Title

Bioconjugation and Delivery of Cas9 RNPs

Author

Lobba, Marco Jackson

Publication Date

2020

Peer reviewed|Thesis/dissertation

Bioconjugation and Delivery of Cas9 RNPs

By

Marco Jackson Lobba

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, Co-Chair
Professor Jennifer A. Doudna, Co-Chair
Professor David Savage

Fall 2020

Bioconjugation for the delivery of Cas9 RNPs

Marco Jackson Lobba

Abstract:

Bioconjugation and Delivery of Cas9 RNPs

By

Marco Jackson Lobba

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Matthew B. Francis, Co-Chair

Professor Jennifer A. Doudna, Co-Chair

The advent of CRISPR based gene editing has transformed both basic scientific research and the development of new medicines and treatments. These complex proteins promise freedom to modify any gene at will but must be carefully delivered to prevent degradation or off-target editing. This dissertation describes work to encapsulate Cas9 within the bacteriophage MS2 protein scaffold, and later to modify the Cas9 protein itself to allow for more efficient delivery and editing. Cas9 was successfully encapsulated within the MS2 VLP, but showed high sensitivity to RNase degradation.

Towards the modification of the Cas9 ribonuclear protein, a novel protein modification reaction using the enzyme tyrosinase was developed, enabling modification and delivery of Cas9 using cationic delivery peptides. The tyrosinase reaction was further used to fuse a variety of proteins to immunoglobulin derivatives including scFvs, nanobodies, and IgG1 Fc domains. Cas9-immunoglobulin fusions were shown to induce tolerance in *in vivo* mouse model studies, highlighting the potential of tyrosinase coupling to mediate Cas9 protein fusions. Prokaryotic tyrosinases were explored for their potential as universal protein activators, of which the tyrosinase from *Bacillus megaterium* showed exciting promise as a universal activator. Using substrate charge preference data from the tyrosinase screens we were able to design charged tyrosine tags that enabled sequential activation and coupling of protein trimers.

Dedicated to my parents Mary and Alex, and to all my friends without whom I could not have done this.

Contents

Table of Abbreviations

Abbreviation	*Definition*
2PCA	2-pyridine carboxaldehyde
abTyr	Agaricus bisporus Tyrosinase
abTyr	tyrosinase from Agaricius bisporus
APC	Antigen Presenting Cell
bmTyr	tyrosinase from Bacillus megaterium
Cas	CRISPR associated protein
cnkTyr	tyrosinase from Candidatus Nitrosopumilus koreensis
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR inhibition, dCas9 is used to bind a gene of interest to turn down it's expression
crRNA	CRISPR RNA
dCas9	Dead Cas9- mutations in the HNH and RuvC domains remove DNA cleavage ability
EPR	Enhanced Permeability and Retention
ESI	Electrospray Ionization
Fc	Immunoglobulin Crystallizable chain
gRNA	guide RNA
IgG	Immunoglobulin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MHC	Major Histocompatability Complex
MS	Mass Spectrometry
MS2	MS2 bacteriophage
NCL	Native Chemical Ligation
NHS	N-hydroxy-succinimide
NLS	Nuclear Localization Sequence
o-quinone	ortho-quinone
RNP	RiboNuclear Protein
SaCas9	Cas9 from Staphylococcus aureus
SpyCas9	Cas9 from S. pyogenes
SV40	Simian Vacuolating Virus 40
TALENS	transcription activator-like effector nucleases
TEM	Transmission Electron Microscopy
TOF	Time of Flight
tracrRNA	trans acting cleavage RNA
UAA	Unnatural Amino Acid
VHH7	MHC Class II targeting nanobody
VLP	Virus Like Particle
WT	Wild Type
ZFNs	Zinc Finger Nucleases

Acknowledgements

I would like to thank Matt and Jennifer for agreeing to co-mentor me on the absurd project of shoving Cas9 inside of MS2 like a miniature turducken. Despite the failure of our initial project I have enjoyed my time in both labs and am tremendously grateful for all the support and mentorship I have received.

Matt, thank you for your unending support and willingness to shovel through strange experimental results. I am very lucky to have worked with someone who values the extracurricular work on diversity I spent so much time on. I hope someday you finally hire that lab manager.

Jennifer, thank you for your continued support for every strange idea I brought to you and for being such an excellent role model on how to do science on a global stage. It has been such a pleasure seeing the lab evolve and grow over the past 5+ years.

Kaihong, the Doudna lab would not function without you. Thank you for helping run so many RNA gels and for letting me steal your gel guides so often.

The Francis Lab- Thank you for being my home base at Berkeley and for all the support and friendship you've all given me during the past few years. You guys helped me when projects failed and life was hard and I can't thank you enough. I will deeply miss the ski trips and summer retreats with you all so much. Special thanks to Emily for letting me collaborate on projects you could have done on your own when my work wasn't going well. Thanks to JCM for also being one of the kindest people I know and letting me collaborate on the tyrosinase work that has been such a huge part of my PhD. Thanks to Ioana for teaching me most of what I know about protein expression and how to be a good scientist. Vanessa, thanks for climbing half dome with me and in general just being a good ear to talk to. I still can't believe we did it! To Sarah, Ayo, Jing, Matt, Kristin, Celine, and everyone who has left the lab I hope to see you in the "real" world soon. To everyone still in the lab: Nick, Dan, Alex, Jamie, Paul, Derek, Wendy, Amanda, Paige, and Deborah I hope you all manage to keep up hope with the craziness that is lab with COVID. You're an amazing group of people and it's been my pleasure to work with you all.

The Doudna lab- Thank you all for your help and patience with the very lost chemist trying to do biology. Special thanks to Christof who went above and beyond to try and help me make the MS2 project work in cells. Thanks to Brett, Cole, and Abby for helping to run all the neural progenitor cell trials, I would have accomplished only a fraction of this work without you. To Gavin, Alex, Janice, Josh, Jenny, Brady and all the others who have been friends and mentors along the way I miss you all and hope we can end up in Hawaii again. To the newer members: Joy, Amy, Hannah, Basem, Riley, Cindy, Connor and everyone else you are so damn impressive and I can't wait to see where you go from here.

To the excellent students I have had the pleasure of mentoring: Florin I hope you continue to be just as hilarious and excited about science. Andrew, your ability to learn new material on your own is greater than anyone I know. Steph, thanks for rescuing Taytay with me, I can't wait to see what you do in your career. Casey, you are endlessly creative and enthusiastic, it has been a

pleasure working with you and I hope you never lose that spark. I could not be more proud of what you all have done. Thank you all for your work and for your trust in my dubious leadership.

To the rotation students I mentored: Isaac and Eli, best of luck out there, I know you will be awesome scientists.

Thanks to everyone from Pomona who helped get me into science. In particular thank you Mal for mentoring me and your patience with my at times glacial research. A huge thanks to Ted who pulled me into research in the first place and taught me that even an undergrad can be the most knowledgeable person in the room. Thanks for Prof. Sea, Prof. Taylor, Prof. O'Leary, Prof. Grieman and all the other amazing scientists I got to learn from. A special shout out to Prof. Seligman who once told me in a prescient moment back in 2010 that "whoever finds a protein that can target any piece of DNA deserves a Nobel Prize."

I would not be where I am today without the love and support of my parents Alex Lobba and Mary Jackson. You always encouraged my love of science since I first started planning to empty out black holes as a toddler. Thank you both for always supporting me and encouraging to go after my passion, you have led by example in your own lives and I cannot imagine a better family to be a part of.

To my sister Emily, thank you for humoring me and at least pretending to be excited about the boring research I always talk about. I will never stop being impressed and amazed by your strength and determination. Growing up together taught me so much about how to keep fighting even when the world doesn't seem to go your way. It has been such a pleasure to see your life's journey so far and I can't wait to see where it goes.

To Maile, thank you for accepting me for who I am. I could not have asked for a better quarantine partner. Thank you for supporting me in the writing of this thesis and bringing me so many snacks. I can't wait to go on more adventures with you.

To my many friends that supported me throughout my PhD, thank you for always being there for me, especially in my darkest moments. Getting divorced during a PhD was not something I had planned for and I really cannot imagine where I would be without you all. I have been blessed by the most supportive friends a person could ask for. From climbing trips, to skiing, to concerts, to the emergency room, and beyond I have had an amazing crew by my side throughout this experience. Thanks homies.

Chapter 1. Delivery and modification of CRISPR Proteins

Abstract

The advent of CRISPR proteins in the past decade has been arguably the most impactful scientific discovery of this author's lifetime and has had a profound impact on biology, healthcare, and even disease detection. The diversity and scope of CRISPR proteins is truly staggering and new applications are being discovered daily, but the most highly lauded application is their potential for gene editing. One of the greatest challenges to turning these fundamental discoveries into therapies has been forcing these CRISPR proteins into cells in the first place. The sheer size of the CRISPR affecter proteins, which range from under 800 amino acids to over 2,000 and require an RNA guide, makes cellular entry a non-trivial matter as most conventional methods of delivery break down at this scale. This review covers the discovery and basic biology of the CRISPR systems and compares the various CRISPR proteins for their suitability as gene editing scaffolds. We then discuss delivery methods for CRISPR proteins and their benefits and drawbacks, before exploring protein modification as a means of screening protein delivery.

1.1 The CRISPR system and CRISPR Biology

Gene therapy has emerged in the 21st century as a possible cure for many genetic disorders with the advent of advanced genetic sequencing. The renewed interest stems largely from a flurry of new methods that enable the precise modification of disease-causing genes without altering the rest of the genome, a process commonly referred to as gene editing.[1] Prior to the last decade, most proposed gene editing methods were either insufficiently specific for therapeutic application as in the case of most naturally occurring endonucleases,[2,3] or required significant engineering for each new target as in the case of transcription activator-like effector nucleases (TALENS) or zinc finger nucleases (ZFNs). [1,2,4,5]

In the 1980's the first signs of a strange repeating sequence in the genome of various bacteria was identified by Ishino el al.[6] It was not until the mid-2000s when several groups showed that these Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) regions actually contained fragments of phage genomes and the cluster was revealed to be a miniature immune system with memory of past infections.[7–9] Originally believed to be a form of bacterial RNA interference (RNAi) which is used in eukaryotes to modulate gene expression. However, examination of the flanking proteins near these CRISPR arrays soon revealed a complex system involved in the targeting, cleavage, and capture of phage genomes.[10] CRISPR associated (Cas) proteins come in a wide variety of sizes and shapes, each with a different function. Some of the earliest identified CRISPR systems (Type I and Type III) use a host of smaller proteins which together in complex are able to bind, recognize and cleave a target DNA strand.[11] However, some of the most puzzling results came out of the Type II CRISPR systems, which contained exceedingly large proteins of unknown function. The Banfield lab at UC Berkeley was one of the early pioneers in the field,[12,13] and a conversation between the two PIs is what led the Doudna lab to study the large Type II effector proteins that would eventually lead to the CRISPR gene editing platform.

In 2012 Martin Jinek found that Cas9 from *Streptococcus pyogenes* (SpyCas9) was capable of cleaving a target strand of DNA based off of the target sequence in the CRISPR array.[14] He showed that Cas9 uses a two-component RNA: a structural CRISPR RNA (crRNA) component which stabilizes the Cas9 protein, and a trans-activating crRNA (tracrRNA) that is derived from the CRISPR array and determines the nuclease target.[15] It is now appreciated that Watson-Crick base pairing is used to determine sequence homology of the interrogated DNA strand with the 20 base long guide where sequence matching is required for nuclease activity, and only limited mismatches are tolerated.[3,16–18] Martin's work in the Doudna lab reengineered the crRNA and tracrRNA into a single strand, known as the single guide RNA (sgRNA).[14] This sgRNA can be expressed easily *in vitro* from a DNA oligomer.[3] The great benefit of the CRISPR gene editing system is that, rather than reengineering the effector protein for each new target, one need simply alter the sgRNA to create highly specific cleavage sites anywhere within the genome.

Nearly all CRISPR proteins also require a constant protospacer adjacent motif (PAM) which helps to prevent cleavage of the host genome.[16,19,20] PAM binding is required in addition to guide RNA complementarity for nuclease activity and is encoded in the sequence of the protein rather than the RNA. In SpyCas9 the PAM is the three base sequence NGG, where N

represents any base, which must be located 2 bases away from the start of the target sequence.

The wildtype Cas9 protein from *S. pyogenes* is by far the most widely employed and best studied version of this system, but many homologues have been discovered in similar organisms with different PAM requirements and activities. Other, smaller versions, of Cas9 have been found recently, but few have been thoroughly vetted and so SpyCas9 remains the most commonly employed tool.

More recently it has been appreciated that Cas9 might also contribute to acquisition and trimming of the CRISPR sequences. A newly submitted manuscript by Jakhanwal et al. shows the role of Cas9 in trimming candidate sequences to size for incorporation into the CRISPR array. Interestingly, this activity is independent of the guide RNA target and requires only the Cas9 HNH domain be active. The author is honored to have helped characterize complexes of Cas1 and Cas2 with Cas9 in a recently submitted work that is cited here but will not be further discussed.

Since 2012 a number of different Cas effector proteins have been discovered using data on microbial communities collected by the Banfield lab and others. Many of these proteins share the common feature with Cas9 of being a single protein capable of binding, recognizing, and cleaving a target nucleic acid sequence, but vary in their structure enough to warrant new characterization. Makarova et al have been diligent in cataloguing these new proteins which have been designated Cas12, 13, and 14.[21–26] More recently, CasΦ has even been discovered in giant phage.[27]

Most relevant between the various Cas proteins is their cleavage behavior. Cas9 always uses both an HNH domain and a RuvC nuclease to create opposing cut sites with a blunt terminus.[28,29] In contrast, Cas12 uses a single HNH domain which cleaves the target strand and then folds over to cleave the opposing strand.[23] This flexibility of cleavage typically results in an overhang cleavage junction, typically considered preferable for homology directed repair of genes. Also of interest is the family of Cas13 proteins which target RNA instead of DNA.[26]

Another of the exciting discoveries from the microbial community databases has been that of the antiCRISPR proteins. Used by phage to evade CRISPR destruction, these small proteins were first identified in *Pseudomonas aeruginosa* in phage capable of replication even in the presence of CRISPR systems.[30] They have since been identified from genomes which held self-targeting sequences in their CRISPR array which should render the host non-viable.[31] This contradictory nature implied the existence of proteins capable of mediating the activity of Cas proteins. Since the first discovery of antiCRISPRs over 50 have been identified, affecting most known Cas effector proteins.[32] A number of Cas12-targeting anti-CRISPRs have been identified with widely diverging mechanisms of action. This author was fortunate to have contributed to the study of complexes formed from AcrVa4 on Cas12 and the unique dimerization effect found in this class of anti-CRISPR.[33,34] These works are referenced here for the reader but will not be further discussed.

1.2 CRISPR for gene editing

Cleavage sites can either be blunt ended or with an overhang. The difference between the two has important ramifications for how they are repaired in eukaryotes. All eukaryotes

rely on two primary repair pathways for the remediation of double stranded breaks, Homology Directed Repair (HDR) and Non Homologous End Joining (NHEJ).[14,35] In the presence of a complementary template and dependent on a variety of factors including the bases surrounding the cut site, cells can repair the cut in the DNA with no trace of the cleavage. HDR is also favored in the case of overhang cut sites.[36,37] This can be used to repair faulty genetic sequences by providing a corrected template to repair off of which can result in a swap of faulty bases. This has been further modified in the case of PRIME editing developed by the Liu lab, where the repair template is included as an extension of the guide RNA and a DNA polymerase is attached to the Cas9 to facilitate the repair.[38] In contrast, NHEJ has no template to use for repair, as a result the repair can produce insertions of 0, 1, or two bases. These insertions are generally random and result in a frame shift mutation, disrupting the register of the gene, effectively turning it off.[35,39]

Because NHEJ results in such disruptive mutations, one of the most important factors in evaluating the potential of Cas proteins as therapeutics is the frequency of off-target activity. Given the requirement of 20 bases and a PAM it is quite possible to identify single sites within most genes that are not repeated elsewhere in the genome for targeting. The problem arises from the fact that Cas9 has variable tolerance for base mismatches based on the position within the target sequence.[18] Generally speaking, mismatch tolerance is lowest near the PAM and around the cleavage site, but increases towards the PAM-distal end of the target region.[40] Since the tail of the target site tolerates mismatches there is a higher chance for off-target cleavage events than might be assumed for the 20 base target provided by the RNA. Most importantly, this means that certain targets may have higher potential for off-target cleavage based on their similarity to other regions within the host genome.[41,42] A number of tools have been developed to evaluate optimal target sites with minimal off-target capability, but the concern of cancer-causing mutations and other deleterious effects remains one of the greatest hurdles to widespread adoption of CRISPR based gene editing.

To minimize the potential for off-target cleavage a number of labs have made modifications to Cas9 around the recognition site to improve the fidelity of the enzyme. In general these modifications have focused on slowing down the cleavage step or k_{cat} of the enzyme.[17] By slowing the rate of cleavage it is possible to ensure that only sites fully bound by the guide RNA are cleaved, as off-target sites will generally be released prior to cleavage. Chen et al have showed the benefit of targeted mutations around the DNA-RNA duplex formed during target recognition and the improvements in specificity that these allow.[17] Interestingly, only alanine substitutions have been thoroughly explored in these studies and a more thorough screen of other amino acid substitutions would likely provide useful data and could yield Cas9 mutants with even greater selectivity.

1.3 Delivery of CRISPR therapeutics

Delivery *in vivo* is a daunting challenge for any therapeutic. Given the rapid degradation of proteins and nucleic acid by the digestive system the major delivery approaches are through inhalation or injection.[43] Given the generally low efficiency of inhalation for delivery of protein cargo it has not been thoroughly explored for CRISPR therapeutics and will not be further discussed in this review, though it could prove attractive once the technology matures.[44] The

vast majority of biological therapeutics are given via injection which can be administered either intravenously through the bloodstream or by direct injection to the target tissue.[45,46] Intravenous delivery allows for broader circulation of the therapeutic and can provide access to regions that would otherwise be hard to target by direct injection.[47] Given the widespread use of intravenous delivery it is generally the preferred approach due to ease and the ability for patients to self-administer if needed. The primary challenge in delivering a therapeutic through the bloodstream is the host's immune system which is often highly efficient at clearing foreign molecules from circulation.[48,49] Even if a therapeutic can reach the target tissue, it must then exit the bloodstream, bind, then enter the target cells.[50,51] In contrast, direct injection is able to provide a concentrated dose of the therapeutic to the site of interest, removing the need to survive in the bloodstream and allowing formulation that focuses on targeting and entering desired cells. The downside with this approach is that only the immediate area around the site of injection is affected, requiring many injections or preventing complete coverage of the desired tissue.[52]

One alternate delivery approach that is widely considered in the case of CRISPR therapeutics is *ex vivo* modification of cells. Where possible, this allows for removal of patient tissue, modification, and screening before the cells are returned to the patient. This has the benefit of ensuring that off-target edits will not compromise the modified cells.[53,54] For *ex vivo* delivery of both nucleic acids and proteins cationic lipid delivery, often complemented by electroporation, is considered the gold standard. Lipofectamine and its various analogues present an efficient cell entry platform, but the lack of specificity, rapid clearance from blood, and potential toxicity have limited *in vivo* applications of this technology.[55,56] However, the high efficiency and ease of use mean that this technique is often used as a benchmark point of comparison in delivery experiments.

Given the astounding size of most Cas proteins and the requirement for a highly anionic RNA guide, delivery has been a major impediment to the development and study of CRISPR as a gene editing platform. Approaches for CRISPR delivery can be divided into delivery of the protein and RNA complex (RNP) and delivery of the nucleic acid sequence encoding the protein and the requisite guide RNA.[11,57,58]

Most *in vivo* approaches near clinical acceptance have used modified viral vectors as carriers for DNA or RNA encoding both the Cas9 protein and the requisite sgRNA, which are then expressed in the infected cell.[1,14,59,60] This approach has the benefit of high delivery efficiency and has the potential for tissue specificity. Most viruses have evolved to evade immune detection and so generally have good circulation times and natively possess the capability of targeting and entering specific cell types. Unfortunately, viral transfection suffers from a number of drawbacks, including size restrictions, dosage, and temporal control.[16] The size restriction is a major impediment, as the most common viral vector for delivery is the adenovirus.[35,61] These well studied viruses are highly efficient and provoke minimal immune response, but they have a genome of only four to five kilobases which is barely enough room for a single Cas protein and allows for no excess cargo.[35] Dosage is another significant problem, as multiple studies have demonstrated that many active copies of Cas9 are produced through viral infection. This in turn leads to off-target cleavage events, increasing the risk of cancer or other diseases.[59,62] More recently, it has been appreciated that viral delivery poses significant

immunological risk to the patient due to the potential for rejection of the cells or tissues that have been edited.[53,63] Together, these drawbacks have limited Cas9 delivery via modified viral vectors to either tissues removed from the body or immunoprivileged regions such as the eye.[64–66]

In recent years several other approaches for delivering Cas9 encoding nucleic acids to cells have emerged as exciting alternatives, carbon nanotubes from the Landry lab and gold nanoparticles from the Murthy lab. Though untargeted, the Landry lab has shown groundbreaking efficiency in their nanotube delivery platform which offers improved rates of delivery even to plant cells but currently lacks cell specificity.[67] The gold nanoparticle approach is similarly untargeted but has seen broader use in gene delivery into a variety of plant and animal systems.[68,69]

Due to the challenges with nucleic acid delivery for CRISPR therapeutics many in the field have focused efforts around delivery of the entire Cas9 and sgRNA ribonuclear protein (RNP) complex to targeted cells rather than rely on transfection/transformation and expression.[70] This reduces the possibility of off-target cleavage events and increase temporal and dosage control while minimizing the chance of rejection.[16]

At the start of this project there were 3 primary delivery paradigms under investigation: delivery of Cas9 protein alone with enhanced uptake through incorporation of cell penetrating peptides or other groups;[57] delivery mediated by charged lipid vesicles,[71] and Cas9 delivery in a protein or polymer cage.[64,72] Because RNase is present in the blood at up to 1 mM concentrations, there is significant difficulty delivering functional Cas9 to cells *in vivo* for therapeutic purposes.[57] Additionally, Cas9 suffers from exceedingly poor pharmacokinetics since it is rapidly removed from the blood which complicate direct delivery of the RNP.[73] Interestingly, recent work from the Wilson lab has shown potential for Cas9 RNP delivery through attachment of various targeting groups.[74–76] However, these approaches still suffer from RNase degradation and poor pharmacokinetic profiles which ultimately limit their applicability for intravenous delivery. Guide RNA modifications to improve stability and reduce the impact of RNase degradation are being explored in most RNP delivery approaches but cost of production currently limits the applicability of these approaches in the majority of cases.[76]

In contrast, direct injection to target tissues continues to be one of the most promising applications for RNP constructs. Work by Staahl et al has shown great promise in delivering Cas9 with multiple copies of the cationic nuclear localization sequence (NLS) of the Simian Vacuolating virus (SV40).[52] This work has shown that neurons and neural progenitor cells are particularly amenable to delivery of CRISPR proteins via cationic peptide fusion. Unfortunately, this technique seems restricted to neural cells as no other cell lines have shown significant editing with similar constructs. This is likely due to the high polarization of the neural membrane, which indicates that cardiac cells might offer another potential application area for these NLS tagged constructs.

Cationic lipid mediated delivery has shown very promising results in *ex vivo* cell delivery studies,[71] however, most lipid delivery systems are limited to the site of injection and/or show poor tissue specificity *in vivo.*[49,77] Here the challenge is tuning lipid delivery profiles, as the rapid fusion with cell membranes drastically reduces the diffusion of the vesicles from the site of administration.[78,79]

Given the difficulties associated with the methods discussed above, protein encapsulation remains a promising line of delivery for the Cas9 RNP. Encapsulation offers superior protection from immune system interrogation,[80,81] improved stability, and enhanced circulation profiles depending on the nature of the encapsulating matrix.[82,83] Of particular interest is the ability to protect the guide RNA from degradation without resorting to prohibitively exotic and expensive RNA derivatives. A further benefit of encapsulation is their ready with targeting groups or immunosuppressant molecules to allow for superior targeting to a variety of cell types or tissues.[84–86]

In the past year Cas9 RNP encapsulation and delivery have been taken into lentiviral capsids in a recently submitted publication by Hamilton et al. and from Lyu et al.[87] This work shows the potential of encapsulation for delivery of Cas9 to target cells. The ability to produce the constructs without the need to first isolate and purify the various components presents a distinct advantage over other methodology discussed in this work. Furthermore, the use of the lentiviral shell offers excellent protection from immune surveillance, and their ability to serotype diverse receptor targets in place of the native binders has shown promising results in the targeting and delivery to various cell types.

1.4 Bacteriophage MS2 and the MS2 virus-like-particle

The bacteriophage MS2 is a well-studied example of RNA bacteriophages prone to infecting *E. coli*. In its wild-type state the MS2 viral capsid contains a membrane insertion domain, and 177 copies of a single coat protein subunit which self-assemble around the RNA genome of the virus.[88] Self-assembly is catalyzed by the MS2 hairpin binding sequence which acts as a nucleation site and is believed to cause a conformational change in the bound MS2 coat protein that facilitates capsid formation.[89–91] The tight binding to the MS2 hairpin has even been used by some groups as a method for binding two proteins together,[92] or for tracking RNA within cells.[93] Early versions of CRISPRi and CRISPRa used MS2 hairpin extensions in the guide RNA sequence of a dead Cas9 (dCas9) to recruit inhibitory or activating proteins fused to MS2 monomers to the gene or region of interest without cleaving the DNA target.[94] More recently, the Wilson lab has used MS2 hairpin extensions of the sgRNA to recruit delivery and targeting groups to Cas9 RNPs.[76]

Within the MS2 capsid, each subunit of has been shown to have some low affinity for other hairpins throughout the RNA genome, facilitating the winding of the genome throughout the entire capsid as revealed by recent TEM images.[91,95] A useful property of the MS2 bacteriophage is that the coat protein alone can be expressed recombinantly in *E. coli* where it self-assembles into virus-like-particles (VLPs), or particles that maintain the structural characteristics of the virus without a competent infection system. Crucially, the MS2 coat protein is capable of forming a capsid around endogenous mRNA transcripts within the cytosol of the bacterium expressing it, allowing for facile production and purification.[96] This mRNA binding ability facilitates screening of MS2 libraries as selections can be applied to a diverse set of constructs followed by mRNA extraction, reverse transcription and sequencing.[96] The author is honored to have contributed to the development of this technique, called Systematic Mutation and Assembled Particle Selection (SyMAPS) which allows for the rapid profiling of

assembly-competent MS2 mutants. It has been employed to screen stability for every single amino acid mutant of the MS2 coat protein, allowing for identification of a number of promising mutants with higher acid sensitivity or reactive groups on the exterior.[96] The development of SyMAPS is catalogued in other works and will not be included here for the sake of brevity.[96,97]

MS2 VLPs are approximately 27 nm in diameter, each with 32 pores (2 nm in diameter) and a 17 nm interior cavity.[86,98] Their robust formation and high stability have made them an attractive scaffold for protein engineering. The Francis lab and others have shown that the capsid is highly tolerant to modification, dimerization of the coat proteins, and even artificial amino acid incorporation. In particular, incorporation of the *p*-aminophenylalanine (*p*aF) group into position T19 of the MS2 coat protein produces capsids with 180 aniline groups on the surface,[99,100] well suited for oxidative coupling of small molecules and proteins up to full antibodies.[51,84] The interior of the MS2 capsid can be similarly modified after incorporation of a cysteine at position N87, allowing for facile attachment of maleimide or other cysteine reactive groups. These two modification approaches have been used to create an array of cargo delivery vehicles for various therapeutic applications from anti-cancer drugs to gold nanoparticles.[cite] Additionally, the ability to attach antibodies to the capsid surface allows for targeting of arterial plaques[101] and potentially other targets. Fused dimers of the MS2 coat protein have been shown to be more tolerant to peptide insertion.[102–104] This characteristic has been employed by several researchers to insert peptides into exposed surface loops for the purposes of peptide display.[104] Combination of this with the SyMAPS screening could yield diverse candidates beyond those identified in the exciting work by Robinson et al.[105]

MS2 VLPs are also an attractive candidate for drug delivery due to their excellent pharmacokinetic properties.[82] Generally remaining in the blood stream for over 24 h, the MS2 VLP is large enough to avoid filtration by the kidney and benefits from a lack of strong innate immune response since it does not naturally target human cells.[106] MS2 VLPs have been used for their high storage capacity for toxic cargo and have been successfully targeted to atherosclerotic plaques. Studies have shown that the MS2 VLP might also benefit from the Enhanced Permeability and Retention (EPR) effect, due to their size they are just small enough to transport into tumors due to the increased permeability of the vasculature within the tumorous tissue, allowing for a more targeted delivery of chemotherapeutic cargo.[82]

Another beneficial property of MS2 VLPS is that they can encapsulate a variety of negatively charged molecules from drugs to gold nanoparticles.[85,107–109] MS2 capsids readily disassemble in the presence of concentrated acetic acid, allowing for removal of the MS2 genome or mRNA, and introduction of new cargo in a controlled environment.[110] Capsid formation occurs rapidly around negatively charged particles in the presence of the stabilizing agent TMAO (Figure 1) and proceeds overnight with little loss of the coat protein monomers. This capability was utilized by Glasgow et al to encapsulate negatively charged enzymes and shows great promise as a potential protein delivery method.[107]

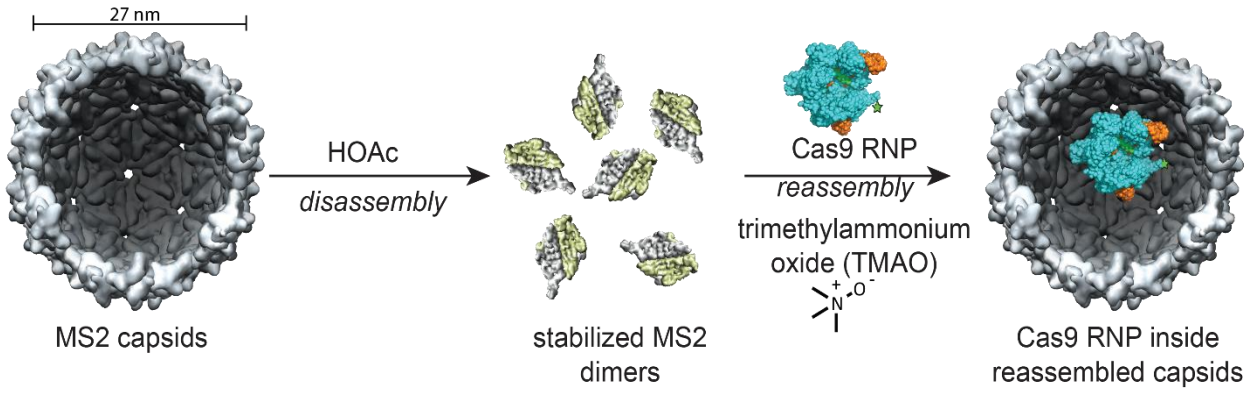


Figure 1: Encapsulating Cas9 within the MS2 viral capsid. MS2 capsids were dissassembled using acetic acid into soluble dimers. Then, in the presence of the stabilizing agent TMAO, capsid formation occurs around negatively charged particles such as RNA or polymers. The presence of the MS2 hairpin may further aid in capsid formation by creating a specific nucleation point for binding to the MS2 monomer.

1.5 Further engineering of CRISPR proteins

The high specificity binding of Cas proteins to specific DNA sequences has been utilized in the creation of a number of highly successful tools based on proteins fused to a Cas effector protein and allows for localized activity in the region surrounding the target. The most successful of these derivatives are the base editors developed by the Liu lab. Here adenine or cytidine deaminases are fused to the N or C terminus of Cas9 and are used to convert bases around the target site. These deaminases effectively swap the base pairing with the effected base, causing A to convert to G on the case of adenine base editors[111] and C to T in the case of cytidine.[112] In the later case it is necessary to fuse a third protein (uracil glycosylase inhibitor or UGI) to the opposite side of Cas9 to inhibit reversion of the edit by uracil glycosylase. Here, the UGI fused to Cas9 remains at the editing site until the next round of DNA replication, at which point the base change is rendered permanent. By varying the linker length the Liu lab has successfully created editors that effect only a narrow window surrounding the target to less than 5 bases, enabling selective modification through single base shifts without the need for DNA cleavage.[111,113] The Liu lab has successfully generated variations on these fused proteins by evolving the deaminase to be more efficient at base conversion. The downside of these efforts has come from reports that these constructs cause significant off-target editing due to transient activity outside of binding sites.[114] Since the base editors are not linked to the conformational change that occurs upon target binding they are able to act indiscriminately. A new generation of these tools that incorporates these base editors in place of the HNH domain in Cas9 is proving an exciting resolution to this problem.

Another exciting technology from the Liu lab is the PRIME editor, a Cas9 derivative with an extended guide RNA and a fused DNA polymerase that allows for much more extensive gene insertions than were previously possible.[38] However, this technology is still quite new and will require further optimization before it is a viable tool.

The final derivative of these Cas9 fusions to note are the as yet unpublished histone modifiers fused to Cas9 by Enrique Lin Shiao in the Doudna lab. These allow for targeted modification of chromatin and promise exciting discoveries in the field of epigenetics.

All of the abovementioned Cas9 fusions have been shown to be highly effective in cells when transformed via plasmid or viral delivery, but few have been successfully expressed and purified due to their large size and complex structure. As the complexity of these CRISPR technologies increases it has become increasingly important to develop methods of creating CRISPR fusions ex vivo for biochemical study. The importance of this type of work has been made abundantly clear by work from Lapinaite et al who were able to show that modifications to TadA in the base editor ABEmax V8 were due to over-activation of the TadA domain rather than optimization in linker length or other modifications.[114] These insights are nearly impossible to uncover without the ability to isolate the protein fusion.

1.6 Methods to create protein fusions

Given the utility of CRISPR conjugates, there is significant interest in developing facile means of protein conjugation. Protein modification is a well-studied field with many alternatives available, but there are few techniques capable of covalently coupling whole proteins together. Despite this fact, fusion proteins have taken on a role of increased significance in modern medicine, showing the importance to further develop tools in this field.[115,116]

1.6.1 Genetic fusions

The most common approach to creating protein fusions, coupling proteins on the genetic level is relatively simple using modern cloning techniques. Once transformed into a cell for production these fused genes are expressed as a single protein product and can be purified accordingly.[2] This approach is readily scalable, produces uniform product, and can be modified rapidly to optimize aspects of the fused product. The primary drawback of fused protein expression is well summarized in the challenges faced in isolating Cas9 base editors and other CRISPR fusions: the larger and more complex the construct, the lower the chance of successful expression.[116] This is often caused by miss-folding in one of the subunits but can also be due to a lack of chaperone proteins or other challenges. Optimizing the spacer sequence between the fused proteins can help with miss-folding in some cases but this can take several rounds of trial and error.

Fused protein expression only worsens when considering multi-subunit proteins or differing glycosylation requirements as in antibody fusions and other similar constructs.[117] Genetic fusions are also limited in their orientation, as the two proteins must be attached via the N and C terminus. This prevents attachment at other points without significant engineering and structural re-organization.

1.6.2 Unnatural amino acid incorporation

One of the primary methods of joining proteins together is through the incorporation of unnatural amino acids (UAAs). The Francis group has extensively used the artificial amino acid *p*aF which acts as an aniline reactive group as a basis for the oxidative coupling reactions.[99,118] Other unnatural amino acids include azido groups for Cu(I) catalyzed click reaction with alkynes or through copper-free click reactions to strained cyclooctynes as pioneered by the Bertozzi lab.[119,120]

Unnatural amino acids are typically introduced using the Schultz method of amber codon suppression. Here, tRNA synthases have been evolved to produce tRNA targeted at the UAG stop codon and loaded with a variety of unnatural amino acids.[118] The UAG stop codon can be substituted at any position in the gene of interest, resulting in a protein that either terminates at the selected site or contains the artificial amino acid. The benefit of this approach is that proteins containing a poly histidine tag at the C-terminus will all have successful incorporation of the artificial amino acid, as unsuccessful translations will result in shortened proteins without the histidine tag.[118]

The drawback to this method is the significant reduction in complete protein transcripts. On average proteins with artificial amino acids are expressed at 10% the rate of their wild type counterparts. Due to the reliance on multiple plasmids this approach is only really feasible on bacterial expression systems and is quite challenging to implement in mammalian expression systems for antibodies or other therapeutics.[118] The final downside is that most of these reactions require a complementary reactive group on the coupling partner. For couplings of two proteins, both must be expressed with the requisite unnatural amino acid, greatly decreasing the ease of production.

1.6.3 Chemical modification of native amino acids

Several native functional groups are regularly targeted for modification. The most are lysine residues, where the primary amine readily couples with NHS esters and other reactive groups. The drawback of lysine modification is that it is a ubiquitous amino acid, comprising over 5% of most protein surfaces.[121] Due to its ubiquity, lysine modification typically results in a mixture of products containing variable numbers of modifications.[122]

Cysteine residues are also frequently targeted due to their selectivity and relative rarity. Cysteines comprise only 2% of surface amino acids on average and in most cases are involved in disulfide bonding.[121] While involved in the disulfide bond cysteines are non-reactive to most common coupling chemistry. Maleimides are the most common choice for cysteine coupling, as they are selective to thiols and are relatively efficient.[123] Recently, maleimide couplings have fallen out of favor in therapeutic development due to their ability to hydrolyze, potentially releasing the bound cargo.[124,125]

Cysteines are also employed in native chemical ligation (NCL), where an N-terminal cysteine allows coupling to peptides with a C-terminal thioether and results in the formation of a stable peptide bond.[126] NCL has been used to synthesize whole proteins from individual

peptides and allow for incorporation of unnatural reactive groups which would otherwise be impossible to introduce. Unfortunately, this approach is limited to the synthesis of smaller proteins and is not feasible for the production of larger protein-protein conjugates. [126]

Glutamic and aspartic acids can also be targeted by certain chemistries but these often require much more damaging conditions and are again prone to over-modification.[127]

The N-terminal amino group poses a unique chemical handle due to the nearby substitution. The Francis lab has utilized this privileged reactivity to great effect to target the N-terminus with a variety of oxidative coupling reactions. The most successful of these is the reaction with 2-pyridine-carboxaldehyde (2PCA) which results in a covalent bond through rearrangement to the cyclized imine.[128,129]

The primary challenge in coupling proteins together using native amino acids is the requirement for a bifunctional linker. In these reactions, it is imperative that the linker attach only once to the protein of interest, and again to the target. This proves exceptionally challenging when targeting lysines, as any lysine modification strategy is capable of targeting both the initial protein and the coupling partner. The most common workaround is the use of heterobifunctional linkers that possess both NHS and maleimide reactive groups on opposing sides.[130,131] In this case, the starting protein cannot contain surface cysteines and is modified by the NHS group at surface lysines, then the small molecule is removed. In a subsequent reaction a cysteine containing coupling partner is introduced, resulting in linkage between the two proteins. Unfortunately these reactions are often inefficient, requiring large excess of the small molecule coupling partner as well as the cysteine containing protein. These linkages are also not directed, in that it is nearly impossible to control which surface lysine is modified which results in a random orientation between the two proteins.[131,132]

1.6.4 Enzymatic protein-protein coupling

An emerging field of work focuses on using enzymes to couple whole proteins together. There are several approaches of note: the use of recognition domains with target peptides, terminal coupling, and native amino acid activation.

Recognition domain technology such as SpyCatcher/SpyTag uses a targeting peptide or molecule that is recognized by the receiving domain and fused via covalent bonding.[133,134] These domain based fusion technologies have been used to produce highly interesting constructs including vaccine candidates and enzyme fusions.[134] The primary benefit of these techniques is that they can be incorporated without the need for artificial amino acids, and in many cases are bio-orthogonal allowing for application in studying various aspects of cell biology.[135–137] This selectivity requires fairly large recognition domains to work, and since coupling domain is 20-30 kDa in size this can equal or exceed the mass of the desired coupling partner.[133] The size of these domains once again limits the orientation of the fused proteins and ensures that there will always be a significant space between the fused partners. A final concern in therapeutic applications is the potential immunological risk of the large domains,

since these are bacterial in origin it is possible that these constructs could provoke strong immune response.[137]

Terminal coupling is primarily accomplished via the sortase system. Here a C-terminal tag of LPXTG is recognized by the sortase enzyme and fused to an N-terminal poly-glycine partner protein.[138] Sortase coupling produces peptide bonds between the fused proteins and leaves behind only the minimal scar of the LPXTG tag between the two.[138] The downside of this approach is that the enzyme operates through an equilibrium mechanism and catalyzes the reaction in both directions.[139] As a result, the reaction tends not to go to completion without significant effort, complicating purification of the fused product and lowering the yield.[140] The other drawback of sortase coupling is that is again only couples proteins at the N and C termini, limiting the potential to explore other configurations.[141]

Native amino acid activation uses enzymes to modify functional groups already present on the surface of a protein to provide unique reactivity. An excellent example of this technology is the SMARTag approach developed by the Bertozzi lab and later Redwood Biosciences.[142] SMARTag uses the formylglycine-generating enzyme (FGE) to recognize the CxPxR sequence on a target protein and converts the cysteine to a formyl-glycine.[143] Further reaction via Pictect-Spengler coupling allows for covalent bond formation between the protein and its target.[142] The downside here is the requirement for a hydrazine group or similar in the coupling partner, requiring chemical modification or artificial amino acid incorporation.[143]

1.7 Conclusions

The discovery of CRISPR and its application towards gene editing has had an earthshaking impact on biology and medicine. In order to develop it into useable treatments and tools we need novel delivery methods to provide entry into cells, and a new toolkit for modifying these enzymes. Due to the potential for immunological rejection, off target editing, and ethical concerns involved with producing Cas9 in cells it would be preferable to develop delivery techniques that mediate delivery of the pre-assembled RNP. Of available options, enclosing the Cas9 RNP in a virus packaging offers an exciting opportunity to avoid immune detection, stave off RNase degradation, and engineer in tissue specificity through the attachment of antibodies or other cell binding moieties.

For modification of Cas proteins, new tools are required that enable protein and peptide coupling at specific positions around the protein surface without the need for artificial amino acids or exotic conditions. An ideal technology would proceed rapidly to completion, require no elevated temperature, and should not require the addition of extraneous domains.

1.8 References

1. Wang, L. *et al.* In Vivo delivery systems for therapeutic genome editing. *Int. J. Mol. Sci.* **17**, 626 (2016).

2. Kim, Y. G. & Chandrasegaran, S. Chimeric restriction endonuclease. *Proc. Natl. Acad. Sci.* **91**, 883–887 (1994).

3. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337**, 816–821 (2012).

4. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci.* **93**, 1156–1160 (1996).

5. Bibikova, M. *et al.* Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Mol. Cell. Biol.* **21**, 289–297 (2001).

6. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakatura, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).

7. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).

8. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).

9. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).

10. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* **1**, 7 (2006).

11. Cameron, P. *et al.* Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nat. Biotechnol.* **37**, 1471–1477 (2019).

12. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science (80).* **320**, 1047–1050 (2008).

13. Tyson, G. W. & Banfield, J. F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10**, 200–207 (2007).

14. Sternberg, S. H. & Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol. Cell* **58**, 568–574 (2015).

15. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (80).* **343**, 1247997 (2014).

16. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).

17. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).

18. Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* **3**, eaao0027 (2017).

19. Huang, T. P. *et al.* Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* **37**, 626–631 (2019).

20. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).

21. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).

22. Stella, S. *et al.* Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* **175**, (2018).

23. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).

24. Swarts, D. C. & Jinek, M. Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *Wiley Interdiscip. Rev. RNA* **9**, 1–19 (2018).

25. Tambe, A., East-Seletsky, A., Knott, G. J., Doudna, J. A. & O'Connell, M. R. RNA Binding and HEPN-Nuclease Activation Are Decoupled in CRISPR-Cas13a. *Cell Rep.* **24**, 1025–1036 (2018).

26. Freije, C. A. *et al.* Programmable Inhibition and Detection of RNA Viruses Using Cas13. *Mol. Cell* **76**, 826–837 (2019).

27. Pausch, P. *et al.* Crispr-casf from huge phages is a hypercompact genome editor. *Science (80-.).* **369**, 333–337 (2020).

28. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2013**, 1–9 (2013).

29. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).

30. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).

31. Watters, K. E., Fellmann, C., Bai, H. B., Ren, S. M. & Doudna, J. A. Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science* **362**, 236–239 (2018).

32. Marino, N. D., Pinilla-Redondo, R., Csörgő, B. & Bondy-Denomy, J. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nat. Methods* **17**, 471–479 (2020).

33. Knott, G. J. *et al.* Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nat. Struct. Mol. Biol.* **26**, 315–321 (2019).

34. Knott, G. J. *et al.* Structural basis for AcrVA4 inhibition of specific CRISPR-Cas12a. *Elife* **8**, (2019).

35. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).

36. Li, P. *et al.* Cas12a mediates efficient and precise endogenous gene tagging via MITI: microhomology-dependent targeted integrations. *Cell. Mol. Life Sci.* **77**, 3875–3884 (2020).

37. Swarts, D. C., van der Oost, J. & Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell* **66**, 221-233.e4 (2017).

38. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

39. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).

40. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).

41. Cameron, P. *et al.* Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods* **14**, 600–606 (2017).

42. Boyle, E. A. *et al.* High-throughput biochemical profiling reveals Cas9 off- target binding and unbinding heterogeneity. *Bioarxiv* **114**, 5461–5466 (2016).

43. Pisal, D. S., Kosloski, M. P. & Balu-Iyer, S. V. Delivery of therapeutic proteins. *J. Pharm. Sci.* **99**, 2557–2575 (2010).

44. Kane, C., O'Neil, K., Conk, M. & Picha, K. Inhalation Delivery of Protein Therapeutics. *Inflamm. Allergy-Drug Targets* **12**, 81–87 (2013).

45. Pei, D. *et al.* Anti-Tumor Activity and Pharmacokinetics of AP25-Fc Fusion Protein. *Int. J. Med. Sci.* **16**, 1032–1041 (2019).

46. Miller, J. B. *et al.* Non-Viral CRISPR/Cas Gene Editing In Vitro and In Vivo Enabled by Synthetic Nanoparticle Co-Delivery of Cas9 mRNA and sgRNA. *Angew. Chemie - Int. Ed.* **56**, 1059–1063 (2017).

47. Zhang, L. *et al.* Prolonging the plasma circulation of proteins by nano-encapsulation with phosphorylcholine-based polymer. *Nano Res.* **9**, 2424–2432 (2016).

48. Cheng, G. *et al.* Self-Assembly of Extracellular Vesicle-like Metal–Organic Framework Nanoparticles for Protection and Intracellular Delivery of Biofunctional Proteins. *J. Am. Chem. Soc.* **140**, 7282–7291 (2018).

49. Longmuir, K. J., Robertson, R. T., Haynes, S. M., Baratta, J. L. & Waring, A. J. Effective targeting of liposomes to liver and hepatocytes in vivo by incorporation of a Plasmodium amino acid sequence. *Pharm. Res.* **23**, 759–769 (2006).

50. Dehaini, D., Fang, R. H. & Zhang, L. Biomimetic strategies for targeted nanoparticle delivery. *Bioeng. Transl. Med.* **1**, 30–46 (2016).

51. Aanei, I. L. *et al.* Biodistribution of Antibody-MS2 Viral Capsid Conjugates in Breast Cancer Models. *Mol. Pharm.* **13**, 3764–3772 (2016).

52. Staahl, B. T. *et al.* Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.* **35**, 431–434 (2017).

53. Mehta, A. & Merkel, O. M. Immunogenicity of Cas9 Protein. *J. Pharm. Sci.* **109**, 62–67 (2020).

54. Wu, Y. *et al.* Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nature Medicine* **25**, 776–783 (2019).

55. Cardarelli, F. *et al.* The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery. *Sci. Rep.* **6**, 25879 (2016).

56. Sánchez-Sánchez, L. *et al.* Design of a VLP-nanovehicle for CYP450 enzymatic activity delivery. *J. Nanobiotechnology* **13**, 66 (2015).

57. Ramakrishna, S. *et al.* Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res.* **24**, 1020–1027 (2014).

58. Liu, B.-Y. *et al.* Peptide and Aptamer Decorated Delivery System for Targeting Delivery of Cas9/sgRNA Plasmid To Mediate Antitumor Genome Editing. *ACS Appl. Mater. Interfaces* **11**, 23870–23879 (2019).

59. Wu, Y. *et al.* Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* **13**, 659–662 (2013).

60. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).

61. Wang, D. *et al.* Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. doi:10.1089/hum.2015.087

62. Jo, Y. Il, Suresh, B., Kim, H. & Ramakrishna, S. CRISPR/Cas9 system as an innovative genetic engineering tool: Enhancements in sequence specificity and delivery methods. *Biochim. Biophys. Acta - Rev. Cancer* **1856**, 234–243 (2015).

63. van Haasteren, J., Li, J., Scheideler, O. J., Murthy, N. & Schaffer, D. V. The delivery challenge: fulfilling the promise of therapeutic genome editing. *Nat. Biotechnol.* **38**, 845–855 (2020).

64. Sun, W. *et al.* Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing. *Angew. Chemie - Int. Ed.* **54**, 12029–12033 (2015).

65. Ledford, H. *The human retina. A CRISPR therapy has been inserted directly into a person's eye*.

66. Chrenek, M. A., Nickerson, J. M. & Boatright, J. H. Clustered Regularly Interspaced Short Palindromic Repeats: Challenges in Treating Retinal Disease. *Asia-Pacific Journal of Ophthalmology* **5**, 304–308 (2016).

67. Cunningham, F. J., Goh, N. S., Demirer, G. S., Matos, J. L. & Landry, M. P. Nanoparticle-Mediated Delivery towards Advancing Plant Genetic Engineering. *Trends Biotechnol.* **36**, 882–897 (2018).

68. Lee, K. *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* **1**, 889–901 (2017).

69. Lee, B. *et al.* Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X syndrome from exaggerated repetitive behaviours. *Nat. Biomed. Eng.* **2**, 497–507 (2018).

70. Yin, H. *et al.* Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* **15**, 541–555 (2014).

71. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73–80 (2015).

72. Qazi, S. *et al.* Programmed Self-Assembly of an Active P22-Cas9 Nanocarrier System. *Mol. Pharm.* **13**, 1191–1196 (2016).

73. Harrington, L. B. *et al.* A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1–7 (2017).

74. Rouet, R. *et al.* Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing. *J. Am. Chem. Soc.* **140**, 6596–6603 (2018).

75. He, M. *et al.* A traceless linker for aliphatic amines that rapidly and quantitatively fragments after reduction. *Chem. Sci.* **11**, 8973–8980 (2020).

76. Rouet, R., de Oñate, L., Li, J., Murthy, N. & Wilson, R. C. Engineering CRISPR-Cas9 RNA–Protein Complexes for Improved Function and Delivery. *Cris. J.* **1**, 367–378 (2018).

77. Tan, M. L., Choong, P. F. M. & Dass, C. R. Recent developments in liposomes, microparticles and nanoparticles for protein and peptide drug delivery. *Peptides* **31**, 184–193 (2010).

78. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat. Biotechnol.* **34**, 328–333 (2016).

79. Gao, X. *et al.* Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217–221 (2018).

80. O'Riordan, C. R. *et al.* PEGylation of adenovirus with retention of infectivity and protection from neutralizing antibody in vitro and in vivo. *Hum. Gene Ther.* **10**, 1349–1358 (1999).

