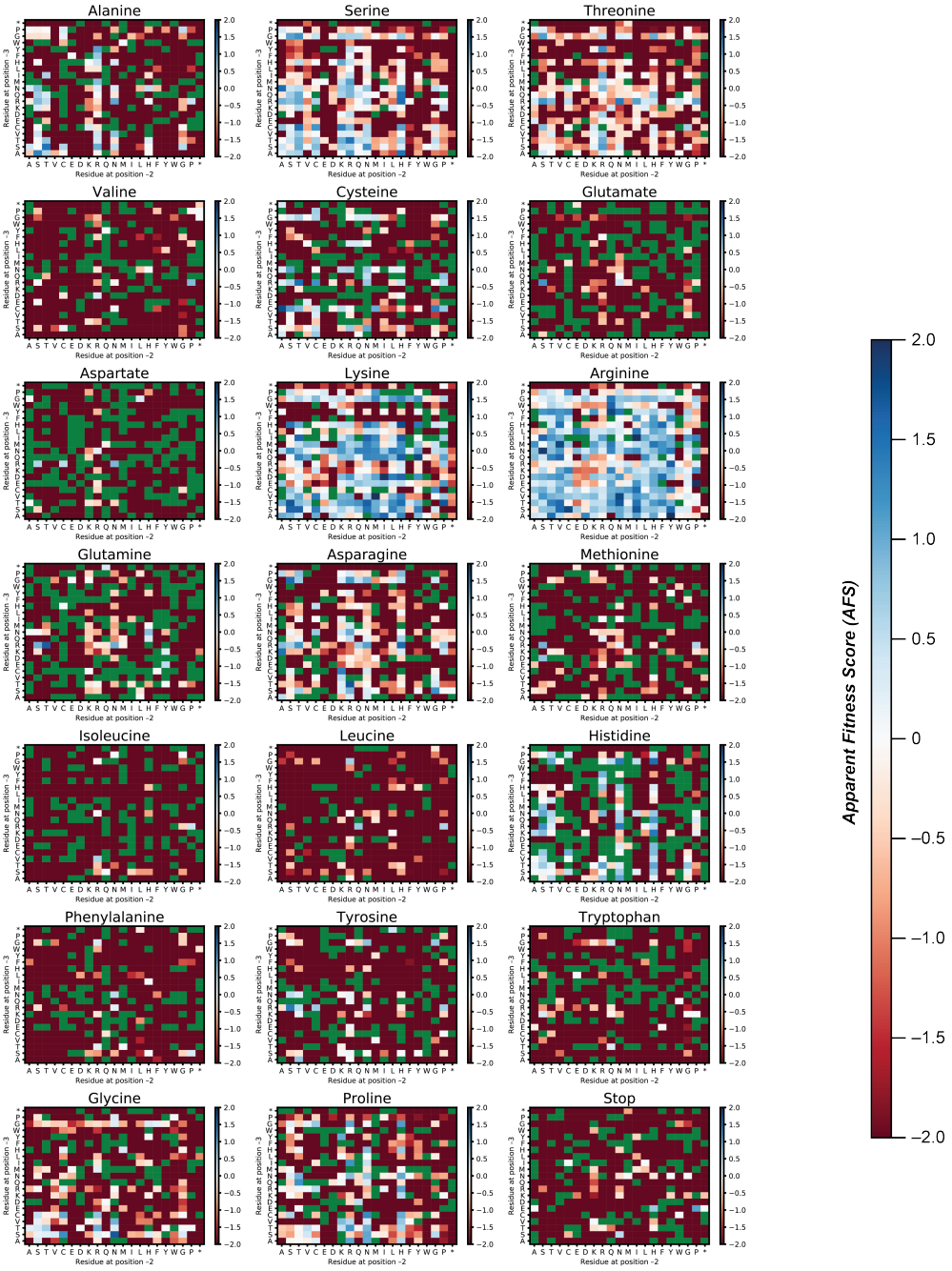

Figure 2.27: Replicate one of the assembly-selected AFL of N-terminal extensions with the pattern P-X-X-X-MS2.

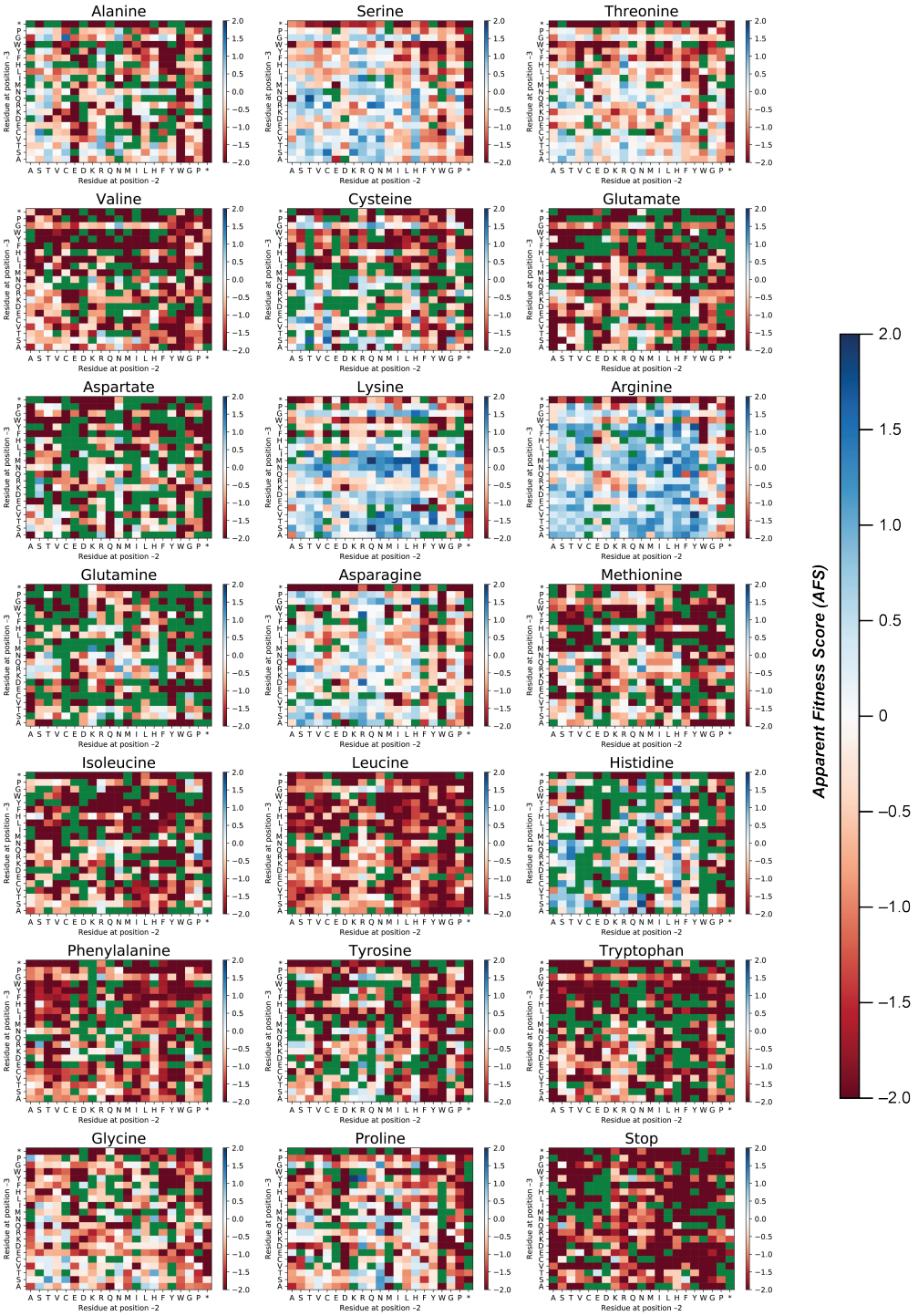

Figure 2.28: Replicate two of the assembly-selected AFL of N-terminal extensions with the pattern P-X-X-X-MS2.

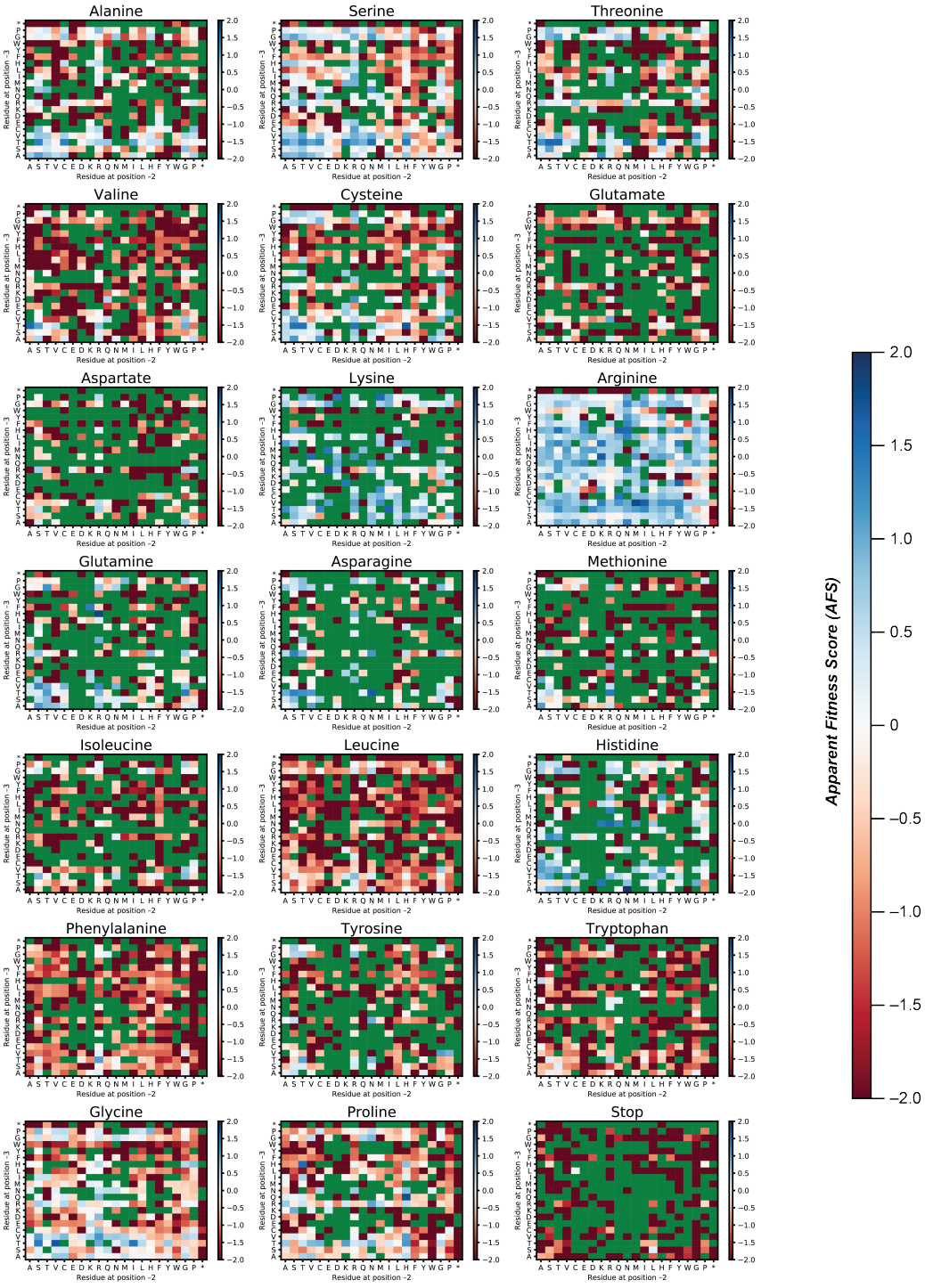

Figure 2.29: Replicate one of the assembly-selected AFL of N-terminal extensions with the pattern X-X-X-MS2.

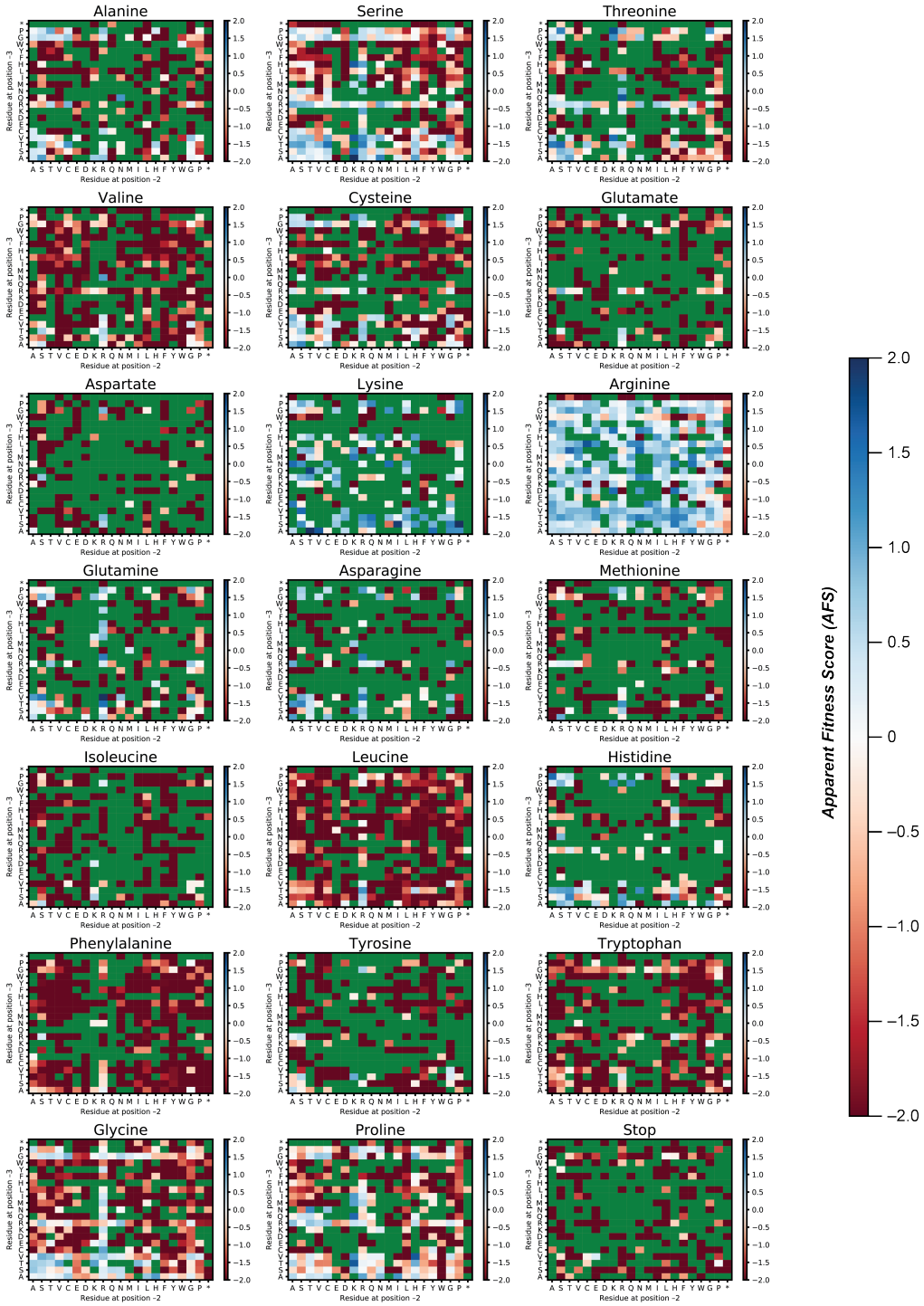

Figure 2.30: Replicate two of the assembly-selected AFL of N-terminal extensions with the pattern X-X-X-MS2.

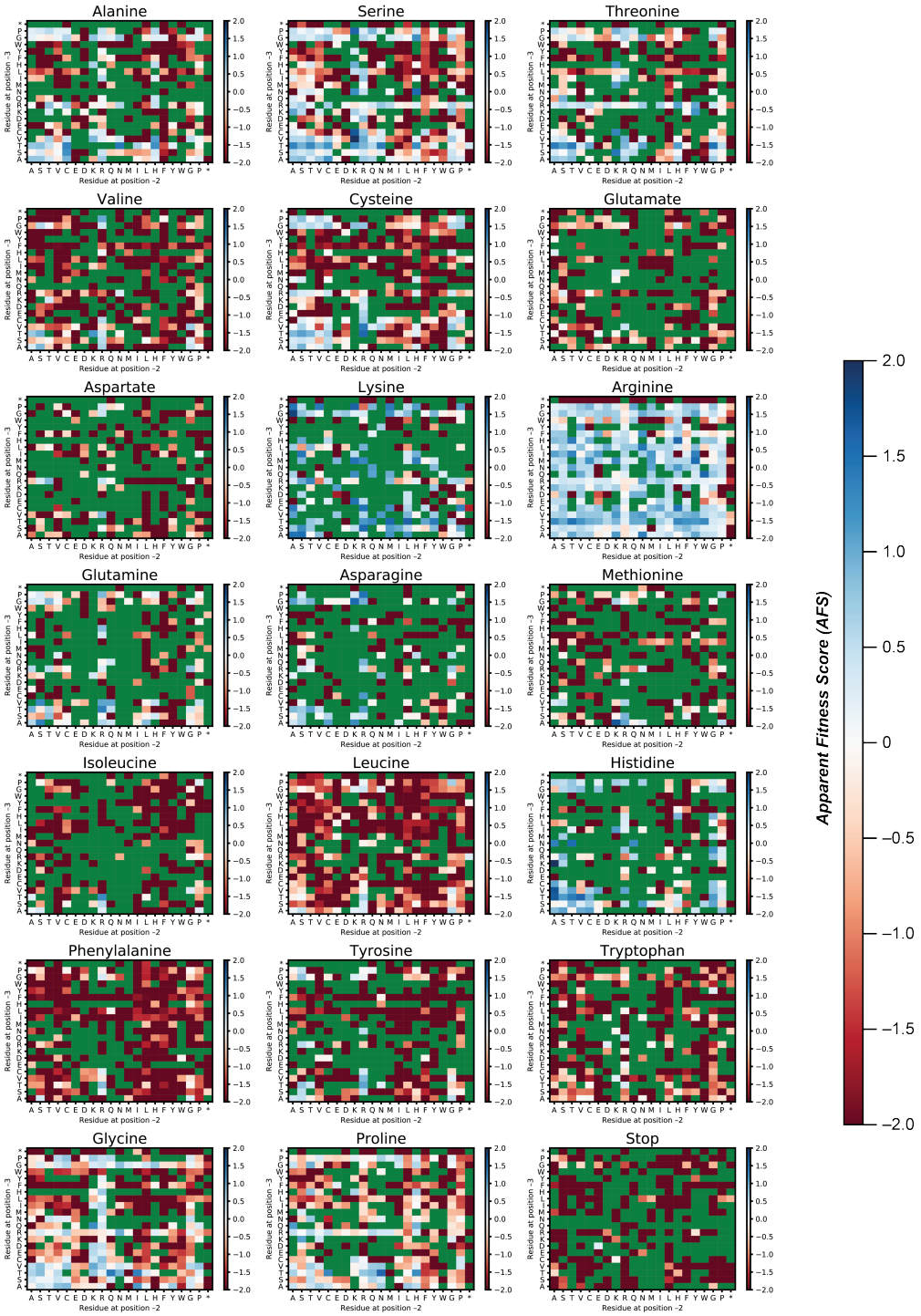

Figure 2.31: Replicate three of the assembly-selected AFL of N-terminal extensions with the pattern X-X-X-MS2.

2.7 Acknowledgements

Illumina next-gen sequencing was carried out by the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant S10OD010786.

2.8 References

(1) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *Journal of the American Chemical Society* **2014**, *136*, 9572–9579.

(2) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nature Chemical Biology* **2015**, *11*, 326–331.

(3) Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M.; Weng, C.-H.; Iavarone, A. T.; Francis, M. B. *Journal of the American Chemical Society* **2013**, *135*, 17223–17229.

(4) Sur, S.; Qiao, Y.; Fries, A.; O'Meally, R. N.; Cole, R. N.; Kinzler, K. W.; Vogelstein, B.; Zhou, S. *Scientific Reports* **2016**, *5*, 18363.

(5) Spicer, C. D.; Pashuck, E. T.; Stevens, M. M. *Chemical Reviews* **2018**, *118*, 7702–7743.

(6) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. *Tetrahedron Letters* **2000**, *41*, 4069–4073.

(7) Geoghegan, K. F.; Stroh, J. G. *Bioconjugate Chemistry* **1992**, *3*, 138–146.

(8) Casi, G.; Huguenin-Dezot, N.; Zuberbühler, K.; Scheuermann, J.; Neri, D. *Journal of the American Chemical Society* **2012**, *134*, 5887–5892.

(9) Palla, K. S.; Witus, L. S.; Mackenzie, K. J.; Netirojjanakul, C.; Francis, M. B. *Journal of the American Chemical Society* **2015**, *137*, 1123–1129.

(10) Rosen, C. B.; Francis, M. B. *Nature Chemical Biology* **2017**, *13*, 697–705.

(11) Li, D.-z.; Han, B.-n.; Wei, R.; Yao, G.-y.; Chen, Z.; Liu, J.; Poon, T. C.; Su, W.; Zhu, Z.; Dimitrov, D. S.; Zhao, Q. *mAbs* **2018**, *10*, 712–719.

(12) Lee, J. P.; Kassianidou, E.; MacDonald, J. I.; Francis, M. B.; Kumar, S. *Biomaterials* **2016**, *102*, 268–276.

