



Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2016

The Development of Bicyclic Peptide Library Scaffolds and the Discovery of Biostable Ligands using mRNA Display

David E. Hacker
Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>

 Part of the [Chemistry Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/4614>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

© David Eugene Hacker _____ **2016**

All Rights Reserved

**The Development of Bicyclic Peptide Library Scaffolds and the Discovery of Biostable
Ligands Using mRNA Display**

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Chemical Biology)

At

Virginia Commonwealth University (2016)

By

David Eugene Hacker

B.S. Chemistry-Biochemistry (2010)

Virginia Commonwealth University, Richmond, VA

Research Advisor: **Matthew C.T. Hartman, Ph.D.**

Associate Professor, Department of Chemistry

Virginia Commonwealth University

and

Massey Cancer Center

Richmond, Virginia

December, 2016

ACKNOWLEDGEMENTS

I would like to begin where it began. My mother is a life-long educator who has served as the consummate role model and supported me unconditionally in everything I have ever attempted, not the least of which was this endeavor. My late father impressed upon me the importance of a strong work ethic and attention to detail. My two wonderful sons, Douglas and Andrew, have been my inspiration in all things and the center of my universe from their first day. My sister Mary Kay, endowed with an indomitable spirit, has provided much-needed encouragement, even amid her own extremely demanding life and personal travails. A special thanks to my brother Larry, his wife Christine and their darling children, Maya and Greyson, for their support and motivation, particularly during the initial stages of the preparation of this manuscript. And of course, none of this could have even been possible without the long-time support of my wife Jennifer.

Entering graduate school as a non-traditional student intending to pursue a life-long dream, I was incredibly fortunate to be accepted into the lab of Dr. Matthew Hartman. Matthew has been the prototypical mentor in all respects. His combination of knowledge, teaching ability, common sense and particularly his patience make his lab an ideal environment for learning and growing. The experience, while at times admittedly frustrating and daunting, has been extremely rewarding as a whole; and for that I owe him a debt of gratitude I could not possibly repay. I would be remiss if I did not prominently mention Jonathan Sheldon, my comrade from Day 1 of graduate school. Jon started out as my ‘third son’ and ended as a peer at the least. It was very rewarding to witness his growth as a scientist and I consider him a dear friend. A thank you to the senior members of the Harman Lab, both past and present, who served as examples to me of how to be a serious scientist; E. Railey White, Tim Reed, Michael Dcona, Daniela Selaya, Deboleena Mitra and Stacie

Richardson. And to the current group of promising, very capable graduate students to whom the torch has already been passed; Nicolas Abrigo, Emil Iqbal and Patrick Dupart.

I'd like to thank the members of my committee for their scientific insight throughout this process. Larry Povirk, Ashton Cropp, Heather Lucas and my committee chair, Vladimir Sidorov were all supportive and encouraging, particularly during the candidacy requirements. I'd also like to thank the VCU Department of Chemistry and Chemical Biology, Massey Cancer Center and Virginia Commonwealth University for the opportunity to work and collaborate at a top-notch institution.

Preface

As advancements in biotechnology unveil ever-expanding information about diseases and disease-associated biochemical pathways, the number of cellular targets attributable to these factors increases proportionally. The traditional approach to therapeutic treatment of diseases has centered around the use of small molecule drugs. However, many disease-associated pathways involve protein-protein interactions (PPIs) with large interacting surfaces. These interfaces are typically not suitable for small molecules, which make relatively few contacts and are therefore not target-specific. As a result, most of these PPIs are considered ‘undruggable’.

Peptides are mini-protein molecules that are capable of spanning a broader surface area than small molecules, are highly target-specific and bind cellular proteins with high affinity. But peptides in their native form are typically unable to enter cells and are not biologically stable. This work focuses on techniques that address the deficiencies of peptides in an effort to make them more attractive therapeutic candidates. By incorporating non-native building blocks and making highly-constrained cyclic peptide conformations, we demonstrate the ability to make peptides more ‘small molecule-like’, with a goal towards increasing biostability. These techniques were adapted to the construction of stable peptide scaffolds that we used in conjunction with our ability to generate peptide libraries containing trillions of unique molecules. These drug-like libraries can then be used to identify peptides that can interact selectively with disease-associated PPIs previously considered undruggable. As a result, the peptides discovered using this technology will be more likely to make it to market as therapeutics.

Table of Contents

Acknowledgements	ii
Preface	iv
Table of Contents	v
List of Figures	ix
List of Tables.....	xiv
Abstract	xvi
Chapter 1: Introduction: Ribosomal Synthesis of Natural Product-Like Peptide Libraries for Drug Discovery	1
1.1 Expanding the genetic code.....	2
1.2 Making peptides more natural product-like	5
1.3 Drug discovery	14
1.4 Future outlook	21
Chapter 2: Ribosomal Translation of Unnatural Peptides: Incorporation of Amino Acid Analogues to Increase Biostability and Chemical Diversity	22
2.1 Introduction	23
2.2 Results and Discussion.....	27
2.2.1 Incorporation of individual analogues in His ₆ -tagged templates	30
2.2.2 Incorporation of individual analogues in Flag-tagged templates	40

2.2.3 Dual incorporations in Flag-tagged templates.....	51
2.2.4 Library validation-multiple incorporations	55
2.3 Summary.....	60
2.4 Future Directions	61
2.5 Experimental	62
Chapter 3: Cyclization methods of synthetic and <i>in vitro</i> translated peptides: Creation of knotted peptides and peptide libraries	66
3.1 Introduction	67
3.2 Results and Discussion	70
3.2.1 Cyclization of solid phase-synthesized peptides using bromomethyl benzene-derivatized linkers	70
3.2.2 Cyclization of <i>in vitro</i> translated peptides using α,α' -dibromo- <i>m</i> -xylene ..	81
3.2.3 Toward CuAAC bicyclization of <i>in vitro</i> translated peptides	83
3.3 Summary.....	90
3.4 Future Directions	91
3.5 Experimental	92
Chapter 4: Mono- vs. Bicyclic Peptide Libraries; <i>In vitro</i> Selection with mRNA Display ..	98
4.1 Introduction	99
4.2 Results	101

4.2.1 Compatibility of CuAAC chemistry with mRNA display	101
4.2.2 Comparative scaffold-diverse <i>in vitro</i> selections	107
4.2.3 Sequence analysis	112
4.2.4 Affinity for streptavidin.....	114
4.2.5 Bicyclization enhances protease stability	118
4.3 Discussion.....	120
4.4 Future Directions	124
4.5 Experimental.....	125
Chapter 5: Scaffold-Diverse Peptide Libraries; <i>In vitro</i> Selection Against Streptavidin #2...	
.....	135
5.1 Introduction	136
5.2 Results	139
5.2.1 Design of semi-random cyclizable libraries	139
5.2.2 <i>In vitro</i> selection against streptavidin #2.....	142
5.2.3 Sequencing results: Streptavidin #2	147
5.2.4 Sequencing results: Streptavidin #2-chymotrypsin	147
5.3 Discussion.....	150
5.4 Future Directions	154
5.5 Experimental.....	155

Chapter 6: <i>In vitro</i> Selection Against XRCC4 using Scaffold-Diverse Peptide Libraries	160
6.1 Introduction	161
6.2 Results	164
6.2.1 Design of semi-random cyclizable libraries	164
6.2.2 <i>In vitro</i> selection against XRCC4.....	164
6.2.3 Sequence analysis: XRCC4.....	167
6.3 Discussion	169
6.4 Future Directions	170
6.5 Experimental.....	171
Bibliography	175
Appendices	191
Appendix A Lanthanide binding peptide for real-time translation monitoring.....	192
Appendix B TRAP: Streamlining mRNA display.....	199
Appendix C HiStrep: Dual affinity epitope tag.....	213
Appendix D Coupled <i>in vitro</i> transcription-translation.....	221
Appendix E Optimization of <i>in vitro</i> translation.....	229
Appendix F Peptide binding to BRCA 1.....	235
Appendix G Miscellaneous experiments.....	238
Appendix H Miscellaneous Tables	242

List of Figures

Chapter 1

Figure 1.1: Ribosomal incorporation of non-canonical amino acids	5
Figure 1.2: Macrocyclic peptide natural products	9
Figure 1.3: Peptide cyclization strategies	13
Figure 1.4: Ligand discovery technologies	20

Chapter 2

Figure 2.1: Incorporation of unnatural amino acids in PURE translation	25
Figure 2.2: Amino acid analogs	26
Figure 2.3: Valine analog incorporation in MHVMH ₆ M template	32
Figure 2.4: Glutamine incorporation in response to the hexahistidine codons	34
Figure 2.5: Tyrosine and aspartic acid analog incorporations in MDYKMH ₆ template	38
Figure 2.6: Optimization of MD incorporation in MDYKMH ₆ template	39
Figure 2.7: Leucine and glutamic acid analog incorporations in MLEPQ-Flag template	42
Figure 2.8: Isoleucine analog incorporations in MTINR-Flag template	44
Figure 2.9: Arginine analog incorporation in MTINR-Flag template	46
Figure 2.10: Tryptophan analog incorporation in MHFSW-Flag template	48
Figure 2.11: Phenylalanine analog incorporations in MHFSW-Flag template	50

Figure 2.12: Dual incorporations in translation templates	52
Figure 2.13: Amino acid analogs for multiple incorporations	54
Figure 2.14: Relative quantification of multiple analog incorporations.....	58
Figure 2.15: Fidelity of multiple analog incorporations in MVLD FEW KIYMH ₆ template.....	59

Chapter 3

Figure 3.1: Mono- and bicyclization routes of synthetic and <i>in vitro</i> translated peptides	69
Figure 3.2: Cyclization of synthetic peptides with α,α' -dibromo- <i>m</i> -xylene	73
Figure 3.3: Cyclization of a peptide containing one cysteine	75
Figure 3.4: Cyclization of a peptide containing one cysteine using a stabilizing linker	77
Figure 3.5: Cyclization with trifunctional bromomethyl benzene linkers.....	80
Figure 3.6: Cyclization of <i>in vitro</i> translated peptide with α,α' -dibromo- <i>m</i> -xylene	82
Figure 3.7: Bicyclic peptide conformations and cyclizable analog incorporation	85
Figure 3.8: Mono- and bicyclization of <i>in vitro</i> translated peptides	87
Figure 3.9: Labeling of azide peptide with an external alkyne reagent.....	88
Figure 3.10: Lack of labeling of azide/alkyne 'pre-clicked' peptide with an external alkyne reagent	89

Chapter 4

Figure 4.1: Scheme of competitive mono- and bicyclic selections	100
Figure 4.2: Copper coordinating ligand effect on mRNA stability	103

Figure 4.3: Compatibility of click reaction with peptides displayed on mRNA	105
Figure 4.4: Compatibility of azide-peptide with dithiothreitol (DTT)	106
Figure 4.5: mRNA display scheme	109
Figure 4.6: Design of mono- and bicyclic libraries	110
Figure 4.7: Selection enrichment.....	111
Figure 4.8: AptaTools round-by-round sequence analysis.....	113
Figure 4.9: MALDI-TOF of <i>in vitro</i> translated selection winners	115
Figure 4.10: Relative binding to streptavidin	116
Figure 4.11: Binding curves of selection winners	117
Figure 4.12: Protease stability	119
Figure 4.13: Selection winner homology	122
Chapter 5	
Figure 5.1: Library scaffolds for <i>in vitro</i> selections	138
Figure 5.2: Semi-random library scaffold generation method	141
Figure 5.3: Scaffold diversity	143
Figure 5.4: Scheme for chymotrypsin digestion in selection	144
Figure 5.5: Selection enrichment.....	146
Figure 5.6: Sequence analysis	148

Figure 5.7: Sequence homology149

Figure 5.8: Protease-stable sequences.....152

Chapter 6

Figure 6.1: Non-homologous end joining (NHEJ)163

Figure 6.2: Selection enrichment.....166

Figure 6.3: Sequence analysis168

Figure 6.4: Purification of GST-XRCC4.....172

Appendix Figures

Figure A1: Lanthanide Binding Tag for monitoring peptide synthesis.....194

Figure A2: Sensitivity of LBT to terbium energy transfer195

Figure B1: TRAP and modified mRNA display201

Figure B2: TRAP display validation scheme205

Figure B3: Comparative Flag-tag enrichment of TRAP and conventional selections206

Figure B4: Modified mRNA display: *in situ* covalent association.....207

Figure C1: HiStrep-tag discovery and scheme for dual affinity purification214

Figure C2: Dual affinity purification of HiStrep-sfGFP216

Figure D1: Coupled transcription-translation reactions with sfGFP DNA224

Figure D2: Comparison of translation and coupled reactions225

Figure D3: Optimization of coupled transcription-translation	226
Figure E1: Buffers and small molecules	230
Figure E2: Ribosomes and RP mix	231
Figure E3: PURE Translation Protein Factors	232
Figure E4: Aminoacyl-tRNA synthetases	233
Figure F1: BRCA1-binding peptide 8.6 with and without canavanine (arginine analog).....	236
Figure F2: Binding of peptide 8.6 to BRCA1	237
Figure G1: mRNA-peptide fusion formation	239
Figure G2: UV crosslinking of mRNA-puromycin linker	240
Figure G3: Tris-tricine peptide SDS-PAGE for cyclization characterization	241
Figure H1: Streptavidin (monocyclic) Top 100 sequence homology (10 random AAs)	244
Figure H2: Streptavidin (bicyclic) Top 100 sequence homology (10 random AAs)	247

List of Tables

Table 2.1: mRNA templates for analog incorporation	28
Table 2.2: Amino acid misincorporation table	29
Table 3.1: Peptide and mRNA sequences for cyclization experiments.....	97
Table 4.1: Oligonucleotides used in mono- vs. bicyclic scaffold selection	125
Table 5.1: Oligonucleotides used for streptavidin #2 and XRCC4 selections.....	155
Table B1: Oligonucleotides used for TRAP display validation	210
Table H1a: Streptavidin (monocyclic) Top 100 DNA/peptide sequences from AptaTOOLS...	242
Table H1b: Streptavidin (monocyclic) Top 100 DNA/peptide sequences from AptaTOOLS ..	243
Table H2a: Streptavidin (bicyclic) Top 100 DNA/peptide sequences from AptaTOOLS.....	245
Table H2b: Streptavidin (bicyclic) Top 100 DNA/peptide sequences from AptaTOOLS	246
Table H3a: Streptavidin #2 Top 100 DNA/peptide sequences from AptaTOOLS	248
Table H3b: Streptavidin #2 Top 100 DNA/peptide sequences from AptaTOOLS.....	249
Table H4a: Streptavidin #2-Chymotrypsin Top 100 DNA/peptide(enrich) from AptaTOOLS	250
Table H4b: Streptavidin #2-Chymotrypsin Top 100 DNA/peptide(enrich) from AptaTOOLS	251
Table H5a: XRCC4 Top 100 DNA/peptide sequences from AptaTOOLS.....	252
Table H5b: XRCC4 Top 100 DNA/peptide sequences from AptaTOOLS	253
Table H6: Fixed-cysteine monocyclic libraries used in our previous selection against XRCC4....	

.....	254
Table H7: DNA mixing ratios for sequencing	255
Table H8: Streptavidin #2 and XRCC4 libraries: Cyclizable codon probability calculations ...	256

Abstract

Peptides are a promising class of therapeutic candidates due to their high specificity and affinity for cellular protein targets. However, peptides are susceptible to protease degradation and are typically not cell-permeable. In efforts to design more effective peptide drug discovery systems, investigators have discovered that incorporation of non-canonical amino acids (ncAAs) and macrocyclization overcome these limitations, making peptides more drug-like.

In this work, we exploit the promiscuity of wild-type aminoacyl-tRNA synthetases (aaRSs) to ‘mischarge’ ncAAs onto tRNA and ribosomally incorporate them into peptides using a cell-free translation system. We have demonstrated the ability to incorporate five ncAAs into a single peptide with near-wild type yield and fidelity. We also demonstrated the *in situ* incorporation of ncAAs containing azide and alkyne functionalities, enabling the use of CuAAC (click chemistry) to generate triazole-bridged cyclic peptides. When combined with bisalkylation of peptides containing two cysteines via an α,α' -dibromo-*m*-xylene linker, we created bicyclic peptides which are structurally similar to the highly bioactive knotted peptide natural products.

Biological display methods, such as mRNA display, are powerful peptide discovery tools based on their ability to generate libraries of $>10^{14}$ unique peptides. We combined our ability to incorporate ncAAs with our bicyclization technique adapted for use with mRNA display to create knotted peptide library scaffolds. We performed side-by-side monocyclic and bicyclic *in vitro* selections against a model protein (streptavidin). Both selections resulted in peptides with mid-nM affinity, and the bicyclic selection yielded a peptide with remarkable protease resistance.

We used a new library that enables the generation of a diverse collection of linear, monocyclic and bicyclic scaffolds in one pot, increasing the likelihood of target-ligand

conformational alignment. We performed a second selection against streptavidin and revealed a nearly unanimous preference for linear peptides containing an HPQ motif, a known streptavidin-binding sequence. However, when we used these libraries for *in vitro* selection against a biological target, DNA repair protein XRCC4, we did not observe convergence.

In summary, we have developed a novel technique for production of bicyclic peptide libraries. These highly-constrained protease-stable scaffolds can be used as platforms to identify high affinity, drug-like ligands using mRNA display.

Chapter 1 Introduction:

Ribosomal Synthesis of Natural Product-Like Peptide Libraries for Drug Discovery

1.1 Expanding the genetic code

The central dogma of biology is based on the storage of genetic information in the form of DNA. The gene is expressed intracellularly in a two-step process, beginning with transcription of the DNA by RNA polymerases to generate RNA. While many non-coding RNA molecules play a part in cellular functions,^{1,2} it is the production of proteins via ribosomal translation of messenger RNA (mRNA) that allows the cell to perform many of the numerous integral functions necessary for survival and proliferation.

In vitro protein synthesis

In prokaryotes, gene expression is a coupled process with the ribosome translating the mRNA long before transcription is completed.³ The concerted nature of these processes enables the study of protein production and function starting from the DNA level. The capability of proteins to be produced outside of the cellular environment was recognized early on.⁴ This discovery enabled researchers to investigate protein synthesis outside the constraints of *in vivo* expression, which typically requires multiple genetic manipulations such as cloning, transformation and overexpression to produce and study the effects of proteins in the cell. Cell-free protein synthesis (CFPS) systems are based on the ability to reproduce translation *in vitro*. One method to do this is based on the use of crude cell extracts (called the S30 fraction) derived from several sources, including *E. coli*,⁵ and are commonly used because of the lack of preparation required and high protein production. An alternative to lysate-derived systems like S30 is the PURE (Protein Synthesis Using Recombinant Elements) system,⁶ which is based on the use of purified recombinant hexaHistidine-tagged proteins. This system requires the individual preparation and addition of all of the necessary protein factors, ribosomes, tRNAs, synthetases,

amino acids and energy sources required for protein synthesis. The controllable nature of this system allows for facile manipulation simply by omission of specific components, supplementation of any endogenous component, or addition of exogenous components.

The genetic code

In a biological context, the genetic alphabet of nucleobases, A, G, C and T are arranged in specifically programmed triplet codons.⁷ There are 61 sense codons that designate the ribosomal incorporation of 20 proteinogenic amino acids. The remaining 3 of the 64 total codons represent a signal to terminate protein expression via release factor recognition. Amino acids are aminoacylated, or charged, onto their cognate tRNA by specific aminoacyl-tRNA synthetases (aaRSs).⁸ The anticodon of the aminoacyl-tRNA is then paired with the complementary triplet codon of the mRNA and the amino acid is incorporated into the nascent peptide. Once the mRNA is decoded, release factors recognize the nonsense codon and catalyze the hydrolysis of the peptidyl-tRNA, resulting in the release of the protein from the ribosome.

Expanding the code

While the biological diversity present in the side chains of the proteinogenic amino acids results in a wide variety of chemical interactions within the cell, from catalytic transformations to protein-protein interactions in signaling pathways, there is a relatively high degree of functional group redundancy. For example, there are two carboxylic acid side chains, two amides and several similar aliphatic residues. In many of these cases, the residues differ only by the presence of a methylene group and the resulting chemical interactions are synonymous. The advent of cell-free systems has allowed researchers to expand the repertoire of amino acid functionality and find new levels of diversity not present within the native proteome.

Nonsense suppression

The three nonsense codons in mRNA, UAG, UAA and UGA, signal the recruitment of release factors and result in termination of protein synthesis and release of the full-length protein. However, the discovery that UGA specifies Trp in mitochondria signaled a deviation from the dogmatic genetic code.⁹ Other organisms have since been shown to incorporate amino acids in response to termination codons,¹⁰ such as the recognition of the opal codon (UGA) by selenocysteine,¹¹ and the amber codon (UAG) by pyrrolysine.^{12,13} The addition of an orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pair to a translation reaction results in the incorporation of the non-proteinogenic amino acid into proteins in response to a nonsense codon (**Figure 1.1a**). This phenomenon is known as stop codon suppression and enables the expansion of the genetic code to include a 21st amino acid and increase the functional diversity of proteins.¹⁴⁻¹⁸ Stop codon suppression, particularly involving the amber codon (UAG), has been used to site-specifically incorporate fluorophores,¹⁹ photo-crosslinkers,²⁰ metal ion chelating groups,^{21,22} and other unique functionalities into proteins to further probe the proteome.

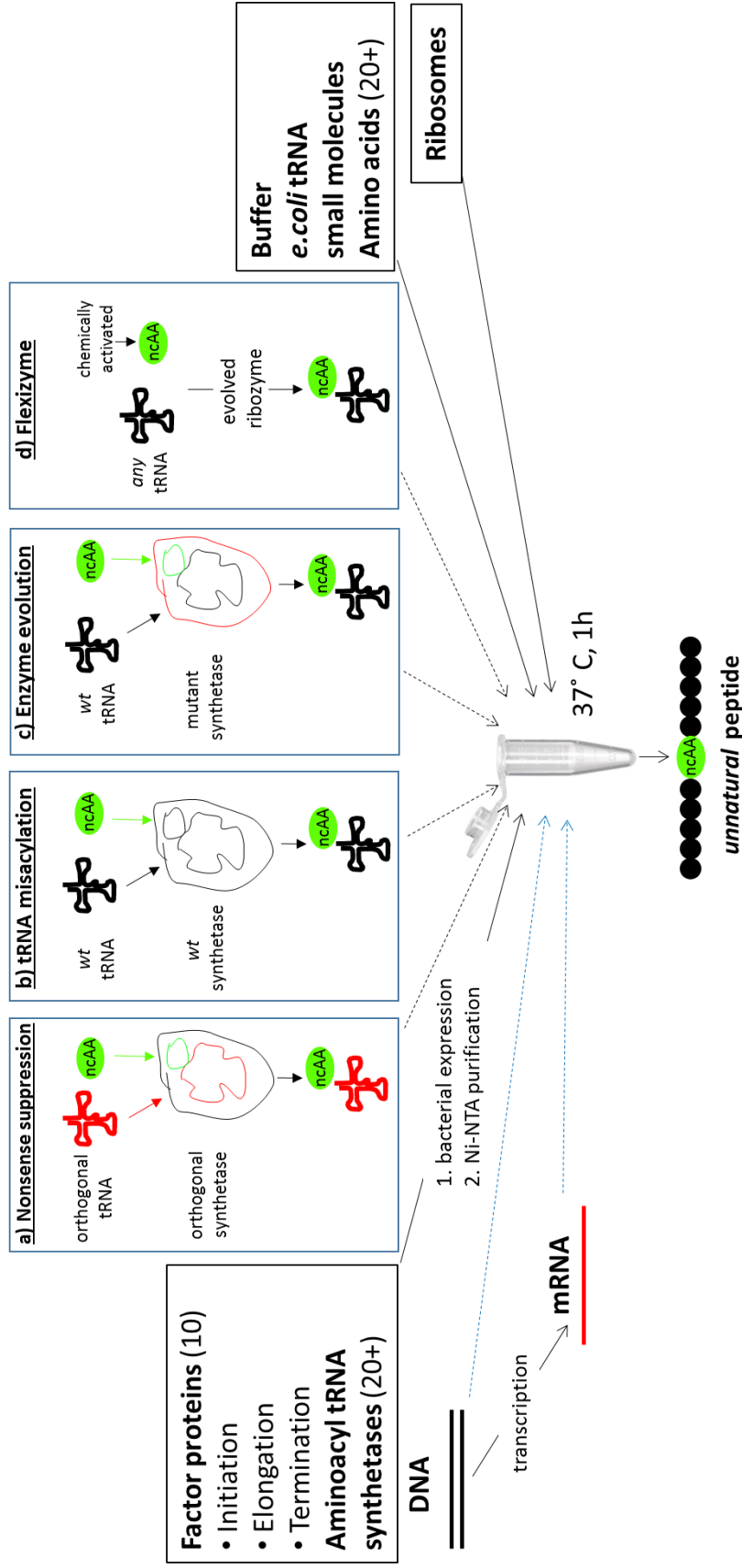


Figure 1.1 Ribosomal incorporation of non-canonical amino acids. *In vitro* methods to create unnatural peptides using the PURE system. a) Nonsense suppression using an orthogonal tRNA/synthetase pair, b) mischarging of tRNA using wild-type translation machinery, c) directed evolution of aaRSs and d) the Flexizyme system, which can charge any tRNA with an activatable amino acid.

Misacylation of tRNAs

The high degree of aaRS substrate specificity results in accuracy of aminoacylation of at least 10,000 to 1.²³ However, it has been shown that some exogenous amino acids can be recognized as substrates by wild-type aaRSs and incorporated into peptides with the concurrent omission of the natural substrate (**Figure 1.1b**). Using the PURE system, Szostak's group succeeded in incorporating 12 amino acid analogs into a peptide simply by substituting the analogs for their wild-type counterparts.²⁴ The same group later identified 59 previously unknown aaRS substrates,²⁵ and over 50 ncAAs, including α,α -disubstituted, N-methylated and β -hydroxyl analogs, that could be ribosomally incorporated into peptides.²⁶ This strategy does not require manipulation of the mRNA and peptides can be synthesized *in situ* under high turnover conditions. However, since enzyme recognition is required for charging, the analogs must be structurally similar to the proteinogenic substrate. Directed evolution of the aaRS has allowed researchers to expand amino acid substrate specificity in an effort to further increase diversity (**Figure 1.1c**).²⁷

Flexizymes

Charging of amino acids onto tRNA by aaRSs requires amino acid specificity and enzyme evolution to expand substrate recognition is laborious. Suga and coworkers evolved ribozymes, termed flexizymes,^{28,29} that catalyze aminoacylation of virtually any tRNA using an amino acid that has been chemically activated with an appropriate leaving group attached via an ester linkage (**Figure 1.1d**).^{30,31} This versatile method has since enabled the Suga group to initiate translation with D-amino acids,³² incorporate multiple N-methyl amino acids,³³ and ribosomally synthesize polypeptoids and peptide-peptoid hybrids.³⁴

1.2 Making peptides more natural product-like

Peptides have advantages over small molecules in terms of target specificity and affinity. Many validated targets that were previously considered ‘undruggable’ by small molecules, due to either an expansive surface area or shallow binding interface, are more amenable to peptide ligands. However, there are intrinsic characteristics of peptides that detract from their attractiveness as biologics. There are >500 proteases in the human degradome.³⁵ As a result, the general lack of serum stability impedes the ability of peptides to reach their intended cellular target. The overall hydrophilicity of peptides prevent them from permeating the cell membrane. Furthermore, peptides generally have molecular masses in excess of Lipinski’s ‘rule of 5’ (RO5)³⁶ limitations. As a result of these combined factors, very few linear peptides derived from library-based selections have had a clinical impact.

Peptide natural products as inspiration

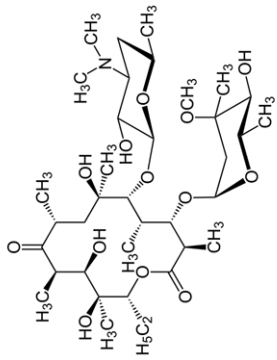
Many natural compounds produced by microbes and plants are radical departures from the general limitations of peptides as potential therapeutics. Natural selection has evolved this class of peptide natural products (PNPs) with potent biological activity and stability. Current macrocyclic drugs are obtained almost exclusively from natural sources. Some examples; erythromycin (**Figure 1.2a**), a macrolide antibiotic, was originally isolated from *Streptomyces erythraeus*; Cyclosporine A (**Figure 1.2b**), isolated from the fungus *Tolypocladium inflatum*, is an orally available immunosuppressant; and the antituberculosis compound rifampin (**Figure 1.2c**), from *Amycolatopsis rifamycinica*, is also administered orally. These potent pharmacores all have evolved to possess profound pharmacological activity, despite the existence of multiple RO5 violations. In many ways, these peptide natural products serve to bridge the divide between linear

peptides and small molecules, possessing the positive attributes of each class. As larger molecules, PNPs have the ability to make multiple contacts over a broad interface, while also possessing the membrane permeability and biostability of small molecules.