81. Pan, Y. *et al.* Development of a microRNA delivery system based on bacteriophage MS2 virus-like particles. *FEBS J.* **279**, 1198–1208 (2012).

82. Farkas, M. E. *et al.* PET Imaging and Biodistribution of Chemically Modified Bacteriophage MS2. *Mol. Pharm.* **10**, 69–76 (2013).

83. Yildiz, I., Shukla, S. & Steinmetz, N. F. Applications of viral nanoparticles in medicine. *Curr. Opin. Biotechnol.* **22**, 901–908 (2011).

84. Elsohly, A. M. *et al.* Synthetically Modified Viral Capsids as Versatile Carriers for Use in Antibody-Based Cell Targeting. *Bioconjug. Chem.* **26**, 1590–1596 (2015).

85. Ashley, C. E. *et al.* Cell-specific delivery of diverse cargos by bacteriophage MS2 virus-like particles. *ACS Nano* **5**, 5729–5745 (2011).

86. Hooker, J. M., Kovacs, E. W. & Francis, M. B. Interior Surface Modification of Bacteriophage MS2. *J. Am. Chem. Soc.* **126**, 3718–3719 (2004).

87. Lyu, P., Javidi-Parsijani, P., Atala, A. & Lu, B. Delivering Cas9/sgRNA ribonucleoprotein (RNP) by lentiviral capsid-based bionanoparticles for efficient 'hit-and-run' genome editing. *Nucleic Acids*

Res. **47**, e99 (2019).

88. Golmohammadi, R., Valegård, K., Fridborg, K. & Liljas, L. The Refined Structure of Bacteriophage MS2 at 2·8 Å Resolution. *J. Mol. Biol.* **234**, 620–639 (1993).

89. Johansson, H. E., Liljas, L. & Uhlenbeck, O. C. RNA recognition by the MS2 phage coat protein. *Semin. Virol.* **8**, 176–185 (1997).

90. Parrott, A. M. RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour. *Nucleic Acids Res.* **28**, 489–497 (2000).

91. Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* **12**, 595–600 (1993).

92. Konermann, S. *et al.* CRISPR-Cas9 complex. *Nature* **517**, 61422–61427 (2015).

93. George, L., Indig, F. E., Abdelmohsen, K. & Gorospe, M. Intracellular RNA-tracking methods. *Open Biology* **8**, 180104 (2018).

94. Liao, H. K. *et al.* In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell* **171**, 1495-1507.e15 (2017).

95. Koning, R. I. *et al.* Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat. Commun.* **7**, 12524 (2016).

96. Hartman, E. C. *et al.* Quantitative characterization of all single amino acid variants of a viral capsid-based drug delivery vehicle. *Nat. Commun.* **9**, 1–11 (2018).

97. Hartman, E. C. *et al.* Experimental Evaluation of Coevolution in a Self-Assembling Particle. *Biochemistry* **58**, 1527–1538 (2019).

98. Anderson, E. A. *et al.* Viral nanoparticles donning a paramagnetic coat: Conjugation of MRI contrast agents to the MS2 capsid. *Nano Lett.* **6**, 1160–1164 (2006).

99. Behrens, C. R. *et al.* Rapid chemoselective bioconjugation through oxidative coupling of anilines and aminophenols. *J. Am. Chem. Soc.* **133**, 16398–16401 (2011).

100. Carrico, Z. M., Romanini, D. W., Mehl, R. A. & Francis, M. B. Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* **0**, 1205 (2008).

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

102. Peabody, D. S. Subunit fusion confers tolerance to peptide insertions in a virus coat protein. *Arch. Biochem. Biophys.* **347**, 85–92 (1997).

103. Plevka, P., Tars, K. & Liljas, L. Structure and stability of icosahedral particles of a covalent coat protein dimer of bacteriophage MS2. *Protein Sci.* **18**, 1653–1661 (2009).

104. Peabody, D. S. *et al.* Immunogenic Display of Diverse Peptides on Virus-like Particles of RNA Phage MS2. *J. Mol. Biol.* **380**, 252–263 (2008).

105. Robinson, S. A., Hartman, E. C., Ikwuagwu, B. C., Francis, M. B. & Tullman-Ercek, D. Engineering a Virus-like Particle to Display Peptide Insertions Using an Apparent Fitness Landscape.

Biomacromolecules **21**, 4194–4204 (2020).

106. Brown, W. L. *et al.* RNA bacteriophage capsid-mediated drug delivery and epitope presentation. *Intervirology* **45**, 371–380 (2002).

107. Glasgow, J. E., Capehart, S. L., Francis, M. B. & Tullman-Ercek, D. Osmolyte-mediated encapsulation of proteins inside MS2 viral capsids. *ACS Nano* **6**, 8658–8664 (2012).

108. Glasgow, J. E., Asensio, M. A., Jakobson, C. M., Francis, M. B. & Tullman-Ercek, D. Influence of Electrostatics on Small Molecule Flux through a Protein Nanoreactor. *ACS Synth. Biol.* **4**, 1011–1019 (2015).

109. Glasgow, J. & Tullman-Ercek, D. Production and applications of engineered viral capsids. *Appl. Microbiol. Biotechnol.* **98**, 5847–5858 (2014).

110. Tong, G. J., Hsiao, S. C., Carrico, Z. M. & Francis, M. B. Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* **131**, 11174–11178 (2009).

111. Gaudelli, N. M. *et al.* Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

112. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

113. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–848 (2018).

114. Lapinaite, A. *et al.* DNA capture by a CRISPR-Cas9-guided adenine base editor. *Science (80-.).* **369**, 566–571 (2020).

115. Neri, D. Antibody–cytokine fusions: Versatile products for the modulation of anticancer immunity. *Cancer Immunol. Res.* **7**, 348–354 (2019).

116. Schmidt, S. R. *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges*. *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges* (John Wiley & Sons, Inc., 2013). doi:10.1002/9781118354599

117. Krishnamoorthy, S. *et al.* Recombinant factor VIII Fc (rFVIIIFc) fusion protein reduces immunogenicity and induces tolerance in hemophilia A mice. *Cell. Immunol.* **301**, 30–39 (2016).

118. Mehl, R. A. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).

119. Baskin, J. M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci.* **104**, 16793–16797 (2007).

120. Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci.* **99**, 19–24 (2002).

121. Gaur, R. K. Amino acid frequency distribution among eukaryotic proteins. *IIOAB J.* **5**, 6–11 (2014).

122. Rosen, C. B. & Francis, M. B. targeting the n terminus for site-selective protein modification. *Nat. Chem. Biol.* **13**, (2017).

123. Boutureira, O. *et al.* Advances in Chemical Protein Modification. *Chem. Rev.* **115**, 2174–2195 (2015).

124. Fontaine, S. D., Reid, R., Robinson, L., Ashley, G. W. & Santi, D. V. Long-Term Stabilization of Maleimide–Thiol Conjugates. *Bioconjug. Chem.* **26**, 145–152 (2015).

125. Ravasco, J. M. J. M. J. M., Faustino, H., Trindade, A. & Gois, P. M. P. P. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. - A Eur. J.* **25**, 43–59 (2019).

126. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science (80-.).* **266**, 776–779 (1994).

127. Schlick, T. L., Ding, Z., Kovacs, E. W. & Francis, M. B. Dual-surface modification of the tobacco mosaic virus. *J. Am. Chem. Soc.* **127**, 3718–3723 (2005).

128. Macdonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **11**, 326–331 (2015).

129. ElSohly, A. M. *et al.* ortho-Methoxyphenols as Convenient Oxidative Bioconjugation Reagents with Application to Site-Selective Heterobifunctional Cross-Linkers. *J. Am. Chem. Soc.* **139**, 3767–3773 (2017).

130. Hermanson, G. T. *Bioconjugate Techniques*. *Bioconjugate Techniques* (Elsevier, 2008). doi:10.1016/B978-0-12-370501-3.X0001-X

131. Wei, B. *et al.* Development of an antisense RNA delivery system using conjugates of the MS2 bacteriophage capsids and HIV-1 TAT cell penetrating peptide. *Biomed. Pharmacother.* **63**, 313–318 (2009).

132. Gerrard, J. A. Protein–protein crosslinking in food: methods, consequences, applications. *Trends Food Sci. Technol.* **13**, 391–399 (2002).

133. Brune, K. D. *et al.* Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Sci. Rep.* **6**, 19234 (2016).

134. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci.* **109**, E690–E697 (2012).

135. Zeng, Y. S., Gao, R. C., Wu, T. W., Cho, C. & Tan, K. T. Fluorescent Probe Encapsulated in SNAP-Tag Protein Cavity to Eliminate Nonspecific Fluorescence and Increase Detection Sensitivity. *Bioconjug. Chem.* **27**, 1872–1879 (2016).

136. Reddington, S. C. & Howarth, M. Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher. *Curr. Opin. Chem. Biol.* **29**, 94–99 (2015).

137. Brune, K. D. & Howarth, M. New routes and opportunities for modular construction of particulate vaccines: Stick, click, and glue. *Front. Immunol.* **9**, (2018).

138. Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).

139. Popp, M. W.-L., Antos, J. M. & Ploegh, H. L. Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. in *Current Protocols in Protein Science* **Chapter 15**, 15.3.1-15.3.9 (John Wiley & Sons, Inc., 2009).

140. Nguyen, G. K. T. *et al.* Butelase-mediated cyclization and ligation of peptides and proteins. *Nat. Protoc.* **11**, 1977–1988 (2016).

141. Theile, C. S. *et al.* Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1800–1807 (2013).

142. Liu, J., Barfield, R. M. & Rabuka, D. Site-Specific Bioconjugation Using SMARTag® Technology: A Practical and Effective Chemoenzymatic Approach to Generate Antibody–Drug Conjugates. in *Methods in Molecular Biology* **2033**, 131–147 (Humana Press Inc., 2019).

143. Wu, P. *et al.* Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3000–3005 (2009).

Chapter 2. Encapsulating CRISPR-Cas9 inside MS2 virus like particles

Abstract

Delivery of preassembled Cas9 RNPs is a major challenge to their use in gene therapies. Due to their sensitivity to RNase and rapid removal from the blood it is ideal to encapsulate Cas9 within a carrier system which can protect against immune clearance and RNase degradation. We show that Cas9 is successfully encapsulated within the MS2 VLP following integration of the MS2 hairpin into the loops of the Cas9 sgRNA, and demonstrate the successful editing in cells of MS2-encapsulated Cas9. We then discovered that the MS2 VLP is not protective against RNase degradation of the Cas9 sgRNA. We validate that RNase is capable of entering the MS2 VLP through the pores in the coat protein structure, and show that degradation does not occur in the presence of larger nucleases or in the case of the wild type MS2 capsid. Finally, we examine a library of MS2 mutants in the loop region surrounding the pore and find that no candidates confer the requisite protection against RNase activity.

Cell work was conducted with the help of Christof Fellmann, and MS2 mutant libraries and library analysis were completed with the help of Emily Hartman and Stephanie Robinson.

2.1 MS2 for Cas9 delivery

The targeted DNA cleavage of CRISPR-Cas9 allows for ready modification of genes within target cells, eliminating problematic proteins or correcting genetic defects.[1] One of the primary challenges in the development of CRISPR therapeutics is delivering the pre-assembled Cas9 Ribonuclear Protein (RNP) through the blood to target cells.[2,3] RNase is the greatest threat to successful delivery as it rapidly degrades the necessary guide RNA, effectively deactivating the complex. Next, there is the fact that recent studies have shown high native immune response to Cas9 and other Cas proteins in patients due to their presence in pathogenic bacteria.[4,5] In some patients, this preexisting immune sensitivity to Cas9 can lead to rejection or inflammation caused by treatment with Cas proteins.[4,5] Finally, as with all protein therapeutics, Cas9 is cleared from the blood rapidly, minimizing the window of effect and decreasing the rate of delivery.[6]

The MS2 Virus Like Particle (VLP) has been used previously to encapsulate a variety of anionic cargo including DNA, RNA, and anionic proteins.[7–11] The ready modification via introduced artificial amino acids such as *p*-aminophenylalanine (*p*aF) allows for facile modification of the capsid exterior with targeting groups or other peptides to facilitate cellular entry.[12–14] The fact that the MS2 VLP naturally forms around RNA containing the MS2 hairpin loop makes it an attractive candidate for its potential to encapsulate the Cas9 RNP.[15–17] This idea is further supported by several reports that utilize MS2 hairpins inserted into the loops of the dCas9 sgRNA to recruit MS2 coat proteins fused to transcription inhibitors, repressing the targeted genes.[18] Since MS2 is an RNA bacteriophage found within humans and in the environment at large one assumes that the MS2 capsid is impervious to RNase activity and should confer this protection to Cas9 RNPs encapsulated within the VLP.[11,19,20] Such a system would provide protection for the cargo, decrease immune surveillance, and improve circulation times due to the excellent pharmacokinetics of the MS2 VLP.

2.2 Encapsulating Cas9 in MS2

2.2.1 Preparation of fluorophore labeled Cas9

As shown in Figure 1, *N*-hydroxysuccinimide (NHS) linked AlexaFluor568 (AF568) was incubated with Cas9-mCherry at a 100:1 molar ratio at room temperature for 1 h in 50 mM phosphate buffer pH 7.5. A spin concentrator was then used to perform a buffer exchange to remove excess dye. Although NHS reactions are typically carried out at pH 8 and above to allow for efficient coupling, initial reactions were carried out at pH 7.5 to prevent excessive modification and potential protein denaturation.

The modified proteins were characterized by ESI-TOF (Figure 1B) and were found to have a maximum of five modifications per protein, producing a highly photoactive Cas9 (Figure 1C).

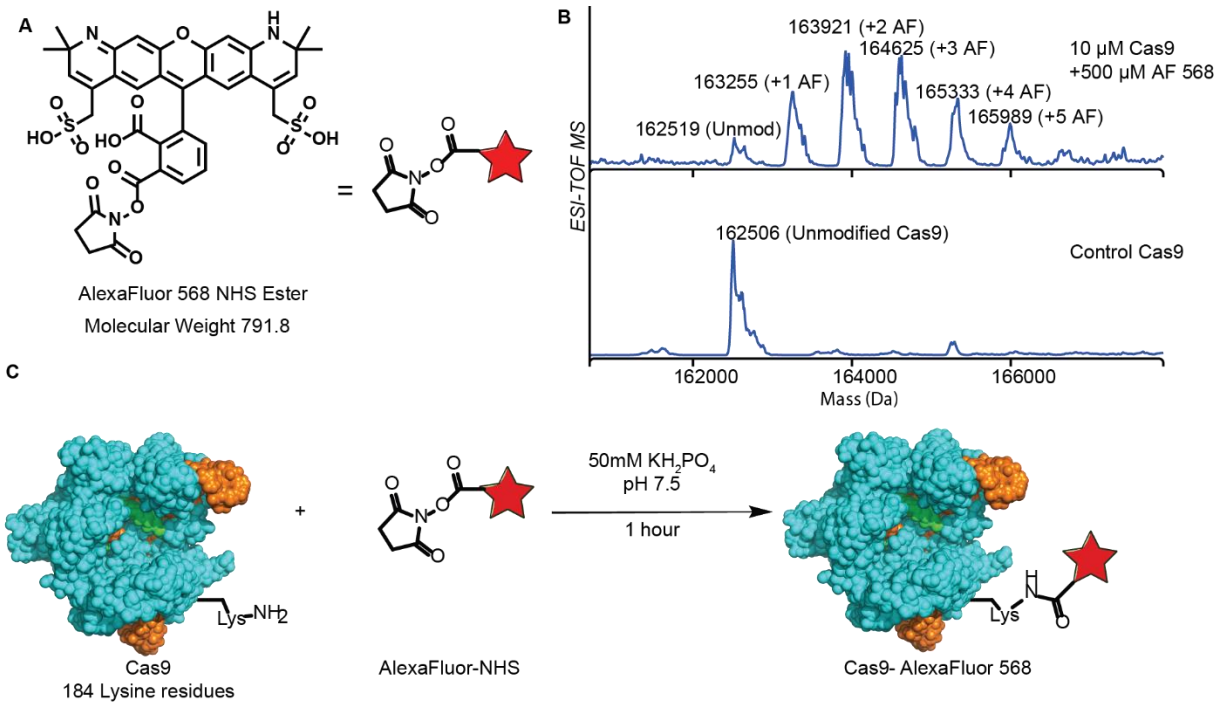


Figure 1: Fluorescent labeling of Cas9**. A** Strcture of AlexaFluor 568 (Excitation: 576 nm Emission: 602 nm) with attached NHS ester. **B**. ESI-TOF spectra for fluorescently labeled Cas9 showing Cas9 with up to 5 modifications per protein. **C**. Scheme illustrating fluorescent labeling conditions for Cas9 (PDBID: 5F9R), which contains 184 lysine residues

2.2.2 Production of modified sgRNA

With the ability to monitor the presence of Cas9 within MS2 capsids, GFP-targeting sgRNA guides were designed with MS2 binding sequences inserted into either loop2 or the tetraloop to enhance MS2 capsid formation (Figure 2A). Successful RNA production yielded up to 521 µg of sgRNA for each of the guides (Supplemental Figure 1). Guides were prepared for 3 potential locations along the GFP gene, designated as 1, 8, and 9, in order to identify the most readily cleaved location. Position 1 is located near the middle of the gene, while 8 and 9 are positioned near the 3' region.

2.2.3 Complexing Cas9 with sgRNA and activity validation

Cas9 was combined with the sgRNA at a 1:1.4 molar ratio according to published protocols.[21] After complexation, the Cas9 with modified guides were electroporated into HEK RT1 or RT6 cell cultures and left for 24 h, then induced with 1 µg/mL of doxycycline for 48 h before analysis via flow cytometry. When gated for GFP fluorescence we observed significant decreases for all editing conditions. There was a noticeable decrease in cleavage efficiency for sgRNA modified in the tetraloop region, which allowed us to move forward using only the sgRNA with the MS2 hairpin in loop 2 targeting position 8 in the GFP gene (Figure 2B, C).

For *in vitro* studies the Cas9 complexed with sgRNA was added to a solution containing an 800 bp long PCR product that had been previously purified by gel electrophoresis. After incubation, the reaction was quenched using stop solution containing EDTA. The cleaved

products were then run through a 1% agarose DNA gel and stained with SybrSafe before imaging. The percent cleavage was determined via densitometry analysis examining the uncleaved band relative to the two cleavage bands. In comparing the three cleavage locations we found that the modification in the tetraloop region slightly reduced cleavage efficiency, while the modification in loop 2 showed comparable activity to the unmodified sgRNA (Figure 2D).

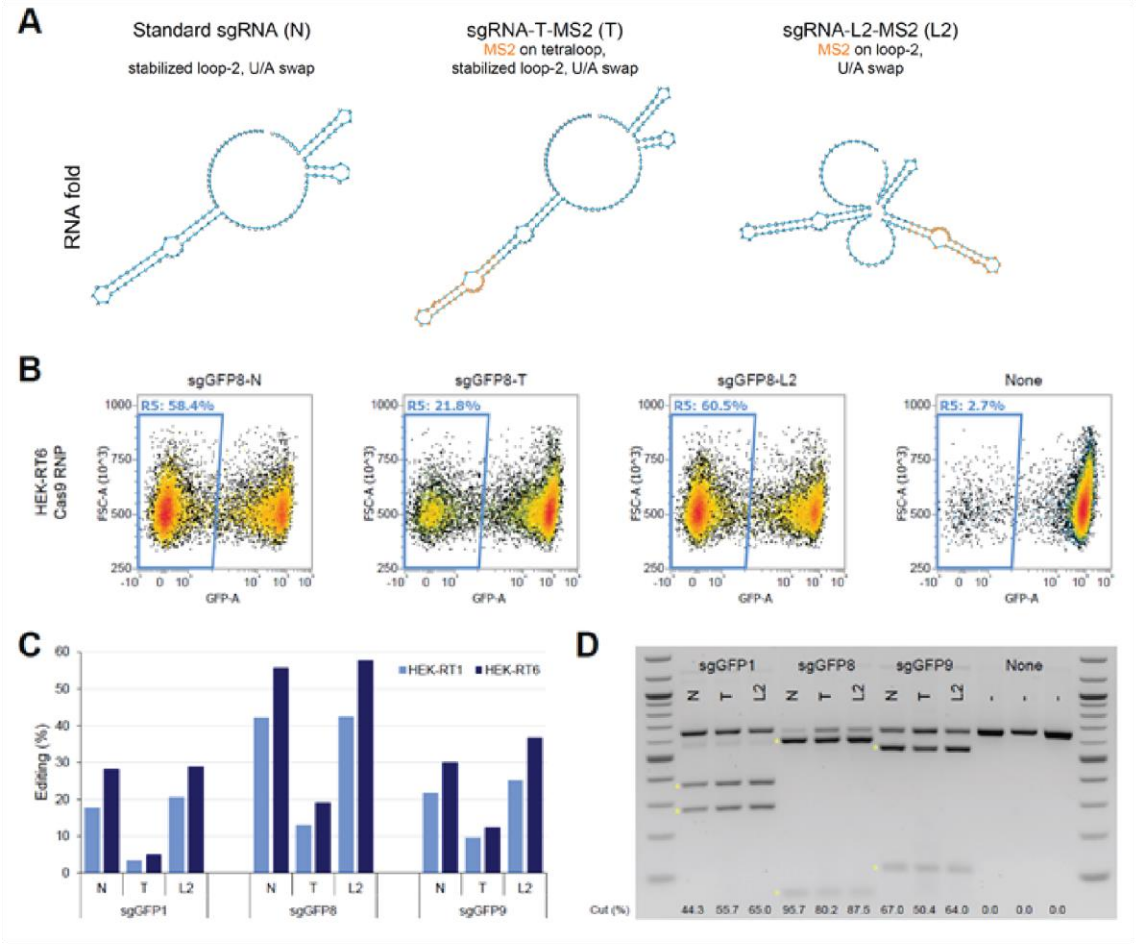


Figure 2: Comparison of Cas9- sgRNA ribonucleoprotein (RNP) complexes seaturing sgRNAs with or without MS2 loops. **A**. Predicted RNA folds of the indicated sgRNA. **B**. Representative flow cytometry of HEK RT6 reporter cells treated with the indicated Cas9-stRNA complexes for 24 h followed by induction of GFP expression with doxycycline (1 µg/mL) for 48 h. **C**. Quantification of flow cytometry data reporting cell culture editing efficiency of the indicated Cas9 RNPs. HEK-RT1 and HEK RT6 reporter cells were treated as in **B**. **D**. In vitro cleavage assay using indicated Cas9-sgRNA complexes. The dsDNA substrate was incubated with 10-fold excess RNP at 37 °C for 30 min. Yellow asterisks (*) indicate expected cleavage products for the sgRNAs.

2.2.4 Encapsulation of Cas9 inside MS2

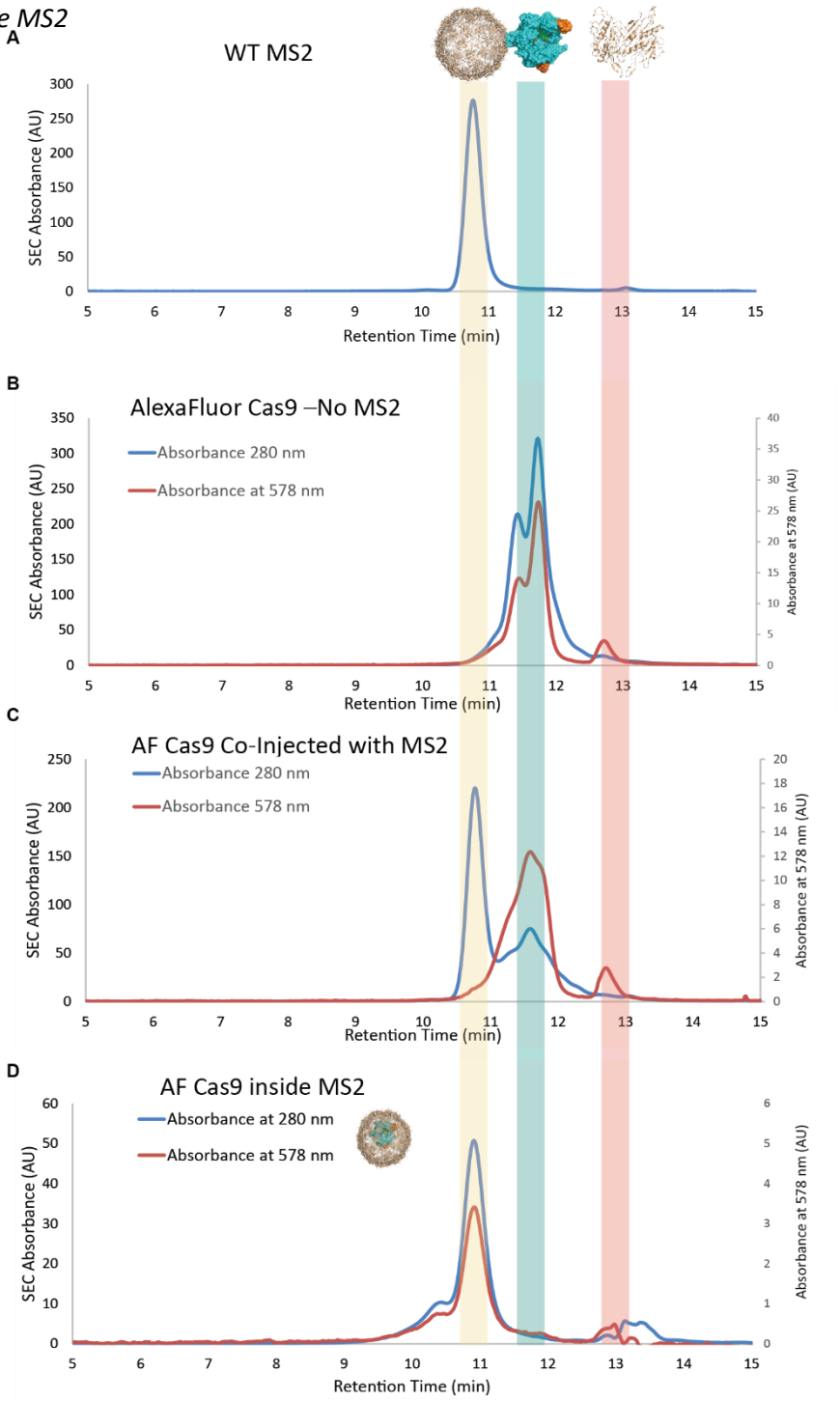


Figure 3: Cas9 encapsulation withing MS2 Capsids via SEC. **A**. SEC trace of wild type MS2 capsids. The beige band highlights capsid peak retention (10.5 min), and the red band highlights monome elution peak (13 min). **B**. SEC trace of Cas9 labeled with AlexaFluor 568 NHS ester. The teal band highlights Cas9 peak retention (11.5 min). **C**. Cas9 and MS2 capsids co-injected without capsid disassembly. The strong peak visible at 10.5 with no absorbance at 578 nm indicates empty capsid with no associated Cas9. **D**. Capsid formed in presence of AF-568 Cas9. The strong absorbance at 578 nm for the capsid peak indicates succesfull RNP encapsulation.

With the MS2 hairpin RNA in hand we next sought to encapsulate Cas9 within an MS2 virus like particle. As previously described by Glasgow et al. MS2 can be disassembled in 66% glacial acetic acid followed by ion removal with a NAP-5 column. Disassembled MS2 and Cas9 were then combined in 50 mM Bis-Tris buffer pH 6.0 with 1.8 M TMAO at 4 °C for 48 h. [9,22,23] Initially capsids were precipitated using saturated ammonium sulfate, but it was found that this procedure damaged the activity of Cas9 and samples were instead capsids were run on a bio-SEC system through size exclusion columns capable of separating the intact capsids from free Cas9. The bio-SEC-5 column was found to be most effective for this process as it results in distinct peaks between free MS2 and Cas9.

To ensure that the re-formed capsids contained Cas9 we took Cas9 modified with AlexaFlour 568 through the encapsulation protocol and monitored both fluorescence and absorbance of the fluorophore during SEC analysis. Controls comprised of MS2 (Figure 3A) and Cas9 alone (Figure 3B), as well as a co-injection of the two combined immediately before injection (Figure 3C). These samples showed clearly that only samples that had undergone the disassembly and reassembly protocol showed fluorescence at the elution time of MS2 (Figure 3D).

Interestingly, all Cas9 sgRNAs were effective at generating MS2 with Cas9 inside, but there was greater consistency in the capsids formed using the Cas9 with sgRNA bearing the MS2 hairpin (Figure 4 A,B). This is likely due to the fact that the Cas9 RNP possesses significant negative charge at the sgRNA loop regions which may be sufficient for capsid assembly.

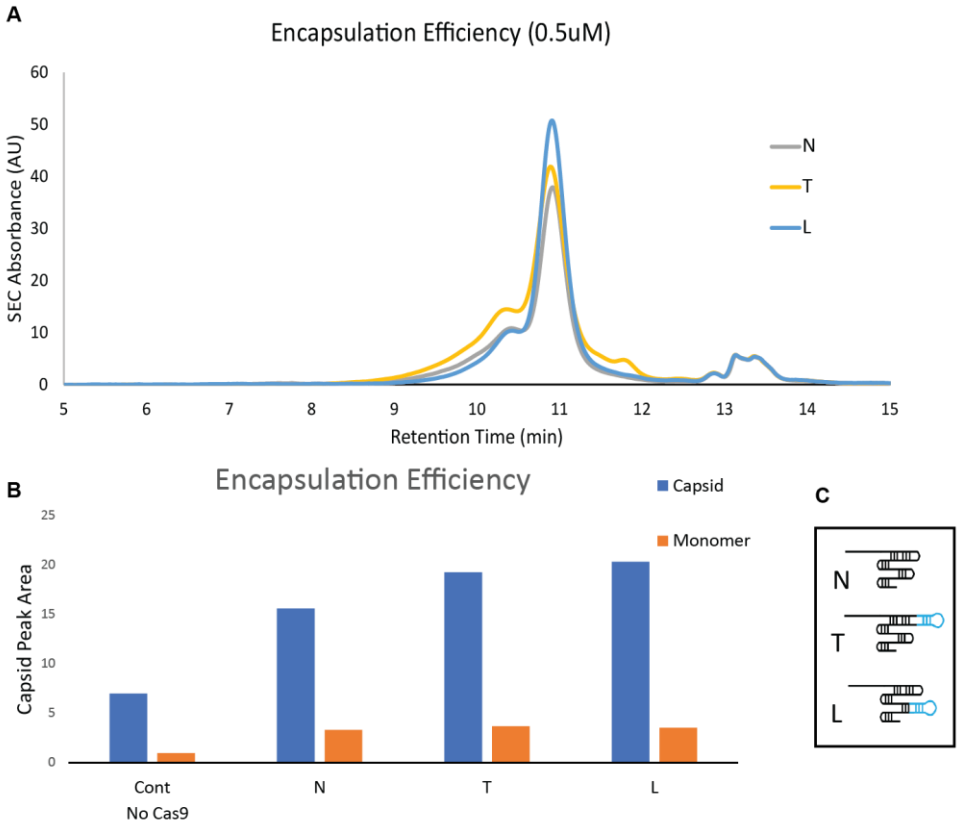


Figure 4: Comparison of sgRNA impact on Cas9 encapsulation withing MS2 capsids via SEC. **A**. SEC trace of MS2 capsids with formed in presence of Cas9 with sgRNA containing the MS2 hairpin loop insertion in positions: None (N), Tetraloop (T) or Loop2 (L). **B**. Integrated area of capsid peaks to assess efficiency of Capsid formation between sgRNAs. **C**. sgRNA structure guide highlighting position of insertion of the MS2 hairpin.

2.2.5 Engineering protease sensitivity into the MS2 VLP

One of the most important design elements for any therapeutic delivery system is release of the cargo from the delivery vector. For small molecule drugs a valine-citroline linker is often used for ready degradation by

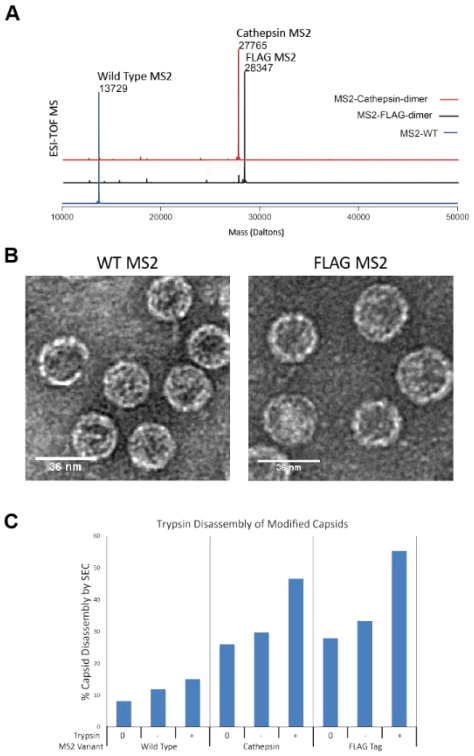


Figure 5: Protease sensitive MS2 dimers. **A**. ESI-TOF MS of MS2 fused dimers with either a cathepsin protease cleavage site or a FLAG tag inserted into the AB loop. **B**. Transmission Electron Microscopy (TEM) images of MS2 and FLAG MS2 capsids. **C**. Quantification of capsid disassembly monitored via SEC UV for wild type MS2 and MS2 dimers with peptide insert.

endosomal proteases.[24] To ensure that encapsulated Cas9 was readily released upon entering the cell it would be beneficial to design capsids which are stable in serum but can be broken down once they reach the cytosol. An efficient approach to force protein degradation is to attach readily degraded sequences into loop regions, promoting the unfolding and degradation of the target.[25,26] However, the MS2 coat protein is very sensitive to peptide insertions, with most attempts resulting in incompetent capsid formation. The Peabody lab has previously shown that adjacent MS2 monomers can be fused, producing an MS2 dimer that shows greater insertion tolerance. Peptides can be inserted into the MS2 AB loop of one of the fused dimers, yielding a capsid with 90 copies of the peptide loop across the surface.[25,27]

Based on previous work the FLAG tag was selected as a control to assess whether competent capsids were formed from the fused MS2 dimer with a short insertion. The most common disease associated protease is cathepsin which is upregulated in cancer cells.[28–30] A cathepsin cleavage sequence was inserted into the AB loop of the fused MS2 dimer and expressed in *E. coli.* Mass spectrum and Transmission Electron Microscopy data showed the purified protein shows competent capsid formation (Figure 5 A,B).

Purified capsids were then exposed to trypsin degradation to assess their sensitivity to generic proteases. Both peptide insertion mutants showed significantly higher degradation as measured via SEC while the wild type capsid showed little evidence of degradation (Figure 5C). These experiments show that peptide insertion MS2 capsids are indeed more readily degraded by proteases than the wild type capsid.

2.3 Cell editing with Encapsulated Cas9

2.3.1 Electroporation into HEK298 cells

With encapsulated samples in hand we next sought to measure the ability of the enshrouded Cas9 to edit mammalian cells. This experiment used the same GFP reporter line as previously described in our sgRNA validation. Here, samples of Cas9 prepared inside MS2, were compared to free Cas9 that had been taken through the encapsulation conditions, as well as Cas9 that had been freshly prepared for this experiment.

For delivery purposes all three Cas9 preparations were co-incubated with one of three cell-entry approaches: the lipid system Lipofectamine-2000 (Lipo-2000), cell penetrating

peptide ppTG21,[31] and cell penetrating peptide LAH4.[32] All three had been shown previously to mediate some form of gene delivery and we hypothesized that the peptide delivery system could be applied to *in vivo* delivery if peptides were encapsulated inside MS2. In contrast, Lipo-2000 is the industry standard for cytosolic delivery and stood for us as a control to examine what would happen if a fully encapsulated Cas9 inside the MS2 VLP was delivered to a cellular cytosol.

As in previous experiments all samples were applied to the cells for 24 h before induction and measured on day 4 of the experiment by flow-cytometry. Figure 6A shows the resulting editing % for each approach. The most striking result from this set of experiments is the fact that the encapsulated Cas9 showed comparable editing to the free Cas9 sample when delivered via Lipofectamine, indicating that even fully encapsulated Cas9 was able to escape the MS2 VLP in the cytosol and mediate editing in the nucleus. The second gratifying result was the confirmation of ppTG21 as a potential delivery peptide for the encapsulated Cas9. Though the delivery efficiency was low, the ppTG21 results indicated that other delivery peptides might be able to mediate efficient cell editing which was an exciting prospect.

2.3.2 MS2 Encapsulated Cas9 and RNase

With the ability to edit cells working, we next examined whether the encapsulation provided material protection against RNase degradation of the sgRNA, a major obstacle to any CRISPR therapeutic. For this round of experiments we took encapsulated Cas9 and compared its editing ability to free Cas9 with the expectation that a shielded Cas9 would be less sensitive to RNase activity. Each sample was exposed to a serial dilution of RNase for 15 min at RT immediately before delivery, then administered to the cells using Lipo-2000 to ensure we were measuring activity rather than delivery. Unfortunately, as evidenced by Figure 6B, not only did the encapsulation not protect Cas9 from RNase breakdown, if anything it appeared to encourage the degradation of Cas9 activity as evidenced by a near complete obliteration of editing at even the most dilute RNase conditions (Figure 6B). This was quite surprising, as the MS2 bacteriophage has an RNA genome that is well protected from RNase and other environmental degradation.

Naturally this surprising result prompted further rounds of experimentation to examine if and why RNase was capable of entering the MS2 VLP, and whether it would be possible to prevent this behavior.

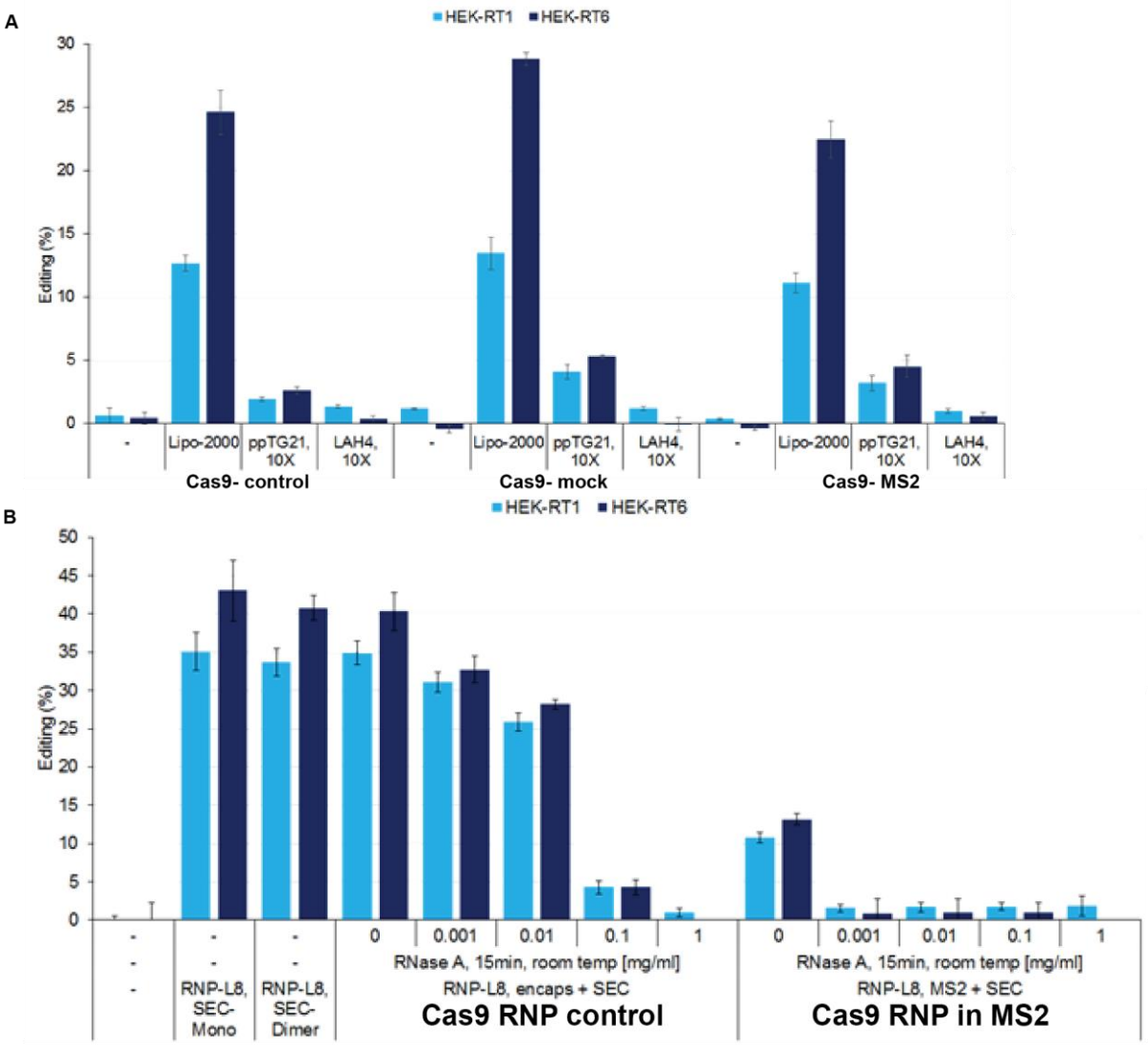


Figure 6: Cell delivery to HEK RT1 and RT6 cells. **A**. Cas9 RNP samples pre-incubated with the indicated delivery vector. Cas9- control samples contained Cas9- sgRNA complexes with no further treatment. Cas9 mock samples comprised Cas9-sgRNA complexes exposed to encapsulation and purification conditions without addition of MS2. Cas9-MS2 samples represent Cas9-MS2 capsids purified via SEC. All samples were allowed to incubate with Lipofectamine or the indicated peptide for 15 min before addition of 50 pmol of Cas9 per well. Samples were incubated 24 h followed by induction of GFP expression with doxycycline (1 µg/mL) for 48 h, then imaging via flow-cytometry. **B**. Editing efficiency of Cas9 RNP after 15 min exposure to RNase. The Cas9 RNP control sample represents the bare RNP, while Cas9 RNP in MS2 is the encapsulated construct. RNase concentrations in mg/mL are listed underneath each sample.

2.4 RNase and the MS2 VLP

We sought to understand the origin of the seemingly contradictory RNase results, where MS2-encapsulated Cas9 showed if anything greater sensitivity to RNase degradation despite the RNA genome being well protected in the wild type capsid. We performed a series of RNA extractions using phenol/chloroform and gel based spin columns to isolate the RNA from the bulk protein in the sample. This process ensures that excess RNase in solution is fully denatured and cannot

contaminate results post extraction. After extraction RNA was loaded into a 10% urea PAGE gel and run for 4 h at 40 W followed by imaging on a BioRad ChemiDoc.

We first examined the impact of RNase on free Cas9 compared to Cas9 in the MS2 capsid. We analyzed both the wild type sgRNA as well as the sgRNA bearing the MS2 stem loop in Loop 2. As seen in Figure 7A we found that in all cases assayed the RNase treatment fully degraded the sgRNA, indicating that this was indeed the cause of our loss of activity in the MS2 encapsulated samples.

Since the RNase entry conflicted with reported protective properties of the MS2 VLP we hypothesized that the re-assembly of the capsid could be at fault, potentially due to structural defects in the re-assembled capsids.[11,19,33–35] As our lab has shown using the SyMAPS system[36], the recombinant MS2 capsids form around endogenous mRNA in the bacteria of origin, encapsulating their own coding mRNA and allowing for downstream readouts of capsid assembly after pH or temperature challenges. To test whether the MS2 VLPs were sensitive to RNase we exposed the purified capsids to a range of RNase concentrations and then ran the samples through the previously described RNA extraction protocol and ran through the urea-PAGE gel. Our results show clear degradation of RNA even at low concentrations of RNase in the MS2 VLP, indicating that even in the case of intact capsids we see RNase degradation (Figure 7A). Capsid stability was confirmed by SEC and TEM which both showed clearly formed capsids, helping to minimize our concerns that we were merely measuring exogenous RNA (Figure 3A, 5B).

To further validate these findings we then repeated the studies with controls containing an RNase inhibitor, as well as samples where RNase was added immediately prior to phenol-chloroform extraction (Figure 7B). These cases showed minimal RNA degradation, confirming that the effect desired was due to RNase accessing RNA within the capsid instead of degradation after the capsid is denatured. The MS2 VLP possesses 32 pores of 2 nm diameter across its surface, which typically do not allow entry of proteins to the interior of the protein shell. However, the Francis lab has shown many times that the capsid is permeable to small molecules, indicating that these pores could allow entry of other hydrophilic molecules.

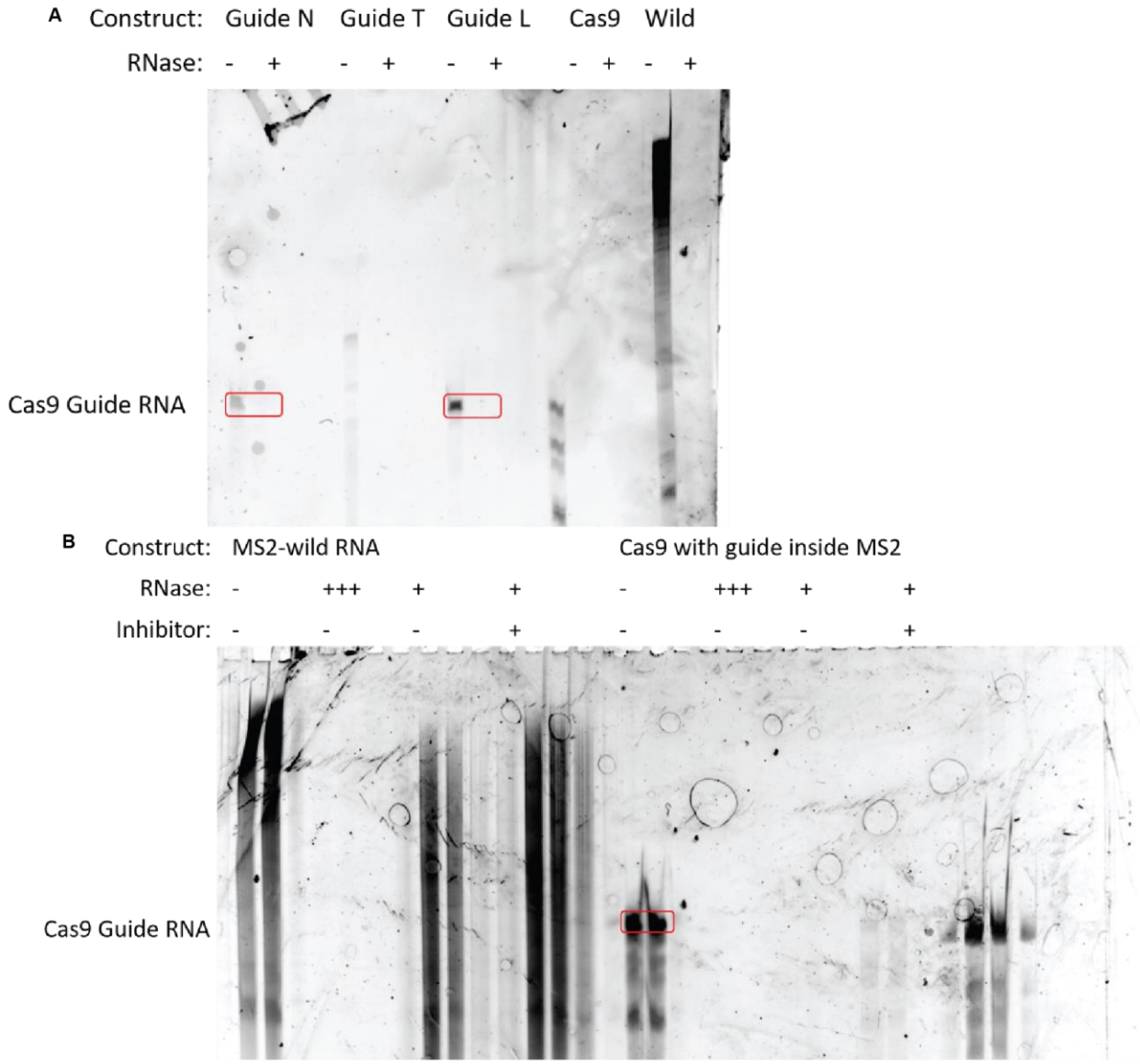


Figure 7: RNA gels stained with SybrGold to visualize RNase degradation of MS2 samples. **A**. Cas9 RNPs and capsids showing RNase degradation of RNA. Samples incubated with (+) or without (-) 1 mg/mL RNase A for 15 min prior to RNA extraction and visualization on a 15% polyacrylamide Urea gel. Samples labeled Guide represent Cas9 with the listed sgRNA, Cas9 represents Cas9 inside MS2, Wild represents MS2 VLPs bearing endogenous RNA. **B**. Analysis of RNase concentration and impact of RNase inhibitor. MS2 capsids with endogenous mRNA or formed around Cas9 were exposed to RNase A at 1 mg/mL (+++) or 0.1 mg/ mL as well as an RNase inhibitor peptide at 1 mg/mL (+) before RNA extraction and visualization.

After validating that intact capsids were RNase sensitive even prior to disassembly, we sought to examine if there was a limit to the size of proteins capable of entering the capsid. This would allow clear determination of whether there were significant flaws in the capsid structure or if the pores are merely flexible enough without the MS2 genome to allow entry of small proteins such as RNase. We found that benzonase from Sigma Aldrich is roughly twice the size of RNase at ~4 nm diameter, according to its crystal structure. This is twice the diameter of the pores in

the MS2 capsid, and should afford a clearer view of whether the capsid is formed defectively or the pores are allowed more flexibility. In the studies comparing Benzonase to RNase we found that benzonase treated MS2 showed no RNA degradation, while un-encapsulated controls were fully degraded. This indicates that Benzonase is capable of degrading RNA samples in our reaction conditions, but not capable of entering the MS2 capsid. (Figure 8D, Supplemental Gel 1, 2, 3).

The surprising result that recombinant capsids were affected by RNase activity led to an interest in whether we could validate our test on the wild type MS2 bacteriophage. Wild type phage were obtained from laboratory stocks and exposed to RNase conditions identical to those for our previous studies then run on a 1% agarose gel. In these samples we observed strong bands corresponding to intact MS2 genomes, supporting the conclusion that recombinant expression yields capsids that are less tightly assembled (Figure 8A).

While surprising, this result is not unprecedented. The MS2 genome has been shown to contact every monomer of the capsid through strong interactions with the cationic residues on the interior.[37–39] It is quite reasonable to assume that the lack of continuous RNA contact might weaken the integrity of the capsid due to charge repulsion or other means. To examine this phenomenon further we performed a series of simulations to monitor the dynamics of the MS2 pores, focusing on the FG loop region which extends into the pores of the capsid. By modelling 10 monomer assemblies for 50 ns at a time we were able to show that there is some flexibility in the FG loop, which showed a maximum separation between opposite loops of 2.3 nm, larger than the minimum diameter of RNase, but still too small for Benzonase entry (Figure 8B).