(13) Esser-Kahn, A. P.; Iavarone, A. T.; Francis, M. B. *Journal of the American Chemical Society* **2008**, *130*, 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. *Bioconjugate Chemistry* **2015**, *26*, 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.; Brinker, C. J.; Caldeira, J. d. C.; Chackerian, B.; Wharton, W.; Peabody, D. S. *ACS Nano* **2011**, *5*, 5729–5745.

(16) Aanei, I. L.; Huynh, T.; Seo, Y.; Francis, M. B. *Bioconjugate Chemistry* **2018**, *29*, 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. *Molecular Pharmaceutics* **2013**, *10*, 69–76.

(18) Crossey, E.; Frietze, K.; Narum, D. L.; Peabody, D. S.; Chackerian, B. *PLOS ONE* **2015**, *10*, ed. by Pizarro, J. C.

(19) Zhai, L.; Peabody, J.; Pang, Y.-Y. S.; Schiller, J.; Chackerian, B.; Tumban, E. *Antiviral Research* **2017**, *147*, 116–123.

(20) Capehart, S. L.; Coyle, M. P.; Glasgow, J. E.; Francis, M. B. *Journal of the American Chemical Society* **2013**, *135*, 3011–3016.

(21) Glasgow, J. E.; Asensio, M. A.; Jakobson, C. M.; Francis, M. B.; Tullman-Ercek, D. *ACS Synthetic Biology* **2015**, *4*, 1011–1019.

(22) Glasgow, J. E.; Capehart, S. L.; Francis, M. B.; Tullman-Ercek, D. *ACS Nano* **2012**, *6*, 8658–8664.

(23) Hartman, E. C.; Jakobson, C. M.; Favor, A. H.; Lobba, M. J.; Álvarez-Benedicto, E.; Francis, M. B.; Tullman-Ercek, D. *Nature Communications* **2018**, *9*, 1385.

(24) Hartman, E. C.; Lobba, M. J.; Favor, A. H.; Robinson, S. A.; Francis, M. B.; Tullman-Ercek, D. *Biochemistry* **2019**, *58*, 1527–1538.

(25) Peabody, D. S. *Archives of Biochemistry and Biophysics* **1997**, *347*, 85–92.

(26) Hooker, J. M.; Esser-Kahn, A. P.; Francis, M. B. *Journal of the American Chemical Society* **2006**, *128*, 15558–15559.

(27) Behrens, C. R.; Hooker, J. M.; Obermeyer, A. C.; Romanini, D. W.; Katz, E. M.; Francis, M. B. *Journal of the American Chemical Society* **2011**, *133*, 16398–16401.

(28) Ni, C.-Z.; Syed, R.; Kodandapani, R.; Wickersham, J.; Peabody, D. S.; Ely, K. R. *Structure* **1995**, *3*, 255–263.

(29) Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S. *Proceedings of the National Academy of Sciences* **1989**, *86*, 8247–8251.

(30) Maza, J. C.; Bader, D. L. V.; Xiao, L.; Marmelstein, A. M.; Brauer, D. D.; ElSohly, A. M.; Smith, M. J.; Krska, S. W.; Parish, C. A.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 3885–3892.

(31) ElSohly, A. M.; MacDonald, J. I.; Hentzen, N. B.; Aanei, I. L.; El Muslemany, K. M.; Francis, M. B. *Journal of the American Chemical Society* **2017**, *139*, 3767–3773.

(32) Furst, A. L.; Smith, M. J.; Lee, M. C.; Francis, M. B. *ACS Central Science* **2018**, *4*, 880–884.

(33) Obermeyer, A. C.; Jarman, J. B.; Netirojjanakul, C.; El Muslemany, K.; Francis, M. B. *Angewandte Chemie International Edition* **2014**, *53*, 1057–1061.

(34) Reddy Chichili, V. P.; Kumar, V.; Sivaraman, J. *Protein Science* **2013**, *22*, 153–167.

(35) Klein, J. S.; Jiang, S.; Galimidi, R. P.; Keeffe, J. R.; Bjorkman, P. J. *Protein Engineering Design and Selection* **2014**, *27*, 325–330.

(36) Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Velu, P. D.; Garcia-Moreno E., B. *Proceedings of the National Academy of Sciences* **2010**, *107*, 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. *Biophysical Journal* **2000**, *79*, 1610–1620.

(38) Tong, G. J.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B. *Journal of the American Chemical Society* **2009**, *131*, 11174–11178.

(39) Stephanopoulos, N.; Tong, G. J.; Hsiao, S. C.; Francis, M. B. *ACS Nano* **2010**, *4*, 6014–6020.

(40) Aanei, I. L.; Francis, M. B. In *Virus-Derived Nanoparticles for Advanced Technologies*, Wege, C., Lomonossoff, G. P., Eds., Series Title: Methods in Molecular Biology; 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. *Nano Letters* **2016**, *16*, 5944–5950.

(42) Gaumet, M.; Vargas, A.; Gurny, R.; Delie, F. *European Journal of Pharmaceutics and Biopharmaceutics* **2008**, *69*, 1–9.

(43) Hietpas, R. T.; Jensen, J. D.; Bolon, D. N. A. *Proceedings of the National Academy of Sciences* **2011**, *108*, 7896–7901.

(44) Pinto, F. L.; Thapper, A.; Sontheim, W.; Lindblad, P. *BMC Molecular Biology* **2009**, *10*, 79.

(45) Raynal, L. C.; Carpousis, A. J. *Molecular Microbiology* **1999**, *32*, 765–775.

(46) Kushner, S. R. *RNA* **2015**, *21*, 673–674.

(47) Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. *PLoS ONE* **2009**, *4*, ed. by Peccoud, J.

(48) Wang, W.; Malcolm, B. A. *BioTechniques* **1999**, *26*, 680–682.

(49) Bolger, A. M.; Lohse, M.; Usadel, B. *Bioinformatics* **2014**, *30*, 2114–2120.

(50) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup *Bioinformatics* **2009**, *25*, 2078–2079.

Chapter 3

Comprehensive Fitness Landscape of a Multi-Geometry Protein Capsid Informs Machine Learning Models of Assembly

3.1 Abstract

Virus-like particles (VLPs) are non-infections viral-derived nanomaterials poised for biotechnological applications due to their well-defined, modular self-assembling architecture. Although progress has been made in understanding the complex effects that mutations may have on VLPs, nuanced understanding of the influence particle mutability has on quaternary structure has yet to be achieved. Here, we generate and compare the apparent fitness landscapes of two capsid geometries (T=3 and T=1 icosahedral) of the bacteriophage MS2 VLP. We find significant shifts in mutability at the symmetry interfaces of the T=1 capsid when compared to the wildtype T=3 assembly. Furthermore, we use the generated landscapes to benchmark the performance of *in silico* mutational scanning tools in capturing the effect of missense mutation on complex particle assembly. Finding that predicted stability effects correlated relatively poorly with assembly phenotype, we used a combination of *de novo* features in tandem with *in silico* results to train machine learning algorithms for the classification of variant effects on assembly. Our findings not only reveal ways that assembly geometry affects the mutable landscape of a self-assembled particle, but also establish a template for the generation of predictive mutational models of self-assembled capsids using minimal empirical training data.

3.2 Introduction

Virus-like particles (VLPs) have emerged as promising scaffolds for many applications in biotechnology, including vaccine development[1–3], targeted drug delivery[4,5], and nanoreactor production[6]. These well-defined, highly symmetrical closed-shell structures self-assemble from discrete protein building blocks that can be engineered to tune the physical and chemical features of the assembly[7]. While great strides have been made in *de novo* design of self-assembling VLP mimetics[8–13], predicting the effects of mutations to the subunit building blocks of a nanocage remains challenging because even small changes can disrupt or drastically alter the complex network of interactions that drive assembly.[14–16]

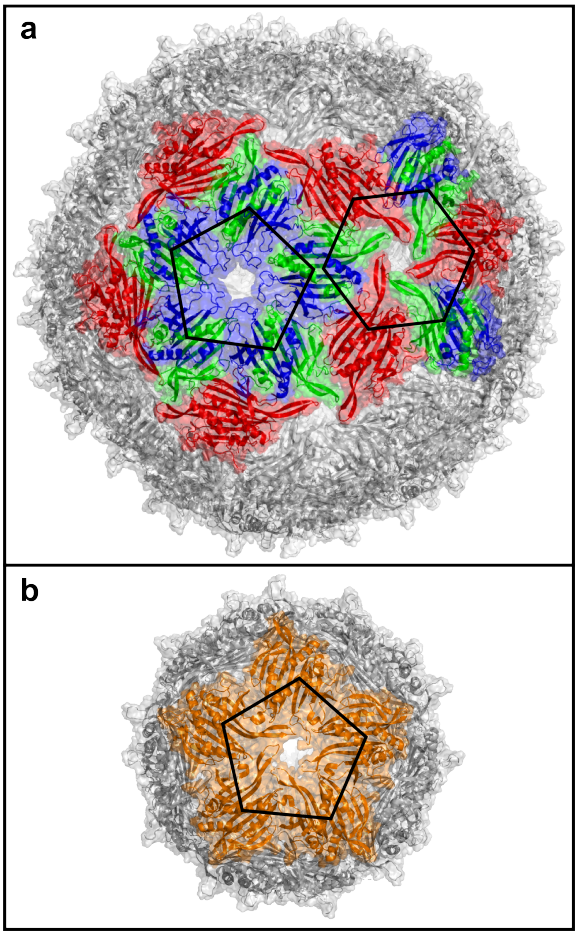

Figure 3.1: Bacteriophage MS2 virus-like particle assemblies. (a) MS2[WT] T=3 icosahedral capsid is shown (PDB ID: 2MS2). C/C and A/B dimer conformers are shown in red and green/blue, respectively. (b) MS2[S37P] T=1 icosahedral capsid (PDB ID: 4ZOR). All dimers in this capsid geometry are equivalent. Assembly capsomers are outlined in black to show constituent symmetry interfaces of each particle geometry.

Systematic study of the effects of mutations on a given protein function has recently been made possible by the advent of deep-mutational scanning, a method that uses next-generation sequencing to assess $> 10^5$ protein variants in a single experiment.[17,18] This technique has previously been applied to self-assembling viral structures such as AAV, HIV,[19,20] influenza,[21] and polio[22] in order to generate fitness landscapes that describe the effects of mutations on viral infectivity. In order to probe the impact of mutations on assembly interactions more directly, we recently developed a strategy called SyMAPS (Systematic Mutagenesis and Assembled Particle Selection), which employs capsid self-assembly as a fitness readout for deep mutational scanning.[23] This method enables a quantitative, systematic understanding of how mutations to the subunits of a closed-shell particle affect the final

assembly state without relying on a selection that bundles many fitness criteria (e.g. infectivity requires successful viral attachment, replication, and assembly).[24] SyMAPS has been successfully utilized to quantify the effects of single mutants, epistatic interactions, loop insertions, and peptide extensions in bacteriophage MS2 VLPs, enabling tuned particle thermostability, acid lability, and chemical reactivity.[25–27]

Here, we employ the SyMAPS approach to generate the apparent fitness landscape of a non-native assembly geometry of an icosahedral VLP. Prior work uncovered a point mutant (MS2[S37P]) in the coat protein of the well-studied MS2 bacteriophage VLP that results in a shift in quaternary assembly of the particle.[16] This shift from a T=3 to T=1 icosahedra dramatically reorganizes the capsid assembly while maintaining nearly identical primary, secondary, and tertiary structures of the constituent coat protein subunits (Figure 3.1).

By generating a 1-dimensional AFL of MS2[S37P] we were able to uncover unprecedented insight into the underlying principles of assembled structure mutability — we find that core residues and residues at the dimer interface of both VLP geometries share similar mutabilities on average. On the other hand, residues mediating interface between dimer subunits or at multiple contact interfaces are significantly more mutable in the T=1 variant. Interestingly, this holds true for residues both at the quasi-6-fold symmetry interface, which only occurs in the wildtype VLP, and the 5-fold symmetry interface, which is present in both capsid phenotypes. We then use the comprehensive variant data to assess computational methods for determining changes to the folding free energy of a protein when point mutants are introduced, and find that correlation between the predicted folding $\Delta\Delta G$ and the experimental effect on assembly viability is relatively low. Lastly, we combine computational prediction results with experimental fitness landscape results to train machine learning models of assembly state classification (Figure 3.2). We find that minimal computational and experimental input may be used to generate well-performing classifiers for mutational effects on VLP assembly state.

3.3 Results

3.3.1 Deep Mutational Scan of a Non-Native Virus-like Particle Assembly

In order to explore the effect of quaternary structure on the mutable landscape of self-assembling VLPs, we generated a deep mutational scanning library of the MS2 coat protein with a fixed S37P backbone point mutation. This backbone, MS2[S37P], shifts the MS2 VLP assembly from a 27 nm, T=3 icosahedral geometry to a 17

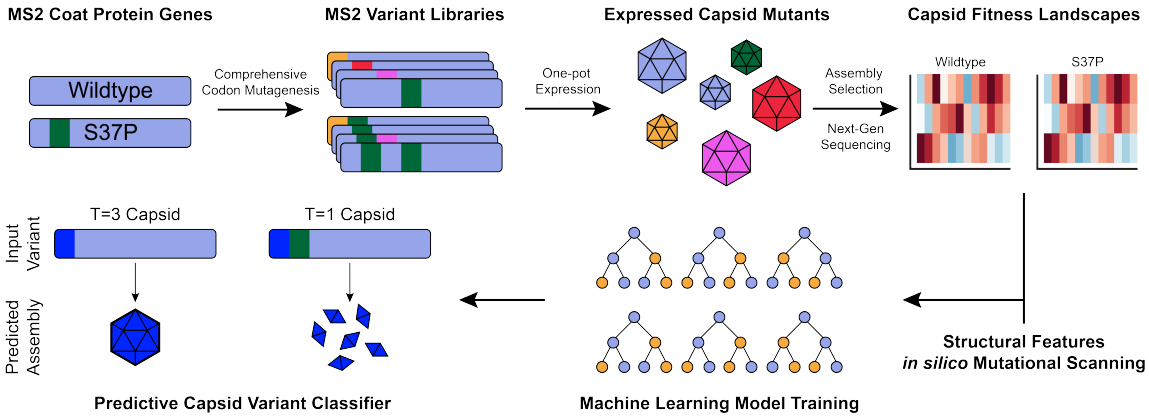

Figure 3.2: Workflow for generation of capsid apparent fitness landscapes and subsequent assembly classifier generation.

nm, T=1 icosahedral assembly.[16] The SyMAPS platform was used to construct triplicate libraries of the MS2[S37P], as well as the wildtype (MS2[WT]) backbone. Next-generation sequencing of the mutagenized plasmid libraries resulted in excellent coverage of the mutational space, with 94% and 96% of possible variants accounted for in the MS2[S37P] and MS2[WT] libraries respectively. In both cases the majority of variants missing from the library were at the first and last position of the coat protein, which is the result of the first 5 base pairs of the Illumina PE300 MiSeq reads falling below our quality score threshold. The mutagenized libraries were expressed and subjected to size-exclusion chromatography to distinguish between assembling T=1 capsids, assembling T=3 capsids, and non-assembling variants. Assembled capsids encapsidate their encoding mRNA, thus the genotype of each well-formed particle was retained for sequencing. A high-temperature challenge was also introduced to remove capsids with compromised thermostability. Following coat protein expression and capsid enrichment, 72% of MS2[S37P] variants and 84% of MS2[WT] variants were recovered via sequencing and assigned an Apparent Fitness Score (AFS) derived from the $\log_10$ change in DNA read abundance between the input plasmid library and assembled capsid library.

Correlation of AFS between replicates of the MS2[WT] was relatively high ($r^2 = 0.70 - 0.83$), indicating strong reproducibility and minimal distortions from biological noise (Figure 3.3a). Interreplicate correlation in the MS2[S37P] libraries was somewhat lower ($r^2 = 0.32 - 0.56$), possibly due to a lower signal-to-noise ratio, as T=1 capsids encapsidate less genetic material than the wildtype capsid (Figure 3.3b).[28] There was a low degree of correlation between the MS2[WT] and MS2[S37P] libraries, meaning that there are both shared mutability preferences and divergences in mutability

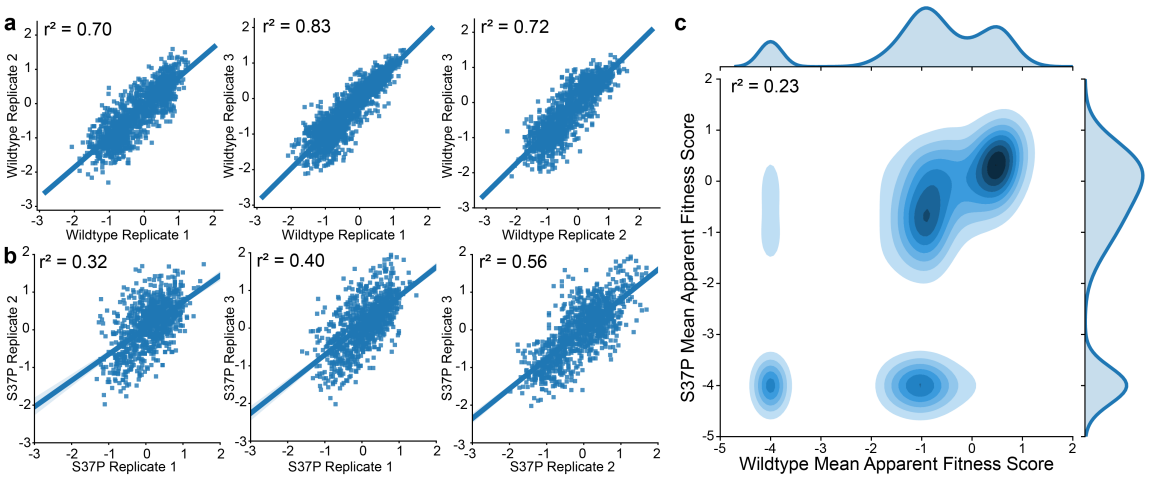

Figure 3.3: Capsid apparent fitness landscape replicate correlation. Correlations between the three biological replicates of the (a) MS2[WT] and (b) MS2[S37P] libraries are shown. (c) Correlation between the mean apparent fitness scores of each variant in the MS2[S37P] and MS2[WT] libraries. Kernel density estimations of apparent fitness score distributions are shown outside of the axes.

between the two capsid geometries. A notable difference in AFS distribution of the libraries is derived from their tolerance to deleterious mutations. A large portion of poorly performing variants in the MS2[WT] library (AFS = –2.0 – 0.0) drop to the lowest possible fitness score in the MS2[S37P] library (AFS = -4) (Figure 3.3c). This suggests that the MS2[WT] capsid may retain some level of assembly competency in response to deleterious mutations, while the MS2[S37P] capsid assembly is completely ablated in such cases. This may however also result from lower mRNA packaging per T=1 capsid, thus dropping deleterious mutations below the detection limit.

To investigate how the mutational landscape of the T=1 assembly phenotype differs from the wildtype T=3 icosahedron further, Shannon Entropy calculations were performed for each position of the capsid monomer backbone.[29] This measure of diversity at a given residue has previously been used to generate a Mutability Index (MI) of each backbone position,[23] which may be used to determine engineering "hotspots" in the capsid coat protein as well as conserved residues likely to mediate key interactions for successful protein folding and particle self-assembly. The difference in MI between the MS2[WT] and MS2[S37P] capsid monomers is shown in (Figure 3.4a). The majority of the coat protein backbone shows a relatively low difference in MI between the wildtype and S37P libraries, indicating that many of the essential folding and assembly interactions are preserved. Regions of defined secondary structure elements, such as the central portion of the alpha helix spanning residues 105-111, show low mutability in both the T=1 and T=3 capsids. Likewise, the well-studied

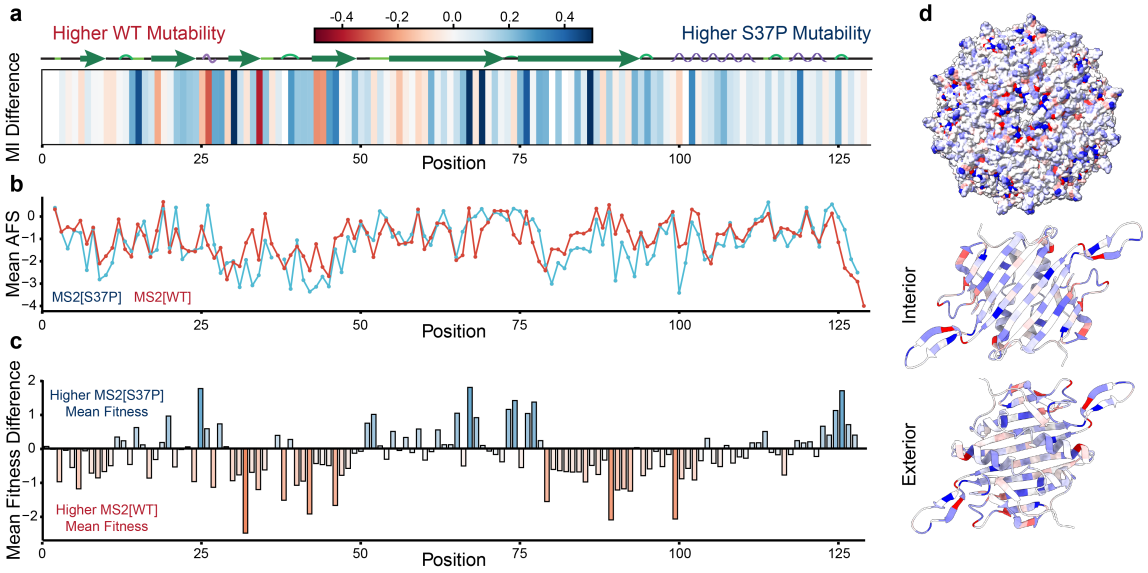

Figure 3.4: Global mutability trends in MS2 capsid assemblies. (a) Per-residue differences in the mutability index between MS2[WT] and MS2[S37P] capsid backbones. (b) Mean apparent fitness scores (AFS) of all mutations at a given backbone position for both mutagenic libraries. (c) Differences in mean AFS for both capsid libraries. (d) Differrences in mutability index mapped to the MS2[S37P] capsid (PDB ID: 4ZOR) as well as its constituent dimer using ChimeraX.[32] In all cases blue represents a higher value in MS2[S37P] and red represents a higher value in MS2[WT].