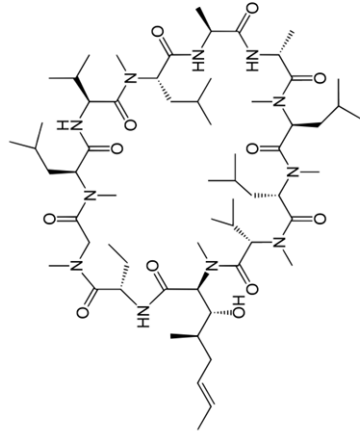
The attractiveness of PNPs has led researchers to probe the structural characteristics of these compounds with an eye toward the development of peptide discovery platforms that ultimately lead to viable drugs.³⁷ Many of these compounds contain *N*-methyl amino acids.³⁸ *N*-methylation reduces the flexibility of the peptide backbone and decreases the hydrogen bond donating ability of the peptide bond to water molecules, which leads to enhanced membrane permeability.³⁹ *N*-methylation of the peptide backbone has also resulted in improved target selectivity,⁴⁰ and has led to increased oral bioavailability of peptides.⁴¹

In addition to *N*-methylated backbones, many PNPs contain non-canonical amino acids (ncAAs). Researchers have shown that incorporation of ncAAs with unique side chains has led to enhanced protease stability,⁴² as well as increased cell permeability.⁴³⁻⁴⁵

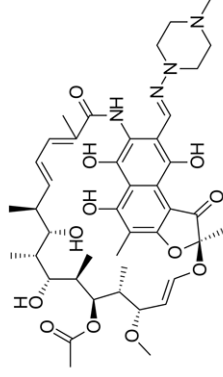
The most distinguishing characteristic of many PNPs that lead to their enhanced potency and biostability is a macrocyclic structure. These cyclic structures are achieved with a variety of linkages, such as backbone head-to-tail (such as cyclosporine), attachment of the N- or C-terminus to a side chain (daptomycin-**Figure 1.2d**) or side chain-to-side chain cyclization (vancomycin-**Figure 1.2e**). There have been many reports emphasizing the significant biological advantages of cyclic over linear conformations, for both peptides and small molecules.⁴⁶



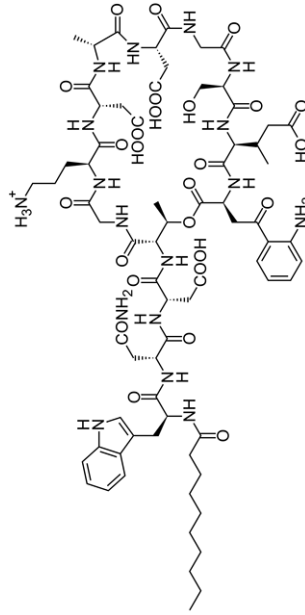
a) erythromycin



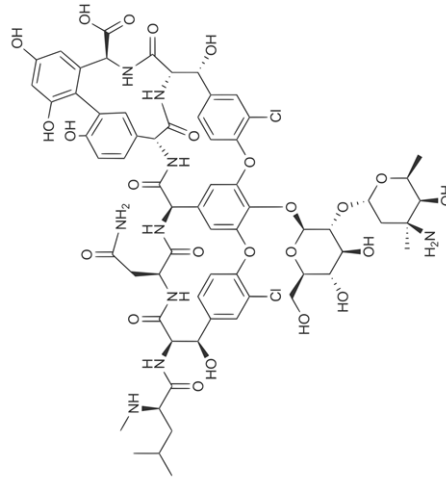
b) cyclosporine A



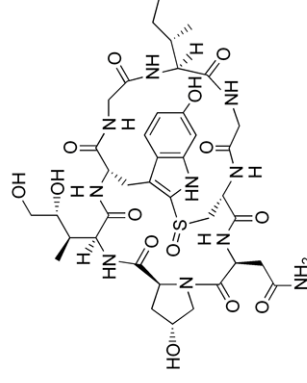
c) rifampin



d) daptomycin



e) vancomycin



f) α -amanitin

Figure 1.2 Macrocyclic Peptide Natural Products.

The conformational constraint enforced on peptides by circularization decreases the relative entropic cost of binding, and as a result target affinity has been shown to be 10-1000 fold higher for cyclic peptides relative to the corresponding linear sequence.⁴⁷ There is also restricted access to the peptide backbone, which leads to increased proteolytic stability.^{42,48-50} Cyclization also increases the degree of intramolecular hydrogen bonding, which reduces interaction with water molecules and increases the potential for permeating the cell membrane.^{51,52}

Production of ribosomally translated macrocyclic peptides

Given the wealth of examples of highly active PNPs, researchers have used these structures as inspiration to create cyclic peptide scaffolds used to discover bioactive ligands. However, cyclization of peptide libraries is challenging. The reaction must be robust and result in high conversion; and specific in order to minimize undesirable side products. Millward and coworkers used a cross-linking agent, disuccinimidyl glutarate (DSG) to post-translationally cyclize an mRNA display library via the N-terminal amine and the ϵ -amine of a downstream lysine (**Figure 2a**).⁵³ In another example of the use of small molecule cross-linking agents to create macrocyclic peptide libraries, Szostak and coworkers first demonstrated the use of α,α' -dibromo-*m*-xylene to cyclize *in vitro* translated peptides containing two cysteines (**Figure 2b**).⁵⁴ The same group later combined this technique with mRNA display libraries containing a majority of ncAAs to discover low-nM inhibitors of thrombin.⁵⁵ Suga and coworkers developed a co-translational cyclization method for head-to-side chain library cyclization using the FIT system. In this method, N^α -(2-chloroacetyl)-Trp is incorporated at the N-terminus via flexizyme-mediated reassignment of the initiator codon. The ncAA reacts spontaneously with a downstream cysteine, resulting in thioether-bridged macrocyclic libraries in a methionine-free *in vitro* translation system (**Figure 2c**).⁵⁶ This group has since used the versatile FIT platform to introduce reactive residues capable of producing

macrocyclic peptides via oxidative coupling,⁵⁷ Michael addition,⁵⁸ Copper-catalyzed azide-alkyne cycloaddition (CuAAC-**Figure 2d**),⁵⁹ and native chemical ligation techniques.⁶⁰

Highly constrained bicyclic peptides

The peptide natural product classes of amatoxins and phallotoxins, produced by the mushroom *Amanita phalloides* (among other mushrooms), are highly toxic agents. Unlike most cyclic natural products, these compounds are translated on the ribosome and post-translationally bicyclized.⁶¹ An example of the amatoxin family, α -amanitin (**Figure 1.2f**) is an octapeptide that has a theta-bridge⁶² or inner loop, between 6-hydroxytryptophan and cysteine within the outer macrocycle, and is a potent inhibitor of RNA polymerase II.⁶³

The beneficial contributions conferred to peptides by macrocyclization techniques have been shown to be supplemented by the addition of a second cycle, creating very highly constrained compounds. Bicyclization of peptides has led to further increases in target affinity,^{64,65} biostability,^{66,67} and cell permeability⁶⁸⁻⁷⁰ relative to the linear and monocyclic counterparts. In an extension of their previous work, the Suga group combined two of their cyclization techniques to produce theta-bridged peptides between orthogonally reactive residues incorporated using flexizymes (**Figure 1.2e**).⁵⁹ Chen and coworkers identified bicyclic peptide ligands among a highly diverse set of bicyclic conformers joined by a pair of disulfide bonds between two sets of cysteines (**Figure 1.2f**).⁷¹ However, the labile nature of the disulfide linkage makes these structures susceptible to ring-opening in the reducing intracellular environment.^{72,73} Heinis and coworkers were the first to create non-reducible bicyclic peptide libraries containing two conjoined loops via alkylation of three fixed cysteines using tris-1,3,5-(bromomethyl)benzene (**Figure 1.2g**). They created this bicyclic library on phage and identified a high affinity bicyclic lead inhibitor (K_i

= 1.5 nM) panning against human plasma kallikrein.⁷⁴ This work was extended by the Heinis group with the use 1,3,5-triacryloyl-1,3,5-triazinane (TATA) and *N,N',N''*-(benzene-1,3,5-triyl)-tris(2-bromoacetamide) (TBAB) as tris-alkylating agents to create bicyclic peptide libraries with more hydrophilic cores that are more conformationally stable as a result of non-covalent interactions between the central linker and peptide.⁷⁵

The examples of peptide cyclization methods given above highlight the progress made in exploring untapped architectural features previously associated only with PNPs. The application of these macrocyclization strategies to peptide library-based combinatorial methods enable the construction of more robust and stable peptide scaffolds, thereby increasing the potential of ligands discovered using these systems to make it to market as viable drugs.

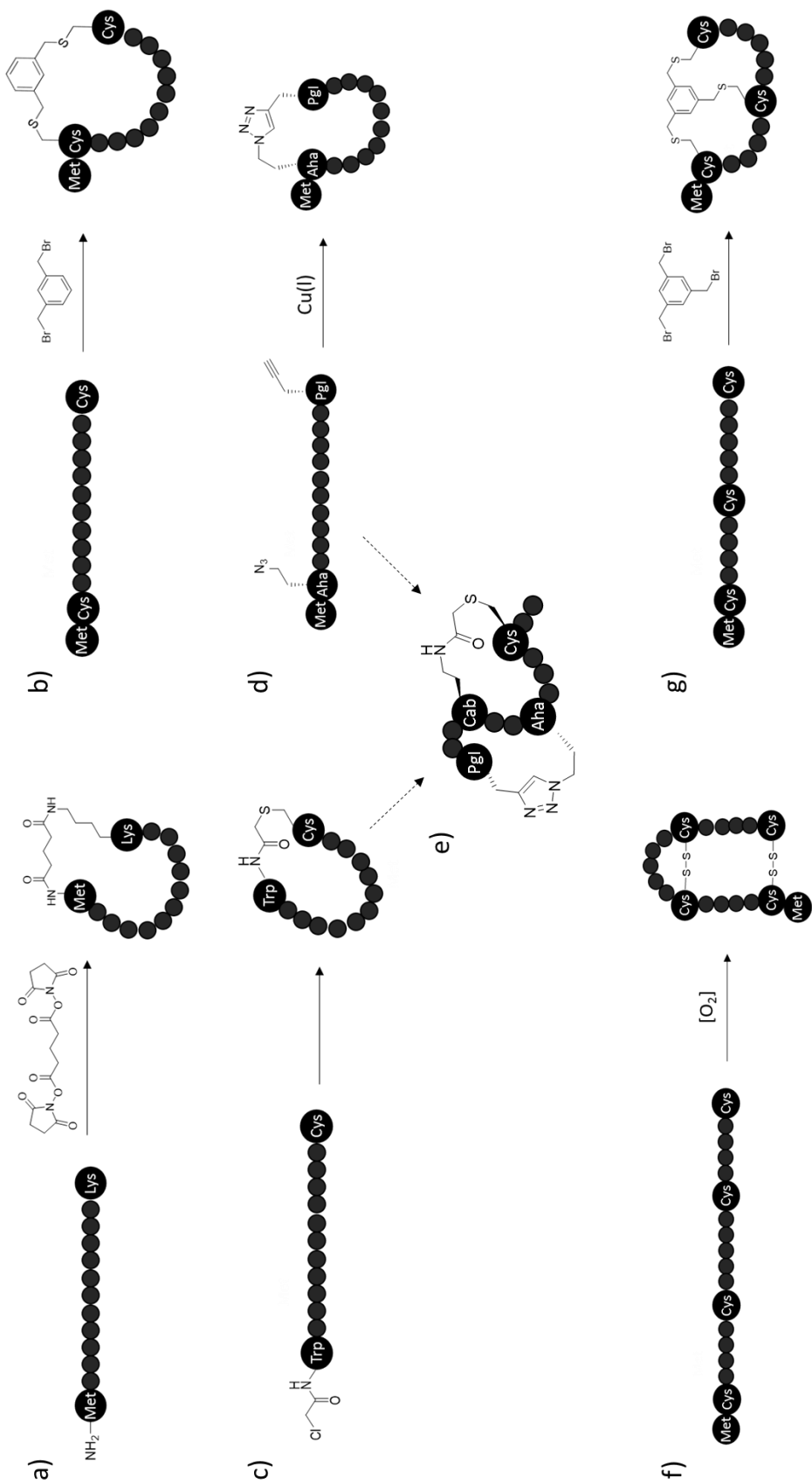


Figure 1.3 Peptide cyclization strategies. a) Dissuccinimidyl glutarate (DSG) used to cyclize peptides produced using rabbit reticulocyte lysate, reacting with the N-terminal amine and a downstream lysine; b) α,α' -dibromo-*m*-xylene cyclization of peptides containing two cysteines; c) Spontaneous reaction of a chloroacetyl derivative with a downstream cysteine to form a thioether bridge; d) CuAAC-mediated cyclization with azide and alkyne amino acids; e) Combination of strategies seen in c) and d); f) Disulfide cyclization and g) Tris-bromomethylbenzene bicyclization of peptides containing three cysteines.

1.3 Drug discovery

The majority of drugs developed in the 20th century stem from what is commonly referred to as the ‘chemistry era’,⁷⁶ when small molecule natural products were either serendipitously discovered or used as leads to generate drugs from screens or rational design based on ligand, receptor, or mechanism-based factors.⁷⁷ This approach led to many successes in the treatment of numerous diseases. This almost exclusive reliance on small molecules as therapeutics led to a number of broad generalizations, most famously Lipinski’s ‘Rule of 5 (RO5)’,³⁶ that have since served as a guide to aid in the search for bioactive compounds. A common approach to small molecule-based drug discovery involves the creation of libraries, most often synthesized on solid phase. These one-bead, one-compound libraries are produced using two general strategies. Retrosynthetic design, based on a known ligand (“target-oriented” or TOS), leads to the synthesis of focused libraries, wherein collections of small molecules with similar structural characteristics known to coordinate with a specific target are produced for screening.^{78,79} Alternatively, a less rational approach involves the synthesis and screening of small molecules libraries that are designed based on a perceived ability to modulate or perturb specific biological pathways, irrespective of any particular protein target. This is known as the “diversity oriented” (DOS)⁸⁰ approach to small molecule library design and has resulted in the identification of proteins or pathways that can serve as therapeutic targets.

The idea of generating large libraries of diverse small molecules took a step forward with the development of “biology-oriented synthesis” (BIOS), a DOS-based approach which focuses on molecular scaffolds of proven biological relevance and is heavily influenced by natural product-like structures.⁸¹ In a first example, Waldmann and coworkers synthesized four natural product-derived or -inspired libraries to identify inhibitors of several tyrosine phosphatases.⁸² These small

molecule libraries have traditionally been considered well-suited for identifying substrate mimics for the inhibition of enzymes or binding to proteins with defined pockets or grooves.

Proteome diversity

The advent of genome sequencing uncovered the complexity that exists in human molecular biology. Assuming the dogmatic “one gene = one protein” hypothesis, there are over 20,000 canonical human proteins.⁸³ However, post-translational modifications, and various polymorphisms act to increase that estimate dramatically, forming an incredibly complex array of proteins in the human proteoform.⁸⁴ Of this vast number, only ~ 10-15% of all human genes are thought to be “druggable” using small molecules,⁸⁵ and of these, only ~ 2% (260-400) have been successfully targeted with small molecule drugs.⁸⁶ Many of these untranslated protein targets participate in protein-protein interactions (PPIs) within disease pathways and consist of broad interfaces. Among all PPIs available in the Protein Data Bank (PDB), the average buried interface area of PPIs is 1600 Å²,⁸⁷ but have been reported to be as large as 4660 Å².⁸⁸ While the surface area of these binding domains are not amenable to specific targeting by individual small molecules, Wells and coworkers were successful in tethering small molecules together to facilitate the coverage of a more significant portion of the chemical landscape of the protein.^{89,90} While the individual fragments discovered using this approach are typically low affinity binders, the synergistic effect caused by tethering results in high ligand efficiency. This fragment-based approach has been an effective way to target PPIs using small molecule combinatorial libraries.

Peptides as attenuators of PPIs

Peptides are ideal candidates for interacting with the expansive interfaces of many PPIs. As miniature proteins comprised of the same amino acid monomers as their targets, they are

capable of forming high affinity, specific interactions with the broad (and oftentimes shallow) binding surfaces that are representative of many PPIs. Macrocyclic PNPs have provided a blueprint in guiding researchers in efforts to discover peptide ligands that are both bioavailable and capable of attenuating PPIs. The majority of known PNPs are synthesized by multi-enzyme complexes called polyketide synthases (PKSs) and nonribosomal peptide synthases (NRPSs). These biosynthetic multimodular complexes use thioester-activated building blocks to produce linear peptides that are subsequently cyclized intramolecularly. While the two complexes use different chemistries to link monomers, there is cooperativity between them, resulting in the generation of hybrid peptide-polyketide products and a natural expansion of chemical and biological diversity in organisms. Researchers have used their understanding of these biosynthetic pathways to produce novel hybrid molecules⁹¹ and peptide-polyketide libraries.⁹² While this approach has resulted in the development of NRPS analogs with increased activity,^{93,94} enzyme engineering is laborious and libraries created using this strategy are limited in diversity (**Figure 1.4a**).

The appeal of NRPS/PKS libraries lies in the ability of these biochemical machines to create potent, biostable products that are attractive therapeutic candidates. However, with strategies now in place to recapitulate these natural stabilizing features provided to peptides by PKs/NRPSs via chemical means, investigators adopted a more DOS-based approach to peptide ligand discovery.

Ribosomal peptide libraries

The development of powerful technologies to create and screen large libraries of DNA-encoded molecules has made peptides increasingly attractive as potential pharmaceuticals.

Moreover, these libraries can be simultaneously screened using display techniques, which use various methods to link or ‘display’ individual peptide molecules to the DNA or RNA that encodes it. This critical coupling of phenotype and genotype allows for sequence elucidation and amplification of functional sequences via polymerase chain reaction (PCR), in an *in vitro* version of Darwinian selection.⁹⁵⁻⁹⁸

Phage display

Phage are viruses that infect bacterial cells. In 1985, Smith and coworkers were the first to develop a method to display peptides on phage by cloning the sequence within gene III, which encodes a minor coat protein (pIII) of filamentous M13-derived bacteriophage.⁹⁹ The fusion protein then displays the peptide in accessible form. The $\sim 10^{10}$ unique peptides displayed using this method (**Figure 1.4a**) are panned against an immobilized biological target. Functional peptides (via associated phage) are subsequently enriched by infection and re-growth in bacteria.⁹⁷ Phage display is a widely used ligand discovery method and it has been successful in the identification of several peptides or proteins that are currently at market or in the later stages of development.¹⁰⁰⁻¹⁰²

***In vitro* display technologies**

The advent of cell-free translation systems enabled the creation of even more highly diverse peptide libraries of up to 10^{14} unique sequences (**Figure 1.4a**). *In vitro* display methods are unaffected by the transformational constraints that limit phage display library diversity and as a result, are capable of generating libraries that are 3-4 orders of magnitude more diverse. Furthermore, the incorporation of ncAAs and post- or cotranslational cyclization chemistries must be compatible with phage and the accompanying bacterial expression. Given the removal of these

constraints, the entire repertoire of the chemical biology toolbox can be utilized to create more drug-like, ultra-high diverse libraries using *in vitro* display methodologies.

Ribosome display

In 1984, Dower and coworkers developed ribosome display, the first fully *in vitro* display technique, using an *E. coli* S30 system.¹⁰³ In ribosome display a DNA library is transcribed and translated using cell-free expression systems. In the absence of nonsense codons, high concentrations of Mg^{2+} and low temperature stall the ribosome at the end of the mRNA with the displayed peptide outside of the exit tunnel tethered to the ribosome. The fully functional, accessible peptide remains associated with its genotype via the ribosome. After selective enrichment, the ternary complex is dissociated by the addition of EDTA, the mRNA is reverse transcribed and the resulting cDNA is PCR amplified to identify enriched sequences. Ribosome display has been used in the selection of high affinity cyclic peptides,¹⁰⁴ and to map areas of antibodies that are tolerant of amino acid replacement,^{105,106} resulting in the development of tralokinumab, an anti-asthmatic currently in phase 3 of clinical trials.

mRNA display

Puromycin, an antibiotic derived from the bacterium *streptomyces alboniger*, is an aminoacyl-tRNA mimic that enters the A-site of the ribosome and is transferred to the nascent peptide. The non-hydrolyzable peptide bond in the peptidyl-puromycin prevents chain extension, resulting in protein truncation. In 1997, Szostak and Roberts linked puromycin to the mRNA via an engineered linker.¹⁰⁷ The processing ribosome reaches the junction of the mRNA/ linker and stalls, allowing the tethered puromycin to enter the A-site and covalently couple to the peptide. This fusion results in the link of phenotype and genotype necessary for post-selection amplification

and sequencing. In 2000, Kurz and Lohse added a psoralen molecule to the puromycin linker, which enabled UV-induced covalent cross-linking of the linker to the mRNA,¹⁰⁸ which is then used to initiate *in vitro* translation to form the mRNA-peptide fusion library. After reverse transcription and target selection, functional cDNA is PCR-amplified and *in vitro* transcribed to form the enriched mRNA for the next round of the iterative mRNA display scheme (**Figure 1.4b**). The Suga group has used mRNA display to identify natural product-like peptide inhibitors of the serine/threonine protein kinase Akt2,¹⁰⁹ and the ubiquitin ligase E6AP.⁴⁴ The Szostak group used the tRNA misacylation technique to create unnatural mRNA-peptide fusion libraries containing 12 ncAAs and cyclized via α,α' -dibromo-*m*-xylene bisalkylation of two fixed cysteines flanking the random region of the library. Using this library to pan against thrombin, they identified drug-like peptides with low-nM K_i s.⁵⁵

In an effort to streamline the mRNA display process, the Suga group re-engineered the hybridization region of the puromycin linker and mRNA to eliminate the need for covalent association, resulting in a 10-fold time-savings relative to a round of mRNA display.¹¹⁰ This method has since been used to identify macrocyclic peptide inhibitors (K_i s = 20 nM and 50 nM) of VEGFR2.¹¹¹

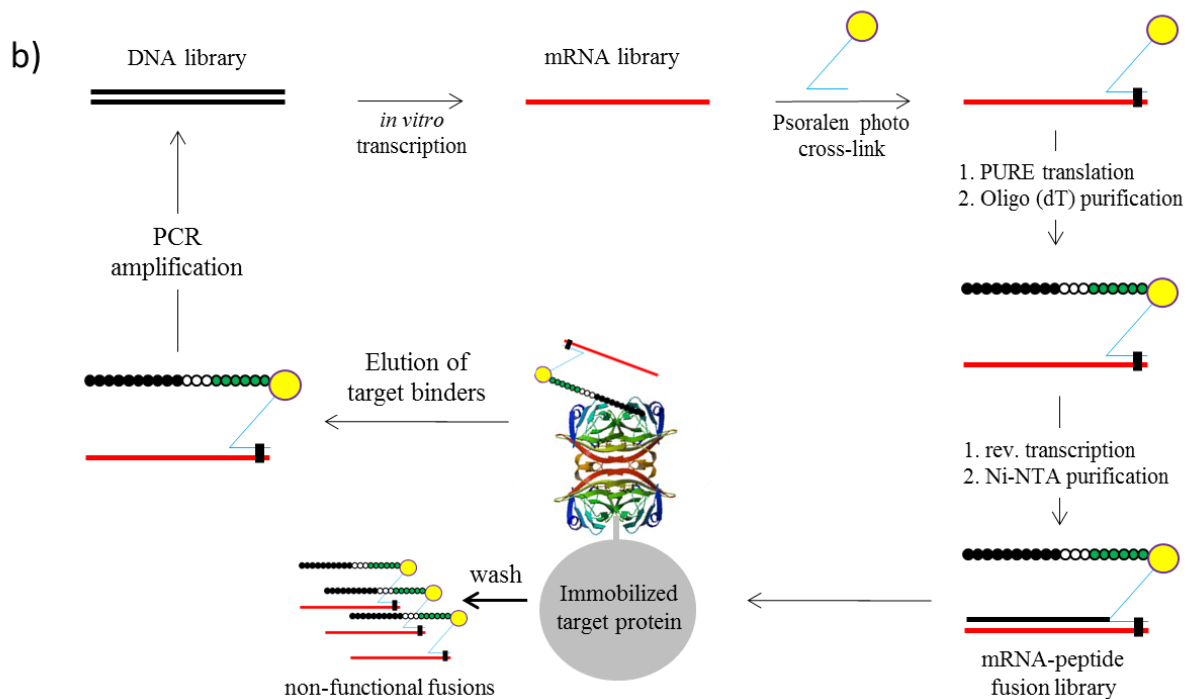
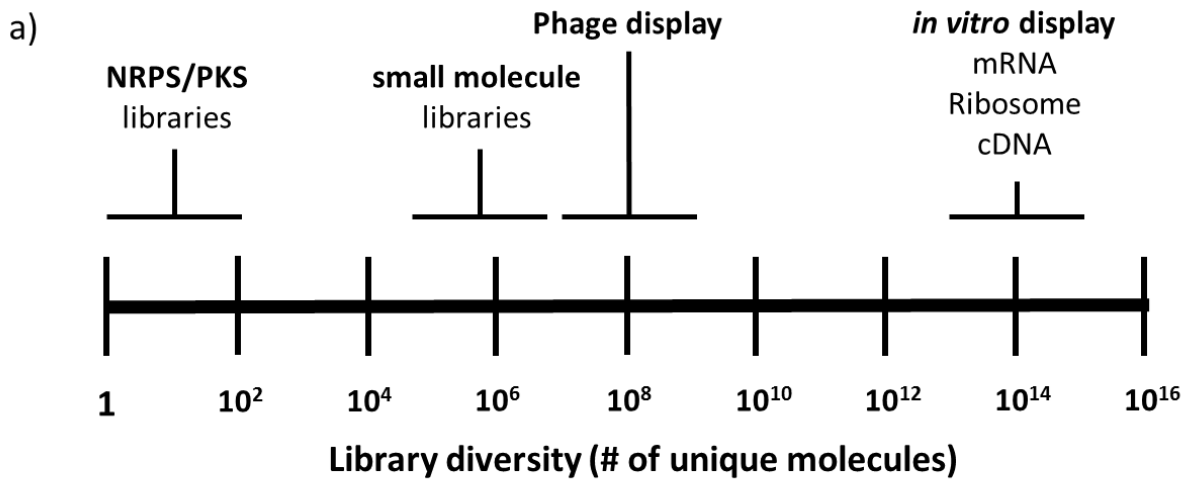


Figure 1.4 Ligand discovery technologies. a) The diversity (# of unique molecules) created using various combinatorial library generation methods. b) mRNA display scheme.

1.4 Future outlook

The rate of technological advancement and innovation related to drug discovery in the 21st century has been astounding. From the sequencing of the human genome to the ever-expanding number of elucidated molecular structures, disease-linked pathways and validated therapeutic targets, investigators are in possession of unprecedented knowledge of human biology. Given these facts, there has been a surprising lack of peptide-based drugs derived from display-based technologies entering development pipelines. However, these technologies are in their relative infancy and the chemical biology tools are now in place to generate trillion-member libraries of increasingly drug-like peptides for the future discovery of bioactive ligands.

**Chapter 2. Ribosomal Translation of Unnatural Peptides: Incorporation of Non-Canonical
Amino Acids to Increase Biostability and Chemical Diversity**

2.1 Introduction

The expansion of the genetic code has enabled researchers to incorporate amino acids with unique functionalities into proteins and peptides to further their understanding of biological systems. One way to achieve genetic code reprogramming is to reassign sense codons to ncAAs via tRNA misacylation. This strategy relies on the surprising degree of promiscuity exhibited by wild-type aminoacyl-tRNA synthetases (aaRS). The misacylation method of generating unnatural libraries has advantages over other techniques. Unlike the flexizyme method, tRNAs can be acylated *in situ* under high turnover conditions, resulting in relatively high peptide yield. In a departure from nonsense codon suppression, which is limited to a single ncAA incorporation that coincides with the orthogonal aaRS/tRNA pair, as many as 12 ncAAs have been incorporated in a single peptide via misacylation.⁵⁵

We can create ribosomally synthesized unnatural peptides simply by substituting a proteinogenic amino acid with a ncAA charged by the same aaRS (**Figure 2.1**). However, since ncAAs are not the natural substrates for the ‘cognate’ aaRS, they typically bind the enzyme with lower affinity and ultimately result in low unnatural peptide yield relative to peptides that contain the wild-type amino acid. However, there are several ways to overcome this limitation. Increasing the concentration of the ncAA and/or the aaRS often bridges the gap between unnatural and natural peptide yield. The unnatural aminoacyl-tRNAs resulting from enzymatic charging are subsequently transported to the ribosome by elongation factor Tu (EF-Tu). Since the affinity of the unnatural aminoacyl-tRNA for EF-Tu is generally lower, the increased presence of EF-Tu results in higher peptide yield as a result of increased ncAA delivery to the ribosome. It has been found that ribosomal accommodation of analogs with bulky side chains, or β -side chains are oftentimes prohibitive,²⁶ or result in low peptide yields,¹¹² unless the ribosomal peptidyl transfer

center is genetically modified.¹¹³ However, we generally do not observe this phenomenon, presumably due to the stringency already present in the form of aaRS accommodation limitations during the aminoacylation reaction. More specifically, we maintain that if the analog can effectively bind its ‘cognate’ aaRS, then the ribosome will accommodate the aminoacyl-tRNA delivered by EF-Tu. Assuming the degree of aminoacylation is sufficiently high, the yield and fidelity of the unnatural peptide generally compares favorably to peptides containing all natural amino acids.

While the natural diversity present in the side chains of the proteinogenic amino acids allows for a wide variety of chemical interactions in the cell, there is a fairly high degree of redundancy. Specifically, there are two carboxylic acid side chains, two amides and several hydrophobic residues that are near-neighbors and will likely behave similarly when interacting with intracellular targets. Simply by substituting a validated ncAA for its wild-type counterpart, we are able to add to the diversity and explore the chemical space in unprecedented ways. Furthermore, since the inclusion of ncAAs in peptides limits protease recognition, unnatural peptides are typically more protease resistant than natural peptides.^{42,44} Hence, when this method is used to create unnatural peptide libraries, we can combine it with *in vitro* selection using mRNA display to generate trillions of uniquely diverse peptides and increase the likelihood of discovering high affinity, protease stable, biologically relevant ligands.

We utilize this strategy to validate a pool of 14 amino acid analogs with unique side chains (**Figure 2.2**) for their ability to incorporate into unnatural peptides.