With the knowledge that pore flexibility and size was the limiting factor affecting RNase sensitivity we decided to employ the SyMAPS approach developed by the Francis lab to evaluate a library of all 2 amino acid mutants into the tip of the FG loop between positions 71 and 76. In conjunction with Emily Hartman and Stephanie Robinson we produced a library of MS2 particles with modifications in the FG loop and exposed them to benzonase or RNase conditions (Supplemental Figure 3, 4). After this, mRNA from the samples was extracted, reverse transcribed, and sequenced. Our goal was to identify capsids that were immune to both RNase and Benzonase conditions in order to find only those competently folded capsids that might allow safe packaging of the mRNA or Cas9 (Supplemental Figure 5). Our assay identified a series of 19 mutants that showed the highest counts in our analysis of RNA sequences present after RNase exposure (Supplemental Figure 5). Here we took prevalence to represent fitness since screens of the library prior to selection (Supplemental Figure 3) did not uncover significant skewing of the library members.

Our 19 candidate mutants were expressed and purified then run through the same RNase assay we used with the native MS2 VLP. Unfortunately, no mutant in isolation was able to prevent degradation by RNase, indicating that our candidates might have some slight protection from RNase activity compared to other members of the library, but insufficient to allow for

protection from a concentrated dose of RNase (Figure 8D, Supplemental Gel 1-3). Given the lack of protection of large mRNA strands we did not pursue further testing of Cas9 encapsulated within these mutants.

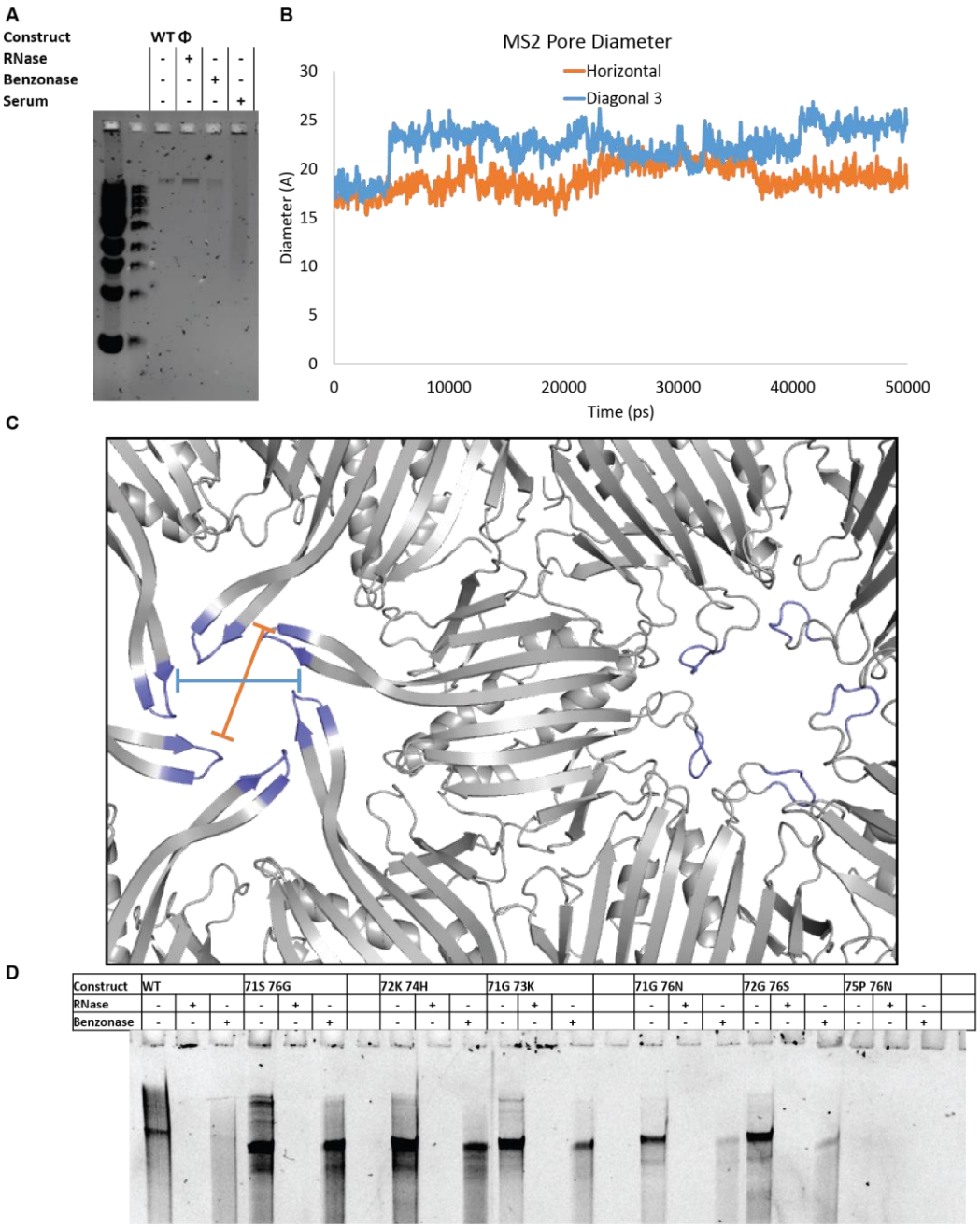


Figure 8: Analysis of the MS2 pore. **A**. Agarose native gel of wild type MS2 bacteriophage exposed to RNase, Benzonase, or control for 15 min prior to visualization and staining with SybrGold. **B**. Protein modeling of the MS2 pore diameter across the horizontal (orange) or diagonal (blue) axis over 50 ns. **C**. Crystal structure of the pores of the MS2 bacteriophage with residues 71-76 highlighted in blue. The pore on left shows the six-fold symmetry arrangement, while the pore on right shows the five-fold symmetry. Horizontal and diagonal axes of the pore marked in blue and gold respectively. **D**. Polyacrylamide urea gel showing MS2 mutants exposed to RNase and Benzonase. Samples were incubated for 30 min at RT prior to extraction and visualization of remaining RNA. Sample names indicate identity of isolated MS2 double mutant.

These disappointing results indicated to us that the MS2 VLP is not an ideal candidate for delivery of RNA containing cargo unless that cargo is a reliable mimic of the MS2 genome in contacting a majority of the capsid interior. Since we were unable to engineer a viable capsid for delivery of the Cas9 RNP we chose to move on from this project and focus instead on the developing tyrosinase chemistry that was beginning in the lab.

2.5 Conclusion and Future Directions

Cas9 RNPs were successfully encapsulated inside of the MS2 VLP, representing the largest protein ever successfully introduced into the capsid. These results show the potential of the MS2 VLP as a delivery system for intact protein cargo. However, the encapsulation process remains inefficient and many MS2 mutants were not able to re-form capsids after disassembly, limiting some of the potential utility of this scaffold.

We have shown that RNase can access RNA within the MS2 capsid if it is not encapsulating the MS2 RNA genome. This runs against earlier works that showed some protective capacity of the MS2 VLP for RNA cargo, but was validated by simulations and exposure to larger nucleases. The potential for malformed capsids was thoroughly investigated and clear TEM structures coupled with DLS and SEC indicate that even intact capsids showed high sensitivity to RNase activity.

All double mutants of the MS2 FG loop were screened for a protective effect against RNase activity, yielding 19 hits that showed promising initial results but ultimately showed similar RNase sensitivity when isolated. The recent paper by Robinson et al. using the library of FG loop peptide inserts provides an attractive option to examine due to the increased steric bulk into the capsid pore, but it is unclear whether these mutants will tolerate reassembly after treatment with acetic acid.

In summary, this work suggests the potential for a flexible model of the MS2 VLP that contrasts with the rigid sphere normally described. In this model the VLP, unsupported by the tightly coiled MS2 genome, is capable of greater dynamic motion –particularly around the 32 pores framed by the FG loop. These results support the idea of a porous shell capable of allowing even small folded proteins into the capsid interior. Along with the challenges faced in re-assembling MS2 mutants, this work highlights the limitations of the MS2 VLP as an engineering platform and care should be taken in future work to account for these failings.

2.6 Materials and Methods

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. Water (ddH_2O) used in biological procedures or as a reaction solvent was de-ionized using a NANOpure purification system (Barnstead, USA). All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Terrific broth used in Cas9 expression was purchased from Mediatech Inc. (Maryland, VA). Spin concentrators were obtained from EMD Millipore (Millipore, Billerica, MA).

Instruments and Characterization

Mass Spectrometry. Protein bioconjugates were analyzed using an Agilent 1200 series liquid chromatograph (Agilent Technologies, Santa Clara, CA) that was connected in-line with an Agilent 6224 Time-of-Flight (TOF) LC/MS system equipped with a turbospray ion source. Samples were run with a Poroshell 300SB-C18 column (Agilent Technologies, USA). Protein mass reconstruction was performed on the charge ladder with Mass Hunter software (Agilent, USA).

Size Exclusion HPLC. HPLC purification of Encapsulated proteins was performed on an Agilent 1100 Series HPLC Systems (Agilent Technologies, USA) outfitted with an Agilent 1200 series automatic fraction collector. Sample analysis for all HPLC experiments was achieved with an inline diode array detector (DAD) and in-line fluorescence detector (FLD). HPLC of capsids and proteins was accomplished using a BioSep-SEC s4000 column 300 x 7.8 mm column (Phenomenex, Torrance, CA). The samples were eluted in isocratic flow of 10 mM or 100 mM pH 6.8 phosphate buffer at 1 mL/min unless otherwise noted.

UV-Vis Spectroscopic Measurements. UV-Vis spectroscopic measurements were conducted on a Thermo Scientific nanoDrop 1000 scan benchtop spectrophotometer (Thermo Scientific, Wilmington, DE). Spectra were collected in Proteins and Labels mode from 200-800nm, while direct protein concentrations were calculated using Protein A280.

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed in a NuPAGE apparatus (Life Technologies, Hercules, CA), using a 10-20% precast linear gradient polyacrylamide gel (Bio-Rad, Hercules, CA). All protein electrophoresis samples were heated for 10 min at 95 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of disulfide bonds. Gels were run for 80 min at 120 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad). Gels were visualized on Gel-Doc EZ Imager (Bio-Rad).

For PCR reactions products were run on a 2% agarose gel in TEA buffer with 1:10000 SyberSafe DNA gel stain (ThermoFisher Scientific, Wilmington, DE). PCR samples were incubated for 5 min with 5x SyberSafe loading buffer before application to the gel. Gels were run for 2 h at 120 V to separate bands. Commercially available markers (1kb DNA Ladder, New England BioLabs, Ipswich, MA) were applied to at least one lane of each gel for assignment of apparent mass. Visualization of DNA bands was accomplished using a Gel-Doc EZ Imager (Bio-Rad).
For RNA separation a 15% polyacrylamide gel containing 6 M urea was made, then IVT products were mixed with a 95% deionized formamide with trace bromphenol blue and xylene cyanol for tracking. Gels were allowed to run for 4.5 h at 40 mV until xylene cyanol bands reached the base of the gel.

Transmission Electron Microscopy (TEM). TEM images were taken at the U.C. Berkeley Electron Microscope Laboratory using a FEI Tecnai 12 transmission electron microscope (FEI, Hillsboro,

OR) with 120 keV accelerating voltage. Samples were prepared by pipetting 5 µL onto Formvar-coated copper mesh grids (400 mesh, Ted Pella, Redding, CA) for 5 min, followed by exposure to 8 µL of a solution of uranyl acetate (15 mg/mL in dd-H_2O) for 2 min as a negative stain. Excess stain was then removed and the grids were allowed to dry in air for 10 min.

Cas9 expression and purification: Cas9 was prepared alone and as a fusion to the reporter protein mCherry to allow for fluorescent detection of Cas9 containing complexes. Both constructs contained two nuclear localization sequences (NLS) linked via a TEV protease site to maltose binding protein (MBP) with a 6xHis tag. The Cas9 samples were analyzed via ESI-TOF and found to have a mass of 162748 Da for Cas9 alone, or 191169 Da for the Cas9- mCherry fusion, both in line with expected values (Supplemental Figure 2). While Cas9 without mCherry was produced at 2 mg protein per 1 L of bacterial culture, the mCherry fusion was not isolated in sufficient yield to continue initial trials. Cas9-mCherry fusions were generously provided by Dr. Romain Rouet and were used in all trials involving Cas9-mCherry.

Protein production

Production of *p*af 19 MS2: Glycerol stocks of *E. coli* containing plasmids for the T19*p*af mutant of MS2 were generously provided by Ioana Aanei. The T19*p*af MS2 expression was carried out in minimal media following the published protocol[13,14,40]. The pellets were thawed and re-suspended in 20 ml of 20 mM taurine buffer (pH 9) containing 6.5 mM DTT, 6 mM MgCl2, and 10 µg/ml each of DNase and RNase. Following sonication for 10 min, the cells were spun down for 45 min at 11,000 rpm. Next, the supernatant was applied to a DEAE-Sephadex column (GE Healthcare, Little Chalfont, United Kingdom). In 20 mM pH 9 taurine buffer, MS2 eluted first from the DEAE column and was collected and precipitated using an equal volume of saturated aqueous ammonium sulfate. The protein pellets were re-suspended in 10 mM pH 7.2 phosphate buffer and applied to a Sephacryl S1000 column (GE Healthcare). The fractions containing MS2 were then collected and concentrated using Amicon Ultra 100 kDa MWCO centrifugal concentrators (Millipore). A yield of ~3 mg/L culture was obtained for T19*p*af MS2 following two rounds of purification.
MS2 *p*af sequence. **T** in bold represents the T19*p*af mutation site:

ASNFTQFVLVDNGGTGDV**T**VAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI

KVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPS

AIAANSGIY

Production of Cas9: Cas9 plasmids were generously provided by Dr. Romain Rouet. Plasmids were transfected into BL21 *E. coli* cells, as per standard protocol and plated. After overnight incubation the bacteria were inoculated into 1 L of Terrific Broth media with 0.1 g/L ampicillin. Through several rounds of optimization, we found that rotation speeds of 160-200 rpm produced maximal pellet size. Cells were cultured at 37 °C until an OD 600 of 0.8, then reduced

to 18 °C for an hour before induction with 0.1 g/L Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested via centrifugation and lysed using sonication (Fischer Scientific FB120) using a 2 s on/ 5 s off cycle at 70% for 10 min total of sonication.

After lysis, cells debris was pelleted, and the supernatant was passed through a Protino Ni-NTA 5 mL column to isolate the his6-tagged Cas9-MBP fusion. The Cas9 fusion was incubated overnight at 4 °C with TEV protease at 20:1 mass ratio inside 3,500 MWCO dialysis cartridges in a pH 7.4 buffer containing 20 mM HEPES, 150 mM KCl, 1 mM TCEP, 10% glycerol, then passed through a Heparin Sulfate column to capture the cationic Cas9 and eluted with a buffer similar to the above with 1M KCl to allow for efficient ion exchange.

The Cas9 was concentrated in a 100,000 MWCO spin concentrator down to a final volume of 0.5 mL, then if further purification was required, injected onto a Superdex 200 10/300 GL size exclusion column and eluted with 2 column volumes. Out of the 4 primary peaks detected and analyzed via ESI-TOF, only the second peak was found to contain Cas9. Lastly, purified Cas9 was concentrated in a 10,000 MWCO spin concentrator and ran on a 4-12% gel to assess purity.

Amino acid sequences for Cas9 with 2 NLS:
SNATCDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRISYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING
IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV
EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDAYPYDVPDYASLGSGSPKKKRKVEDPKKKRKVD

Amino acid sequence for Cas9-mCherry
SNATCDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRISYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN

LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING
IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV
EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDAYPYDVPDYASLGSGSPKKKRKVEDPKKKRKVDGIGSGSNGSS
GSVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQF
MYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVM
QKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNED
YTIVEQYERAEGRHSTGGMDELYK

Prior to TEV cleavage both Cas9 derivatives contained the his6-tagged maltose binding protein to aid in expression, sequence below:

KSSHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDI
IFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIP
ALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHM
NADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKEL
AKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAV
INAASGRQTVDEALKDAQTNSSSNNNNNNNNNLGIEENLYFQ

sgRNA Production

Our guides were purchased as two part oligomers containing the forward and reverse sequences up to 100 nt in length with a minimum overlap of 50 nt. In addition to the MS2 insertion, these guides differ from standard sgRNA sequences in the stabilization of loop 2 by the addition of two GC base pairs at the base of the loop. Full target sequences are as follows:
sgRNA alone (N1)- total length 131 nucleotides, MW: 42730.2 g/mol

5′**TAATACGACTCACTATAG**CCTCGAACTTCACCTCGGCGGTTTAAGAGCTATGCTGGAAACAGCATAGC
AAGTTTAAATAAGGCTAGTCCGTTATCACGCCGAAAGGCGGGCACCGAGTCGGTGCTTTTTTT3′

MS2 on tetraloop (T1)- total length 151 nucleotides, MW: 49191.2 g/mol

5′**TAATACGACTCACTATAG**CCTCGAACTTCACCTCGGCGGTTTAAGAGCTAGGCCAACATGAGGATCAC
CCATGTCTGCAGGGCCTAGCAAGTTTAAATAAGGCTAGTCCGTTATCACGCCGAAAGGCGGGCACCGAG
TCGGTGCTTTTTTT3′

MS2 on loop 2 (L2-1)- total length 161 nucleotides, MW: 52438.2 g/mol

5′TAATACGACTCACTATAGCCTCGAACTTCACCTCGGCGGTTTAAGAGCTATGCTGGAAACAGCATAGC
AAGTTTAAATAAGGCTAGTCCGTTATCAACTTGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAG
TGGCACCGAGTCGGTGCTTTTTTT3′

The guides were combined with forwards and reverse primers containing extended T7 promotor sites to facilitate transcription in subsequent steps.

Forward: 5′GAAATTAATACGACTCACTATAG3′

Reverse: 5′AAAAAAAGCACCGACTCGGTGC3′

PCR was performed using the Phusion polymerase kit from New England Biolabs. Template concentration 20 nM, primer concentration 1 µM. PCR program was conducted as follows:

Initial melt: 95 °C		01:00 (mm:ss)	
Denature:	95 °C	00:10	\
Anneal:	59 °C	00:10	| x40
Extension:	72 °C	00:10	/
Final extension: 72 °C		01:00	
Hold:	4 °C		

After PCR, guides were analyzed via 2% agarose gel as previously described[41] after purification via QIAGEN QIAquick PCR purification kit (QIAGEN, Germany). Purified product was then added to a 1 mL reaction volume of T7 polymerase in vitro transcription (IVT) in 50 mM Tris buffer pH 8 with 30 mM $MgCl_2$. After incubating overnight at 37 °C RNase free DNase (Promega, Madiso, WI) was added to degrade the DNA template and left at 37 °C for 2 h.

DNase reaction was quenched by addition of 95% formamide with trace bromphenol blue, xylene cyanol and 40 mM EDTA at 60 °C for 5 min, then loaded to a 15% polyacrylamide gel containing 6 M urea and run at 40 watts for 4.5 h until dye ran to bottom of gel. Gels were then collected and visualized using UV shadowing over TLC plates to identify product bands. Bands

were then excised and crushed, then added to 15 ml 300 mM sodium acetate (pH 5) overnight at 4 °C with rocking.

Solutions were filtered through a 0.22 µm filter (Corning, 50ml Tube Top Filter) and precipitated with 40 ml cold 100% ethanol before storage at -80 °C for 1-2 h. Next, sgRNA was pelleted at 4000 g for 45 min, washed with 70% ethanol, and dried via speed vacuum.

RNA was then refolded by heating in the presence of buffered $MgCl_2$ and allowed to cool before storage at -80°C.

Encapsulation in MS2

Encapsulation was carried out in 20 mM Bis-Tris buffer, pH 6.0. A 10 mg/mL solution of MS2 coat protein (~720 µM monomer concentration) in BT buffer was mixed 2:1 vol/vol with glacial acetic acid pre-chilled on ice. The reaction mixture was incubated on ice for 30 minutes and white precipitate was observed. The solution was spun down using a microcentrifuge at 13,200 rcf for 15 min at 4 °C. The supernatant was loaded onto a Nap5 desalting column equilibrated with chilled 1 mM acetic acid. The protein was eluted with 1 mM acetic acid and the fractions containing protein (by absorbance at 280 nm) were pooled. MS2 was added to sufficient BT buffer such that the final MS2 monomer concentration was 15 µM, and Cas9 or control targets such as bacterial alkaline phosphatase (PhoA) were added to a final concentration of 1 µM unless otherwise noted. TMAO in BT buffer was added to a final concentration of 1.8 M. The reassembly was allowed to proceed at 4 °C without stirring for 36 – 48 h before precipitation with a 1:1 volume ratio of saturated ammonium sulfate. Precipitation ran for a minimum of 90 min before the capsids were collected via centrifugation for 45 min at 17,000 rcf, 4 °C. Precipitated capsids were then re-suspended in ~50 uL 10 mM phosphate buffer, pH 7.2 and concentration measure via nanoDrop UV/Vis.

Synthesis of Cas9- -NHS Fluorophore conjugates

A solution of 5 mM Cas9- in 50 mM pH 7.5 phosphate buffer was prepared and combined with NHS-AlexaFluor 568 (Life Technologies, Carlsbad, CA) from a 1 mM DMSO stock (50 equiv.). The mixture was briefly vortexed, and the reaction was allowed to proceed at room temperature for 1.5 h. The excess dye molecules were removed by spin concentration with a MWCO of 30 kDa (Millipore, Billerica, MA) in 10 mM pH 7.0 phosphate buffer. The resulting Cas9 - fluorophore conjugates contained 1-5 dye per protein molecule, as quantified by absorbance on a nanoDrop and ESI-TOF.

Generation of MS2 dual mutant library

Much of this work appears previously.[42]

***Strains*.** MegaX DH10B *Escherichia coli* electrocompetent cells (ThermoFisher Scientific, catalog no. C640003) were used for all library experiments, and DH10B chemically competent cells produced in house were used for expression of individual variants of interest. Overnight cultures from a single colony were grown for 16−20 h at 37 °C while being shaken at 200 rpm in LB-Miller medium (Fisher Scientific, catalog no. BP1426-2) with 32 mg/L chloramphenicol. Expressions were subcultured in a 1:100 ratio into 2×YT medium (Teknova, catalog no. Y0210)

with 32 mg/L chloramphenicol and allowed to express overnight at 37 °C while being shaken at 200 rpm.

***Library Generation*.** To generate libraries with two-amino acid mutations in the FG loop, we modified a library generation strategy developed by the Bolon lab, known as EMPIRIC cloning.[43] EMPIRIC cloning uses a plasmid with a self-encoded removable fragment (SERF) surrounded by inverted BsaI restriction sites. With this setup, BsaI digestion simultaneously removes both SERF and BsaI sites. These plasmids are termed entry vectors, and the SERF in this study encodes constitutively expressed GFP to permit green/white screening. We used a previously described entry vector that replaced a 26-codon segment flanking the FG loop in the MS2 CP with the SERF.[43,44] Single-stranded DNA primers with all 15 combinations of degenerate codons were purchased, enabling overlap extension polymerase chain reaction (PCR) to generate double-stranded DNA with all possible pairwise combinations in this six-residue region. These primers were resuspended in water, pooled, and diluted to a final concentration of 50 ng/µL. The reverse strand was filled in by overlap extension PCR with a corresponding forward primer. The amplified, double-stranded DNA was purified using a PCR Clean-up Kit (Promega, catalog no. A9282). The purified DNA was diluted to 1−5 ng/µL and cloned into the described entry vector using established Golden gate cloning techniques.[45] The ligated plasmids were desalted on membranes (Millipore Sigma, catalog no. VSWP02500) for 20 min and then transformed into MegaX DH10B *E. coli* electrocompetent cells (ThermoFisher Scientific, catalog no. C640003). Following electroporation and recovery, cells were plated onto two large LB-A plates (VWR, catalog no. 82050- 600) with 32 µg/mL chloramphenicol and allowed to grow at 37 °C overnight. The colony number varied, but every transformation yielded a number of colonies that was at least 50 times greater than the library size. This protocol was repeated in full for three total biological replicates that are fully independent from library generation through selection.

Size Selection. Colonies were scraped from plates into LB-M and allowed to grow for 2 h. Each library was then subcultured at a 1:100 ratio into 1 L of 2×YT (Teknova, catalog no. Y0210) and allowed to grow to an OD600 of 0.6, when they were induced with 0.1% arabinose. Libraries of variants were expressed overnight at 37 °C. Cultures were then harvested, resuspended in 10 mM phosphate buffer (pH 7.2) with 2 mM sodium azide, and sonicated. Libraries were subjected to two rounds of 50% (w/v) ammonium sulfate precipitation, followed by fast protein liquid chromatography (FPLC) size exclusion chromatography purification to select for well-formed VLPs**.**

FPLC SEC (assembly selection). MS2 CP libraries or individual variants were purified on an Akta Pure 25 L FPLC system with a HiPrep Sephacryl S-500 HR column (GE Healthcare Life Sciences, catalog no. 28935607) size exclusion chromatography (SEC) column via isocratic flow with 10 mM phosphate buffer (pH 7.2), 200 mM sodium chloride, and 2 mM sodium azide. Fractions containing MS2 coat protein were collected for further analysis.

***Heat Selection*.** Following assembly selection (FPLC purification), libraries were incubated at 50 °C for 10 min. The buffer from FPLC purification [10 mM phosphate buffer (pH 7.2), 200 mM sodium chloride, and 2 mM sodium azide] was used for these studies. Precipitated VLPs were pelleted by centrifugation, and well-formed VLPs were isolated by semipreparative high-performance liquid chromatography (HPLC) SEC. Fractions containing VLP were combined and subjected to RNA extraction, barcoding, and high-throughput sequencing.

Cell Culture Protocols

All mammalian cell cultures were maintained in a 37°C incubator, at 5% CO2. HEK293T (293FT; Thermo Fisher Scientific, #R70007) human kidney cells and derivatives thereof were grown in Dulbecco's Modified Eagle Medium (DMEM; Corning Cellgro, #10-013-CV) supplemented with 10% fetal bovine serum (FBS; Seradigm, #1500-500), and 100 Units/ml penicillin and 100 µg/ml streptomycin (100-Pen-Strep; GIBCO #15140-122).

HEK-RT1 and HEK-RT6 cells were produced as previously described[46] and tested for absence of mycoplasma contamination (UC Berkeley Cell Culture facility) by fluorescence microscopy of methanol fixed and Hoechst 33258 (Polysciences #09460) stained samples.

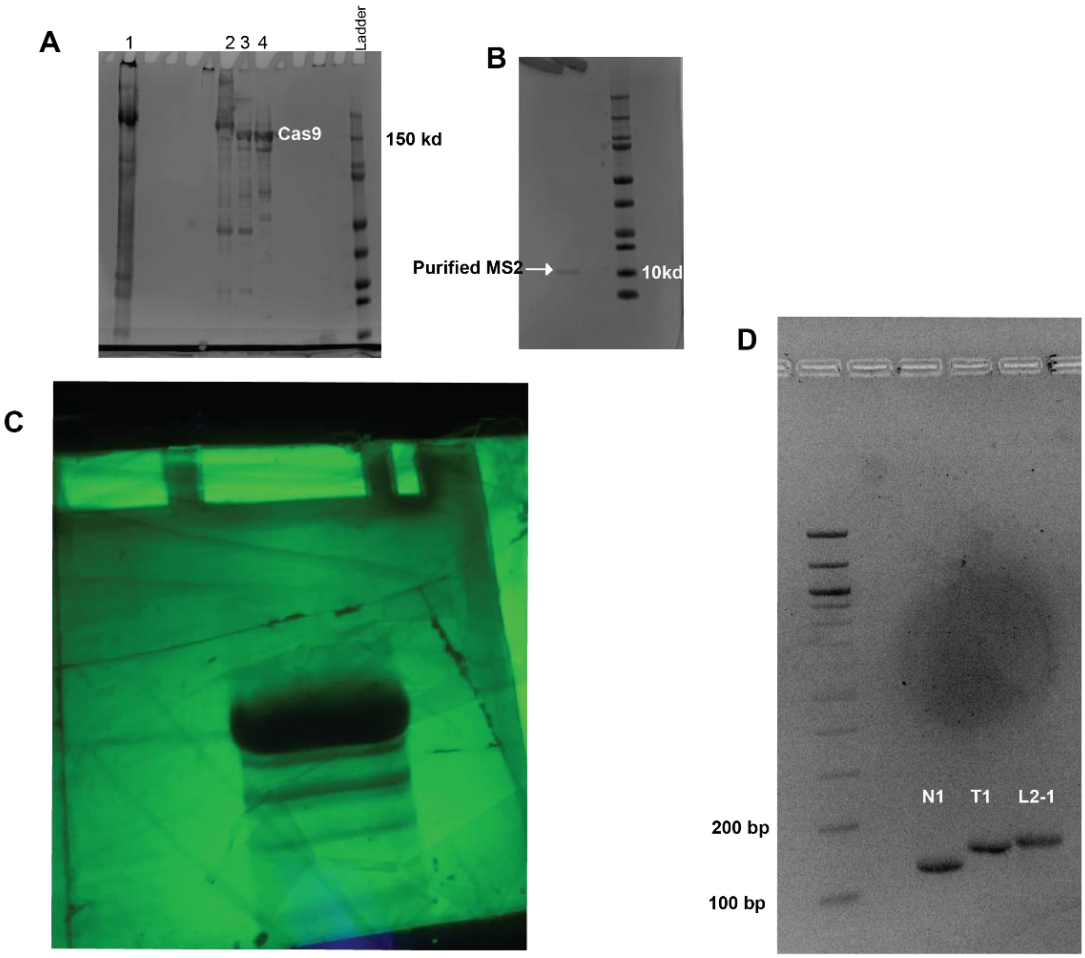
Preparation of Cas9 for cell culture trials*.* Cas9 RNP was prepared based on previously described methods [47]. To prepare the Cas9 RNP complexes, modified Cas9 protein was incubated with sgRNA at a 1:1.2 molar ration. Briefly, we added sgRNA to Buffer #1 (25mM NaP_i, 150 mM NaCl, 200 mM trehalose, 1 mM $MgCl_2$), then added the Cas9 to the sgRNA, slowly with swirling, and incubated at 37 °C for 10 min to form RNP complexes. RNP complexes were filtered prior to use through a 0.22 µm Costar 8160 Filter pre-wet with 200 µL of Buffer #1. If needed, the RNP sample was concentrated with a 0.5 mL Millipore Ultra 100 kDa cutoff filter, part # UFC510096, until the desired volume was obtained.

Human Endothelial Kidney (HEK) cell line assay*.* HEK cells were treated based on a previously described protocol[47]. Briefly, NPCs were isolated from cortices from embryonic day 13.5 Ai9-tdTomato homozygous mouse embryos. Cells were cultured as neurospheres in NPC medium: DMEM/F12 with glutamine, Na-Pyruvate, 10 mM HEPES, non-essential amino acid, penicillin and streptomycin, 2-mercaptoethanol, B-27 without vitamin A, N2 supplement, bFGF and EGF, both 20 ng/ml as final concentration. NPCs were passaged using MACS Neural Dissociation Kit (Papain) cat# 130-092-628 following manufacturer's protocol. bFGF and EGF were refreshed every other day and passaged every 6 days. The NPC line was authenticated by immunocytochemistry marker protein staining. They were tested for mycoplasma using Hoechst stain with visual analysis and were negative.

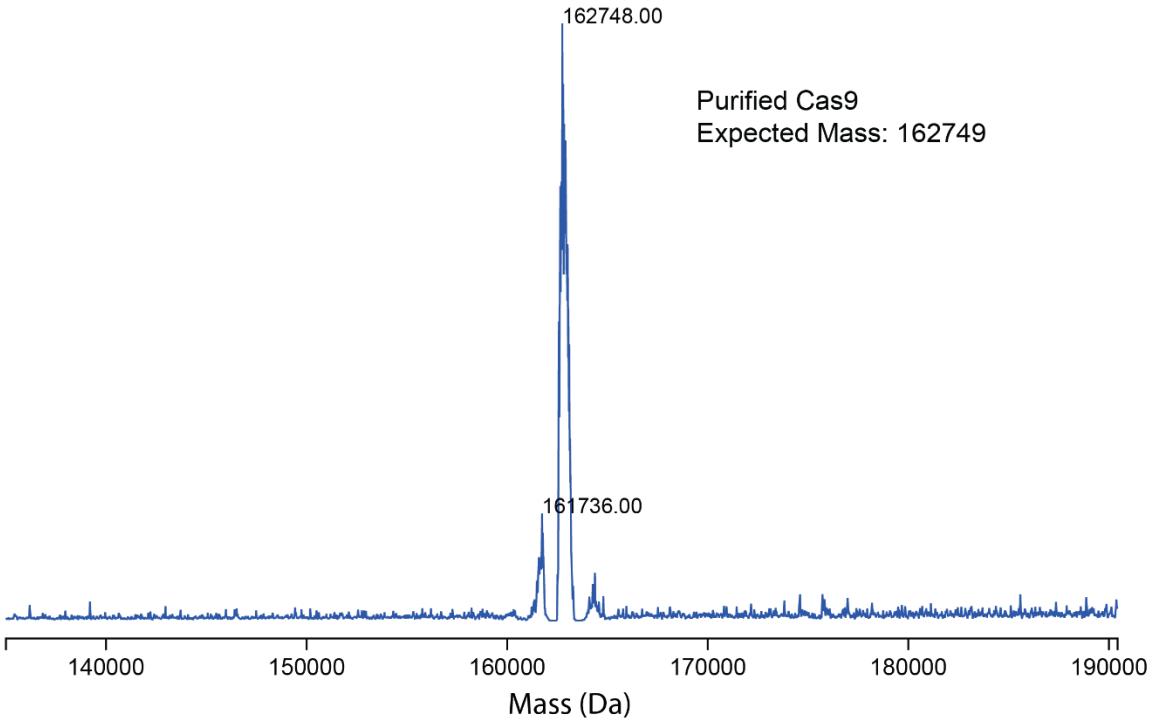
Direct delivery of Cas9 RNP. To test direct delivery of Cas9 ribonucleoprotein (RNP) complexes, RNPs targeting the STOP cassette at the tdTomato locus of cells derived from Ai9-tdTomato transgenic mice (sgRNA-298, 5'-AAGTAAAACCTCTACAAATG-3')[47] were pre-assembled. Assembled RNPs were added to the media of Ai9 neural progenitor cells (NPCs) and incubated for 24 h. Cells were then washed twice with 200 U/ml heparin in DMEM-based media and allowed to grow for an additional 24-48 h before analysis. The total amount of RNP per well of a 96-well plate was titrated from 10 pmol to 100 pmol. Percentages of tdTomato-positive (successfully edited) NPCs were quantified by flow cytometry (Attune Nxt acoustic focusing cytometer,

Thermo Fisher Scientific) with tdTomato fluorescence measured at the 561nm laser line using the integrated analysis tool (Attune NxT software, v3.1.0). We routinely acquired at least 10,000-30,000 events per sample.

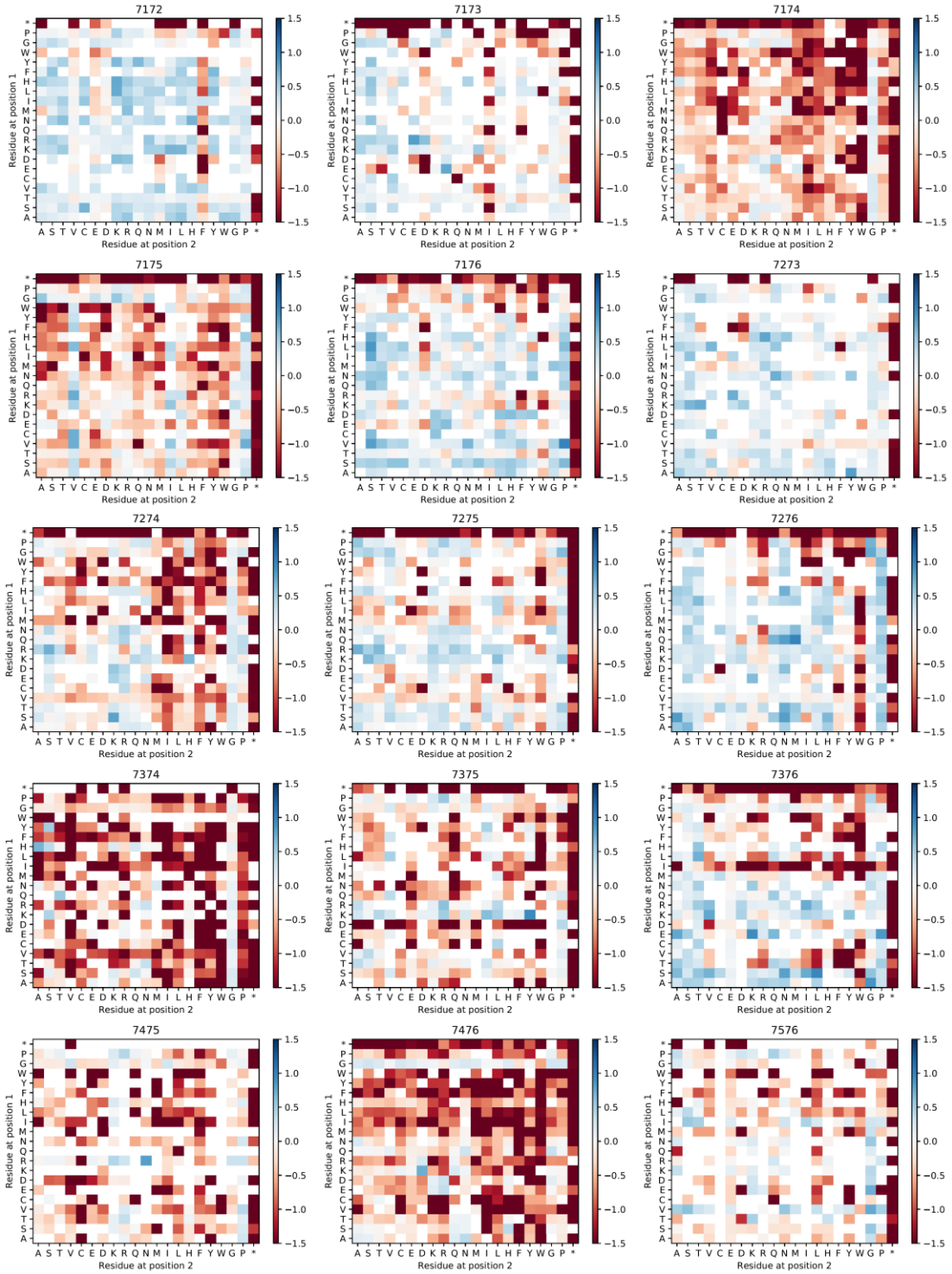
2.7 Supplemental Figures



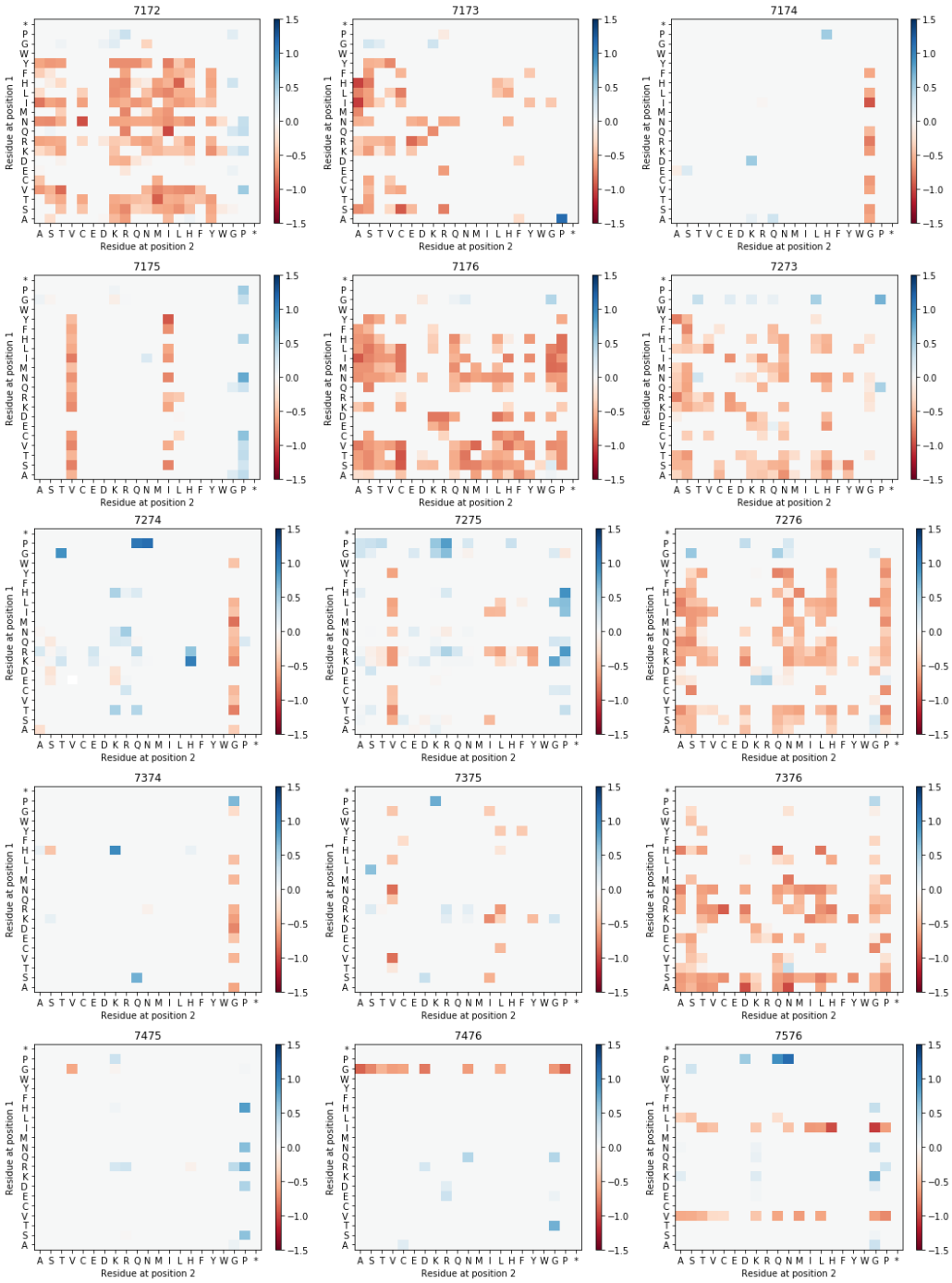
Supplementary Figure 1. Protein and sgRNA Gels A) Cas9 Purification starting from crude lysate (1) through NiNTA purification (2) TEV cleavage and subsequent nickle cleanup (3) and final gel filtration (3). B) Purified MS2 running at 13kd. C) UV shadow of sgRNA produced from IVT, the large band at top is the desired product. D) Gel analysis of IVT DNA templates produced via PCR. Clear bands at 131, 151, and 161 indicate success.



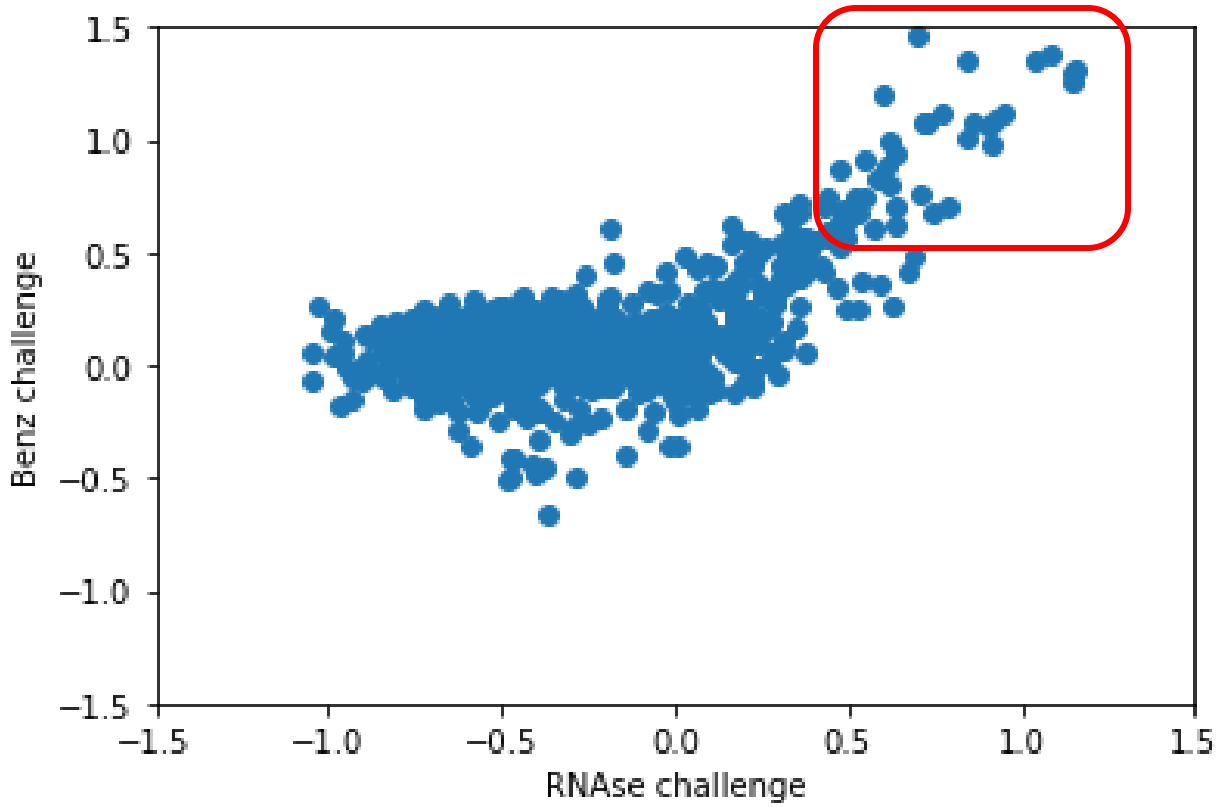
Supplementary Figure 2: ESI-TOF spectra for Cas9 expression after TEV cleavage of MBP. Isolated peak at 162748 Da closely matches the expected mass of 162794 Da.



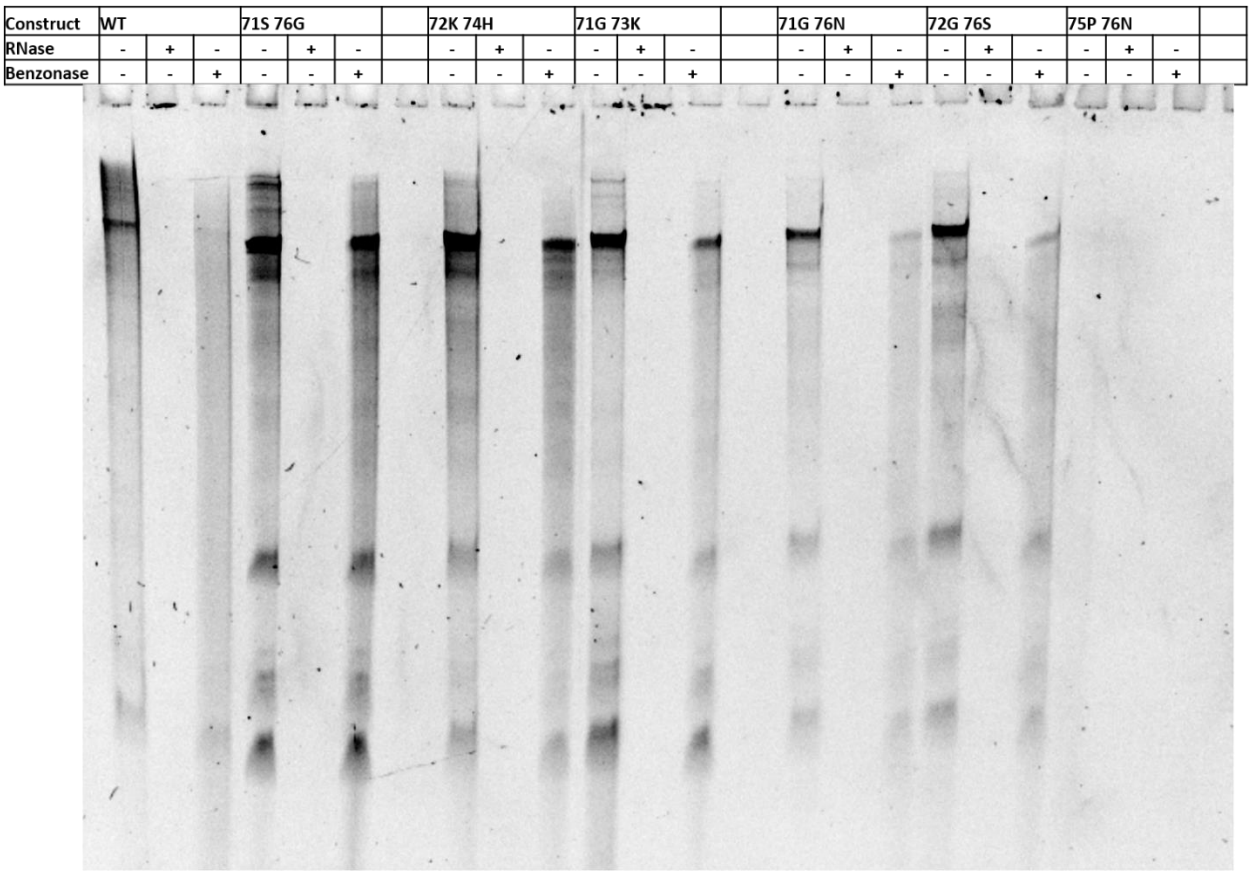
Supplemental Figure 3- library heatmap after exposure to temperature control and size exclusion, showing enrichment or loss in blue and red respectively. Each square represents two residues that were co-varied in the analysis, with the axis on left representing all 20 amino acids at the first position, and the bottom axis representing all 20 amino acids at the second position.