FG loop region[30,31] (residues 71-76) of the coat protein retains high mutability in both capsid assemblies. However, individual positions with a significant shift in MI (< -0.2 or > 0.2) span the length of the backbone.Mapping of MI to the capsid structures reveals the variety of secondary structural motifs with shifts in mutability between the two capsid phenotypes — residues in beta sheets, alpha helices, turns, and disordered loops show unique preferences in the MS2[S37P] capsid (Figure 3.4d).

Surprisingly, it is much more common that positions become more permissive to mutations in the T=1 capsid than become more strongly conserved, which is reflected in a higher mean MI for the MS2[S37P] than MS2[WT] (-0.24 and -0.33, respectively). This may seem counterintuitive, given that the mean AFS of the MS2[S37P] library was lower than that of the MS2[WT] library (-1.14 and -0.92, respectively). However, the lower mean AFS of the MS2[S37P] library results from the abundance of mutants at the minimum -4.0 AFS.

Disaggregation of AFS distribution by structural context (core, interface, or surface) reveals that the capsid AFLs differ most at subunit interfaces (Figure 3.5a). The MS2[WT] capsid shows a hierarchy whereby core residues are most conserved, followed by subunit interface residues, with surface residues (at either the interior or

exterior surface) being most permissive to mutation. These preferences are in line with previous trends determined from multiple sequence alignment of T=3 VLPs.[33] MS2[S37P] capsids also possess a highly conserved core; however the mutability indicies of the interface and surface residues are statistically similar.

Closer examination of contact interface context reveals that interdimer contact residues have similar mutabilities in both MS2[WT] and MS2[S37P]. Meanwhile, residues along intradimer contacts or residues that participate in multiple interface contacts are more permissive to mutation in the MS2[S37P] capsid library.

The symmetry elements of MS2[WT] and MS2[S37P] were next explored to understand better differences in the AFL of both capsid phenotypes. Casper and Klug's theory of quasi-equivalence dictates that T=3 capsids will form from a mixture of pentameric and hexameric subunits at 5-fold and quasi-6-fold axes of symmetry, while T=1 icosahedral capsids form only from pentamer subunits at 5-fold symmetry axes.[34] Prior work has also isolated two key intermediates in the MS2[WT] assembly pathway: a 12mer subunit at the quasi-6-fold symmetry axis (6CP2) and a 20mer subunit at the 5-fold symmetry axis (10CP2).[35] While assembly intermediates of MS2[S37P] have not been directly observed, the 10CP2 subunit is a viable assembly intermediate, while the 6CP2 subunit presumably is not due to the geometry of the T=1 capsid. As such, we hypothesized that residues involved in the pentameric interfaces of the MS2 capsid would be similarly conserved in MS2[WT] and MS2[S37P], while the hexameric interfaces would be more flexible in MS2[S37P] as they do not mediate productive intermediates in the assembly

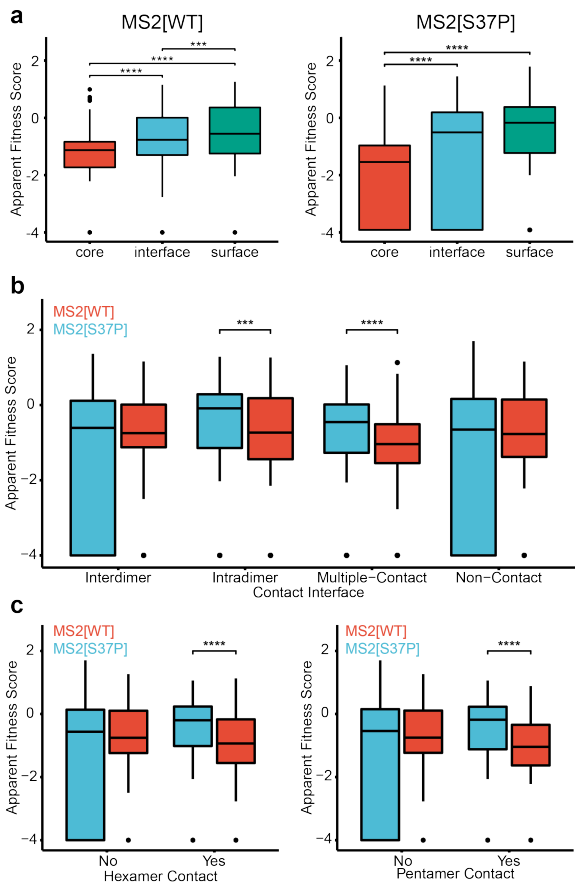

Figure 3.5: Capsid mutability trends based on structural context. (a) Boxplots of MS2[WT] and MS2[S37P] apparent fitness score (AFS) distribution by residue context. (b) Boxplot of AFS distribution by interface context, further separated in (c) by symmetry contact. MS2[WT] and MS2[S37P distributions are shown in red and blue, respectively. *** indicates $p < 0.001$ and **** indicates $p < 0.0001$.

pathway. Surprisingly, we find that both hexameric and pentameric interface residues are significantly more flexible in MS2[S37P]. This may suggest that the assembly pathway of the T=1 capsid may be more reliable than the T=3 capsid once initial folding and dimerization has occurred. There are indeed more potential misassembly pathways in T=3 icosehedra due to their larger number of unique interfaces,[33] although the relationship between the number of improper assembly pathways and the mutational flexibility of VLPs has not been previously established. These trends and observations demonstrate how comprehensive fitness landscapes may reveal mutability preferences that challenge expectations from sequence conservation in nature or known assembly pathways.

3.3.2 Benchmarking Computational Residue Scanning Methods for Variant Fitness Prediction

We used these datasets to benchmark computational tools that calculate predicted changes in folding free energy of protein variants. *In silico* deep mutational scans of the MS2[WT] and MS2[S37P] coat proteins were performed using BioLuminate,[36] the DynaMut2 web server[37] and the PoPMuSiC 2.1 web server[38]. As modeling the monomer subunit only accounts for one potential mode of assembly failure (protein misfolding), computational scans of the dimeric, 6CP2, and 10CP2 subunits of the assembled capsid were performed to achieve better coverage of the assembly deficits related to changes in intra-dimer and inter-dimer affinity. The 6CP2 capsomer was not modeled for MS2[S37P] as it is not likely to be formed in the assembly pathway. Correlation between computationally predicted $\Delta\Delta$G of coat protein variants and experimentally determined AFS was low across all tested methods and modeled subunits. This suggests that calculated fitness values of the capsid assemblies depends on a more complex set of factors than are captured by the endpoint Gibbs free energy of its constituent subunits.

Although the absolute AFS is not well correlated with computational $\Delta\Delta$G scoring, we hypothesized that these methods may be successful in a binary classification of variant assembly competency. Previously established AFS thresholds for assembly were used to label variant phenotypes (AFS > 0.2 = "assembling", AFS < -0.2 = "non-assembling"). Predictions were classified as "assembling" if $\Delta\Delta G <= 0$ (i.e. computation predicts the mutation is neutral or stabilizing) and "non-assembling" if $\Delta\Delta G > 0$ (i.e. computation predicts the mutant is destabilizing). Binary classification results are shown in Figure 3.6. The accuracy of the tested classification methods was modest, ranging from 0.62-0.73. There were no large differences in accuracy between the three computational scanning methods. Gratifyingly, models of assembly

subunits improved classification accuracy over models of the coat protein monomer for both MS2[WT] and MS2[S37P] in all three computational methods. Accuracy can, however, provide an overly optimistic measure of performance in cases where the tested dataset has a class imbalance (i.e. there are many more non-assembling mutations than assembling mutations).[39] Thus, the Matthews correlation coefficient (MCC) was reported for each classification case, as this metric is a more reliable measure for classification of imbalanced datasets.[40,41] MCC was calculated with values derived from Table 3.1 using equation 3.1.

Table 3.1: Confusion matrix describing performance of computational scanning classification.

		Fitness Landscape	
		Assembling	Non-Assembling
Computational Scan	Assembling	TP	FP
	Non-Assembling	FN	TN

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP}) \times (\mathrm{TP} + \mathrm{TN}) \times (\mathrm{TN} + \mathrm{FP}) \times (\mathrm{TN} + \mathrm{FN})}} \tag{3.1}$$

MCC ranges from -1 to +1, where MCC = +1 represents a perfect classifier, MCC = -1 represents perfect misclassification, and MCC = 0 represents the expected value of a coin toss classification. Evaluation by MCC reveals that the physics-based prediction via Bioluminate scored much higher than the other two methods. While the overall accuracy of each method was relatively similar, Bioluminate correctly predicted the minority class (i.e., "assembling" mutation) in more cases than DynaMut2 and PoPMuSiC 2.1. Subsequent tuning efforts focused on Bioluminate's computational scanning results, since predicting assembly-competent mutations is more valuable for the production of capsids with engineered properties or new sequence motifs. In order to improve classifier performance, the $\Delta\Delta$G threshold for designation of a mutation as "assembling" or "non-assembling" was scanned to optimize the MCC scores (Figure 3.6c). Threshold scanning revealed that a higher $\Delta\Delta$G cutoff ($\Delta\Delta$G = +5-8 kcal/mol) resulted in a higher MCC score for capsid phenotype classification. This indicates that MS2 capsid assembly can readily occur in spite of moderate increases to the folding free energy of the capsid subunits. Though tuning the cutoff threshold improved model performance, we sought to improve predictive power further by employing machine learning to capitalize on the rich feature set derived from these computational methods.

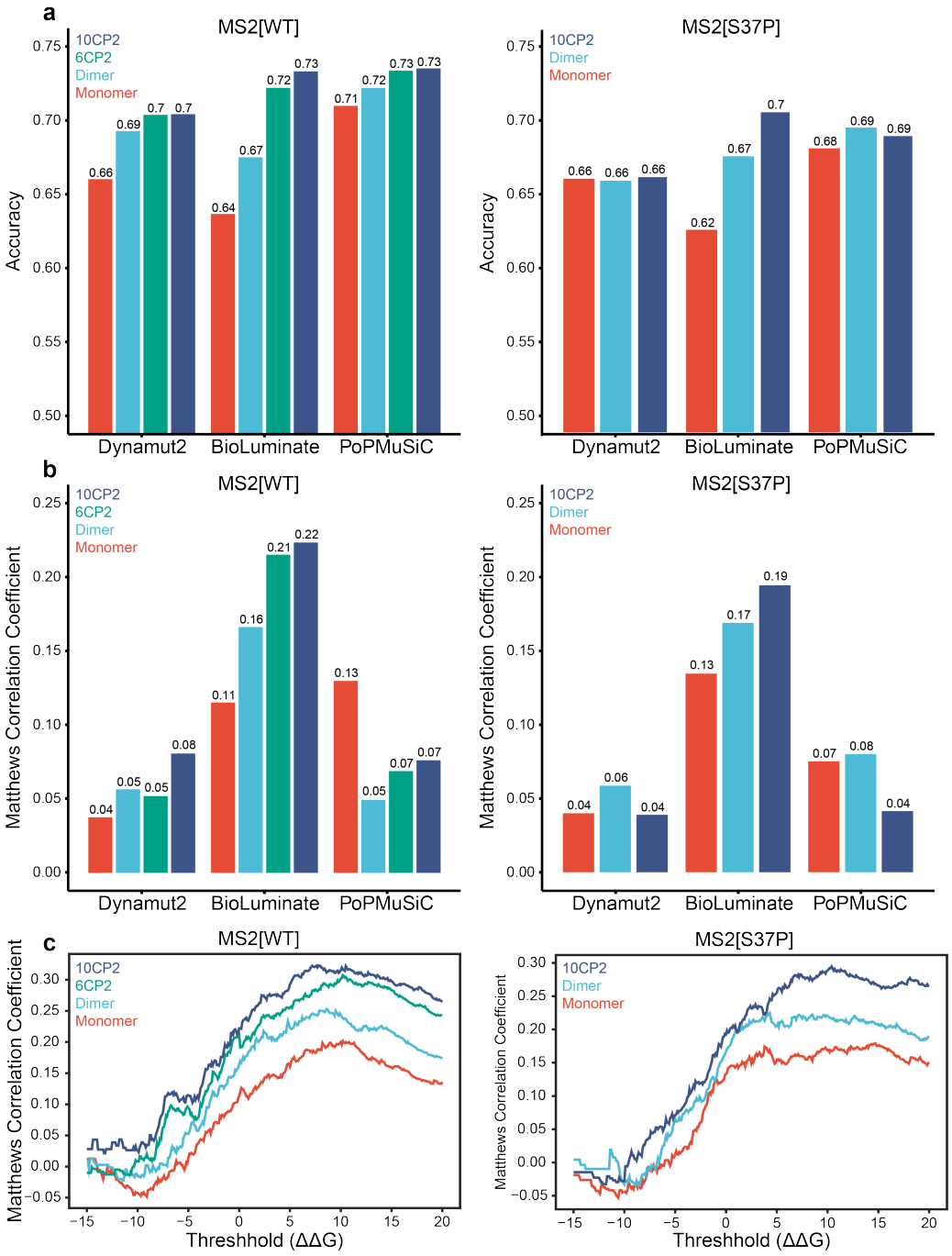

Figure 3.6: Computational residue scanning classification performance. The (a) accuracies and (b) Matthews correlation coefficients of three computational residue scanning methods are shown for various subunit models of the MS2[WT] and MS2[S37P] capsid. Calculated $\Delta\Delta G$ values were compared to experimental apparent fitness scores. (c) Matthews correlation coefficients are shown as a function of classification $\Delta\Delta G$ threshhold.

3.3.3 Development of Supervised Learning Models for Capsid Assembly Classification

A total of 106 features for the MS2[WT] set and 81 features for the MS2[S37P] set were prepared for each amino acid variant represented in the experimental fitness landscapes of each capsid. Initial models were prepared via four supervised learning algorithms (K-Nearest Neighbor,[42] logistic regression,[43] random forest,[44] and XGBoost[45]) These models were trained with default hyperparameters using a stratified 75/25 train/test split of the landscape data. Models were evaluated by 10-fold cross validation, and binary classifier performance was measured using the MCC scores (Figure 3.7). The random forest and XGBoost models outperformed the K-nearest neighbor and logistic regression models in both training sets, as measured by MCC. Bayesian Optimization was subsequently implemented to tune the hyperparameters used for XGBoost or random forest model training.[46] The resultant tuned models showed strong predictive performance on the unseen test set of landscape data, with the MS2[WT] model slightly outperforming the MS2[S37P] model (accuracy = 0.86, MCC = 0.58 and accuracy = 0.82, MCC = 0.52, respectively).

Ensemble tree-based algorithms, such as XGBoost and random forest, furthermore enable calculation of the importance of each feature with respect to prediction of the target variable, thus offering interpretability of the final MS2[WT] and MS2[S37P] models (Figure 3.8).[47,48]

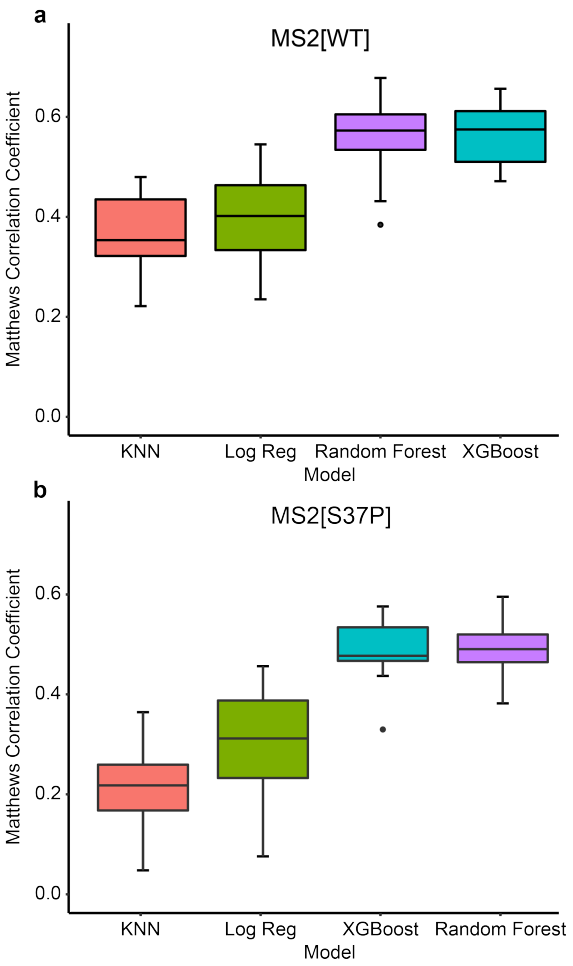

Figure 3.7: Initial machine learning model classification performance. Four machine learning algorithms were benchmarked for predicting the assembly state of (a) MS2[WT] or (b) MS2[S37P] capsid variants. Boxplots of the resultant Matthews correlation coefficients from 10-fold cross validation of models are shown.

Across all four models, the most important feature for prediction was the ΔStability of the 10CP2 subunit calculated using Bioluminate. Many *de novo* features, such as residue solvent-accessible surface area, hydropathy, potential energy, and internal energy, contributed significantly to the final models. In the MS2[WT] models, ΔStability of the 6CP2 subunit was much less important than the 10CP2 subunit, which may indicate either that the endpoint stability of the 6CP2 complex does not capture inhibition of the subunit assembly pathway, or that this complex is more resilient to perturbation than the 10CP2 complex. Interestingly, no intersubunit affinity terms ranked highly in importance, while intradimer affinity ranked in the top ten of 2 out of 4 of the final models.

We next aimed to establish the minimal input data set necessary for model training, as reduction of the training set size or initial computational modeling steps may enable accurate construction of assembled capsid fitness landscapes with reduced experimental and computational cost. Training features were reduced to include only *de novo* amino acid features as well as the most predictive feature, ΔStability derived from the 10CP2 Bioluminate modeling. A random forest model was trained on randomly generated subsets of the fitness landscape (1-2000 training samples) and evaluated by predicting the remaining landscape assembly states. Learning curves of both models displaying the gain in MCC and accuracy as a function of training samples are

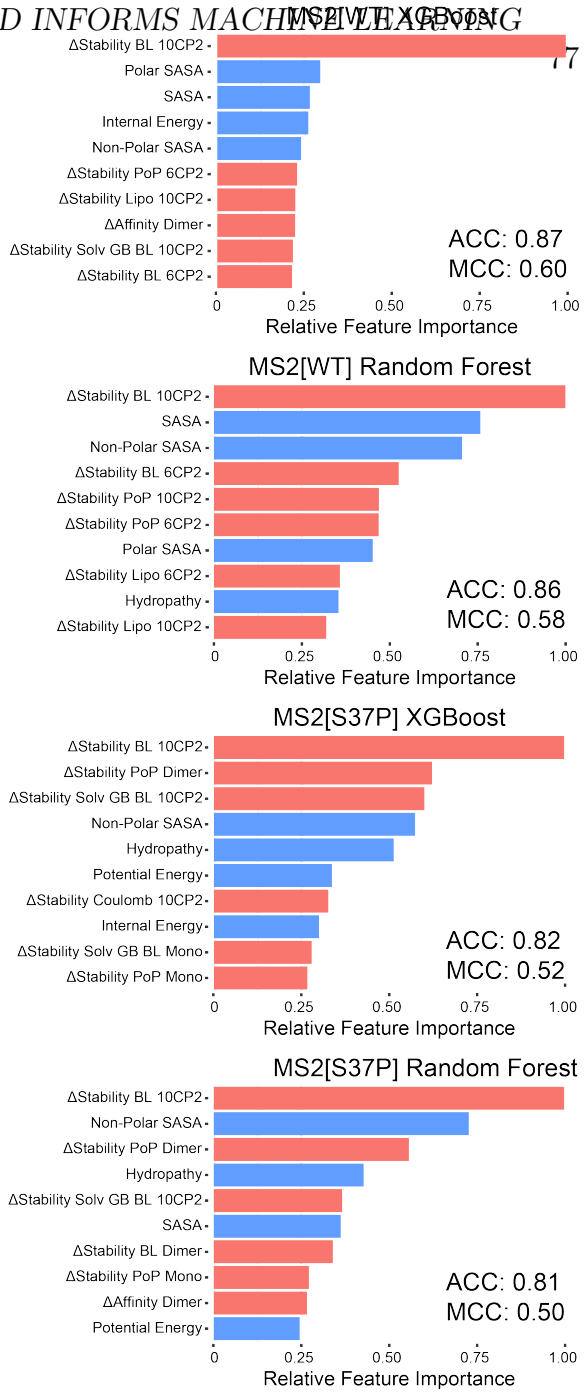

Figure 3.8: Performance of XGBoost and random forest models of (a), (b) MS2[WT] and (c), (d) MS2[S37P] assembly classification are shown. The top 10 most informative features for classification are shown. Model-derived and *de novo* features are shown in red and blue, respectively.