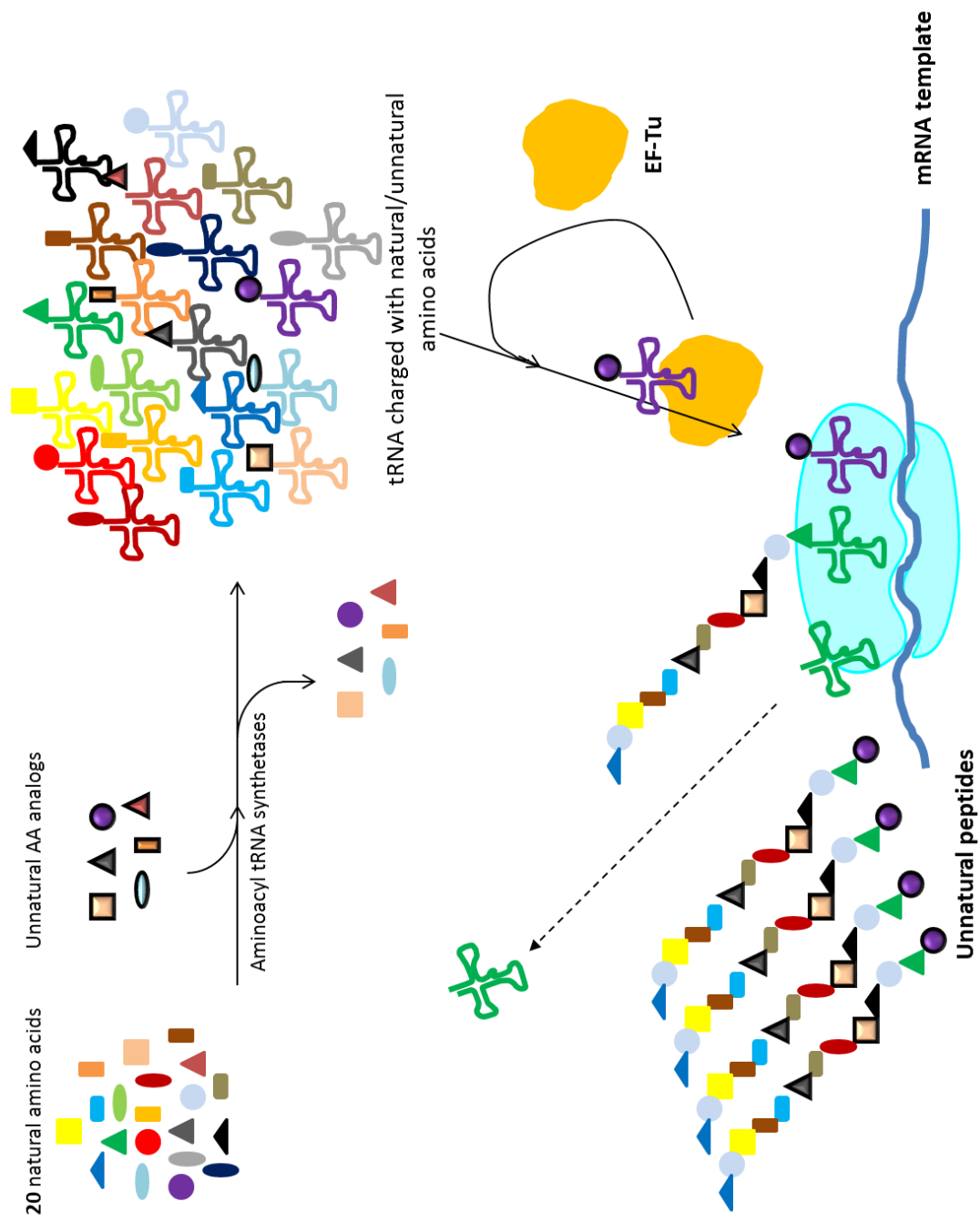


Figure 2.1 Incorporation of non-canonical amino acids in PURE translation. Unnatural peptides can be ribosomally synthesized by introducing a viable ncAA while concurrently omitting its wild-type counterpart

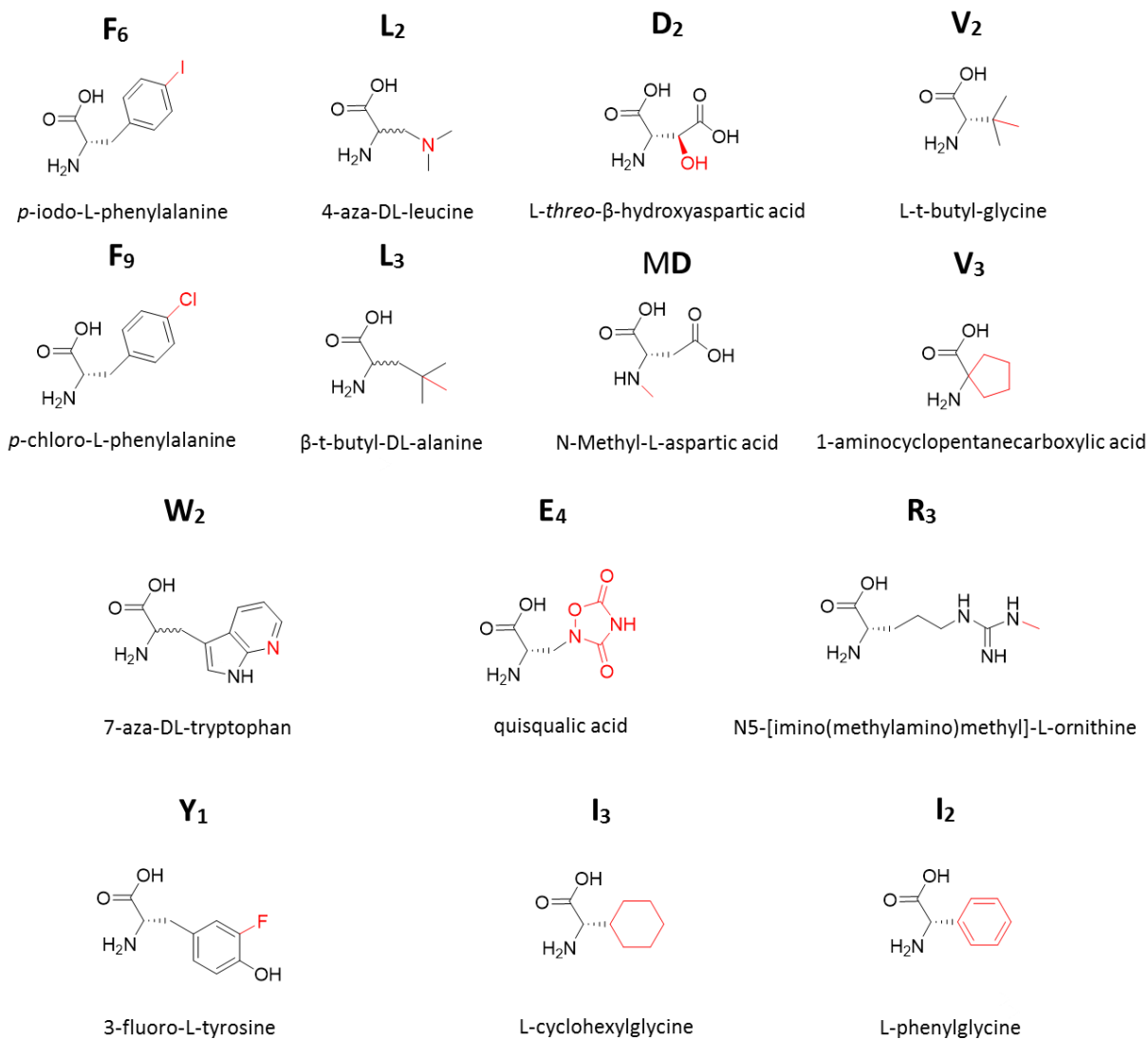


Figure 2.2 Amino acid analogs. The 14 ncAAs tested for ribosomal incorporation in translation templates. The letter above each structure denotes the aminoacyl-tRNA synthetase responsible for charging onto its cognate tRNA, and hence the wild-type amino acid each would replace in translation. Red indicates the portion of each amino acid that differentiates it from the wild-type.

2.2 Results and Discussion

To quantify the ability of each analog to be incorporated into short peptides strategically designed for analog validation (**Table 2.1**), we titrated them (0.4, 1.25 and 4.0 mM) into *in vitro* translation reactions containing ^{35}S -methionine and lacking the corresponding wild-type amino acid. While the optimized proteinogenic amino acid concentration is 0.2 mM, we have found that increasing the concentration of ncAAs often compensates for the lower affinity of the synthetase for the non-natural substrate. To confirm the fidelity of translated peptides containing the analogs, we utilized MALDI-TOF and compared the spectra to peptides containing all natural amino acids. We then identify any potential amino acid substitutions during ribosomal translation using a misincorporation table (**Table 2.2**). This approach enables the stringent testing and evaluation of ncAA candidates using both quantitative and qualitative processes, necessary components for unnatural peptide library validation.

Coding region	C-term. epitope	mRNA sequence
MHFSW	DYKDDDDK	GGGAGACCACAAACGGUUUCCUCUAGAGAAUUAUUUUUUUAACUUUAAGAAGGAGAUUAU CAUAUGCAUUUAGCUGGGGAUUUAAAAGACGACGAUGACAAAUAAGAGGGAUCCGGCUGC UAAACAAGCCCGAAAAGGAGUGAGUUGGUGCUGCCACCCTGGAGCAUAUAAACUAG
MVHM	HHHHHM	GGGAGACCACAAACGGUUUCCUCUAGAGAAUUAUUUUUUUAACUUUAAGAAGGAGAUUAU CAUAUGCACGUAAUGCAUCACCAUCACCACCAUAUUGUAGUAGGGAUCCGGCUGCUAACAA AGCCGAAAGGAAGCUGAGUUGGUGCUGCCACCCTGGAGCAUAUAAACUAG
MTINR	DYKDDDDK	GGGAGACCACAAACGGUUUCCUCUAGAGAAUUAUUUUUUUAACUUUAAGAAGGAGAUUAU CAUAUGAUUAAACCGUGAUUAAAAGACGACGAUGACAAAUAUAGGGAUCCGGCUGCUAA CAAAGCCCGAAAAGGAGCUGAGUUGGUGCUGCCACCCTGGAGCAUAUAAACUAG
MLEPQ	DYKDDDDK	GGGAGACCACAAACGGUUUCCUCUAGAGAAUUAUUUUUUUAACUUUAAGAAGGAGAUUAU CAUAUGCUGGAAACCGCAGGAUUUAAAAGACGACGAUGACAAAUAUAGGAGGAGGAUCCGGC UGCUAACAAAAGCCCGAAAAGGAGCUGAGUUGGUGCUGCCACCCTGGAGCAUAUAAACUAG
MDYKM	HHHHHH	GGGAGACCACAAACGGUUUCCUCUAGAGAAUUAUUUUUUUAACUUUAAGAAGGAGAUUAU CAUAUGGAUUUAAAUGCAUCACCAUCACCACUACUAGAGGAUUUAAAAGACGACGAU GACAAAUAUAGGGAGGAUCCGGCUGCUAACAAAAGCCCGAAAAGGAGCUGAGUUGGUGCUG UGCCACCGGAGCAUAUAAACUAG
MCFCAIKVAQRSSADVPGSGLG	HHHHHRL	GGGUAACGCCAGGGUUUCCCGAGUCAGGAGCGUUGUAAAACGCCGCCAGUGAGCGCGCGU AAUACGACUCACUAUUGGGAAUUGGAGCUCGCCGGUGCGCCGUUNUAGAUAUAGGG GUUCCCCGGGGCCACCCCAAUGGGAAUUCGCCUUUAUANGCCUCAUAUAGGGUU ACCUUAGUAGGGGGCCAGUAAAUGUCCAGUUCUUGUCCCAUCCGCCUUUCCGA UUCCGUGGUGUGUAGCUUAGCCCCCAUCCACCCTGGCUGUAGGUAG
MCRYFSLWLHEPTGCAGGSGSLG	HHHHHRL	GGGUAACGCCAGGGUUUCCCGAGUCAGGAGCGUUGUAAAACGCCGCCAGUGAGCGCGCGU AAUACGACUCACUAUUGGGAGCUCGCCGGUGCGCCGCUAGAAUAGUUGAUCCAGG GCGAAACCCGGGUCAGCCCAAUGUGAAUUCGCCUUUAUACGACUCAUAUAGGGUU AAUUUAGUAAGGAGGACAGCUAAAUGUGCAUGCAUGGAGUUCUGGGCAGCAUAUAGCUAC AUCAUCGGGUCUGUAGCUUAGGCCACCACCAUCCACCAUCCACCCTGGCUGUAGGUAG GGGAGACCACAAACGGUUUCCUCUAGAGAAUUAUUUUUUUAACUUUAAGAAGGAGAUUAU CAUAUGGUUCUUGACUUUGAAUUGGAAAUUUUAUUGCAUUAUCCACCAUUAUAGUAGG GAUCCGGCUGCUAACAAAAGCCCGAAAAGGAGCUGAGUUGGUGCUGCCACCCTGGAGCAUA AACUAG
MVLDFEWKIYM	HHHHHH	GGGUAACGCCAGGGUUUCCCGAGUCAGGAGCGUUGUAAAACGCCGCCAGUGAGCGCGCGU AAUACGACUCACUAUUGGGAGCUCGCCGGUGCGCCGCUAGAAUAGUUGAUCCAGG GCGAAACCCGGGUCAGCCCAAUGUGAAUUCGCCUUUAUACGACUCAUAUAGGGUU AAUUUAGUAAGGAGGACAGCUAAAUGUGCAUGCAUGGAGUUCUGGGCAGCAUAUAGCUAC AUCAUCGGGUCUGUAGCUUAGGCCACCACCAUCCACCAUCCACCCTGGCUGUAGGUAG GGGAGACCACAAACGGUUUCCUCUAGAGAAUUAUUUUUUUAACUUUAAGAAGGAGAUUAU CAUAUGGUUCUUGACUUUGAAUUGGAAAUUUUAUUGCAUUAUCCACCAUUAUAGUAGG GAUCCGGCUGCUAACAAAAGCCCGAAAAGGAGCUGAGUUGGUGCUGCCACCCTGGAGCAUA AACUAG

Table 2.1 mRNA templates for ncAA incorporation. mRNA sequences, with epitope tags, used to initiate *in vitro* translations to test the incorporation efficiency of ncAAs.

		Expected Amino acid																			
		G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
Substituted by	G	0	-14	-30	-40	-42	-44	-46	-56	-56	-57	-58	-71	-71	-72	-74	-80	-90	-99	-106	-129
	A	14	0	-16	-26	-28	-30	-32	-42	-42	-43	-44	-57	-57	-58	-60	-66	-76	-85	-92	-115
	S	30	16	0	-10	-12	-14	-16	-26	-26	-27	-28	-41	-41	-42	-44	-50	-60	-69	-76	-99
	P	40	26	10	0	-2	-4	-6	-16	-16	-17	-18	-31	-31	-32	-34	-40	-50	-59	-66	-89
	V	42	28	12	2	0	-2	-4	-14	-14	-15	-16	-29	-29	-30	-32	-38	-48	-57	-64	-87
	T	44	30	14	4	2	0	-2	-12	-12	-13	-14	-27	-27	-28	-30	-36	-46	-55	-62	-85
	C	46	32	16	6	4	2	0	-10	-10	-11	-12	-25	-25	-26	-28	-34	-44	-53	-60	-83
	I	56	42	26	16	14	12	10	0	0	-1	-2	-15	-15	-16	-18	-24	-34	-43	-50	-73
	L	56	42	26	16	14	12	10	0	0	-1	-2	-15	-15	-16	-18	-24	-34	-43	-50	-73
	N	57	43	27	17	15	13	11	1	1	0	-1	-14	-14	-15	-17	-23	-33	-42	-49	-72
	D	58	44	28	18	16	14	12	2	2	1	0	-13	-13	-14	-16	-22	-32	-41	-48	-71
	K	71	57	41	31	29	27	25	15	15	14	13	0	0	-1	-3	-9	-19	-28	-35	-58
	Q	71	57	41	31	29	27	25	15	15	14	13	0	0	-1	-3	-9	-19	-28	-35	-58
	E	72	58	42	32	30	28	26	16	16	15	14	1	1	0	-2	-8	-18	-27	-34	-57
	M	74	60	44	34	32	30	28	18	18	17	16	3	3	2	0	-6	-16	-25	-32	-55
	H	80	66	50	40	38	36	34	24	24	23	22	9	9	8	6	0	-10	-19	-26	-49
	F	90	76	60	50	48	46	44	34	34	33	32	19	19	18	16	10	0	-9	-16	-39
	R	99	85	69	59	57	55	53	43	43	42	41	28	28	27	25	19	9	0	-7	-30
	Y	106	92	76	66	64	62	60	50	50	49	48	35	35	34	32	26	16	7	0	-23
	W	129	115	99	89	87	85	83	73	73	72	71	58	58	57	55	49	39	30	23	0

Table 2.2 Amino acid misincorporation table. Numerical values in the table correspond to the calculated mass difference of peptides observed via MALDI-TOF when an encoded amino acid in the top row is substituted with one in the left column.

2.2.1 Ribosomal incorporation of individual analogs in hexahistidine-tagged templates

Valine analog incorporations

We have shown that L-t-butyl-glycine (V_2) and 1-aminocyclopentylcarboxylic acid (V_3) are charged onto tRNA-Val by valyl-tRNA synthetase (VRS).²⁵ To test the efficiency of ribosomal incorporation, we used the MVHMH₆M mRNA to initiate standard translation reactions which included valine (0.2 mM) or each titrated valine analog. The radiolabeled yield of V_2 was equivalent to valine incorporation, even at 0.4 mM (**Figure 2.3a**). Unexpectedly, the peptide yield decreased as concentration of V_2 in translation increased. The incorporation efficiency of V_3 was much lower (55% relative to wild-type valine), and was not rescued by increasing analog concentration up to 10-fold (58%). Since the constrained nature of V_3 makes it a less near-natural variant than V_2 , we propose that it is a less efficient substrate for VRS and subsequently there is a lower abundance of V_3 -tRNA_{Val} available for incorporation. We thus opted for V_2 instead of V_3 as the valine analog used for further study.

The fidelity of translation of the affinity purified MVHMH₆M peptide was not affected by the substitution of either of the analogs for valine (**Figure 2.3b-d**). We observed a global incorporation of the amino acid added to incorporate in response to the valine codon. However, in each of the MALDI-TOF scans we observed a series of peaks with decreasing abundance with a mass/charge (m/z) difference of -9 relative to the observed peptide. Since these peaks are also present in the peptide containing valine, we concluded that both analogs were incorporated with fidelity equivalent to that of *wt* valine. The observed -9 peak is a phenomenon that we have encountered previously when monitoring fidelity of templates containing a hexahistidine tag¹¹⁴ and we hypothesized that the peaks correspond to the incorporation of glutamine for histidine or

Q→H (this nomenclature represents the misincorporation of Q for the encoded, or expected, H and is determined from the misincorporation table in Table 2.2). However, we chose to investigate this phenomenon further.

Misincorporation of glutamine for histidine in hexahistidine-tagged peptides

The proposed misincorporation of glutamine for histidine is based on several factors: 1) The high proportional number of histidine codons in hexahistidine-tagged peptides may result in the depletion or unavailability of either tRNA_{His} and/or His-tRNA_{His} during the course of a standard translation reaction; 2) glutamine [CA(U/G)] incorporation is a 3rd base wobble insertion for histidine [CA(U/C)], the most common reason for amino acid misincorporation, and 3) glutamine insertion for histidine would have very little effect on Ni-NTA affinity of the peptide,¹¹⁵ and thus would not be selected against during the affinity purification of the peptides. To test this hypothesis, we first performed a standard translation initiated with MDYKMH₆ mRNA as a control and as expected, we observed the typical series of -9 peaks (**Figure 2.4a**). We then conducted translations in the absence of glutamine (**Figure 2.4b**) or glutaminyl-tRNA synthetase (QRS) (**Figure 2.4c**). In each case, the -9 m/z peaks were not present in the spectrum, leaving only the peak associated with incorporation of histidine in response to the histidine codon. We thus conclude that the series of -9 m/z peaks are indeed attributable to glutamine misincorporation for histidine and that the omission of either glutamine or QRS from translation would both improve MALDI-TOF sensitivity and make spectral deconvolution simpler. Since both omission of the amino acid or the synthetase yield

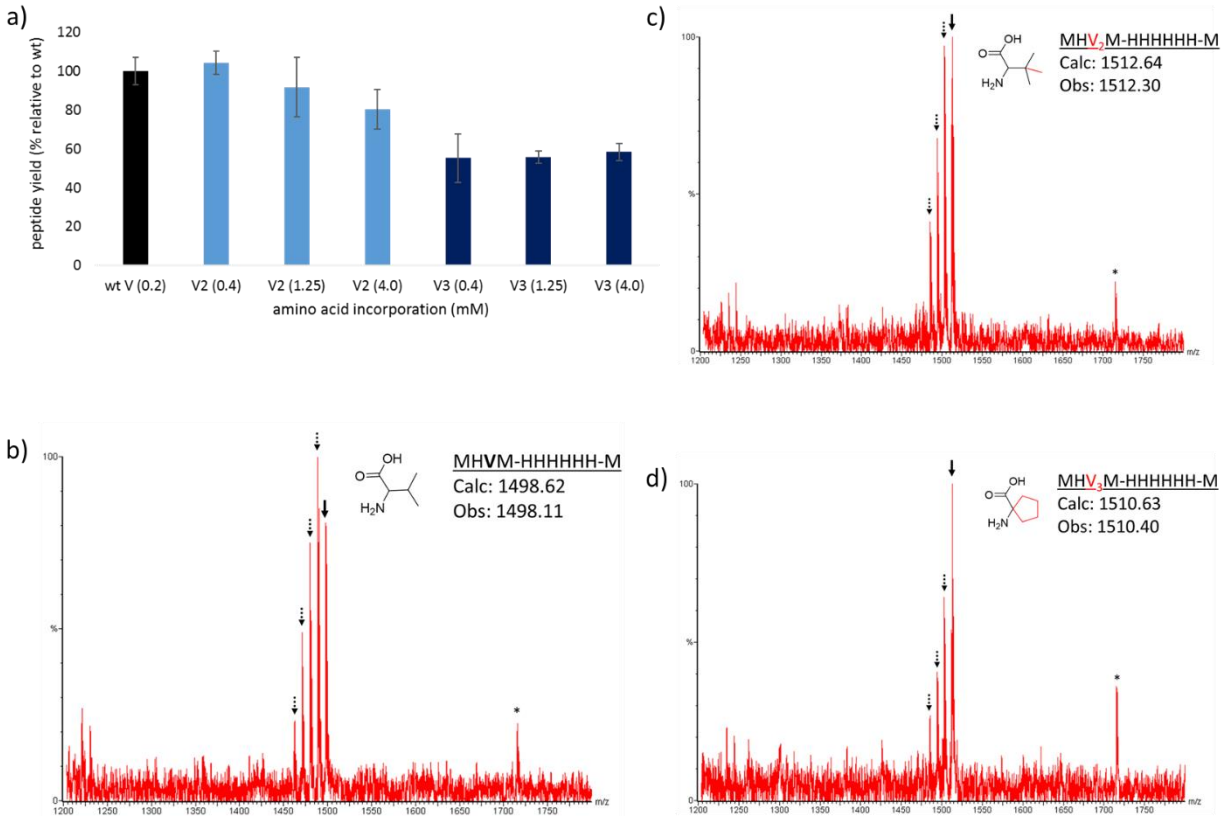


Figure 2.3 Valine analog incorporation in MHVMH₆M template. a) Relative radiolabeled peptide yield obtained when analogs were titrated in translation initiated by the addition of 1 μ M MHVMHHHHHHM mRNA. MALDI-TOF spectra of translations containing b) valine (0.2mM), c) V₂ (0.4mM) and d) V₃ (0.4mM). Solid arrows represent the expected peptide mass and dotted arrows represent the typically observed Q \rightarrow H incorporation in templates containing a hexaHistidine epitope. Asterisk (*) indicates an unidentified peak present in all spectra.

similar results, we decided to omit glutamine when necessary rather than QRS because it requires less manipulation of the PURE system. However, when the incorporation of glutamine is necessary, or when this phenomenon is observed in a standard translation containing glutamine, the -9 m/z peaks will be assigned to a glutamine misincorporation.

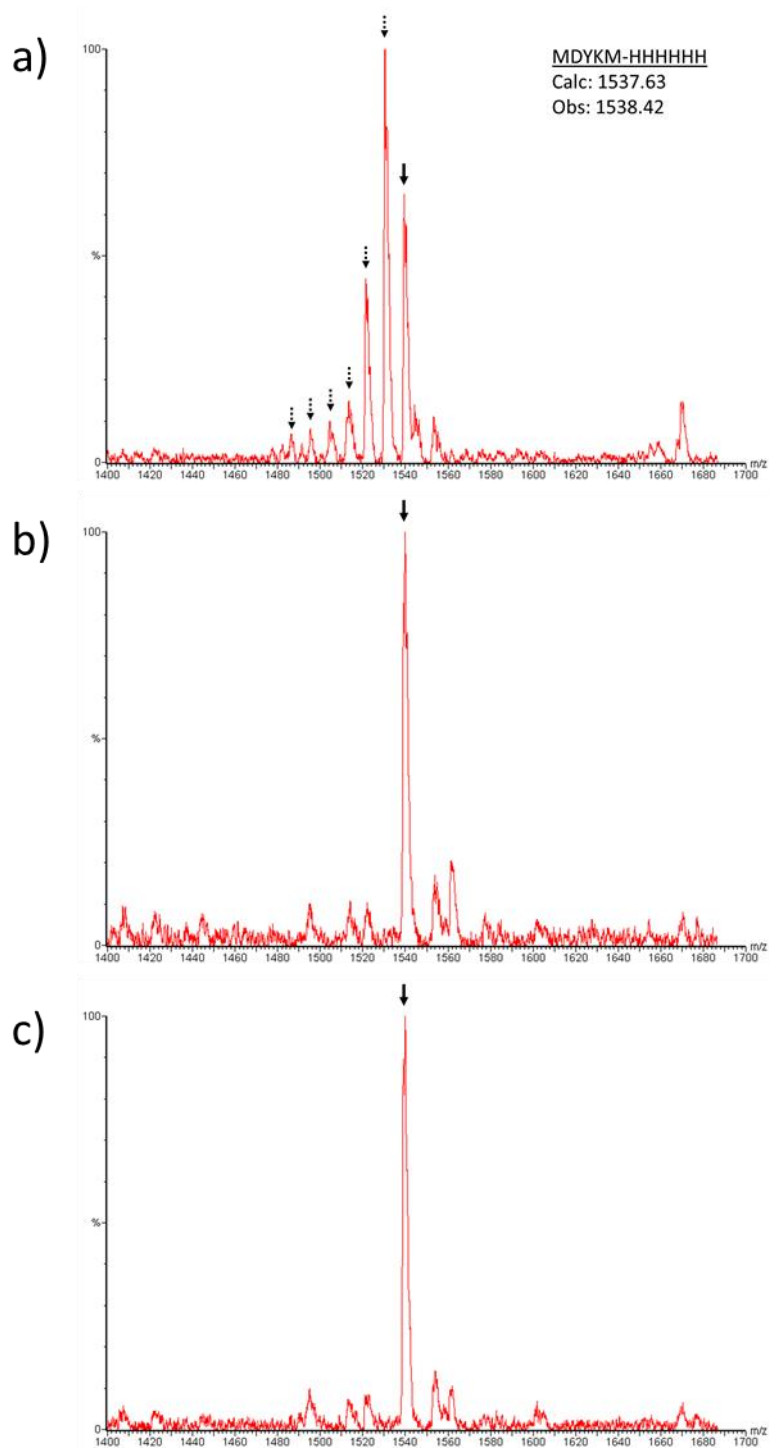


Figure 2.4 Glutamine incorporation in response to the hexahistidine codons. MALDI-TOF spectra of *in vitro* translations containing a) all amino acids and those performed in the absence of b) glutamine and c) glutaminyl-tRNA synthetase (QRS). Solid arrows represent the expected peptide mass and dotted arrows indicate the expected series of -9 m/z peaks.

Tyrosine analog incorporation

We tested the incorporation efficiency of 3-fluoro-L-tyrosine (Y_1), a tyrosine analog, in the MDYKMH₆ template. There was a concentration-dependent increase in peptide yield that resulted in 84% relative incorporation at 4 mM of Y_1 (**Figure 2.5a**). The peptide containing Y_1 was synthesized with high fidelity, with a small degree of an unexpected I/L→H misincorporation (**Figure 2.5c**). While these results warranted further investigation as a multiple incorporation candidate, we ultimately decided to eliminate Y_1 since tyrosine has been shown to be an important residue in many protein-protein interactions. However, we consider it a viable future candidate for peptide libraries containing ncAA building blocks.

Aspartic acid analog incorporation

We tested aspartic acid analogs *L-threo*- β -hydroxyaspartic acid (D_2) and N-methyl-L-aspartic acid (MD) for incorporation efficiency in the MDYKMH₆ template. These analogs offer intriguing potential to increase diversity and stability to unnatural peptides libraries. The β -hydroxyl group of D_2 may serve to provide additional polar contacts in binding interactions or even potentially act to enhance the phosphomimetic nature of *wt* L-aspartic acid, while N-methylation has been shown to increase protease resistance in peptides. We observed an analog concentration-dependent increase in D_2 yield, with 100% relative incorporation at 4.0 mM (**Figure 2.5a**). Incorporation of MD was equivalent to aspartic acid at 0.4 mM (**Figure 2.5a**).

While MALDI-TOF sensitivity was problematic with D_2 , the calculated peak containing the analog was the only significant peak detected (**Figure 2.5d**). While the yield for MD was sufficient, the observed base peak in the spectrum obtained from the MD incorporation translation agreed with the calculated mass of the peptide containing aspartic acid (**Figure 2.5e**). We

hypothesized that this was due to one of two possible causes: 1) The stock of N-methyl-L-aspartic acid contained a small amount of aspartic acid, or 2) the stock of asparagine had decomposed to form aspartic acid to some degree. In either case, even if only trace amounts of the wild-type amino acid were present, the natural substrate for the synthetase would presumably have a much higher affinity than the analog would have for the enzyme. This would result in the preferential charging of aspartic acid and hence the incorporation of the wild-type amino acid into peptides. To address these possible causes, we took two approaches. To remove any aspartic acid from the stock of MD, we performed a charging assay. When incubated with a catalytic amount of aspartyl-tRNA synthetase and tRNA_{Asp}, the natural substrate will be preferentially aminoacylated onto its cognate tRNA. The Asp-tRNA_{Asp} is then precipitated with ethanol, leaving only N-methyl-aspartic acid in solution. This decontaminated MD was then used in a translation reaction with the MDYKMH₆ template. Surprisingly, while peptide yield with decontaminated MD was higher than the wild-type peptide (**Figure 2.6a**), the charging assay approach resulted in only about 30% of the MD-containing peptide relative to aspartic acid peptide as determined by MALDI-TOF (**Figure 2.6c**). We next investigated the potential of asparagine (N) degradation to aspartic acid. Since the MDYKMH₆ peptide contained no asparagine residues, we omitted N. We performed translation reactions (without N) but in the presence of either the initial stock of MD, or with the analog that was decontaminated using the charging assay. While both assays yielded relative peptide amounts at least equal to that of the wild-type (**Figure 2.6a**), only the assay that both omitted N and used the decontaminated stock of MD resulted in a base peak that represented the peptide containing MD (**Figure 2.6e**). The two combined approaches significantly improved the fidelity of MD incorporation and we expect future optimization of the charging assay would allow us to further purify the MD stock. Since the prospect of including a backbone N-methylated analog

in our library was intriguing, we chose MD over D₂ for further study since only one analog per proteinogenic amino acid can be used to construct a library.