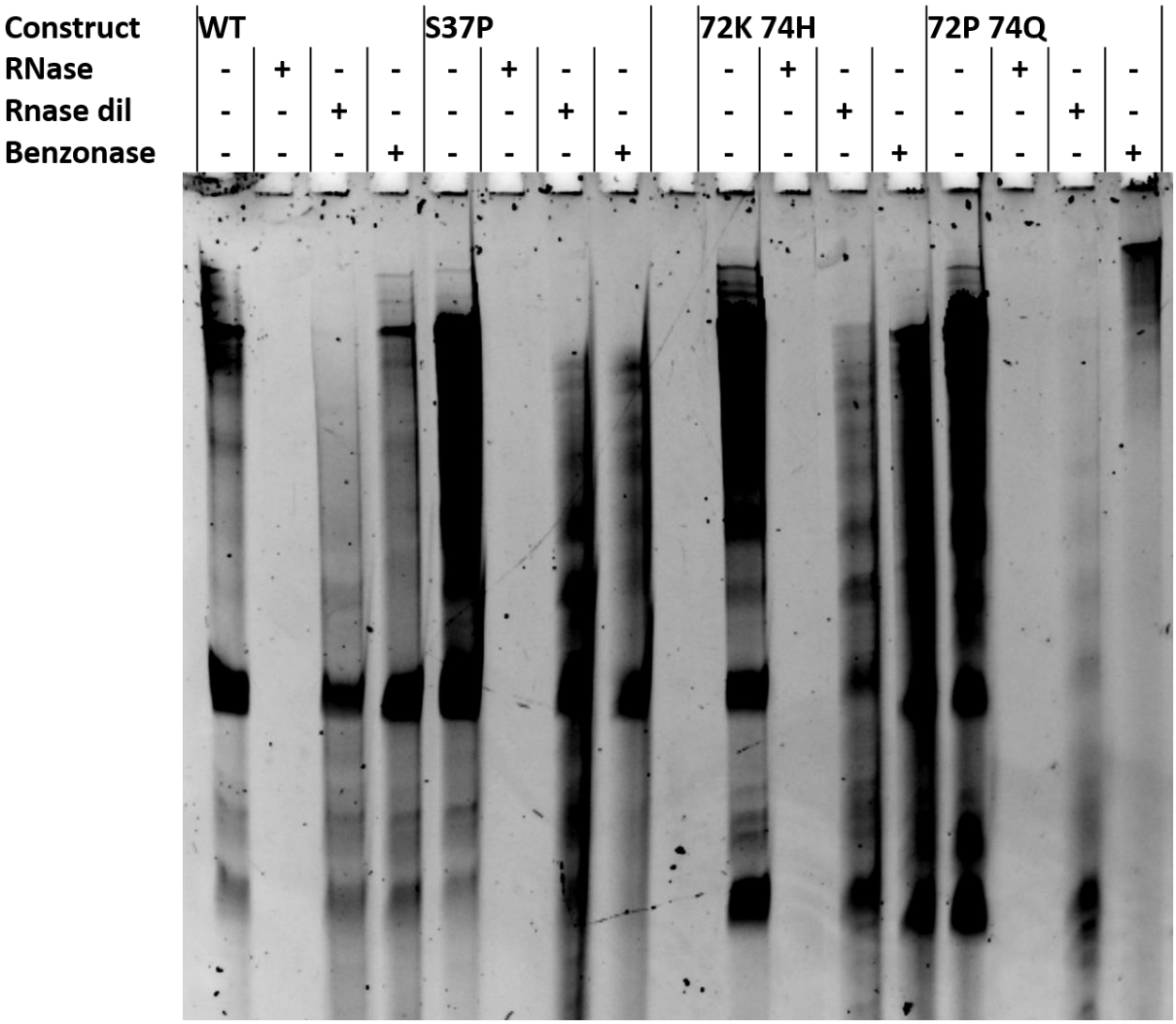
Supplemental Figure 4- library heatmap after exposure to RNase conditions. Enrichment or loss shown in blue and red respectively while grey represents insufficient reads of the given mutant to detect. Each square represents two residues that were co-varied in the analysis, with the axis on left representing all 20 amino acids at the first position, and the bottom axis representing all 20 amino acids at the second position. The consistency of proline residues in these assays shows potential protective effect.



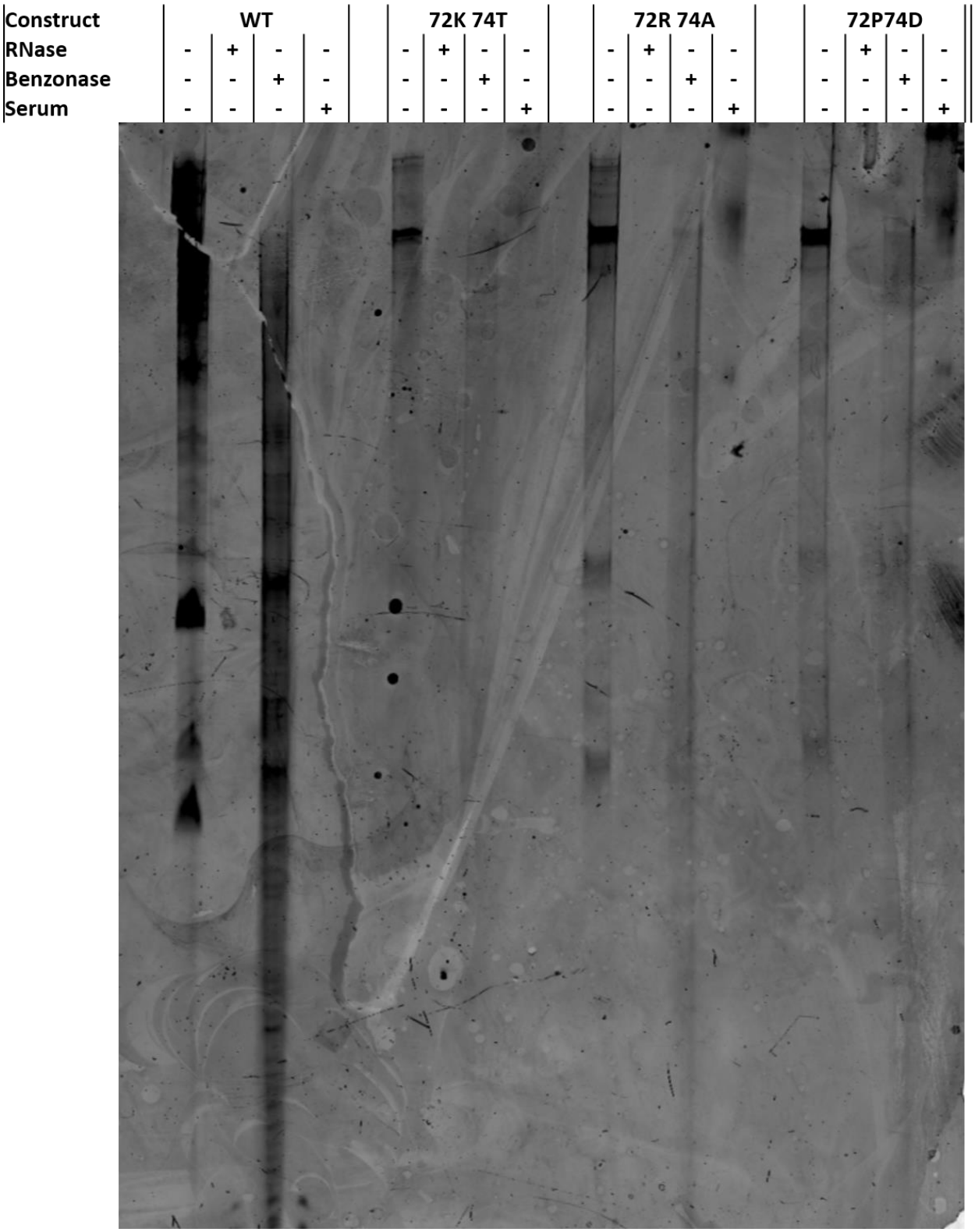
Supplemental Figure 5 – plot of relative abundance of each double mutant in the RNase and Benzonase challenges. Mutants enriched in both appear in the top right corner of the figure in the red square.



Supplemental Gel 1- polyacrylamide urea gel showing MS2 mutants exposed to RNase and Benzonase. Samples were incubated for 30 min at RT prior to extraction and visualization of remaining RNA. Sample names indicate identity of isolated MS2 double mutant.



Supplemental Gel 2- polyacrylamide urea gel showing MS2 mutants exposed to RNase and Benzonase. Samples were incubated for 30 min at RT prior to extraction and visualization of remaining RNA. Sample names indicate identity of isolated MS2 double mutant.



Supplemental Gel 3- polyacrylamide urea gel showing MS2 mutants exposed to RNase and Benzonase. Samples were incubated for 30 min at RT prior to extraction and visualization of remaining RNA. Sample names indicate identity of isolated MS2 double mutant.

2.8 References

1. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2013**, 1–9 (2013).
2. Mout, R., Ray, M., Lee, Y. W., Scaletti, F. & Rotello, V. M. In Vivo Delivery of CRISPR/Cas9 for Therapeutic Gene Editing: Progress and Challenges. *Bioconjug. Chem.* **28**, 880–884 (2017).
3. Jo, Y. Il, Suresh, B., Kim, H. & Ramakrishna, S. CRISPR/Cas9 system as an innovative genetic engineering tool: Enhancements in sequence specificity and delivery methods. *Biochim. Biophys. Acta - Rev. Cancer* **1856**, 234–243 (2015).
4. Mehta, A. & Merkel, O. M. Immunogenicity of Cas9 Protein. *J. Pharm. Sci.* **109**, 62–67 (2020).
5. Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, 249–254 (2019).
6. Moreno, A. M. *et al.* Immune-orthogonal orthologues of AAV capsids and of Cas9 circumvent the immune response to the administration of gene therapy. *Nat. Biomed. Eng.* **3**, 806–816 (2019).
7. Witus, L. S. & Francis, M. B. Using synthetically modified proteins to make new materials. *Acc. Chem. Res.* **44**, 774–783 (2011).
8. Elsohly, A. M. *et al.* Synthetically Modified Viral Capsids as Versatile Carriers for Use in Antibody-Based Cell Targeting. *Bioconjug. Chem.* **26**, 1590–1596 (2015).
9. Farkas, M. E. *et al.* PET Imaging and Biodistribution of Chemically Modified Bacteriophage MS2. *Mol. Pharm.* **10**, 69–76 (2013).
10. Fu, Y. & Li, J. A novel delivery platform based on Bacteriophage MS2 virus-like particles. *Virus Res.* **211**, 9–16 (2016).
11. Pan, Y. *et al.* Development of a microRNA delivery system based on bacteriophage MS2 virus-like particles. *FEBS J.* **279**, 1198–1208 (2012).
12. ElSohly, A. M. *et al.* ortho-Methoxyphenols as Convenient Oxidative Bioconjugation Reagents with Application to Site-Selective Heterobifunctional Cross-Linkers. *J. Am. Chem. Soc.* **139**, 3767–3773 (2017).
13. Mehl, R. A. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).
14. Carrico, Z. M., Romanini, D. W., Mehl, R. A. & Francis, M. B. Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* **0**, 1205 (2008).
15. Johansson, H. E., Liljas, L. & Uhlenbeck, O. C. RNA recognition by the MS2 phage coat protein. *Semin. Virol.* **8**, 176–185 (1997).
16. Golmohammadi, R., Valegård, K., Fridborg, K. & Liljas, L. The Refined Structure of Bacteriophage MS2 at 2·8 Å Resolution. *J. Mol. Biol.* **234**, 620–639 (1993).
17. Peabody, D. S. & Lim, F. Complementation of RNA binding site mutations in MS2 coat protein heterodimers. *Nucleic Acids Res.* **24**, 2352–2359 (1996).
18. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).

19. Legendre, D. & Fastrez, J. Production in Saccharomyces cerevisiae of MS2 virus-like particles packaging functional heterologous mRNAs. *J. Biotechnol.* **117**, 183–194 (2005).

20. Li, J. *et al.* Messenger RNA vaccine based on recombinant MS2 virus-like particles against prostate cancer. *Int. J. Cancer* **134**, 1683–1694 (2014).

21. Anders, C. & Jinek, M. In Vitro Enzymology of Cas9. in *Methods in enzymology* **546 BT**-, 1–20 (Elsevier Inc., 2014).

22. Capehart, S. L., Coyle, M. P., Glasgow, J. E. & Francis, M. B. Controlled integration of gold nanoparticles and organic fluorophores using synthetically modified MS2 viral capsids. *J. Am. Chem. Soc.* **135**, 3011–3016 (2013).

23. Glasgow, J. E., Capehart, S. L., Francis, M. B. & Tullman-Ercek, D. Osmolyte-mediated encapsulation of proteins inside MS2 viral capsids. *ACS Nano* **6**, 8658–8664 (2012).

24. Doronina, S. O. *et al.* Novel peptide linkers for highly potent antibody-auristatin conjugate. *Bioconjug. Chem.* **19**, 1960–1963 (2008).

25. Peabody, D. S. Subunit fusion confers tolerance to peptide insertions in a virus coat protein. *Arch. Biochem. Biophys.* **347**, 85–92 (1997).

26. Spokoyny, A. M. *et al.* A Perfluoroaryl-Cysteine S N Ar Chemistry Approach to Unprotected Peptide Stapling. (2013). doi:10.1021/ja400119t

27. Peabody, D. S. *et al.* Immunogenic Display of Diverse Peptides on Virus-like Particles of RNA Phage MS2. *J. Mol. Biol.* **380**, 252–263 (2008).

28. Biniossek, M. L., Nägler, D. K., Becker-Pauly, C. & Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *J. Proteome Res.* **10**, 5363–5373 (2011).

29. Choe, Y. *et al.* Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *J. Biol. Chem.* **281**, 12824–12832 (2006).

30. Ren, G. *et al.* Non-invasive imaging of cysteine cathepsin activity in solid tumors using a 64Cu-labeled activity-based probe. *PLoS One* **6**, 1–9 (2011).

31. Rittner, K. *et al.* New Basic Membrane-Destabilizing Peptides for Plasmid-Based Gene Delivery in Vitro and in Vivo. *Mol. Ther.* **5**, 104–114 (2002).

32. Kichler, A., Leborgne, C., Marz, J., Danos, O. & Bechinger, B. Histidine-rich amphipathic peptide antibiotics promote efficient delivery of DNA into mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1564–1568 (2003).

33. Brown, W. L. *et al.* RNA bacteriophage capsid-mediated drug delivery and epitope presentation. *Intervirology* **45**, 371–380 (2002).

34. Ashley, C. E. *et al.* Cell-specific delivery of diverse cargos by bacteriophage MS2 virus-like particles. *ACS Nano* **5**, 5729–5745 (2011).

35. Wei, B. *et al.* Development of an antisense RNA delivery system using conjugates of the MS2 bacteriophage capsids and HIV-1 TAT cell penetrating peptide. *Biomed. Pharmacother.* **63**, 313–318 (2009).

36. Hartman, E. C. *et al.* Quantitative characterization of all single amino acid variants of a viral capsid-based drug delivery vehicle. *Nat. Commun.* **9**, 1–11 (2018).

37. ElSawy, K. M. The impact of viral RNA on the association free energies of capsid protein assembly: bacteriophage MS2 as a case study. *J. Mol. Model.* **23**, (2017).

38. Koning, R. I. *et al.* Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat. Commun.* **7**, 12524 (2016).

39. Ni, C. Z. *et al.* Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure* **3**, 255–263 (1995).

40. Tong, G. J., Hsiao, S. C., Carrico, Z. M. & Francis, M. B. Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* **131**, 11174–11178 (2009).

41. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (80-.).* **339**, 823–826 (2013).

42. Hartman, E. C. *et al.* Experimental Evaluation of Coevolution in a Self-Assembling Particle. *Biochemistry* **58**, 1527–1538 (2019).

43. Hietpas, R. T., Jensen, J. D. & Bolon, D. N. A. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci.* **108**, 7896–7901 (2011).

44. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. *PLoS One* **4**, e5553 (2009).

45. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).

46. Oakes, B. L. *et al.* CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification. *Cell* **176**, 254-267.e16 (2019).

47. Staahl, B. T. *et al.* Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.* **35**, 431–434 (2017).

Chapter 3. Site-specific bioconjugation through enzyme catalyzed tyrosine-cysteine bond formation

Abstract:
The synthesis of protein-protein and protein-peptide conjugates is an important capability for producing vaccines, immunotherapeutics, and targeted delivery agents. Herein we show that the enzyme tyrosinase is capable of oxidizing exposed tyrosine residues into *ortho*-quinones that react rapidly with cysteine residues on target proteins. This coupling reaction occurs under mild aerobic conditions and has the rare ability to join full-size proteins in under 2 h. The utility of the approach is demonstrated for the attachment of cationic peptides to enhance the cellular delivery of CRISPR-Cas9 by 20-fold, and for the coupling of reporter proteins to a cancer-targeting antibody fragment without losing its cell-specific binding ability. The broad applicability of this technique provides a new building block approach for the synthesis of protein chimeras.

3.1 Tyrosinases for protein modification

The covalent modification of proteins while preserving native function has long been a goal of chemical biology research[1–3]. Unfortunately, it can be difficult to target chemically distinct functional groups in protein sequences, leading to heterogeneous product mixtures with many modification strategies. This challenge is particularly notable for the synthesis of protein-protein and protein-peptide conjugates, which can serve as highly valuable constructs for vaccines, immunotherapy agents, signaling peptides with enhanced circulation properties, and targeted delivery vehicles. Currently, the coupling of two biomolecules is usually achieved using heterobifunctional crosslinking agents (which are rarely site-selective for both reactive species)[1] or Click-type reactions[4] between bioorthogonal functional groups[5] that must be introduced using artificial amino acid incorporation, metabolic engineering, or additional synthetic steps. The coupling of native functional groups on proteins and peptides can be achieved through native chemical ligations and sortase-based techniques[6,7], but these approaches can only be used at the N- and C-termini and can limit the quantities of bioconjugates that can be obtained. In some cases it is possible to fuse multiple protein sequences at the genetic level for recombinant expression[8,9]. While this is a time-tested approach for producing chimeric proteins, such methods often require significant engineering to determine optimal linker lengths[10], can encounter folding problems during the expression of constructs of increasing size[11], and are again limited to attachment via the N and C termini.

In this report, a new method is described for the efficient generation of protein-peptide and protein-protein conjugates through the direct coupling of solvent-exposed tyrosine residues to cysteine sulfhydryl groups. This reaction requires only catalytic amounts of tyrosinase and adventitious oxygen, affording complex, multicomponent bioconjugates in under 1 h under mild reaction conditions and low stoichiometric excess. This strategy is demonstrated as a convenient method for the installation of cell penetrating peptides on proteins, such as Cas9, and the coupling of entire protein domains. Most importantly, this reaction demonstrates the very rare ability to couple full-size proteins together through positionally defined linkages using native amino acids, providing a powerful new tool for the construction of multifunctional biomolecular constructs.

3.2 Tyrosinase background and oxidative coupling

In previous work, our lab has shown that proteins bearing aniline and N-terminal proline nucleophiles can be modified with *o*-quinone and *o*-iminoquinone electrophiles in high yields[12,13]. The oxidized intermediates in these reactions have typically been generated through the *in situ* oxidation of *o*-aminophenols or similar species using periodate[12] or ferricyanide anions[13]. While effective in many contexts, these stoichiometric reagents can exhibit collateral oxidation of methionine thioethers and surface exposed cysteines to varying degrees. A recent breakthrough in this chemistry has involved the use of the tyrosinase enzymes to generate the same reactive *o*-quinone electrophiles using simple phenols and oxygen[14]. The small amount of molecular oxygen that is dissolved in the buffers is sufficient for the reaction to proceed, and water is the only byproduct. This reaction addresses two important limitations of the previous

versions of this chemistry. First, it is no longer necessary to prepare and install the oxygen-sensitive *o*-aminophenol and *o*-catechol groups; instead, simple tyramine and tyrosine derivatives can be used. Second, the need for an additional oxidant that must later be removed is avoided. Importantly, the enzymatic method for triggering these reactions is selective for phenols that are exposed in solution and does not cause off-target oxidation of sensitive functional groups.

It has been known for some time that tyrosinases can act upon tyrosine residues in proteins and peptides[15]. Typically these reactions have been used to achieve protein crosslinking at high substrate concentrations, involving indiscriminate conversion of tyrosine residues within the peptide backbone to *o*-quinines followed by reactions with a variety of nucleophiles[16–19]. Previous reports have followed reaction progress using UV or other spectroscopic analyses, and have explored the impact of tyrosine oxidation on enzyme activity; however the sequence specificities and sites of oxidation have not been characterized well. In terms of site-selective coupling chemistry, other labs have recently shown that tyrosine-tagged immunoglobulins can be oxidized and subjected to subsequent hetero-Diels-Alder reactions with cyclooctynes[20], and abTYR has been used to activate an ovalbumin peptide tag for coupling to substituted boronic acids[21]. Work from our lab has showed that abTYR and a new tyrosinase derived from *Bacillus megatarium* can activate phenols in both small molecule and protein substrates for coupling to N-terminal proline residues and anilines (Figure 1a, top and bottom reaction pathways)[14,22]. Taken together, these studies highlight the utility of tyrosinase enzymes to generate transient *o*-quinone electrophiles in dilute aqueous solutions, allowing highly chemoselective couplings that are not realized at the higher substrate concentrations evaluated previously.

Key to the present work is the observation that cysteine thiols on full-size biomolecules participate readily in this chemistry at low equivalencies, allowing the first general coupling reaction for two native and readily encodable functional groups on protein surfaces (Figure 1a, middle pathway). Although we have known that these residues can participate in oxidative coupling reactions to varying degrees, the reliance on stoichiometric oxidants in our previous studies led to numerous reaction byproducts that are completely avoided using the new tyrosinase oxidation system. Other reports have noted reactivity between small molecule thiols and *o*-quinones, but with minimal product characterization[23,24].

3.3 Tyrosinase and MS2

In the context of site-selective bioconjugation, this reactivity pathway was first tested on genome-free MS2 bacteriophage capsids bearing N87C mutations in each of the subunit proteins[25,26]. Each of these assemblies provides 180 cysteine residues facing the inside of a 27 nm sphere[27]. The capsid structures possess a series of 2 nm pores that are large enough for peptides and small molecules to access the inner surface (Figure 1b)[26,28,29]. Modification was achieved by exposing 10 μM (based on monomer) MS2 N87C to 200 μM α-endorphin peptide (Figure 1c) in the presence of 167 nM abTYR for 30 min at room temperature in 20 mM phosphate buffer, pH 6.5. Clean conversion of each capsid monomer to a single peptide conjugate was observed (Figure 1d). This represents the installation of >170 peptide groups inside each capsid despite the high degree of steric crowding. Subsequent exposure to the

common cysteine capping reagent *N*-ethylmaleimide (NEM) showed no further modification (Figure 1e, Figure S1), supporting the cysteine selectivity of the reaction. Similarly, no α-endorphin addition was observed for MS2 N87C that had been previously coupled to NEM, and there was no modification of wild type MS2 capsids lacking exposed cysteine residues (Figure S1). Equally importantly, there was no background oxidation observed on any of the MS2 proteins, which each contain four tyrosine residues (Figure 1c, Figure S1). This is presumably because tyrosine residues embedded in secondary and tertiary structural elements cannot access the deep binding pocket of the enzyme active site[30]. These results demonstrate that covalent bonds can be formed between these two native amino acid residues using an efficient and operationally simple procedure.

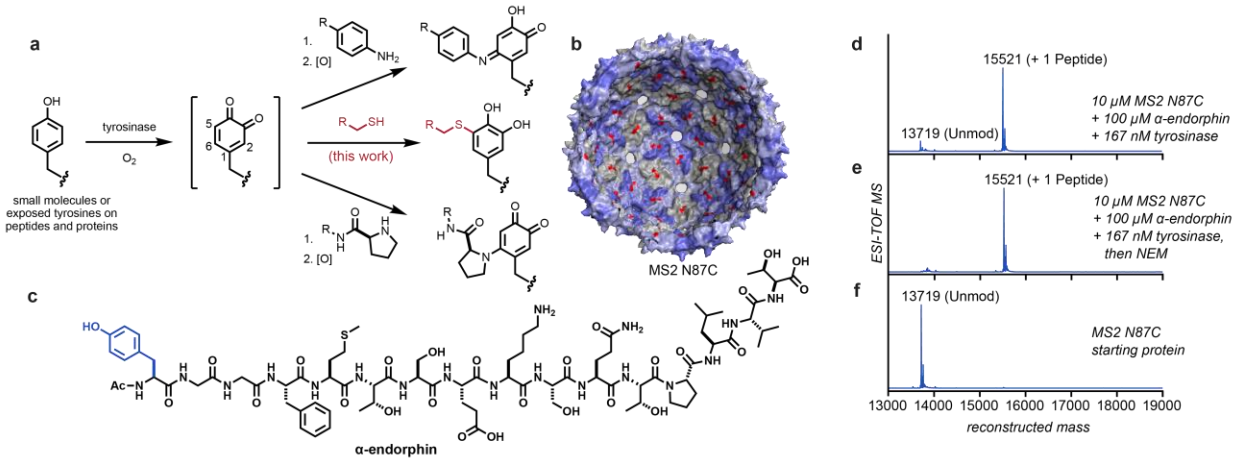


Fig. 1. Tyrosinase-mediated oxidative coupling reactions. (a) Phenols are oxidized by tyrosinase to yield *o*-quinone intermediates that couple with a variety of nucleophiles on biomolecules. In this work, the addition of cysteine thiolates is explored (shown in red). (b) An MS2 viral capsid is shown, with 180 interior cysteine residues (N87C) indicated in red. PDB ID: 2MS2. Pores in the capsid shell allow the entry of peptides and small molecules. (c) The structure of α-endorphin is shown, with the targeted tyrosine residue in blue. (d-f) Coupling reactions were screened using ESI-TOF MS, showing full modification of the MS2 N87C capsid without off-target oxidation. Reaction conditions: pH 6.5 phosphate buffer, RT, 30 min. Expected mass values: MS2 N87C $[M+H]^+$ = 13719; MS2 N87C–*N*-ethylmaleimide (NEM) adduct $[M+H]^+$ = 13844; MS2 N87C–endorphin *o*-hydroquinone product $[M+H]^+$ = 15521.

3.4 NMR and modelling of tyrosinase thiol adducts

In order to characterize the chemical structure of the bioconjugation product, we conducted NMR experiments for a small molecule analog. Somewhat surprisingly, the product results from thiol addition to the 5 position of the *o*-quinone ring (numbering from the amino acid connection point as shown in Figure 1a), and it remains predominantly in the catechol form (Figure 2a, Figures S2, S3). Additional MS data further support the catechol structure. This

product contrasts with the previously observed structures of proline and aniline additions to *o*-quinone species that add at the 6-position and tend to remain in the oxidized quinone form (Figure 1a)[12–14]. MS data confirmed that the thiol addition products on larger biomolecules also existed predominantly in the reduced catechol form.

To establish the degree of quinoid versus catechol nature of the product further, we exposed both thiol and N-terminal proline coupling adducts of Y200C GFP to CH_3ONH_2 (methoxyamine) overnight to effect oxime formation with any ketone groups that were present, as has been recently shown[31]. Notably, only the proline adducts exhibited any detectable reactivity to the methoxyamine, (Figure S4).

The energetics of the *o*-quinone-thiol coupling reaction were investigated using DFT calculations. Briefly, geometry optimized structures were evaluated using B3LYP-D3/6-31G** to obtain vibrational spectra, and ωB97M-V/6-311G^{++}3df-3pd was used to obtain accurate ground state electronic energies[32]. These results were combined to yield estimates of the thermodynamic parameters for the reaction. A CPCM solvent model[33–35] was used for all of the calculation results appearing in Figure 2b; gas phase versions of the same calculations are included in Figure S5. This study indicated that the reaction is highly favorable ($\Delta H° = -38.0$ kcal/mol and $\Delta G° = -25.3$ kcal/mol), which is a common feature associated with Click-type reactions and serves to validate why the reactions proceed despite the high degree of building steric bulk[4]. For comparison, the addition of a thiol nucleophile to a maleimide group affords more modest values ($\Delta H° = -24.4$ kcal/mol and $\Delta G° = -10.1$ kcal/mol) using the same computational approach (Figure S5). Calculations were also used to compare the relative oxidation propensities for the initial thiol- and aniline-addition products relative to a methyl-substituted *o*-quinone. While such an equilibrium was strongly favorable for the formation of the oxidized aniline product (Figure 2c), it was disfavored for the thiol addition product (Figure 2d). This relative comparison is in line with the observation that the thiol addition product remains predominantly in the reduced catechol form.

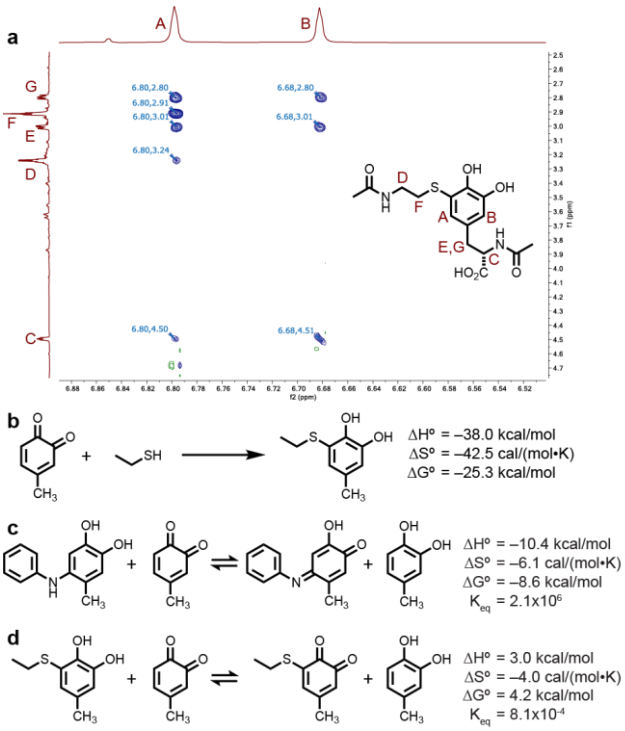


Fig. 2. Structural analysis of thiol-*o*-quinone coupling products. (a) ROESY ^{1}H NMR data (900 MHz) show a clear correlation between proton A and the corresponding signals from both the alkyl chain in *N*-acetylcysteamine (F) and the *N*-acetyltyrosine alkyl chain (E,G). (b) The energetics of the coupling reaction were estimated using DFT calculations. Geometry optimization and vibrational spectral calculations were conducted using B3LYP/6-31G**, and final electronic energies were calculated using ωB97M-V/6-311G^{++}3df-3pd. A CPCM solvation model was used in these calculations (see Supporting Information for details and molecular coordinates). These data were combined to obtain the reported thermodynamic values. For comparison purposes, the reaction energies and equilibrium constants were estimated using this approach for the oxidation of the initial (c) aniline and (d) thiol coupling products with a methyl-substituted *o*-quinone.

3.4 Tyrosinase stability studies and substrate screening

To test protein conjugation product stability, we subjected α-endorphin modified MS2 to buffers ranging from pH 5 to pH 8, potassium ferricyanide, 5 mM tris-carboxyethyl phosphine (TCEP), and 5 mM n-mercaptobutanol for up to 6 days at room temperature. None of the conditions led to greater than 5% loss of the attached peptides (Supplementary Table 1, Figure S6). Additionally, GFP modified with an RRRRY peptide showed no reversion when kept at 37 °C for up to 7 days, even when exposed to up to 4 mM glutathione (GSH). It should be noted that small amounts of GSH and other small molecule thiols were shown to add into the *o*-hydroquinone ring over the course of several days (Figure S7), presumably after aerobic oxidation to form small equilibrium amounts of the *o*-quinone; however, these adducts did not appear to affect the integrity of the linkage formed in the initial coupling reaction.

Stability tests were also conducted in human blood serum after coupling Y200C sfGFP to a biotin moiety through either maleimide- or tyrosinase-based conjugation. The resulting samples were kept in the serum at room temperature or at 37 °C for up to 7 days. At several time points during incubation, aliquots were removed and subjected to streptavidin bead capture, followed by a final elution using 80% acetonitrile (ACN), 5% formic acid (FA) and 2 mM biotin (Figure S8)[36]. We found that the maleimide coupled product decreased over time, and at 7 days was no longer detectable in either sample (Figure S9). This result is consistent with literature observations for these types of adducts on antibodies and other proteins[37]. In contrast, the tyrosinase-coupled product was recovered throughout the course of the experiment, indicating that the tyrosinase coupling product has substantially greater stability in serum.

In addition to α-endorphin (Figure 3a), the tyrosine-cysteine coupling chemistry has proven remarkably successful for the site-selective attachment of additional peptides. Both super-folder GFP (sfGFP) bearing the Y200C mutation (GFP Y200C, Fig. 3a,b) and the N87C MS2 coat protein (Figure S10) were coupled to tyrosines at the ends of peptides derived from interleukin-13 (IL-13), the Simian Vacuolating Virus 40 (SV40) large T-antigen[38,39], HIV Tat[40], and polyarginine[40] with excellent conversion (Fig. 3b). It is important to note that N-terminal acylation or an additional glycine was needed in these cases to prevent dopachrome-like cyclization of the oxidized N-terminal tyrosine residues. Substrate compatibility tests using the tyrosinase activation/thiol addition protocol also showed that small molecule phenols, including biotin, rhodamine, and 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) all resulted in high conversion of the starting protein to the expected products within 30 min (Figure S11). Some secondary adducts were observed with these small molecule substrates for sfGFP, which are believed to be the products of competing N-terminal modification (Figure S11)[14]. The diversity of small molecules and peptides compatible with this coupling strategy demonstrates the utility of the reaction as a general tool for protein modification. Our previous studies of aniline-based oxidation coupling reactions have indicated that *o*-quinone formation is rate-limiting[41]. This is likely the case for the thiol additions as well, complicating the calculation of the second-order rate constant for the key coupling step. Even considering the enzymatic activation requirement, these reactions still reach completion in 1 h with coupling partner concentrations well below 100 μM.

3.5 Delivery of Cas9 and protein fusions

An application of this chemistry for the modification of large proteins and enzymes was demonstrated for CRISPR-Cas9 ribonucleoprotein (RNP) complexes. Cas9, an RNA-guided DNA endonuclease used widely for genome editing, contains two surface cysteines at positions 80 and 574, thus providing two native bioconjugation handles for peptide attachment. To generate a modified form of Cas9 bearing two peptides, coupling reactions were conducted using Cas9 bearing two copies of the SV40 large T-antigen nuclear localization sequence (NLS) to facilitate nuclear entry. Quantitative conversion was observed to the doubly modified protein in both the apo and RNA bound forms (Figure 3c). Importantly, tyrosinase-modified Cas9 RNPs fully retained their target cleavage activity when compared to unmodified Cas9 RNPs (Figure 3d). Building off of the success of peptide conjugates, we applied the tyrosinase conjugation method

to the coupling of whole protein domains. We found that tyrosinase could be used to ligate Cas9 to green fluorescent protein with no detriment to DNA cleavage ability (MYGGS-GFP, Figure 2c,e, Figure S12). It is important to note that the mass spectral data revealed no evidence of Cas9 oxidation, nor did they indicate significant amounts of off-target modification. These experiments underscore the exquisite chemoselectivity of this reaction as the Cas9 constructs contain 158 lysine and 58 tyrosine residues in addition to representatives of the 18 other native amino acids, providing ample opportunities for off-target reactions. Despite having a highly reactive system, this method has exceptionally predictable conjugation outcomes.

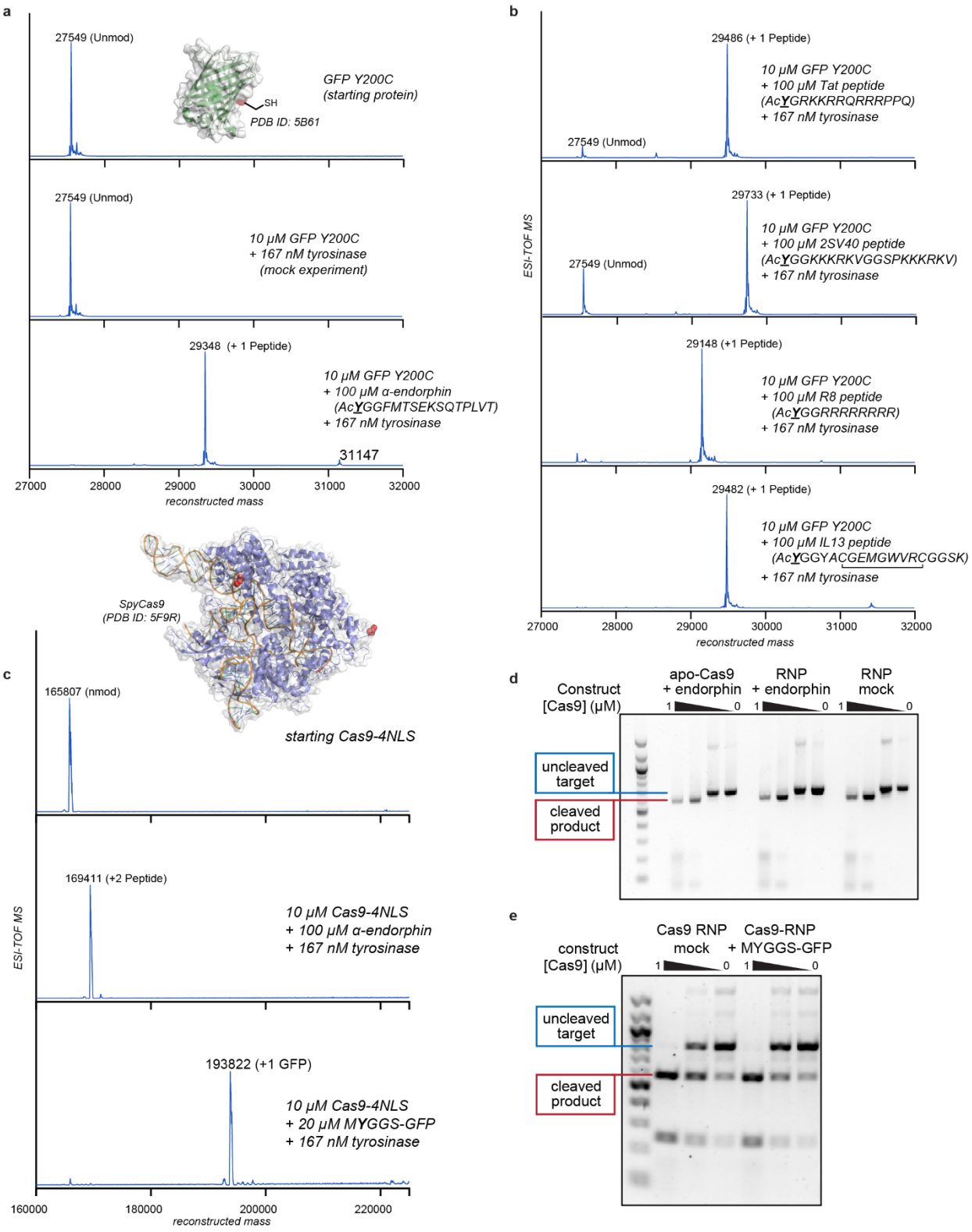


Fig. 3. Synthesis of protein-peptide and protein-protein conjugates using thiol-directed enzymatic oxidative coupling reactions. (a,b) A superfolder GFP thiol mutant (Y200C) was used as a protein component to evaluate peptide coupling reactions. All reactions were run for 30 min at RT and

characterized using ESI-TOF MS. (c,d,e) A CRISPR-Cas9 variant with two C-terminal and four N-terminal nuclear localization sequences (Cas9-4NLS) was modified with peptide and protein coupling partners. Cas9 has two surface thiols (cys 80 and cys 574, red in the attached structure) in the native sequence. (c) ESI-TOF MS data indicated complete coupling at both sites with the α-endorphin peptide. The coupling reaction was successful even when the coupling partner was superfolder GFP bearing an exposed tyrosine residue near the N-terminus (MYGGS-GFP). All Cas9 couplings were run on ice for 1 h. (d) The site-specific DNA cleaving ability of the Cas9-peptide conjugate was unchanged relative to an untreated control. This was true when the pre-assembled ribonucleoprotein (RNP) was modified directly, or when the RNA-free protein (apo-Cas9) was used for the modification reaction. In the latter case, the single guide RNA strand was added before the DNA cleavage experiment. (e) The Cas9-GFP conjugate retained site-specific DNA cleaving ability despite the high degree of added steric bulk.

With these results in hand, we examined whether we could attach peptides capable of mediating Cas9 delivery. A genetic fusion of four consecutive SV40 NLSs to the N-terminus of Cas9 bearing two additional C-terminal NLSs has been shown to enable delivery of Cas9 RNPs to human neural progenitor cells (NPCs) in culture and *in vivo*[38]. However, the production of such complexes is lengthy, requires dedicated cloning and protein purification, and limits the locations of the delivery peptides to the native N- and C-termini. Moreover, most commercially produced Cas9 RNPs demonstrate little-to-no cell penetrating capabilities[38]. Thus, we examined whether standard Cas9 RNP complexes could be chemically modified with peptides or proteins for editing in human NPCs[38]. Three different cell-penetrating peptides were coupled to the native cysteine residues of a Cas9 RNP bearing two C-terminal NLS sequences, which does not appreciably edit cells in the NPC assay on its own (Figure 4a). After the reaction, ESI-TOF MS analysis confirmed that two copies of each tested peptide coupled quantitatively to Cas9 (Figure S13). These constructs, plus Cas9RNP with sfGFP, were then assayed for their editing efficiencies[42–45]. Cas9 conjugates bearing +36 GFP proteins precipitated upon guide RNA addition, and thus could not be evaluated in this experiment. The Cas9RNP construct bearing two 2SV40 peptides (each peptide is composed of two consecutive SV40 NLSs) showed a 20-fold increase in editing efficiency (Figure 4a and Figure S14) relative to the unmodified 2NLS-Cas9RNP. This indicates that the delivery of Cas9 RNP can be mediated by direct peptide coupling to the protein surface of a readily available standard Cas9 protein. The addition of other peptides and sfGFP did not improve editing efficiency. In addition, experiments involving the original 4NLS Cas9 construct showed successful modification, but the added protein domains did not increase editing activity further (Figure S15**)**.

Though Cas9-GFP conjugates showed no change in cell uptake, the capability to couple two protein domains so efficiently presents an attractive prospect that has been pursued for some time by the field. To test the scope of the reaction, we generated versions of nano-luciferase (nanoLuc) and a HER2 binding scFv with tyrosine tags and combined them with Y200C sfGFP. The nanoLuc was produced with either an N-terminal or C-terminal tag, and coupling was performed using 10 µM nanoLuc with 20 µM Y200C GFP in pH 6.5 phosphate buffer for 30 min at RT. Both constructs yielded over 80% conversion, (as determined using ESI-TOF MS) to the singly modified nanoLuc-GFP conjugates (Figure S16). The reaction was similarly successful

for the coupling of Y200C sfGFP (~7 µM by absorbance at 485 nm using an extinction coefficient of 83,300 $M^{-1}cm^{-1}$) to a HER2-binding short chain variable fragment (scFv)-GGSY construct (5 µM) in pH 6.5 phosphate buffer at room temperature for 30 min. These conditions quantitatively converted the scFv into a tagged GFP (Figure 4b,c). Tagged GFP constructs were then incubated with resuspended $HER2^+$ and $HER2^-$ cells (SK-BR-3 and MDA-MB-468, respectively), at 2 µg per 100 µL of cell suspension. Binding was tracked via flow cytometry, where the scFv-GFP construct bound specifically to HER2 expressing cells, but not HER2-negative cells (Figure 4d). No cellular fluorescence was observed with either GFP alone or in controls where GFP was not coupled to the scFv (Figures S17, S18). These and corresponding control experiments showed that only highly exposed tyrosines near the N- or C-termini are capable of participating in the reaction due to the lack of further substrate oxidation[22]. This was quite important, as the scFv is known to have 15 native tyrosine residues, with 8 located in the antigen binding site[46]. Furthermore, these results demonstrate the utility of tyrosinase mediated conjugation for site-specific protein-protein coupling while preserving targeted immunoglobulin binding affinity.

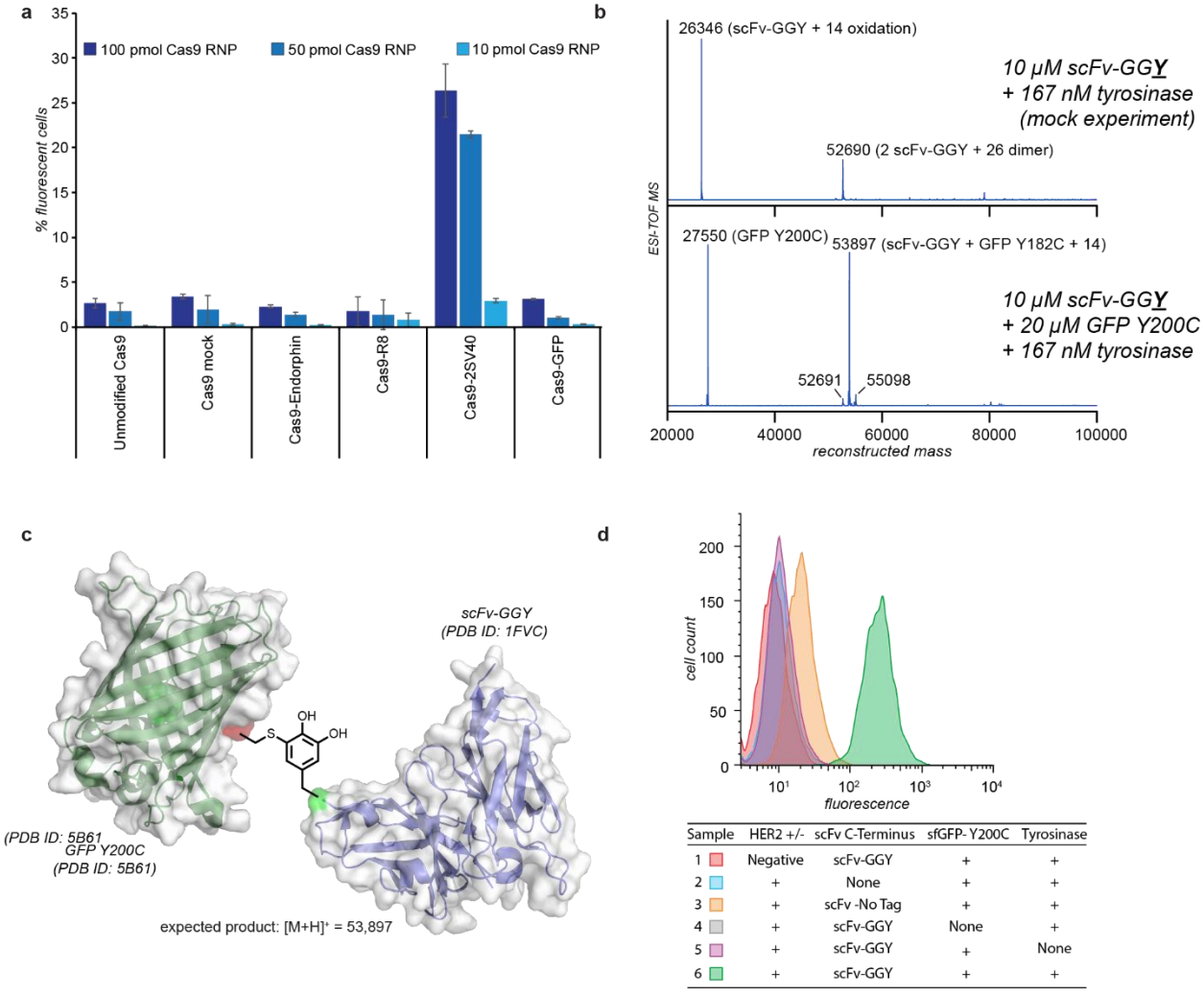


Fig. 4. Evaluation of tyrosine-cysteine bioconjugates in cell culture assays. (a) Evaluation of modified CRISPR-Cas9 ribonucleoproteins (RNP) for gene editing in neural progenitor cells. The assay measured the increase in fluorescent protein expression following successful editing at a tdTomato locus. Cas9 modified with two copies of the SV40 nuclear localization sequence (Cas9-2SV40) showed a 20-fold increase in editing efficiency compared to unmodified Cas9. Data were collected at 72 h post-treatment. Error bars show the standard deviation of triplicates. (b) The coupling reaction was successful for the coupling of GFP Y200C to a HER2-binding scFv-GGY construct, as measured by ESI-TOF. The observed product mass was consistent with the linkage depicted in (c). (d) Flow cytometry analysis of SK-BR-3 ($HER2^+$) and MDA-MB-468 ($HER2^-$) breast cancer cells treated with the Trastuzumab scFv-sfGFP construct showed HER2 specific cell binding. Gating and statistics are shown in Figure S18.

3.6 Conclusion

This study establishes that tyrosinase activation can achieve site-specific protein-protein coupling through the formation of cysteine-tyrosine linkages. This simple approach allows for the targeted modification of proteins with a diverse array of molecules in potentially any desired location using an inexpensive, commercially available enzyme as the catalyst. It does not require additional modification steps or exposure to aggressive reaction conditions, conserving protein folding and function. It also offers the advantage of moving the attachment sites by changing the location of the introduced thiol component, which is an inherent advantage of cysteine-based modification chemistry. In particular, the success of Cas9 delivery via peptide conjugation highlights a potential application for tyrosinase mediated coupling to facilitate the screening of candidate peptides for the cellular entry of large biomolecules without the need to clone, express, and purify individual genetic fusions for each combination. Finally, its ability to couple full-sized proteins together with high yields and minimal stoichiometric excesses suggests broad potential for the generation of protein chimeras for biotherapeutic, vaccine, and bioanalytical applications.

3.7 General Procedures and Materials

General Methods and Instrumentation

All reagents were obtained from commercial sources and used without further purification. Tyrosinase isolated from *Agaricus bisporus* (abTYR, both 25 kU [SKU = T3824-25KU] and 50 kU [SKU = T3824-50KU] were used in these studies) was purchased from Sigma-Aldrich. DOTA-NHS was purchased from Macrocyclics (Plano, Tx). Tyrosine containing peptides were purchased from GeneScript (Piscataway, NJ) and used without further purification. Spin concentrators with 10 and 100 kDa molecular weight cutoffs (MWCO) and sterile spin filters with 0.22 µm pores were purchased from Millipore (Billerica, MA). Milli-Q H_2O was obtained from a Millipore purification system. Small molecule phenol substrates were synthesized as previously reported[14].

UV-VIS measurements. A NanoDrop 1000 (Thermo Fisher Scientific) was used to quantify the samples in this work based on absorbance values at 280 nm (or 488 nm for sfGFP).

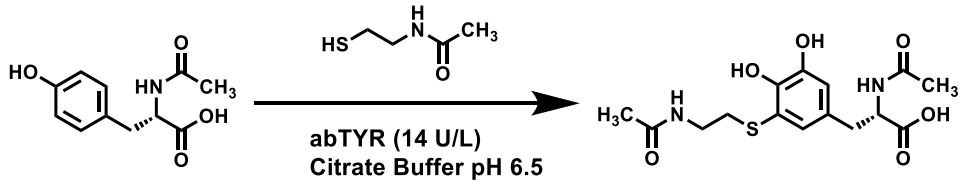
Electrospray Ionization Time of Flight Liquid chromatography mass spectrometry (ESI-TOF LCMS) analysis. Acetonitrile (Optima grade, 99.9%, Fisher,Waltham, MA), formic acid (1 mL ampules, 99+%, Pierce, Rockford, IL), and water purified to a resistivity of 18.2 MΩ·cm (at 25 °C) using a Milli-Q Gradient ultrapure water purification system (Millipore, Billerica, MA) were used to prepare mobile phase solvents for LC-MS. Electrospray ionization mass spectrometry (ESI-MS) of protein bioconjugates was performed using an Agilent 1260 series liquid chromatograph outfitted with an Agilent 6224 time-of-flight (TOF) LC-MS system (Santa Clara, CA). The LC was equipped with a Proswift RP-4H (monolithic phenyl, 1.0 mm × 50 mm, Dionex) analytical column. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). For each sample, approximately 15 to 30 picomoles of analyte were injected onto the column. Following sample injection, a 5-100% B elution gradient was run at a flow rate of 0.30 mL/min over 8 min. Data were collected and analyzed by deconvolution of the entire elution profile in order to provide reconstructed mass spectra that are representative of the entire sample using Agilent MassHunter Qualitative Analysis B.05.00. Percent modification was determined through integration of MS peaks using opensource Chartograph software (www.chartograph.com). The integration of the completely unmodified protein peak served as an internal standard in determining the percent modification.