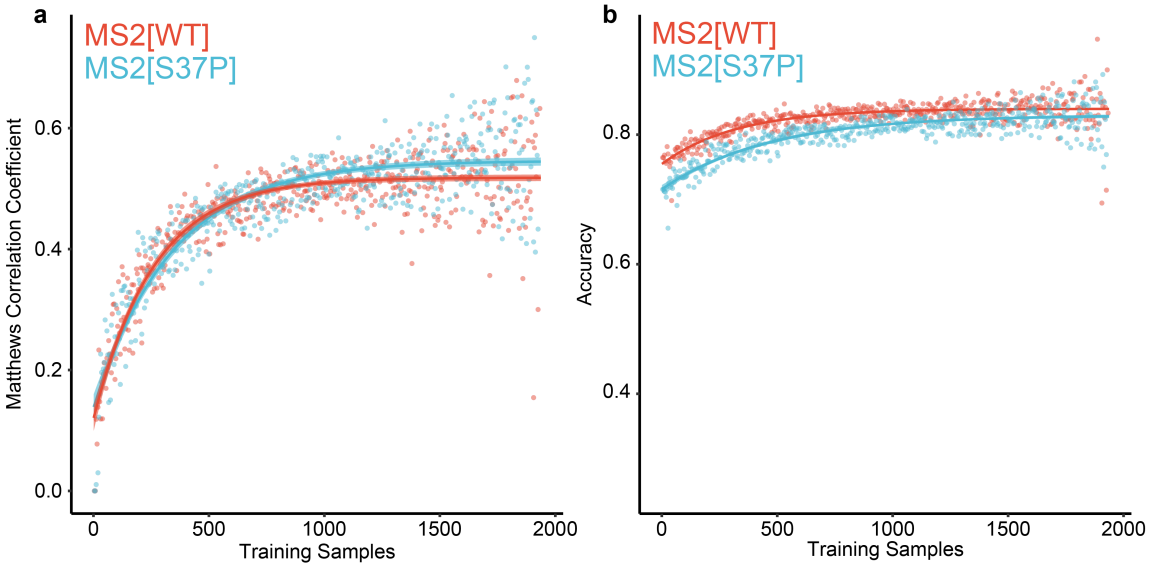

Figure 3.9: Determination of minimized input model performance. (a) MCC gain plot and (b) accuracy gain plot of both capsid models as functions of training set size. Models were generated using only ΔStability BL 10CP2 and *de novo* parameters for training. Random subsets of the fitness landscape data were used to simulate low-sample training information. MS2[WT] and MS2[S37P] learning curves are shown in red and blue, respectively.

shown in Figure 3.9. Interestingly, the model achieves reasonable accuracy after very few samples, and only marginally improves as the sample numbers increase. Viewing MCC gain, however, displays a sharp improvement that plateaus after roughly 500 samples. This suggests that small sublibraries of self-assembled protein capsids may enable reasonably accurate predictions of capsid fitness landscapes, likely by learning that the majority of mutations inhibit capsid assembly. To obtain predictions that correctly classify assembly-competent mutants, a moderate sampling of the full mutational space of the VLP (about 20% of variants) will be needed. The extent that this sampling ratio may be generalized to VLPs of different levels of homology to MS2 remains to be determined.

3.4 Conclusion

Here, we have characterized the apparent fitness landscape of a non-native (T=1 icosahedral) VLP assembly geometry. Comparison to the apparent fitness landscape of the wild type (T=3 icosahedral) particle yielded insight into the influences quaternary

structure has on the mutability of a self-assembling material. Residues at the core of the monomer subunit, as well as residues mediating the dimer interfaces of the capsid showed similar mutability preferences in both capsid libraries. Surprisingly, residues along both the five-fold and quasi-six-fold symmetry interfaces of the particle were significantly more mutable in the T=1 capsid than the wildtype assembly, suggesting that the T=1 capsid may be a more versatile platform for engineering efforts. Further, the deep mutational scanning data produced were used to benchmark the performance of various computational residue scanning tools for estimating the effects of missense mutations on protein folding energetics. While correlations between calculated $\Delta\Delta G$ values and measured assembly competency were relatively low, modeling known assembly intermediates using BioLuminate's MM/GBSA method offered the best overall performance. Finally, machine learning models were trained using *de novo* structural features in tandem with computational modeling data to produce predictive assembly state classifiers for both capsid assemblies. The relationship between input training size and classifier performance demonstrates that relatively small input datasets can be used to produce predictive classifiers for self-assembly with strong performance. As such, this study may represent a generalizable strategy for semi-rational engineering and systematic mutation of proteinaceous self-assembled nanomaterials.

3.5 Experimental

3.5.1 Strains

All library experiments were conducted with MegaX DH10B *E. coli* electrocompetent cells (ThermoFisher Scientific, Cat# C640003). Overnight growths from a single colony were incubated for 18 h at 37 °C with shaking at 200 RPM in LB-Miller media (Fisher Scientific, Cat# BP1426-2) with chloramphenicol at 32 mg/L. Expressions were subcultured 1:100 into 2xYT media (Teknova, Cat# Y0210) with 32 mg/L chloramphenicol, allowed to grow to an OD600 of 0.6, then induced with 0.1% arabinose. Expressions continued overnight at 37 °C with shaking at 200 RPM.

3.5.2 FPLC SEC

MS2 CP libraries and select individual variants were purified on an Akta Pure 25 L Fast Protein Liquid Chromatography (FPLC) system with a HiPrep Sephacryl S-500 HR column (GE Healthcare Life Sciences, Cat# 28935607) Size Exclusion

Chromatography (SEC) column via isocratic flow with 10 mM phosphate buffer at pH 7.2 with 200 mM sodium chloride and 2 mM sodium azide.

3.5.3 HPLC SEC

Variants or libraries were analyzed on an Agilent 1290 Infinity HPLC with an Agilent Bio SEC-5 column (5 um, 2000A, 7.8x300mm) with isocratic flow of 10 mM phosphate buffer at pH 7.2 with 200 mM sodium chloride and 2 mM sodium azide. Fractions were harvested at 11.2 min, or the characteristic elution time for wild-type MS2. Harvested VLPs were then subjected to RNA extraction and high-throughput sequencing sample preparation.

3.5.4 Library generation

Deep mutational scanning libraries of both capsids were performed as previously described via SyMAPS.[23] SyMAPS uses the EMPIRIC cloning developed in the Bolon lab.[49] In EMPIRIC cloning, a plasmid contains a self-encoded removable fragment (SERF) flanked by inverted BsaI recognition sites. Thus, BsaI digestion simultaneously removes the SERF and BsaI sites. Five previously generated entry vectors each designed to accept a 26 codon region of the MS2 coat protein as an insert were used to prepare both the MS2[WT] and MS2[S37P] libraries. Four of the entry vectors bearing the position S37 codon were mutated to S37P via Quikchange mutagenesis for the MS2[S37P] library. Single-stranded DNA primers spanning each 26-codon region with an NNK codon at each position were pooled and diluted to a final concentration of 50 ng/uL. Touchdown PCR was used to amplify the mutagenic inserts and the double-stranded DNA was purified by PCR clean-up (Promega, Cat# A9282). Cleaned dsDNA inserts were diluted to 5 ng/µL, then cloned into the entry vector using Golden Gate cloning. Ligated plasmids were incubated on desalting membranes (Millipore Sigma, Cat# VSWP02500) for 20 min, followed by transformation into electrocompetent DH10B E. coli cells (Invitrogen, Cat# 18290015). Following electroporation and recovery, cells were plated onto LB-A containing 32 µg/mL chloramphenicol and grown overnight at 37 °C. Dilutions (100-fold) of each replicate were plated and counted to ensure a minimum of 3x colony-forming units relative to the number of possible variants.

3.5.5 Assembly selection

Colonies were harvested by scraping plates into LB-M and growing the mixture for 2 h. Each library was subcultured 1:100 into 1 L of 2xYT (Teknova, Cat# Y0210)

and grown to an OD of 0.6, then induced with 0.1% arabinose. Variant libraries were expressed overnight at 37 °C. Cells were harvested by centrifugation and lysed by sonication. Libraries were subjected to two rounds of ammonium sulfate precipitation (50% saturation), followed by FPLC size exclusion chromatography purification to select for well-assembled T=3 or T=1 VLPs. Collected fractions were then subjected to 50 °C thermostability challenges in a H_2O bath for 10 min. Precipitated VLPs were pelleted via centrifugation, and assembled VLPs were isolated via semi-preparative HPLC size exclusion chromatography. Fractions containing assembled VLPs (at the characteristic elution time of 6.5 min for MS2[WT] or 8.0 min for MS2[S37P]) were combined, subjected to RNA extraction, barcoded, and identified via high-throughput sequencing.

3.5.6 High-throughput sequencing sample preparation

Plasmid DNA was isolated prior to expressions using Zyppy Plasmid Miniprep Kit (Zymo, Cat# D4036). RNA was extracted from the assembly-selected libraries using previously published protocols.[50] Briefly, homogenization was carried out with TRIzol (Thermo Fisher Cat# 15596026), followed by chloroform addition. The sample was centrifuged to separate into aqueous, interphase, and organic layers. The aqueous layer, containing RNA, was isolated, and the RNA was precipitated with isopropanol and washed with 70% ethanol. RNA was dried and resuspended in RNAse free water. cDNA was synthesized with the Superscript III first strand cDNA synthesis kit from Life (Cat# 18080051, random hexamers primer).

cDNA and plasmids were both barcoded for high-throughput sequencing. Both types of samples were amplified with two rounds of PCR (10 cycles, followed by 8 cycles) to add barcodes and Illumina sequencing handles, following Illumina 16S Metagenomic Sequencing Library Preparation recommendations. Libraries were quantified with Qubit and combined in equal molar ratio. Libraries were analyzed by PE300 MiSeq in collaboration with the UC Davis Sequencing Facilities.

3.5.7 High-throughput sequencing data processing and analysis

Data were trimmed and processed described previously,[23] with minor variation. Briefly, data were trimmed with Trimmomatic[51] with a 4-unit sliding quality window of 20 and a minimum length of 32. Reads were merged with FLASH (fast length adjustment of short reads) with a maximum overlap of 300 base pairs. Reads were sorted and indexed with SAMtools[52] and unmapped reads were filtered with the

Picard function CleanSam. Reads were processed with in-house code to produce various AFLs.

3.5.8 AFL calculations

Cleaned and filtered high-throughput sequencing reads were analyzed using in-house Python code (available at https://github.com/dan-brauer/MS2-Assembly-Classification). Briefly, the start codon and stop codons on either end of the MS2 gene were parsed and the correct gene length was verified. For MS2[S37P], codon encoding for the S37P mutation was also verified. Codons were then translated into amino acids to generate counts for each tripeptide combination

Calculations were repeated for all replicate plasmid pools and assembled selections. Relative abundances were calculated similarly to the previously described protocol.[26] Briefly, the relative abundance of a selection was calculated by dividing the abundance generated from an assembled VLP by the plasmid abundance from its respective biological replicate.

The mean relative abundance across three biological replicates was calculated for the final fitness landscape of each library. All NaN (null) values, which indicate variants that were not identified in the plasmid library were ignored. Scores of zero, which indicate variants that were sequenced in the unselected library but absent in the VLP library, were replaced with an arbitrary score of 0.0001. The log10 relative abundance value for an individual variant in a particular challenge is referred to as its Apparent Fitness Score (AFS).

3.5.9 Computational fitness scanning

Assembly subunits (monomer, dimer, 6CP2, and 10CP2) were constructed in the Maestro molecular modeling environment based on the crystal structures of MS2[WT] (PDB:2MS2) and MS2[S37P] (PDB:4ZOR). Resultant structures were preprocessed and minimized using the OPLS3e forcefield. Computational scanning for minimized subunit structures via DynaMut2 and PoPMuSiC 2.1 was performed using their respective web servers. BioLuminate calculations were performed symmetrically across chains using side-chain prediction with backbone minimization at a 0.00 Å cutoff. Classifier performance was evaluated using in-house Python code (available at https://github.com/dan-brauer/MS2-Assembly-Classification).

3.5.10 Machine learning model training

Machine learning models were trained and assessed using the Tidymodels[53] framework in R (available at https://github.com/dan-brauer/MS2-Assembly-Classification). Continuous training variables were normalized, nominal variables (e.g. amino acid identity) were encoded as binary dummy variables, and zero variance variables were filtered in data preprocessing. Algorithm comparisons (K-Nearest Neighbor, logistic regression, random forest, and XGBoost) were performed using default hyperparameters. Initial random forest and XGBoost model training used 1000 trees. Cross validation (10-fold) was performed on a 75/25 split of the fitness landscaping data. Random forest and XGBoost models were further optimized through 100 rounds of Bayesian hyperparameter tuning, with 10 initial parameter sets and an early stop after 20 consecutive iterations with no improvement. Optimized parameters were used to train final models which were evaluated on the 25% of data held out for final evaluation. Model feature importance was determined using the R 'vip' package.[54]

3.6 Additional Figures

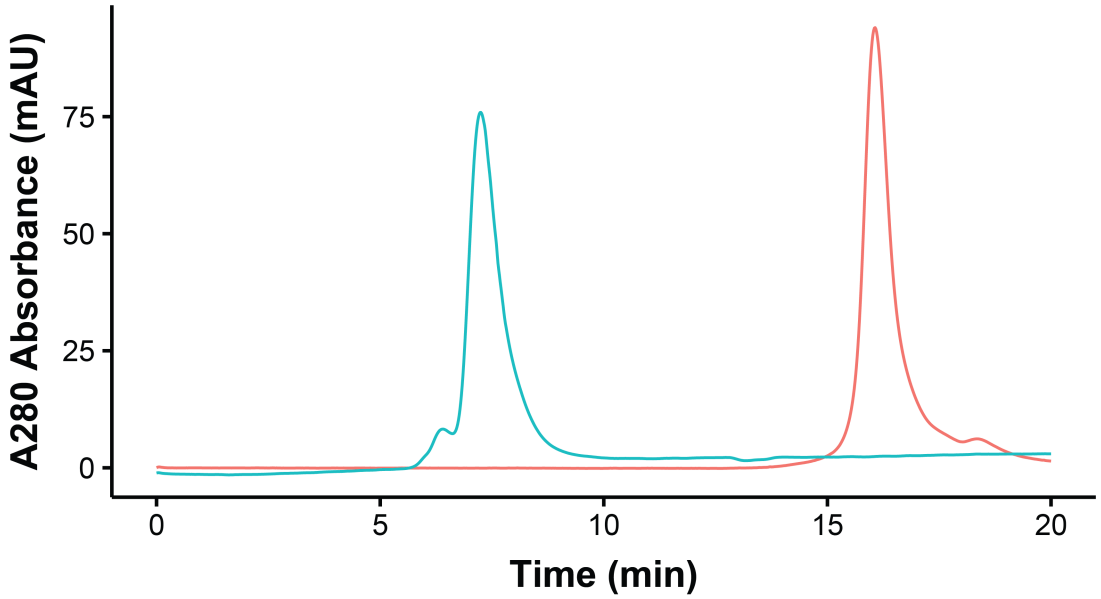

Figure 3.10: Separation of purified MS2[WT] and MS2[S37P] capsids by size exclusion chromatography. MS2[WT] and MS2[S37P are shown in blue and red traces, respectively

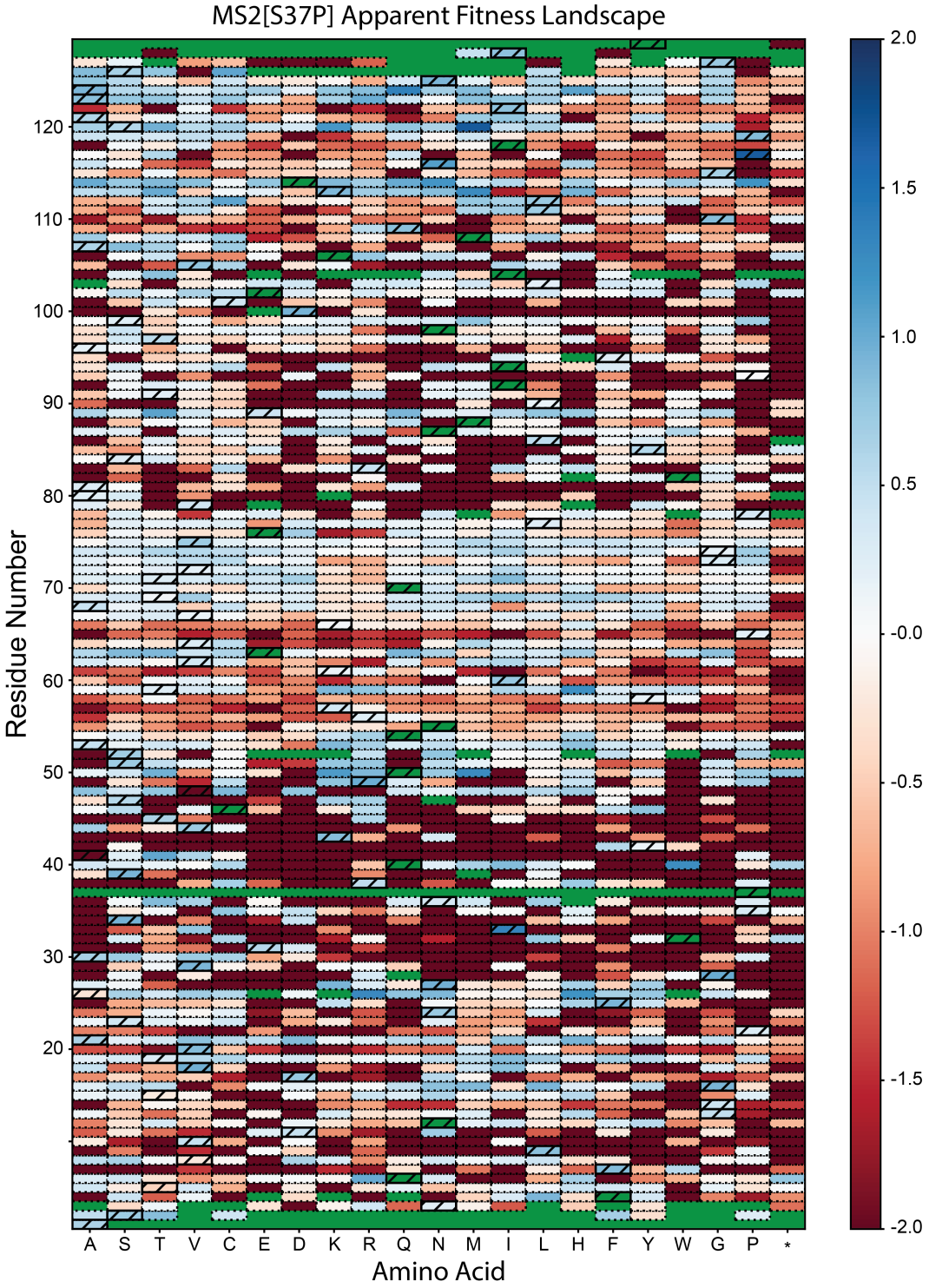

Figure 3.11: Apparent fitness landscape of MS2[S37P]. Blue indicates enriched amino acids, while red indicates combinations that are not enriched. Variants that were present in the plasmid library but absent in the VLP library are indicated in dark red while missing values are shown in green. Nonsense mutations are marked with an asterisk.

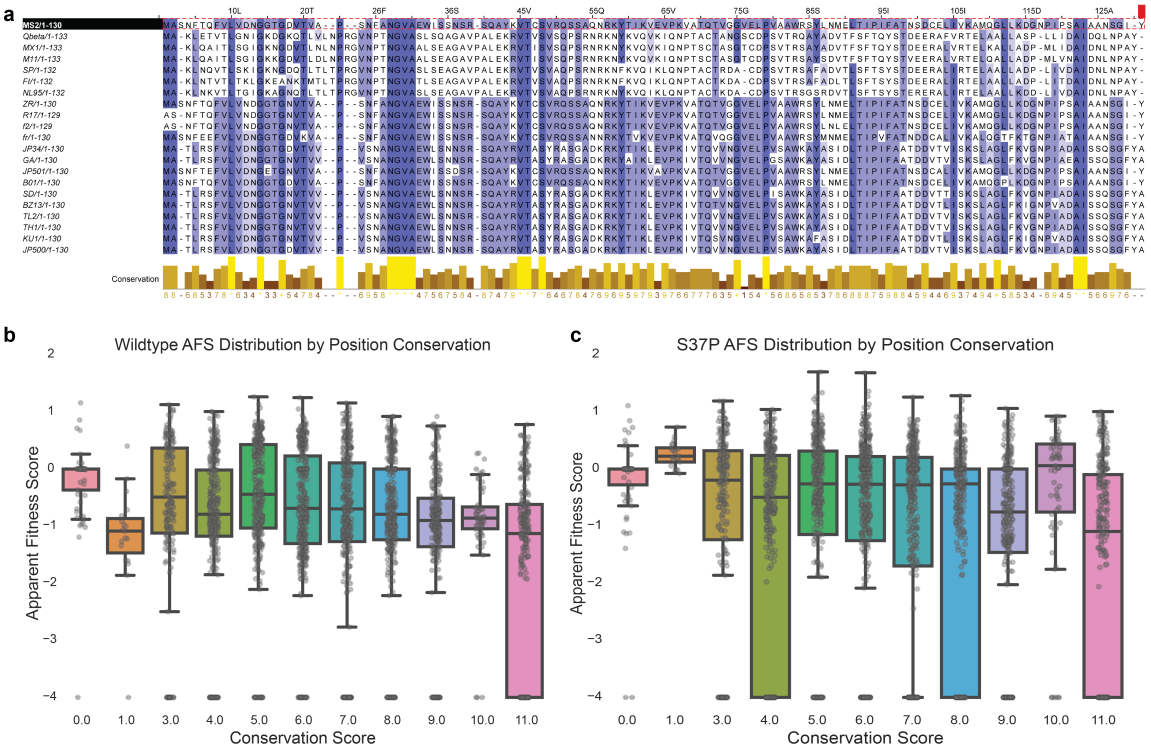

Figure 3.12: Sequence conservation of MS2 compared to apparent fitness scores. (a) Multiple sequence alignment of bacteriophage MS2 with structural homologues. Apparent fitness score distribution for (b) MS2[WT] and (c) MS2[S37P] sorted by conservation score is shown. A conservation score of 11 indicates a position was completely conserved, while a score of 0 indicates a position was highly variable.