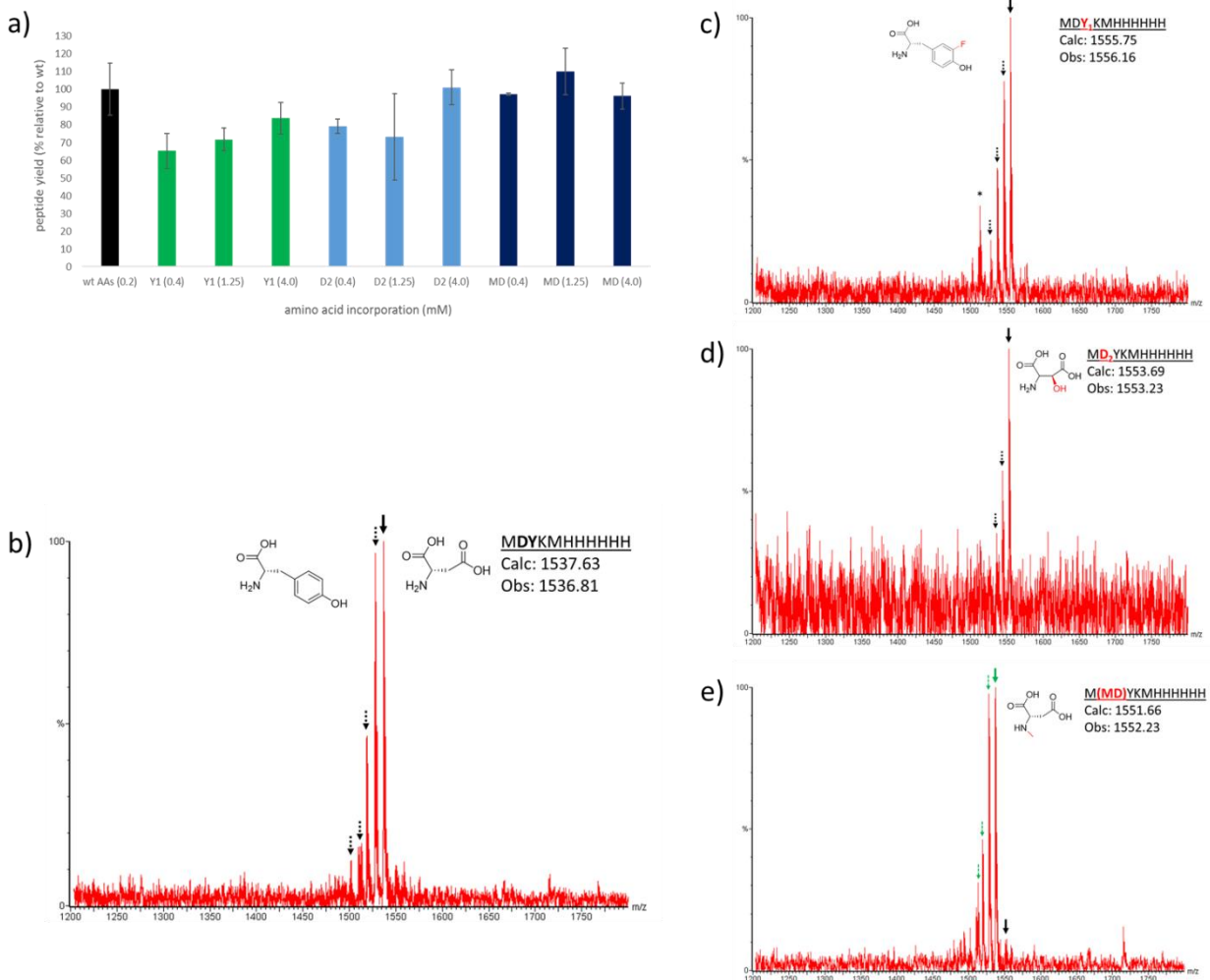


Figure 2.5 Tyrosine and aspartic acid analog incorporations in MDYKMHH₆ template. a) Relative radiolabeled peptide yield obtained when analogs were titrated in translation initiated by the addition of 1 μ M MDYKMHHHHHH mRNA. MALDI-TOF spectra of translations containing b) tyrosine and aspartic acid (0.2 mM each), c) Y₁ (4.0 mM), d) D₂ (4.0 mM) and e) MD (1.25 mM); Green arrows represent incorporation of aspartic acid. Solid black arrows represent the expected peptide mass and dotted black arrows represent the typically observed Q \rightarrow H incorporation in templates containing a hexahistidine epitope. Asterisk (*) indicates an I/L \rightarrow H misincorporation.

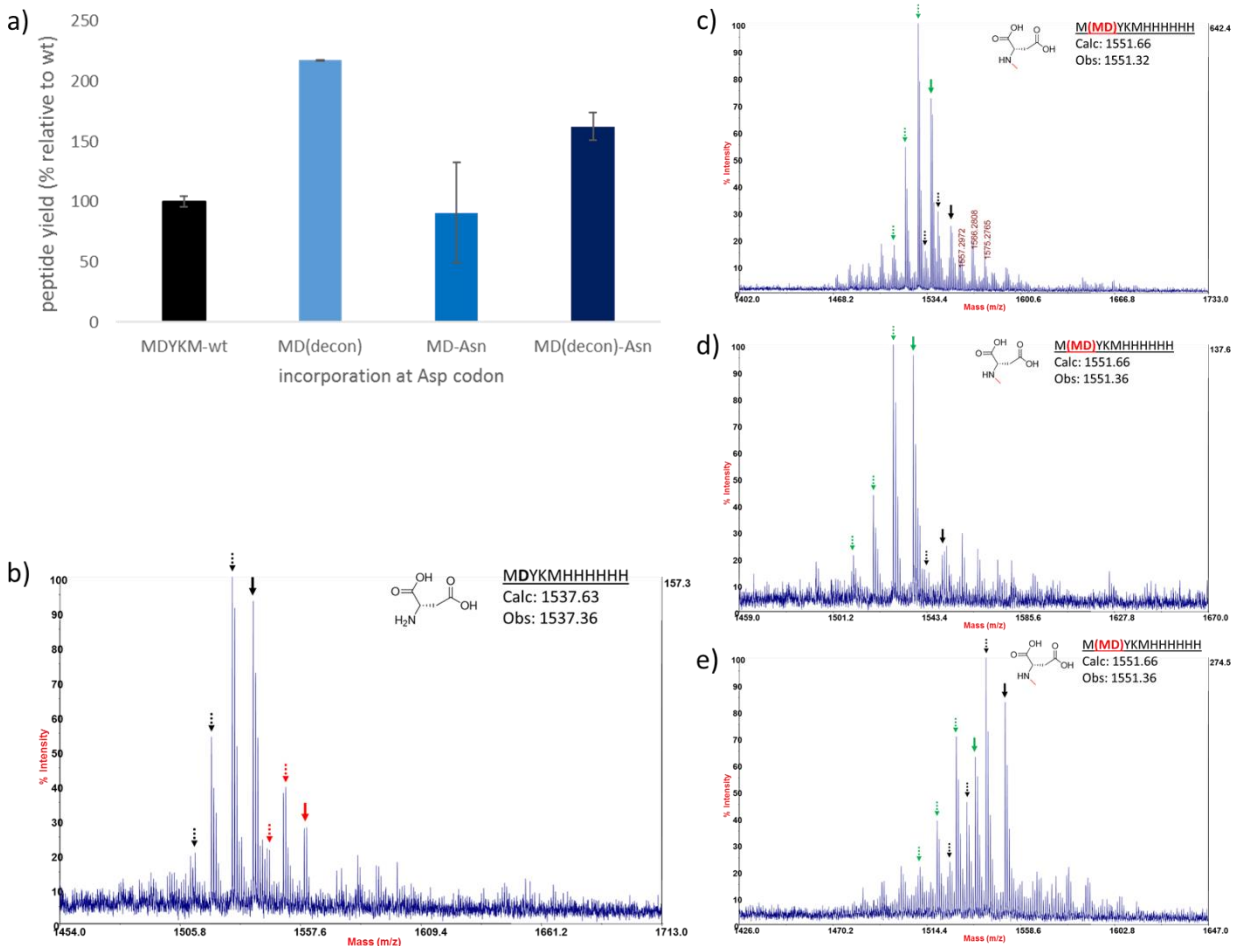


Figure 2.6 Optimization of MD incorporation in MDYKMH₆ template. a) Relative radiolabeled yield of optimized translation reactions. b) MDYKMH₆ control, red arrows indicate relative + Na⁺ peaks, c) incorporation of MD with aspartic acid contaminant removed via the charging assay, d) stock of MD with Asn omitted and e) decontaminated MD with Asn omitted. Black arrows indicate MD incorporation in the peptide, green arrows in (c-e) indicate incorporation of aspartic acid.

2.2.2 Incorporation of individual analogs in Flag-tagged templates

Leucine analog incorporations

We tested the amino acids shown to be aminoacylated by leucyl-tRNA synthetase (LRS); 4-aza-DL-leucine (L_2) and β -t-butyl-DL-alanine (L_3).²⁵ We used the MLEPQ-Flag mRNA to initiate *in vitro* translations containing either leucine or the leucine analogs. Both of these analogs are structurally very similar to leucine, so it wasn't surprising to observe a yield for L_2 (with a tertiary amine) and L_3 (with a t-butyl group like the V_2 analog) that was similar to the leucine-containing peptide (**Figure 2.7a**). L_2 (1.25 mM) was incorporated at 85% and L_3 (1.25 mM) at 100%. Unlike the inverse effect observed with V_2 incorporation, there was a direct relationship between analog concentration and peptide yield.

MALDI-TOF assessment of fidelity of L_2 showed a small degree of arginine incorporation for L_2 ($R \rightarrow L_2$), a low amount of serine incorporation for glutamic acid ($S \rightarrow E$) as well as a glutamic acid for aspartic acid incorporation ($E \rightarrow D$) that we typically observe when translating Flag-tagged peptides (**Figure 2.7c**). The addition of L_3 to translation resulted in fidelity as high as the leucine-containing peptide (**Figure 2.7d**). Although we observed wild-type levels of both yield and fidelity associated with L_3 incorporation, we had already selected a t-butyl analog (V_2) for our library and the unique nature of the tertiary amine of L_2 , along with its reasonable fidelity and yield, caused us to select L_2 for further study.

Glutamic acid analog incorporation

While many of the amino acids charged onto tRNA by their wild-type synthetases can be referred to as near-natural due to structures similar to the proteinogenic substrates, quisqualic acid (E_4) as a substitute for glutamic acid is very different and thus an intriguing candidate. The efficient

incorporation of the highly functionalized heterocyclic side chain of E₄ would bring with it interesting possibilities from a chemical interaction perspective. We observed 70% relative incorporation of E₄ at 1.25 mM (**Figure 2.7a**).

Relative MALDI-TOF fidelity was similar to the wild-type peptide (**Figure 2.7e**) with the exception of a peak that corresponded to an E₄→D incorporation in the Flag epitope, further evidence that this commonly observed substitution is indeed a wobble insertion for reasons similar to the Q→H incorporation observed in His-tagged peptides. While we presume that this substitution would have a greater impact on binding the α-Flag antibody, the fact that both are carboxylic acids and we only see a single misincorporation (even though there are 5 encoded aspartic acids in the tag) via MALDI-TOF strengthens our conclusion. Since E₄ is incorporated with high fidelity, the potential impact on library diversity presented by this unique analog led to our decision to carry this amino acid forward despite the slight decrease in yield.

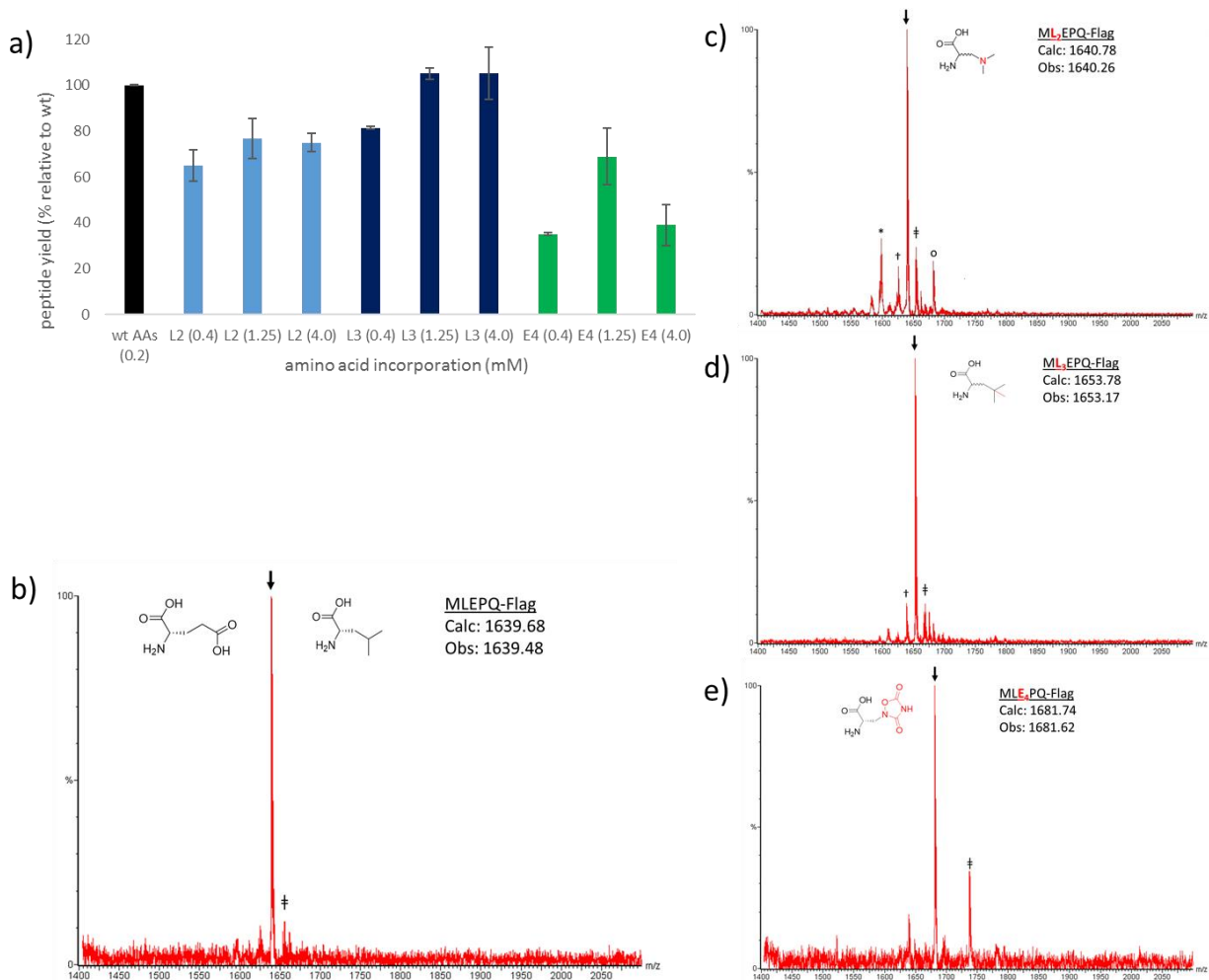


Figure 2.7 Leucine and glutamic acid analog incorporations in MLEPQ-Flag template.

a) Relative radiolabeled peptide yield obtained when analogs were titrated in translation initiated by the addition of 1 μ M MLEPQ-Flag mRNA. MALDI-TOF spectra of translations containing b) leucine and glutamic acid (0.2 mM each), c) L₂ (1.25 mM), d) L₃ (1.25 mM) and e) E₄ (1.25 mM). Solid arrows represent the expected peptide mass, (‡) represents the typical +14 m/z peak associated with E(or E₄) \rightarrow D misincorporation in templates containing a Flag epitope, (†) indicates D \rightarrow E misincorporation, (*) indicates S \rightarrow E, (o) indicates R \rightarrow L₂.

Isoleucine analog incorporations

We tested the amino acids known to be aminoacylated by isoleucyl-tRNA synthetase (IRS); L-phenylglycine (I₂) and L-cyclohexylglycine (I₃).²⁵ We used the MTINR-Flag mRNA to test the analogs relative to isoleucine. Surprisingly, these cyclic variants were both incorporated more efficiently than isoleucine, with I₂ yield at up to 150% and I₃ yield up to 116% relative to the wild-type peptide (**Figure 2.8a**). However, when translation was performed in the absence of both isoleucine or the analogs, we still observed a 106% yield, indicating that there was a surprisingly high amount of misincorporation in this template.

Fidelity for I₂ revealed the base peak from MALDI-TOF was actually an F→I₂ misincorporation and the peak corresponding to I₂ incorporation was present in approximately 80% relative abundance (**Figure 2.8d**). When F was substituted with the phenylalanine analog *p*-chloro-L-phenylalanine (F₉), we observed the disappearance of the peak associated with F misincorporation and the fidelity of the peptide containing I₂ was greatly improved (**Figure 2.8e**). MALDI-TOF fidelity of I₃ in translation showed only 10% relative abundance and misincorporations of M→I₃ (likely a 3rd base wobble) as well as a similar F→I₃ mischarging were observed (**Figure 2.8f**). To answer the question of how the peptides could have such poor fidelity yet still have a higher radiolabeled yield than the wild-type, we obtained the MALDI-TOF spectrum of the peptide translated in the absence of isoleucine or IleRS. As expected, we did not observe a peak associated with isoleucine, only the two misincorporations (M→I, F→I) seen in the presence of the analogs (**Figure 2.8c**). As a result, based on these experiments, we eliminated I₃ from consideration and chose to carry on I₂ based on the high level of fidelity when F₉ was substituted for F in translation.

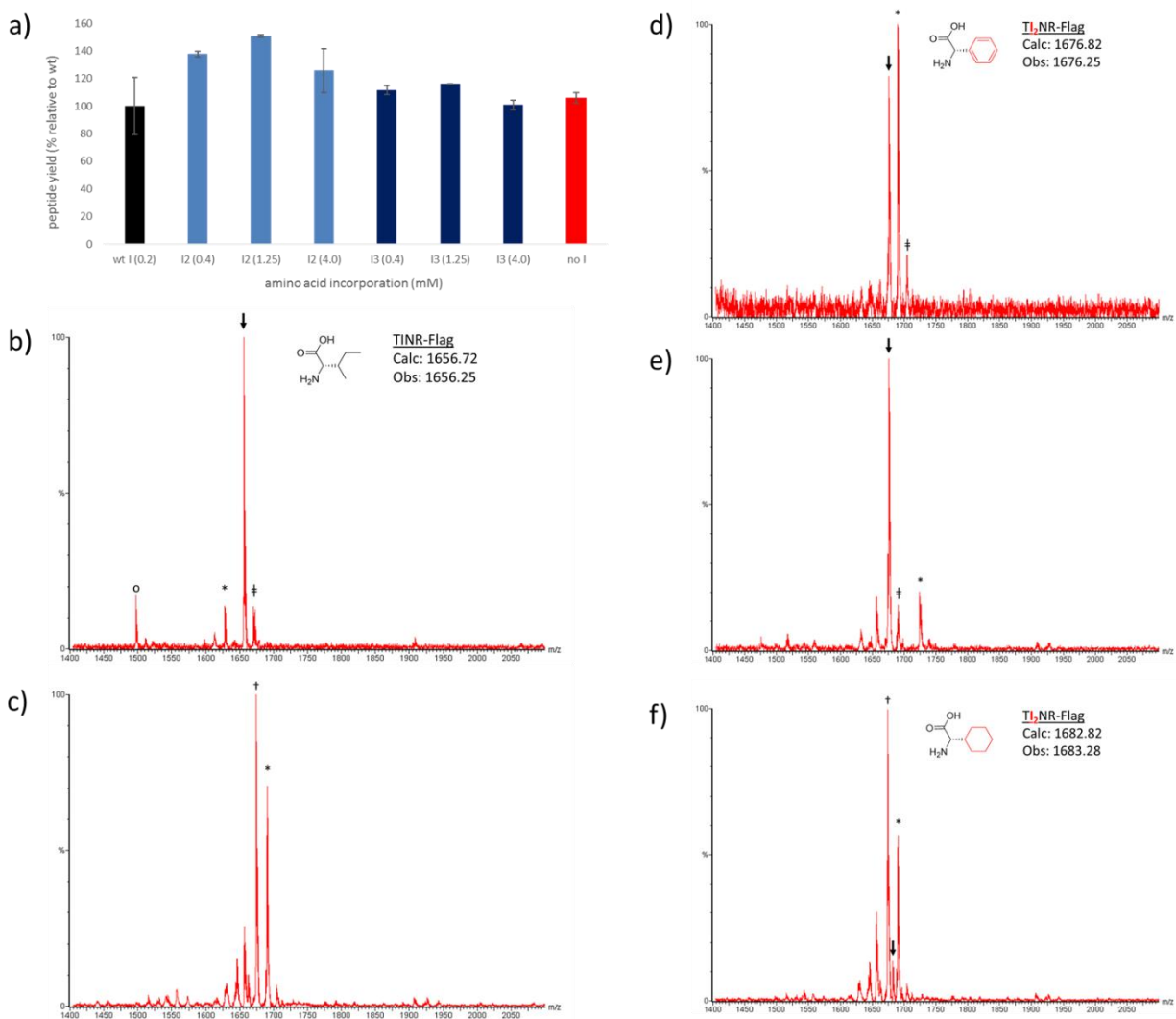


Figure 2.8 Isoleucine analog incorporations in MTINR-Flag template. a) Relative radiolabeled peptide yield obtained when analogs were titrated in translation initiated by the addition of 1 μ M MTINR-Flag mRNA. MALDI-TOF spectra of translations containing; b) isoleucine (0.2 mM), (o) is an unidentified peak not observed in analog spectra; c) no I, (*) = F \rightarrow I, (†) = M \rightarrow I; d) I₂ (0.4 mM), (*) = F \rightarrow I₂; e) I₂ with F₉ analog (minus F) in translation (*) = F₉ \rightarrow I₂; f) I₃ (0.4 mM), (*) = F \rightarrow I₃, (†) = M \rightarrow I₃. Solid arrows represent the expected peptide mass, (‡) represents the typical +14 m/z peak associated with E(or E₄) \rightarrow D misincorporation in templates containing a Flag epitope.

Arginine analog incorporation

The arginine analog, N5-[imino(methylamino)methyl]-L-ornithine (R_3) contains an interesting methylated amine in the guanidinium side chain. We tested the incorporation efficiency of R_3 in the MTINR-Flag translation template. A translation with the analog at 1.25 mM resulted in peptide synthesis at 74% relative to the peptide containing arginine (**Figure 2.9a**). The MALDI-TOF spectrum of 0.4 mM R_3 revealed a significant amount of $C \rightarrow R_3$ misincorporation, a small amount of $H \rightarrow R_3$, as well as what appears to be a peak corresponding to the incorporation of arginine (**Figure 2.9c**), presumably a contaminant in the R_3 stock similar to the earlier MD contamination with aspartic acid (**2.2.1**). When R_3 is added at 4.0 mM, C misincorporation decreases, the $H \rightarrow R_3$ peak disappears, while the relative abundance of the arginine peak remains unchanged (**Figure 2.9d**). We also observed the appearance of an unknown peak at 4.0 mM R_3 that was not present at 0.4 mM. Furthermore, when translation is performed in the absence of any programmed response to the Arg codon, peptide yield was 58% relative to the wild-type and the $C \rightarrow R$ is the base peak (**Figure 2.9a,e**). The $H \rightarrow R$, and a new $H \rightarrow N$ peak is present. To confirm the hypothesis that R_3 is contaminated with arginine, the peaks corresponding not only to peptide containing R_3 , but the peak associated with arginine disappears. While the yield of R_3 is sufficient, the degree of misincorporation makes fidelity of peptide libraries created with this analog a concern. However, our translation templates are short and somewhat limited in scope, so we decided to test this analog further due to the unique N-methylated side chain.

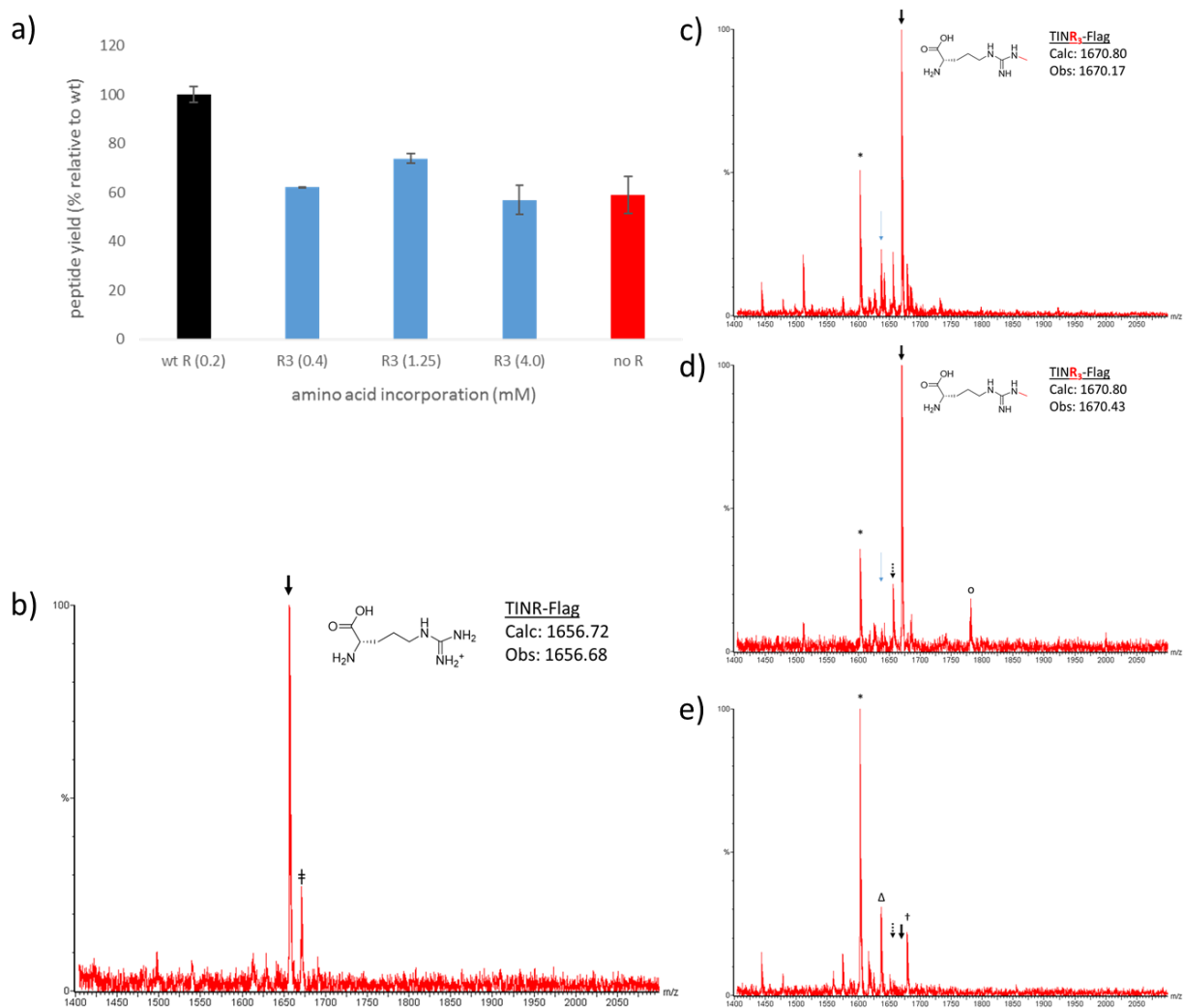


Figure 2.9 Arginine analog incorporation in MTINR-Flag template. a) Relative radiolabeled peptide yield obtained when each analog was titrated in translation initiated by the addition of 1 μ M MTINR-Flag mRNA. MALDI-TOF spectra of translations containing; b) arginine (0.2 mM); c) R₃ (0.4 mM), (*) = C→R₃, dotted arrow represents arginine contamination in R₃ stock, blue arrow indicates H→R₃; d) R₃ (4.0 mM), (o) = an unidentified peak not present at 0.4 mM R₃; e) no R or R₃, (*) = C→R₃, (Δ) = H→R, (†) = H→N. Solid arrows represent the expected peptide mass, (‡) represents the typical +14 m/z peak associated with E(or E₄)→D misincorporation in templates containing a Flag epitope.

Tryptophan analog incorporation

We tested the incorporation efficiency of the tryptophan analog 7-aza-DL-tryptophan (W_2) in the MHFSW-Flag template. Surprisingly, when we performed a control *in vitro* translation with tryptophan, we observed a MALDI-TOF base peak of the peptide associated with arginine incorporation in place of tryptophan (MHFSR-Flag) (**Figure 2.10a**). We considered that perhaps the low activity of tryptophanyl-tRNA synthetase (WRS) in our PURE system resulted in a low abundance of Trp-tRNA_{Trp}. To test this hypothesis, we attempted to rescue activity by supplementing the reaction with 5x WRS, with limited success (**Figure 2.10b**). We next omitted arginine from translation and the MHFSR-Flag peak disappeared (**Figure 2.10c**), confirming our conclusion that arginine was incorporated in lieu of tryptophan. While this is problematic for maintaining the necessary fidelity when performing selections using mRNA display, for the purposes of testing relative analog incorporation efficiency, it is a variable which can be accounted for due to the control translations performed in each experiment. When we added W_2 to translation under the same conditions as tryptophan, we observed similar effects. Based on the MALDI-TOF spectra, there is a higher relative abundance of W_2 as compared to W in a standard translation (**Figure 2.10d**) and with 5x WRS supplemented (**Figure 2.10e**). When arginine is omitted, W_2 is incorporated with equal fidelity to tryptophan (**Figure 2.10f**). Interestingly, peptide yield with W_2 in translation is unaffected by the presence of arginine or supplemental WRS (96%-105% relative to tryptophan incorporation). Given these relative results, we selected W_2 as a candidate for further testing.