Preparation and storage of abTYR stock solutions. Tyrosinase isolated from common button mushroom (*Agaricus bisporus*) was purchased as a lyophilized dark-red powder from Sigma-Aldrich (either 25 kU or 50 kU). Upon receipt, the powder was portioned as preweighed dry samples in Eppendorf tubes. These were stored at -20 °C until use. Stock solutions were prepared by dissolving a dry portion of enzyme to yield a concentration of 2 mg/mL in 50 mM phosphate buffer at pH 6.5 (corresponding to ~16.7 µM enzyme). Portions of this solution (10 µL each) were transferred to 250 µL Eppendorf tubes, which were then capped and stored at -80 °C until use. These storage conditions were shown to be the optimal for abTYR in a previous study [14]. Stock solutions were used as single-use aliquots and diluted to desired working concentrations in 50 mM phosphate buffer at pH 6.5 prior to use.

Small Molecule and Protein Procedures

Preparation and structural characterization of tyrosine-thiol conjugates. Separate stock solutions of 100 mM *N*-acetyltyrosine, 100 mM *N*-acetylcysteamine, and 2 µM abTYR were prepared in 100 mM citrate buffer, pH 6.5. The base of a 15 mL 3k MWCO spin concentrator was pre-loaded with 5 mL of 100 mM citrate and 10 mM *N*-acetylcysteamine prepared by dilution of the stock solution. In the top section of the filter, 1 mL of 10 mM *N*-acetyltyrosine solution was combined with 1 mL of 2 µM abTYR in 100 mM citrate buffer at pH 6.5. The solution was incubated at RT for 30 min before centrifugation at 3000 rpm for 25 min. After this step, an additional 1 mL of 10 mM *N*-acetyltyrosine in 100 mM citrate buffer at pH 6.5 was added to the tyrosinase in the top section and allowed to sit for a further 20 min on ice before centrifugation as above. This process was repeated an additional three times, for a total added volume of 4 mL

of 10 mM *N*-acetyltyrosine solution. After the final centrifugation the system was spun an additional 25 min at 4000 rpm, yielding a total of 11 mL of purple-red liquid.



Citrate buff. (pH 6.5)			N-Ac-Tyr			N-Ac-Cysteamine			A. bisporus Tyrosinase			H2O	Total Vol
i (mM)	f (mM)	vol (µL)	i (mM)	f (mM)	vol (µL)	i (mM)	f (mM)	vol (µL)	i (U/L)	f (U/L)	vol (µL)	(µL)	(µL)
100	20	3000.0	100.0	5.00	750.0	100.0	5.0	750.0	982	14.0	213.8	10286	15000

A secondary protocol was also used, which generated lower yields of product: To a 10 mL aliquot of Milli-Q H_2O in a Falcon tube were added stock solutions of sodium citrate buffer (pH 6.5), *N*-acetyltyrosine, and *N*-acetylcysteamine with initial (stock solution) concentrations (i) and final (reaction mixture) concentrations (f) given in the table above. The abTYR stock solution was added last. After gentle mixing, the resulting reaction solution was allowed to stand at room temperature for 2 h and then was stored at 4 °C until purification. After passing the rxn solution through a 0.22 µm Millipore polyethersulfone (PES) syringe filter, the crude reaction mixture was subjected to analytical and preparatory HPLC.

Analysis and purification of the crude small molecule oxidative coupling product. Purification was performed on a Waters UPLC-MS system with the following modules: Waters 2545 Binary Gradient Module pumps, a Waters 2767 Sample Manager for sample injection and fraction collection, a Waters System Fluidics Organizer, a Waters 2489 UV/Visible Detector, and a Waters Acquity Ultra Performance LC mass spectrometer with SQ detector. Waters Mass Lynx software V4.1 (SCN919) was used to operate the instrument. For gradient separations, solvent A was: Milli-Q H_2O with 0.1% TFA, and solvent B was: HPLC-grade acetonitrile with 0.1% TFA. This instrument is owned and maintained by the LBNL/UCB Catalysis Group Instrumentation Facility and is supported by DOE Basic Energy Sciences program KC0302010.

For analytical HPLC, a Waters X-Bridge C18, 3.5 µm, 4.6 x 100 mm column was used. 20 µL injections of the filtered reaction mixture were subject to a linear gradient of 5 to 20% solvent B in solvent A over 15 min at a flow rate of 1.00 mL/min.

For preparatory HPLC, a Waters X-Bridge C18, 10 µm OBD, 4.6 x 250 mm column was used. Injections of the filtered reaction mixture (1 mL) were subjected to a linear gradient of 5 to 20% solvent B in solvent A over 15 min at a flow rate of 30.0 mL/min.

The product was then isolated through evaporation of the solvent. This yielded several flakes of red-brown powder. The dried product was resuspended in D_2O and submitted to NMR analysis on a Bruker 900 MHz system.

***Computational Studies*.** Molecular mechanics methods (Macromodel, OPLS3e force field) were used to generate and minimize large populations of conformers to identify the lowest energy candidates. For each compound under study, the geometries of all conformers within 3 kcal/mol of the global minimum were optimized using at the B3LYP-D3/6-31G** level[32]. At this stage, the resulting global minima were subjected to a second geometry optimization and vibrational spectrum calculation (B3LYP-D3/6-31G**) to determine the zero-point energies, enthalpy corrections, and the internal entropy values at 298.15 K. Finally, refined electronic energy calculations were performed on each optimized geometry using an expanded basis set (ωB97M-V/6-311G-3df-3pd++). For calculations including solvation, a CPCM model was used with Bondi van der Waals radii and a SAS solvent probe radius of 1.4 Å.[33,34] In these cases, solvation was introduced in the geometry optimization, frequency, and final electronic energy calculations.

Expression of sfGFP. Expression was carried out as previously reported [47]. Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of SpCas9. Expression was carried out as previously reported [48]. Purified protein was spin concentrated into 20 mM phosphate buffer with 200 mM trehalose and 300 mM NaCl at pH 7 using 100k MWCO spin filters prior to reaction. Cas9 sequences for 2NLS and 4NLS were identical to those used previously [38].

Expression of N87C MS2. Expression and purification was carried out as previously reported [49]. Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of Trastuzumab scFv. Expression was carried out according to established protocols [22]. Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of Y200C sfGFP and N-Terminal MYGGSH$_6$-sfGFP. Following sequence verification by sequencing, mutants were transformed into Rosetta DE3 *E. coli* cells for expression. Expression cultures at 1000 mL in TB broth were prepared using 10 mL of confluent culture. These were grown at 37 °C until cultures reached an OD600 value between 0.6-0.8, upon which protein expression was induced with a final concentration of 0.1 g/L IPTG.

Cells were then shaken overnight at 37 °C and pelleted. After freezing at -80 °C, cell pellets were resuspended in 15 mL of an equilibration buffer (20 mM sodium phosphate, 2 M NaCl, 20 mM imidazole at pH = 7.4) and then lysed via sonication for 30 min at 60% amplitude. The cell lysate was centrifuged at 14,000 rpm for 30 min, then the supernatant decanted before loading on to a HisTrap FF Crude column. After binding, the bound protein was washed with 4 portions

of two resin bed volumes of wash buffer (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole at pH = 7.4) and subsequently eluted with four resin bed volumes of elution buffer (20 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole at pH = 7.4). The purified sfGFP variant was then spun concentrated into 20 mM phosphate buffer at pH 7.2 using 10k MWCO filters. Purified sfGFP samples were flash frozen and stored at -20 °C until use.

Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of nano luciferase (nanoLuc). The gene encoding nanoLuc was cloned from plasmids provided by Promega (pNL1.1) using primers that encoded the specified N- or C-terminal modifications as well as a golden gate cloning site. The modified product was then transferred into a modified pet28b cloning vector (pECH081) and transformed into Rosetta DE3 pLysS *E. coli*. Overnight cultures were grown at 37 °C in LB with 50 mM kanamycin, then transferred into 1 L TB media with 50 mM kanamycin. Once cultures reached an OD600 of 0.6 protein production was induced by addition of 100 mg/mL IPTG and left shaking overnight at 37 °C. After 16 h of incubation, the cultures were centrifuged at 7000 x g, and the pellets collected were then re-suspended in 50 mM phosphate, 200 mM NaCl, 20 mM imidazole, pH 7 buffer. Samples were lysed by sonication, then centrifuged at 17000 x g. The supernatant was collected and run through a HiTrap NiNTA column from GE in a miniAkta, and eluted with a 50 mM phosphate, 200 mM NaCl, 250 mM imidazole buffer at pH 7. Purified nanoLuc samples were flash frozen and stored at -20 °C until use.

Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of +36 MYGGS sfGFP. The +36 GFP gene[50] with addition of MYGGS-H_6 at the N-terminus was purchased as a gene block from Integrated DNA Technologies (IDT) and used as follows:

A pET28b vector that was cleaved by XbaI and XhoI and column purified was then combined with the +36GFP gene block and cloned using Gibson 2x Master Mix from New England BioLabs (NEB) using standard procedure [51].

Following sequence verification, plasmids were transformed into Rosetta DE3 *E. coli* cells via heat shock and streaked onto LB-agar plates with 1 mM kanamycin. Expression and purification were carried out according to previous literature [51].

abTYR-mediated oxidative coupling of phenols to Y200C sfGFP. To a 10 µM protein solution in 20 mM phosphate buffer (pH = 6.5) was added the phenol coupling partner to a final concentration of 100 µM. A 2 µM solution of abTYR in 50 mM phosphate buffer (pH=6.5) was added to reach a final enzyme concentration of 200 nM. The reaction was left unstirred for 30 min at RT. The reactions were subsequently quenched through the addition of tropolone (final concentrations of 1.9 mM), and allowed to stand at RT for 1 min before spin concentration

against 10 kDa MWCO filters with 10 mM bisTris at pH 7.2 (three cycles at 13,000 rpm for 3 min) for small molecule removal. The resulting samples were analyzed using ESI-TOF MS.

Cysteine modification of N87C MS2. To a 10 µM solution of protein was added the phenol- or tyrosine-containing coupling partner (100 µM) in 20 mM phosphate buffer (pH = 6.5). To this was then added abTYR to a final concentration of 200 nM. After 30 min the reaction was quenched with tropolone to a final concentration of 2 mM and the product was analyzed by ESI-MS (see **Figure 1**).

Stability study of cysteine modification products. To assess the stability of the oxidative coupling products, a solution of endorphin modified MS2 N87C (**Pdt**) was incubated in a variety of conditions. To assess the impact of pH on the linkage stability, a 10 µM solution of **Pdt** with 1 mM tropolone (to prevent further oxidation via tyrosinase) was incubated in 50 mM phosphate buffer at pH 5.0, 7.0 and 8.0 at RT for 0, 3, or 6 days. At each time point a 10 µL sample of the protein was diluted 1:1 with 50 mM pH 7.0 bisTris buffer and spin concentrated against a 10 kDa MWCO filter. The resulting samples were analyzed using ESI-TOF MS and the percent labeling was determined by peak integration using open source Chartograph software (www.chartograph.com). To assess the impact of nucleophiles on the linkage stability, a 10 µM solution of **Pdt** with 1 mM tropolone was incubated in 50 mM phosphate buffer at pH 7.4 at RT for 0, 3, and 6 days in the presence of 10 mM concentrations of various nucleophiles. At each time point, the proteins were spin concentrated against 10 kDa MWCO filters with 10 mM bisTris at pH 7.2 for three cycles at 13,000 rpm for 3 min each. The resulting samples were analyzed using ESI-TOF MS and the percent labeling was determined via peak integration using the opensource Chartograph software. Note that starting modifications differ as each condition was treated as a biological replicate.

Cas9-tyrosinase coupling conditions. Cas9 coupling occurred under the following conditions: 20 mM Tris HCl, 300 mM KCl, 200 mM trehalose pH 7.0 (Buffer A), 4 °C, 1 h, 10 µM Cas9. All samples were quenched with the tropolone solution as above, then solvent exchanged into Buffer A three times using a 100,000 kDa MWCO spin concentrator. For peptide coupling, peptides were added in five-fold excess, yielding 50 µM peptide concentrations. In protein-protein coupling we found that a 1:1 ratio of Cas9:target protein yielded near-quantitative conversion to the singly modified Cas9 after filtration, as characterized using ESI-TOF MS. A 1:5 ratio of Cas9:target yielded full conversion to the doubly modified product. It is of particular interest that the trehalose buffer is tolerated as previous oxidative strategies were incompatible with saccharides in solution [49]. Additionally, we have found that the internal tyrosine residues within both of these proteins are not affected by this reaction, as revealed by a lack of detectable oxidation for either Cas9 or a sfGFP that lacks the N-terminal tyrosine residue by ESI-TOF MS. Further supporting this is the lack of dimer formation between Cas9 molecules or peptide substrates in these reactions. Thus, it appears that only highly exposed tyrosines, ideally near the N- or C-termini are capable of participating in reactions [22].

Cas9 in vitro cleavage assay. Two gene fragments were produced via PCR for use as target sequences (Cas9 recognition sites shown underlined in red):

Cas9 cleavage sequence 1: 5'-
TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCT
GACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC
ATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCAC
AACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCAC
AACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC
GGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCA
ACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCA
TGGACGA-3'

Cas9 Cleavage sequence 2: 5'-
TGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAA
AACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGG
CCACAACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC
TACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCC
TCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCA
CGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAA
CTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTC
AAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACC
CCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGA
GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
TCACTCTCGGCATGGACGA-3'

This was done using a plasmid containing EGFP and the following PCR primers:

PCR forward primer 1: *5'-TAAACGGCCACAAGTTCAGC-3'*

PCR forward primer 2: *5'-TGTATGGGATCTGATCTGGGG-3'*

PCR reverse primer: *5' -TCGTCCATGCCGAGAGTGATCC- 3'*

Cas9 was combined with sgRNA (targeting sequence *5'-CAGGGTCAGCTTGCCGTAGG-3')* produced by *in vitro* transcription as previously published to produce the active RNP[48]. A 0.4 pmol sample of cleavage sequence 1 or 2 was then exposed to varying concentrations of Cas9 RNP in 20 mM mM Tris-HCl, pH 7.5, 75 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 5% glycerol buffered solution and incubated at 37 °C for 30 min, followed by denaturation at 95 °C for 2 min. Samples were then loaded onto a 1% agarose TAE gel with SYBR-safe and visualized on a BioRad EZDoc Gel imager using the ImageLab 6.0 software package.

Cell Culture Procedures

Cell labeling with Trastuzumab scFv-sfGFP conjugate. SK-BR-3 and MDA-MB-468 cells were obtained in separate T-25 flasks from the Berkeley Cell Culture Facility, transferred to a 37 °C, 5% CO_2 incubator, and grown to 90% confluency. Growth medium was removed, and cells were detached by incubation with 6 mL of 0.13% trypsin in DPBS at 37 °C for 10 min. After visually confirming detachment, each cell sample was diluted to 15 mL with Binding Buffer (DPBS with 10% FBS and 1% NaN_3). Cells were added to 15 mL Falcon tubes and subject to two cycles of centrifugation (300 g, 5 min, room temperature), removing the supernatant after each cycle and diluting with 14 mL of Binding Buffer before the second cycle. Cells were re-suspended in 0.8 mL of Cell Binding Buffer and counted with a BioRad TC20 automated cell counter. Live cell counts were 99 to 100% as indicated by trypan blue staining. Cell suspension density was adjusted to 1.9 $x10^6$ cells/mL with additional Binding Buffer. A series of 100 µL aliquots was distributed to Eppendorf tubes and incubated on ice. A 40 µL sample of the 0.05 µg/µL scFv-sfGFP construct or appropriate negative control in PBS was added to each aliquot, and these samples were allowed to sit on ice, in the dark for 1 h. Samples were then subjected to 3 cycles of centrifugation (300 g, 3 min), re-suspending cells in 1000 µL of Binding Buffer before each cycle and discarding supernatant after each cycle. After centrifugation, cell pellets were re-suspended in 500 µL of DPBS with 0.5% paraformaldehyde and kept on ice in the dark for 1-3 h until flow cytometry analysis.

Flow cytometry analysis for cell labeling. Cell samples were diluted 2-3x in PBS and analyzed on to a BD LSR II flow cytometer (BD Biosciences). Fluorescence excitation at the 488 nm laser line and FITC filter set were used. Data was acquired using BD FACSDiva software (version 6.2, 2000). A total of 10,000 events were recorded for each sample. Data were processed using FlowJo software (version 10.0).

Preparation of Cas9 for cell culture trials. Cas9 RNP was prepared based on previously described methods [38]. To prepare the Cas9 RNP complexes, modified Cas9 protein was incubated with sgRNA at a 1:1.2 molar ration. Briefly, we added sgRNA to Buffer #1 (25mM NaP_i, 150 mM NaCl, 200 mM trehalose, 1 mM $MgCl_2$), then added the Cas9 to the sgRNA, slowly with swirling, and incubated at 37 °C for 10 min to form RNP complexes. RNP complexes were filtered prior to use through a 0.22 µm Costar 8160 Filter pre-wet with 200 µL of Buffer #1. If needed, the RNP sample was concentrated with a 0.5 mL Millipore Ultra 100 kDa cutoff filter, part # UFC510096, until the desired volume was obtained.

Neural progenitor cell (NPCs) line assay. NPCs were treated based on a previously described protocol[38]. Briefly, NPCs were isolated from cortices from embryonic day 13.5 Ai9-tdTomato homozygous mouse embryos. Cells were cultured as neurospheres in NPC medium: DMEM/F12 with glutamine, Na-Pyruvate, 10 mM HEPES, non-essential amino acid, penicillin and streptomycin, 2-mercaptoethanol, B-27 without vitamin A, N2 supplement, bFGF and EGF,

both 20 ng/ml as final concentration. NPCs were passaged using MACS Neural Dissociation Kit (Papain) cat# 130-092-628 following manufacturer's protocol. bFGF and EGF were refreshed every other day and passaged every 6 days. The NPC line was authenticated by immunocytochemistry marker protein staining. They were tested for mycoplasma using Hoechst stain with visual analysis and were negative.

Direct delivery of Cas9 RNP. To test direct delivery of Cas9 ribonucleoprotein (RNP) complexes, RNPs targeting the STOP cassette at the tdTomato locus of cells derived from Ai9-tdTomato transgenic mice (sgRNA-298, 5'-AAGTAAAACCTCTACAAATG-3')[38] were pre-assembled. Assembled RNPs were added to the media of Ai9 neural progenitor cells (NPCs) and incubated for 24 h. Cells were then washed twice with 200 U/ml heparin in DMEM-based media and allowed to grow for an additional 24-48 h before analysis. The total amount of RNP per well of a 96-well plate was titrated from 10 pmol to 100 pmol. Percentages of tdTomato-positive (successfully edited) NPCs were quantified by flow cytometry (Attune Nxt acoustic focusing cytometer, Thermo Fisher Scientific) with tdTomato fluorescence measured at the 561nm laser line using the integrated analysis tool (Attune NxT software, v3.1.0). We routinely acquired at least 10,000-30,000 events per sample.

Sequences of Proteins and Peptides

MYGGS-H_6 +36 GFP:

MYGGSHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTR
AEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRKNGIKAKFKIRHNVK
DGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMVLLEFVTAAGIKHGR
DERYK*

MYGGSH_6-sfGFP amino acid sequence:

MYGGSHHHHHHRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLV
NRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK*

Y200C sfGFP amino acid sequence:

MRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTY
GVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHCQQNTPIGDGPVLLPD
NHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKHHHHHH*

N87C MS2 amino acid sequence:

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVE
VPKVATQTVGGVELPVAAWRSYLCMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAY*

scFv amino acid sequence:

EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTISA
DTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSGGGGSGGGGSGGGGSDIQM
TQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSRSGTDFTLTISSL
QPEDFATYYCQQHYTTPPTFGQGTKLEIKRTGGY*

MYGGS-nanoLuc amino acid sequence:

MYGGSHHHHHHVFTLEDFVGDWRQTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVII
PYEGLSGDQMGQIEKIFKVVYPVDDHHFKVILHYGTLVIDGVTPNMIDYFGRPYEGIAVFDGKKITVTGTLW
NGNKIIDERLINPDGSLLFRVTINGVTGWRLCERILA*

nanoLuc $(AP)_4$GGY amino acid sequence:

MHHHHHHVFTLEDFVGDWRQTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVIIPYEG
LSGDQMGQIEKIFKVVYPVDDHHFKVILHYGTLVIDGVTPNMIDYFGRPYEGIAVFDGKKITVTGTLWNGNKI
IDERLINPDGSLLFRVTINGVTGWRLCERILAAPAPAPAPGGY*

Cas9 constructs and sequences are identical to those previously published[38]. In short, both Cas9 constructs had a solubilizing MBP leader sequence that was subsequently removed by TEV protease cleavage. MBP leader sequence:

MKSSHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGP
DIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEE
IPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKH
MNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNK
ELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRT
AVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEENLYFQ

Cas9- 2NLS post TEV cleavage:

SNATMDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR
RKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA
DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA

KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGSPKKKRKVEDPKKKRKVD*

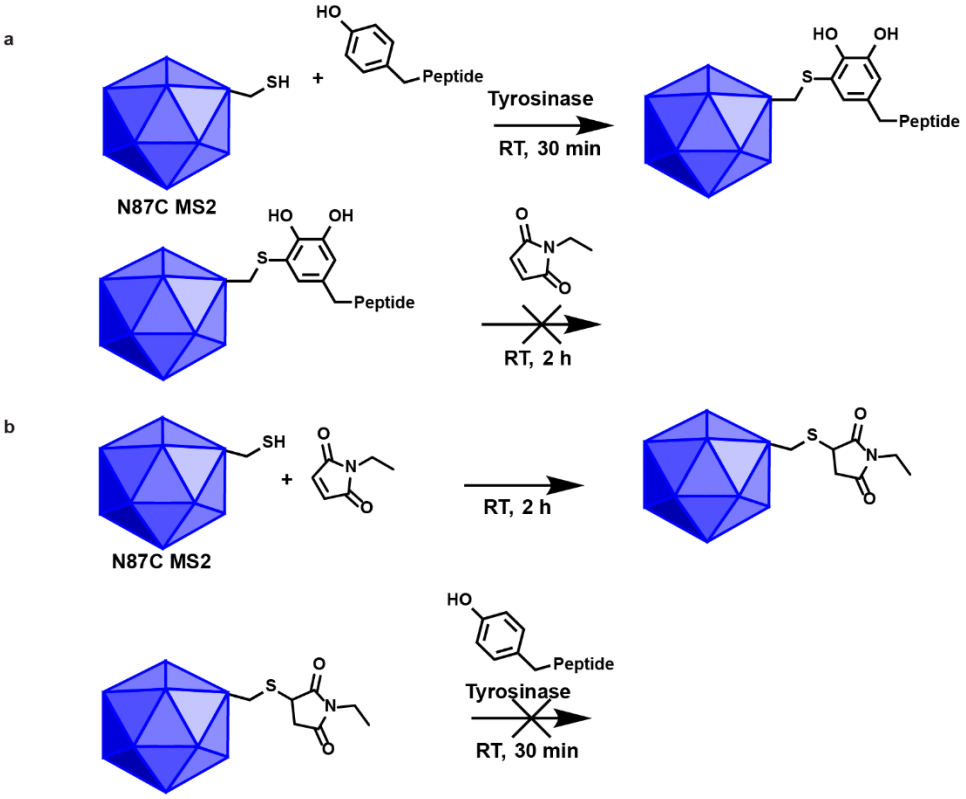
Cas9 4NLS post TEV cleavage:

SNATPKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVGIHGVPAATMDKKYSIGLDIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR
PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDGSPKKKRKVEDPKKKRKV*

Peptide Sequences:

α-endorphin:	Ac-YGGFMTSEKSQTPLVT
R8:	Ac-YGGRRRRRRRR
IL-13:	Ac-YGGYACGEMGWVRCGGSK (C-C disulfide bonded)
HIV-Tat:	Ac-YGRKKRRQRRRPPQ
2SV40:	Ac-YGGKKKRKVGGSPKKKRKV

3.8 Supplementary Figures



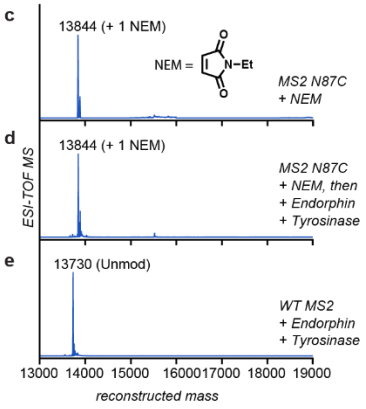


Figure S1. Flowchart representing maleimide capping experiments. (a) MS2 modified with endorphin using abTYR should not react with an additional maleimide coupling partner when exposed to coupling conditions in subsequent steps. Similarly, (b) maleimide-modified MS2 should be untouched by the tyrosinase reaction. Mass spectral data support these hypotheses, as (c) shows NEM capped MS2, and (d) shows NEM capped MS2 exposed to tyrosinase coupling conditions is not further modified by endorphin. Finally, (e) shows WT MS2 is unmodified by tyrosinase conditions.

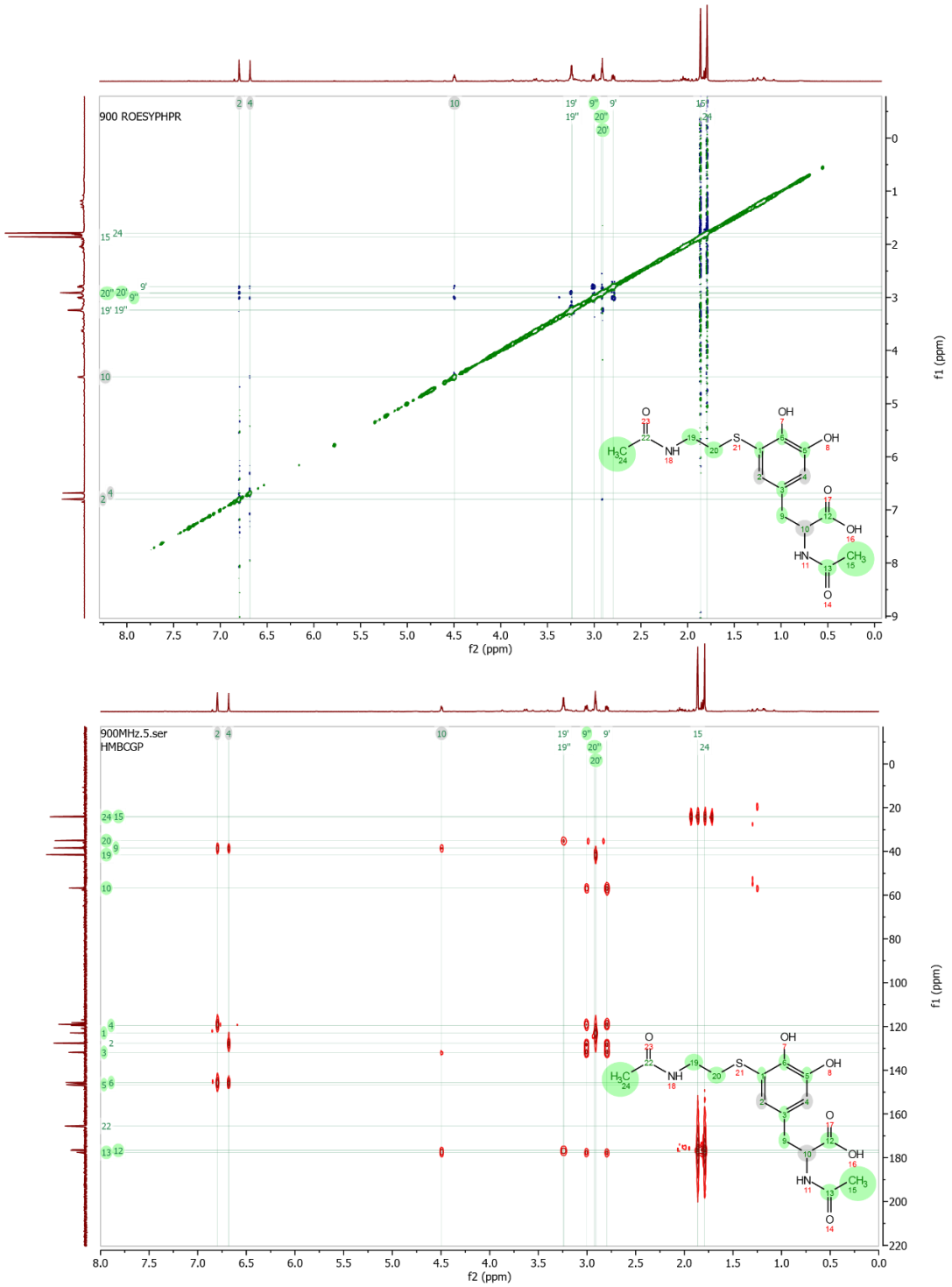


Figure S2. ROESY and HMBC data (900 MHz) for *N*-acetyltyrosine coupled to *N*-acetylcysteamine. ROESY data (top) show a clear correlation between proton (2) and the corresponding signal from both the alkyl chain in cysteamine (20) and the tyrosine alkyl chain (9). HMBC corroborates the proton assignments.

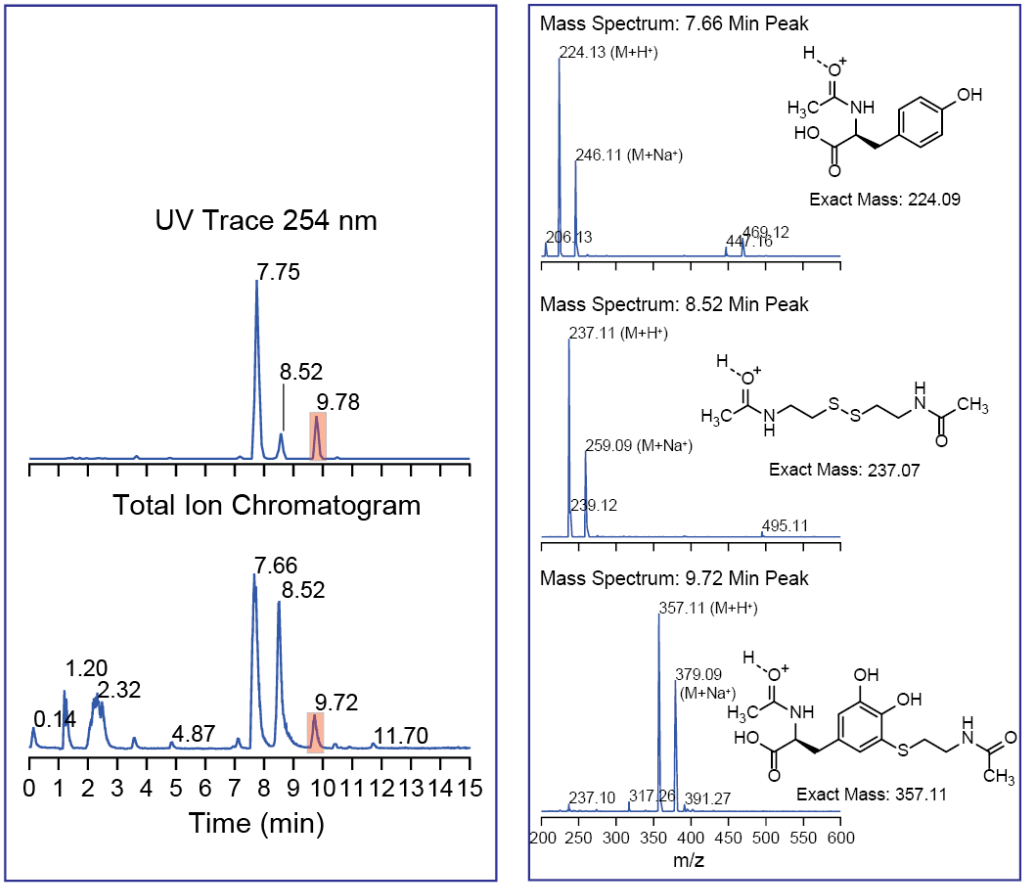


Figure S3. Analytical HPLC chromatogram of the small molecule thiol addition reaction sample. Red boxes indicate single adduct peaks. The primary peak indicates unreacted tyrosine. The cystine disulfide dimer elutes at 8.52 min. The peak at 9.7 was collected and used for NMR analysis as seen in Figure S2.

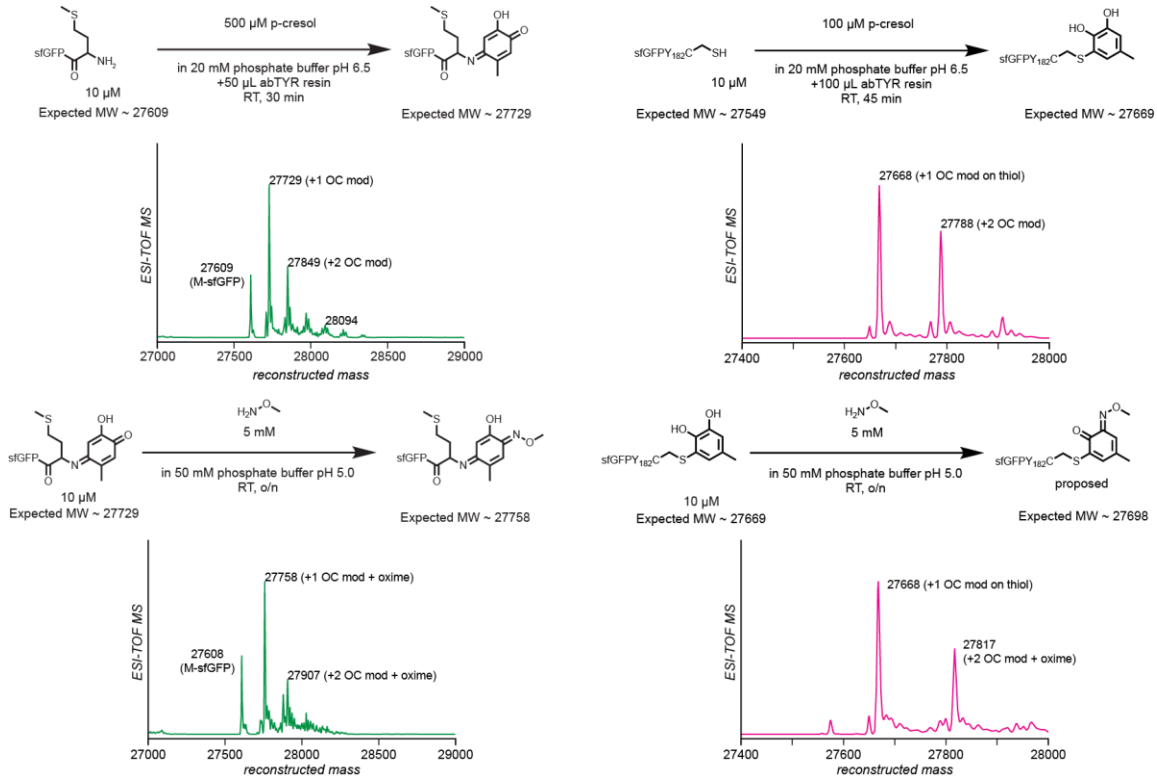


Figure S4. Exploring oxime reaction susceptibility of *o*-quinone-modified endogenous amino acids. *Ortho*-quinones can readily be installed on primary amines at N-termini (green trace, top left) or on free cysteine residues on proteins (pink trace, top right). Structures shown are based on those previously reported for *o*-quinone-modified amine N-termini[14] and now thiols. In addition, small molecule *o*-quinones have been observed to overly modify proteins if the concentration of *o*-quinone is too high (secondary peaks in top green and pink traces, oxidative coupling products marked as "OC"). While the exact nature of these over modification products is not known, they are presumably highly reactive nitrogenous residues on sfGFP. To 10 µM of the various *o*-quinone-modified products was added 5 mM of CH_3ONH_2 and the mix was allowed to react overnight in 50 mM phosphate buffer at pH 5.0. Oxime ligation was only observed on *o*-quinones that had modified the primary amine N-terminus of met-sfGFP (green trace, bottom left), and the adduct between thiols and *o*-quinones did not undergo oxime formation (pink trace; bottom right). The higher order modifications between *o*-quinones and sfGFP also underwent oxime formation, further corroborating that these modifications must be on highly reactive nitrogenous bases on sfGFP.

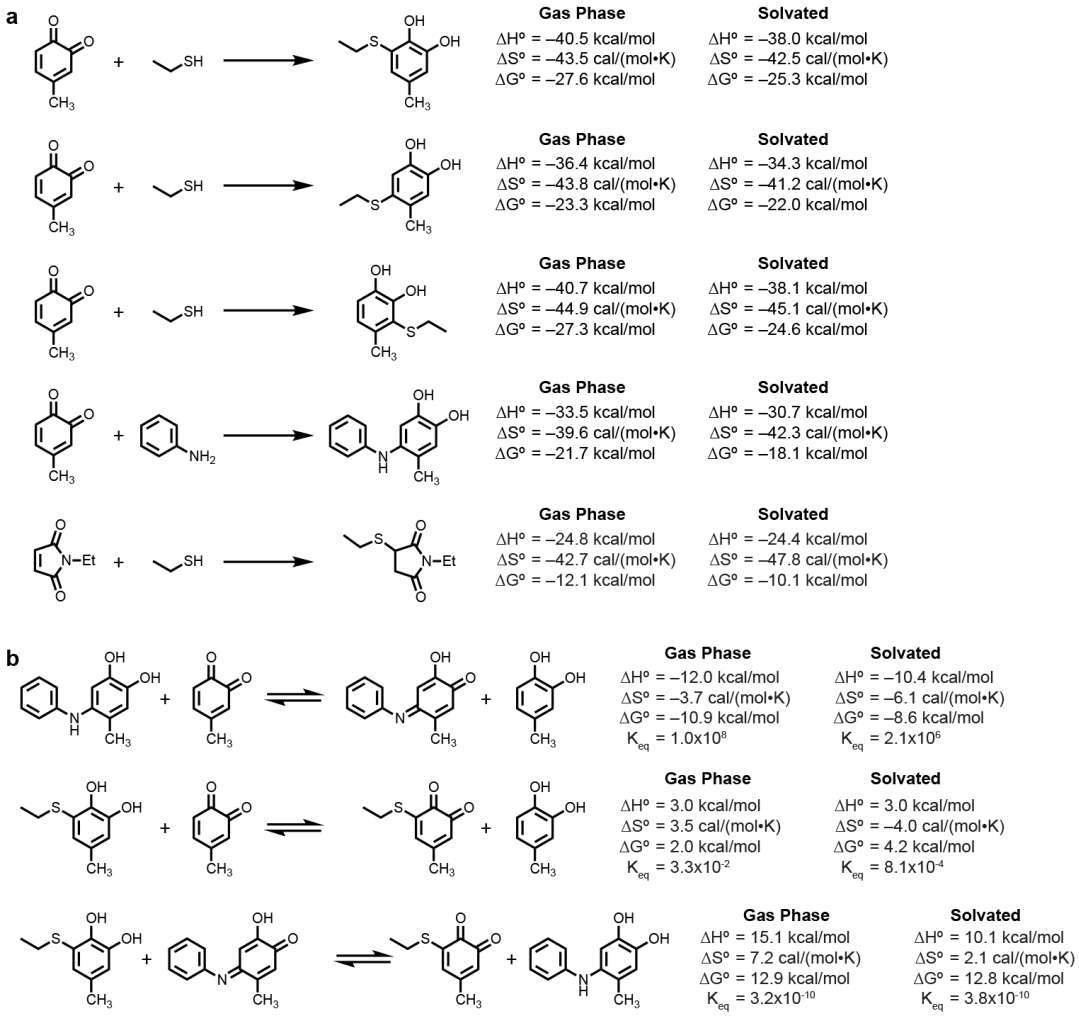


Figure S5. Computational analysis for thiol addition products. (a) The energetics of the lowest energy product conformers are compared for three isomeric thiol addition reactions, along with the addition of an aniline. A thiol addition to a maleimide is also included for comparison. The geometry for each species was optimized using B3LYP/6-31G**. Vibrational spectra were calculated using B3LYP/6-31G** to obtain entropies and zero-point energies. For the "Solvated" data, a CPCM model was used through self-consistent reaction field optimizations.[ref] Final electronic energies were calculated using ωB97M-V/6-311G^{++}3df-3pd, and these data used to obtain the final thermodynamic values. All calculations were performed using QChem software. (b) Redox reaction energies and equilibrium constants were estimated using the data calculated above.

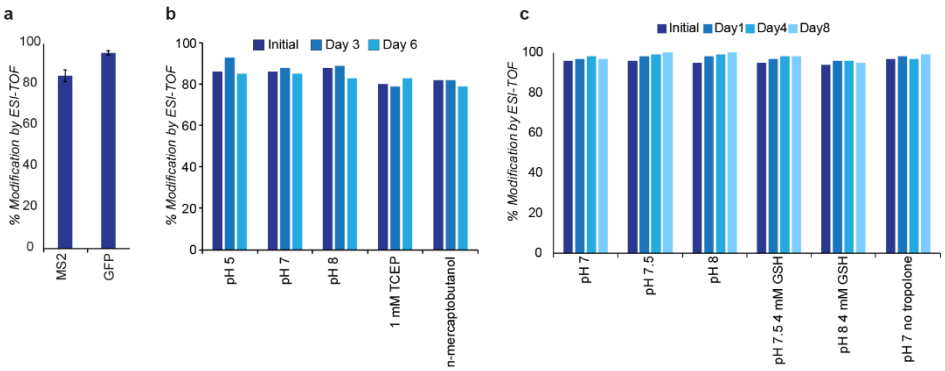


Figure S6. Stability studies for N87C MS2 and Y200C sfGFP coupled to peptides. In all cases percent modification was defined as the integrated area of peaks bearing modification to the unmodified peaks. Unless otherwise noted, bars represent the results of single measurements. (a) Standard deviation for MS2 and GFP modifications. Error bars represent standard deviation between 6 replicate reactions. (b) The stabilities of the N87C MS2 -endorphin conjugates under various pH conditions at RT, with small molecule nucleophiles, and reducing agents show no more than 5% reversion for any given sample. (b) The stability of Y200C GFP-RRRRY conjugates to a series of pH and glutathione conditions at 37 °C again showed no significant reversion over an 8 day period within the limits of quantitation for these experiments.

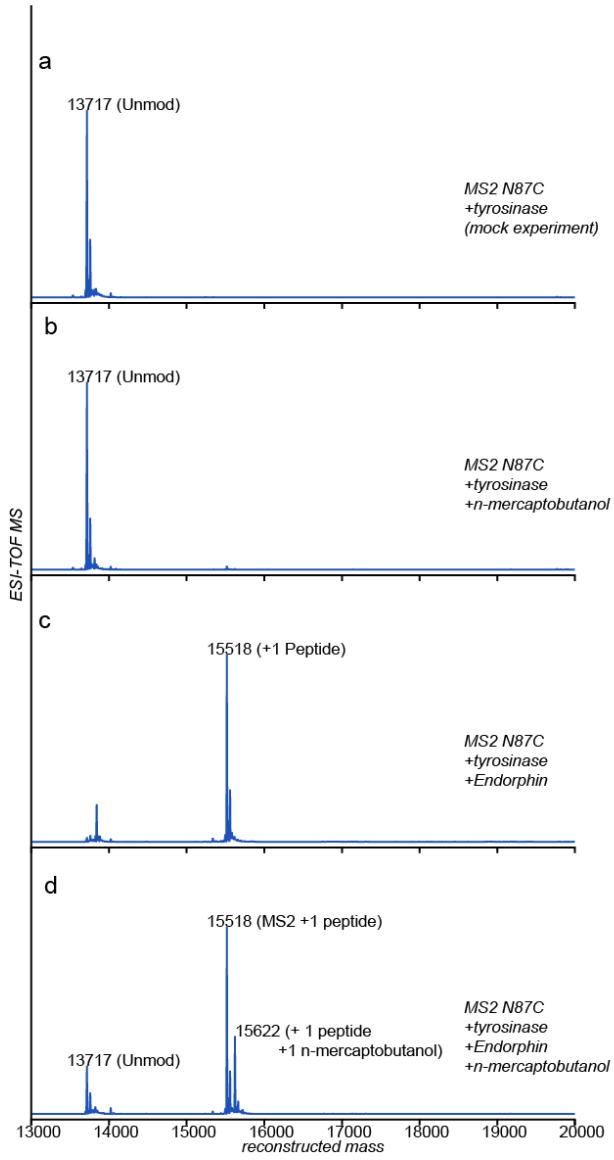


Figure S7. Addition of small molecule thiols to tyrosinase-modified substrates. (a) Sample of unmodified MS2 N87C and (b) MS2 N87C exposed to tyrosinase and n-mercaptobutanol yielded no addition products. (c) MS2 N87C combined with endorphin and tyrosinase yields the expected +1 peptide modification. (d) MS2 N87C with endorphin exposed to n-mercaptobutanol yielded some +1 peptide and some +1 n-mercaptobutanol product.

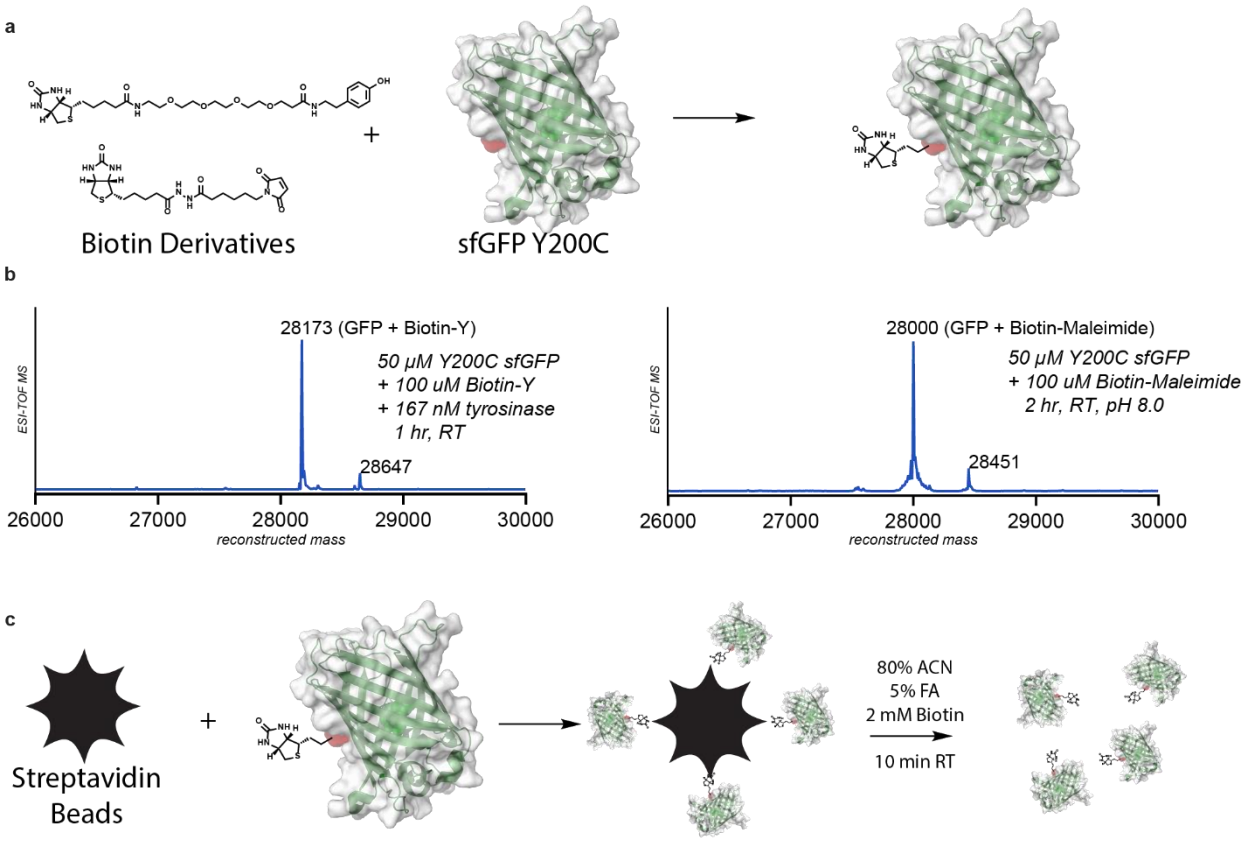


Figure S8. Serum stability experimental outline. (a) A reaction scheme is shown for the biotin labelling of Y200C sfGFP with tyrosine- or maleimide-tagged biotin. (b) ESI-TOF data are shown for tyrosine and maleimide adducts to Y200C sfGFP. The expected mass for biotin-tagged GFP = 28172 Da, and the expected mass for maleimide-tagged GFP = 28000 Da. (c) To extract the biotin-tagged GFP samples from serum, aliquots were incubated with streptavidin beads for 1 h, washed 3x, then eluted using 80% CAN, 5% FA, and 2 mM biotin.

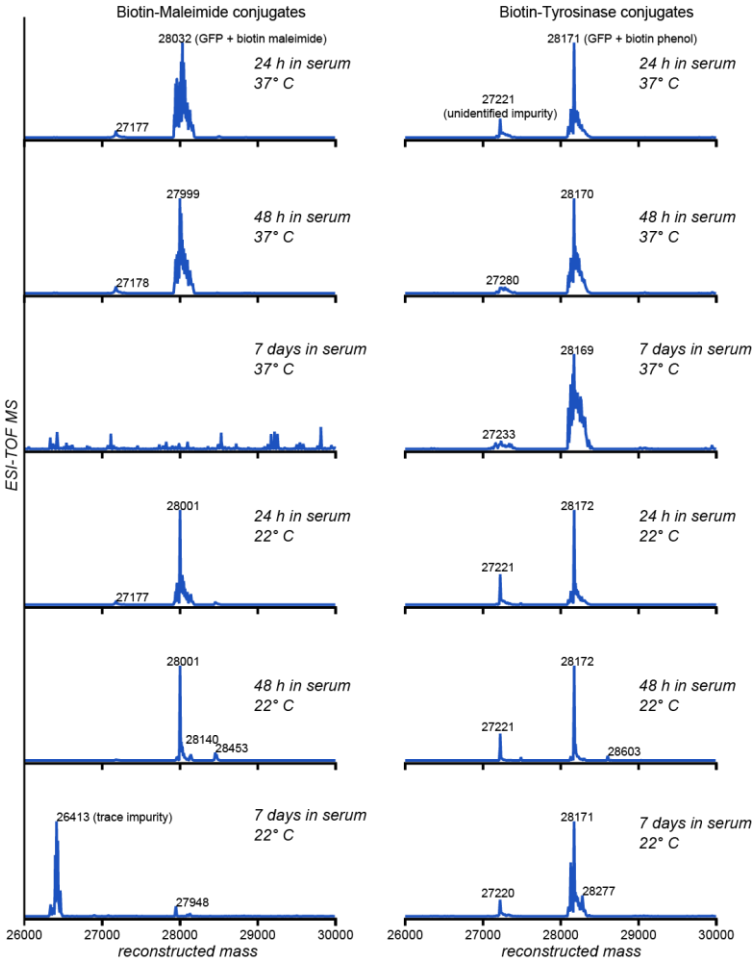


Figure S9. ESI-TOF data for the serum stability study of biotin tagged GFP. The left column shows the results for the maleimide construct, while the right column shows results for the tyrosinase conjugates. The top three panels for each show the results after 1,2, or 7 days at 37 °C, while the bottom three rows show the same samples at room temperature. Most significant to the present study is the fact that the maleimide conjugate was not recovered by day 7 in either condition.