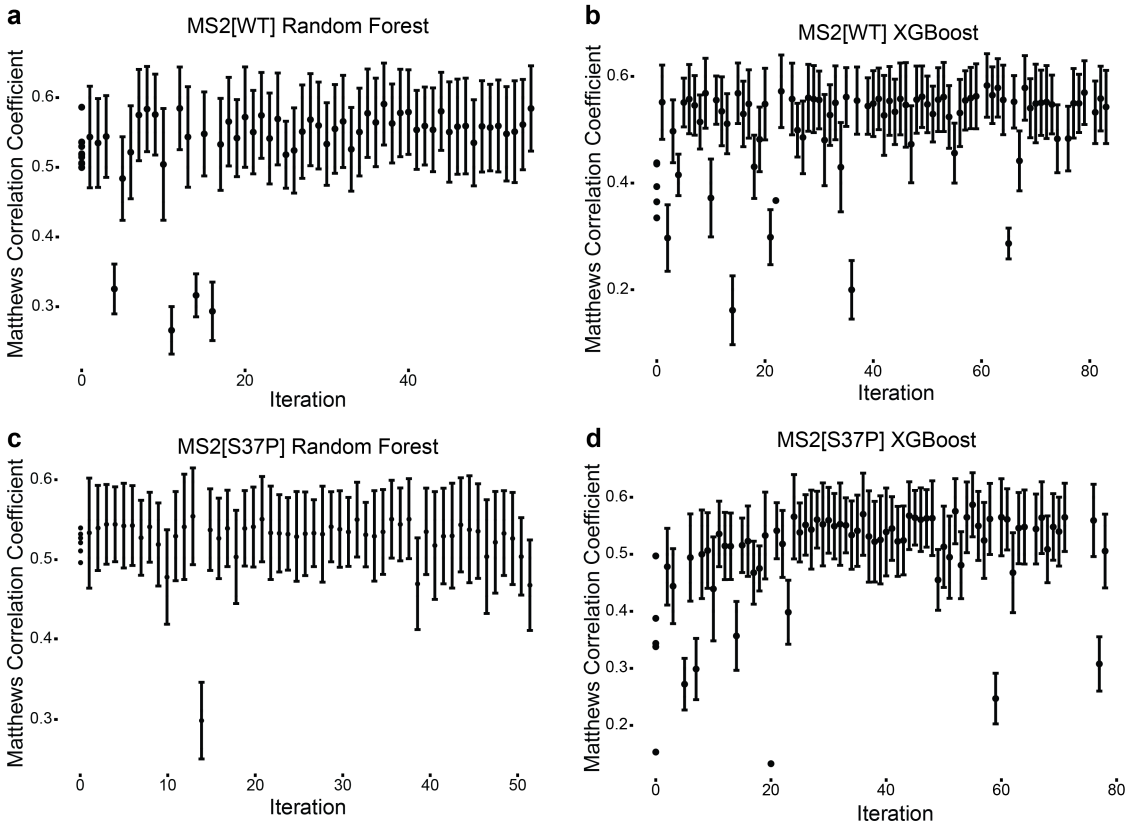

Figure 3.13: Bayesian tuning of machine learning models for assembly classification. Mean Matthews correlation coefficient of 10-fold cross validation for each parameter iteration is plotted for (a) MS2[WT] random forest, (b) MS2[WT] XGBoost, (c) MS2[S37P] random forest, and (d) MS2[S37P] XGBoost models. Tuning tested 10 initial parameter sets and stopped after 20 iterations with no improvement.

3.7 Acknowledgements

The authors thank Dr. Kathleen Durkin at the UC Berkeley Molecular Graphics and Computation Facility for assistance in running BioLuminate computational Scanning through Maestro. The sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01.

3.8 References

(1) Glasgow, J.; Tullman-Ercek, D. *Applied Microbiology and Biotechnology* **2014**, *98*, 5847–5858.

(2) Molino, N. M.; Wang, S.-W. *Current Opinion in Biotechnology* **2014**, *28*, 75–82.

(3) Garg, H.; Mehmetoglu-Gurbuz, T.; Joshi, A. *Scientific Reports* **2020**, *10*, 4017.

(4) Nooraei, S.; Bahrulolum, H.; Hoseini, Z. S.; Katalani, C.; Hajizade, A.; Easton, A. J.; Ahmadian, G. *Journal of Nanobiotechnology* **2021**, *19*, 59.

(5) Rohovie, M. J.; Nagasawa, M.; Swartz, J. R. *Bioengineering & Translational Medicine* **2017**, *2*, 43–57.

(6) Wilkerson, J. W.; Yang, S.-O.; Funk, P. J.; Stanley, S. K.; Bundy, B. C. *New Biotechnology* **2018**, *44*, 59–63.

(7) Lázaro, E.; Arribas, M.; Cabanillas, L.; Román, I.; Acosta, E. *Scientific Reports* **2018**, *8*, 8080.

(8) Tetter, S.; Terasaka, N.; Steinauer, A.; Bingham, R. J.; Clark, S.; Scott, A. J. P.; Patel, N.; Leibundgut, M.; Wroblewski, E.; Ban, N.; Stockley, P. G.; Twarock, R.; Hilvert, D. *Science* **2021**, *372*, 1220–1224.

(9) Khmelinskaia, A.; Wargacki, A.; King, N. P. *Current Opinion in Microbiology* **2021**, *61*, 51–57.

(10) Wargacki, A. J.; Wörner, T. P.; van de Waterbeemd, M.; Ellis, D.; Heck, A. J. R.; King, N. P. *Nature Communications* **2021**, *12*, 883.

(11) Hsia, Y. et al. *Nature Communications* **2021**, *12*, 2294.

(12) Laniado, J.; Cannon, K. A.; Miller, J. E.; Sawaya, M. R.; McNamara, D. E.; Yeates, T. O. *ACS Nano* **2021**, *15*, 4277–4286.

(13) Laniado, J.; Meador, K.; Yeates, T. O. *Protein engineering, design & selection: PEDS* **2021**, *34*, gzab008.

(14) Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W. *Science (New York, N.Y.)* **1992**, *255*, 178–183.

(15) Cordonnier, A.; Montagnier, L.; Emerman, M. *Nature* **1989**, *340*, 571–574.

(16) Asensio, M. A.; Morella, N. M.; Jakobson, C. M.; Hartman, E. C.; Glasgow, J. E.; Sankaran, B.; Zwart, P. H.; Tullman-Ercek, D. *Nano Letters* **2016**, *16*, 5944–5950.

(17) Araya, C. L.; Fowler, D. M. *Trends in biotechnology* **2011**, *29*, 435–442.

(18) Starita, L. M.; Fields, S. *Cold Spring Harbor Protocols* **2015**, *2015*.

(19) Al-Mawsawi, L. Q.; Wu, N. C.; Olson, C. A.; Shi, V. C.; Qi, H.; Zheng, X.; Wu, T.-T.; Sun, R. *Retrovirology* **2014**, *11*, 124.

(20) Ferguson, A. L.; Mann, J. K.; Omarjee, S.; Ndung'u, T.; Walker, B. D.; Chakraborty, A. K. *Immunity* **2013**, *38*, 606–617.

(21) Visher, E.; Whitefield, S. E.; McCrone, J. T.; Fitzsimmons, W.; Lauring, A. S. *PLOS Pathogens* **2016**, *12*, Publisher: Public Library of Science.

(22) Acevedo, A.; Brodsky, L.; Andino, R. *Nature* **2014**, *505*, 686–690.

(23) Hartman, E. C.; Jakobson, C. M.; Favor, A. H.; Lobba, M. J.; Álvarez-Benedicto, E.; Francis, M. B.; Tullman-Ercek, D. *Nature Communications* **2018**, *9*, 1385.

(24) Qi, H.; Wu, N. C.; Du, Y.; Wu, T.-T.; Sun, R. *Current Opinion in Virology* **2015**, *14*, 62–70.

(25) Hartman, E. C.; Lobba, M. J.; Favor, A. H.; Robinson, S. A.; Francis, M. B.; Tullman-Ercek, D. *Biochemistry* **2019**, *58*, 1527–1538.

(26) Brauer, D. D.; Hartman, E. C.; Bader, D. L. V.; Merz, Z. N.; Tullman-Ercek, D.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 3875–3884.

(27) Robinson, S. A.; Hartman, E. C.; Ikwuagwu, B. C.; Francis, M. B.; Tullman-Ercek, D. *Biomacromolecules* **2020**, *21*, 4194–4204.

(28) Kafader, J. O. et al. *Nature Methods* **2020**, *17*, 391–394.

(29) Stewart, J. J.; Lee, C. Y.; Ibrahim, S.; Watts, P.; Shlomchik, M.; Weigert, M.; Litwin, S. *Molecular Immunology* **1997**, *34*, 1067–1082.

(30) Stonehouse, N. J.; Valegård, K.; Golmohammadi, R.; van den Worm, S.; Walton, C.; Stockley, P. G.; Liljas, L. *Journal of Molecular Biology* **1996**, *256*, 330–339.

(31) Koning, R. I.; Gomez-Blanco, J.; Akopjana, I.; Vargas, J.; Kazaks, A.; Tars, K.; Carazo, J. M.; Koster, A. J. *Nature Communications* **2016**, *7*, 12524.

(32) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. *Protein Science: A Publication of the Protein Society* **2021**, *30*, 70–82.

(33) Bahadur, R. P.; Janin, J. *Proteins: Structure, Function, and Bioinformatics* **2008**, *71*, 407–414.

(34) Caspar, D. L. D.; Klug, A. *Cold Spring Harbor Symposia on Quantitative Biology* **1962**, *27*, 1–24.

(35) Knapman, T. W.; Morton, V. L.; Stonehouse, N. J.; Stockley, P. G.; Ashcroft, A. E. *Rapid communications in mass spectrometry : RCM* **2010**, *24*, 3033–3042.

(36) Beard, H.; Cholleti, A.; Pearlman, D.; Sherman, W.; Loving, K. A. *PLOS ONE* **2013**, *8*.

(37) Rodrigues, C. H.; Pires, D. E.; Ascher, D. B. *Protein Science* **2021**, *30*, 60–69.

(38) Dehouck, Y.; Kwasigroch, J. M.; Gilis, D.; Rooman, M. *BMC Bioinformatics* **2011**, *12*, 151.

(39) Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2012**, *42*, 463–484.

(40) Chicco, D.; Jurman, G. *BMC Genomics* **2020**, *21*, 6.

(41) Boughorbel, S.; Jarray, F.; El-Anbari, M. *PLOS One* **2017**, *12*.

(42) Altman, N. S. *The American Statistician* **1992**, *46*, 175–185.

(43) Walker H., S.; Duncan, D. B. *Biometrika* **1967**, *54*, 167–179.

(44) Breiman, L. *Machine Learning* **2001**, *45*, 5–32.

(45) Chen, T.; Guestrin, C. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery: New York, NY, USA, 2016, pp 785–794.

(46) Snoek, J.; Larochelle, H.; Adams, R. P. In *Advances in Neural Information Processing Systems*, Curran Associates, Inc.: 2012; Vol. 25.

(47) Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. *BMC Bioinformatics* **2008**, *9*, 307.

(48) Auret, L.; Aldrich, C. *Chemometrics and Intelligent Laboratory Systems* **2011**, *105*, 157–170.

(49) Hietpas, R. T.; Jensen, J. D.; Bolon, D. N. A. *Proceedings of the National Academy of Sciences* **2011**, *108*, 7896–7901.

(50) Pinto, F. L.; Thapper, A.; Sontheim, W.; Lindblad, P. *BMC Molecular Biology* **2009**, *10*, 79.

(51) Bolger, A. M.; Lohse, M.; Usadel, B. *Bioinformatics* **2014**, *30*, 2114–2120.

(52) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup *Bioinformatics* **2009**, *25*, 2078–2079.

(53) Kuhn, M.; Wickham, H., *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.* 2020.

(54) Greenwell, B. M.; Boehmke, B. C. *The R Journal* **2020**, *12*, 343–366.

Chapter 4

Improving the Academic Climate of an R1 STEM Department: Quantified Positive Shifts in Perception

The following is adapted from Stachl, Brauer, Mizuno, Gleason, Francis, and Baranger; J Chem Ed, 2021 with permission.

4.1 Abstract

While calls for increased diversity have echoed through STEM for decades, the field continues to exclude groups based on race, gender, sexual orientation and ability. Efforts to improve diversity have historically failed because they focus on increasing recruitment without repairing the systemic barriers within academic communities that prevent marginalized individuals from thriving. Presented here are the results of a multiyear effort to improve the climate of the UC Berkeley Department of Chemistry. We use student-led, department-specific, faculty-supported initiatives to quantitatively assess and address issues of departmental climate. These results provide evidence that data-driven community discussions alongside cooperative efforts to remedy persistent concerns are effective methods for driving positive change. Longitudinal assessment of academic climate from 2018 to 2020 via annual department-wide surveys indicates that these interventions have succeeded in shifting student and faculty perceptions. This work confirms the positive outcomes of having a sustainable, data-driven framework for affecting change within a graduate community.

4.2 Introduction

Although the relative representation of women in science, technology, engineering, and mathematics (STEM) has improved since 1993, women continue to be underrepresented in all career stages across nearly every STEM field.[1] Moreover, individuals who identify as Black, Hispanic, American Indian, Alaskan Native, Native Hawaiian, and Pacific Islander make up just 7.6% of researchers at all levels of doctorate-granting research universities and are systematically disadvantaged within STEM.[2] These numbers further decline at the most senior levels of academia and industry.[1–4]

There is growing evidence that systemic patterns of bias, discrimination, and inequity discourage women and members of other historically marginalized groups from entering or persisting in STEM.[5–9] For example, archaic stereotypes suggesting that women have less innate scientific ability, and implying that whiteness is correlated more strongly with ability than any other race, have cemented gender and racial/ethnic disparities as natural outcomes.[10–12] The resulting, persistent culture heavily influences perceptions of who can and cannot thrive in STEM[5,13–22] and creates structural barriers that can impede the success of women and gender, racial, and ethnic minorities—for example, by making the doctoral experience socially isolating, research groups inhospitable, and mentoring interactions less than satisfactory.[5,23–29]

As members of the academic community, we have a moral imperative to ensure equitable opportunities and inclusive environments for would-be scientists from all backgrounds. Top-down administrative approaches have gained traction as solutions to many of the aforementioned issues, but such efforts can fail to identify and improve the unique situation of a department. Department-level initiatives are thus necessary to combat factors that negatively impact equity and inclusion within a community.[1,24,30–32] Furthermore, grassroots departmental efforts have the ability to engage passionate stakeholders to drive lasting institutional change.

4.2.1 Theoretical Framework

A change framework has grounded the goals of a graduate student-led, stakeholder supported, grassroots initiative to shift the academic climate of the University of California, Berkeley Department of Chemistry in a positive direction. This framework requires: (1) making longitudinal data collection an institutional priority, to diagnose specific problems within a department's academic climate;[30] (2) empowering committed administrative and graduate student leadership for the collaborative design and implementation of targeted, evidence-based interventions; (3) institutionalizing the developed interventions so they can persist through institutional or leadership disruptions (e.g. COVID-19 pandemic); and (4) soliciting regular feedback via annual

data collection, to monitor the department climate and determine whether the interventions are effective. (3) The application of this theory of change has successfully enabled us to address the following research questions:

1. Can longitudinal changes in academic climate be quantified?
2. Have perceptions of academic climate improved as data-driven interventions have been implemented?

To the best of our knowledge, no other large-scale, coordinated, longitudinal efforts to improve a STEM department climate exist. Thus, the approach we use, and the results presented herein, expand the scope of foundational knowledge and existing methods for improving academic culture in a quantitative way.

4.3 Results

Since 2018, graduate students in the Department of Chemistry at the University of California, Berkeley, have been leading efforts to assess the issues that affect diversity, equity, and inclusion within the department. Thorough discussion of the founding of these initiatives and their multitude of outcomes are discussed elsewhere.[24,30,32,33] The first of these efforts is an annual department academic climate survey—designed to obtain data indicative of the department sentiment on key issues affecting inclusion, diversity, and well-being among graduate students, postdoctoral researchers, and faculty—which has been administered every spring since 2018.[30]

The total response rates were 43.1% (2018), 35.7% (2019), and 39.4% (2020). Graduate student and postdoctoral researcher respondent demographics have been representative of the department population across all 3 years of data, as 40% of graduate student and postdoctoral researcher respondents identify as female, and 40% of the graduate student researchers in the UC Berkeley Department of Chemistry are female. We note that 55% of respondents identified as belonging to underrepresented groups (URGs) across all three climate surveys. While this number is high, our definition of URGs is broad—it includes, but is not limited to, individuals who identify as female; are from underrepresented racial, religious, ethnic, sexual orientation, and international groups; have a disability(ies) (defined as a physical or mental impairment that substantially limits one or more major life activities); and have low socio-economic status. (4) Given the underrepresentation of women and racial/ethnic minority scholars in STEM, the term URG was used to enable a general comparison of URG and majority respondent populations, while still maintaining a balanced representation of study participants. Racial/ethnic minority

respondent responses are not analyzed separately due to concerns about compromised confidentiality from low total numbers of these students and faculty in the department. It is important to note that we did not collect demographic information from faculty members, as the low numbers of female and minority faculty in the department may compromise respondent confidentiality.

Each year after the climate survey closes, the areas of concern most frequently highlighted by department members in survey responses are compiled and used to ground open, active discussion among community members at the annual 'Chemistry Department Information and Brainstorming Session' (cDIBS) forum. cDIBS occurs every spring and has been a critical aspect of our academic climate initiative because it encourages community members to generate practical solutions to the issues that are revealed in our department's own data collaboratively.[24,30] cDIBS is attended by a range of stakeholders (graduate students, faculty, postdocs, and staff), which enables rapid implementation and institutionalization of the new initiatives and interventions that result from this event every year.

4.3.1 Interventions

Since 2018, the following evidence-based interventions have been designed and implemented to directly combat disparities and increase inclusivity within the Department of Chemistry academic community:

1. Discussions of mental health, cultural adaptation, and student identity are included throughout new graduate student orientation.

2. The graduate student handbook has been updated to include a comprehensive overview of the departmental policies and resources for students who do not pass their qualifying exam or decide to leave the program, to reduce the stigmas that surround those options.

3. Graduate students and College of Chemistry administration host an annual crowdfunding campaign to raise awareness of and money for these new initiatives.

4. Formal methods for inclusion of graduate student feedback in the faculty hiring processes were developed: graduate student search committee members interview candidates about their research, mentorship, diversity, service, and teaching; the student committee's feedback is then compiled and given to the faculty hiring committee.

5. Sexual violence and sexual harassment (SVSH) prevention training for new graduate students was redesigned. All College of Chemistry incoming graduate

students attend a peer-led, in-person SVSH prevention workshop during new student orientation. The workshop includes a small group discussion of real SVSH scenarios that occurred on the University of California, Berkeley campus. Scenarios are restructured in conjunction with the PATH to Care Center. Within their small groups, workshop attendees assess the problem in each scenario, discuss the underlying structural issues that led to those situations, and brainstorm strategies to resolve the issue.

6. A monthly event called the Diversity and Inclusion Focus Group (DIFG) was developed as a recurring forum where faculty, graduate students, staff, and postdoctoral researchers discuss challenging issues in academic culture. DIFG is intended to foster literature-based discussions about systemic inequities within academia, help shift social norms, and positively influence their confidence when engaging with peers, mentors, and mentees in difficult topics of conversation. Past DIFG topics include: sexism and racism in science, LGBTQ+ community inclusion, unconscious bias in hiring and letters of recommendation, the ramifications of sexual assault and harassment, mental health, managing work–life balance, and more.

While we recognize that many factors can contribute to improved perceptions of the academic climate, these interventions were specifically implemented between the 2018 and 2020 climate surveys. Thus, we speculate that these interventions largely contributed to the resulting, overall positive change in community perceptions of the Berkeley Chemistry academic climate.

4.3.2 Longitudinal Analysis of Academic Climate

Annual department climate survey data have been critical to measuring perceptions within the chemistry community. Longitudinal analyses were performed on the response distributions from each core survey question to understand shifts in the perception of the academic climate from 2018 to 2020. Encouragingly, all statistically significant shifts in responses were positive.