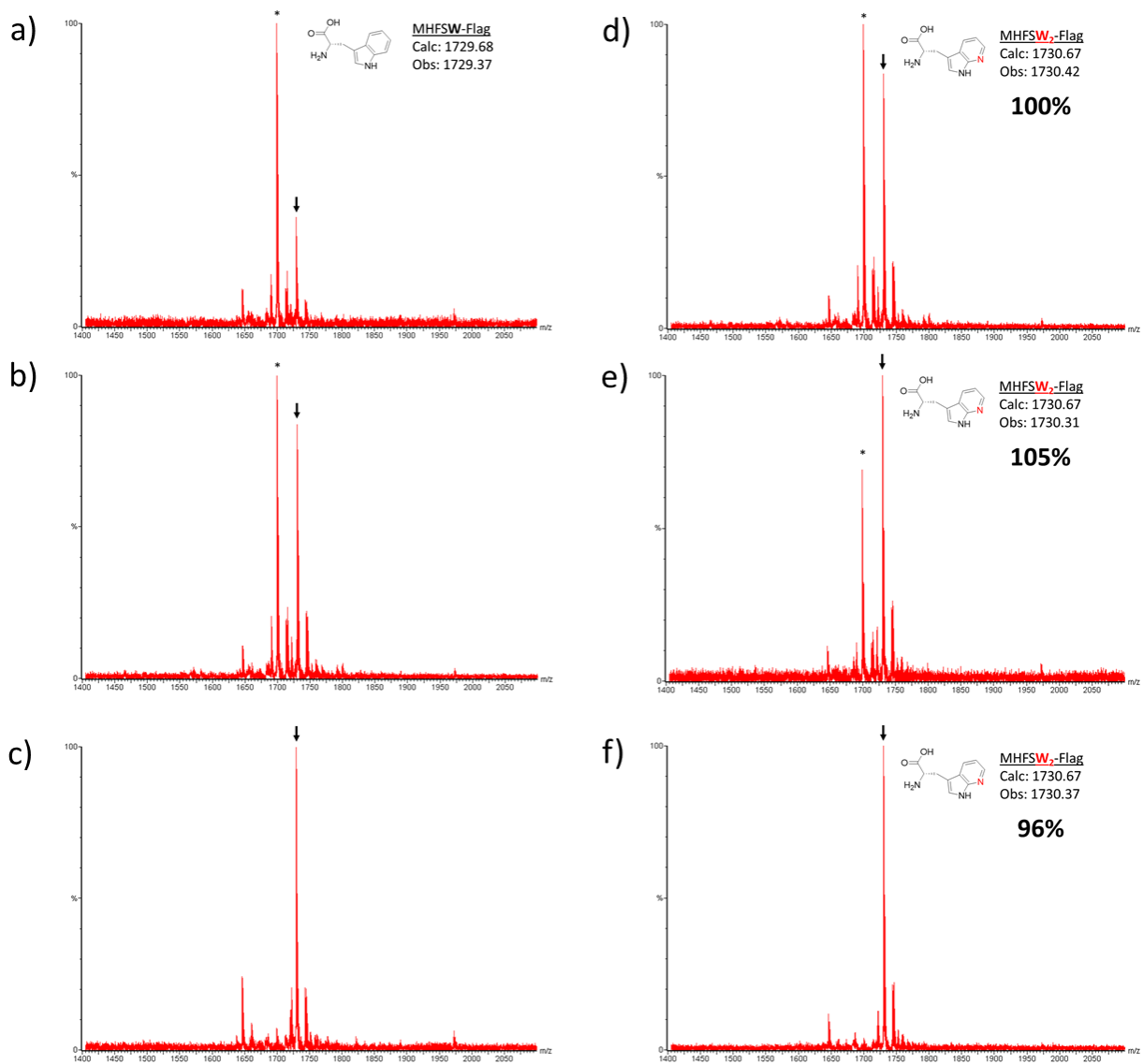


Figure 2.10 Tryptophan analog incorporation in MHFSW-Flag template. (Left column) MALDI-TOF spectra of MHFSW-Flag peptide with tryptophan in translation. a) standard *in vitro* translation, b) translation with 5x tryptophanyl-tRNA synthetase (WRS), c) translation with arginine omitted. (Right column) MALDI-TOF spectra of MHFSW-Flag peptide with 7-aza-DL-tryptophan (W₂) in translation. a) standard *in vitro* translation, b) translation with 5x tryptophanyl-tRNA synthetase (WRS), c) translation with arginine omitted. Solid arrows represent the expected peptide mass. (*) represents R→(W or W₂) misincorporation.

Phenylalanine analog incorporations

We next tested the incorporation efficiency of the phenylalanine analogs *p*-iodo-L-phenylalanine (F₆) and *p*-chloro-L-phenylalanine (F₉) in the MHFSW-Flag template. Even at 4.0 mM F₆, the translation containing the analog resulted in the synthesis of only 47% of the radiolabeled peptide relative to the phenylalanine control (**Figure 2.11a**). Similar to the experiment for W₂ incorporation in MHFSW-Flag, we observed a MALDI-TOF base peak associated with R→W misincorporation in the control reaction containing phenylalanine (**Figure 2.11b**). The spectrum of the translation including F₆ indicated the presence of peaks associated with both F→F₆ and bromophenylalanine→F₆ (previously seen by others-unpublished data) misincorporations (**Figure 2.11c**). We presume that these misincorporations are a result of contamination of the F₆ stock. Given both the low yield and lack of fidelity, we chose to eliminate F₆ as a candidate for further testing.

The relative yield of F₉ in a standard translation was 78% and the fidelity was similar to the control (**Figure 2.11d**). However, to further investigate the effect of the arginine misincorporation in this template and the effect on F₉ yield and fidelity, we performed a translation in the absence of arginine and obtained 96% relative yield with high fidelity (**Figure 2.11e**). Since we did not desire to omit arginine from our unnatural peptide libraries, and given that we previously observed that adding the arginine analog to translation resulted in the disappearance of an arginine-associated incorporation, we performed a translation with F₉ (1.25 mM) and R₃ (4.0 mM). While this approach resulted in only 50% relative yield, we observed no R₃ incorporation and overall fidelity was high (**Figure 2.11f**). As a result of these findings, we chose to carry F₉ forward for further testing.

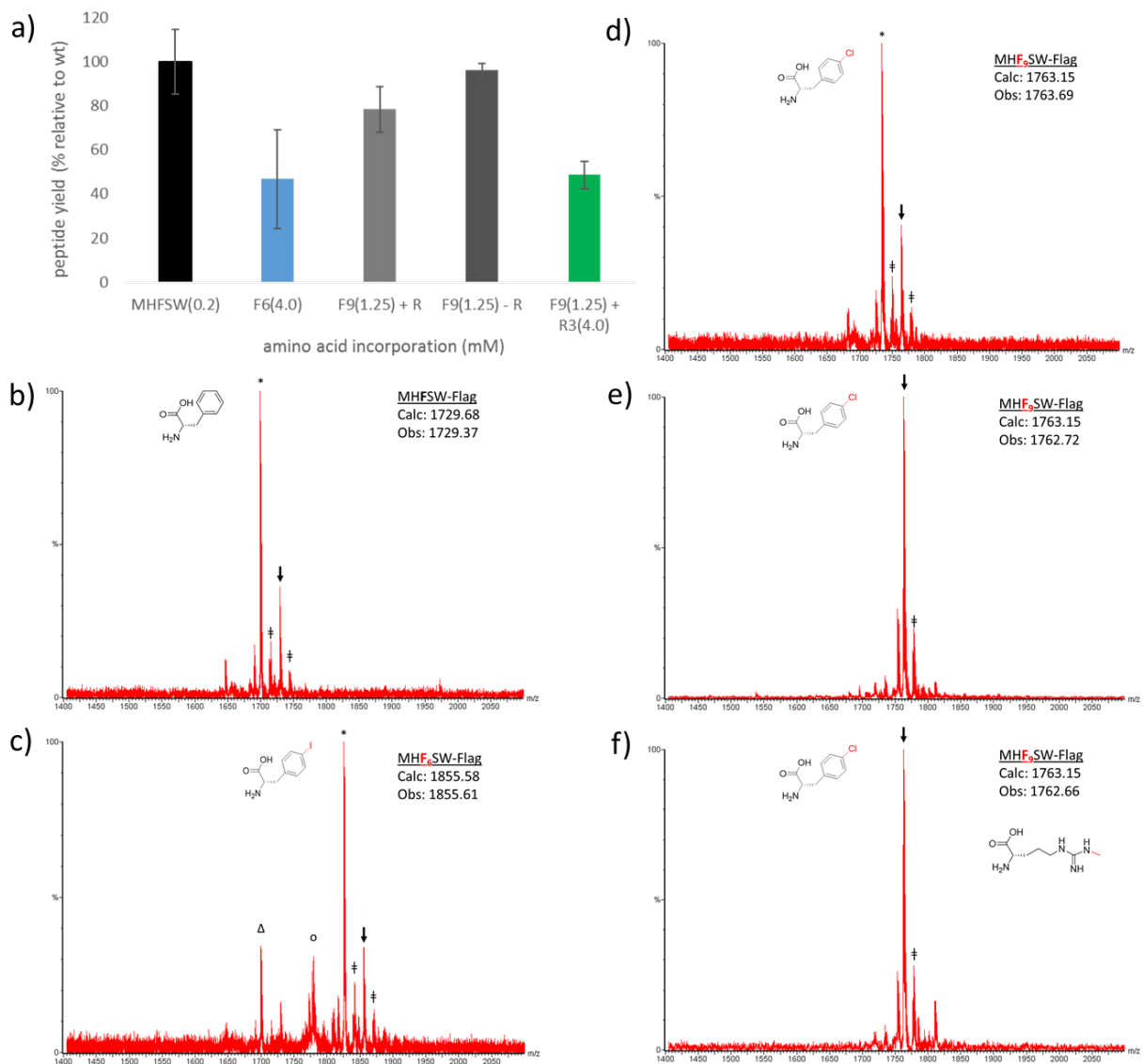


Figure 2.11 Phenylalanine analog incorporations in MHFSW-Flag template. a) Relative radiolabeled peptide yield obtained when amino acids were added to translation initiated by the addition of 1 μ M MHFSW-Flag mRNA. MALDI-TOF spectra of translations containing; b) phenylalanine (0.2 mM), c) F₆ (4.0 mM) d) F₉ (1.25 mM) with R in translation (1.25 mM), e) F₉ (1.25 mM) minus R, f) F₉ (1.25 mM) with R₃ (4.0 mM) in translation. Solid arrows represent the expected peptide mass, (*) = R \rightarrow W, (Δ) = F \rightarrow F₆, (o) = bromophenylalanine/F₆, (\ddagger) represents the typical +14 m/z peak associated with E \rightarrow D misincorporation in templates containing a Flag epitope.

2.2.3 Dual incorporations in Flag-tagged templates

Prior to testing the compatibility of the remaining analogs in a longer template, we chose to test pairs of analogs which were expressed in the same template. Analog F₉ and W₂ are encoded in the MHFSW-Flag template. We performed a translation reaction containing the optimized concentrations of the two analogs (F₉ = 1.25 mM, W₂ = 0.4 mM). The peptide containing both analogs yielded 82% radiolabeled peptide relative to the control and translated with slightly higher fidelity than the wild-type peptide (**Figure 2.12a,b**). Given the individual incorporation efficiencies of F₉ and W₂, as well as their compatibility in this template, we carried both forward for further testing.

MTINR-Flag encodes both I₂ and R₃, so we chose to test their compatibility in the same translation reaction at their optimized concentrations (I₂ = 0.4 mM, R₃ = 1.25 mM). We observed only 41% relative yield when the two were co-translated and although the desired mass was the base peak, fidelity was low (**Figure 2.12c,d**). The contamination of R₃ with arginine again appeared problematic as well as a small amount of F₉→I₂ misincorporation. Surprisingly, there was a high relative abundance of the wild-type peptide detected. Given the low yield and fidelity of this dual incorporation, we decided to eliminate R₃, given its relatively low yield and fidelity in individual testing. We chose to carry I₂ forward given its acceptable individual incorporation and since most of the issues in the dual incorporation were attributable to R₃.

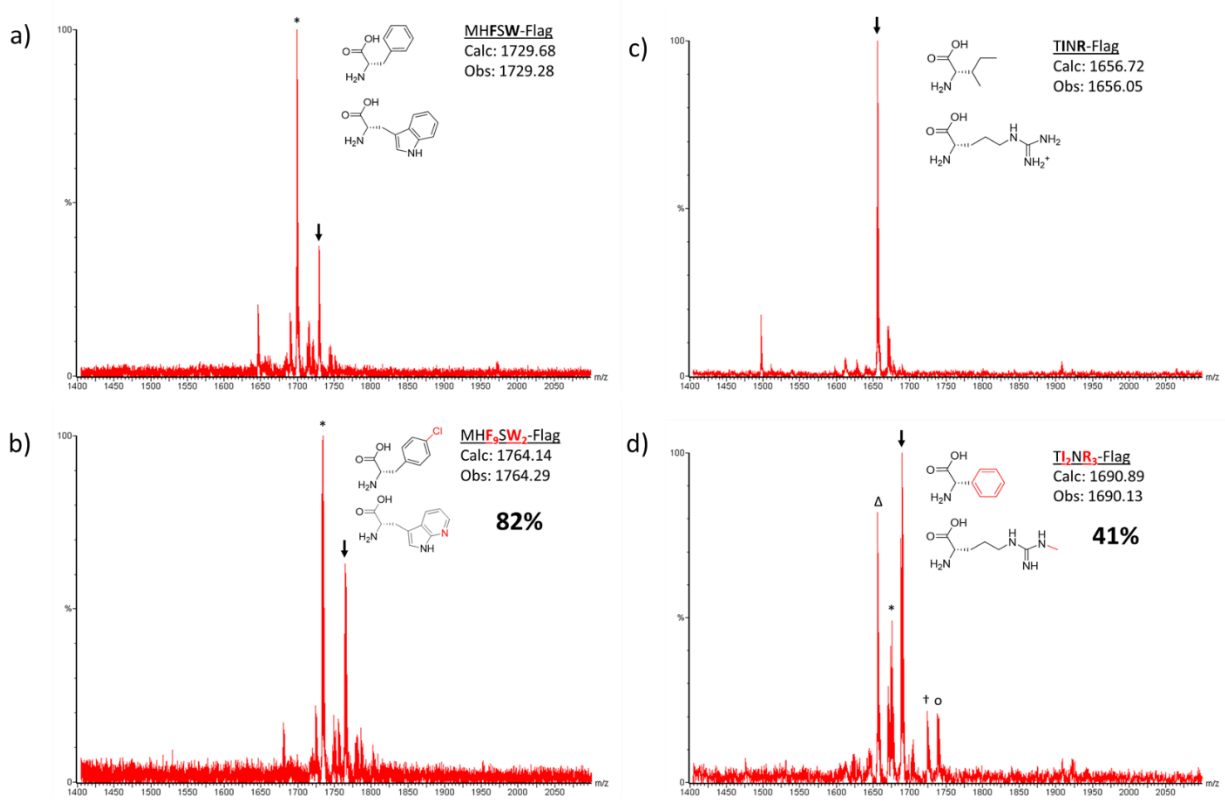


Figure 2.12 Dual incorporations in translation templates. MALDI-TOF spectrum of *in vitro* translation reaction of MHFSW-Flag peptide containing a) all natural amino acids and b) F₉ and W₂ in place of F and W, respectively. MALDI-TOF spectrum of *in vitro* translation reaction of MTINR-Flag peptide containing c) all natural amino acids and d) I₂, R₃ and F₉ in place of I, R and F. Numbers in boxes in b) and d) represent radiolabeled peptide yield (% relative to the control translation). Solid arrow represents the expected peptide, (*) = arginine misincorporations, (†) = R→R₃ and F₉→I₂, (o) = F₉→R₃, (Δ) = MTINR-Flag

Summary of individual and dual incorporations

Our initial pool of non-proteinogenic amino acids consisted of 14 analogs. After testing each of the residues individually in the translation templates, we carried 8 through to the next step of validation for unnatural peptide library creation. Since several of the analogs were encoded in the same template, we chose to test 4 of them in two sets of dual incorporation experiments using our short translation templates. This process resulted in the elimination of arginine analog R₃, which left seven remaining candidates that we chose to carry through to multiple incorporation experiments (**Figure 2.13**).

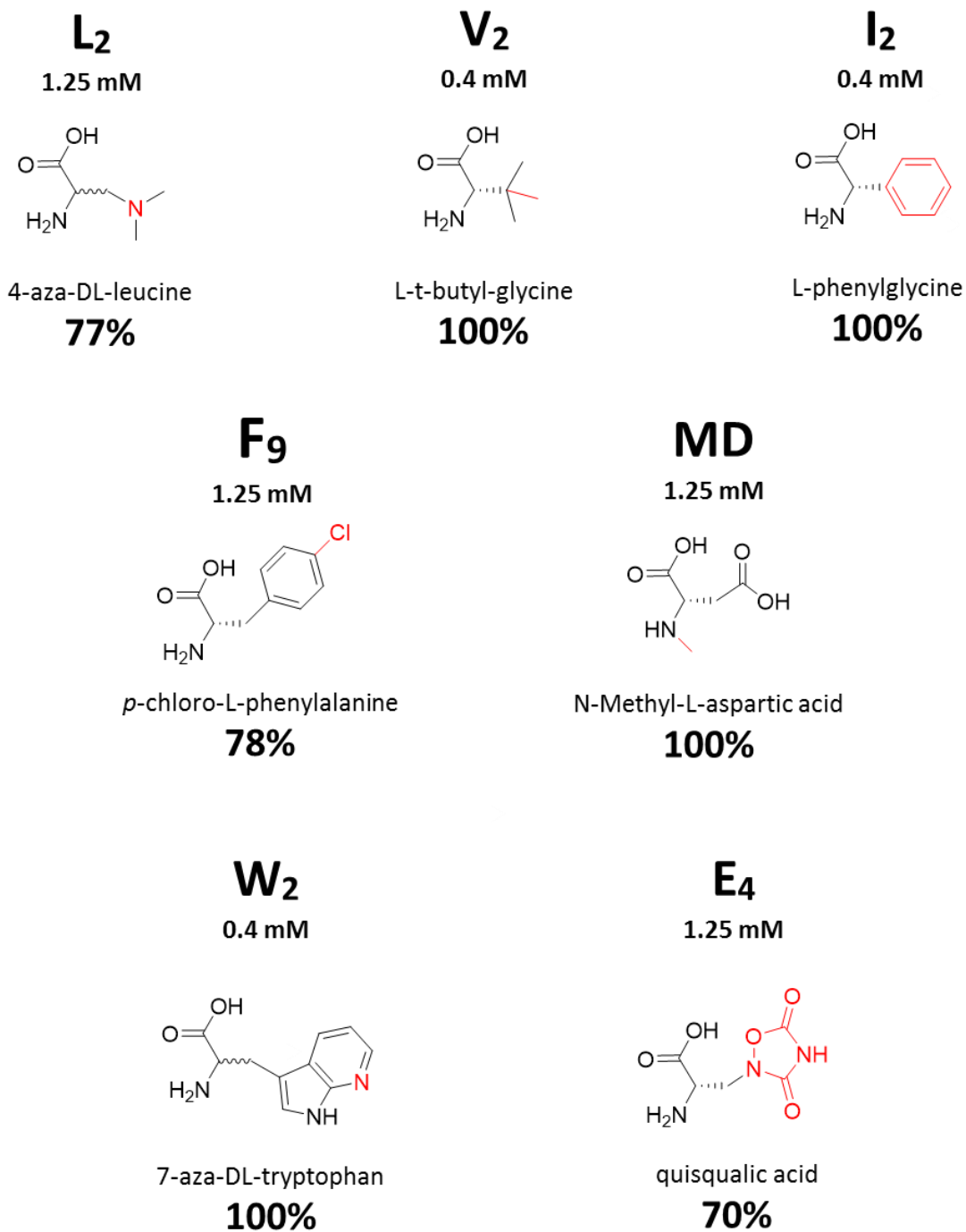


Figure 2.13 Amino acid analogs for multiple incorporations. The ncAAs that incorporated efficiently in individual and dual translations. The letter indicates the wild-type amino acid each replaces in translation along with the optimized concentration determined by titration in individual tests. The percent incorporation based on relative radiolabeled yield is below each residue.

2.2.4 Library validation: Multiple analog incorporations

Incorporation of all analogs in library templates

In creating a peptide library with non-canonical building blocks, it's critical that the residues be compatible with each other and that incorporation be robust enough to generate peptides that may contain several of the residues, either separately or in succession. In both the individual and dual tests, we were able to make assessments based on independent incorporations. To test the ability of the analogs to exist and function in conjunction with each other and all of the components of the PURE system, we used a pair of library templates previously identified in a selection in our lab (unpublished results) that included all of the analogs to be tested. The translated region of sequence E6 encoded the peptide MCFCAIKVAQRSSADVPGSGSLGHHHHHHRL and sequence C6 expressed the peptide MCRYFSLWLHEPTGCAGGSGSLGHHHHHRL (analog candidates are underlined). We performed standard 50 μ L translation reactions initiated by 0.5 μ M each of E6 and C6 mRNA. As with our individual template incorporation experiments, we performed a control reaction with all natural amino acids and a test reaction which included all 7 of the remaining analogs (and omitted the wild-type residues). 35 S-Met labeling resulted in 66% yield of the peptide containing all 7 analogs relative to the natural peptide (**Figure 2.14a**). We considered this to be sufficient for library creation, since the generation of sufficient quantities of peptide necessary to synthesize at least one copy of peptide per encoded mRNA (coverage of theoretical diversity) can be accounted for by adjusting the volume of translation. Fidelity is equally important since the phenotype must accurately reflect the sequenced genotype to ensure that the identified sequence was actually the peptide that interacted with the target. However, we were unable to obtain fidelity data. Since there were two mRNA templates in translation and some of the residues required multiple incorporations each, this was a very stringent test. Furthermore,

the peptide yield for both reactions was less than 2 pmols, glutamine was necessary, and there were likely some misincorporation events. Thus, detection failure was likely due to the quantity of peptide generated being below the sensitivity of the MALDI-TOF instrument.

Incorporation of all analogs in one template

To address the issues presented by the use of the library templates, we designed a sequence that contained all of the analogs (MVLDFEWKIYMH⁹HHHHH-analog candidates are underlined). This single sequence was 14 residues shorter than E6 or C6 and we omitted Q from the sequence. Reaction volumes were doubled and we took a more cautious approach to incorporation. Since four analogs (F₉, W₂, V₂ and I₂) were shown to be incorporated with the highest yield and fidelity, we incorporated this group in one peptide. We then created three separate peptides using this mix with E₄, L₂ and MD added individually to each reaction. A peptide containing all 7 analogs was also synthesized. The relative yield of the peptide containing F₉, W₂, V₂ and I₂ was 73% (**Figure 2.14b**). The addition of each of the remaining individual analogs resulted in an expected decrease in yield, with MD addition creating the lowest relative amount of peptide synthesized (56%). However, we were encouraged to discover that the addition of all 7 analogs to translation resulted in the highest relative peptide yield (85%).

We next investigated fidelity via MALDI-TOF. We were able to detect the desired peak of the peptide containing all natural amino acids, however the base peak represented the desired peak with an R→W misincorporation (**Figure 2.15a**). While this was an encouraging improvement from the previous experiment, we were again unable to detect the peptides containing the analogs. Glutamine was included in this experiment since we wanted to closely simulate *in vitro* selection conditions when validating our analogs. However, given the quantity of -9 m/z peptides created for each sequence (**Figure 2.15b**), it was again likely that we fell below the detection limit despite

creating 3-5 pmols in each reaction. We then omitted glutamine in our next attempt at obtaining fidelity data from our multiple incorporations. Aside from leaving glutamine out of translation, we performed the identical non-radiolabeled portion of the experiment. We were able to detect a similar spectrum for the natural peptide, albeit without the -9 m/z peaks associated with glutamine incorporation (**Figure 2.15c,d**). We were successful in detecting only the peptide containing F₉, W₂, V₂, L₂ and I₂ (5-analog) and the peptide containing all 7 analogs. The 5-analog peptide was the base peak (**Figure 2.15e**), and while there were a pair of misincorporations (R→F₉ and E→D) the results were similar to the peptide containing all natural amino acids (**Figure 2.15d**). The peptide containing all 7 analogs (5-analog plus MD and E₄) had more significant misincorporations.

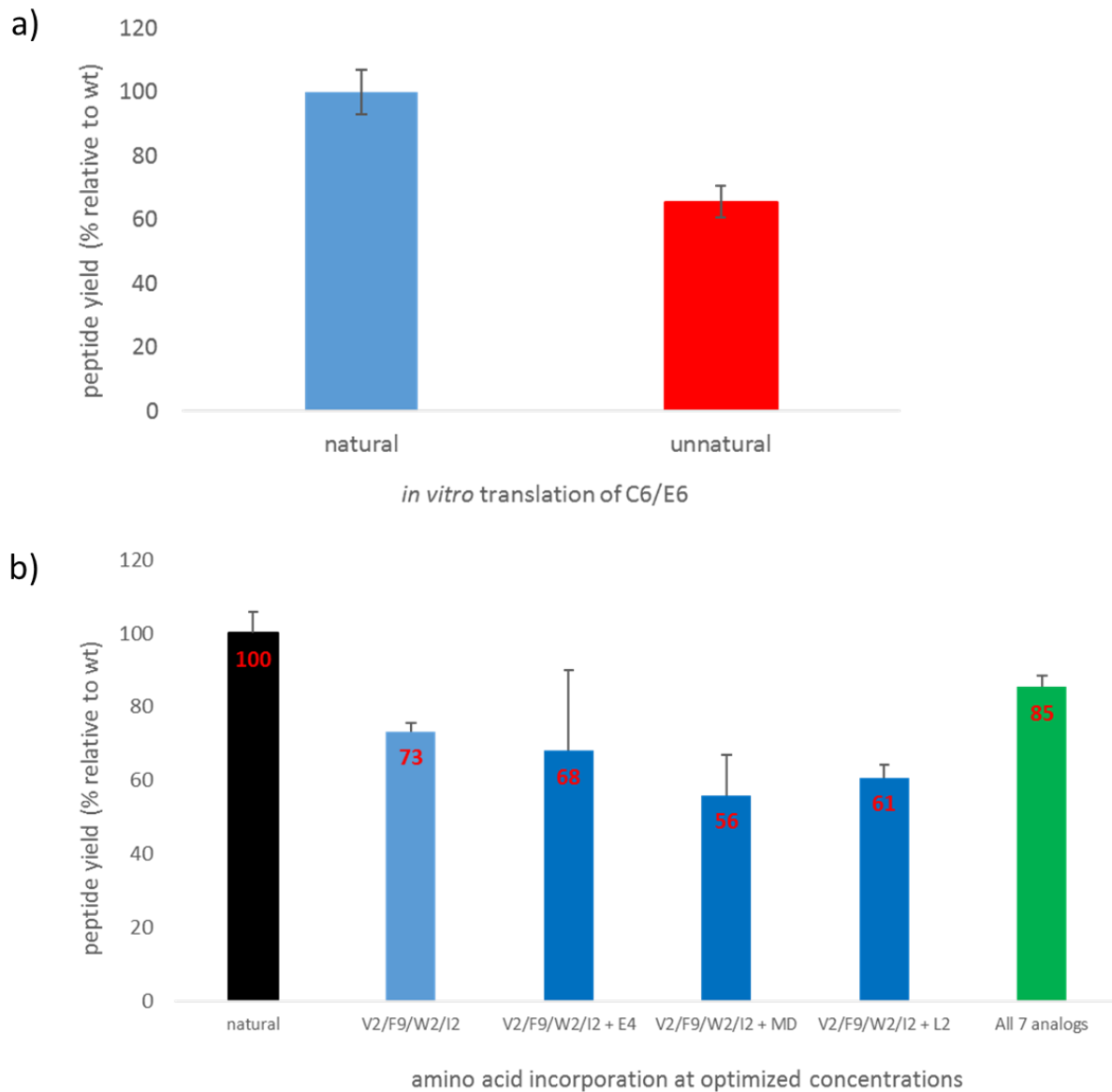


Figure 2.14 Relative quantification of multiple analog incorporations. ³⁵S-Met relative quantification of translations initiated by a) an equimolar mix of C6/E6 library templates and b) **MVLDFEWKIYMH₆** template containing the indicated amino acids. Red numbers indicate the relative yield of each reaction.

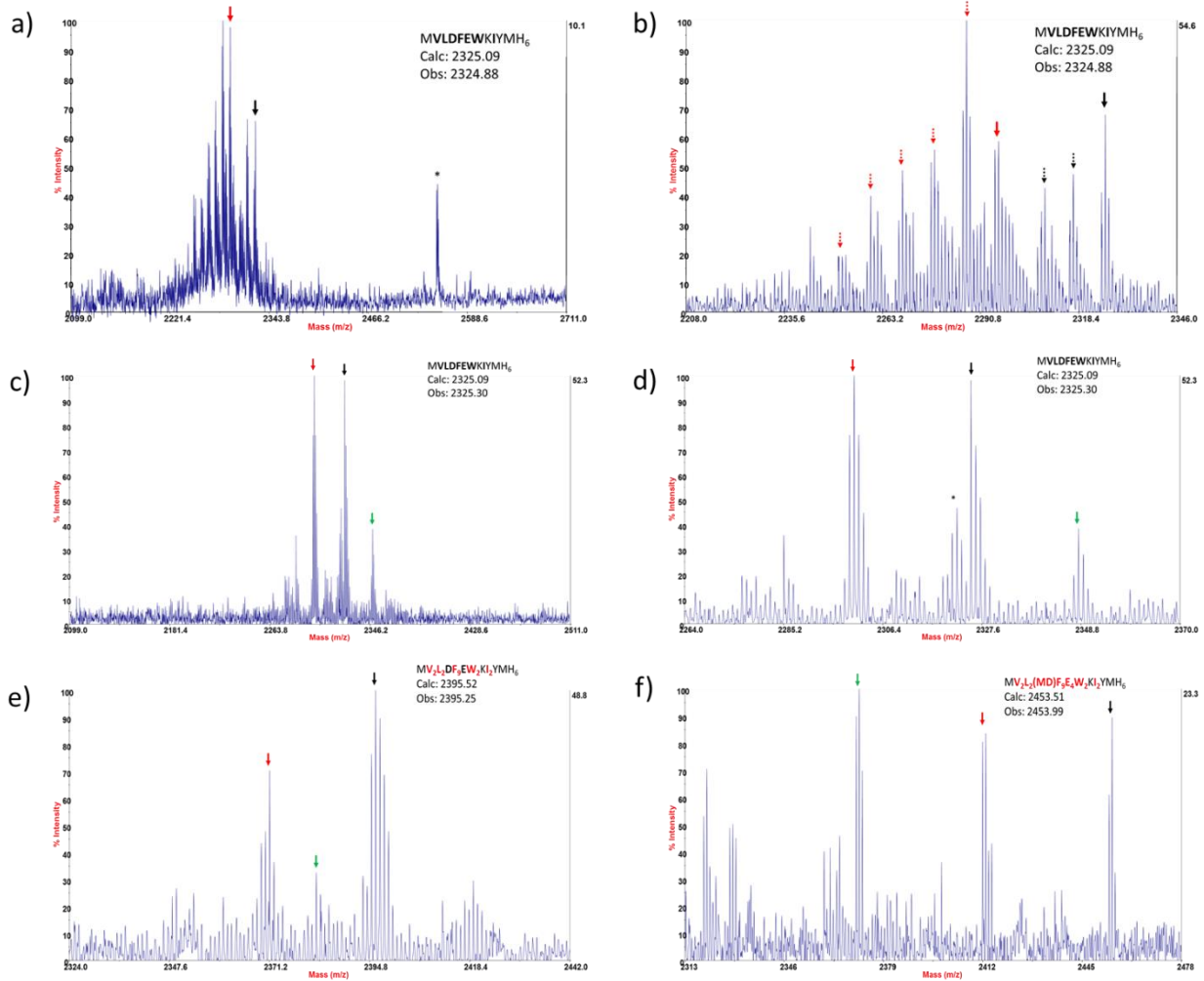


Figure 2.15 Fidelity of multiple analog incorporations in MVLDFEWKIYMH₆ template.