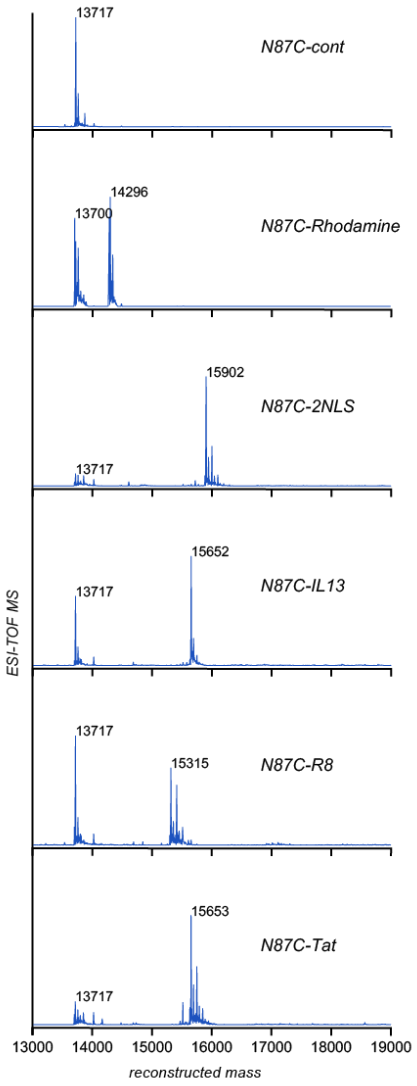


Figure S10. Addition of small molecule phenols and tyrosine containing peptides to MS2 N87C. A 10 µM sample of protein was combined with a phenol coupling partner (at 100 µM) and 167 nM abTYR in pH 6.5 phosphate buffer for 30 min at RT, then quenched with 2 mM final concentration of tropolone and analyzed via ESI-TOF.

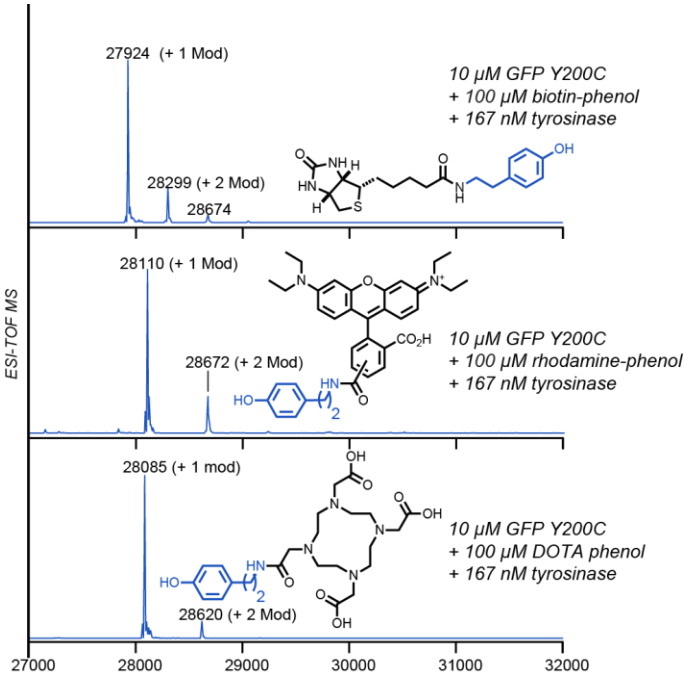


Figure S11. Small molecule additions to Y200C sfGFP. biotin, rhodamine, and DOTA phenol were coupled quantitatively to GFP with some excess secondary modification. The secondary additions are assumed to be at the N-termini, but due to the small yields of these adducts, it was not possible to isolate them and determine their location.

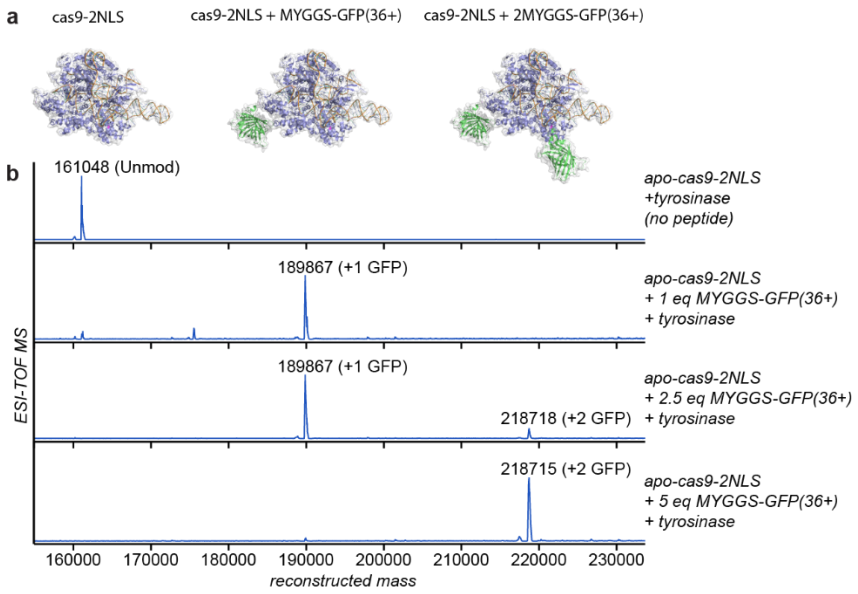


Figure S12. Coupling of MYGGS +36 GFP to Cas9. (a) Structural representations of the coupled proteins are shown using PDB IDs 5F9R and 6B9C for Cas9 and GFP, respectively. (b) ESI-MS data indicate the quantitative conversion to singly or doubly modified Cas9 when 1 to 5 equivalents of GFP are added to tyrosinase reaction conditions.

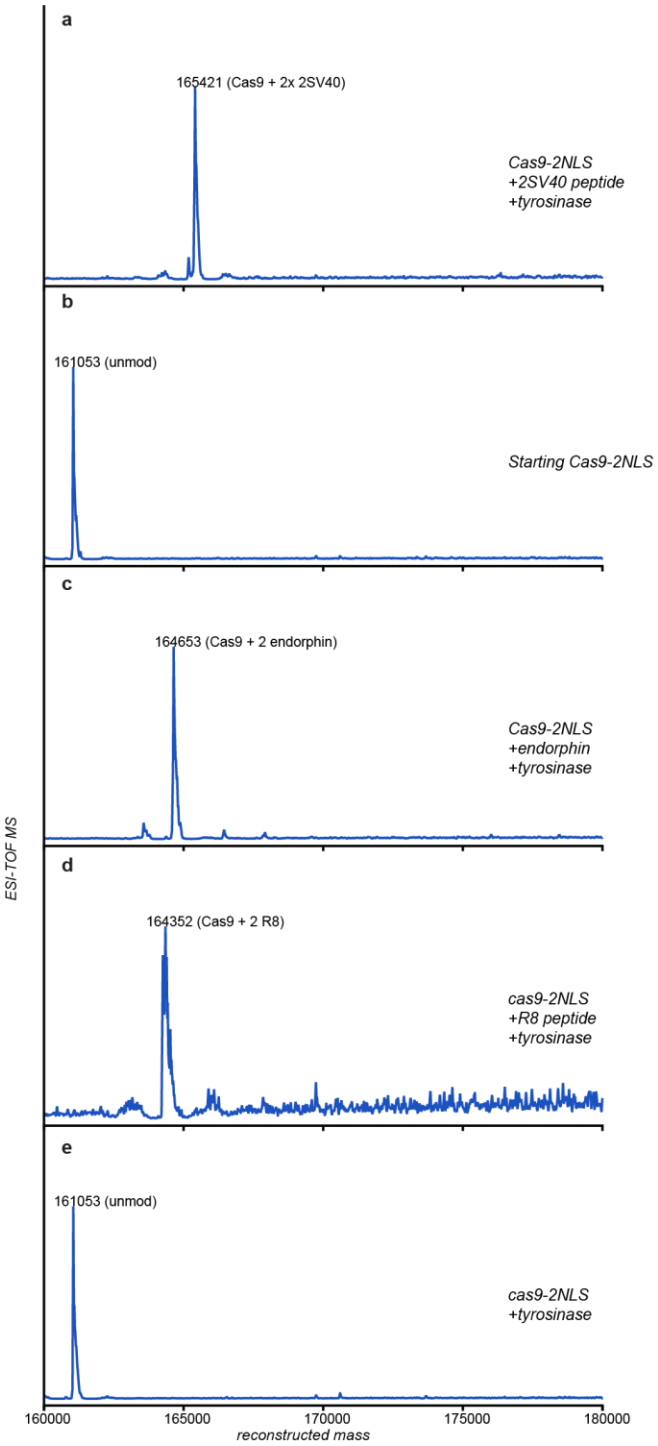


Figure S13. Cas9 peptide modifications. Cas9 at 10 µM combined with 100 µM of the respective peptides for 1 h at 4 °C in Buffer 1. Sample include (a) Cas9 coupled to the 2SV40 peptide, (b) a control sample of uncoupled Cas9-2NLS, (c) Cas9-2NLS coupled to endorphin, (d) Cas9-2NLS coupled to an R8 peptide. As shown in (e), Cas9 exposed to tyrosinase conditions with no added phenols exhibits no oxidation or addition products.

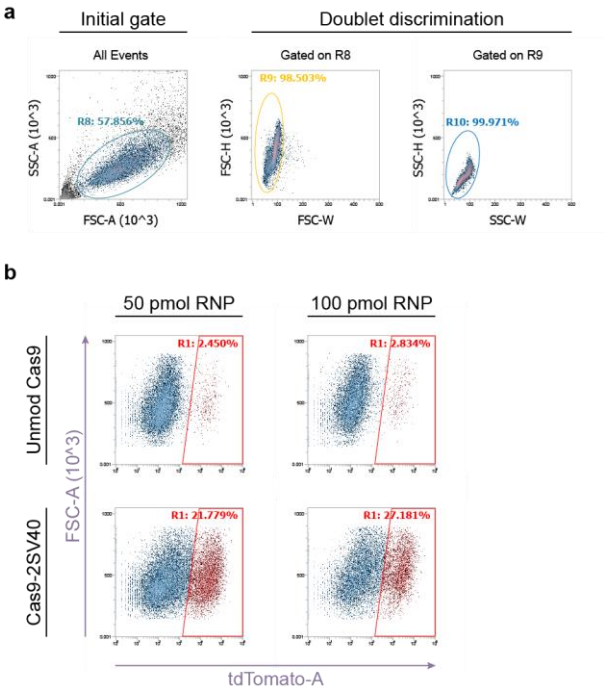


Figure S14. Delivery of Cas9 RNP into neural progenitor cells (NPCs). (a) Flow cytometry gating was used as indicated. Shown are the flow cytometry gates used for selection of viable cells and doublet discrimination, based on one representative sample (Ai9 NPCs treated with 100 pmol unmodified Cas9). (b) Genome editing was measured for Ai9 NPCs treated with various amounts of Cas9 RNPs conjugated with the indicated peptides. Shown are select representative examples to illustrate gating and quantification procedures. The scatter plots represent data gated on R10, shown in panel (a). Unmod Cas9: unmodified Cas9. Cas9-2SV40: Cas9 conjugated with two peptides, each representing two consecutive SV40 nuclear localization sequences.

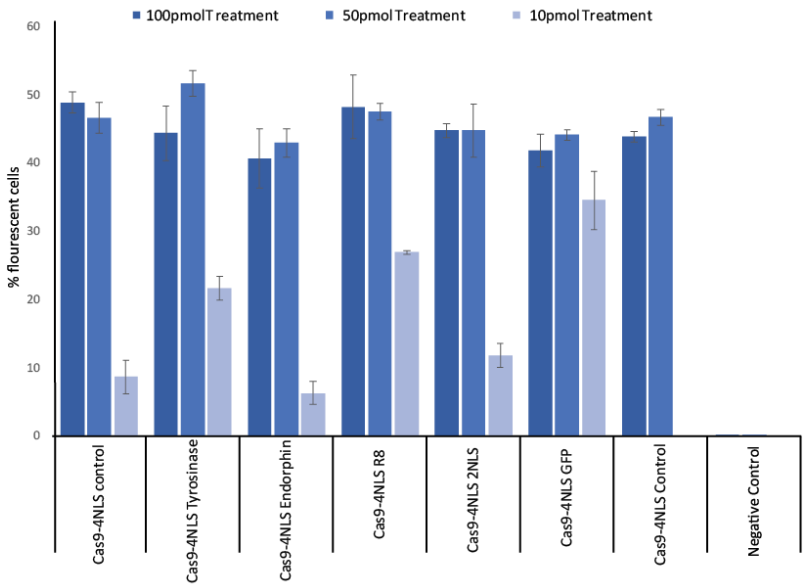


Figure S15. Editing of Ai9 neural progenitor cells with Cas9-4NLS conjugated to the indicated peptides. All treatment conditions showed similar editing despite the diverse peptides. Error bars show the standard deviation of triplicates.

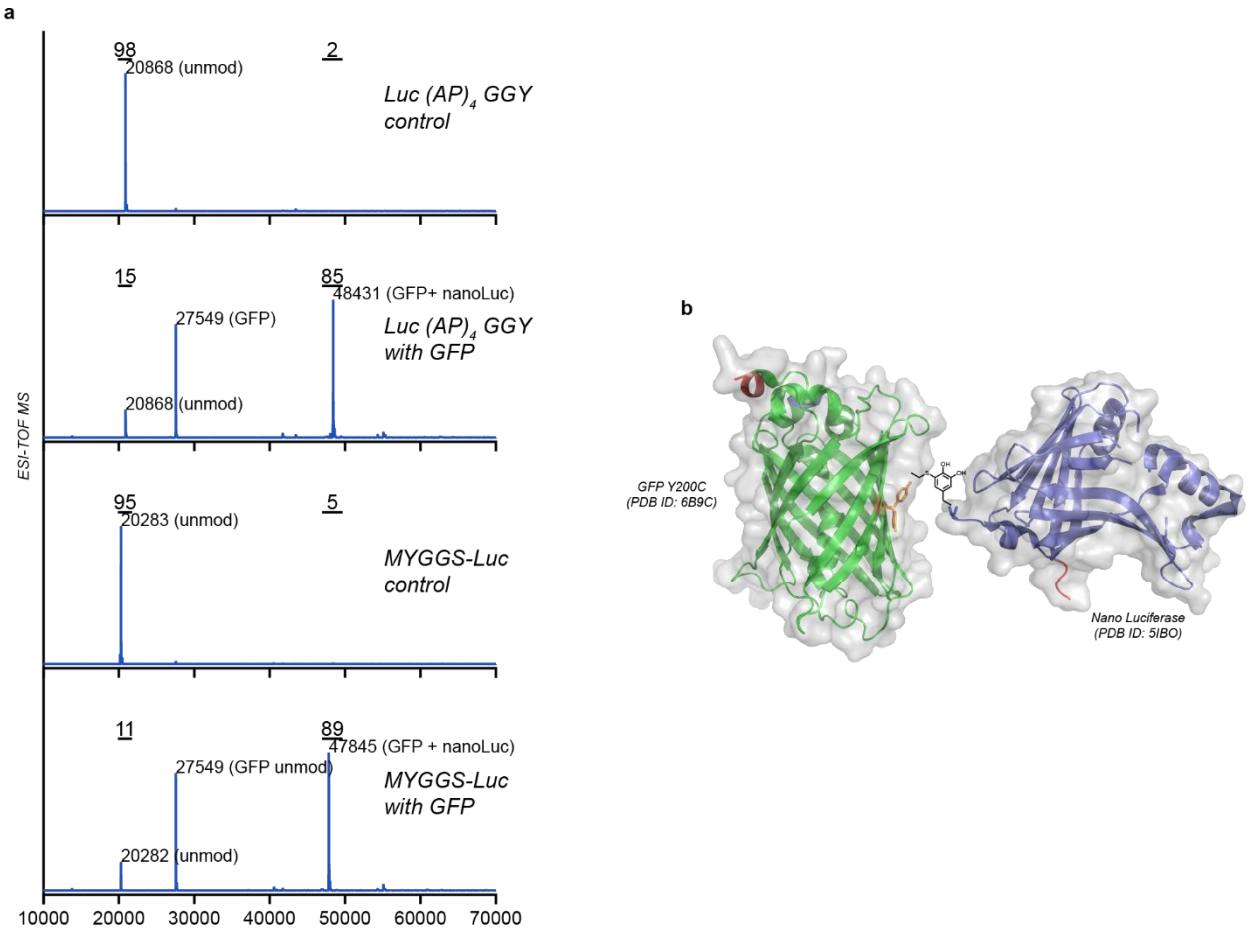


Figure S16. Coupling of two nano-luciferase constructs to Y200C sfGFP. In each reaction 10 µM of nano-luciferase was combined with 20 µM Y200C sfGFP and 200 nM abTYR for 30 min at RT. The $(AP)_4$GGY Luciferase possessed a C-terminal tyrosine tag consisting of the sequence nanoLuc-APAPAPAPGGY. The MYGGS-Luc construct consisted of nanoLuc with an N-terminal MYGGS tag. Both N- and C-terminally tagged constructs showed over 80% conversion to the singly modified product.

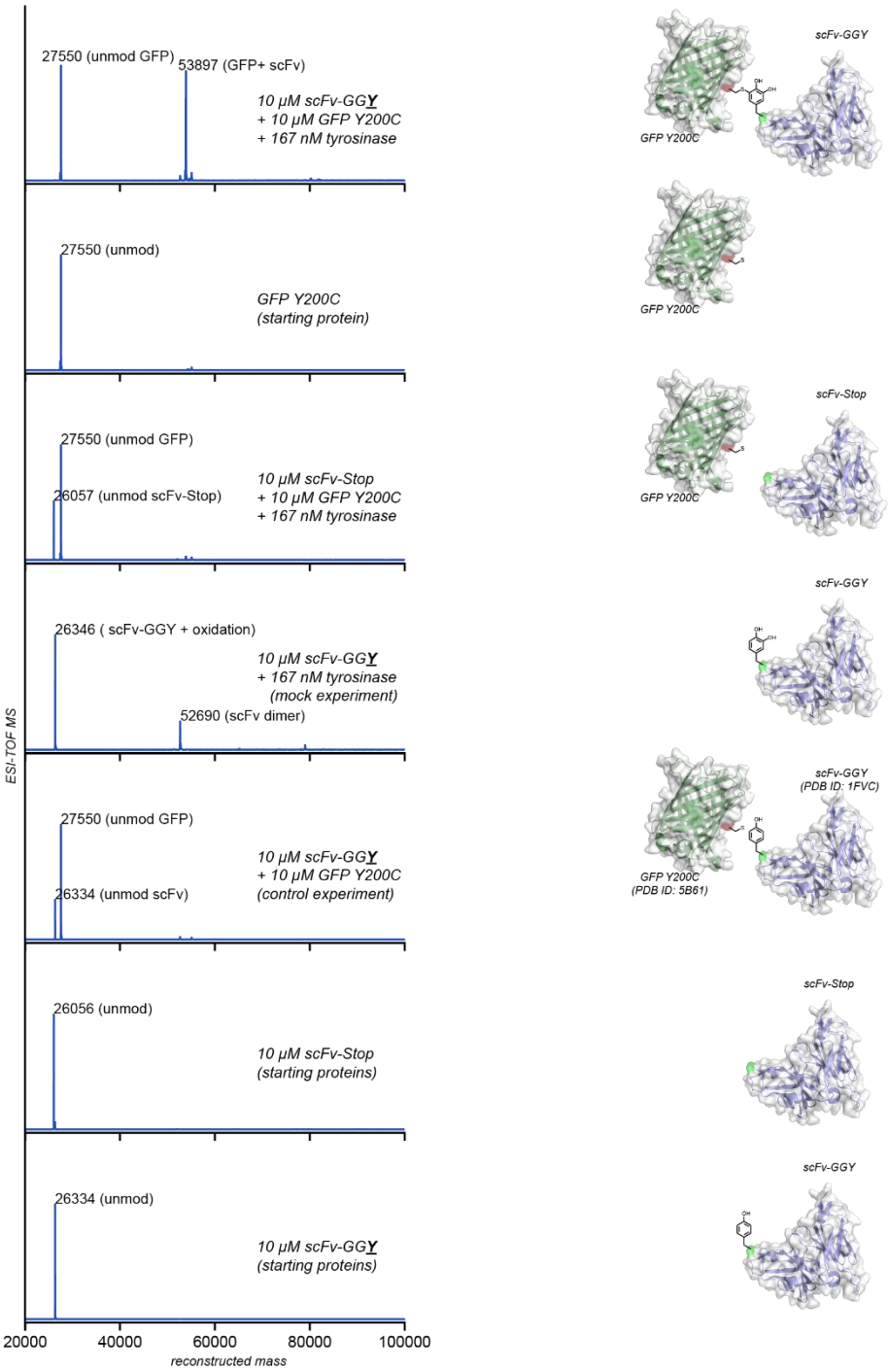
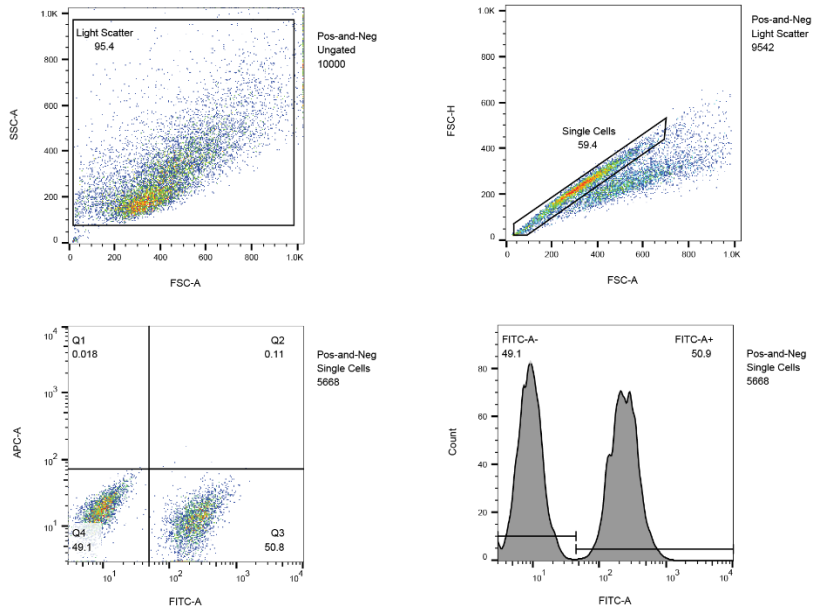


Figure S17. Coupling of Trastuzumab-derived scFV (Tras) and GFP Y200C. ESI-TOF mass spectra show coupling between the scFV and GFP Y200C. A 5 µM solution of scFv was combined with 5 µM of GFP-Y200C to yield near-quantitative conversion to the GFP-scFV dimer.



Gate Depth	Name	Statistic (%)	#Cells
	Pos-and-Neg		10000
>	Pos-and-Neg/Light Scatter	95.4	9542
> >	Pos-and-Neg/Light Scatter/Single Cells	59.4	5668
> > >	Pos-and-Neg/Light Scatter/Single Cells/FITC-A+	50.9	2886
> > >	Pos-and-Neg/Light Scatter/Single Cells/FITC-A-	49.1	2782
	Sample 1		10000
>	Sample 1/Light Scatter	98	9796
> >	Sample 1/Light Scatter/Single Cells	80.2	7852
> > >	Sample 1/Light Scatter/Single Cells/FITC-A+	0.23	18
> > >	Sample 1/Light Scatter/Single Cells/FITC-A-	99.8	7834
	Sample 2		10000
>	Sample 2/Light Scatter	90.4	9041
> >	Sample 2/Light Scatter/Single Cells	62.7	5665
> > >	Sample 2/Light Scatter/Single Cells/FITC-A+	0.19	11
> > >	Sample 2/Light Scatter/Single Cells/FITC-A-	99.8	5654
	Sample 3		10000
>	Sample 3/Light Scatter	94.5	9446
> >	Sample 3/Light Scatter/Single Cells	67.9	6416
> > >	Sample 3/Light Scatter/Single Cells/FITC-A+	5.77	370
> > >	Sample 3/Light Scatter/Single Cells/FITC-A-	94.2	6046

Gate Depth	Name	Statistic (%)	#Cells
	Sample 4		10000
>	Sample 4/Light Scatter	93.3	9332
> >	Sample 4/Light Scatter/Single Cells	67.2	6270
> > >	Sample 4/Light Scatter/Single Cells/FITC-A+	0.21	13
> > >	Sample 4/Light Scatter/Single Cells/FITC-A-	99.8	6257
	Sample 5		10000
>	Sample 5/Light Scatter	93.9	9387
> >	Sample 5/Light Scatter/Single Cells	67.3	6314
> > >	Sample 5/Light Scatter/Single Cells/FITC-A+	0.33	21
> > >	Sample 5/Light Scatter/Single Cells/FITC-A-	99.7	6293
	Sample 6		10000
>	Sample 6/Light Scatter	91.4	9138
> >	Sample 6/Light Scatter/Single Cells	62.5	5707
> > >	Sample 6/Light Scatter/Single Cells/FITC-A+	99.2	5662
> > >	Sample 6/Light Scatter/Single Cells/FITC-A-	0.79	45

Figure S18. Labeling of HER2^{+} cells using Trastuzumab scFv-sfGFP conjugate. Representative plots show the gating scheme on a mixture of 1:1 untreated MDA-MB-468 (HER2^{-}) cells and SK-BR-3 (HER2^{+}) cells treated with Trastuzumab scFv-sfGFP conjugate (as in Sample 6). The gating scheme was applied uniformly to all samples and statistics were tabulated. The fluorescence intensity of single cells was measured using the 488 nm laser line and the FITC filter set.

Table S1.

Cysteine-Tyrosine coupling stability			
Reaction Condition	Day 1	Day 3	Day 6
pH 5	86%	93%	85%
pH 7	86%	88%	85%
pH 8	88%	89%	83%
TCEP + Tropolone	80%	79%	83%
N-mercapto-butanol	82%	82%	79%

Modification levels for MS2-N87C coupled to α-endorphin over a 6-day time course, corresponding to Figure S6. Data points represent the analysis of individual runs analyzed by peak integration using chartograph.com online software.

3.9 References

1. Hermanson, G. T. *Bioconjugate Techniques*. *Bioconjugate Techniques* (Elsevier, 2008). doi:10.1016/B978-0-12-370501-3.X0001-X

2. Witus, L. S. & Francis, M. B. Using synthetically modified proteins to make new materials. *Acc. Chem. Res.* **44**, 774–783 (2011).

3. Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nature Chemical Biology* **7**, 876–884 (2011).

4. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie Int. Ed.* **40**, 2004–2021 (2001).

5. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chemie - Int. Ed.* **48**, 6974–6998 (2009).

6. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science (80-.).* **266**, 776–779 (1994).

7. Theile, C. S. *et al.* Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1800–1807 (2013).

8. Kim, Y. G. & Chandrasegaran, S. Chimeric restriction endonuclease. *Proc. Natl. Acad. Sci.* **91**, 883–887 (1994).

9. Bolukbasi, M. F. *et al.* Orthogonal Cas9-Cas9 chimeras provide a versatile platform for genome editing. *Nat. Commun.* **9**, 4856 (2018).

10. Oakes, B. L. *et al.* CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification. *Cell* **176**, 254-267.e16 (2019).

11. Schmidt, S. R. *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges*. *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges* (John Wiley & Sons, Inc., 2013). doi:10.1002/9781118354599

12. Behrens, C. R. *et al.* Rapid chemoselective bioconjugation through oxidative coupling of anilines and aminophenols. *J. Am. Chem. Soc.* **133**, 16398–16401 (2011).

13. Obermeyer, A. C., Jarman, J. B. & Francis, M. B. N-Terminal Modification of Proteins with o -Aminophenols. *J. Am. Chem. Soc.* **136**, 9572–9579 (2014).

14. Maza, J. C. *et al.* Enzymatic Modification of N-Terminal Proline Residues Using Phenol Derivatives. *J. Am. Chem. Soc.* **141**, 3885–3892 (2019).

15. Cory, J. G. & Frieden, E. Differential Reactivities of Tyrosine Residues of Proteins to Tyrosinase. *Biochemistry* **6**, 121–126 (1967).

16. Ito, S., Kato, T., Shinpo, K. & Fujita, K. Oxidation of tyrosine residues in proteins by tyrosinase. Formation of protein-bonded 3,4-dihydroxyphenylalanine and 5-S-cysteinyl-3,4-dihydroxyphenylalanine. *Biochem. J.* **222**, 407–411 (1984).

17. Selinheimo, E., Autio, K., Kruus, K. & Buchert, J. Elucidating the Mechanism of Laccase and Tyrosinase in Wheat Bread Making. *J. Agric. Food Chem.* **55**, 6357–6365 (2007).

18. Lantto, R., Puolanne, E., Kruus, K., Buchert, J. & Autio, K. Tyrosinase-Aided Protein Cross-Linking: Effects on Gel Formation of Chicken Breast Myofibrils and Texture and Water-Holding of Chicken Breast Meat Homogenate Gels. *J. Agric. Food Chem.* **55**, 1248–1255 (2007).

19. Thalmann, C. R. & Lötzbeyer, T. Enzymatic cross-linking of proteins with tyrosinase. *Eur. Food Res. Technol.* **214**, 276–281 (2002).

20. Bruins, J. J. *et al.* Inducible, Site-Specific Protein Labeling by Tyrosine Oxidation–Strain-Promoted (4 + 2) Cycloaddition. *Bioconjug. Chem.* **28**, 1189–1193 (2017).

21. Montanari, E. *et al.* Tyrosinase-Mediated Bioconjugation. A Versatile Approach to Chimeric Macromolecules. *Bioconjug. Chem.* **29**, 2550–2560 (2018).

22. Marmelstein, A. M. *et al.* Tyrosinase-Mediated Oxidative Coupling of Tyrosine Tags on Peptides and Proteins. *J. Am. Chem. Soc.* **142**, 5078–5086 (2020).

23. Ito, S. & Prota, G. *A facile one-step synthesis of cysteinyldopas using mushroom tyrosinase*. *Experientia* **33**, 1118–1119 (Birkhäuser-Verlag, 1977).

24. Agrup, G., Hansson, C., Rorsman, H. & Rosengren, E. The effect of cysteine on oxidation of tyrosine dopa, and cysteinyldopas. *Arch. Dermatol. Res.* **272**, 103–115 (1981).

25. Wu, W., Hsiao, S. C., Carrico, Z. M. & Francis, M. B. Genome-free viral capsids as multivalent carriers for taxol delivery. *Angew. Chemie - Int. Ed.* **48**, 9493–9497 (2009).

26. Capehart, S. L., Coyle, M. P., Glasgow, J. E. & Francis, M. B. Controlled integration of gold nanoparticles and organic fluorophores using synthetically modified MS2 viral capsids. *J. Am. Chem. Soc.* **135**, 3011–3016 (2013).

27. Valegård, K., Liljas, L., Fridborg, K. & Unge, T. The three-dimensional structure of the bacterial virus MS2. *Nature* **345**, 36–41 (1990).

28. Farkas, M. E. *et al.* PET Imaging and Biodistribution of Chemically Modified Bacteriophage MS2. *Mol. Pharm.* **10**, 69–76 (2013).

29. Carrico, Z. M., Romanini, D. W., Mehl, R. A. & Francis, M. B. Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* **0**, 1205 (2008).

30. Muñoz-Muñoz, J. L. *et al.* Catalytic oxidation of o-aminophenols and aromatic amines by mushroom tyrosinase. *Biochim. Biophys. Acta - Proteins Proteomics* **1814**, 1974–1983 (2011).

31. Maza, J. C. *et al.* Secondary Modification of Oxidatively-Modified Proline N-termini for the Construction of Complex Bioconjugates. *Org. Biomol. Chem.* **18**, 1881–1885 (2020).

32. Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals. *Molecular Physics* **115**, 2315–2372 (2017).

33. Cossi, M., Rega, N., Scalmani, G. & Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **24**, 669–681 (2003).

34. Barone, V. & Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **102**, 1995–2001 (1998).

35. Takano, Y. & Houk, K. N. Benchmarking the conductor-like polarizable continuum model (CPCM) for aqueous solvation free energies of neutral and ionic organic molecules. *J. Chem. Theory Comput.* **1**, 70–77 (2005).

36. Schiapparelli, L. M. *et al.* Direct detection of biotinylated proteins by mass spectrometry. *J. Proteome Res.* **13**, 3966–3978 (2014).

37. Fontaine, S. D., Reid, R., Robinson, L., Ashley, G. W. & Santi, D. V. Long-Term Stabilization of Maleimide–Thiol Conjugates. *Bioconjug. Chem.* **26**, 145–152 (2015).

38. Staahl, B. T. *et al.* Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.* **35**, 431–434 (2017).

39. Pandya, H., Gibo, D. M., Garg, S., Kridel, S. & Debinski, W. An interleukin 13 receptor α 2-specific peptide homes to human Glioblastoma multiforme xenografts. *Neuro. Oncol.* **14**, 6–18 (2012).

40. Wender, P. A. *et al.* The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: Peptoid molecular transporters. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13003–13008 (2000).

41. ElSohly, A. M. *et al.* ortho-Methoxyphenols as Convenient Oxidative Bioconjugation Reagents with Application to Site-Selective Heterobifunctional Cross-Linkers. *J. Am. Chem. Soc.* **139**, 3767–3773 (2017).

42. Allen, J., Brock, D., Kondow-McConaghy, H. & Pellois, J.-P. Efficient Delivery of Macromolecules into Human Cells by Improving the Endosomal Escape Activity of Cell-Penetrating Peptides: Lessons Learned from dfTAT and its Analogs. *Biomolecules* **8**, 50 (2018).

43. Cronican, J. J. *et al.* Potent Delivery of Functional Proteins into Mammalian Cells in Vitro and in Vivo Using a Supercharged Protein. *ACS Chem. Biol.* **5**, 747–752 (2010).

44. Lönn, P. *et al.* Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.* **6**, 32301 (2016).

45. Futaki, S. *et al.* Arginine-rich peptides. An abundant source of membrane-permeable peptides

having potential as carriers for intracellular protein delivery. *J. Biol. Chem.* **276**, 5836–40 (2001).

46. Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A. A. X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling. *J. Mol. Biol.* **229**, 969–995 (1993).

47. Obermeyer, A. C., Jarman, J. B. & Francis, M. B. N-Terminal Modification of Proteins with o -Aminophenols. *J. Am. Chem. Soc.* **136**, 9572–9579 (2014).

48. Anders, C. & Jinek, M. In Vitro Enzymology of Cas9. in *Methods in enzymology* **546**, 1–20 (Elsevier Inc., 2014).

49. Tong, G. J., Hsiao, S. C., Carrico, Z. M. & Francis, M. B. Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* **131**, 11174–11178 (2009).

50. Li, M. *et al.* Discovery and Characterization of a Peptide That Enhances Endosomal Escape of Delivered Proteins in Vitro and in Vivo. *J. Am. Chem. Soc.* **137**, 14084–14093 (2015).

51. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).

Chapter 4. Tolerance induction towards CRISPR Cas9

Abstract:

CRISPR based gene editing has had an unprecedented impact in the study of medicine and biology over the past decade. One of the major hurdles in the adoption of CRISPR therapies is the high prevalence of immune response against CRISPR proteins, with over 70% of patients sensitive to SpyCas9 alone. In order to prevent inflammation or rejection of edited cells it is vital to develop tools to minimize the immune response towards these proteins. Recent work has shown a number of immunoglobulin derivatives capable of inducing immune tolerance towards fused partners, but attempts to fuse these proteins to Cas9 have been stymied due to its size. We show that resin bound tyrosinase is capable of catalyzing efficient coupling of SpyCas9 to both IgG Fc domains and the MHC Class II targeting VHH7 nanobody. Initial mouse studies support strong tolerance induction, highlighting the versatile capabilities of tyrosinase coupling and provide a tantalizing view of modular therapeutic design.

This work conducted in collaboration with the Raulet and Barton Labs. Mouse work conducted by Gabrielle Reiner in the Raulet lab. Optimization of nanobody coupling and endotoxin removal was done in collaboration with Johnathan Maza.

4.1 Immunogenicity of Cas9 therapies

SpyCas9 is one of the most studied of the CRISPR (Clustered Regularly Interspersed Short Palindromic Repeat) effector proteins and the most promising for a variety of clinical implementations.[1–3] Unfortunately, a number of recent studies have emerged showing high levels of immune response to cells treated with Cas9.[3,4] This is exacerbated in cases were cells endogenously express the protein, as higher expression levels are correlated with Major Histocompatability Complex (MHC) display.[3] The immune response to SpyCas9 is perhaps unsurprising as the protein originates from a pathogenic bacterium, and even though the CRISPR system is not directly involved in bacterial pathogenesis it is still a candidate for macrophage and T-cell recognition.[5] Unfortunately, adopting Cas9 orthologues from less pathogenic bacteria does not guarantee to remedy this issue for most therapeutic applications as repeat dosing are often required, ultimately leading to immune recognition and rejection.[5,6] Using a delivery vector consisting of Ribonuclear Protein (RNP) alone appears to improve outcomes by decreasing the cytosolic dosage and resultant MHC display, however RNP delivery systems suffer from lower delivery efficiency, making repeat dosing even more important.[7]

The Catch-22 of CRISPR proteins as therapeutics indicates an urgent need to limit immune response to these proteins in order to minimize the risk of rejection of edited cells. To this end, it is necessary to develop technologies capable of reducing the immune reaction to CRISPR proteins. Tolerance induction has long been a goal of immunological research, and is key for the treatment of autoimmune disorders, allergies, and organ transplants.[8–11] One of the primary difficulties in understanding the mechanisms of tolerance induction is that very few tolerance inducing systems have been discovered compared to the wide range of adjuvants and other immune stimulating molecules used in vaccine production and other applications.[12,13] The lack of tolerance inducing molecules has made it challenging to fully understand the mechanism behind this process and it is still poorly understood, further impacting our ability to generate new tolerance induction strategies.[10,14,15]

One of the most well studied tolerance induction systems is the use of Immunoglobulin (IgG) Fc domains fused to a target.[15,16] In these studies the Fc domain is recombinantly attached to the antigen via genetic fusion and then administered to the mouse model after which immune response is monitored by interferon gamma (IFNγ) response to the antigenic protein.[15] The mechanism of this tolerance induction is not well understood but it has been shown to increase the number of regulatory T cells and lower Antigen Presenting Cell (APC) response to the target protein.[16] The primary drawback of the Fc system is the challenge of producing Fc genetic fusions, as the proteins must be expressed in mammalian cells for proper glycosylation, and the attached antigenic target must be expressed and folded correctly to ensure tolerance. This is feasible for smaller proteins, but for something the size of Cas9 becomes a daunting challenge.

In a similar manner, unpublished work from the Ploegh lab indicates that nanobodies targeting MHC II could help induce tolerance to a coupled protein.[17] Specifically, the VHH7 nanobody is believed to target tolerogenic antigen presenting cells (APCs), allowing for selective tolerance induction towards coupled proteins. Nanobodies do not require glycosylation and can be expressed in bacteria. They also benefit from being only 15 kDa in size,

allowing for more facile production of genetic fusions. However, they still require periplasmic expression for proper disulfide bond formation and cannot be purified when expressed as a fused protein with Cas9. In similar experiments, the previously published sortase based coupling method was also shown to be inefficient in coupling to Cas9.[17,18] Thus, it would be ideal to have a method of easily attaching these MHC II targeting nanobodies (VHH7) to Cas9 directly.

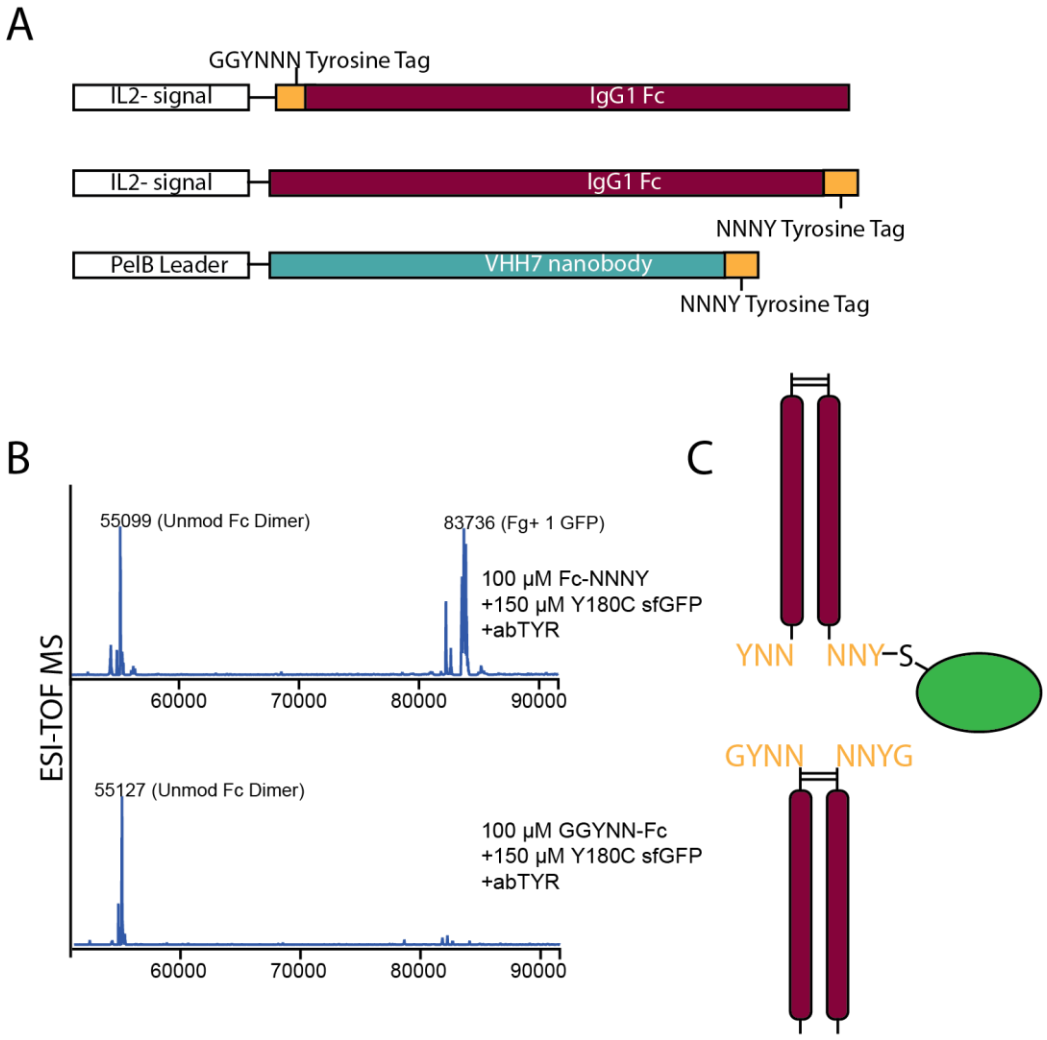


Figure 1. Initial design and modification of IgG Fc domains with tyrosine tags. **A**. Protein design for tyrosine tagged immunoglobulin derivatives. IgG Fc domains with tyrosine tags at N- and C-termini. Due to the presence of the cleavable IL-2 signal sequence we chose to use the N-terminal tag GGYNNN. In contrast, the C-terminus was modified with the tag NNNY. For the VHH7 nanobody a C-terminal NNNY tag was used due to prior success with ScFvs. **B**. ESI-TOF spectra showing products of a reaction between 100 µM Fc and 150 µ M Y200C sfGFP. The C-terminal Fc construct showed efficient coupling to the GFP, while the N-terminal Fc construct showed little modification. **C**. Representation of the reaction products shown in **B**.

Recent work from our lab has shown that the Tyrosinase coupling reaction can be used to rapidly produce functional protein conjugates, but we had not shown successful coupling of

Cas9 to a nanobody.[19–21] This work endeavors to establish a protocol for the creation of Cas9-immunoglobulin fusions to induce immune tolerance of SpyCas9.

4.2 Design of Fc constructs and Fc coupling

N-terminal and C-terminal tyrosine tagged IgG Fc constructs (Figure 1A) were produced by the Raulet and Barton labs and screened for complementarity with the tyrosinase coupling protocol. C-terminal tyrosine tags were found to be most reactive for activation and coupling to a cysteine GFP (Y182C sfGFP) (Figure 1 B,C).

Initial trials for the attachment of Cas9 to the IgG Fc constructs emphasized the C-terminal construct due to the favorable GFP coupling. We changed buffer conditions to a trehalose solution and ran the reaction on ice to increase stability of Cas9 during reaction.[19] Due to concerns about Cas9 stability early reactions were conducted with the sgRNA- Cas9 ribonucleoprotein (RNP) as seen in Figure 2A. Despite screening a range of equivalencies there was little improvement in yield from baseline conditions.

Unfortunately, incorporation of sgRNA is non-ideal due to potential confounding factors from immune response to the RNA.[22] Several time course studies were conducted to determine if full conversion could be reached with the free Cas9 and Fc domain but found a strict plateau after roughly 1.5 h (Figure 2B). Incomplete reactivity poses a major barrier to these experiments as any uncoupled Cas9 has the potential to act as an adjuvant and stimulate a greater immune response in our assay. Unfortunately, common methods of isolating Fc domains such as using a protein A column require harsh conditions for elution, which would denature the sensitive apo-Cas9.

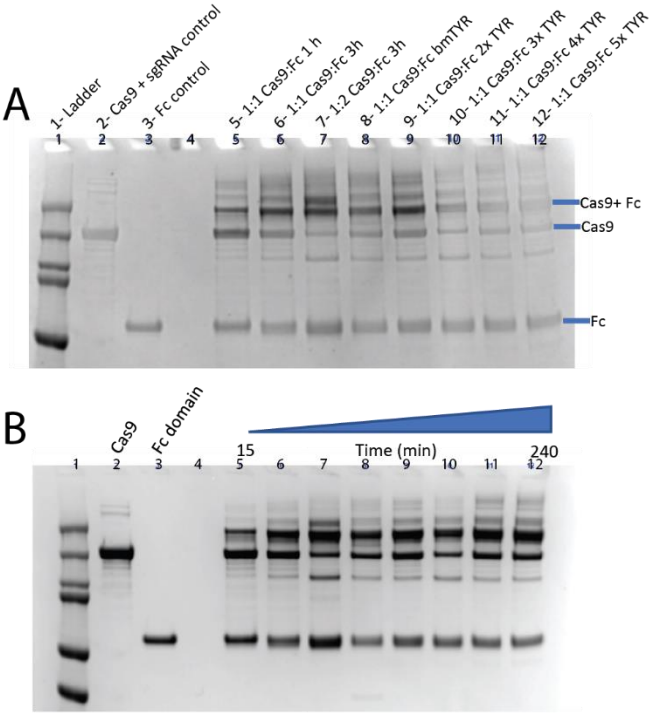


Figure 2. Gels monitoring Cas9-Fc conjugation reactions. **A**. Screen of reaction conditions to couple Cas9 with sgRNA to Fc-NNNY. Ratio of Cas9:Fc listed above each well along with reaction time and tyrosinase equivalents, where 1x TYR is equivalent to 12 U/L of abTYR. **B**. Time course of apo-Cas9 conjugation to Fc-NN-NY taken in every 15 min for the first hour followed by every 30 min for the remainder of the reaction.

After comparison between reducing and non-reducing conditions, we found that the majority of the uncoupled Fc domains were in the dimer form, indicating that oxidation had occurred but that some form of cross reactivity was causing the two arms of the Fc domain to self-couple, preventing their reaction with Cas9 cysteines (Figure 3A). Nonetheless, We were still able to confirm the creation of mono and di-Fc modified Cas9 (Figure 3A,B).

The cross coupling of the Fc domains is not unprecedented, previous works by the d'Ischia lab have shown ready dimerization of *o*-quinone and catechols. [23,24] Lysine residues on the protein surface may also be capable of reacting with the *o*-quinone if sufficiently reactive. Upon closer examination of the structure it is apparent that there is ample flexibility to allow for cross reactivity of the C-terminal tags (Figure 3C,D). Future work will focus on changing the C-terminal tyrosine tags to a shorter or less flexible linker in hopes of generating non-self-coupling Fc domains.[25,26]

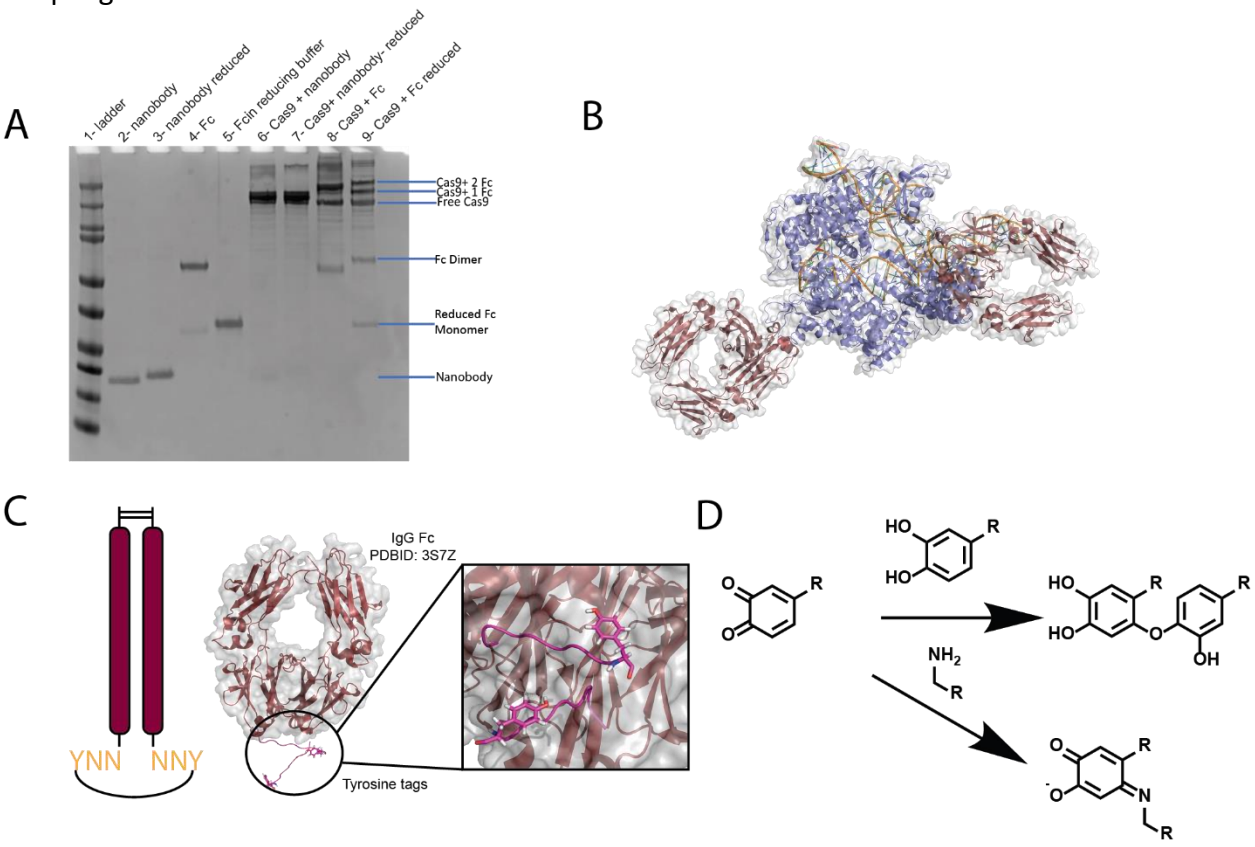


Figure 3. Analysis of side products in Fc coupling reactions. **A**. SDS-PAGE showing both reducing and non-reducing conditions for Cas9 immunoglobulin reactions. All reactions were conducted on 10 μM Cas9 with 100 μM of the corresponding Fc or nanobody in pH 6.5 phosphate buffer with trehalose on ice for 1 h. Prior to loading on gel samples were combined with 4X loading buffer with or without β-mercaptoethanol (BME) as reducing agent. remaining Fc dimer after reduction is indicative of Fc cross coupling. **B**. Structure of Cas9 (shown in purple. PDBID: 5F9R) coupled to two Fc domains (shown in red. PDBID: 3S7Z) . **C**. Representation of the self coupling of tyrosine tags in the Fc dimer. Crystal structures with modeled Tyrosine tags shown in pink show high potential for tyrosine tag cross-reactivity. **D**. Reactions for cross coupling of Fc domains showing coupling of the *o*-quinone with catechol or free amines.