4.3.3 Equity and Inclusion

We report a significant increase from 2018 to 2020 in graduate student and postdoctoral researcher perceptions of there being sufficient discussion of equity and inclusion issues ($p \leq 0.01$; Figure 4.1A) and action toward improving equity and inclusion ($p \leq 0.01$; Figure 4.1B). The data also indicate a significant increase in respondent perceptions of feeling valued and included as a member of the department from 2018 to 2020 ($p \leq 0.01$; Figure 4.1C).

These data were further analyzed to determine whether any differences in perception exist between those who identify as belonging to a URG or not across survey years. The results suggest persistent differences in perceptions of inclusion between respondents that belong to URGs and their majority counterparts (Figure 4.2). This is also true with respect to the department's overall tolerance of exclusionary behavior and harassment (Figure 4.2). In 2020, for example, non-URG respondents felt generally more valued and included than respondents belonging to URGs ($p \leq 0.05$). Interestingly, not all of these perception gaps existed among 2018 and 2019 respondents. The significant perception gaps between URGs and their majority counterparts in 2018 and 2019 are shown in Figure 4.2 as well. Data from 2018 and 2019 questions that did not indicate statistically significant differences between URG and non-URG respondents are omitted.

The data also indicate that faculty

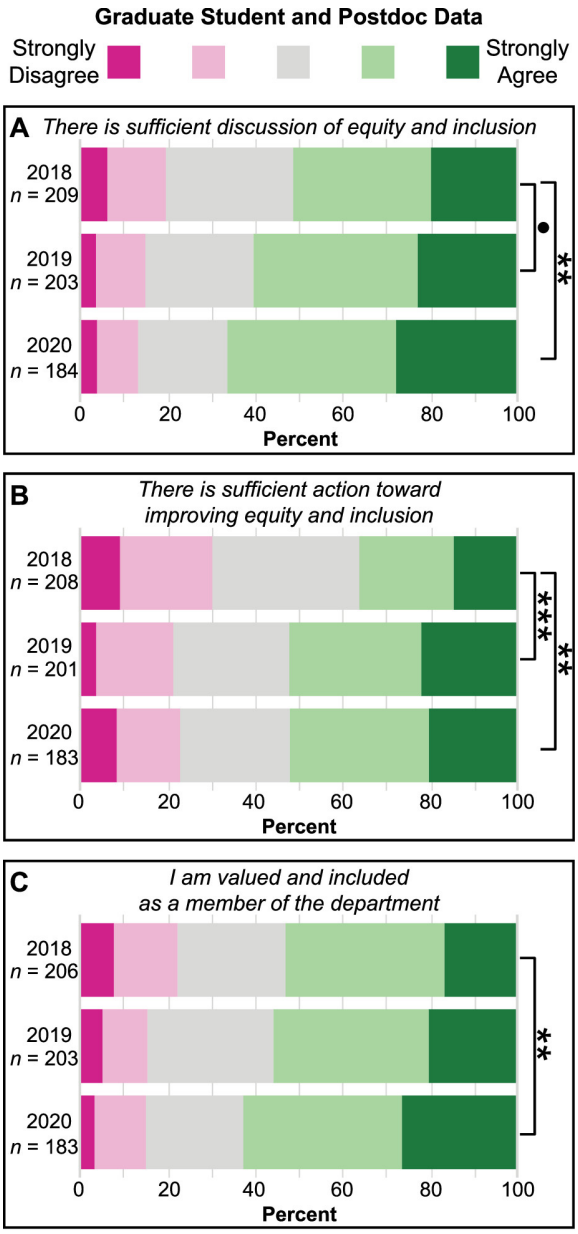

Figure 4.1: Significant increase in graduate student and postdoctoral researcher perceptions of equity, inclusion, and their sense of value from 2018 to 2020. Responses regarding (A) discussion of equity and inclusion, (B) action toward improving equity and inclusion, and (C) perceptions of feeling valued and included as a member of the department are shown.

perceptions of mutual respect, cooperation, and collaboration with their colleagues improved significantly since 2018 ($p \leq 0.05$). Perceptions that faculty from URGs are treated the same as all other faculty members during the tenure process also increased significantly ($p \leq 0.01$).

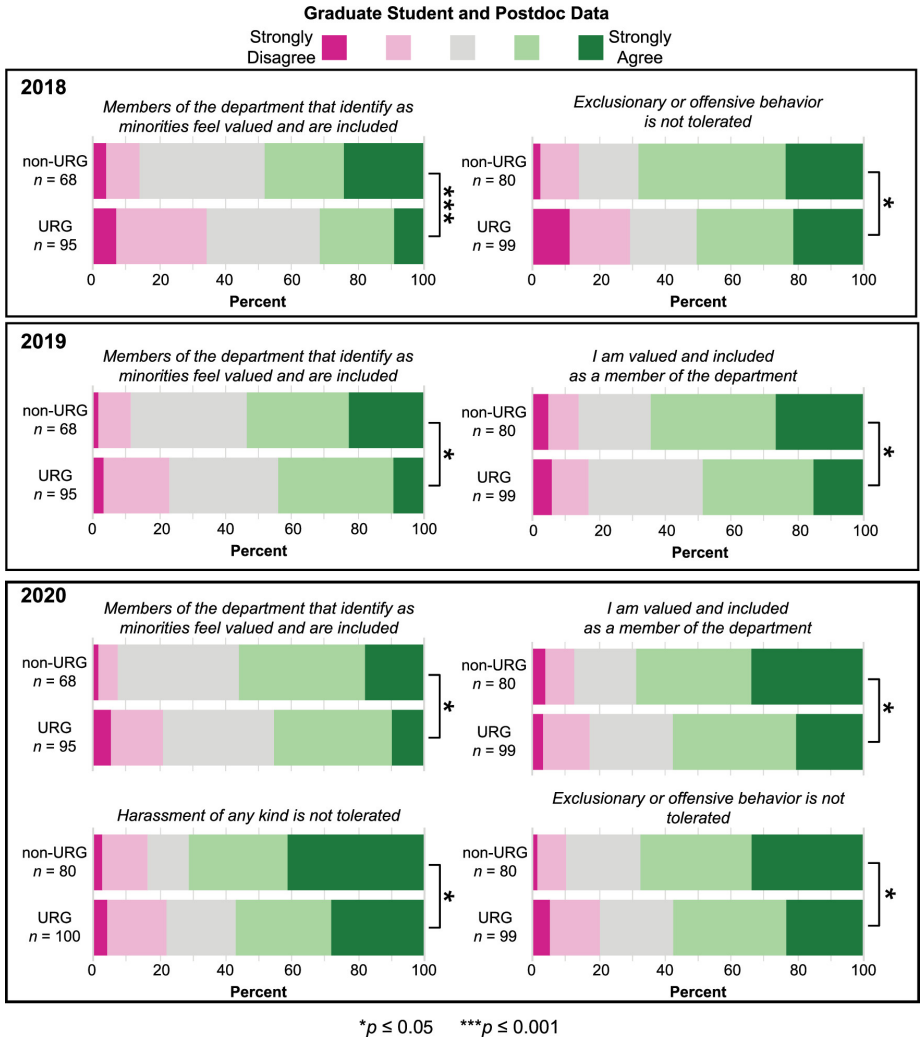

Figure 4.2: Perception gaps between URG and non-URG respondents. The entire distribution of responses from 2018 (top panel), 2019 (middle panel), and 2020 (bottom panel) climate survey data regarding 'peer and community interaction' questions, disaggregated based on URG-identity. Note that "URG" includes both female-identifying and racial/ethnic minority individuals, as the low total numbers of racial/ethnic minority members of the department may compromise confidentiality.

4.3.4 Mentorship and Mental Health

Graduate student and postdoctoral researcher respondents were asked about the research-, career-, and mental health-related support they receive from chemistry faculty. Encouragingly, faculty were generally rated highly in their ability to provide research-related support across all three years of data. There was an increase in mentee perceptions of having a research advisor who treats their ideas with respect (2018–2019, $p < 0.10$; 2018–2020, $p \leq 0.01$; Figure 4.3). All other mentoring interaction questions showed no statistically significant changes from 2018 to 2020.

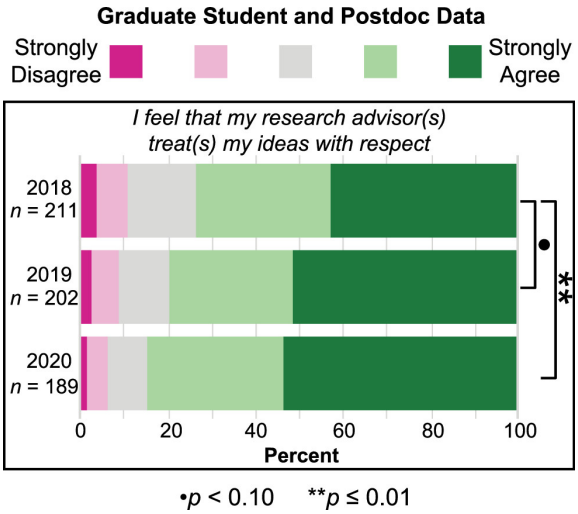

Figure 4.3: Significant increase in graduate student and postdoctoral researcher perceptions of mentors treating their ideas with respect. The entire distributions of responses from each year are compared for this 'mentor interactions' question.

4.3.5 Department Resources

All department members indicated increased knowledge of whom to approach regarding concerns about department climate since 2018 (Figure 4.4). This knowledge continued to increase significantly for graduate students and postdoctoral researchers from 2019 to 2020. This indicates that community members may be more empowered to engage with agents of change throughout the department when issues or new ideas arise.

4.4 Results and Discussion

Annual department climate survey data have supported the collaborative development of a number of department interventions, including the monthly, student-led Diversity and Inclusion Focus Groups (DIFGs), which began in 2018 and have succeeded in building community among attendees and lowering the barrier to engaging in challenging conversations about mental health, sexism, racism, unconscious bias, and more. While we recognize that many factors can contribute to improved perceptions of the academic climate, we speculate that these interventions contributed largely. We also believe that periodic, publicized initiatives (e.g., annual cDIBS event, crowdfunding, and the recent Stachl et al. publication have been instrumental in displaying the ongoing commitment to equity and inclusion, as well as ensuring transparency with

the department community. Multilevel stakeholder participation in these initiatives, combined with a formative assessment approach to maximizing their efficacy,[34,35] has been critical for improving our academic climate. These grassroots methods for creating tailored solutions to departmental concerns are poised to improve graduate community climate more nimbly and precisely than top–down administrative approaches.

Data from the past three years indicate that respondents have felt an increase in their overall sense of value and inclusion within the department. Still, it is clear that URG respondents feel these improvements less strongly than their majority counterparts (Figure 4.2). While it is heartening that there continues to be near-unanimous agreement that representation should improve at all levels within our department, we must continue efforts to improve the academic climate in order to attract and retain individuals of all identities and from all backgrounds. This is particularly important at the faculty level, as our faculty population currently includes just 18% women. Additionally, it is critical to take further action to educate our community about the biases that negatively affect the experiences of members of historically marginalized groups in STEM. One way to do this is to teach mentors (at all levels, but particularly faculty with large research groups) how to make use of inclusive approaches to mentoring (e.g., active listening, cultural awareness and responsiveness, and how to reflect on biases and prejudices that may impact trust between mentors and mentees—especially those with marginalized identities).[1] In a previous sense of belonging study, faculty acknowledged that it is most difficult for them to mentor all of their students effectively, suggesting that improved mentor training would significantly

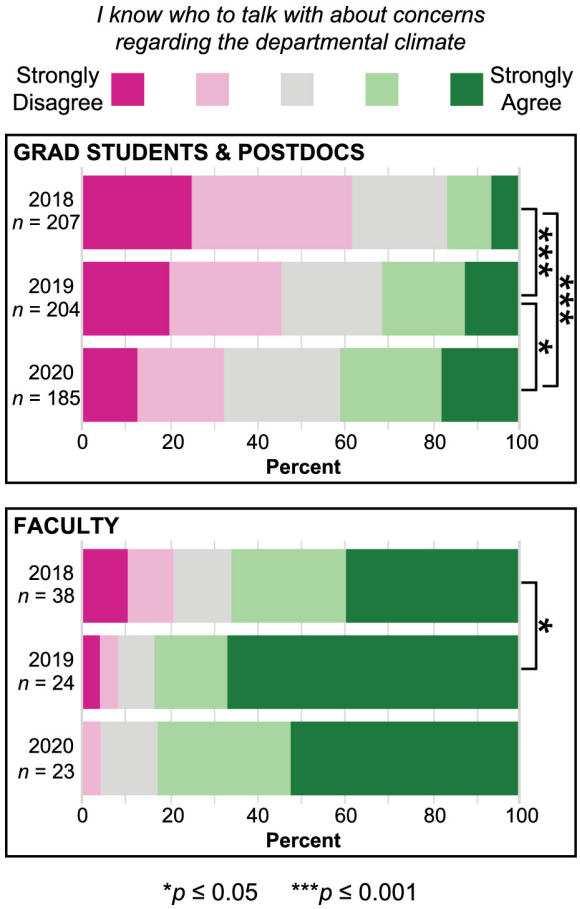

Figure 4.4: Significant increase in all department members' knowledge of whom to contact regarding department climate concerns from 2018 to 2020. The entire distribution of graduate student and postdoctoral researcher responses (top panel) and faculty responses (bottom panel) for each year of climate survey data.

benefit both mentors and mentees.[24] Mentor training could also contribute positively to faculty perceptions of how prepared they feel directing mentees to mental and physical health resources on campus, which have not changed since 2018.

Allocating resources to understand the experiences of members of historically marginalized groups is key to further improving the academic climate, because such experiences do not stand out when data are collected in aggregate in a predominantly white department.[1] The small number of racial/ethnic minority members within our academic community necessitates the aggregation of those data with those of female respondents, which limits investigation of the specific needs of members of historically marginalized groups. Future qualitative studies will be carried out to improve understanding of the needs of members of historically marginalized groups and help shape policies to improve their experiences. Additionally, reimagining faculty recruitment, evaluation, and promotion can also have a profound effect on diversity and inclusion at all levels—specifically because being mentored by faculty with similar identities can elevate mentees from URGs, helping them develop positive self-perceptions about their academic capabilities.[36–38]

Perceptions of faculty membership have improved overall since 2018, and there was a significant increase in mentee's feelings that their research advisor(s) provide(s) emotional support when necessary. Mentees continue to find it considerably easier to discuss research-related topics with their mentors than to solicit non-academic career support or discuss mental/physical health concerns. Therefore, we suggest offering more opportunities for mentorship in external groups, either through peer-to-peer mentoring or mentee-non-advisor faculty interactions. Finding a way to formalize and improve mentoring at all levels would not only help accommodate the needs and personalities of more mentees, but may also help students feel more comfortable communicating openly with their mentors about research, career goals, and their general well-being. The latter is particularly important for eliminating stigmas surrounding the discussion of mental health, and may help faculty advisors become more supportive in all facets of mentee development.[39]

In response to the 2020 climate survey and cDIBS (Chemistry Department Information and Brainstorming Session; during which these longitudinal data were presented and discussed), the COVID-19 pandemic, and the resurgence of the Black Lives Matter movement, the UC Berkeley College of Chemistry has implemented at least four new interventions to continue making significant forward progress toward a more diverse, inclusive, and equitable academic climate and culture. These include the following: (1) appointing an inaugural Associate Dean for Diversity, Equity, and Inclusion, who is already developing a 5-year strategic diversity, equity, inclusion, and belonging plan; (2) hiring its first Chief Diversity, Equity, and Inclusion Officer, who will continue developing and directing programs to engage faculty, staff, and

students at all levels in the ongoing diversity, equity, inclusion, and belonging efforts; (3) institutionalizing a "Graduate Diversity Program", to provide financial, social, and educational support—as well as College-wide recognition—for graduate students who carry out projects to improve diversity, equity, inclusion, and belonging within the College of Chemistry; and (4) ensuring more frequent discourse between community members at all levels and administrative leadership, particularly concerning the impact of COVID-19 on the Ph.D. program and College as a whole. We are excited to continue this grassroots work and make an enduring impact on the Berkeley Chemistry community.

4.5 Conclusion

Although it may be appealing to seek standardized solutions to resolve issues of institutional climate and inclusion, the particular needs and constraints of a given department vary widely. Highlighted here is a flexible framework for generating and implementing tailored solutions for problems related to diversity, equity, and inclusion in a STEM department. Periodic, grassroots initiatives in particular drive stakeholder engagement and culture change over time.

In contrast to strictly administrative policy changes, a community-driven, holistic approach enables rapid identification and addressing of the most pressing climate issues. The cyclical nature of the survey along with consistent respondent feedback allows for integrated, data-rich problem identification and solution evaluation. Three years of results from UC Berkeley's Department of Chemistry gives evidence that statistically significant, positive changes in departmental culture are possible using the climate survey and cDIBS framework. We hope these methods can act as a template for any academic institution seeking to develop a departmental culture that better serves its community, ultimately changing academia over time to be more welcoming for all.

4.6 Materials and Methods

4.6.1 Academic Climate Survey Instrument

The academic climate survey used in this study has been modified slightly from the original instrument designed by Stachl et al.[30,32] to assesses the experiences of graduate students, postdoctoral researchers, and faculty within the Department of Chemistry at the University of California, Berkeley. Graduate students and postdoctoral researchers took one version of the academic climate survey, and the

faculty took a different version. The content of the faculty survey mirrors that of the graduate student and postdoctoral researcher survey but from the perspective of a faculty member.

Twenty-five questions in the graduate student and postdoctoral researcher survey, and 18 questions in the faculty survey, were measured on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree), and six questions were measured on a 3-point Likert scale (1 = Not Important, 2 = Somewhat Important, 3 = Very Important). Twenty-one and 15 of these questions, respectively, were identical across all 3 years of survey data collected to enable a longitudinal comparison of data and changes of the perception of department climate over time. The remaining, non-core survey data were added to the 2020 survey, to assess the publication culture within our department. These data are beyond the scope of this manuscript. All survey items were coded so that a higher score indicates a more positive perception or experience of the academic climate within our community.

Additionally, the following questions were added to the survey in 2019 to gauge whether the administrative changes to the Department of Chemistry that have been implemented since the inception of the academic climate initiative have been noticed by the community:

"Since the 2018 and 2019 climate surveys, the Chemistry Graduate Life Committee (CGLC) and Department of Chemistry administration have worked together to:

1. Update the first-year handbook
2. Ensure discussions of mental health were incorporated into Fall orientation
3. Incorporate graduate student input in the faculty hiring process
4. Ensure non-alcoholic beverages and snacks in our weekly chemistry social hour (Chem Keg)
5. Incorporated peer-led sexual violence and sexual harassment training into new student orientation
6. Established a monthly diversity and inclusion focus group

Did you notice any of these changes?" (yes/no response choices) and "Do you have any feedback regarding the changes listed above?" (open-ended question).

The reliability of our academic climate survey was evaluated using the item response theory[40] and has a value of 0.84. This indicates that the items in this survey relate to each other and do provide a reliable measure of academic climate across all 3 years of data collection.[41,42]

4.6.2 Academic Climate Survey Administration

The surveys were fielded using the Qualtrics LLC platform. They were administered confidentially and distributed electronically via email to all graduate students, postdoctoral researchers, and faculty in the UC Berkeley Department of Chemistry using the "individual link" function in Qualtrics during the Spring semester of every academic year. The "anonymize responses" function in Qualtrics was used to retroactively delete all identifying information from survey responses. Additionally, since 2019, a departmental survey "release party" has been hosted—with free coffee, snacks, and sweets—to incentivize survey participation. All other details of the survey administration are the same as those reported by Stachl et al.[30] This longitudinal study was authorized by the University of California, Berkeley institutional review board, protocol ID#2019-01-11732. All survey respondents were informed that completion of the surveys is voluntary, and they all completed informed consent.

4.6.3 Longitudinal Data Analysis

The 2018 data used in this study were previously published by Stachl et al. "Prefer not to answer" data was omitted from our analysis, and nonbinary gender responses were removed from the gender category because of the low overall number of responses. We note that the term "URG" as used in this paper inclusive of both female-identifying and racial/ethnic minority populations. We did not separate these populations' data due to concerns about compromising confidentiality from low total numbers of racial/ethnic minority trainees. Additionally, we did not collect demographic information from faculty members, as low overall numbers of faculty in the department may compromise the confidentiality of responses. We also note that 2020 data were collected prior to the mandated shelter-in-place order in California due to the COVID-19 pandemic.

The Kruskal–Wallis H test (nonparametric significance test for ordinal data with 3+ independent variables) was used to carry out longitudinal comparison of data by question, and the Mann Whitney U test (nonparametric significance test for ordinal data with 2 independent variables) was used to compare demographic data for any given question within 1 year's dataset. In cases where pairwise comparisons are included, the Kruskal–Wallis H test was carried out on the entire dataset; if this analysis indicated that the distributions are not the same for each independent group ($p < 0.05$), then pairwise comparisons were carried out using the Mann Whitney U Test to determine the significance level between two groups within the dataset. All of these statistical analyses were completed using IBM SPSS Statistics Version 26. In general, changes were considered significant for $p \leq 0.10$. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.7 Acknowledgements

This effort would not have been possible without the support of the Department of Chemistry at UC Berkeley. Dr. Emily Hartman Guthrie in particular offered extensive contribution and support. The members of the Chemistry Graduate Life Committee, the cDIBS planning committee and group discussion leaders, and the Diversity and Inclusion Focus Group team volunteered significant time and effort for this initiative.