Black arrows represent the desired peptide peak in all cases. MALDI-TOF spectra of a) All natural amino acids, including glutamine (red arrow is R → W, dotted red arrows represent the typical Q → H, * is an unidentified peak); b) Zoom in of a); c) All natural amino acids except glutamine (red arrow is likely H → Y, green arrow is likely H → N); d) Zoom in of c); e) Translation containing F₉, V₂, W₂, I₂, and L₂ analogs (red arrow = R → F₉, green arrow is D → E); f) Translation containing all 7 analogs (red arrow is likely MD → E₄, green arrow is likely both MD → E₄ and S → MD). Black arrows represent the desired peptide peak.

2.3 Summary

The fidelity of translation that occurs in nature is remarkably accurate. The faithful expression of the genetic code ensures that proteins are accurately translated and cellular functions and homeostasis are maintained. In iterative, enrichment-based ligand discovery systems such as mRNA display, where the sequencing of the enriched DNA obtained at the end of the selection is critical in identifying the effector peptides, fidelity is equally important.

In identifying amino acid analogs that would increase the diversity and proteolytic stability of our libraries, we stringently measured and assessed fidelity via MALDI-TOF to ensure that the phenotype was representative of the encoded sequence. In our incorporation experiments, we found that each of the analogs that were carried through the individual tests was translated with high relative yield (77-100%) and fidelity that coincided with wild-type peptide synthesis. The remaining pool consisted of some intriguing functional groups, including a chlorinated phenyl ring (F₉), tertiary amine (L₂), backbone N-methyl (MD) and a multi-functional polar heterocycle (E₄). The amine-functionalized tryptophan (W₂), t-butyl group of the valine analog (V₂) and an additional phenyl side chain (I₂) all add diversity to the hydrophobic group of residues that are already present. The 5 analogs (F₉, V₂, W₂, I₂, and L₂) that incorporate with wild-type fidelity have been demonstrated to be highly compatible and ready for use in constructing an unnatural library. While yield was acceptable, the remaining two analogs (MD and E₄) appear less compatible based on relative fidelity in the MVLDFEWKIYMH₆ template. However, this template did not translate well (even with wild-type residues) and it is a highly stringent test, since all seven analogs are incorporated (with six being consecutive), a situation which would be very rare in the generation of a library for *in vitro* selection. To include these analogs in an unnatural peptide library, MD incorporation could be improved by further purifying the MD stock via the charging

assay and by using a fresh stock of pure asparagine. E₄ incorporation could potentially be improved by increasing the concentration of glutamyl-tRNA synthetase (ERS) to increase the amount of E₄-tRNA_{Glu} available for incorporation. Also, it is possible that E₄ is not a good substrate for EF-Tu and the aminoacylated tRNA is not delivered to the ribosome in sufficient quantities. While EF-Tu concentration is already fairly high (10 μM), we have shown that further increasing [EF-Tu] increases peptide yield.

We have demonstrated the ability to incorporate uniquely functional amino acids into peptides *in situ* without the laborious manipulations of the translation apparatus, such as enzyme or ribosome evolution, pre-charging or the need to chemically activate the analogs. While the incorporation of the analogs must be optimized and rigorously tested, the use of wild-type translation machinery enables a generalizable approach to the creation of unnatural peptide libraries with interesting functionalities. The uniqueness of these side chains acts to increase the diversity of our peptide libraries, makes them more protease-resistant and introduces the potential for unprecedented chemical interactions at binding interfaces that we propose will increase the likelihood of high affinity ligand discovery from an unnatural peptide library selection-based approach.

2.4 Future Directions

The 5 analogs (F₉, V₂, W₂, I₂, and L₂) that were validated for yield and fidelity can be used for the design of unnatural peptide libraries for *in vitro* selection using mRNA display. The other two analogs (E₄ and MD) that were validated in individual testing and with the 5 analogs based on yield, will require further optimization and testing prior to inclusion in an unnatural library.

2.5 Experimental

2.5.1 Amino Acids and reagents

Natural amino acids were purchased from Fluka. 1-aminocyclopentanecarboxylic acid (V₃) and 3-fluoro-L-tyrosine (Y₁) were purchased from Fluka. *p*-chloro-L-phenylalanine (F₉), *p*-iodo-L-phenylalanine (F₆), quisqualic acid (E₄), 7-aza-L-tryptophan (W₂) and 4-aza-DL-leucine (L₂) were purchased from Sigma. L-phenylglycine (I₂), L-*t*-butyl-glycine (V₂) and L-cyclohexylglycine (I₃) were purchased from Bachem. β -*t*-butyl-DL-alanine (L₃) and N-methyl-L-aspartic acid (MD) were purchased from ChemImpex. L-NMMA (N5-[imino(methylamino)methyl]-L-ornithine)(R₃) was purchased from Cayman Chemical Co. L-*threo*- β -hydroxy aspartic acid (D₂) was purchased from MP Biomedicals. All amino acids were dissolved in water to at least 10 mM, adjusted with KOH to pH 7.4-7.8 and passed through a 0.22 μ m sterile syringe filter. All amino acids, except Met and Cys (freshly prepared) were stored in aliquots at -20°C. ³⁵S-cysteine (specific activity: >1000 Ci (37.0 TBq/mmol), 11mCi/mL) was purchased from Perkin Elmer. Ni-NTA agarose was purchased from MCLAB (NINTA-400); M2 anti-Flag agarose (Sigma # A2220).

2.5.2 Preparation of mRNA translation test templates

Previously prepared pET12b vectors containing the sequences used for individual incorporations²⁶ were linearized and amplified by PCR with standard T7 forward and reverse primers. The resulting dsDNA was used as the template for T7-mediated runoff transcription as previously described.¹¹⁶ For multiple incorporations, bottom strand DNA was purchased from Integrated DNA technologies (IDT), annealed to the T7 forward primer (70 °C for 5 min., 0.1 °C/s gradient to 25 °C) and *in vitro* transcribed.¹¹⁶ All mRNA was gel-purified, electroeluted, ethanol-precipitated and resuspended in ddH₂O to a stock concentration of 50 μ M.

2.5.3 Enzymes and ribosome preparation

All PURE recombinant His-tagged proteins were expressed and purified as previously described¹¹⁷ and stored at -80° C in enzyme storage buffer (50 mM HEPES-KOH pH 7.6, 50 mM KCl, 50 mM potassium acetate, 10 mM MgCl₂, 7 mM BME, 30% glycerol) except MetRS, which was stored at -20° C in enzyme storage buffer containing 50% glycerol. Ribosomes were prepared as previously described.¹¹⁷ Inorganic pyrophosphatase, myokinase and nucleoside diphosphate kinase were purchased from Sigma. Creatine kinase was purchased from Roche. Enzyme and ribosome concentrations were determined from UV absorbance at 280 and 260 nm, respectively using extinction coefficients calculated from amino acid composition for proteins and OD₂₆₀ = 1000 being equal to 23 nmol of ribosomes.¹¹⁷

2.5.4 *In vitro* translation for incorporation of ncAAs acids in test templates

Ribosomal incorporation and optimization of ncAAs was performed using the PURE cell-free translation system which contained the following reagents: HEPES-KOH, pH 7.6 (50 mM), potassium acetate (100 mM), magnesium acetate (6 mM), creatine phosphate (20 mM), spermidine (2 mM), ATP (1 mM), GTP (1 mM), *E. coli* total tRNA (2.4 mg/mL), deacylated as previously described,¹¹⁴ (6R,S)-methenyl-5,6,7,8-tetrahydrofolic acid chloride, adjusted to pH 8 (0.1 mM), inorganic pyrophosphatase (1 µg/mL), creatine kinase (4 µg/mL), myokinase (3 µg/mL), nucleoside diphosphate kinase (1.1 µg/mL), ribosomes (1.2 µM), EF-Tu (10 µM), EF-Ts (8 µM), EF-G (0.52 µM), IF-1 (2.7 µM), IF-2 (0.4 µM), IF-3 (1.5 µM), methionyl-tRNA transformylase (0.6 µM), RF-1 (0.3 µM), RF-3 (0.17 µM), RRF (0.5 µM), RRS (0.03 µM), CRS (0.11 µM), ERS (0.32 µM), QRS (0.02 µM), IRS (0.06 µM), LRS (0.04 µM), YRS (0.04 µM), WRS (0.47 µM), DRS (0.15 µM), NRS (0.06 µM), GRS (0.02 µM), HRS (0.2 µM), KRS (0.07 µM), PRS (0.01

μM), TRS (0.05 μM), VRS (0.04 μM), SRS (0.06 μM), ARS (D777K) (0.04 μM), FRS (A294G) (0.02 μM), and the necessary amino acids for each experiment (200 μM each). Met and Cys were prepared fresh for each translation.

To determine relative analog yield, standard 50 μL translation reactions containing 0.11 μCi ^{35}S -methionine were performed as previously described,¹¹⁴ with the following exceptions: For each analog incorporation experiment, the corresponding wild-type amino acid was omitted from the reaction and the analog was titrated in independent reactions. Reactions were incubated for 1 h at 37°C and Flag-tagged peptides were quenched with 350 μL of 50 mM Tris-HCl pH 8, 300 mM NaCl (TBS). His-tagged peptides were quenched with TBS with 5 mM BME (TBS-BME). His-tagged peptides were then added to 40 μL of Ni-NTA agarose resin (MCLab) in a 500 μL microcentrifuge filter tube (VWR) and rotated at 4°C for 1h. The reactions were spun at 5900 rpm on a benchtop centrifuge, and washed thrice with 500 μL TBS-BME, centrifuged at 5900 rpm each time to remove the filtrate. The peptides were eluted with 50 μL of 1% TFA. 45 μL of each elution was added to 2 mL of EconoSafe scintillation fluid and counted for 2 min. on a Beckman scintillation counter. Flag-tagged peptides were treated similarly, with the following exceptions: TBS was used for binding and washing, 10 μL of M2 α -Flag agarose (Sigma) was used to isolate the peptides. To determine fidelity, non-radiolabeled peptides were *in vitro* translated with 200 μM methionine, affinity purified as described above and desalted by C-18 reverse phase microchromatography as previously described.¹¹⁴ Peptides were characterized using a Micromass MALDI-R MALDI-TOF mass spectrometer.

2.5.5 Aminoacyl-tRNA synthetase charging assay

To remove the aspartic acid from the stock of MD, a 2 mL reaction mixture containing 20 mM Tris-HCl (pH 7.8), 7 mM MgCl₂, 2 mM ATP, 0.1 mM EDTA, 50 µg/mL BSA, 2.5 mM β-mercaptoethanol, 2 mM N-methyl-L-aspartic acid (MD), 10 µM *e. coli* total tRNA, and 1.3 µM DRS was incubated at 25 °C for 15 min.¹¹⁸ Since the natural substrate, aspartic acid, is rapidly and preferentially charged by DRS, the resulting L-Asp-tRNA_{Asp} was selectively precipitated with 0.1 volume of 3 M KOAc (pH 5.2) and 3 volumes of ethanol. The supernatant, containing MD, was lyophilized, resuspended in water and adjusted with 1 M KOH to pH 7.6.

**Chapter 3. Cyclization Methods of Synthetic and *In vitro* Translated Peptides:
Creation of Knotted Peptides and Peptide Libraries**

3.1 Introduction

Peptides are widely considered to be promising drug candidates due to their very high specificity and activity. However, protease degradation and low bioavailability are major obstacles to the adoption of peptides as therapeutics. Cyclization is a promising approach to address these shortcomings, and the imposed constraint has also led to increased target affinity in many cases. There are many approaches currently used to create cyclic peptides,¹¹⁹⁻¹²² including the use of bromomethylbenzene linkers, originally described by Timmerman and coworkers,¹²³ to bisalkylate peptides containing two cysteines, resulting in the creation of side chain-to-side chain macrocycles. The use of tri-substituted bromomethylbenzene derivatives for use as bifunctional cyclization and labeling reagents has also been reported.¹²⁴ The bromomethylbenzene reaction has been shown to be highly specific for thiols, yet it is unknown how, or if, peptides with only one cysteine are cyclized. In this work, we probed the versatility of bifunctional bromomethylbenzene reagents and investigated the ‘cyclizability’ of peptides that contain only a single cysteine (**Figure 3.1**).

Highly constrained bicyclic peptides are structurally similar to the highly potent knotted peptide natural products, an attractive and emerging class of therapeutics with enhanced stability and bioavailability derived from their knotted structures.¹²⁵⁻¹²⁷ Knotted peptides and mini-proteins have displayed a wide scope of biological activities, including uses as antimicrobials,^{128,129} protease inhibitors^{48,130} and as highly potent modulators of a variety of ion channels.¹³¹⁻¹³³ Furthermore, knotted peptides and peptide conjugates have been adapted for use as imaging agents^{134,135} and drug delivery vehicles.¹³⁶

We envisioned a strategy for creating theta-bridged⁶² bicyclic peptides that could ultimately be used as a ligand discovery platform. However, creating ribosomally translated bicyclic peptides

is challenging. To control the independent cyclization steps to create bicyclic peptides with directable topologies, it is important that the chemistries involved be robust and orthogonal. We describe a method to create bicyclic *in vitro* translated peptides that act as structural mimics for knotted peptide natural products, using two sequential, orthogonal chemistries: bisalkylation with an α,α' -dibromo-*m*-xylene linker¹²³ and the widely applied CuAAC (click) reaction (**Figure 3.1**).^{59,137}

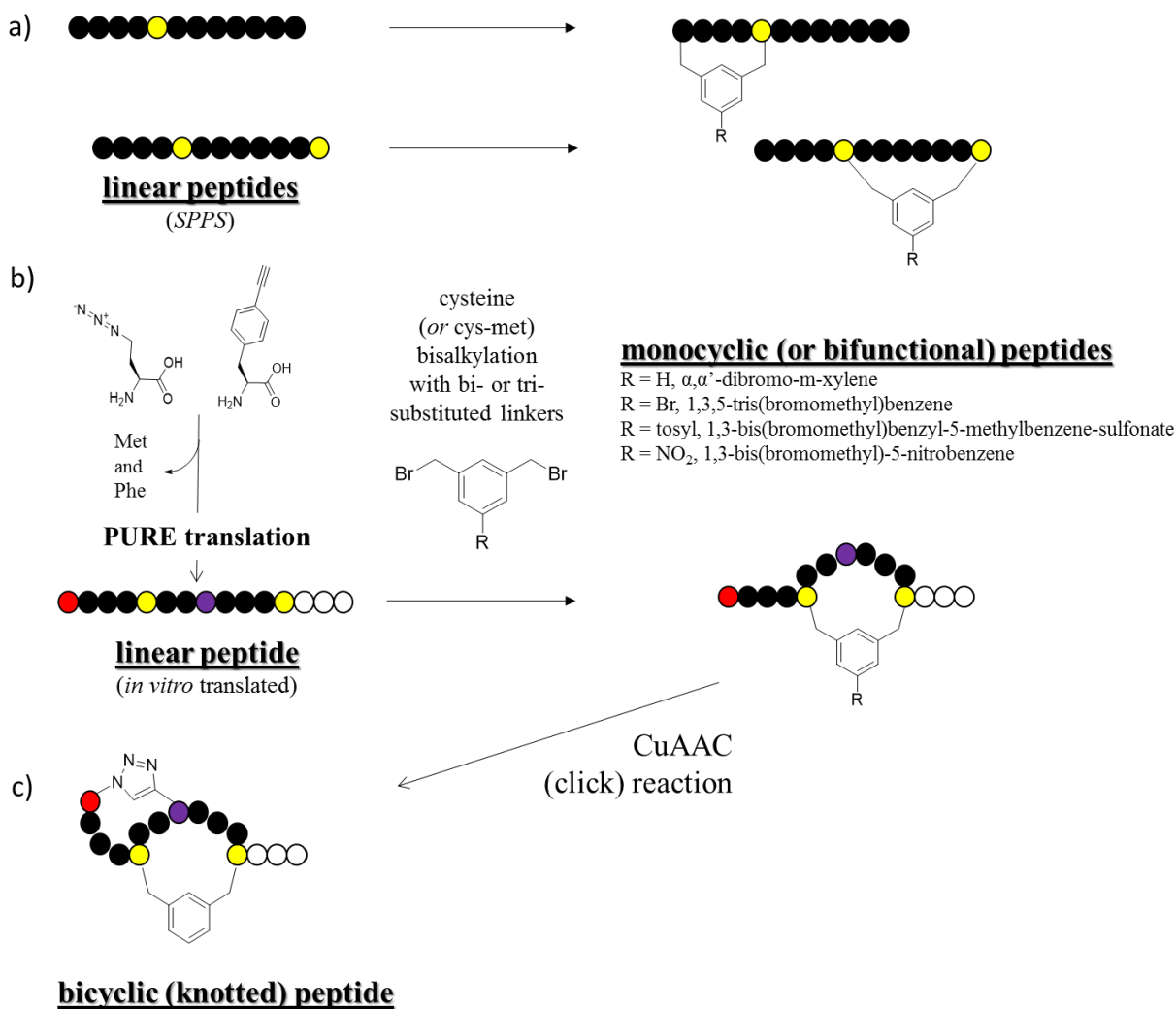


Figure 3.1 Mono- and bicyclization routes of synthetic and *in vitro* translated peptides. Bisalkylation with bromomethylbenzene linkers results in mono- and bifunctional macrocycles when reacted with a) synthetic peptides containing one or two cysteines in solution or b) *in vitro* translated peptides immobilized on oligo(dT) resin. C) Highly constrained bicyclic peptides generated from bisalkylated monocycles containing an azide and alkyne via CuAAC. Yellow circles indicate cysteine, red circles are β -azido-L-homoalanine (AzHA) and purple circles denote *p*-ethynyl-L-phenylalanine (F-yne).

3.2 Results and Discussion

3.2.1 Cyclization of solid phase synthesized peptides using bromomethylbenzene linkers

Given the robust nature of the bisalkylation of peptides containing two cysteines via bromomethylbenzene linkers, we sought to probe the degree of versatility of the reaction using a variety of derivatives and bifunctionalities. We also sought to investigate the reactivity of the linker in peptides that did not contain two cysteines.

Cyclization of a solid phase synthesized peptide containing two cysteines

We reacted a synthetic linear peptide, R2 (EACCARV1AACEAAARQ, 1 = α -amino isobutyric acid) which contained two cysteines in the $i, i + 7$ positions, with α, α' -dibromo-*m*-xylene as previously described.¹³⁸ As expected, the cyclization reaction resulted in the complete conversion of the linear peptide (**Figure 3.2a**) to the bithioether benzene-bridged cyclic peptide after 1 h (**Figure 3.2b**).

Cyclization of a solid phase synthesized peptide containing only one cysteine

The bisalkylation reaction using dibrominated benzene reagents was developed and validated (and has since been used by us and others)^{55,117,123,124,139} to produce numerous side chain-to-side chain macrocycles of peptides containing two cysteines. The reaction is highly specific for thiols. However, what would result if the cyclization reaction was performed on a peptide containing only one cysteine, but numerous other nucleophilic side chains? Others in our lab performed an *in vitro* mRNA display selection against BRCA1 with a library containing one fixed cysteine (random region = MC(NNS)₁₀).¹³⁹ The odds of the library containing a second cysteine were 40%. This design allowed for both target selection, rather than fixed codon imposition, of the optimal macrocyclic scaffold. Alternatively, if the binding interface was not amenable to a constrained ligand, the linear option remained available. After 7 rounds of selection, sequencing revealed that the most abundant peptides contained only one cysteine. A sequence from the most

abundant family, peptide 8.6, contained a single cysteine. We produced a longer Fmoc-based synthetic peptide containing a portion of the nearby hexaHistidine tag (HH), a flexible spacer (GS) along with a nuclear localization sequence (PKKKRKV) that would facilitate transportation to the nucleus to interact with the BRCA1 protein. The full length synthetic peptide 8.6 (MCNDFTFDKNLNHHGSPKKKRKV, **Figure 3.2c**) contained multiple nucleophilic candidates for a second cyclization point with the dibrominated linker (underlined residues). The mRNA display scheme for this selection included a cyclization step in the mRNA-peptide fusion preparation process.¹¹⁷ Even though there was only one cysteine in the most abundant peptides and the bromomethylbenzene reaction presumably requires two cysteines, the target (BRCA1) would have selected the peptides after the bisalkylation step. We were confident that the single cysteine in 8.6 would alkylate the linker, but which residue, if any, would alkylate at the second benzylic position? Given this uncertainty, we chose to perform the bisalkylation reaction with 8.6 prior to testing the peptide for binding to BRCA1.

The protocol for the bisalkylation reaction contains a step that requires the addition of an excess of a thiol reagent, β -mercaptoethanol (BME) to quench any unreacted linker. After workup of the reaction, MALDI-TOF deconvolution indicated a peak corresponding to the addition of BME to the peptide containing the linker, likely forming a peptide 8.6-linker-BME adduct (**Figure 3.2d**). However, when the quenching step was omitted from the bisalkylation reaction, the base peak corresponded to 8.6-linker, or cyclic 8.6. To confirm that the base peak was indeed attributable to the addition of a thiol adduct to the second electrophilic position of the linker, we used dithiothreitol (DTT) to quench the reaction, instead of BME. The base peak indicated the formation of an 8.6-linker-DTT adduct (**Figure 3.2e**), which confirmed our reducing agent adduct hypothesis. While we demonstrated the clear cyclization of 8.6, it appeared to be a transient, or

unstable, association that can be displaced by the nucleophilic quenching reagents. We hypothesized that the second cyclization point could be with the nearby methionine through the formation of a relatively unstable sulfonium ion.

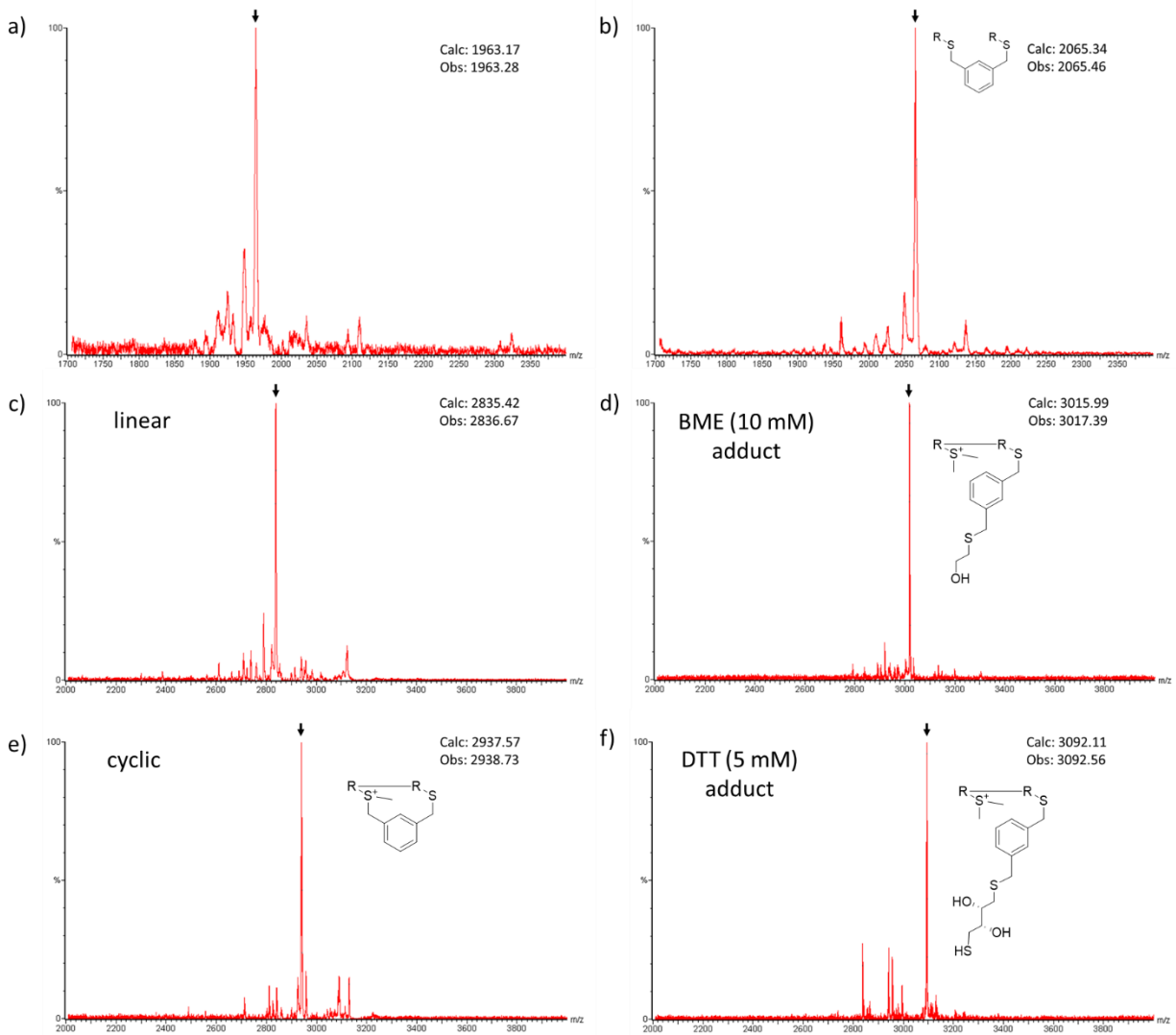


Figure 3.2 Cyclization of synthetic peptides with α,α' -dibromo-*m*-xylene. MALDI-TOF images of a) synthetic peptide R2 containing two cysteines. b) peptide R2 after bisalkylation reaction with α,α' -dibromo-*m*-xylene. c) Linear peptide 8.6, which contains only one cysteine, d) Peptide 8.6 after bisalkylation reaction with α,α' -dibromo-*m*-xylene. e) Peptide 8.6 after bisalkylation reaction with the protocol step requiring quenching of the unreacted linker with BME omitted. f) Peptide 8.6 after bisalkylation reaction with the unreacted linker quenched with dithiothreitol (DTT).

To further test this hypothesis, we performed ESI-MS/MS on cyclic-8.6, which was bisalkylated without quenching (**Figure 3.2e**). Tandem mass spectrometry, or fragmentation, is frequently used to characterize cyclic peptides.¹⁴⁰⁻¹⁴³ When a peptide bond in a linear peptide is broken, the result is two fragments whose mass can be analyzed to determine the fragmented sequences. However, when a peptide bond in a cyclic peptide is broken, the residues will still be connected via the linking residue or reagent, and no change in mass is observed. Our analysis of the fragment ions in the ESI spectrum (**Figure 3.3a**) revealed the presence of all daughter ions except those *b* and *y* ions associated with the fragmentation of the bond between Met and Cys (**Figure 3.3b**). This led us to the conclusion that the second cyclization point in the reaction of dibrominated linkers with peptides containing one cysteine is methionine (**Figure 3.3c**), although it is relatively unstable and can be disrupted or exchanged with the subsequent addition of nucleophilic reducing agents. This conclusion has since been confirmed by others.¹⁴⁴

While the cyclization protocol in selection contained a BME quenching step, it is likely, based on our tests on the synthetic peptides, that the mRNA-peptide fusions that contained only one cysteine were selected by BRCA1 in their mRNA-peptide-BME adduct form. Thus the preparation and testing of the cyclic 8.6-BME adduct construct would most closely mimic selection conditions and lead to the highest affinity peptide during testing.