Due to challenges encountered in coupling the Fc domain, we chose to instead focus on coupling between Cas9 and the VHH7 nanobody. Unfortunately, separation of the Fc-Cas9 conjugates from free Cas9 would be quite challenging as protein A columns which could capture the Fc tails requires elution using highly acidic media which would have denatured the Cas9. The one potential alternative we did not explore is the use of ion exchange

chromatography which poses similar limitations due to the requirement of extreme pH conditions for elution.

4.3 VHH7-nanobody conjugates

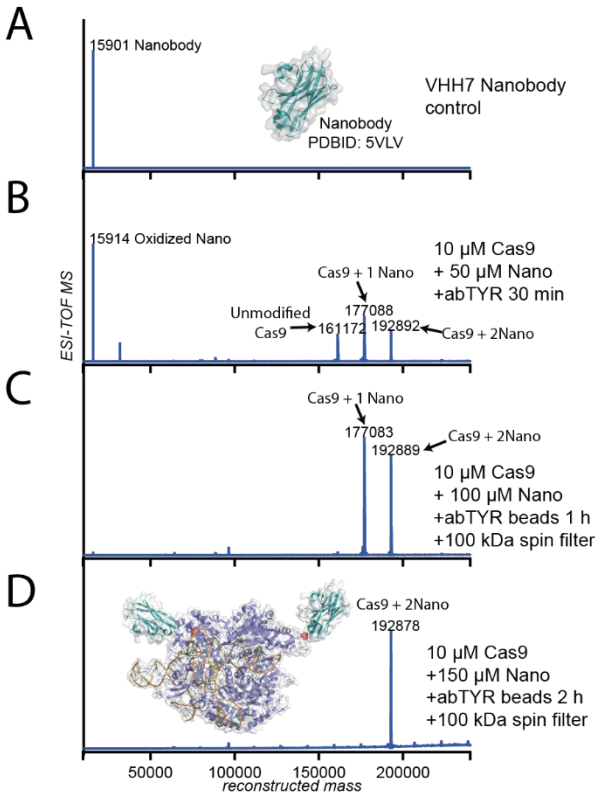


Figure 4. ESI TOF analysis of Cas9-VHH7 nanobody coupling. **A**. Control of VHH7 nanobody with crystal structure in teal (PDBID: 5VLV). **B**. Initial trial of Cas9 coupling to VHH7. Cas9 at 10 μM combined with 50 μM VHH7 and abTYR for 30 min, showing the unmodified Cas9 with both the single and double addition of VHH7. **C**. Optimization of filtering for Cas9 samples with abTYR bound to resin beads.Cas9 at 10 μM was combined with 100 μM VHH7 and abTYR beads on ice for 1 h. Spin filter with 100 kDa MWCO used to remove excess nanobody. **D**. Optimized production of Cas9 with 2 nanobodies. Reaction run at 2 mL scale to produce ~1 mg of modified Cas9. Similarly 10 μM Cas9 combined with 150 μM VHH7 and abTYR beads at 4 °C for 2 h with end-over-end shaking.

VHH7 nanobodies were produced with both N and C-terminal tyrosine tags. We found that the requirement for the pectate lyase B (PelB) leader peptide and its variable cleavage led to inefficient coupling of tyrosine tags on the N terminus, so C-terminal tags were again used in these experiments. In contrast to the Fc combination, attachment of the VHH7 nanobodies was highly efficient. As seen in Figure 4, the coupling was efficient from an early stage, yielding high degrees of the mono a di-modified Cas9. After optimizing conditions, we were able to achieve full conversion to a doubly modified Cas9 using a surface immobilized tyrosinase (Figure 4D)

Scale-up of production was relatively trivial, requiring a slight increase in nanobody:Cas9 molar ratio to 9:1. Concerned with the potential for contamination from free tyrosinase in our reactions, we changed our protocol to include tyrosinase bound to resin beads. By incubating for 2 h with end-over-end rotation of the reaction vessel we achieved quantitative yield of the doubly modified Cas9-VHH7. With the pure Cas9-VHH7 construct in hand we began trials to remove excess endotoxin. Unfortunately, the filters employed by Berkeley's Macro lab to remove endotoxin from Cas9 prep (Pall Corp's Mustang Syringe filters) adhered tightly to the free nanobody and Cas9-VHH7 conjugates. Endotoxin spin filters from Pierce proved less deleterious to our samples, producing low endotoxin samples after 2 h of incubation

Mouse studies with the sample were conducted, however, it became readily apparent that significant aggregation of the Cas9 control sample was present leading to immediate mouse death. This aggregation

was likely a result of disulfide bond formation between Cas9 in solution. Since other samples had the surface cysteines capped we did not observe significant aggregation.

To understand the source of the aggregation we performed a series of DLS experiments that confirmed that the non-reduced Cas9 samples showed particle sizes as large as 400 nm, while the freshly prepared Cas9 had a homogenous peak at around 20 nm (Figure 5A, B, C).

Despite unfortunate losses, we were able to obtain promising data in our initial mouse trial, with the Cas9-nanobody treated mouse producing significantly less IFNγ response, and thereby greater tolerance than the nanobody and vehicle controls (Figure 5D). However, further experiments are required due to a lower than expected response in mice treated with the Cas9 control and a higher than expected result from the nanobody control group.

To avoid future complications, all samples were doped with 1mM DTT after quenching to prevent aggregate formation and sterile filtered through 0.2 µm membranes to remove any large particulates from the sample before freezing. Three months after the initial round of mouse experiments the protocol was repeated with reducing buffer and included Cas9 and the nanobody co-administered as a control. These data showed a more expected result, with Cas9 and buffer treated mice showing the greatest IFNγ via cytokine staining when exposed to a fresh dose of Cas9. We saw a significant reduction in IFNγ response between the Cas9 alone and the Cas9-nanobody fusion, indicating partial tolerance induction. However, the IFNγ signal in nanobody treated controls and the Cas9 with nanobody control was comparable to the Cas9-nanobody fusion. Repeat experiments at different dosages will likely be required to validate these results.

In discussions with the Ploegh lab it appears that 100 ug of conjugate is often delivered for their experiments on tolerance induction. One potential source of error in our experiment is that the Cas9 constructs are roughly 2-5x the size of comparable products from their tests, indicating the need to increase dosage. This work is ongoing and will likely require several more trials before publishable data is obtained.

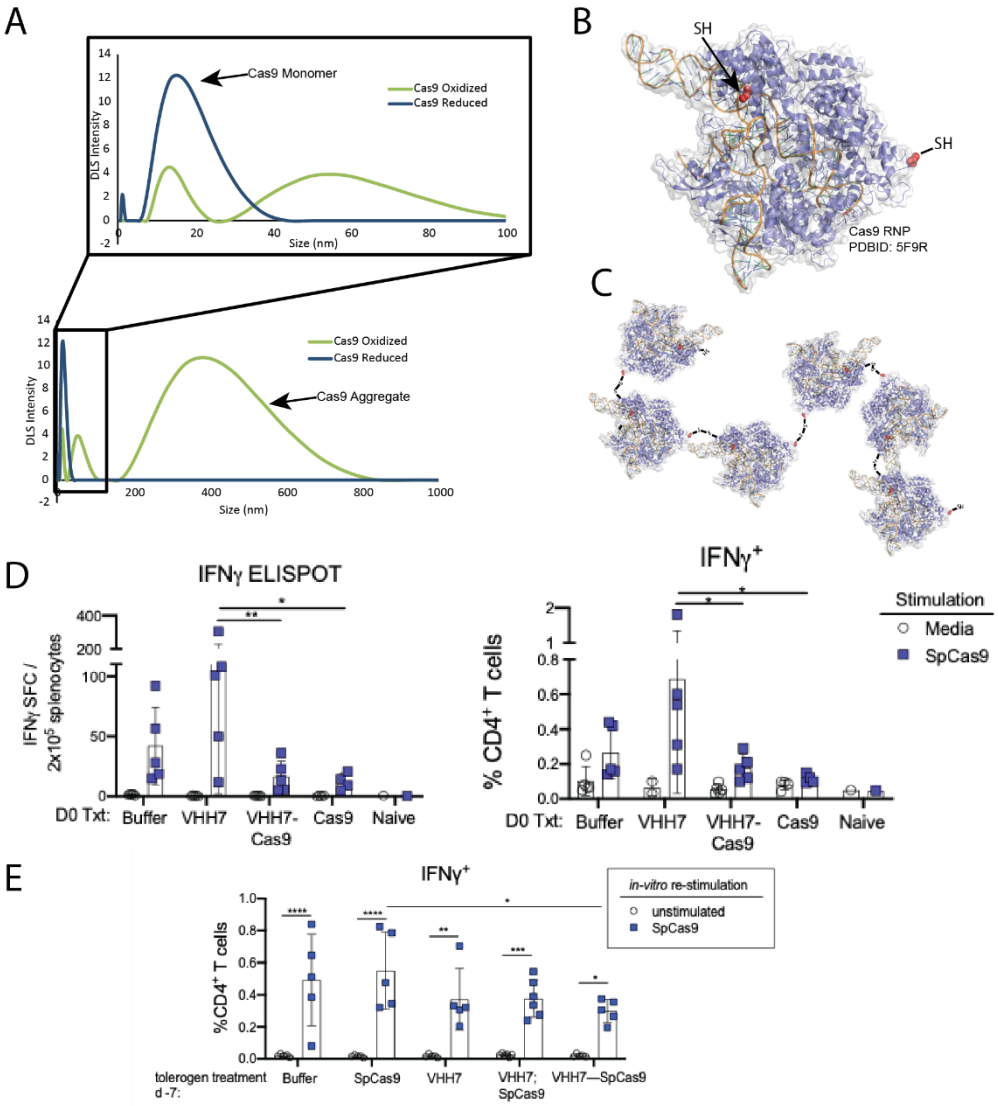


Figure 5. Cas9 aggregation analysis and mouse model results **A**. DLS data showing particle size distribution for Cas9 solutions in oxidized and reducing conditions. Lower trace shows distribution from 0 to 1000 nm, while top trace shows a zoomed in trace from 0 to 100 nm. Oxidized sample shown in light green exhibits a significant peak at 400 nm, indicative of aggregate formation. Reduced peak shows a monodisperse species at 19 nm. **B**. Crystal structure of Cas9 (PDBID 5F9R) with surface cysteines shown in red. **C**. Representation Cas9 aggregates due to disulfide bonding. **D**. IFNγ response in mice treated with Cas9 constructs as measured via ELISPOT (Left) or intrcellular cytokine staining (Right). Mice were treated systemically via retro-orbital injection on day 0 with the VHH7-Cas9 reagent or controls (buffer, VHH7 nanobody alone, Cas9 alone, Cas9 combined with VHH7). Mice were then immunized with Cas9 protein + adjuvant on day 7 and 14. On day 21, spleens were harvested from the treated animals as well as a naive, untreated mouse. The splenocytes were then left unstimulated (media, open circles) or were stimulated with Cas9 protein (blue squares). . **E**. A repeat of the experiment in **D** with freshly prepared samples in reducing buffer. Here we see a significant reduction in immune response comparing Cas9 and nanobody conjugated samples, however the small difference between the VHH7 alone and the conjugate indicates further experimentation is required.

4.4 Conclusion and Future Directions

Targeted tolerance induction is a longstanding goal of immunological research with enormous potential outside of applications towards CRISPR therapies.[10,15,27] The development of a trivial coupling protocol to convey tolerance on a generic protein cargo would have a staggering impact on our approach to allergy treatments, organ transplants, and all manner of autoimmune disorders. Given recent appreciation for the broad prevalence of Cas9 immunogenicity, this work has particular import to the field of CRISPR therapeutics. The ability to pre-treat patients with a tolerance inducing compound could prevent rejection of cells modified with Cas9 and might even be used to prevent immune build up following repeat injections.[3,11]

In this work, we show that tyrosinase coupling is capable of attaching a variety of proteins including Cas9 to both IgG Fc domains and VHH7 nanobodies. While IgG Fc coupling efficiency remains low, modification of the tyrosine tag could enable higher efficiency coupling in future work. It is worth noting that modification of the C-terminus of the Fc domain could prevent binding to relevant receptors, but this remains unclear without further analysis. We lay out a robust protocol capable of generating large volumes of Cas9-nanobody conjugates using resin-bound tyrosinase and validate endotoxin removal protocols for the conjugation products.

Early results from mouse trials with these conjugates are promising but require further experimentation to determine the source of low signal in control samples. If successful, this technique can be readily adapted for a variety of allergens and may offer an alternative to chronic allergy medication.

4.5 General Procedures and Materials

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. Water (ddH_2O) used in biological procedures or as a reaction solvent was de-ionized using a NANOpure purification system (Barnstead, USA). All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Terrific broth used in Cas9 expression was purchased from Mediatech Inc. (Maryland, VA). Spin concentrators were obtained from EMD Millipore (Millipore, Billerica, MA).

Instruments and Characterization

Mass Spectrometry. Protein bioconjugates were analyzed using an Agilent 1200 series liquid chromatograph (Agilent Technologies, Santa Clara, CA) that was connected in-line with an Agilent 6224 Time-of-Flight (TOF) LC/MS system equipped with a turbospray ion source. Samples were run with a Proswift RP-4H chromolith column (Agilent Technologies, USA). Protein mass reconstruction was performed on the charge ladder with Mass Hunter software (Agilent, USA).

UV-Vis Spectroscopic Measurements. UV-Vis spectroscopic measurements were conducted on a Thermo Scientific nanoDrop 1000 scan benchtop spectrophotometer (Thermo Scientific, Wilmington, DE). Spectra were collected in Proteins and Labels mode from 200-800nm, while direct protein concentrations were calculated using Protein A280.

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed in a NuPAGE apparatus (Life Technologies, Hercules, CA), using a 4-12% precast linear gradient polyacrylamide gel (Bio-Rad, Hercules, CA). All protein electrophoresis samples were heated for 10 min at 75 °C in the presence or absence as noted of 1,4-dithiothreitol (DTT) to ensure reduction of disulfide bonds. Gels were run for 34 min at 200 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad). Gels were visualized on Gel-Doc EZ Imager (Bio-Rad).

Protein production
VHH7 and IgG Fc were obtained from the Roulet and Barton labs as a pre-isolated proetins.

Production of Cas9: Cas9 used in mouse studies was obtained from the UC Berkeley Macro Lab in endotoxin-free form.

For test reactions Cas9 plasmids were generously provided by Dr. Brett Staahl. Plasmids were transfected into BL21 *E. coli* cells, as per standard protocol and plated. After overnight incubation the bacteria were inoculated into 1 L of Terrific Broth media with 0.1 g/L ampicillin. Cells were cultured at 37 °C until an OD 600 of 0.8, then reduced to 18 °C for an hour before induction with 0.1 g/L Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested via centrifugation and lysed using sonication (Fischer Scientific FB120) using a 2 s on/ 5 s off cycle at 70% for 10 min total of sonication.

After lysis, cells debris was pelleted, and the supernatant was passed through a Protino Ni-NTA 5 mL column to isolate the his6-tagged Cas9-MBP fusion. The Cas9 fusion was incubated overnight at 4 °C on ice with his6 labeled TEV protease at 20:1 mass ratio, then sent through a nickel column prepared with buffered media, where the MBP and TEV were retained on the column, along with any un-cleaved Cas9-MBP.

The Cas9 was concentrated in a 100,000 MWCO spin concentrator down to a final volume of 0.5 mL, then injected onto a Superdex 200 10/300 GL size exclusion column and eluted with 2 column volumes. Out of the 4 primary peaks detected (Supplementary Figure 5) and analyzed via ESI-TOF, only the second peak was found to contain Cas9. Lastly, purified Cas9 was concentrated in a 10,000 MWCO spin concentrator.

Amino acid sequences for Cas9 with 2 NLS:

SNATCDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRISYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING
IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV
EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDAYPYDVPDYASLGSGSPKKKRKVEDPKKKRKVD

Expression and purification of Y200C-sfGFP: After sequence verification, a 10 mL overnight culture of BL21 (DE3) Codon Plus RIL *E. coli* in TB medium with 50 µg/mL kanamycin was inoculated into a 2.8 L trident Erlenmeyer flask with 1 L of sterile Terrific Broth (TB) medium and kanamycin. The cultures were grown at 37 °C with 220 rpm shaking until O.D. at 600 nm reached 0.6 – 0.8. To induce expression, IPTG was added from a 1.0 M stock to a final concentration of 1.0 mM. The cultures were incubated at 37 °C for 12-16 hours. Cells were then incubated for another 12-16 hours at 37 ºC and pelleted. After freezing at -80 ºC, cell pellets were resuspended in 15 mL of an equilibration buffer (20 mM sodium phosphate, 2 M NaCl, 20 mM imidazole at pH = 7.4) and then lysed via sonication for 30 min at 60% amplitude. Cell lysate was spun down at 14,000 rpm for 30 min, then the supernatant decanted before loading on to a HisTrap FF Crude column. After binding, the bound protein was washed with 4 portions of two resin bed volumes of wash buffer (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole at pH = 7.4) and subsequently eluted with four resin bed volumes of elution buffer (20 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole at pH = 7.4). The purified sfGFP variant was then spun concentrated into 50 mM phosphate buffer at pH 7.2 using 10k MWCO filters. Purified sfGFP samples were flash frozen and stored at -20 ºC until use. Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7.0 using 10k MWCO spin filters prior to reaction.

Y200C-sfGFP-His_6-EEEEY Protein sequence:
ATGTCGAATAAATATCGTGTGCGTAAGAACGTCTTGCATCTGACGGACACAGAGAAACGCGACTTTGTT
CGCACCGTTCTTATCTTAAAAGAAAAAGGGATTTATGATCGCTACATCGCATGGCACGGAGCCGCAGGC
AAGTTCCACACGCCGCCTGGAAGCGATCGCAATGCGGCGCACATGTCGTCAGCTTTCTTACCATGGCAC

CGCGAATATTTATTGCGTTTTGACGCGACCTTCAAAGCATTAACCCCGAGGTCACACTGCCCTATTGGGA
GTGGGAAACCGATGCTCAAATGCAAGATCCTCTCAATCTCAGATTTGGTCCGCAGACTTTATGGGGGGA
AACGGTAATCCGATCAAGGATTTCATCGTGGACACGGGACCATTTGCGGCAGGGCGCTGGACCACGATT
GATGAGCAAGGAAACCCTAGTGGGGGTCTGAAACGTAATTTCGGTGCGACAAAAGAAGCGCCCACATT
GCCTACTCGTGACGATGTCTTGAACGCACTTAAAATCACGCAGTATGACACGCCGCCGTGGGATATGAC
TTCGCAGAACTCTTTCCGCAATCAGCTGGAAGGCTTTATTAACGGGCCTCAGCTTCACAATCGCGTCCAC
CGCTGGGTAGGCGGACAAATGGGTGTTGTGCCCACTGCCCCGAATGACCCCGTCTTCTTCTTGCATCATG
CTAACGTCGACCGTATCTGGGCCGTCTGGCAGATCATCCACCGCAACCAGAATTACCAACCGATGAAGA
ACGGCCCATTCGGGCAGAACTTCCGCGATCCTATGTACCCCTGGAATACAACACCCGAAGACGTTATGA
ATCATCGCAAATTAGGTTACGTTTATGACATCGAACTTCGCAAGAGTAAGCGCTCCTCTctcgagcaccacca
ccaccaccactga

Corresponding to:
MSNKYRVRKNVLHLTDTEKRDFVRTVLILKEKGIYDRYIAWHGAAGKFHTPPGSDRNAAHMSSAFLPWHRE
YLLRFERDLQSINPEVTLPYWEWETDAQMQDPSQSQIWSADFMGGNGNPIKDFIVDTGPFAAGRWTTIDE
QGNPSGGLKRNFGATKEAPTLPTRDDVLNALKITQYDTPPWDMTSQNSFRNQLEGFINGPQLHNRVHRWV
GGQMGVVPTAPNDPVFFLHHANVDRIWAVWQIIHRNQNYQPMKNGPFGQNFRDPMYPWNTTPEDVM
NHRKLGYVYDIELRKSKRSSLEHHHHHH*

IgG1 Fc sequences
YNNN-Fc
ATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAATTCGTACAACAACA
ACATATCGGCCATGGTTAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGG
GGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTC
ACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGT
GGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGC
GTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCC
CTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACC
CTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTAT
CCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCC
CGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCA
GGGGAACGTCTTCTCATGCTCCGTGATGCACGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTC
CCTGTCTCCGGGTAAATGA
Corresponding to an amino acid sequence (with Tyrosine tag in bold):
MYRMQLLSCIALSLALVTNS**YNNN**ISAMVRSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCV
VVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFL
YSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK*

Fc-NNNY
ATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAATTCGATATCGGCCAT
GGTTAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGT
CTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTG

GTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAA
TGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCC
TGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCA
TCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCC
GGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATC
GCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTC
CGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTT
CTCATGCTCCGTGATGCACGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGT
AAAAACAACAACTACTGA

Corresponding to an amino acid sequence (with Tyrosine tag in bold):
MYRMQLLSCIALSLALVTNSISAMVRSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVS
HEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKA
KGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTV
DKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK**NNNY***

VHH7 Nanobody sequence:
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCGGCCCAGCCGGCCATGGCCCAG
GTGCAGCTGCAGGAGTCAGGGGGAGGATTGGTGCAGGCTGGGGACTCTCTGAGACTCTCCTGCGCAGC
CTCTGGACGCACCTTCAGTCGCGGTGTAATGGGCTGGTTCCGCCGGGCTCCAGGGAAGGAGCGTGAGT
TTGTAGCAATCTTTAGCGGGAGTAGCTGGAGTGGTCGTAGTACATACTATTCAGACTCCGTAAAGGGCC
GATTCACCATCTCCAGAGACAACGCCAAGAACACGGTGTATCTGCAAATGAACGGCCTGAAACCTGAGG
ACACGGCCGTTTATTACTGTGCAGCGGGATATCCGGAGGCGTATAGCGCCTATGGTCGGGAGAGTACAT
ATGACTACTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCAGGAGGACTGCCGGAAACCGGCGGCCAC
CACCATCACCATCACGGCGGCTATTAA

Corresponding to an amino acid sequence (with Tyrosine tag in bold):
MKYLLPTAAAGLLLLAAQPAMAQVQLQESGGGLVQAGDSLRLSCAASGRTFSRGVMGWFRRAPGKEREFV
AIFSGSSWSGRSTYYSDSVKGRFTISRDNAKNTVYLQMNGLKPEDTAVYYCAAGYPEAYSAYGRESTYDYWG
QGTQVTVSSGGLPETGGHHHHHH**GGY***

Cas9 to IgG1 Fc domain coupling protocol

In all trials unless otherwise noted Cas9 was held at 10 µM with Fc concentration varying from 1x to 100x cas9 concentration. Reactions typically ran for 1 h on ice, in a buffer containing 20 mM HEPES, 300 mM NaCl and 200 mM Trehalose at pH 6.5.

Tyrosinase resin protocol

In order to retain maximum binding capacity of the pre-activated medium prior to the coupling step, use cold (0 °C to 4 °C) solutions. The time interval for all washing steps must be minimized; *prepare all required solutions prior to coupling the protein.*

1. Make tyrosinase solution at 2 mg/mL in 50 mM PB, pH 8.0.
2. For **X** µL of NHS-activated Sepharose 4 Fast Flow, wash sepharose resin 2**X** µL of cold 1 mM HCl and then 9 x **X** µL immediately before use. (Use small SPPS columns for washing.)

3. Drain last portion of HCl, and add **X** µL of the appropriate tyrosinase solution. Check pH to ensure it is around pH 8.0. (Can skip checking pH – found it was around 7 the first time and have never needed to adjust it.
4. Wrap in parafilm and rotate at 4 °C overnight.
5. Drain + keep tyrosinase coupling solutions (keep coupling solution only if you want to quantify how much protein is on resin). Wash 4 x with 100 µL of 50 mM PB, pH 8.0 (again keep solution if you want to quantify).
6. Block unreacted NHS by incubating resin with 2**X** µL of 0.5 M ethanolamine, 0.5 M NaCl, pH 8.3 for 2 hours at RT.
7. Drain block solution and wash resin by alternating between 0.1 M Tris-HCl buffer pH 8 to 9 and 0.1 M acetate buffer, 0.5 M NaCl pH 4 to 5. Wash with 3 x **X** µL Tris buffer followed by 3 x **X** µL acetate buffer. Repeat this cycle 4 times. Wash with 50 mM PB, pH 8.0 x 1.
8. Re-suspend final resin in 50 mM PB, pH 8.0.

Using the above protocol, were **X**= 500 µL, 500 µL of resin was prepared and used in all reactions.

Endotoxin removal

The UC Berkeley MacroLab has shown excellent results using the PallCorp Mustang E-Membrane Acrodisc©. Cas9 obtained from the MacroLab was purified using this filter system and shows roughly 20% loss per filter cycle. Unfortunately, we found that the VHH7 nanobody and nanobody-Cas9 fusions bound tightly to the Acrodisc and were resistant to elution. Pierce endotoxin removal resin was found to release the nanobody samples and was used for all samples produced. Pierce resin was used according to manufacturer's protocols where the sample buffer used contained 20 mM HEPES, 300 mM NaCl and 200 mM Trehalose at pH 6.5. Samples were incubated for 2 hours on resin before final elution to maximize endotoxin removal. Nanobodies were filtered twice, once upon arrival, and again before sample delivery.

Cas9-nanobody Tyrosinase Coupling Protocol

Endotoxin free Cas9 was obtained from the UC Berkeley MacroLab. Nanobodies were obtained from the Barton/Raulet labs and endotoxins removed as above prior to reaction. Cas9 was diluted to 5.15 mg/mL or 40 µM, and nanobodies were diluted to 1.5 mg/mL or 100 µM in a contained 20 mM HEPES, 300 mM NaCl and 200 mM Trehalose at pH 6.5 buffer. For the reaction 300 µL of Cas9 was combined with 1050 µL of VHH7 nanobody and 135 µL tyrosinase beads. Samples were rotated end over end for 2 h prior to quenching by addition of 100 µL 25 mM tropolone with 10 mM dithiothreitol (DTT). Samples were then passed through a 0.2 µM sterile filter to remove tyrosinase beads and applied to endotoxin filters as above, before dilution with sterile buffer to 1 mg/mL final concentration by UV 280 and a final filtration through 0.2 µM sterile filter and flash freezing. Control samples were prepared by combining 500 uL Cas9 with 100 uL of VHH7 nanobodies to replicate a 1:2 molar ratio of Cas9:nanobody and treated identically to above, as were Cas9 and nanobody alone controls.

Endotoxin Assay

Pierce™ Chromogenic Endotoxin Quant Kits (Catalog no. A39552) were obtained from Thermo Fischer and used per manufacturers instructions, with samples diluted 2-10x before quantiation.

Mouse Treatment protocol (Conducted by Gabrielle Reiner in the Barton lab)
We treated mice systemically (retro-orbital injection) on day 0 with 100 μg of VHH7-Cas9 reagent or controls (buffer, VHH7 nanobody alone, Cas9 alone). Then we immunized the mice with Cas9 protein + adjuvant on day 7 and 14. On day 21, spleens were harvested from the treated animals as well as a naive, untreated mouse. The splenocytes were then left in unstimulated (media, open circles) or were stimulated with Cas9 protein (blue squares). T cells that recognize Cas9 will make IFNγ upon stimulation which was assayed via ELISPOT or intracellular cytokine staining. All mouse work was completed under Berkeley Animal use Protocol AUP-2017-03-9679-1.

4.6 References

1. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat. Biotechnol.* **34**, 328–333 (2016).
2. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
3. Mehta, A. & Merkel, O. M. Immunogenicity of Cas9 Protein. *J. Pharm. Sci.* **109**, 62–67 (2020).
4. Charlesworth, C. T. *et al.* Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. *bioRxiv* 243345 (2018). doi:10.1101/243345
5. Crudele, J. M. & Chamberlain, J. S. Cas9 immunity creates challenges for CRISPR gene editing therapies. *Nat. Commun.* **9**, 9–11 (2018).
6. Burstein, D. *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
7. Lyu, P., Javidi-Parsijani, P., Atala, A. & Lu, B. Delivering Cas9/sgRNA ribonucleoprotein (RNP) by lentiviral capsid-based bionanoparticles for efficient 'hit-and-run' genome editing. *Nucleic Acids Res.* **47**, e99 (2019).
8. Li, P. Y., Fan, Z. & Cheng, H. Cell Membrane Bioconjugation and Membrane-Derived Nanomaterials for Immunotherapy. *Bioconjug. Chem.* **29**, 624–634 (2018).
9. Naso, F. *et al.* Alpha-Gal Inactivated Heart Valve Bioprostheses Exhibit an Anti-Calcification Propensity Similar to Knockout Tissues. *Tissue Eng. - Part A* **23**, 1181–1195 (2017).
10. Elias, N., Benedict Cosimi, A. & Kawai, T. Clinical trials for induction of renal allograft tolerance. *Current Opinion in Organ Transplantation* **20**, 406–411 (2015).
11. Schep, S. J., Schutgens, R. E. G., Fischer, K. & Boes, M. L. Review of immune tolerance induction in hemophilia A. *Blood Reviews* **32**, 326–338 (2018).

12. Yin, Z. *et al.* Boosting immunity to small tumor-associated carbohydrates with bacteriophage Qβ capsids. *ACS Chem. Biol.* **8**, 1253–1262 (2013).

13. Thoms, F. *et al.* Immunization of cats to induce neutralizing antibodies against Fel d 1, the major feline allergen in human subjects. *J. Allergy Clin. Immunol.* **144**, 193–203 (2019).

14. Calvani, M., Ospedaliera, A., Forlanini, S. C., Giorgio, V. & Sopo, S. M. *Specific oral tolerance induction for food. A systematic review*. (2010).

15. Blumberg, R. S. & Lillicrap, D. Tolerogenic properties of the Fc portion of IgG and its relevance to the treatment and management of hemophilia. *Blood* **131**, 2205–2214 (2018).

16. Krishnamoorthy, S. *et al.* Recombinant factor VIII Fc (rFVIIIFc) fusion protein reduces immunogenicity and induces tolerance in hemophilia A mice. *Cell. Immunol.* **301**, 30–39 (2016).

17. Fang, T. *et al.* Structurally Defined αMHC-II Nanobody-Drug Conjugates: A Therapeutic and Imaging System for B-Cell Lymphoma. *Angew. Chemie - Int. Ed.* **55**, 2416–2420 (2016).

18. Fang, T. *et al.* Targeted antigen delivery by an anti-class II MHC VHH elicits focused αmUC1(Tn) immunity. *Chem. Sci.* **8**, 5591–5597 (2017).

19. Lobba, M. J. *et al.* Site-Specific Bioconjugation through Enzyme-Catalyzed Tyrosine-Cysteine Bond Formation. *ACS Cent. Sci.* **6**, 1564–1571 (2020).

20. Maza, J. C. *et al.* Enzymatic Modification of N-Terminal Proline Residues Using Phenol Derivatives. *J. Am. Chem. Soc.* **141**, 3885–3892 (2019).

21. Marmelstein, A. M. *et al.* Tyrosinase-Mediated Oxidative Coupling of Tyrosine Tags on Peptides and Proteins. *J. Am. Chem. Soc.* **142**, 5078–5086 (2020).

22. Kim, S. *et al.* CRISPR RNAs trigger innate immune responses in human cells. *Genome Res.* **28**, 367–373 (2018).

23. Manini, P., Napolitano, A., Westerhof, W., Riley, P. A. & D'Ischia, M. A reactive ortho-quinone generated by tyrosinase-catalyzed oxidation of the skin depigmenting agent monobenzone: Self-coupling and thiol-conjugation reactions and possible implications for melanocyte toxicity. *Chem. Res. Toxicol.* **22**, 1398–1405 (2009).

24. De Lucia, M., Panzella, L., Pezzella, A., Napolitano, A. & D'Ischia, M. Oxidative chemistry of the natural antioxidant hydroxytyrosol: Hydrogen peroxide-dependent hydroxylation and hydroxyquinone/o-quinone coupling pathways. *Tetrahedron* **62**, 1273–1278 (2006).

25. Bruins, J. J., Blanco-Ania, D., Van Der Doef, V., Van Delft, F. L. & Albada, B. Orthogonal, dual protein labelling by tandem cycloaddition of strained alkenes and alkynes to: Ortho -quinones and azides. *Chem. Commun.* **54**, 7338–7341 (2018).

26. Bruins, J. J. *et al.* Inducible, Site-Specific Protein Labeling by Tyrosine Oxidation-Strain-Promoted (4 + 2) Cycloaddition. *Bioconjug. Chem.* **28**, 1189–1193 (2017).

27. Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, 249–254 (2019).

Chapter 5. Identification and evolution of prokaryote tyrosinases for selective tyrosine oxidation

Abstract:
The most commonly available form of tyrosinase is extracted from the mushroom *Agaricus bisporous* and has been used to great effect in protein coupling reactions. However, it is limited in application and modification since it is obtained as an extract. We screened a several tyrosinase candidates and identified a bacterial tyrosinase from the literature (bmTYR) which shows ideal behavior and stable recombinant expression. We find that bmTYR is not only capable of activating proteins previously untouched by abTYR, but also shows a broader range of substrate tolerance based on peptide charge. We use the expanded substrate preference as a basis for evolution and show that single point mutants of bmTYR are capable of substrate discrimination and show potential for the creation of protein oligomers.

This work was conducted in collaboration with Casey Mogilevsky and would not have been possible without his efforts. bmTYR validation on protein L was conducted by Alan Marmelstein. Portions of this text appeared previously in a paper and are reproduced herein with permission from Alan Marmelstein, Casey Mogilevksy, Johnathan Maza, Daniel Brauer, and Matthew Francis. Tyrosinase-Mediated Oxidative Coupling of Tyrosine Tags on Peptides and Proteins. *J. Am. Chem. Soc. 142*, 5078–5086 (2020).

5.1 Tyrosinases for protein modification

Tyrosinase enzymes have emerged over the past several years as exciting tools for native amino acid activation and protein-protein fusions. Tyrosinases are present in most forms of life as they play an important role in melanin synthesis which is vital for protection from UV damage to DNA.[1,2] Representing the first step of the melanin synthesis pathway tyrosinase oxidizes free tyrosine into an *ortho*-quinone (*o*-quinone) which then cyclizes with the free amine, producing eumelanin.[3] Due to their central role in melanin production, tyrosinases have been the subject of much study for the field of cosmetics and whitening products abound with tyrosinase inhibitors to prevent tanning.[4,5]

Starting in the 1980s it was appreciated that tyrosinases could act on tyrosine residues within a peptide backbone.[6] This observation sparked interest in tyrosinase applications for protein functionalization. Several works have reported the use of tyrosinases for activation of tyrosines in protein backbones utilizing a variety of functional groups.[7,8]

Since Maza et al. showed the potential to utilize tyrosinase for protein modification,[9] the Francis lab has published a number of papers showing the utility of tyrosinase from *Agaricus bisporus* (abTYR) in tyrosine activation.[9–11] Due to its applications in food homogenization abTYR is available at the industrial scale as an extract from mushrooms.[12,13] There are several protein homologues of the protein expressed by *Agaricus bisporus* believed to be responsible for tyrosinase activity and commercially available extracts are comprised of a mixture of these proteins.[5,13,14] Furthermore, since abTYR proteins are not amenable to recombinant expression, it was impossible to evolve or modify the enzyme. To remedy this we sought out a number of tyrosinases that had been identified in the literature which could be recombinantly expressed and possessed the phenol oxidase behavior required for our goal of protein coupling.

We identified tyrosinases from *Bacillus megaterium* (bmTYR) and *Candidatus Nitrosopumilus koreensis* (cnkTYR) as optimal candidates for recombinant expression.[15,16] BmTYR was selected due to a strong literature presence including several studies which showed that the enzyme tolerated point mutations while retaining high activity on free tyrosine.[15] The published structures for bmTYR were another benefit for protein engineering approaches as they would allow us the rational design of point mutants and facilitate computational analysis of the protein.[15] cnkTyr was recently shown to have excellent monophenol monooxygenase activity and to tolerate low temperatures without significant loss in catalytic efficiency.[16] Given our focus on Cas9 modification that requires reaction temperatures of 4 °C, the low temperature tolerance of cnkTyr was an exciting candidate, despite the lack of available crystal structure.

5.2 Production and screening of prokaryote tyrosinases

The catalyst for these explorations was the realization that abTYR was incapable of activating phenols attached to a single stranded DNA (ssDNA) substrate. Our first experiment with bmTYR showed significant coupling between GFP and the phenol-DNA (Figure 1A). From

the results of the DNA coupling trials we determined that some aspect of the substrate was preventing activation by amTYR but was tolerated by bmTYR (Figure 1B). We sought to examine the substrate preference of the new tyrosinases, specifically their charge and steric tolerance. A peptide library with tyrosines at the C-terminus was ordered from GeneScript, shown in Supplemental Table 4.1, which contained a variety of charged and neutral peptides to assess charge preference and steric tolerance of the tyrosinases. Peptides were coupled to Y200C GFP in 30 min reactions and analyzed via ESI-TOF MS to determine conversion to modified protein for each peptide. Comparing abTYR to the two prokaryotic tyrosinases it is readily apparent that abTYR is significantly more sensitive to negative charge. In fact, abTYR shows no activity towards any peptide with more than a single acidic residue. Also of note is the fact that bmTYR shows high activity towards all peptide substrates, with a drop in activity only in the case of the RRRRY peptide. These results supported the earlier challenge activating ssDNA substrates using abTYR, and suggest that bmTYR may serve as a highly permissive tyrosine activator.

One odd result observed in our study was that cnkTyr samples showed excess oxidation which was present without the tyrosine containing peptides. This indicates that cnkTyr might be capable of activating tyrosines along the peptide backbone of GFP. Unfortunately we were unable to identify a consistently modified residue, indicating potential non-specific activity. The crystal structures of the three tyrosinases were examined to determine the origin of their selectivity, with cnkTyr represented by a homology model.[16] abTYR is found as a hetero-tetramer and shows significant steric occlusion of the active site,[17] while bmTYR and cnkTyr show a much more open active site, giving rise to their more permissive activities. Due to this complication and the lack of off target activity of bmTYR, cnkTyr was dropped from further experiments. However, further exploration or engineering of cnkTyr could yield a more specific phenol monooxygenase capable of acting with higher precision.

The smaller size and more exposed active site of bmTYR were utilized in recent publication by Marmelstein, which showed that protein L derivates with tyrosine tags could not be activated by abTYR despite significant engineering of the protein C-terminus (Supplemental Figure 1).[10] Using bmTYR we were able to mediate efficient coupling between protein L and small molecule cargo, showing the utility of the smaller tyrosinase.[10] It is probable that difference in size was the primary factor in these experiments as a library of tyrosine tags on protein L were tried and all were efficiently modified by bmTYR.[10] The size difference is readily apparent when the bmTYR monomer is compared to the tyrosinase active complex (Figure 2A). In further experiments we have shown that bmTYR efficiently mediates coupling between GFP and luciferase, indicating that this is a versatile tool capable of equivalent activity to abTYR. This is of significant benefit to the field as bmTYR is readily produced at 50-100 mg per L of culture, whereas abTYR must be purchased at $150 for 18 mg. The other benefit of recombinant expression is the ready mutation of the enzyme, allowing for novel functionality to be explored.

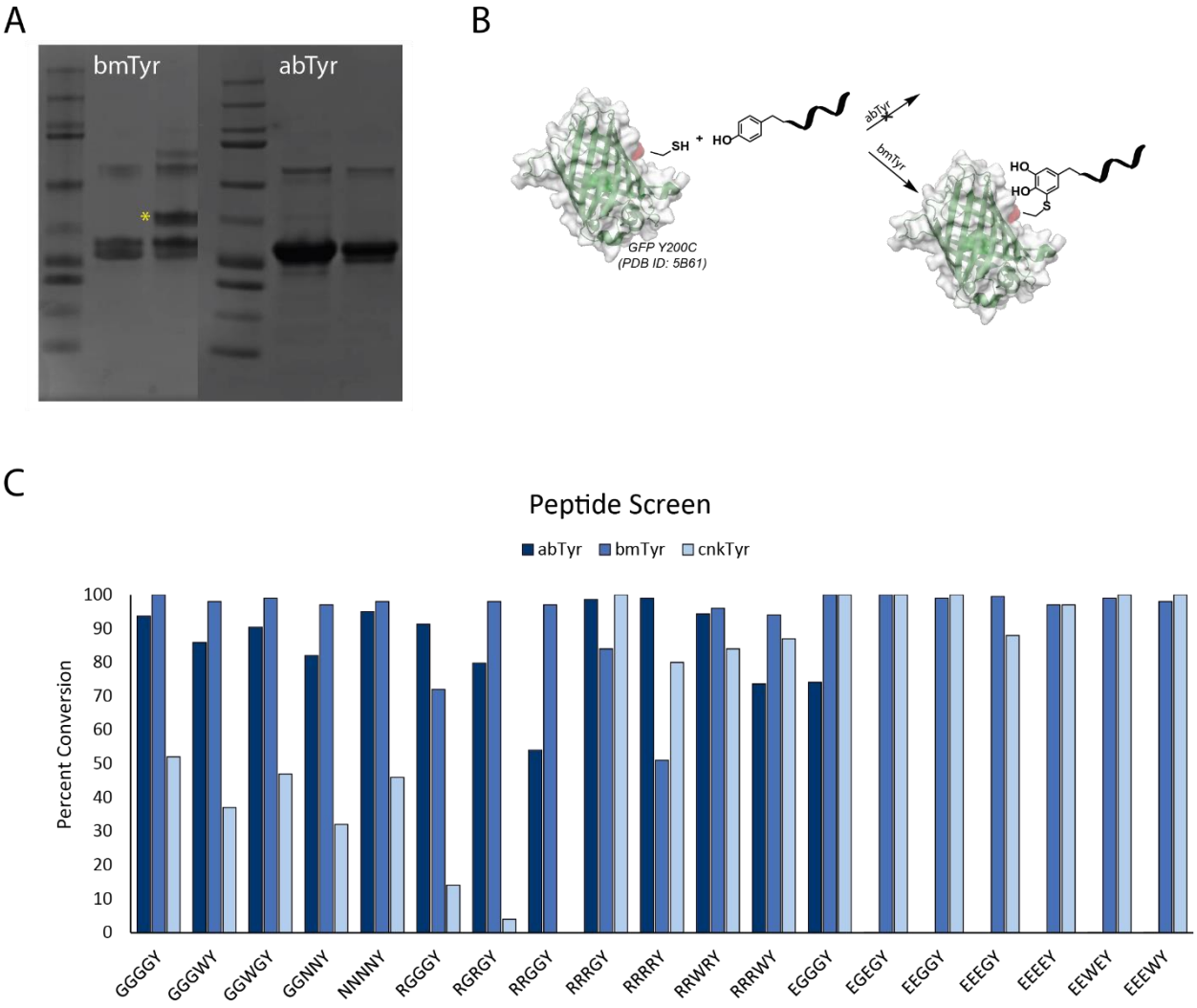


Figure 1. Identification of charge preferences in tyrosinase samples. **A**. SDS-PAGE gel showing modification of sfGFP Y128C with phenol modified ssDNA sample. Each gel shows a ladder, followed by the GFP control, then the GFP-DNA reaction with the indicated tyrosinase. Product band marked with a yellow asterisk (*). **B**. Reaction scheme showing reaction depicted in **A**. **C**. Peptide reaction screen showing percent conversion of Y200C sfGFP with the indicated peptide. All reactions carried out at 50 μM Y200C sfGFP, 250 μM peptide, and 12 U/L equivalent of either abTyr (dark blue), bmTyr (blue), or cnkTyr (grey) in 50 mM pH 6.5 phosphate buffer for 30 min before quench with 4 mM tropolone and analysis via ESI-TOF-MS. Deconvoluted data were analyzed on chartograph.com with percent conversion calculated as percentage of peak area for the modified vs unmodified samples.

5.3 Evaluation of bmTYR mutants

There exists a significant literature on the impact of point mutants on the tyrosinase activity of bmTYR.[15] We selected R209H, D55K, and V218G as potential mutants of interest due to their impacts on diphenolase activity, charge, and copper binding respectively.[15,18] After initial screening it appeared that R209H decreased the charge preference of bmTYR, while D55K appeared to increase the preference for anionic substrates as shown by the higher rate of modification with the EEEEY peptide than in other mutants (Figure 2B, structures with residues highlighted in Figure 2D). To assess the degree of substrate preference we ran competition

assays containing both EEEEY and RRRRY peptide with low tyrosinase concentrations to minimize substrate saturation. In these assays abTYR showed no activity towards EEEEY, while R209H showed similar preference for the two. Interestingly, wild type bmTYR showed a 90% preference for EEEEY, while the D55K mutant pushed this to 95% (Figure 2C). These results indicated that increasing the net positive charge on bmTYR might yield a tyrosinase incapable of activating cationic substrates in a similar manner to abTYR with anionic targets.

Generating orthogonal tyrosinases is of great interest to the field as it would enable the facile creation of protein chains by coupling anionic and cationic tagged proteins sequentially in a manner similar to peptide synthesis but using entire proteins as building blocks. Controlled production of multi-protein polymers is of use for protein based materials, biomedical technologies, and enzyme engineering.

As a first step towards this goal we generated two double mutant versions of bmTYR, D55K, V218K and D55K, E141K. The enzymes were successfully produced but showed minimal change to substrate preference in our peptide competition assay (Figure 2C). These slightly disappointing results indicated that we would need a more rigorous approach to generating charged mutants of bmTYR.

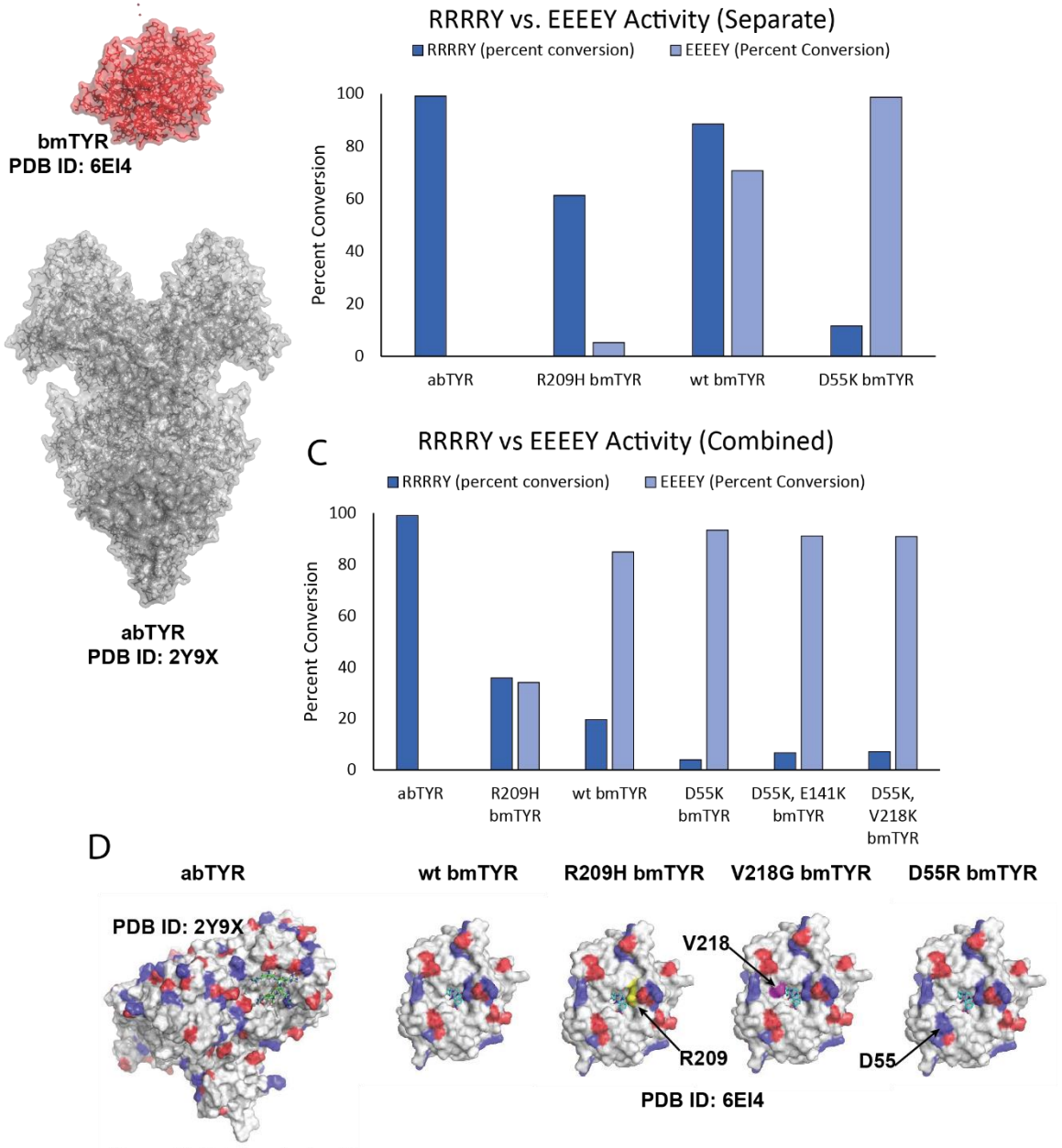
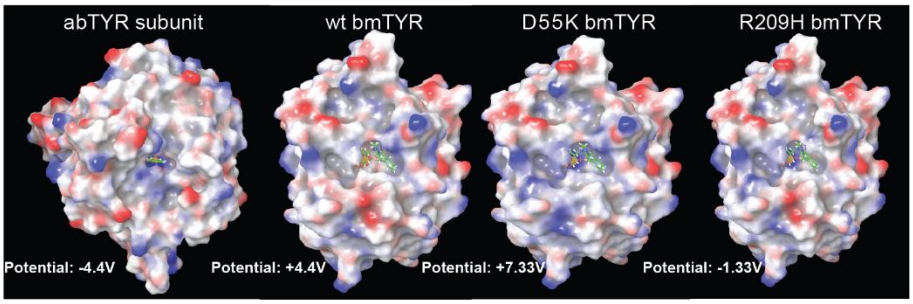


Figure 2. Point mutants of bmTyr and their impact on substrate preference. **A**. Size comparison between the abTYR hetero-octamer complex and the abTYR monomer to illustrate the size difference between the active form of the two enzymes. **B**. Comparison of reactivity of abTYR vs single mutants of bmTYR on peptides EEEEY and RRRRY. Reaction carried out 30 min at RT. in 50 mM pH 6.5 phosphate buffer, 12 U/L tyrosinase, 50 μM Y200C sfGFP and 250 μM peptide. Reaction quenched with 4 mM tropolone. **C**. Competition reaction between RRRRY and EEEEY peptides to assess substrate preference of tyrosinase mutants. Reactions carried out for 30 min at RT in 50 mM pH 6.5 phosphate buffer. Tyrosinase concentration reduced to 3 U/L to reduce substrate saturation, with 50 μM Y200C sfGFP and 250 μM peptide. Reaction quenched with 4 mM Tropolone. All samples analyzed via ESI-TOF-MS and quantified using chartograph.com for peak integration. D. Structural comparison of abTYR vs bmTYR with locations of point mutants highlighted. R209 shown in yellow, V218 in magenta, and D55 in blue. Structures show Asp and Glu residues in red and Lys and Arg in blue.