4.8 References

(1) National Academies of Sciences, E.; Medicine, et al., *Promising practices for addressing the underrepresentation of women in science, engineering, and medicine: Opening doors*; National Academies Press: 2020.

(2) Rivers, E. *National Science Foundation* **2017**.

(3) Shen, H. *Nature* **2013**, *495*, 22–24.

(4) Hernandez, R. *Open Chemistry Collaborative in Diversity Equity (OXIDE)*; tech. rep.; National Science Foundation, 2019.

(5) Moss-Racusin, C. A.; Dovidio, J. F.; Brescoll, V. L.; Graham, M. J.; Handelsman, J. *Proceedings of the National Academy of Sciences* **2012**, *109*, 16474–16479.

(6) Finson, K. D.; Beaver, J. B.; Cramond, B. L. *School Science and Mathematics* **1995**, *95*, 195–205.

(7) Cheryan, S.; Master, A.; Meltzoff, A. N. *Frontiers in Psychology* **2015**, *6*.

(8) Master, A.; Cheryan, S.; Meltzoff, A. N. *Journal of Educational Psychology* **2016**, *108*, 424–437.

(9) Cheryan, S.; Ziegler, S. A.; Montoya, A. K.; Jiang, L. *Psychological Bulletin* **2017**, *143*, 1–35.

(10) Nosek, B. A.; Banaji, M. R.; Greenwald, A. G. *Group Dynamics: Theory, Research, and Practice* **2002**, *6*, 101–115.

(11) Gaertner, S. L.; McLaughlin, J. P. *Social Psychology Quarterly* **1983**, *46*, 23.

(12) Dovidio, J. F.; Evans, N.; Tyler, R. B. *Journal of Experimental Social Psychology* **1986**, *22*, 22–37.

(13) Chanderbhan-Forde, S.; Heppner, R. S.; Borman, K. M. *Journal of Women and Minorities in Science and Engineering* **2012**, *18*, 179–198.

(14) Kulis, S.; Sicotte, D.; Collins, S. *Research in Higher Education* **2002**, *43*, 657–691.

(15) Clancy, K. B. H.; Lee, K. M. N.; Rodgers, E. M.; Richey, C. *Journal of Geophysical Research: Planets* **2017**, *122*, 1610–1623.

(16) Benya, F.; Widnall, S.; Johnson, P. National Academies of Sciences, Engineering, and Medicine Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine, 2018.

(17) Aycock, L. M.; Hazari, Z.; Brewe, E.; Clancy, K. B. H.; Hodapp, T.; Goertzen, R. M. *Physical Review Physics Education Research* **2019**, *15*.

(18) Tran, N.; Hayes, R. B.; Ho, I. K.; Crawford, S. L.; Chen, J.; Ockene, J. K.; Bond, M.; Rayman, P.; Dean, B.; Smith, S.; Thorndyke, L.; Frankin, P.; Plummer, D.; Pbert, L. *Psychology of Women Quarterly* **2019**, *43*, 509–525.

(19) Greene, J.; Stockard, J.; Lewis, P.; Richmond, G. *Journal of Chemical Education* **2010**, *87*, 381–385.

(20) Bronstein, P.; Farnsworth, L. *Research in Higher Education* **1998**, *39*, 557–585.

(21) Settles, I. H.; Cortina, L. M.; Malley, J.; Stewart, A. J. *Psychology of Women Quarterly* **2006**, *30*, 47–58.

(22) Milkman, K. L.; Akinola, M.; Chugh, D. *Journal of Applied Psychology* **2015**, *100*, 1678–1712.

(23) Newsome, J. L. *A report for the UK Resource Centre for Women in SET and the Royal Society of Chemistry* **2008**, 1–38.

(24) Stachl, C. N.; Baranger, A. M. *PLOS One* **2020**, *15*, ed. by Blanch, A.

(25) Eaton, A. A.; Saunders, J. F.; Jacobson, R. K.; West, K. *Sex Roles* **2020**, *82*, 127–141.

(26) Urry, M. *Nature* **2015**, *528*, 471–473.

(27) Grogan, K. E. *Nature Ecology & Evolution* **2019**, *3*, 3–6.

(28) Alper, J. *Science* **1993**, *260*, 409–411.

(29) Bejerano, A. R.; Bartosh, T. M. *Journal of Women and Minorities in Science and Engineering* **2015**, *21*, 107–124.

(30) Stachl, C. N.; Hartman, E. C.; Wemmer, D. E.; Francis, M. B. *Journal of Chemical Education* **2019**, *96*, 2149–2157.

(31) Puritty, C.; Strickland, L. R.; Alia, E.; Blonder, B.; Klein, E.; Kohl, M. T.; McGee, E.; Quintana, M.; Ridley, R. E.; Tellman, B.; Gerber, L. R. *Science* **2017**, *357*, 1101–1102.

(32) Stachl, C. N. From Water to Human Dynamics: Probing Chemical Systems in Aqueous Nanodrops, Quantifying Graduate Academic Culture, and Building Community Within an R1 STEM Department, Doctoral Thesis, University of California, Berkeley, 2020.

(33) Stachl, C. N.; Brauer, D. D.; Mizuno, H.; Gleason, J. M.; Francis, M. B.; Baranger, A. M. *ACS Omega* **2021**, *6*, 14410–14419.

(34) Coffey, J. E.; Hammer, D.; Levin, D. M.; Grant, T. *Journal of Research in Science Teaching* **2011**, *48*, 1109–1136.

(35) Van Den Heuvel-Panhuizen, M. *For the Learning of Mathematics* **2005**, *25*, Publisher: FLM Publishing Association, 2–23.

(36) Jimenez, M. F.; Laverty, T. M.; Bombaci, S. P.; Wilkins, K.; Bennett, D. E.; Pejchar, L. *Nature Ecology & Evolution* **2019**, *3*, 1030–1033.

(37) Santos, S. J.; Reigadas, E. T. *Journal of Hispanic Higher Education* **2002**, *1*, 40–50.

(38) Mireles-Rios, R.; Garcia, N. M. *Journal of Latinos and Education* **2019**, *18*, 376–386.

(39) Evans, T. M.; Bira, L.; Gastelum, J. B.; Weiss, L. T.; Vanderford, N. L. *Nature Biotechnology* **2018**, *36*, 282–284.

(40) Wilson, M., *Constructing Measures: An Item Response Modeling Approach: An Item Response Modeling Approach*; Routledge: New York, 2004.

(41) Bond, T. G.; Fox, C. M., *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition*, 2nd ed.; Psychology Press: New York, 2007.

(42) Wright, B.; Masters, G. *Measurement and statistics* **1982**.

Chapter 5

Mismatch in Perceptions of Success: Investigating Academic Values among Faculty and Doctoral Students

The following is adapted from Brauer, Mizuno, Stachl, Gleason, Bumann, Yates, Francis, and Baranger; J Chem Ed, 2021 with permission.

5.1 Abstract

Many cultural and institutional barriers have prevented Chemistry from realizing greater calls for diversity in academia. Though recent work has elucidated how the measures of success used in academia can disadvantage students from underrepresented groups at the undergraduate level, understanding of how success metrics are valued by minoritized students at the graduate level is lacking. Here, we use data generated from the UC Berkeley Department of Chemistry's student-led climate survey to investigate both how graduate students prioritize and how faculty employ common metrics for graduate student success. Results revealed that faculty undervalued metrics preferred by students from underrepresented groups (URGs) in STEM such as underrepresented people of color, women, LGBTQ+ students, and first-generation students. Priorities of students that do not identify as underrepresented displayed no statistically significant differences to faculty values. Questions regarding publication record, a common measure of success in STEM academia, suggest that graduate students, particularly those belonging to URGs, challenge the use of publication record as the primary metric of success in graduate school. These findings highlight some of the ways that definitions of academic success can be exclusionary for graduate

students from underrepresented backgrounds and encourage re-envisioning graduate school success in ways that reflect the values of diverse student populations.

5.2 Introduction

Despite receiving significant attention and widespread awareness, the crisis of marginalization of women and underrepresented people of color (URPOC) continues to persist in STEM disciplines. In the US, 30% of the population identify as Black, Indigenous, and People of Color (BIPOC), yet they account for only 10% of chemistry doctoral recipients and 6% of faculty at research institutions. Women in chemistry likewise experience significant attenuation in higher education, representing 41% of chemistry graduate students and 20% of faculty at research institutions.[1,2] Myriad research has been devoted to understanding this marginalization, highlighting institutional barriers,[3] discrimination,[4] stereotype threat,[5] and belonging uncertainty[6,7] as factors to address.

Recent work has examined definitions of student success in STEM education through the lens of critical race theory, asserting that historical views of student success enforce cultural hegemony and ultimately translate to policies and structures that harm minoritized students.[8] Students from underrepresented backgrounds often define success differently from the most literature-prevalent success metric, the grade point average (GPA). Through interviews, first-generation undergraduates revealed qualitative, emotional themes for success, such as self-efficacy and passion, and explicitly rejected notions of success related to exam performance or GPA.[9] Minoritized students are thus forced to either conform to a value system designed and enforced by the majority, underperform in these chosen metrics, or choose to leave the system entirely.[10]

In higher levels of STEM academia, quantitative metrics shift to those related to research productivity and output. For decades, publication rate has been the key indicator for success in the academy, often defined by the mantra "publish or perish."[11] Publication records are a key factor in hiring and tenure decisions, academic promotions, and allocation of grant funds.[12,13] The field of bibliometrics has employed statistical methods to generate various models for research productivity beyond raw publication count that have gained widespread use, including journal impact factor, h-index, and g-index.[14–16] These measures attempt to capture the significance of a scientist's contributions by weighing the number of publications with citation count over time. The application of these measures in an increasingly competitive field has resulted in a rise of publication-related fraud and manipulation, as scientists struggle to secure scarce funding.[17] Furthermore, it is well-established that these metrics

contribute to the marginalization of underrepresented groups in STEM: women and underrepresented minorities publish at lower rates, are less represented in high impact journals, and are less likely to be cited for their work than their well-represented counterparts.[18–20]

While efforts have been made to incorporate undergraduate student input in defining notions of academic success, work to understand how STEM graduate students prioritize various metrics of success during their doctoral program is lacking.[21] Graduate program climate surveys have shown recent success as mechanisms for understanding graduate student perceptions and implementing evidence-based interventions, particularly with regard to improving student mental health.[22,23] Here, we present an examination of how graduate students in the UC Berkeley Department of Chemistry view graduate student success using primary data generated via the Department's annual climate survey.[24,25] We sought to address two research questions:

1. How do graduate students and faculty in chemistry prioritize the varied measures of success for graduate students?

2. Are success metric preferences influenced by graduate student identity?

Graduate student rankings of success metrics reveal differences in priorities between students from various underrepresented groups (URGs) and non-URG students. In this study, URG encompasses URPOC, women, first-generation students, LGBTQ+ individuals, students with disabilities, and low-socioeconomic status (low-SES) students. Student values were also compared to success metrics employed by faculty in their assessment of graduate students. Analyses reveal that, on average, priorities diverge between faculty and URG graduate students while faculty and non-URG graduate student priorities align. These results highlight the importance of setting clear expectations between graduate students and faculty in STEM and underscore the importance of elevating minoritized student voices in rethinking frameworks for success in chemistry higher education.

5.2.1 Theoretical Framework

This work draws on the frameworks of cultural capital, social reproduction, and critical race theory (CRT) to investigate value systems in STEM academia. Theories of cultural capital and social reproduction posit the perpetuation of social inequalities via socialization of culturally relevant skills, preferences, and norms.[26] Educational settings reinforce and reward cultural capital that aligns with the values of the dominant class, providing institutional recognition and conversion social hierarchies

into academic hierarchies.[27] Some educational research has used theories of cultural capital to construct deficit notions of minoritized groups — minoritized groups are to blame for unequal outcomes due to failure in required metrics.[28] However, more recent work has reconceptualized cultural capital through CRT, rejecting deficit notions of minoritized individuals.[29] CRT challenges the dominant ideology and looks to dismantle "notions of colorblindness, meritocracy, deficit thinking, linguicism, and other forms of subordination".[30] Intersectionality is a core commitment of CRT, recognizing that discrimination crosses multiple planes of identity, including race, gender, class, sexuality and disability.[31] Framing cultural capital through CRT recognizes that structures center the power with the majority and fail to recognize the values of minoritized communities. With this understanding, dominant modes of cultural capital in settings such as higher education may be challenged with skills, abilities, and knowledge valued by marginalized voices. This work in particular aims to garner input from underrepresented doctoral students to reframe modes of success in graduate school that may better serve the priorities of all graduate students.

5.3 Survey Methods

Since 2018, the Department of Chemistry at UC Berkeley has administered an annual climate survey to understand faculty and graduate student perspectives on issues pertaining to inclusivity, diversity, and wellness in the department.[24,25] Discussions held in the accompanying Chemistry Department Information and Brainstorming Session (cDIBS) in which faculty, postdocs, and graduate students discuss survey results led to initial conception of a survey section concerning publication and success metrics. Success metrics questions were subsequently designed by a graduate student committee responsible for climate survey design and administration. Questions were finalized following input from faculty stakeholders, including the Chair of the Department of Chemistry and Associate Dean for Diversity, Equity, and Inclusion.

Survey data were collected in the 2021 academic climate survey administered over 14 days in March 2021. Graduate students and faculty received a unique link to the online survey via their UC Berkeley email using the Qualtrics LLC platform. All respondents offered informed consent and were notified that completion of the survey was voluntary. The "anonymize responses" function in Qualtrics was used to remove identifying information from survey responses. The study was approved by the UC Berkeley institutional review board, protocol #2019-10-11732. The 2021 survey included the new success metric section: graduate students were asked four questions related to their perception of publications as a metric for graduate student success; faculty were asked three questions related to publication as a metric of assessment

for their own careers as well as how important publication record is in evaluating their graduate students. Graduate students and faculty were also asked to rank seven common metrics for assessing graduate student success.

Demographic information on identification with underrepresented group(s) was collected for graduate students but not for faculty as such information could compromise anonymity. Ensuring anonymity has been crucial for encouraging continued engagement of graduate students and faculty. The graduate student response rate for the 2021 survey was ∼40% (154/387), and the faculty response rate was ∼40% (26/63), which is consistent with response rates from previous years.[24,25]. The statistical significance of differences in distribution between groups was calculated using the Mann–Whitney U test. All statistical analyses were performed using IBM SPSS Statistics 27.

5.4 Results and Discussion

5.4.1 Academic Publication as the Key Metric for Graduate Student Success

The full distribution of graduate student and faculty responses to publication questions are presented in Figure 5.1. In general, members of the department agree that publications are an important metric of their personal successes (Figure 1a). However, there are large disparities in how publications should be and are

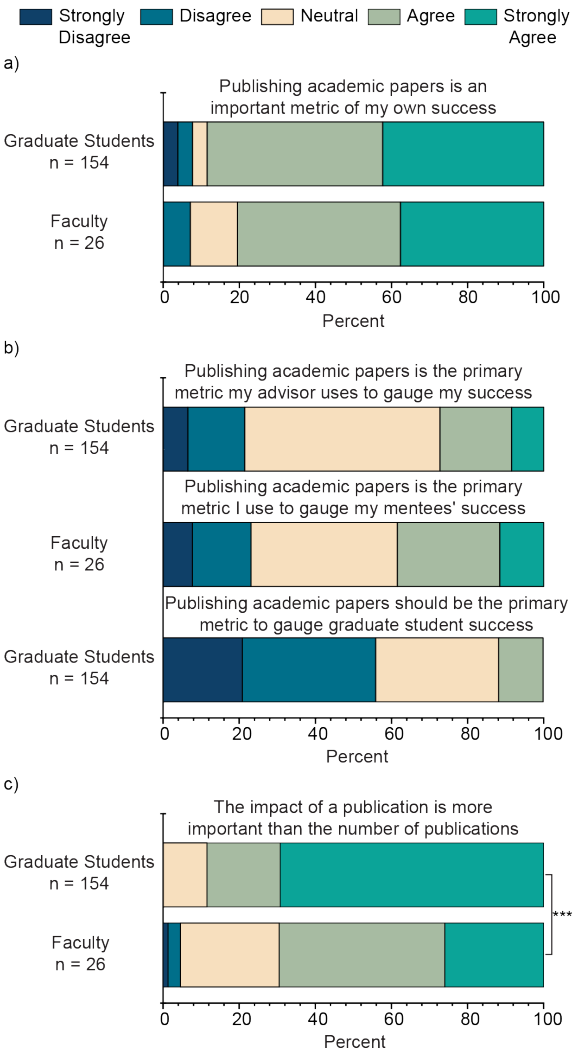

Figure 5.1: Distribution of responses regarding publication as a metric for success. (a) Graduate student and faculty responses pertaining to publication as a personal success metric. (b) Responses related to use of publication by faculty advisors to judge graduate student success. (c) Responses evaluating if publication impact is more important than quantity. *** indicates $p < 0.001$.

utilized in the assessment of graduate student success. About 40% of faculty respondents agree that they use publication record as the primary metric to gauge mentees' success, while only ∼27% of graduate students believe their advisor assesses them in this manner (Figure 5.1b). Furthermore, very few (∼12%) graduate students believe academic papers should be used as the primary metric to gauge their success (Figure 5.1c). Though the majority of graduate students (∼70%), and faculty (∼88%) agree that the impact of a publication is more important than the number of publications, faculty tend to endorse impact over quantity less strongly (Figure 5.1c).

These data were further analyzed based on self-identified URG status. As our definition of URG encompasses many underrepresented groups, including women, they constitute the majority of respondents (67% of graduate student responses). However, the constituent groups are underrepresented across multiple facets of STEM, ranging from the undergraduate to the faculty level, and thus are considered URG members despite their majority designation in our responses. Graduate students who identified as members of a URG were less likely than their non-URG peers to agree with the statement: "publishing academic papers should be the primary metric used to gauge graduate student success" ($p < 0.05$, Figure 5.2). No other questions about publication values achieved statistical significance when disaggregated by URG status.

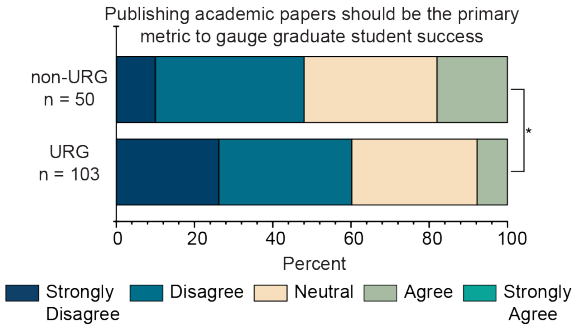

Figure 5.2: Distribution of responses regarding publication as a metric for success. (a) Graduate student and faculty responses pertaining to publication as a personal success metric. (b) Responses related to use of publication by faculty advisors to judge graduate student success. (c) Responses evaluating if publication impact is more important than quantity. *** indicates $p < 0.001$.

5.4.2 Divergent Priorities of Graduate Students by Demographic Group

Histograms containing the rank order distribution of graduate student success metrics as determined by faculty, URG graduate students and non-URG graduate students are shown in Figure 5.3. Both faculty and graduate students generally indicated the "ability to develop new research projects" as the most important factor for assessing graduate student success. "Publication record", "Mentorship", and "Scientific presentation" were given moderate priority while "Teaching", "Time spent working" and "DEI and outreach work" were ranked lower on average. Non-URG students'

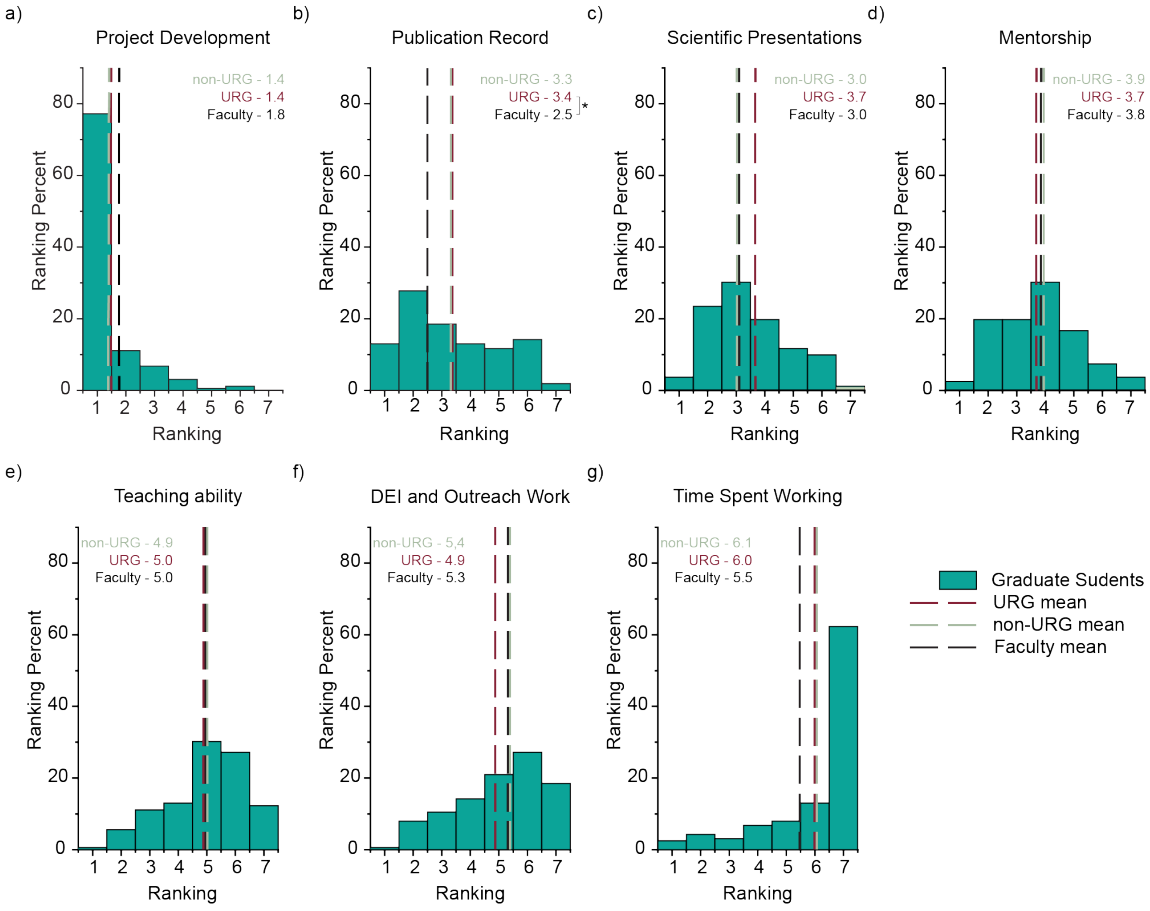

Figure 5.3: Graduate student ranking distribution of success metrics. Students ordered their priorities from 1 to 7, with 1 indicating the highest priority metric. Faculty, URG-identifying graduate student, and non-URG graduate student mean rankings for each metric are indicated via black, red, and green dotted lines, respectively.

priorities aligned more closely with the values of faculty; no metrics displayed a statistically significant difference in distribution between faculty and well-represented graduate students. URG students, however, ranked publication record as a lower priority success metric than did faculty ($p = 0.031$). It is worth noting that non-URG students ranked publications lower than faculty as well, but the shift was just outside of statistical significance ($p = 0.058$). Other visible preferences that were not statistically significant include a lower priority of scientific publication by URG students, an increased priority of DEI and outreach work by URG students, and lower priority of time spent working by both graduate student groups relative to faculty.