The transient nature of the cyclic peptide would likely lead to ring-opening in the reducing cellular environment. In our effort to increase the stability of the cyclic peptides containing a linkage between cysteine and methionine and allow testing in reducing conditions, we bisalkylated peptide 8.6 with 1,3-bis(bromomethyl)-5-nitrobenzene. We reasoned that the electron withdrawing nitro group in the 5-position would make the 1,3 benzylic positions more electrophilic

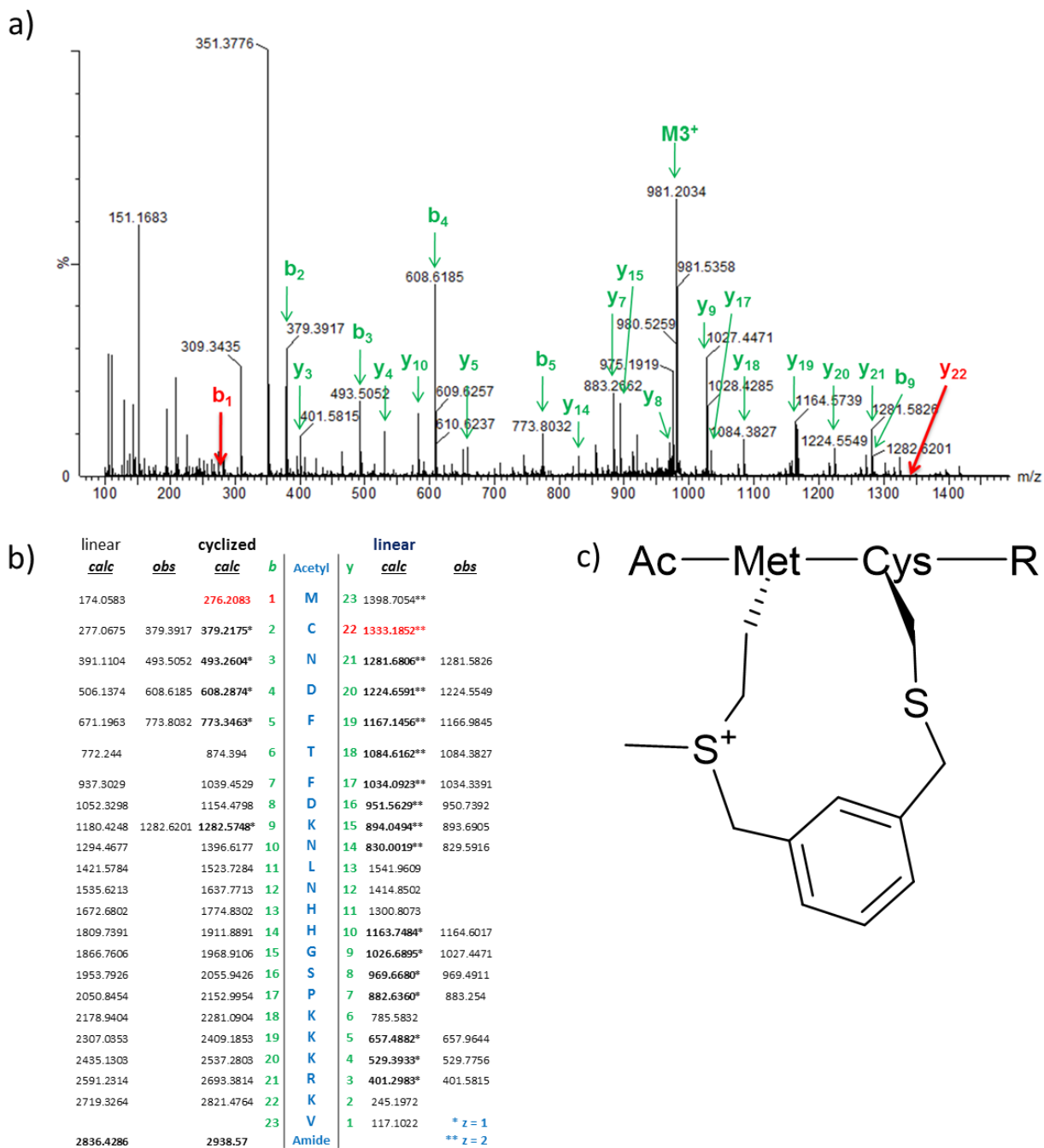


Figure 3.3 Cyclization of a peptide containing one cysteine. a) ESI-MS/MS fragmentation of peptide 8.6. b) Assignment of *b* and *y* ion peaks. The absence of *b* peak 276.21 and *y* peak 1333.19 leads to the conclusion that b) the second residue involved with cysteine in peptide cyclization is methionine, via formation of a sulfonium ion.

and thus form the sulfonium more rapidly. We observed the complete conversion of linear 8.6 (**Figure 3.4a**) to the nitrobenzyl-linked cyclic 8.6, along with an associated peak corresponding to the laser ablation-mediated conversion of the nitro to nitroso group commonly observed when characterization is performed via MALDI-TOF.^{145,146} Surprisingly, the use of this stabilizing linker enabled the cyclic peptide to survive the quenching step of the standard bisalkylation protocol intact (**Figure 3.4b**). The stability of the nitro-substituted cyclic structure compared to the unsubstituted linker likely stems from positive charge (δ^+) accumulation at the benzylic position during the ring opening reaction. These results imply that the 1,3-bis(bromomethyl)-5-nitrobenzene linker is a promising tool for use in the formation of more stable macrocycles when the bisalkylation reaction is performed with peptides containing only one cysteine.

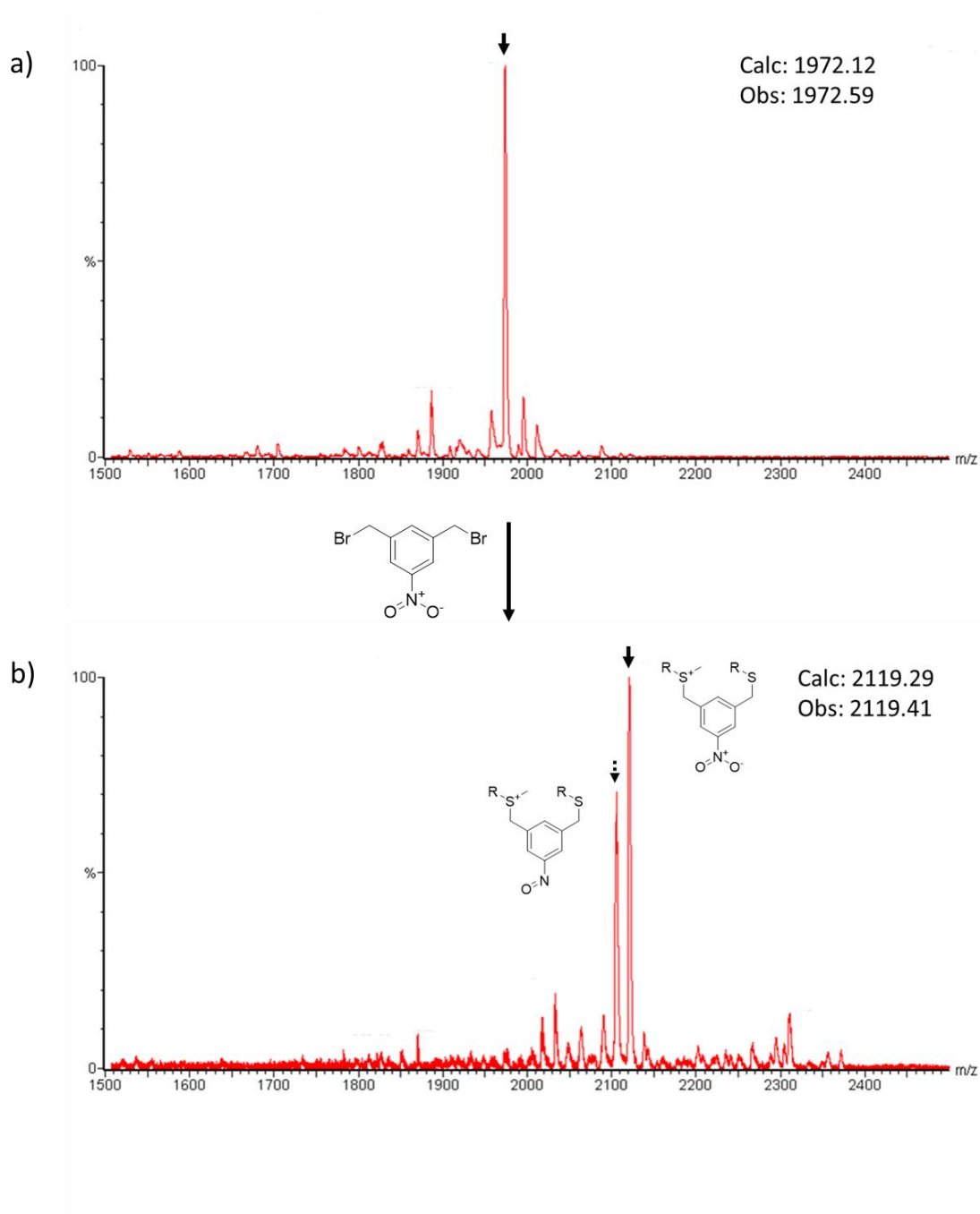


Figure 3.4 Cyclization of a peptide containing one cysteine using a stabilizing linker. a) Linear peptide 8.6, which contains one cysteine. b) After bisalkylation reaction (including quenching with BME). Solid arrow indicates the desired product; Dotted arrow represents the laser ablation-mediated formation of the nitrosobenzyl-linked cyclic peptide.

Cyclization using tri-substituted, bifunctional bromomethylbenzene linkers

Our lab has demonstrated the ability to simultaneously label and cyclize peptides containing two cysteines by using 1,3,5-substituted bromomethyl benzene derivatives,¹²⁴ with the 1,3-positions of the benzene ring containing bromomethyl groups while the 5-position remains available for a variety of potential uses. Examples include the nitro group for electron-withdrawing (**3.2.1.2**), fluorescent molecules for labeling, hydrophobic substituents to increase lipophilicity or adding an additional bromomethyl to create bicyclic peptides.⁷⁴ To explore ways to increase peptide affinity, we considered the potential of adding a tosyl group to the 5-position of the ring, as well as a 3rd bromomethyl group. We reasoned that if the functional group attached in the 5-position could survive the cyclization reaction, it could potentially serve as a way to covalently bind to proteins with a cysteine in the binding interface or any nearby nucleophile in the case of the tosyl-activated linker. We performed the reactions under mild conditions, using a low concentration of a mild reducing agent (200 μ M TCEP) to ensure the thiols were reduced and did not quench the excess linker with BME. We used synthetic linear peptide 8.5 (MCYDFDTTDHTFIGSPKKRKY) from the BRCA1 selection for our tests (**Figure 3.5a**). This peptide had only one cysteine, but based on our previous results (**Figure 3.2**) and given the mild conditions we intended for these experiments, we considered this the best option. Specifically, if there were two cysteines, the possibility was high that methionine would attack the third electrophilic position and produce the bicyclic peptide, or if the quenching step was performed, we would produce the BME adduct.

When the peptide was reacted with 1,3,5-tris(bromomethyl)benzene we detected only about 30% of the cyclic peptide with the intact 5-bromomethyl (**Figure 3.5b**). The base peak coincided with the addition of TCEP to the 5-position of the cyclic peptide (cyclic 8.5-TCEP).

Surprisingly, there was a high degree of a debrominated 5-position, presumably indicated a 3rd attachment to the peptide, forming bicyclic-8.5. Our previous work has demonstrated that the reaction was orthogonal, specific to cysteine and under mild conditions, methionine. While we did not probe the reaction further, we demonstrated that the 5-bromomethyl group could survive the reaction conditions, and further optimization (perhaps by decreasing the concentration of TCEP since the linear peptide was fully reduced and reactive) could increase the formation of the cyclic peptide with an intact bromomethyl in the 5-position.

We next reacted linear peptide 8.5 with 1,3-bis(bromomethyl)benzyl-5-methylbenzene-sulfonate under the same mild conditions. Again, the bicyclic product was present, this time as the base peak. The amount of intact tosyl-functionalized cyclic peptide was similar to the bromomethyl product. However, for both of the peaks, there was an associated larger peak that correlated with an oxidation of the peptide, presumably the thioether forming a thiosulfonate, that was not present in the reaction with tris-(bromomethyl)benzene. We did not pursue the reaction further and additional optimization would need to be performed. The utility and generalizability of a completely intact bifunctional cyclic peptide would be an intriguing development as it would open up the potential for covalent inhibition. However, the lability of these reactive leaving groups would necessitate mild test conditions.

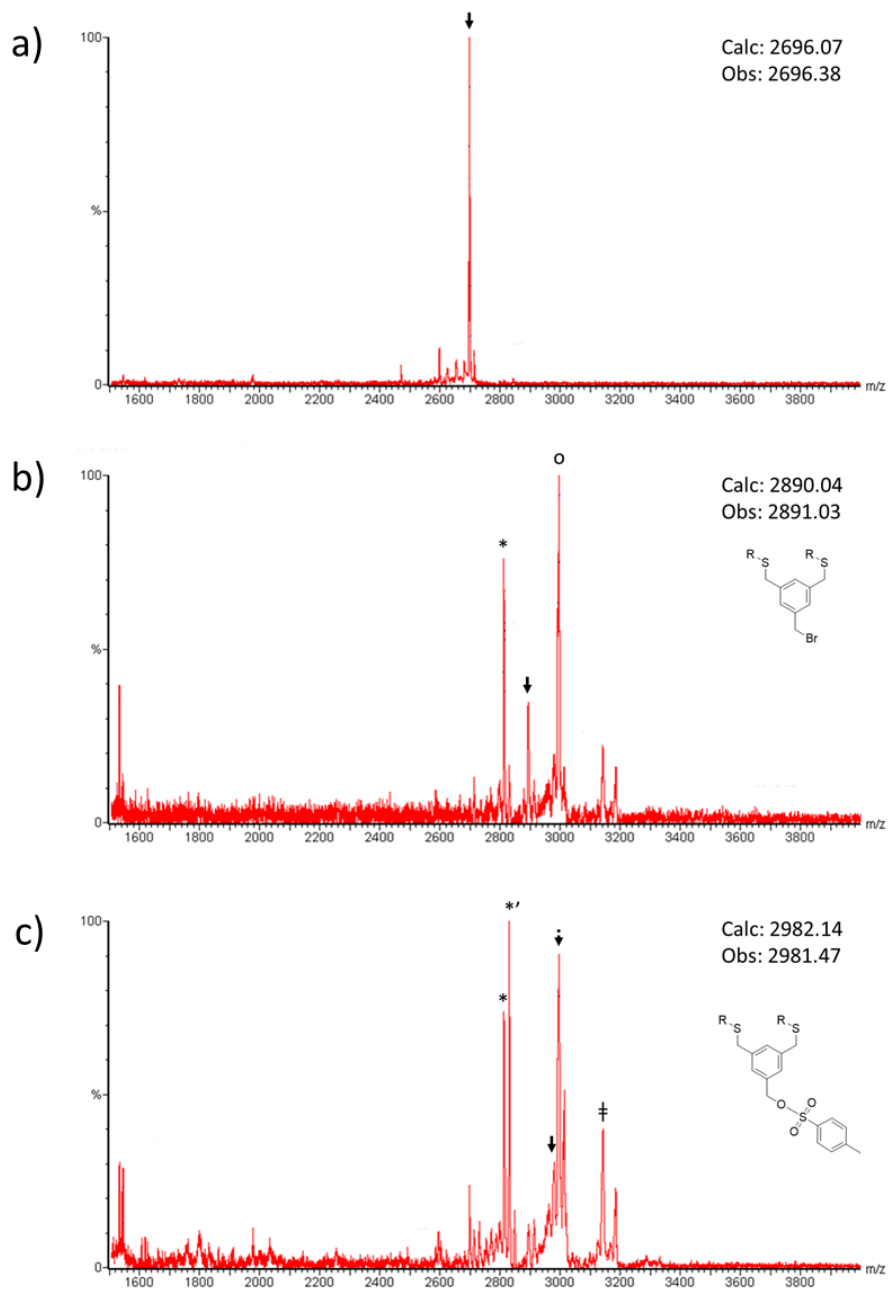


Figure 3.5 Cyclization with trifunctional bromomethylbenzene linkers. MALDI-TOF spectra of a) linear peptide 8.5, b) 8.5 reacted with 1,3,5-tris(bromomethyl)benzene, and c) 8.5 reacted with 1,3-bis(bromomethyl)benzyl-5-methylbenzenesulfonate. Solid arrows represent expected mass, (*) = bicyclic peptide, (o) = monocyclic B2 with TCEP attached, (*') and dotted arrow indicate oxidized versions of (*) and solid arrow respectively. (‡) indicates an unassigned peak.

3.2.2 Cyclization of *in vitro* translated peptides using α,α' -dibromo-*m*-xylene

Having demonstrated the versatility and robust nature of the bisalkylation reaction for cyclizing synthetic peptides in solution, we turned our attention to creating cyclic ribosomally translated peptides. Since PURE translation contains a complex mixture of proteins, nucleic acids and various small molecules, we chose to capture the peptides via the epitope tag and perform the chemistry on affinity resin. We designed a sequence containing two cysteines that expressed a model peptide containing a hexaHistidine epitope tag, OvLap (MVTNCSVFTSVCGGG-H₆) for our *in vitro* experiments. We initiated a standard translation with OvLap mRNA containing all natural amino acids (see 3.5.5), then performed the bisalkylation with α,α' -dibromo-*m*-xylene on OvLap immobilized on Ni-NTA resin, which allowed removal of most of the PURE components and the filtering of excess linker. After eluting the peptide from the resin, it was concentrated and desalted by C₁₈ microchromatography and subsequently characterized by MALDI-TOF. We observed > 90% conversion of the linear peptide (**Figure 3.6a**) to cyclic-OvLap (**Figure 3.6b**), demonstrating that the bisalkylation reaction is a highly efficient method for cyclizing *in vitro* translated his-tagged peptides immobilized on nickel resin.

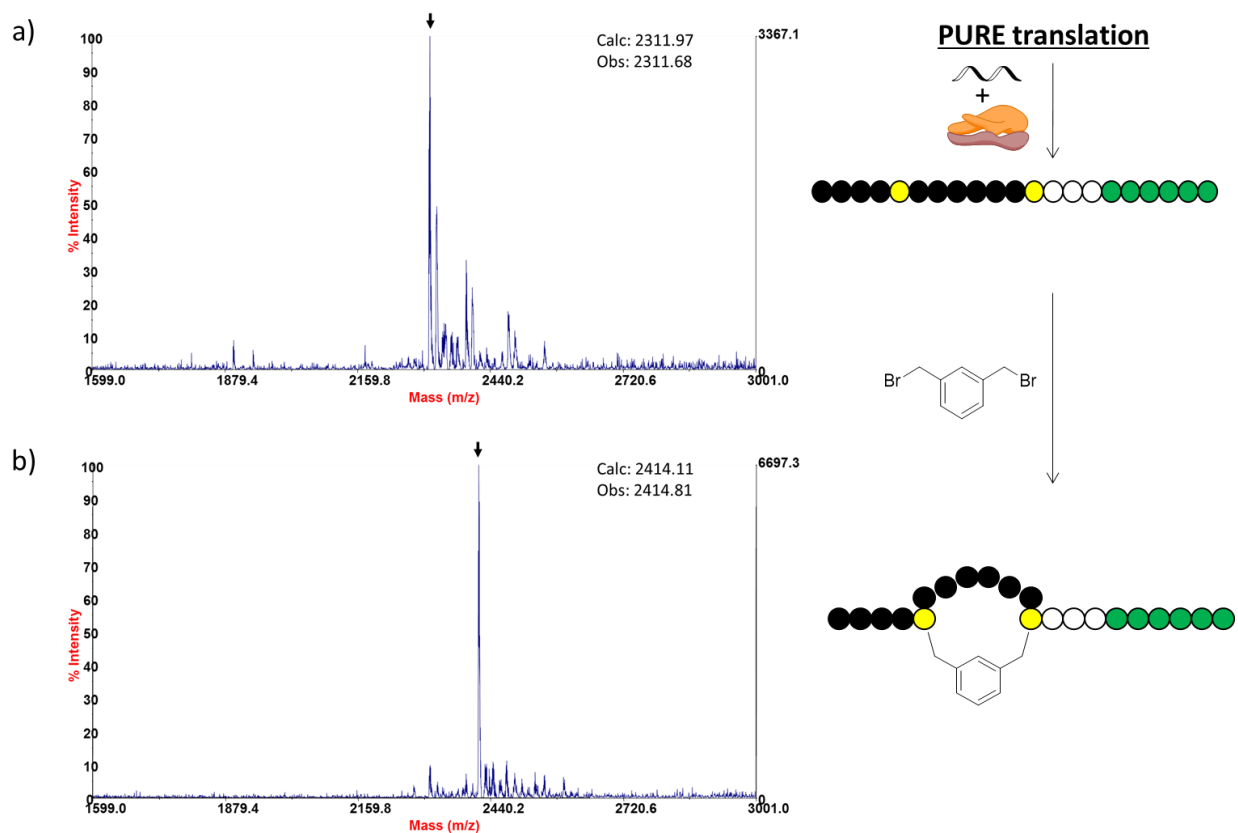


Figure 3.6 Cyclization of an *in vitro* translated model peptide with α,α' -dibromo-*m*-xylene. MALDI-TOF spectra of a) linear test peptide containing two cysteines (yellow circles) and b) the same peptide after the bisalkylation reaction performed on nickel resin.

3.2.3 Toward CuAAC bicyclization of *in vitro* translated peptides

The copper(I) catalyzed azide-alkyne cycloaddition reaction (CuAAC or click reaction), pioneered by Kolb, Sharpless and Finn,¹³⁷ is widely utilized in chemical biology due to its regioselectivity, orthogonality and robustness under mild, aqueous conditions. Having demonstrated the ability to create monocyclic *in vitro* translated peptides (which contain two cysteines) via bisalkylation with an α,α' -dibromo-*m*-xylene linker, we chose to pursue the click reaction as an orthogonal second chemistry to generate programmable bicyclic peptides with virtually unlimited topologies, including both manacle and theta-bridged conformers (**Figure 3.7a**).⁶²

Incorporation of azide and alkyne functionalized amino acids in translation

In order to utilize CuAAC cyclization, it is necessary to incorporate α -amino acids bearing an azide and an alkyne. We and others have previously shown that L- β -azidohomoalanine (AzHA) and *p*-ethynyl-L-phenylalanine (F-yne, **Figure 3.7b**) act as substrates for aminoacylation onto their cognate tRNAs by methionyl-tRNA synthetase (MetRS) and a mutant phenylalanyl-tRNA synthetase (PheRS A294G), respectively, and can be incorporated independently into peptides in the absence of the wild-type substrates.^{24,26,147-149} The bromomethylbenzene chemistry only requires two cysteines. To demonstrate the compatibility of these two cyclization methods and optimize conditions for the cyclization chemistries, we used our model peptide OvLap which contained two cysteines and encoded fixed clickable residues and would result in a knotted peptide in the theta (θ) conformation when the cyclization chemistries were applied (**Figure 3.7c**). We measured the yield and fidelity of incorporation of these analogs within a PURE translation lacking methionine and phenylalanine (**Figure 3.7d-h**); both were incorporated with good fidelity and

yield individually. Incorporation of both amino acids led to a slightly diminished yield (**Figure 3.7d**), with maintenance of fidelity (**Figure 3.7e**).

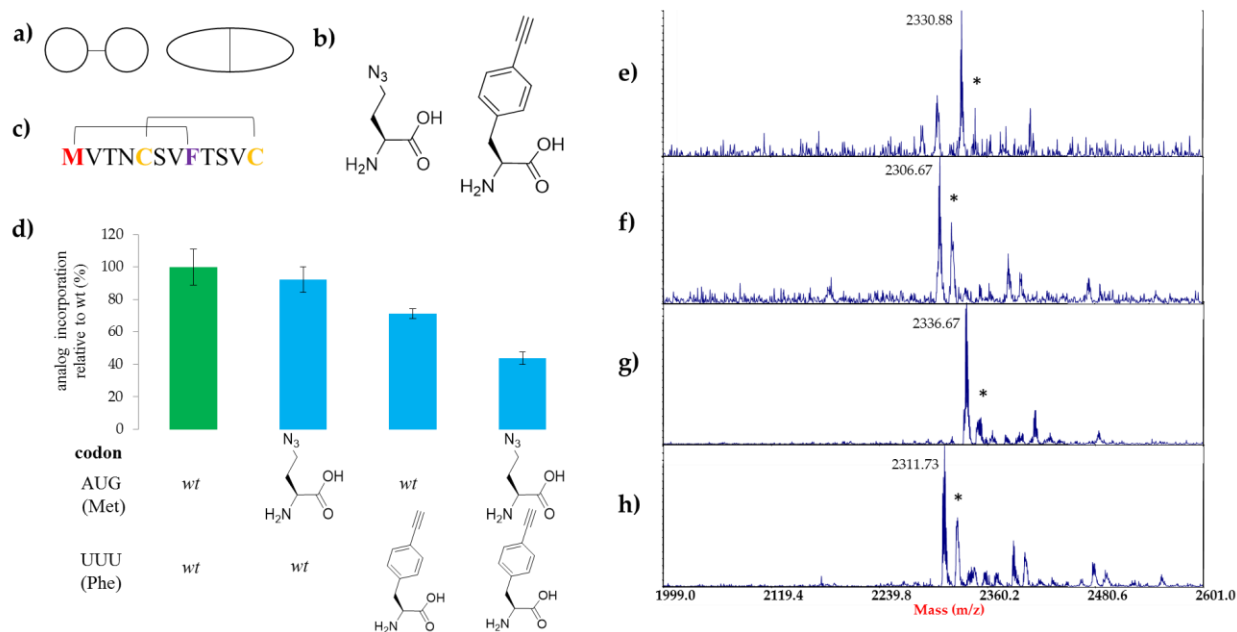


Figure 3.7 Bicyclic peptide conformations and cyclizable analog incorporation. a) Bicyclic peptide conformations obtained from two orthogonal chemistries; manacle (left) and theta-bridged (right) b) Analogs β -azidohomoalanine (AzHA, left) and *p*-ethynyl phenylalanine (F-yne, right) incorporated in place of methionine and phenylalanine, respectively. c) Overlap model peptide sequence, indicating the knotted-like theta-bridged conformation. d) Radiolabeled peptides were synthesized in sufficient quantities and MALDI-TOF confirmed appropriate mass for peptides containing e) both analogs (calc-2331.20), f) AzHA (calc-2306.98), g) F-yne (calc-2336.11), relative to h) all wild-type amino acids (calc-2311.97). * Indicates a ubiquitous +16 peak, presumably an artifact of sulfur oxidation during Zip-tip desalting.

Creation of bicyclic peptides using orthogonal chemistries. We next worked to ensure that we could perform both CuAAC and cysteine bisalkylation on this model peptide. We reasoned that performing the mild bisalkylation reaction first would allow for complete conversion to the monocyclic peptide and minimize any potential decrease in yield and/or fidelity from oxidative by-products generated during the click reaction.^{137,150} We reacted unnatural versions of the model peptide (containing AzHA and/or F-yne) with α,α' -dibromo-*m*-xylene and observed complete conversion to the bithioether benzene-bridged macrocyclic peptide via MALDI-TOF (**Figure 3.8a**). The generation of triazole-linked cyclic peptides does not result in a mass change; therefore, we employed ‘clickable’ external reagents to optimize our click conditions. We prepared a version of our peptide containing only F-yne. When we titrated this peptide with an azide-functionalized sulforhodamine 101 (Texas Red-azide, **Figure 3.8b**), we observed concentration-dependent labeling in the presence of click reagents (**Figure 3.8c**). Similarly, an AzHA-labeled peptide could be successfully reacted with hex-5-yne-oic acid (**Figure 3.9**). We then prepared a peptide containing both AzHA and F-yne. When this peptide was treated with CuAAC reagents prior to addition of 250 μ M Texas Red azide, very little labeling occurred (**Figure 3.8, lane 9**). Similarly, we observed no labeling when hex-5-yne-oic acid was added to the ‘pre-clicked’ peptide (**Figure 3.10**). These data, taken together, demonstrate that we can perform both CuAAC and α,α' -dibromo-*m*-xylene cyclizations on *in vitro* translated peptides, resulting in the creation of bicyclic peptides.

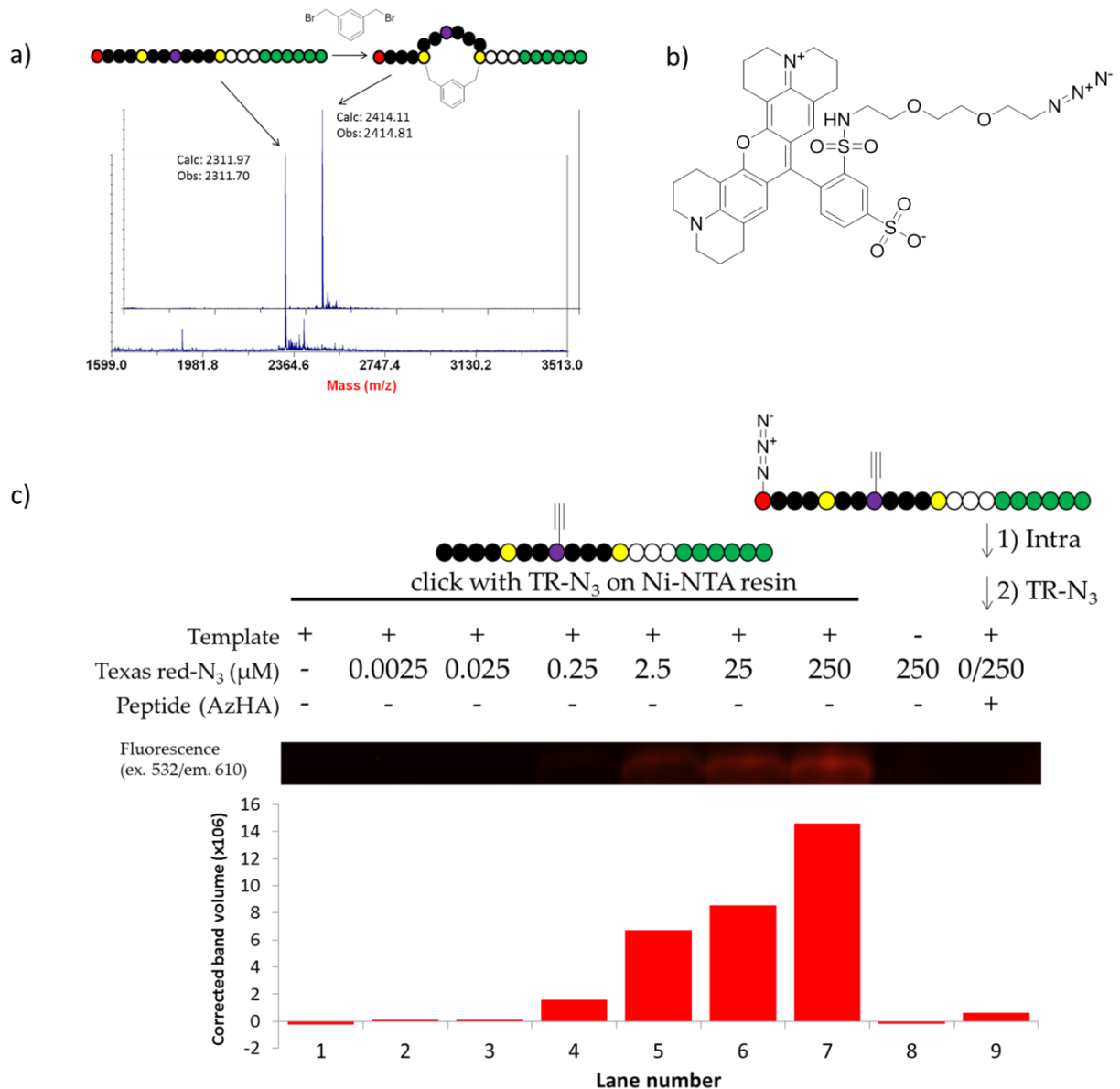


Figure 3.8 Mono- and click bicyclization of *in vitro* translated peptides. a) MALDI-TOF spectrum of linear OvLap reacted with α,α' -dibromo-*m*-xylene, b) azide-functionalized sulforhodamine-101 (Texas Red-azide or TR-N₃) used for fluorescent labeling. c) Concentration-dependent labeling of the (AzHA) peptide by TR-N₃. Negligible fluorescence is observed when the same reaction is performed on a ‘pre-clicked’ peptide containing both the azide and alkyne (far right lane), evidence that the intramolecular click cyclization has already occurred. Lanes 1-7: Concentration-dependent fluorescent labeling of alkyne-containing peptide with TR-N₃, indicating that the click reaction is compatible with Ni-NTA agarose-immobilized *in vitro* translated peptides. Lane 8: No mRNA template control reaction. Lane 9: TR-N₃ click-reacted with ‘pre-clicked’ azide/alkyne-containing peptide.