To assess the utility of modulating the tyrosinase charge abTYR, bmTYR, and the bmTYR mutants were modeled in Maestro and used the charge mapping tool from MacroModel to calculate the surface potential for each of the proteins compared to a point charge placed 10

Ångstroms from the active site copper atoms (Figure 3A). We found that abTYR monomers had a net potential of -4.4 due to high concentrations of acidic residues on the surface and corresponds to an isoelectric point (pI) of only 4.41. When the active hetero-tetramer of abTYR is considered the net potential drops to -15.7.[17] In contrast, bmTYR was shown to have a net potential of 4.4 and a pI of 8.66, correlating well with the observed charge preference. Modeling the point mutants of bmTYR showed that R209H had a potential of 3.0, a significant drop compared to the wild type. D55K corresponded to a potential of 7.33, with the double mutants showing variable change in net change compared to D55K (Figure 3B). Given the magnitude of the difference in potential between abTYR and bmTYR we hypothesized that a bmTYR mutant with a surface potential of +15 would be sufficient to confer similar levels of selectivity against cationic substrates as seen with abTYR and anionic substrates. A complete listing of all tyrosinase mutants evaluated is available in Supplementary Table 4.5.

A



B

Protein	Isoelectric point (pI)	Net Charge	Solvated Potential (V)	Δ Potential (V)
abTYR (monomer)	4.41	-13	-4.4	NA
abTYR (complex)	4.41	-39	-15.7	NA
bmTYR	8.66	7	4.4	NA
R209H	8.02	6	3	-1.33
D55K	9.13	9	7.33	2.9
D55R	9.15	9	7.49	3.1
D55K V218K	9.26	10	5.65	1.2
D55K E141K	9.37	11	9.25	4.8

Figure 3. Calculations on tyrosinase mutants. **A**. Surface potential maps of abTYR and several bmTYR mutants with negative charge in red and positive charge in blue. **B**. Table showing results of charge potential calculations for the various tyrosinase mutants. For each tyrosinase a point charge was placed 10 Ångstroms from the copper atoms within the active site and potential charge was calculated usig the Macromodel module of Maestro developed by Schroedinger. Δ potential calculated as change in potential between the wild type and the listed mutant. pI and net charge calculated using the protparam tool at web.expasy.org

The STRIDE algorithm was used to map 15 amino acid mutations that together yielded a protein with a net +30 charge. This sequence was ordered and transformed into cells but unfortunately did not yield stable protein. Future experiments will focus on combining one to two of the mutations with the D55R mutant already shown to possess greater charge selectivity to hopefully create bmTYR with no activity towards cationic targets.

We were curious to examine the rate constants of these reactions to understand the origin of the charge preference observed with abTYR and the mutants of bmTYR. Previous work has shown that amine addition into the *ortho*-quinone ring produces a chromophore with strong absorbance at 475 nm. This has been used with free tyrosine to track the rate constants of many previously studied tyrosinases and we developed a similar assay using proline as an amine to produce an analogous chromophore. Trials were successful in the case of the GGGGY peptide, but we quickly discovered that the RRRRY and EEEEY peptides showed spectral properties that differed from the expected values. Even after controlling for pH we were unable to replicate similar values, indicating incompatability with our approach. We next ordered a series of N-terminal peptides YGGGG, YRRRR, YEEEE and YNNNN in the hopes that the N-terminal tyrosine could self-cyclize, producing the desired dopachrome-like chromophore. Unfortunately the YRRRR peptide formed an opaque solution when exposed to tyrosinase, and the other peptides displayed divergent spectral properties. At the time of this writing we are unsure of the cause of these issues and will continue further investigation.

5.4 Protein trimers through tyrosinase selectivity

Although the bmTYR mutants to date were not fully prevented from acting on cationic substrates, the range in selectivity between abTYR and bmTYR proved a tantalizing target. The cysteine containing Y200C sfGFP was modified with an EEEEY C-terminal extension to evaluate whether abTYR was incapable of activating an anionic tag when attached to a full protein (Figure 4A, B). After purification, the Y182C sfGFP-EEEEY (GFP-E4Y) was modified with a biotin-maleimide, then exposed to abTYR for 30 min prior to quenching and analysis. No oxidation was observed when analyzed via ESI-TOF MS, indicating that the EEEEY tag was not susceptible to abTYR activation (Figure 4C). We then exposed the GFP-E4Y to a phenol-biotin in the presence of abTYR and showed full conversion to the singly-modified GFP-E4Y-biotin without sign of further oxidation (Figure 4D). After purification using a 10 kDa MWCO spin concentrator, GFP-E4Y-biotin was combined with Y200C sfGFP and 12 U/L wt bmTYR for 30 min and analyzed via ESI-TOF-MS. While some higher order aggregates appear due to over-reactivity, the data show over 70% conversion to the trimer composed of two GFP proteins and a biotin molecule, making this one of the first techniques capable of creating 3-component fusions without the need for solid support tethering or removable protective groups (Figure 4E).

To evaluate the potential of the alternating tyrosinase reaction for the creation of protein trimers GFP-E4Y was combined with a tyrosine tagged nano-luciferase and 12 U/L abTYR to activate the luciferase tag. The resulting GFP- E4Y-nano-luciferase solution was purified by a 30 kDa MWCO spin concentrator prior to subsequent reaction with Y200C sfGFP and 12 U/L bmTYR, resulting in the GFP-GFP-luciferase trimer. We again saw almost 80% conversion from the dimer to the final product, indicating that bmTYR is capable of catalyzing tyrosinase oxidation on anionic tags even in the presence of multi-protein complexes.

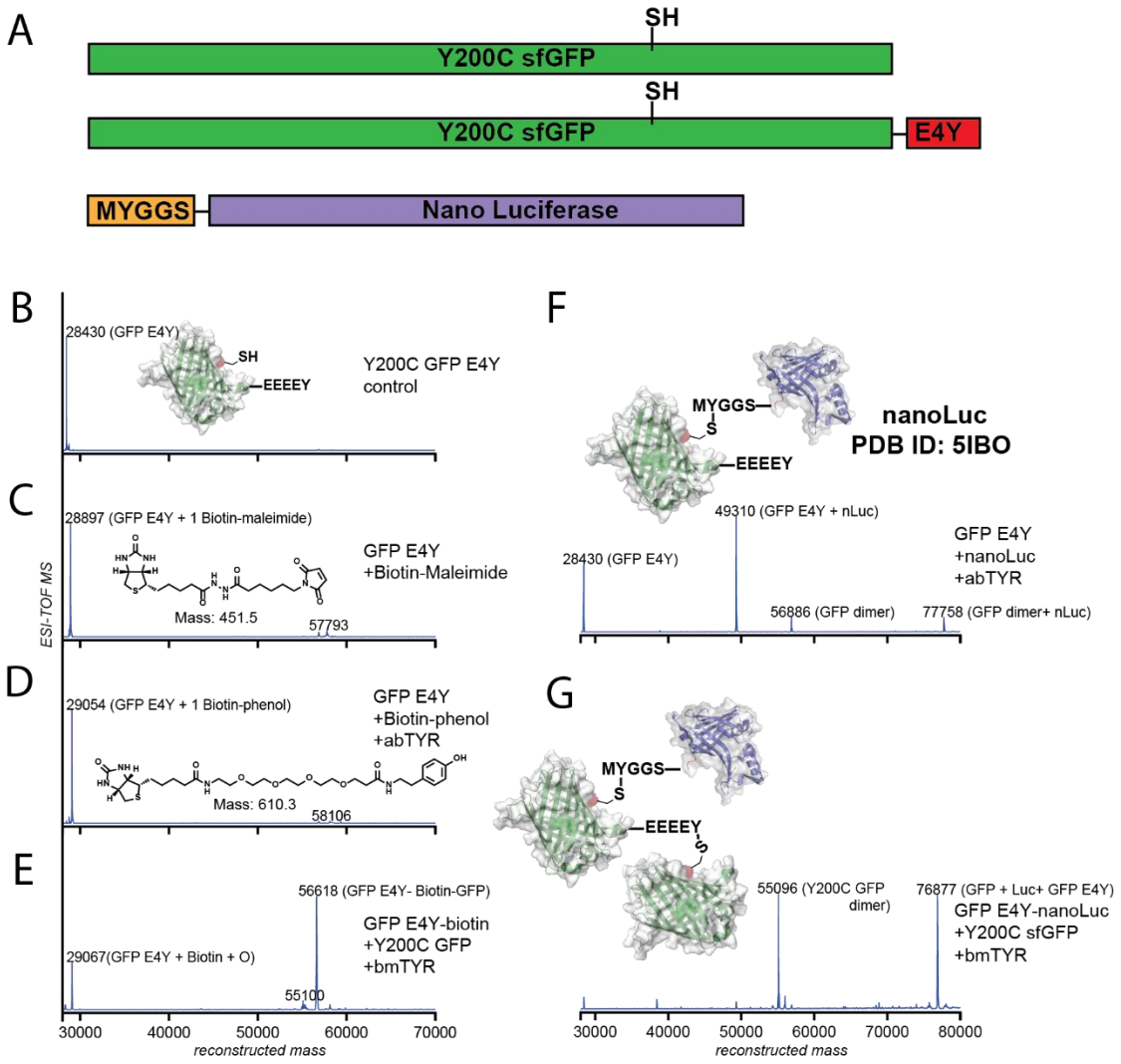


Figure 4. Use of tyrosinase charge selectivity for synthesis of ternary protein complexes. **A**. Protein maps for Y200C sfGFP (Y200C), Y200C with EEEEY tag (GFP E4Y) and MYGGS tagged nano Luciferase (nanoLuc). **B**. Biotin modification of GFP E4Y. Control shows mass of GFP E4Y (28430 Da). **C**. Biotin-maleimide coupled to GFP E4Y by combining 50 μM GFP E4Y with 500 μM biotin-maleimide at pH 7.5 for 2 h. **D**. GFP E4Y was coupled to biotin-phenol by combining 50 μM GFP E4Y with 200 μM biotin-phenol with 12 U/L abTYR in 50 mM phosphate pH 6.5 for 30 min at RT. Product mass expected = 29054 Da. **E**. Reaction of 10 μM GFP-E4Y-biotin-phenol with 50 μM Y200C GFP with 12 U/L bmTYR for 30 min at RT. Expected mass for GFP+GFP+biotin = 56618. **F**. Reaction of 50 μM GFP-E4Y with 100 μM nanoLuc and 12 U/L abTYR for 30 min. GFP + nanoLuc structure shown above chart. Expected mass = 49310. Additional conversion is likely possible with higher equivalent of nanoLuc. Large GFP dimer peaks observed **G**. GFP E4Y-nanoLuc (10 μM) purified by spin down using a 30 kDa MWCO spin concentrator combined with 100 μM Y200C GFP and 12 U/L bmTYR for 30 min at RT yielded nearly complete conversion of E4Y GFP-nanoLuc to the GFP-GFP-nanoLuc trimer. Trimer structure shown above spectrum. Expected mass = 76877 Da. Large excess of Y200C GFP produced a significant dimer peak at 55096 Da.

5.5 Conclusion and Future Directions

Here we show that recombinantly expressed prokaryote tyrosinases are successful activators of terminal tyrosine residues on a variety of protein substrates. While cnkTyr showed insufficient

selectivity for our purposes, its ability to oxidize tyrosine residues within peptide backbones remains a promising result that warrants future examination and could be utilized as a basis for engineering sequence specific tyrosine activators.

The results from the bmTYR mutant screen show that bacterially derived tyrosinases possess sufficient selectivity to mediate protein-protein coupling through tyrosine-cysteine bond formation. Our success at modulating the charge preference of this enzyme indicates that it is tolerant to mutation and could be engineered towards orthogonal activity compared to abTYR. To this end, single and double mutant versions of the enzyme should be assayed to determine compatible mutations that will enable supercharging of the enzyme while retaining activity. The protein trimers created using the combination of abTYR and bmTYR in sequence show that minimal engineering is required to produce a system capable of creating higher order protein constructs.

5.6 General Procedures and Materials

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. Water (ddH_2O) used in biological procedures or as a reaction solvent was de-ionized using a NANOpure purification system (Barnstead, USA). All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Terrific broth used in Cas9 expression was purchased from Mediatech Inc. (Maryland, VA). Spin concentrators were obtained from EMD Millipore (Millipore, Billerica, MA).

Instruments and Characterization

Mass Spectrometry. Time of flight electrospray ionization mass spectrometry (TOF-ESI-MS) of proteins was performed using an Agilent 1260 series liquid chromatography (LC) system with an Agilent 6224 TOF LC-MS system (Santa Clara, CA). For mobile phases, the system utilized Milli-Q H2O with 0.1% (v/v) formic acid (solvent A) and HPLC grade ACN with 0.1% (v/v) formic acid (solvent B). Before submission to analysis, each sample was desalted and passed through a 0.22 µm cellulose acetate centrifugal filter. Samples of 2 -10 µL were injected, corresponding to 0.2 to 1.0 µg of protein. The sample was separated on a Proswift RP-4H (monolithic phenyl, 1.0 mm × 50 mm, Dionex) analytical column with a flow rate of 0.30 mL/min and a gradient of 5-100% solvent B in solvent A over 8 minutes. Agilent Mass Hunter Workstation software (Qualitative Analysis Version B.05.00, Build 5.0.519.13, ©Agilent Technologies Inc. 2011) was used to extract total ion chromatograms from raw data and the Maximum Entropy deconvolution algorithm was used to reconstruct the original analyte masses. CSV files of spectra were plotted using the open source Chartograph software (www.chartograph.com). Integration of an unmodified protein sample served as an internal standard for determining conversion.

Size Exclusion HPLC. HPLC purification of Encapsulated proteins was performed on an Agilent 1100 Series HPLC Systems (Agilent Technologies, USA) outfitted with an Agilent 1200 series automatic fraction collector. Sample analysis for all HPLC experiments was achieved with an inline diode array detector (DAD) and in-line fluorescence detector (FLD). HPLC of capsids and proteins was accomplished using a BioSep-SEC s4000 column 300 x 7.8 mm column

(Phenomenex, Torrance, CA). The samples were eluted in isocratic flow of 10 mM or 100 mM pH 6.8 phosphate buffer at 1 mL/min.

UV-Vis Spectroscopic Measurements. UV-Vis spectroscopic measurements were conducted on a Thermo Scientific nanoDrop 1000 scan benchtop spectrophotometer (Thermo Scientific, Wilmington, DE). Spectra were collected in Proteins and Labels mode from 200-800nm, while direct protein concentrations were calculated using Protein A280.

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed in a NuPAGE apparatus (Life Technologies, Hercules, CA), using a 10-20% precast linear gradient polyacrylamide gel (Bio-Rad, Hercules, CA). All protein electrophoresis samples were heated for 10 min at 95 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of disulfide bonds. Gels were run for 80 min at 120 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad). Gels were visualized on Gel-Doc EZ Imager (Bio-Rad).

For PCR reactions products were run on a 2% agarose gel in TEA buffer with 1:10000 SyberSafe DNA gel stain (ThermoFisher Scientific, Wilmington, DE). PCR samples were incubated for 5 min with 5x SyberSafe loading buffer before application to the gel. Gels were run for 2 h at 120 V to separate bands. Commercially available markers (1kb DNA Ladder, New England BioLabs, Ipswich, MA) were applied to at least one lane of each gel for assignment of apparent mass. Visualization of DNA bands was accomplished using a Gel-Doc EZ Imager (Bio-Rad).

Protein Expression and Purification

Expression and purification of Bacillus megaterium Tyrosinase:

A 10 mL overnight culture of BL21 (DE3) Codon Plus RIL *E. coli* in Terrific Broth (TB) medium with 50 µg/mL kanamycin was inoculated into a 2.8 L trident Erlenmeyer flask with 1 L of sterile TB with kanamycin. The culture was incubated at 37 °C, 220 rpm and cell density was monitored until the O.D. at 600 nm was between 0.4 and 0.6. The cells were then induced by adding IPTG to a final concentration of 0.4 mM and subsequently incubated for another 12 hours at 37 °C. The cells were then pelleted and stored at -80 °C.

The cell pellet was thawed on ice in lysis buffer (500 mM NaCl, 20 mM imidazole, 20 mM Tris-HCl pH 7.5) and PMSF (0.1 mM), and cells were lysed by sonication for 10 minutes (2 sec on, 4 sec off, 60% amplitude, with a Fisher Scientific Sonic Dismembrator). The cellular debris was removed with centrifugation (17,000 g, 15 min, 4 °C), and the supernatant was applied to a Ni-NTA gel spin column (ThermoFisher HisPur). The column was washed 3 times with 6 mL of lysis buffer (lysis buffer without PMSF), and fractions were collected. The protein was then eluted using four 6 mL portions of elution buffer (500 mM NaCl, 500 mM imidazole, 20 mM Tris-HCl, 0.02 mM $CuSO_4$, pH 7.5) for a total of 24 mL. All fractions were analyzed using SDS-PAGE, and the tyrosinase-containing fractions were collected. The buffer was exchanged for protein

storage buffer (PBS with 15% glycerol and 0.02 mM $CuSO_4$) using a spin concentrator (10 KDa MWCO, 15 mL, Amicon Ultra) and then concentrated to 150 µM. The tyrosinase solutions were stored at -80 ^oC.

Expression and purification of Y200C-sfGFP variants:
After sequence verification, a 10 mL overnight culture of BL21 (DE3) Codon Plus RIL *E. coli* in TB medium with 50 µg/mL kanamycin was inoculated into a 2.8 L trident Erlenmeyer flask with 1 L of sterile Terrific Broth (TB) medium and kanamycin. The cultures were grown at 37 °C with 220 rpm shaking until O.D. at 600 nm reached 0.6 – 0.8. To induce expression, IPTG was added from a 1.0 M stock to a final concentration of 1.0 mM. The cultures were incubated at 37 °C for 12-16 h. Cells were then incubated for another 12-16 hours at 37 °C and pelleted. After freezing at -80 °C, cell pellets were resuspended in 15 mL of an equilibration buffer (20 mM sodium phosphate, 2 M NaCl, 20 mM imidazole at pH = 7.4) and then lysed via sonication for 30 min at 60% amplitude. Cell lysate was spun down at 14,000 rpm for 30 min, then the supernatant decanted before loading on to a HisTrap FF Crude column. After binding, the bound protein was washed with 4 portions of two resin bed volumes of wash buffer (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole at pH = 7.4) and subsequently eluted with four resin bed volumes of elution buffer (20 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole at pH = 7.4). The purified sfGFP variant was then spun concentrated into 50 mM phosphate buffer at pH 7.2 using 10k MWCO filters. Purified sfGFP samples were flash frozen and stored at -20 °C until use. Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7.0 using 10k MWCO spin filters prior to reaction.

Expression and purification of paF MS2:
Glycerol stocks of *E. coli* containing plasmids for the T19*p*af mutant of MS2 were generously provided by Ioana Aanei. The T19*p*af MS2 expression was carried out in minimal media following the published protocol.[19–21] The pellets were thawed and re-suspended in 20 ml of 20 mM taurine buffer (pH 9) containing 6.5 mM DTT, 6 mM $MgCl_2$, and 10 µg/ml each of DNase and RNase. Following sonication for 10 min, the cells were spun down for 45 min at 11,000 rpm. Next, the supernatant was applied to a DEAE-Sephadex column (GE Healthcare, Little Chalfont, United Kingdom). In 20 mM pH 9 taurine buffer, MS2 eluted first from the DEAE column and was collected and precipitated using an equal volume of saturated aqueous ammonium sulfate. The protein pellets were re-suspended in 10 mM pH 7.2 phosphate buffer and applied to a Sephacryl S1000 column (GE Healthcare). The fractions containing MS2 were then collected and concentrated using Amicon Ultra 100 kDa MWCO centrifugal concentrators (Millipore). A yield of ~3 mg/L culture was obtained for T19*p*af MS2 following two rounds of purification.

MS2 *p*af sequence. **T** in bold represents the T19*p*af mutation site:
ASNFTQFVLVDNGGTGDV**T**VAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI

KVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPS

AIAANSGIY

Expression of nano luciferase (nanoLuc). The gene encoding nanoLuc was cloned from plasmids provided by Promega (pNL1.1) using primers that encoded the specified N- or C-terminal modifications as well as a golden gate cloning site. The modified product was then transferred into a modified pet28b cloning vector (pECH081) and transformed into Rosetta DE3 pLysS *E. coli*. Overnight cultures were grown at 37 °C in LB with 50 mM kanamycin, then transferred into 1 L TB media with 50 mM kanamycin. Once cultures reached an OD600 of 0.6 protein production was induced by addition of 100 mg/mL IPTG and left shaking overnight at 37 °C. After 16 h of incubation, the cultures were centrifuged at 7000 x g, and the pellets collected were then re-suspended in 50 mM phosphate, 200 mM NaCl, 20 mM imidazole, pH 7 buffer. Samples were lysed by sonication, then centrifuged at 17000 x g. The supernatant was collected and run through a HiTrap NiNTA column from GE in a miniAkta, and eluted with a 50 mM phosphate, 200 mM NaCl, 250 mM imidazole buffer at pH 7. Purified nanoLuc samples were flash frozen and stored at -20 °C until use.

Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Protein Sequences and DNA Manipulation

Protein gene blocks were ordered from Integrated DNA Technologies, codon-optimized for *E. coli* expression. Bsa1 cut sites were present at either end of the gene sequence for cloning into the pET28b golden gate entry vector.[22] This vector enabled green / white screening for colonies successfully transformed with plasmids bearing the inserted gene and provided resistance to kanamycin. PCR primers used to make bmTYR and sf-GFP variants are listed in Supporting Table 2.

DNA and Protein Sequences:

Y200C-sfGFP-His_6 Protein sequence:

MRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFA
RYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNF
NSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHCLSTQSVLSKDPNEKRD
HMVLLEFVTAAGITHGMDELYKHHHHHH*

Plasmid was generously provided by Jonathan Maza and expressed as previously reported[9]

Y200C-sfGFP-His_6-EEEEY Protein sequence:

ATGTCGAATAAATATCGTGTGCGTAAGAACGTCTTGCATCTGACGGACACAGAGAAACGCGACTTTGTT
CGCACCGTTCTTATCTTAAAAGAAAAAGGGATTTATGATCGCTACATCGCATGGCACGGAGCCGCAGGC
AAGTTCCACACGCCGCCTGGAAGCGATCGCAATGCGGCGCACATGTCGTCAGCTTTCTTACCATGGCAC
CGCGAATATTTATTGCGTTTTGACGCGACCTTCAAAGCATTAACCCCGAGGTCACACTGCCCTATTGGGA
GTGGGAAACCGATGCTCAAATGCAAGATCCTCTCAATCTCAGATTTGGTCCGCAGACTTTATGGGGGGA
AACGGTAATCCGATCAAGGATTTCATCGTGGACACGGGACCATTTGCGGCAGGGCGCTGGACCACGATT
GATGAGCAAGGAAACCCTAGTGGGGGTCTGAAACGTAATTTCGGTGCGACAAAAGAAGCGCCCACATT
GCCTACTCGTGACGATGTCTTGAACGCACTTAAAATCACGCAGTATGACACGCCGCCGTGGGATATGAC
TTCGCAGAACTCTTTCCGCAATCAGCTGGAAGGCTTTATTAACGGGCCTCAGCTTCACAATCGCGTCCAC

CGCTGGGTAGGCGGACAAATGGGTGTTGTGCCCACTGCCCCGAATGACCCCGTCTTCTTCTTGCATCATG
CTAACGTCGACCGTATCTGGGCCGTCTGGCAGATCATCCACCGCAACCAGAATTACCAACCGATGAAGA
ACGGCCCATTCGGGCAGAACTTCCGCGATCCTATGTACCCCTGGAATACAACACCCGAAGACGTTATGA
ATCATCGCAAATTAGGTTACGTTTATGACATCGAACTTCGCAAGAGTAAGCGCTCCTCTctcgagcaccacca
ccaccaccactga
Corresponding to:
MSNKYRVRKNVLHLTDTEKRDFVRTVLILKEKGIYDRYIAWHGAAGKFHTPPGSDRNAAHMSSAFLPWHRE
YLLRFERDLQSINPEVTLPYWEWETDAQMQDPSQSQIWSADFMGGNGNPIKDFIVDTGPFAAGRWTTIDE
QGNPSGGLKRNFGATKEAPTLPTRDDVLNALKITQYDTPPWDMTSQNSFRNQLEGFINGPQLHNRVHRWV
GGQMGVVPTAPNDPVFFLHHANVDRIWAVWQIIHRNQNYQPMKNGPFGQNFRDPMYPWNTTPEDVM
NHRKLGYVYDIELRKSKRSSLEHHHHHHGGSEEEEY*

B. megaterium Tyrosinase-His_6
ATGTCGAATAAATATCGTGTGCGTAAGAACGTCTTGCATCTGACGGACACAGAGAAACGCGACTTTGTT
CGCACCGTTCTTATCTTAAAAGAAAAAGGGATTTATGATCGCTACATCGCATGGCACGGAGCCGCAGGC
AAGTTCCACACGCCGCCTGGAAGCGATCGCAATGCGGCGCACATGTCGTCAGCTTTCTTACCATGGCAC
CGCGAATATTTATTGCGTTTTGACGCGACCTTCAAAGCATTAACCCCGAGGTCACACTGCCCTATTGGGA
GTGGGAAACCGATGCTCAAATGCAAGATCCTCTCAATCTCAGATTTGGTCCGCAGACTTTATGGGGGGA
AACGGTAATCCGATCAAGGATTTCATCGTGGACACGGGACCATTTGCGGCAGGGCGCTGGACCACGATT
GATGAGCAAGGAAACCCTAGTGGGGGTCTGAAACGTAATTTCGGTGCGACAAAAGAAGCGCCCACATT
GCCTACTCGTGACGATGTCTTGAACGCACTTAAAATCACGCAGTATGACACGCCGCCGTGGGATATGAC
TTCGCAGAACTCTTTCCGCAATCAGCTGGAAGGCTTTATTAACGGGCCTCAGCTTCACAATCGCGTCCAC
CGCTGGGTAGGCGGACAAATGGGTGTTGTGCCCACTGCCCCGAATGACCCCGTCTTCTTCTTGCATCATG
CTAACGTCGACCGTATCTGGGCCGTCTGGCAGATCATCCACCGCAACCAGAATTACCAACCGATGAAGA
ACGGCCCATTCGGGCAGAACTTCCGCGATCCTATGTACCCCTGGAATACAACACCCGAAGACGTTATGA
ATCATCGCAAATTAGGTTACGTTTATGACATCGAACTTCGCAAGAGTAAGCGCTCCTCTctcgagcaccacca
ccaccaccactga
Corresponding to:
MSNKYRVRKNVLHLTDTEKRDFVRTVLILKEKGIYDRYIAWHGAAGKFHTPPGSDRNAAHMSSAFLPWHRE
YLLRFERDLQSINPEVTLPYWEWETDAQMQDPSQSQIWSADFMGGNGNPIKDFIVDTGPFAAGRWTTIDE
QGNPSGGLKRNFGATKEAPTLPTRDDVLNALKITQYDTPPWDMTSQNSFRNQLEGFINGPQLHNRVHRWV
GGQMGVVPTAPNDPVFFLHHANVDRIWAVWQIIHRNQNYQPMKNGPFGQNFRDPMYPWNTTPEDVM
NHRKLGYVYDIELRKSKRSSLEHHHHHH*

Candidatus Nitrosopumilus koreensis Tyrosinase-His_6
ATGACCCGTAAAAACGCTAAAGATTTCCTTCCCGACGAGCGCCAACGCTACTGTAACGCTTTGCTGACAC
TGAAGCAGACGATCGAACCTGGGCACACGTTGAGCAAGTATGACGAATTTGTAGCTATCCACTACGGTG
TCACCCGCCGCTTACGCAACGGCATTCCAATCGGAGATGGAGCCCACTTCGTGCCGGGCTTTCTTGCTTG
GCATCGTGAGTATCTGAATCGCTTTGAGAAGGCAATTCGCACAGTGGACAGTACCTTATCACTGCCTTAC
TGGAATTGGTCTAGTGGTGATGATACAGATACGACCGAGATCTTCACAGACGATTTTATGGGACCCCCG
GGAGATCCGAACAATGGCAACAAAATTACATCCGGCTATTTCGTAGAGAATAACTGGGAGGTTCACTCT
GAGCTGGACGGCGGGAACAATGGTTCGGTTCTTGTCCGTGACTCCACACTGTTATCTTCATCAAAGTTAT
CTCAGGTCTCGGGTTACGGCGAATTGGCTATGGATGCCGTAAATGGTGATAACGACTTCGACTCATTTCT

GCCTGGCTTAGAGGGTCCACATGGTTCCATCCACATGTGGATCGGTGGACATATGACCAGTATGACCAG
TCCCAACGATCCTATTTTCTTCCTTCATCACGCCAATATTGATCGTTTGTGGAGTAAGTGGCAAGAGTTGC
ATCCTGGGCCGGAGAACTATAATCCAAACAACGACGGTTCATATGGCAGTCGCCTTAATGACCGTATGT
GGCCGTGGGACGGCTCCGAGGACACCACAACTACACGCACCGGAACCGCAACAGGGATCTCTCTGCAA
GGTTTACTTCCTACCTTTAGTGATTATGATATTGTAACACCGCGCCATGTGCTTGATAACAATTTGACGAT
TCCGCACCACCATCACCACCATTAA
Corresponding to:
MTRKNAKDFLPDERQRYCNALLTLKQTIEPGHTLSKYDEFVAIHYGVTRRLRNGIPIGDGAHFVPGFLAWHRE
YLNRFEKAIRTVDSTLSLPYWNWSSGDDTDTTEIFTDDFMGPPGDPNNGNKITSGYFVENNWEVHSELDGG
NNGSVLVRDSTLLSSSKLSQVSGYGELAMDAVNGDNDFDSFLPGLEGPHGSIHMWIGGHMTSMTSPNDPI
FFLHHANIDRLWSKWQELHPGPENYNPNNDGSYGSRLNDRMWPWDGSEDTTTTRTGTATGISLQGLLPTF
SDYDIVTPRHVLDNNLTIPHHHHHH*

MYGGS His_6 *NanoLuciferase*
ATGTATGGTGGTTCACACCATCACCATcaccatGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACA
GACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGG
GGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGT
CATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTA
CCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCC
GAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGT
AACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCT
GTTCCGAGTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGTAA
Corresponding to:
MYGGSHHHHHHVFTLEDFVGDWRQTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVII
PYEGLSGDQMGQIEKIFKVVYPVDDHHFKVILHYGTLVIDGVTPNMIDYFGRPYEGIAVFDGKKITVTGTLWN
GNKIIDERLINPDGSLLFRVTINGVTGWRLCERILA*

Supporting Tables

Protein	Modification	Unmodified Mass (Da)	Methionine Excision Mass (Da)
Y200C GFP	None	27,549.04	
	RRRRY tag	28,336.96	
	GGSEEEEY tag	28,429.86	
pAF MS2	None	13,789.00	
bmTYR	Wild Type	35,474.80	35,343.61
	R209H	35,455.75	35,324.56
	D55K	35,487.89	35,356.70
	D55R	35,515.90	35,384.71
	D55K, V218K	35,516.93	35,385.74
	D55K, E141K	35,486.94	35,355.75

Supporting Table 4.1: Calculated masses of expressed proteins (fix above masses)

Y200C GFP			pAF MS2		
Peptide Added	Single Addition Mass (Da)	Double Addition Mass (Da)	Peptide Added	Single Addition Mass (Da)	Double Addition Mass (Da)
None	27,549.04	N/A	None	13,789.00	N/A
GGGGY	27,940.42	28,331.80	GGGGY	14,180.38	14,571.76
NNGGY	28,054.52	28,560.00	NNGGY	14,294.48	14,799.96
NNNNY	28,168.63	28,788.22	NNNNY	14,408.59	15,028.18
GGGWY	28,069.58	28,590.12	GGGWY	14,309.54	14,830.08
GGWGY	28,069.58	28,590.12	GGWGY	14,309.54	14,830.08
RGGGY	28,039.55	28,530.06	RGGGY	14,279.51	14,770.02
RGRGY	28,138.69	28,728.34	RGRGY	14,378.65	14,968.30
RRGGY	28,138.69	28,728.34	RRGGY	14,378.65	14,968.30
RRRGY	28,237.83	28,926.62	RRRGY	14,477.79	15,166.58
RRRRY	28,336.96	29,124.88	RRRRY	14,576.92	15,364.84
RRRWY	28,366.99	29,184.94	RRRWY	14,606.95	15,424.90
RRWRY	28,366.99	29,184.94	RRWRY	14,606.95	15,424.90
EGGGY	28,012.48	28,475.92	EGGGY	14,252.44	14,715.88
EGEGY	28,084.55	28,620.06	EGEGY	14,324.51	14,860.02
EEGGY	28,084.55	28,620.06	EEGGY	14,324.51	14,860.02
EEEGY	28,156.61	28,764.18	EEEGY	14,396.57	15,004.14
EEEEY	28,228.67	28,908.30	EEEEY	14,468.63	15,148.26
EEEWY	28,285.77	29,022.50	EEEWY	14,525.73	15,262.46
EEWEY	28,285.77	29,022.50	EEWEY	14,525.73	15,262.46

Supporting Table 4.2: Calculated masses of oxidative coupling products

Y200C GFP Charge Tags	Reverse Primer
RRRRY	aGGTCTCcTTTAATAGCGTCGCCGACG GTGATGGTGATGGTGATGTTTGTACAG
GGSEEEEY	aGGTCTCcTTTAATATTCCTCTTCCTCGCTGCCACC GTGATGGTGATGGTGATGTTTGTACAGTTCA
Protein L, New C-Term	Forward Primer
All	aGGTCTCcCATGcgtaaaggcgaagagcTGTTCACTGG

Supporting Table 4.3: Primers used for preparing C-terminally charge tagged GFP variants. The resulting DNA was then transformed into a pET-28b vector using a Golden Gate protocol.

bmTYR Mutants	Forward Primer	Reverse Primer
R209H (1)	GCGTCCACCaCTGGGTAG	gcgataatgtcgggcaatcaggtgcgac
R209H (2)	gtcgcacctgattgcccgacattatcgc	CTACCCAGtGGTGGACGC
D55K	aGGTCTCcAAGcGCAATGCGGCGCACATGTCG	aGGTCTCcgCTTGCTTCCAGGCGGCGTGTGG

D55R	aGGTCTCcCGTcGCAATGCGGCGCACATGTCG	aGGTCTCcgACGGCTTCCAGGCGGCGTGTGG
E141K	aGGTCTCTAAGCAAGGAAACCCTAGTGGGGGT	aGGTCTCTGCTTATCAATCGTGGTCCAGCGCCC
V218K	aGGTCTCTAAGCCCACTGCCCCGAATGACCCC	aGGTCTCGGCTTAACACCCATTTGTCCGCCTAC
N205K	aGGTCTCcAAACGCGTCCACCGCTGGGTAGGC	aGGTCTCcgTTTGTGAAGCTGAGGCCCGTTAAT
E195K	aGGTCTCGAAAGGCTTTATTAACGGGCCTCAG	aGGTCTCCCTTTCAGCTGATTGCGGAAAGAGTT

Supporting Table 4.4: Primers used for preparing bmTYR point mutations from the DNA sequence shown below. Initial attempts at inserting mutations used 4 primers, two for each half of the plasmid, followed by a Gibson Assembly protocol to reform the complete plasmid. However, this proved to be highly cost ineffective, which prompted a switch to a Golden Gate Assembly point mutation protocol.

Computational methods and electrostatic mapping

Tyrosinase crystal structures from abTYR (PDB ID: 2Y9X) and bmTYR (PDB ID: 3NM8) were uploaded into the Schrodinger Maestro software suite. Following this, the proteins were then prepared and charges were assigned using the protein preparation wizard (PPW) tool in Maestro using the standard MGCF presets found in the Glide-Intro to Docking document (https://docs.google.com/document/d/1U4Oath1Lj7JFBt0qOH_g8s4GPwrwIMJahzifztCp4U0/). abTYR is a heterotetramer, and bmTYR is a monomer, and so each protein was prepared as such, and all metal ions other than those found in the type 3 di-nuclear copper active sites were deleted. In addition, a monomeric version of abTYR was constructed by deleting all but one subunit and subsequently put through the same PPW protocol again. Following this, single and double point mutants of the bacterial tyrosinase were created using the mutation tool found in the Maestro suite and all structures were passed through the PPW protocol again prior to performing any calculations.

To create the surface ESP plots, the Poisson-Boltzmann ESP tool was applied to each of the surfaces with the following presets, solute dielectric constant: 1, solvent dielectric constant: 80, solvent radius: 1.4 Å, and temperature: 298 K. These structures were then saved as .png image files. Next, the active site potential was calculated using the following protocol. First, a +1 lithium ion test charge was placed at a distance of 14 Å away from the active site of each enzyme variant. This distance was chosen as it is roughly equivalent to the position of the first adjacent residue to the reactive tyrosine in a peptide in the reactive conformation, the conformation in which the tyrosine phenol moiety is placed into the enzyme active site. Then the Macromodel current energy tool was used to calculate the energies of each enzyme variant both with and without the lithium ion test charge. These calculations were performed using the OPLS3e force field with charges coming from the force field, and with water as the solvent. To calculate the ESP, the difference between the electrostatic energy contribution to the total

energy for each structure were converted to an electrostatic potential difference using the following formula.

$$V = \frac{(E_{electrostatic\ with\ lithium} - E_{electrostatic\ w/o\ lithium})(\frac{kj}{mol}) \times 1000}{96\,485.3329(\frac{C}{mol})}$$

Finally, positively and negatively supercharged variants of the bacterial tyrosinase (bmTYR) were constructed using the online protein analysis tool STRIDE to calculate the level of solvent exposure of all residues in the crystal structure. Then to make the +30 variant, the negatively charged glutamate and aspartate residues were ordered by solvent exposure. The most solvent exposed residues were then chosen and swapped to lysine to achieve the desired charge. Conversely, to make the -30 variant, the most solvent exposed lysine and arginine residues were changed to glutamate to achieve the desired charge. For the STRIDE protocol, the structures were made by modifying the wild type enzyme using the mutation tool in Schrodinger Maestro. All structures were then minimized using the PPW protocol, and the surface ESP plots, and active site ESPs were calculated, similarly to what was done previously.

Tyrosinase-Mediated Oxidative Coupling Reactions of Proteins

Tyrosinase stocks and activity assay:

Agaricus bisporus Tyrosinase (abTYR) was obtained from Sigma Aldrich as a lyophilized powder and stored at -20 °C when received. Stock solutions (at 1 mg/mL) were prepared in 50 mM phosphate buffer, in 15% glycerol in water, pH 6.5, stored at -80 °C. *Bacillus megaterium* Tyrosinase was obtained by recombinant expression from *E. coli* as described above and stored as a stock solution and -80 °C in a similar buffer after purification. Tyrosinase aliquots were thawed on ice shortly before use and the activity of each new stock was assayed by monitoring conversion of L-Tyrosine to dopachrome via UV-Vis as described previously.[9] A stability study has indicated that the activity of abTYR aliquots prepared in this manner is stable over months of storage at -80 °C.[9]

Tyrosinase peptide charge screen for subsequent TOF-LCMS analysis:

Na2HPO4 pH 6.5			Protein Nucleophile			Peptide (mg/mL)			Tyrosinase			H2O	Total Vol
i (mM)	f (mM)	vol (µL)	i (µM)	f (µM)	vol (µL)	i (µM)	f (µM)	vol (µL)	i (U/L)	f (U/L)	Vol (µL)	(µL)	(µL)
50	10	10.0	500	50.0	5	1000	250	12.5	200	20.0	5.00	17.5	50.0

Volumes of each stock solution indicated in the table above were added to an Eppendorf tube with the tyrosinase enzyme added last. The resulting solution was allowed to stand at room temperature for 30 min before adding 5.0 µL of a 100 mM

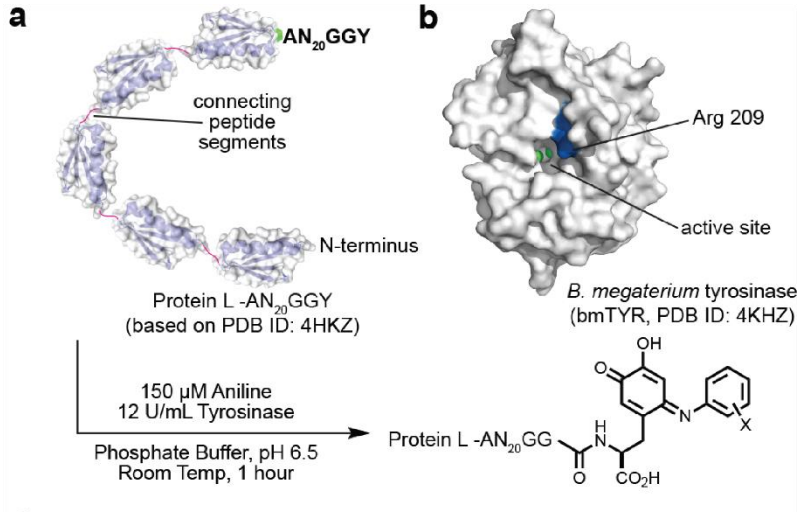
aqueous solution of tropolone. The quenched reaction was then diluted with 150 µL of pH 6.5 50 mM Tris-HCl buffer and submitted to TOF-LCMS analysis. This standard procedure was followed unless indicated otherwise.

Tyrosinase-mediated oxidative coupling charge screen kinetics assay

Tyrosinase aliquots were thawed on ice shortly before use and the activity of each new stock was assayed by monitoring reaction of c-terminal tyrosine peptides and free proline to form a dopachrome-like product via UV-Vis as described previously,[9] with the following modifications. As a substitute for free tyrosine, C-terminal tyrosine peptides were used for this experiment, resulting in the need for a separate nucleophile. Proline was chosen as a close mimic for the amine in free tyrosine since previous studies have shown it to produce a similarly colored product (ask Johnathan for citation). In order to minimize the effects of proline concentration on reaction rate, proline concentration was held at 50 mM for all reactions.

Tyrosinase-mediated protein trimer formation

Tyrosinase aliquots were thawed on ice shortly before use and diluted to a 10x stock solution. In the first reaction 10 µM Y200C GFP-GGSEEEEY was combined with 100 µM MYGGS-nanoluciferase and 200 nM abTYR for 40 min at RT before quenching with tropolone. Solution was then spin-concentrated 6 times using a 50 kDa MWCO spin filter from Milipore to remove excess luciferase. The retained sample was then combined at 10 µM with 100 µM of Y200C sfGFP with wild type bmTYR (12 U/L equivalent) for 45 min before quenching and spin concentration using a 50 kDa MWCO spin filter.



c

Variant	Type	Linker	Length	% Conversion
1	Flexible	-GGY	3	94
2	Flexible	-$(G_4S)_1$GGY	8	93
3	Flexible	-$(G_4S)_2$GGY	13	98
4	Flexible	-$(G_4S)_3$GGY	18	98
5	Alpha-helical	-A$(EAAAK)_2$AGGY	15	99
6	Rigid Ala/Pro	-$(AP)_3$GGY	9	99
7	Poly-Asn	-AN_{20}GGY	24	98
8	scFv C-Term. Seq.	-EIKRTGGY	8	97
9	ΔDomain 5	-(Domain 4-5 Linker)GGY	10	96
10	ΔDomain 5	-(Domain 4-5 Linker)G_4SGGY	15	95
11	None	None	0	0

Supplemental Figure 1. Modification of protein L C-termini using bmTYR. a) Crystal structure of protein L. Tyrosine tags were applied to the C terminus as shown including 2 that utilized the natural inter-domain linker sequence between domains 4 and 5 (Δ Domain 5) b) bmTYR structure c) conversion efficiency for bmTYR activation as measured through aniline addition via ESI-TOF MS. Figure taken from Marmelstein et al. with permission of all authors.[10]

Supporting Table 4.5: Calculated surface potential for bmTYR mutants			
Protein	Active Site Potential (V)	Charge Difference (from wt)	Delta Potential Difference (V)
abTYR a subunit	-4.432871517	N/A	N/A
abTYR full	-15.69927675	N/A	N/A
wt bmTYR	4.411211231	0	0
S54K	5.465411835	1	1.055618436
S54R	5.570592584	1	1.160799185
D55K	7.331465336	2	2.921671937
D55K, V218K	5.650227483	3	1.240434084
D55K, E141K	9.256587917	4	4.846794518

D55R	7.492515715	2	3.082722316
N57K	4.248702599	1	-0.1610908
N57R	4.023036184	1	-0.386757215
D123K	5.525005346	2	1.115211947
D123R	5.508730334	2	1.098936935
D140K	5.91658091	2	1.506787511
D140R	5.920912138	2	1.511118739
E141K	6.338844038	2	1.929050639
E141R	6.382729463	2	1.972936064
T156K	4.442343423	1	0.032550024
T156R	5.371160207	1	0.961366808
E158K	6.459611612	2	2.049818213
E158R	6.447344463	2	2.037551064
D181K	5.260717518	2	0.850924119
D181R	5.360472573	2	0.950679174
Q187K	5.057278315	1	0.647484916
Q187R	5.059220579	1	0.64942718
E195Q	5.498771308	1	1.088977909
E195R	7.044504633	2	2.634711234
N199K	5.263105446	1	0.853312047
N199R	5.191851095	1	0.782057696
N205R	5.629053281	1	1.219259882
R209H	3.076285068	-1	-1.333508331
Q214K	4.39331732	1	-0.016476079
Q214R	5.177722524	1	0.767929125
V217K	5.59261562	1	1.182822221
V217R	5.744882281	1	1.335088882
V218K	2.703413902	1	-1.706379497
V218R	4.113886259	1	-0.29590714
T220K	5.377436809	1	0.96764341
T220R	5.512172308	1	1.102378909
STRIDE Charge Swap +30	19.21950492	24	14.80971153
STRIDE Charge Swap -30	-15.61879094	-36	-20.02858434
Liu-AvNAPSA +30	24.94103651	24	20.53124311
Liu-AvNAPSA -30	-11.93816119	-36	-16.34795459
Kuhlman-AvNAPSA +30	17.44944218	24	13.03964878
Kuhlman-AvNAPSA -30	-14.7665744	-36	-19.1763678

Supporting Table 4.5: Computational charge screen for cationic mutants of bmTYR. Point mutants were generated in Maestro and minimized before addition of a lithium atom near the active site to calculate the active site potential. Potential difference measures the net change from wild type protein. The final entries represent different supercharging algorithms used to generate highly charged mutants of bmTYR.

5.7 References

1. Fairhead, M. & Thöny-Meyer, L. Bacterial tyrosinases: old enzymes with new relevance to biotechnology. *N. Biotechnol.* **29**, 183–191 (2012).

2. Zaidi, K. U., Ali, A. S., Ali, S. A. & Naaz, I. Microbial tyrosinases: Promising enzymes for pharmaceutical, food bioprocessing, and environmental industry. *Biochemistry Research International* **2014**, (2014).

3. Land, E. J., Ramsden, C. A. & Riley, P. A. Tyrosinase autoactivation and the chemistry of ortho-quinone amines. *Acc. Chem. Res.* **36**, 300–308 (2003).

4. Choi, J., Lee, Y.-M. & Jee, J.-G. Thiopurine Drugs Repositioned as Tyrosinase Inhibitors. *Int. J. Mol. Sci.* **19**, (2017).

5. Rescigno, A., Zucca, P., Flurkey, A., Inlow, J. & Flurkey, W. H. Identification and discrimination between some contaminant enzyme activities in commercial preparations of mushroom tyrosinase. *Enzyme Microb. Technol.* **41**, 620–627 (2007).

6. Marumo, K. & Waite, J. H. Optimization of hydroxylation of tyrosine and tyrosine-containing peptides by mushroom tyrosinase. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **872**, 98–103 (1986).

7. Mattinen, M. L., Lantto, R., Selinheimo, E., Kruus, K. & Buchert, J. Oxidation of peptides and proteins by Trichoderma reesei and Agaricus bisporus tyrosinases. *J. Biotechnol.* **133**, 395–402 (2008).

8. Montanari, E. *et al.* Tyrosinase-Mediated Bioconjugation. A Versatile Approach to Chimeric Macromolecules. *Bioconjug. Chem.* **29**, 2550–2560 (2018).

9. Maza, J. C. *et al.* Enzymatic Modification of N-Terminal Proline Residues Using Phenol Derivatives. *J. Am. Chem. Soc.* **141**, 3885–3892 (2019).

10. Marmelstein, A. M. *et al.* Tyrosinase-Mediated Oxidative Coupling of Tyrosine Tags on Peptides and Proteins. *J. Am. Chem. Soc.* **142**, 5078–5086 (2020).

11. Maza, J. C. *et al.* Secondary modification of oxidatively-modified proline N-termini for the construction of complex bioconjugates. *Org. Biomol. Chem.* **18**, 1881–1885 (2020).

12. Lantto, R., Puolanne, E., Kruus, K., Buchert, J. & Autio, K. Tyrosinase-Aided Protein Cross-Linking: Effects on Gel Formation of Chicken Breast Myofibrils and Texture and Water-Holding of Chicken Breast Meat Homogenate Gels. *J. Agric. Food Chem.* **55**, 1248–1255 (2007).

13. Inlow, J. K. Homology models of four Agaricus bisporus tyrosinases. *Int. J. Biol. Macromol.* **50**, 283–293 (2012).

14. Pretzler, M., Bijelic, A. & Rompel, A. Heterologous expression and characterization of functional mushroom tyrosinase (AbPPO4). *Sci. Rep.* **7**, 1810 (2017).

15. Shuster Ben-Yosef, V., Sendovski, M. & Fishman, A. Directed evolution of tyrosinase for enhanced monophenolase/diphenolase activity ratio. *Enzyme Microb. Technol.* **47**, 372–376 (2010).

16. Do, H., Kang, E., Yang, B., Cha, H. J. & Choi, Y. S. A tyrosinase, mTyr-CNK, that is functionally

available as a monophenol monooxygenase. *Sci. Rep.* **7**, 17267 (2017).

17. Ismaya, W. T. *et al.* Crystal structure of agaricus bisporus mushroom tyrosinase: Identity of the tetramer subunits and interaction with tropolone. *Biochemistry* **50**, 5477–5486 (2011).

18. Kanteev, M., Goldfeder, M., Chojnacki, M., Adir, N. & Fishman, A. The mechanism of copper uptake by tyrosinase from Bacillus megaterium. *J. Biol. Inorg. Chem.* **18**, 895–903 (2013).

19. Carrico, Z. M., Romanini, D. W., Mehl, R. A. & Francis, M. B. Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* **0**, 1205 (2008).

20. Tong, G. J., Hsiao, S. C., Carrico, Z. M. & Francis, M. B. Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* **131**, 11174–11178 (2009).

21. Mehl, R. A. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).

22. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).