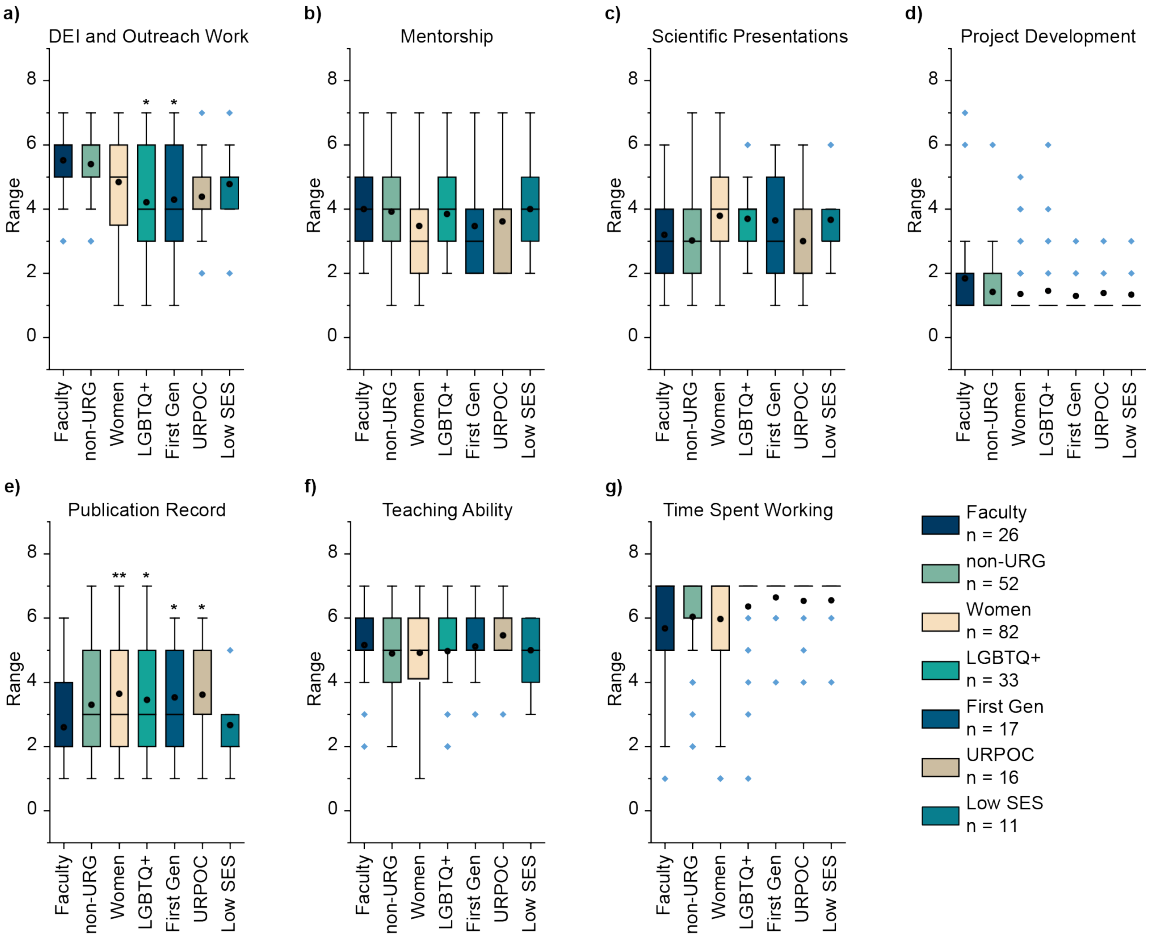

Figure 5.4: Box plot of success metric priority rankings separated by graduate student identity. A response of 1 indicates the highest priority metric, and 7 indicates the lowest priority. Mean ranking of each group is indicated by a black circle, while outliers are shown with blue circles. * indicates $p < 0.05$, ** indicates $p < 0.01$.

When separated by underrepresented group affiliations (URPOC, women, LGBTQ, first-generation students, and low socio-economic status), various statistically significant shifts in priorities emerge between graduate student affinity groups as well as relative to the metrics faculty use for student assessment. Significant disparities are shown in Figure 5.4. Graduate students who identified as URPOC de-emphasized the value of publication record ($p = 0.019$) while placing more value on DEI and outreach work ($p = 0.039$) compared to faculty. Women graduate students showed a similar tendency towards publication record, rating this metric as much lower

priority than did faculty ($p = 0.0083$). LGBTQ+ students likewise had a strong preference for DEI and outreach work ($p = 0.0046$) and de-emphasized publication record ($p = 0.043$). This pattern is also observed in first-generation students (outreach $p = 0.048$, publication $p = 0.037$). Interestingly, the values of low socio-economic status students statistically align with faculty.

5.4.3 The Gap in Graduate Student and Principal Investigator Success Metrics

These data display the discrepancy between the metrics of success faculty apply to graduate student mentees and the priorities reported by graduate students in the Chemistry Department at University of California, Berkeley. UC Berkeley graduate students from backgrounds that are historically well-represented in STEM have goals and priorities that more closely align with the assessment metrics of faculty than students from URGs. Various underrepresented groups, including underrepresented people of color (URPOC), LGBTQ+ individuals and first-generation students value DEI and outreach work significantly more than such work is valued by faculty. This may be tied to the well-established phenomenon of "cultural taxation" or "invisible work", where URGs are asked and expected to spend significantly more time devoted to diversity-related mentorship and service often at the expense of time spent researching.[32–34] Underrepresented individuals must not only shoulder this burden of additional service, but also find their work unrecognized as essential by their advisors and peers.

The data presented here indicate that our graduate student population as a whole rejects the use of publishing as a primary indicator for graduate student success. This finding aligns with many challenges to hegemony of publishing Ph.D. evaluation.[8,35–39] We find this effect, however, is more pronounced for students from minoritized groups. Historically marginalized students are significantly less likely to agree with the statement "Publishing academic papers should be the primary metric to gauge graduate student success" than their well-represented peers. Almost every minoritized group of graduate students valued publication record significantly lower than faculty advisors. Considering the notable disparities in publication for women and BIPOC scientists, this challenge to publication as a graduate student's primary evaluative metric should be expected. Though other underrepresented groups such as LGBTQ+ students face discrimination and systemic inequalities in STEM, information about publication rate is unavailable due to a lack of data.[40,41] Taken together, these results demonstrate how 'common sense' modes of evaluation and success in STEM can contribute to exclusion and alienation of underrepresented individuals. URGs

in STEM must reconcile their values with the evaluation metrics of their faculty advisors, who mostly represent the dominant class (white, heterosexual, male). This dissonance of values can harm graduate students' sense of belonging in STEM, which is linked to lowered retention and achievement.[42,43] Indeed, the culture of publishing is understood to impact URG sense of belonging negatively at all levels of academia.[6] Meanwhile, our data suggest that non-URG trainees, on average, find their values more closely reflected in the values of their advisors. This mismatch in values may stem from the ambiguity in what a 'successful' outcome is for a graduate student.[37] Some may define success as the attainment of a PhD, while others may define success based on their career goal. Thus, the path to achieving each different success outcome will naturally necessitate different benchmarks and milestones throughout graduate school. It is well established that graduate students in Chemistry find research faculty careers to be most strongly encouraged by their advisors, despite waning interest in academic careers across STEM.[44] Yet this supported path is much less likely to be a desired path for women in STEM, and in particular women from underrepresented racial or ethnic groups prefer non-research careers following completion of their degree.[45]

We note that the culture and expectations of graduate students and faculty at a large research institution such as UC Berkeley may not be representative of those across Chemistry academia. However, graduate students from top-ranking institutions such as UC Berkeley produce a larger share of future faculty than other research institutions. This may affect the real and perceived career interests of graduate students and faculty towards academia, where publication metrics remain the strongest indicator for successful faculty appointment.[46] Thus, we hope this work inspires increased collection of community-level, quantitative data to explore the evaluative structure of Chemistry.

5.4.4 Building Inclusive Evaluation in Chemistry

Steps to remove this barrier to equity can take place both on the individual and institutional scale. Clear communication of expectations between graduate students and faculty advisors can ensure values of both parties are aligned, and has been shown to improve graduate student confidence and motivation.[47] Additionally, systematic feedback between mentees and mentors is known to empower graduate students to address areas of their experience that need improvement, while training in "Culturally Aware Mentoring" shows promise for teaching mentors to better serve mentees from minoritized populations.[48] Even the act of affirming one's own values can positively impact their sense of belonging.[49] Taken together, there are many practices that

individual research groups can adopt to ensure the goals and priorities of graduate student mentees are reflected throughout their training.

Deeper scrutiny of publishing and promotion in STEM via principles from CRT sheds light on how the current system upholds the power of the majority. Moving away from a publication-dominant values system may better serve groups with diverse priorities. However, the current structure of STEM academia may make it difficult for faculty to shift their focus away from publication as the dominant notion of success for themselves and for mentees. Publication metrics not only factor in hiring, promotion, and tenure of faculty, but are also correlated with successful grant funding.[50–52] Accumulation of funding enables financial support of more graduate student and postdoc researchers, leading to increased research output, establishing a publication "feed-forward loop".[53–57] Many funding agencies have incorporated criteria to emphasize mentorship and outreach, including the National Science Foundation and the Natural Sciences and Engineering Research Council of Canada Athena Scientific Women's Academic Network. This represents meaningful progress towards valuation of non-publication metrics, and further commitment to such efforts better supports marginalized researchers.

Alternative systems, such as those employed at the University Medical Center Utrecht, seek to rectify the "evaluation gap" by shifting the Ph.D. candidate evaluation such that it no longer focuses on publications, abstracts, or academic awards. Instead, assessment relies on a holistic self-evaluation of accomplishments of which the candidate is most proud, including those historically defined as "invisible work".[58] Recently, Utrecht University has extended this shift to the faculty level by replacing impact factor and h-indices in hiring and promotion decisions with standards such as commitment to teamwork and efforts to promote open science.[59] This initiative aims to shirk measures that "contribute to a 'productification' of science" and instead focus on factors that promote open science. Other avenues for reimagining scientific impact focus on centering the mentorship as an evaluation metric, whereby measures of acquired skills, self-efficacy and wellbeing of mentees may be quantitatively tracked using surveys.[60] Incorporating the values and voices of historically marginalized groups into a new academic model of success is key to establishing Chemistry as an equitable pursuit for all.

5.5 Acknowledgements

I would like to thank the entire Department of Chemistry at UC Berkeley for their continued commitment and support of the Climate Survey initiative. In particular, the Chemistry Graduate Life Committee (CGLC) and Diversity and Inclusion Focus

Group (DIFG) teams have offered invaluable input.

5.6 References

(1) Open Chemistry Collaborative in Diversity Equity (OXIDE) survey 2017–18.

(2) US National Science Foundation Survey of Earned Doctorates 2018.

(3) Wilson, M. A.; DePass, A. L.; Bean, A. J. *CBE Life Sciences Education* **2018**, *17*, DOI: 10.1187/cbe.17-09-0210.

(4) Moss-Racusin, C. A.; van der Toorn, J.; Dovidio, J. F.; Brescoll, V. L.; Graham, M. J.; Handelsman, J. *Science* **2014**, *343*, 615–616.

(5) Van Veelen, R.; Derks, B.; Endedijk, M. D. *Frontiers in Psychology* **2019**, *10*, DOI: 10.3389/fpsyg.2019.00150.

(6) Stachl, C. N.; Baranger, A. M. *PLOS One* **2020**, *15*, ed. by Blanch, A.

(7) Deiglmayr, A.; Stern, E.; Schubert, R. *Frontiers in Psychology* **2019**, *10*, Publisher: Frontiers, DOI: 10.3389/fpsyg.2019.01114.

(8) Weatherton, M.; Schussler, E. E. *CBE—Life Sciences Education* **2021**, *20*, DOI: 10.1187/cbe.20-09-0223.

(9) O'Shea, S.; Delahunty, J. *Higher Education Research & Development* **2018**, *37*, 1062–1075.

(10) McGee, E. O. *Educational Researcher* **2020**, *49*, Publisher: American Educational Research Association, 633–644.

(11) Wilson, L., *The Academic Man: A Study in the Sociology of a Profession*; Transaction Publishers: 1942.

(12) Carpenter, C. R.; Cone, D. C.; Sarli, C. C. *Academic Emergency Medicine* **2014**, *21*, 1160–1172.

(13) Abbott, A.; Cyranoski, D.; Jones, N.; Maher, B.; Schiermeier, Q.; Van Noorden, R. *Nature* **2010**, *465*, 860–862.

(14) Hirsch, J. E. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 16569–16572.

(15) Sahel, J.-A. *Science Translational Medicine* **2011**, *3*, 84cm13.

(16) Egghe, L. *Scientometrics* **2006**, *69*, 131–152.

(17) Edwards, M. A.; Roy, S. *Environmental Engineering Science* **2016**, *34*, 51–61.

(18) Larivière, V.; Ni, C.; Gingras, Y.; Cronin, B.; Sugimoto, C. R. *Nature News* **2013**, *504*, 211.

(19) Bendels, M. H. K.; Müller, R.; Brueggmann, D.; Groneberg, D. A. *PLOS ONE* **2018**, *13*, e0189136.

(20) Ginther, D. K.; Basner, J.; Jensen, U.; Schnell, J.; Kington, R.; Schaffer, W. T. *PLOS One* **2018**, *13*.

(21) Thompson, J. J.; Jensen-Ryan, D. *CBE—Life Sciences Education* **2018**, *17*, ed. by Schinske, J.

(22) Beasley, M. S.; Lumley, M. A.; Janicki, T. D.; Fernandez, R. L.; Manger, L. H.; Tucholski, T.; Thomas, N. C.; Whitmire, L. D.; Lawson, A.; Buller, A. R. *Journal of Chemical Education* **2020**, *97*, 643–650.

(23) Mousavi, M. P. S.; Sohrabpour, Z.; Anderson, E. L.; Stemig-Vindedahl, A.; Golden, D.; Christenson, G.; Lust, K.; Bühlmann, P. *Journal of Chemical Education* **2018**, *95*, 1939–1946.

(24) Stachl, C. N.; Hartman, E. C.; Wemmer, D. E.; Francis, M. B. *Journal of Chemical Education* **2019**, *96*, 2149–2157.

(25) Stachl, C. N.; Brauer, D. D.; Mizuno, H.; Gleason, J. M.; Francis, M. B.; Baranger, A. M. *ACS Omega* **2021**, *6*, 14410–14419.

(26) Bourdieu, P. In *Power and Ideology in Education*; Oxford University Press: New York, 1971, pp 487–511.

(27) Bourdieu, P. In *Knowledge, education and cultural change: Papers in the sociology of education*; Taylor & Francis: London, 1973, pp 71–112.

(28) Winkle-Wagner, R. *ASHE Higher Education Report* **2010**, *36*, 1–144.

(29) Yosso, T. J. *Race Ethnicity and Education* **2005**, *8*, 69–91.

(30) Ledesma, M. C.; Calderón, D. *Qualitative Inquiry* **2015**, *21*, Publisher: SAGE Publications Inc, 206–222.

(31) Crenshaw, K. *Stanford Law Review* **1991**, *43*, 1241.

(32) Wijesingha, R.; Ramos, H. *Canadian Journal of Higher Education / Revue canadienne d'enseignement supérieur* **2017**, *47*, 54–75.

(33) Jimenez, M. F.; Laverty, T. M.; Bombaci, S. P.; Wilkins, K.; Bennett, D. E.; Pejchar, L. *Nature Ecology & Evolution* **2019**, *3*, 1030–1033.

(34) Group, S. S. F. N. R. I. *Humboldt Journal of Social Relations* **2017**, *39*, 228–245.

(35) Yeung, N. *Nature Human Behaviour* **2019**, *3*, 1036–1036.

(36) Coriat, A.-M. *Nature Human Behaviour* **2019**, *3*, 1007–1007.

(37) Hartnett, R. T.; Willingham, W. W. *Applied Psychological Measurement* **1980**, *4*, 281–291.

(38) Wilsdon, J.; Allen, L.; Belfiore, E.; Campbell, P.; Curry, S.; Hill, S.; Jones, R.; Kain, R.; Kerridge, S.; Thelwall, M.; Tinkler, J.; Viney, I.; Wouters, P.; Hill, J.; Johnson, B. **2015**.

(39) Harshman, J. *Journal of Chemical Education* **2021**, *98*, 259–269.

(40) Cech, E. A.; Waidzunas, T. J. *Science Advances* **2021**, *7*, eabe0933.

(41) Freeman, J. B. *Policy Insights from the Behavioral and Brain Sciences* **2020**, *7*, 141–148.

(42) Tibbetts, Y.; Harackiewicz, J. M.; Canning, E. A.; Boston, J. S.; Priniski, S. J.; Hyde, J. S. *Journal of personality and social psychology* **2016**, *110*, 635–659.

(43) Hurtado, S.; Alvarado, A. R.; Guillermo-Wann, C. *JCSCORE* **2015**, *1*, 59–81.

(44) Sauermann, H.; Roach, M. *PLOS One* **2012**, *7*.

(45) Gibbs, K. D.; McGready, J.; Bennett, J. C.; Griffin, K. *PLoS ONE* **2014**, *9*, ed. by Launois, P.

(46) Van Dijk, D.; Manor, O.; Carey, L. B. *Current Biology* **2014**, *24*, R516–R517.

(47) Hund, A. K.; Churchill, A. C.; Faist, A. M.; Havrilla, C. A.; Stowell, S. M. L.; McCreery, H. F.; Ng, J.; Pinzone, C. A.; Scordato, E. S. C. *Ecology and Evolution* **2018**, *8*, 9962–9974.

(48) Byars-Winston, A.; Womack, V. Y.; Butz, A. R.; McGee, R.; Quinn, S. C.; Utzerath, E.; Saetermoe, C. L.; Thomas, S. B. *Journal of Clinical and Translational Science* **2018**, *2*, 86–94.

(49) Tibbetts, Y.; Harackiewicz, J. M.; Priniski, S. J.; Canning, E. A. *CBE Life Sciences Education* **2016**, *15*, DOI: 10.1187/cbe.16-01-0001.

(50) Venable, G. T.; Khan, N. R.; Taylor, D. R.; Thompson, C. J.; Michael, L. M.; Klimo, P. *World Neurosurgery* **2014**, *81*, 468–472.

(51) Rezek, I.; McDonald, R. J.; Kallmes, D. F. *Academic Radiology* **2011**, *18*, 1337–1340.

(52) Neumann, Y. *Research in Higher Education* **1978**, *9*, 115–122.

(53) Kaatz, A.; Lee, Y.-G.; Potvien, A.; Magua, W.; Filut, A.; Bhattacharya, A.; Leatherberry, R.; Zhu, X.; Carnes, M. *Academic Medicine* **2016**, *91*, 1080–1088.

(54) Taffe, M. A.; Gilpin, N. W. *eLife* **2021**, *10*.

(55) Győrffy, B.; Herman, P.; Szabó, I. *Journal of Informetrics* **2020**, *14*, 101050.

(56) Rosenbloom, J. L.; Ginther, D. K.; Juhl, T.; Heppert, J. A. *PLOS One* **2015**, *10*, ed. by Amaral, L. A. N.

(57) Jacob, B. A.; Lefgren, L. *Journal of Public Economics* **2011**, *95*, 1168–1177.

(58) Changing recognition and rewards for PhD candidates | Universiteit Utrecht.

(59) Woolston, C. *Nature* **2021**.

(60) Davies, S. W. et al. *PLOS Biology* **2021**, *19*, DOI: 10.1371/journal.pbio.3001282.