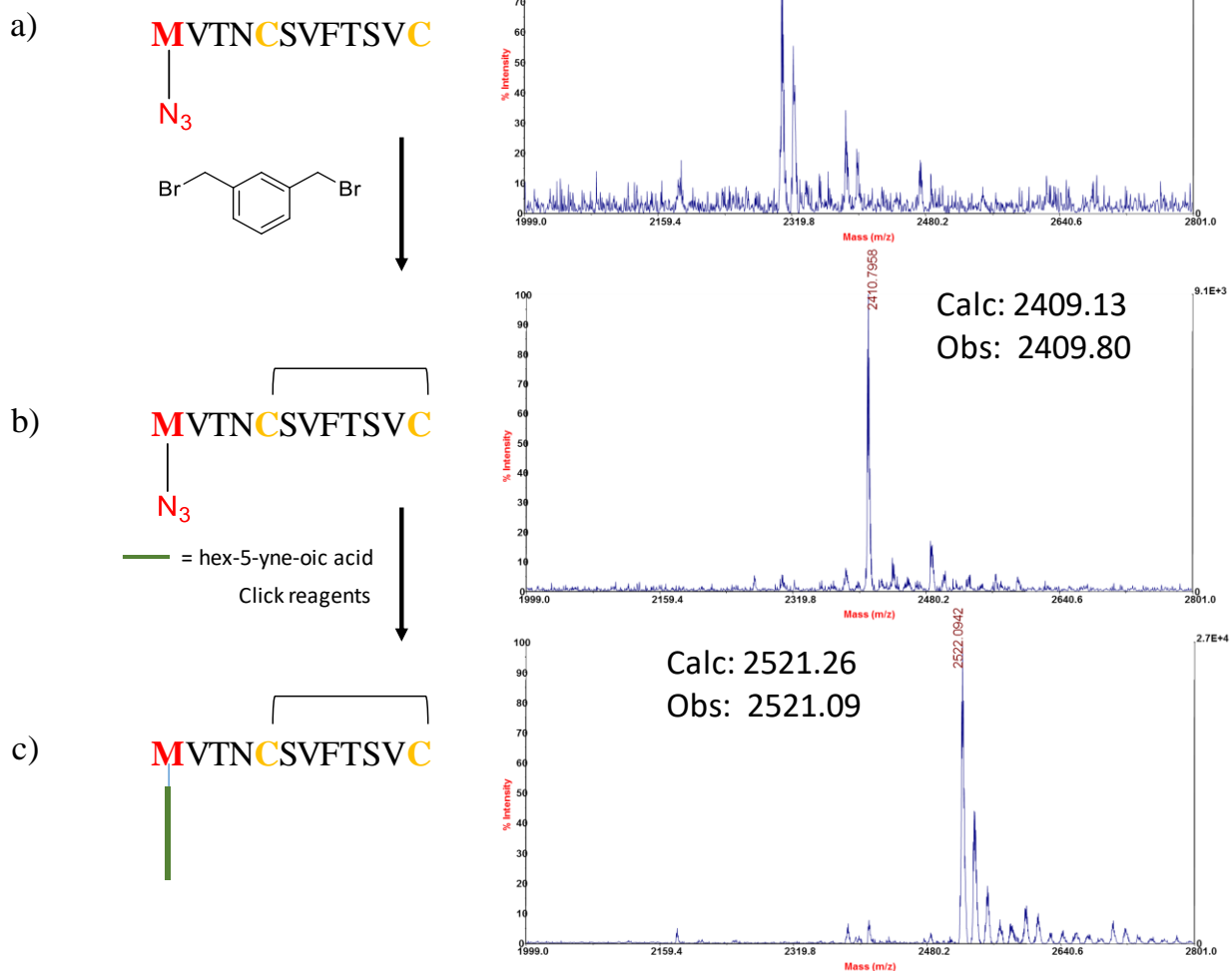


Figure 3.9 Labeling of azide peptide with an external alkyne reagent. Model peptide containing AzHA; a) in vitro translated peptide, b) bisalkylation reaction with α,α' -dibromo-*m*-xylene, followed by c) CuAAC with hex-5-yne-oic acid.

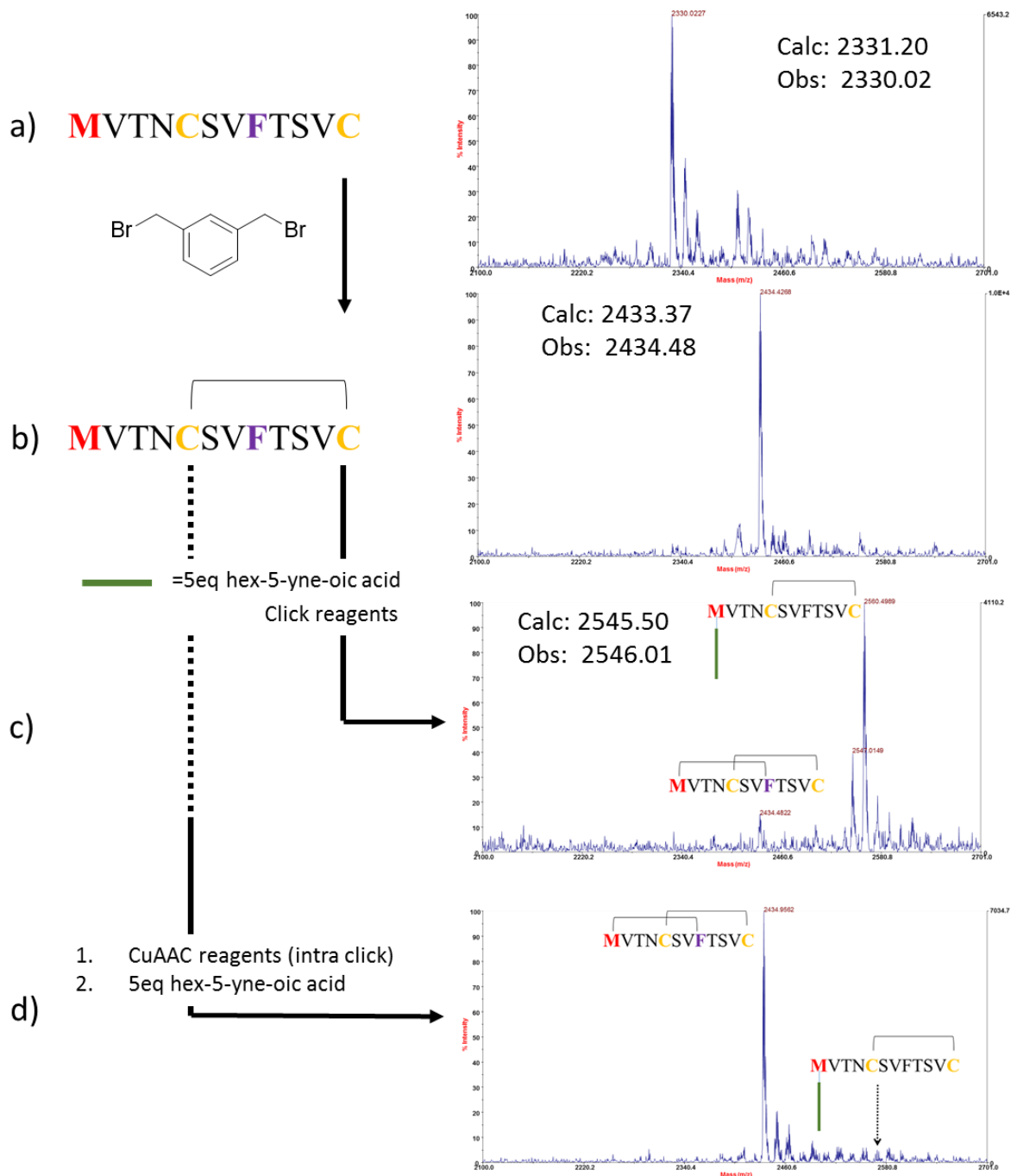


Figure 3.10 Lack of labeling of azide/alkyne ‘pre-clicked’ peptide with an external alkyne. Model peptide containing both AzHA and F-yne; a) in vitro translated peptide, b) bisalkylation reaction with α, α' -dibromo-*m*-xylene, c) reaction with 5 equivalents of hex-5-yne-oic acid (+14 base peak is an unidentified side product) and d) the pre-clicked peptide is unreactive in a subsequent CuAAC with hex-5-yne-oic acid.

3.3 Summary

Peptide cyclization is widely regarded as a critical component in providing peptides the requisite level of protease stability and target affinity enhancement required for application as therapeutics. Bromomethylbenzene reagents have long been applied for this purpose. In this study, we further investigated the versatility of bromomethylbenzene derivatives to cyclize and also to concurrently cyclize and functionalize peptides. We demonstrated the ability of a bromomethylbenzene linker, specifically α,α' -dibromo-*m*-xylene, to completely convert a linear synthetic peptide (R2) containing two intrafacial cysteines ($i, i + 7$) to cyclic R2. We previously identified a peptide with only one cysteine from an α,α' -dibromo-*m*-xylene -cyclized library from sequencing following *in vitro* selection against BRCA1. When treated with α,α' -dibromo-*m*-xylene, we determined that the peptide formed a small cycle between the initiator methionine and the adjacent fixed cysteine, but the relatively unstable sulfonium ion could be displaced by a strong reducing agent attacking the benzylic carbon to form the peptide-linker-reducing agent adduct. However, if the reaction is performed in the presence of a low concentration of a weak reducing agent, the Met to Cys cyclic peptide is preserved. The use of an electron withdrawing group in the 5-position (NO₂) of the bromomethylbenzene linker acted to stabilize the Met-Cys linked peptide and enabled it to survive the presence of a strong reducing agent using the standard bromomethylbenzene cyclization protocol. The 5-position of the bromomethylbenzene derivatives offers a host of opportunities to append labels, chemical handles, or to introduce functionalities that tune polarity, solubility or lipophilicity. The addition of reactive handles to potentially react with and covalently label binding domain residues in target proteins is an exciting possibility. We were able to introduce a tosylate and a bromomethyl group into a cyclic peptide. While the yields of each were relatively low, it is possible that optimization of the reaction (such

as adjusting concentration of linker and reducing agent and decreasing time) will make this a more viable route for the development and application of bifunctional cyclic-covalent inhibitor peptides.

We adapted and optimized the bromomethylbenzene protocol to cyclize *in vitro* translated peptides immobilized on Ni-NTA resin with α,α' -dibromo-*m*-xylene. The on-resin method facilitates the removal of the majority of the PURE system components simply by washing and subsequent centrifugation using a microcentrifuge filter tube. Furthermore, the excess linker can be removed via the same process, negating the necessity of the quenching step, and hence strong reducing agents, that open the cycle created in peptides containing one cysteine. This quick and simple protocol allows full conversion to the cyclic product from linear peptides containing one or two cysteines.

Using our established method of ncAA incorporation into peptides, we were able to incorporate amino acids containing azide and alkyne functional groups *in situ*, which enabled the use of the orthogonal CuAAC reaction. When CuAAC is performed in conjunction with bisalkylation with α,α' -dibromo-*m*-xylene, we are able to create bicyclic, ribosomally translated peptides. The directable nature and orthogonal reactivity of the two reactions allows for the generation of nearly unlimited topologically diverse bicyclic peptides, including theta-bridged conformations which act as structural mimics of the highly potent knotted peptide natural products.

3.4 Future Directions

Further optimization is necessary to develop bromomethylbenzene derivatives that act as bifunctional linkers and potential covalent inhibitors (3.2.1.3). The method we developed to

generate bicyclic peptides (3.2.3.2) will be adapted and optimized to be compatible with mRNA display. This will enable the generation of mRNA-peptide fusion libraries used to perform *in vitro* selection using bicyclic or knotted scaffolds.

3.5 Experimental

3.5.1 Amino acids and translation reagents

p-chloro-L-phenylalanine (F₁₁) and L-β-azidohomoalanine (M₄) were obtained with appreciation from the Tirrell Laboratory at Cal Tech. For natural amino acid information, see 2.5.1. The synthesis of Texas Red (sulforhodamine 101) azide will be reported elsewhere. Hex-5-yne-*oic* acid was purchased from Sigma. Preparation of *in vitro* translation components and reagents is described in 2.5.3.

3.5.2 Standard solid phase Fmoc-based microwave-assisted peptide synthesis

The peptides were synthesized using a Liberty Automated Peptide Synthesizer with a Discover microwave module (CEM). Fmoc-PAL-PEG-PS resin (Applied Biosystems) was chosen as the solid support for a standard 0.1 mmol scale synthesis using N-α-Fmoc-protected amino acids (CEM). The peptides were cleaved from the resin using a cocktail containing trifluoroacetic acid (TFA-Chem Impex)/triisopropylsilane (TIS-Sigma)/3,6-dioxa-1,8-octanedithiol (DODT-Sigma)/water in a ratio of 92.5/2.5/2.5/2.5 and incubated at RT for 4 h. The TFA was bubbled off of the filtrate with argon, the peptide was precipitated with 10 volumes of cold diethyl ether and collected by centrifugation. The supernatant was discarded and the pellet was dissolved in 25% acetonitrile

(Fisher Optima grade), followed by freezing and lyophilization. Peptides were purified by reverse phase HPLC using a Shimadzu Prominence instrument with a Vydac Protein & Peptide (P/N 218TP52210) semi-preparative column with a 0.1% TFA in water (A) and 0.1% TFA in acetonitrile (B) as the mobile phase monitoring at 215 nm and 280 nm.

3.5.3 Standard protocol for peptide cyclization with bromomethylbenzene reagents

A 50 mL oven-dried flask was charged with water (19.34 mL) and acetonitrile (5.28 mL) and degassed under argon for 10 min. Then 200 mM ammonium bicarbonate (3.5 mL, pH 8.4), tris-carboxyethylphosphine (TCEP) (2.0 mg, 7.0 μmol) in 3.5 mL water and peptide (10 mg, 3.5 μmol) were added and the reaction was kept under argon. After 30 min, bromomethylbenzene linker (10.2 mg, 38.6 μmol) was added as a solution in 3.5 mL of acetonitrile. The reaction was incubated at RT and monitored by MALDI-TOF. After 2 h, β -mercaptoethanol (BME) (13.7 mg, 176 μmol) or dithiothreitol (DTT) (27.1 mg, 176 μmol) was added to quench unreacted linker. The reaction was then frozen and lyophilized. When purification of the cyclic peptides was required, the resulting white powder was dissolved in 25% acetonitrile and purified by reverse phase semi-preparative HPLC under the following conditions. Column: Vydac 218TP52210 22 x 100 mm: Flow rate: 10 mL/min: Solvents: A=water/0.1% TFA, B=acetonitrile/0.1% TFA).

3.5.4 Peptide cyclization with bromomethylbenzene reagents (without quenching)

A 50 mL oven-dried flask was charged with water (19.34 mL) and acetonitrile (5.28 mL) and degassed under argon for 10 min. Then 200 mM ammonium bicarbonate (3.5 mL, pH 8.4), tris-carboxyethylphosphine (TCEP) (2.0 mg, 7.0 μmol) in 3.5 mL water and peptide (10 mg, 3.5 μmol) were added and the reaction was kept under argon. After 30 min, bromomethylbenzene linker (10.2 mg, 38.6 μmol) was added as a solution in 3.5 mL of acetonitrile. The reaction was

incubated at RT and monitored by MALDI-TOF. The reaction was then frozen and lyophilized. When purification of the cyclic peptides was required, the resulting white powder was dissolved in 25% acetonitrile and purified by reverse phase semi-preparative HPLC under the following conditions. Column: Vydac 218TP52210 22 x 100 mm: Flow rate: 10 mL/min: Solvents: A=water/0.1% TFA, B=acetonitrile/0.1% TFA).

3.5.5 Cyclization of immobilized *in vitro* translated peptides with α,α' -dibromo-*m*-xylene

A standard translation reaction (2.5.3-100 μ L each) initiated with mRNA encoding the model peptide (OvLap) was quenched by diluting with 300 μ L of 50 mM Tris-HCl pH 8.0, 300 mM NaCl (TBS) with 5 mM β -mercaptoethanol (BME) and captured with 50 μ L Ni-NTA agarose slurry (MCLAB) in a 0.5 mL centrifugal filter (VWR #82031-358). After 1 h, immobilized peptides were washed twice with 500 μ L TBS containing 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), and 500 μ L cyclization buffer was added (20 mM Tris-HCl pH 7.8, 300 mM NaCl, 3 mM α,α' -dibromo-*m*-xylene, 33% acetonitrile (v/v), 0.5 mM TCEP). The reactions were rotated at RT for 30 min. The bound peptides were washed twice with 500 μ L TBS containing 5 mM BME to quench the unreacted linker. For MALDI-TOF characterization, cyclized peptides synthesized from translation reactions containing 200 μ M cysteine were eluted with 50 μ L of 1% TFA, desalted and concentrated by reverse phase micro-chromatography using C₁₈ Zip-Tips (Millipore) per manufacturer's instructions and eluted with 10 mg/mL 4-chloro- α -cyanocinnamic acid (CHCl) in 7:3:0.01 acetonitrile: water: trifluoroacetic acid. Mass analysis was determined with Applied Biosystems Voyager DE PRO MALDI-TOF operated in reflector positive mode.

3.5.6 Incorporation of clickable amino acids in test templates

Ribosomal incorporation and optimization of non-canonical amino acids was performed using the PURE cell-free translation system containing the following reagents: HEPES-KOH, pH 7.6 (50 mM), potassium acetate (100 mM), magnesium acetate (6 mM), creatine phosphate (20 mM), spermidine (2 mM), ATP (1 mM), GTP (1 mM), E. coli total tRNA (2.4 mg/mL), deacylated as previously described,¹¹⁴ (6R,S)-methenyl-5,6,7,8-tetrahydrofolic acid chloride, adjusted to pH 8 (0.1 mM), inorganic pyrophosphatase (1 µg/mL), creatine kinase (4 µg/mL), myokinase (3 µg/mL), nucleoside diphosphate kinase (1.1 µg/mL), ribosomes (1.2 µM), EF-Tu (10 µM), EF-Ts (8 µM), EF-G (0.52 µM), IF-1 (2.7 µM), IF-2 (0.4 µM), IF-3 (1.5 µM), methionyl-tRNA transformylase (0.6 µM), RF-1 (0.3 µM), RF-3 (0.17 µM), RRF (0.5 µM), RRS (0.03 µM), CRS (0.11 µM), ERS (0.32 µM), QRS (0.02 µM), IRS (0.06 µM), LRS (0.04 µM), YRS (0.04 µM), WRS (0.47 µM), DRS (0.15 µM), NRS (0.06 µM), GRS (0.02 µM), HRS (0.2 µM), KRS (0.07 µM), PRS (0.01 µM), TRS (0.05 µM), VRS (0.04 µM), SRS (0.06 µM), ARS (D777K) (0.04 µM), FRS (A294G) (0.02 µM), and 17 amino acid mix (200 µM each) lacking Met, Cys and Phe. To create unnatural peptides, we added 1.25 mM AzHA and/or 0.8 mM F-yne, while omitting methionine and phenylalanine from translation. For generation of natural peptides, we added 200 µM of each natural amino acid. Peptide yield was determined by 50 µL translation reactions containing 0.11 µCi ³⁵S-Cysteine (along with 3.5 µM cysteine) and fidelity was determined by MALDI-TOF of peptides synthesized in translation reactions containing 200 µM unlabeled cysteine.

3.5.7 CuAAC reactions with *in vitro* translated model peptide

3.5.7.1 CuAAC peptide-alkyne labeling with PEG-azide functionalized sulforhodamine 101 (TR-azide)

Ni-NTA immobilized *in vitro* translated OvLap peptide, cyclized with α,α' -dibromo-*m*-xylene (see 3.5.5) was washed once with 500 μ L TBS with 5 mM BME and the following reagents were added (500 μ L total volume): 100 mM phosphate pH 8, 1 mM CuSO₄, 500 mM NaCl, 2 mM Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA), 166 μ L DMSO. For CuAAC reaction of alkyne-peptide, azide-functionalized sulforhodamine 101 (TR-azide) was dissolved in dimethyl sulfoxide (DMSO) and titrated (0.0025-250 μ M) against the immobilized peptide. Argon was bubbled through the reaction followed by addition of 10 mM sodium ascorbate. An argon blanket was placed over the reaction and the tube was sealed and rotated at RT for 2h. After incubation, the peptides were washed once with 500 μ L of 10% (v/v) DMSO in TBS w/5 mM BME (to remove any remaining TBTA) and twice more with 500 μ L TBS w/5 mM BME. Bicyclic peptides were subsequently eluted similarly to the monocyclic peptides above. For validation of TR-azide labeling of the model peptide, the titration reactions were analyzed on a 15% Tris-Tricine peptide gel.¹⁵¹ TR-azide fluorescence was measured using a Chemi-Doc MP imaging system (Bio-Rad) with excitation and emission wavelengths set at 532 and 605 nm, respectively.

3.5.7.2 CuAAC peptide-azide labeling with hex-5-yne-oic acid

Ni-NTA immobilized *in vitro* translated OvLap peptide, cyclized with α,α' -dibromo-*m*-xylene (see 3.5.5) was washed once with 500 μ L TBS w/5 mM BME and the following reagents were added (500 μ L total volume): 100 mM phosphate pH 8, 1 mM CuSO₄, 500 mM NaCl, 2mM TBTA, 166 μ L DMSO. For CuAAC reaction of azide-peptide, hex-5-yne-oic acid was added to the aqueous reagents above at a final concentration of 5 mM. Argon was bubbled through the reaction followed by addition of 10 mM sodium ascorbate. An argon blanket was placed over the reaction and the tube was sealed and rotated at RT for 2 h. After incubation, the peptides were washed once with 500 μ L of 10% (v/v) DMSO in TBS w/5 mM BME (to remove any remaining

TBTA) and twice more with 500 μ L TBS w/5 mM BME. To confirm intramolecular pre-click cyclization of the peptide that contained both the azide and the alkyne, hex-5-yne-oic acid was omitted from the cyclization buffer during the initial 2h reaction, the peptide was washed as above, then the hex-5-yne-oic acid was included in 500 μ L of cyclization buffer and reacted for an additional 2h, then washed as above. Bicyclic peptides were eluted with 50 μ L of 1% TFA and collected by centrifugation through a spin filter following a 15 min. incubation at RT. For validation of labeling, peptides were analyzed via MALDI-TOF mass shift following C₁₈ Zip-tip microchromatography.

3.5.8 Preparation of OvLap mRNA for *in vitro* translation

Bottom strand oligo was purchased from IDT. T7-mediated *in vitro* transcription was performed and mRNA was purified and prepared for translation initiation as described in 2.5.2.

Name	Peptide sequence	Section
R2	EAC <u>AR</u> V1AA <u>CE</u> AAARG	3.2.1.1
8.6	MC <u>ND</u> FTFDKLNHHGSPKKR <u>KV</u>	3.2.1.2
8.5	MC <u>YD</u> FDTTDTHTFIGSPKKR <u>KV</u>	3.2.1.3
OvLap (natural)	MVTN <u>C</u> SVFTSV <u>C</u> GGGHHHHHH	3.2.2
OvLap (azide)	(AzHA)VTN <u>C</u> SVFTSV <u>C</u> GGGHHHHHH	3.2.3.2
OvLap (alkyne)	MVTN <u>C</u> SV(F-yne)TSV <u>C</u> GGGHHHHHH	3.2.3.2
OvLap (azide/alkyne)	(AzHA)VTN <u>C</u> SV(F-yne)TSV <u>C</u> GGGHHHHHH	3.2.3.2
Name	RNA sequence	
OvLap	GGGUUAAACUUUAAGAAGGAGAUAUACAUAUGGUGACCA ACUGCAGCGUUUUUACGAGUGUCUGUGCGGUGGCAU CAUCACCAUCACCAUUAGCACCGGCUAUAGG	3.2.2-3.2.3

Table 3.1 Peptide and mRNA sequences for cyclization experiments.

Chapter 4. Mono- vs. Bicyclic Peptide Libraries;

***In vitro* Selection with mRNA Display**

4.1 Introduction

We have demonstrated the ability to create highly constrained bicyclic, or knotted, *in vitro* translated peptides (**Chapter 3**). Our stringent validation method consisted of a model unnatural peptide that contained a highly-constrained theta-bridge linkage when sequentially bicyclized via α,α' -dibromo-*m*-xylene bisalkylation and CuAAC. Given this exciting development, the difficulties in creating large libraries of bicyclic, or knotted peptides have proven to be a significant obstacle. Cyclization reactions for this purpose must be: (1) robust and high-yielding to maximize enrichment, (2) mild enough to be compatible within the selection system used for ligand discovery, and (3) orthogonal to guarantee the generation of a stable, high yielding product. For example, the use of 1,3,5-trisbromomethylbenzene (TBMB) to generate bicyclic libraries of tricysteine peptides used for phage display selection has been reported.⁷⁴ While the size of each macrocycle can be varied using this technique, conformations are limited to two independent loops, because of the reliance on a central linker. The formation of bicyclic peptides with disulfide bond linkages has been reported,^{66,71,152} but the lack of stability in the intracellular reducing environment limits the application of this technique as a viable peptide library cyclization strategy.^{72,73}

Here we describe the adaptation of our peptide bicyclization technology to the creation of $>10^{13}$ member libraries of knotted peptides using mRNA display.^{107,153} By using two orthogonal chemical cyclization steps, we generated libraries that contain bicyclic peptides with highly variable and directable theta-bridged topologies. We conducted parallel mono- vs. bicyclic *in vitro* selections to compare the relative fitness of the respective library scaffolds to bind to a model target, streptavidin, (**Figure 4.1**) as well as the ability to resist protease degradation.

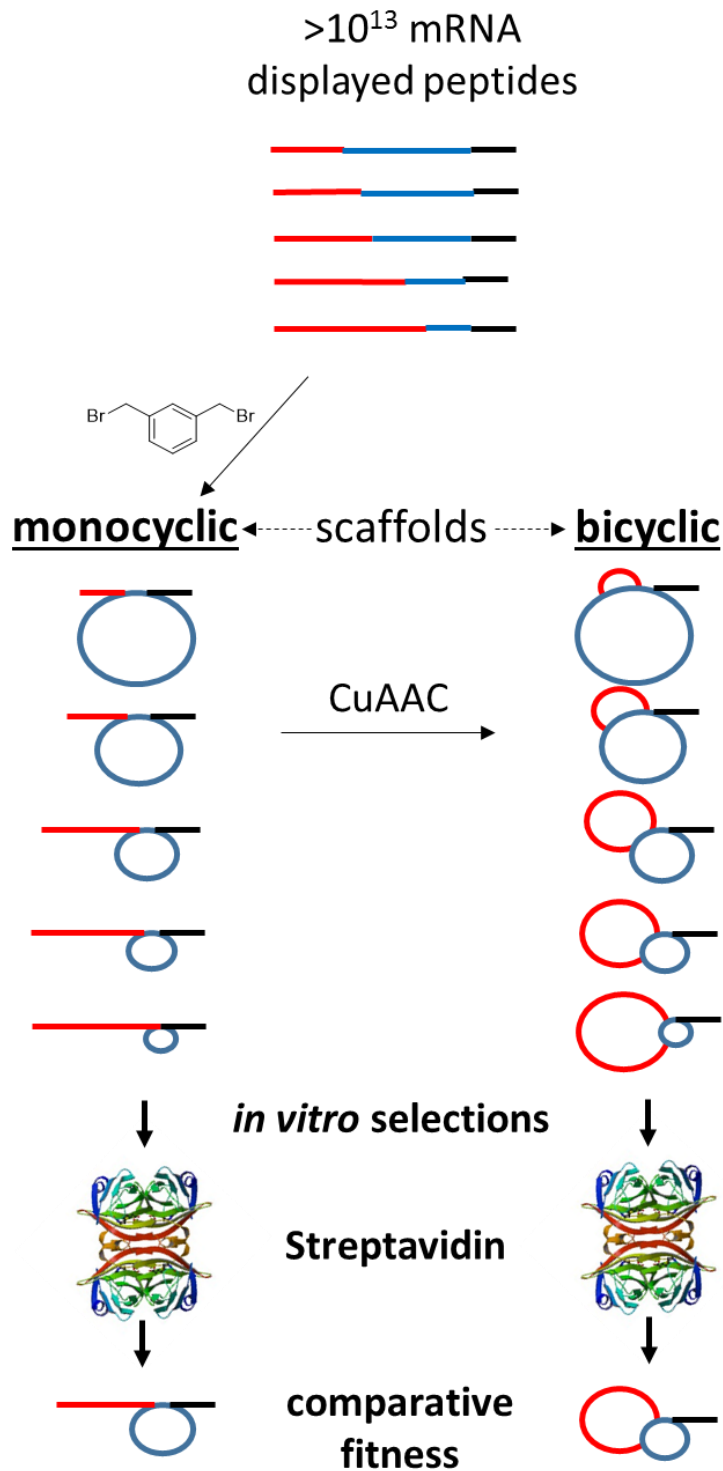


Figure 4.1 Scheme of competitive mono- and bicyclic selections. Parallel mRNA-peptide fusion preparations with bisalkylation to form the monocyclic library and a subsequent CuAAC reaction to generate the bicyclic library. Side-by-side panning against streptavidin yields enriched sequences which are tested for affinity and protease stability.

4.2 Results

We previously demonstrated the ability to ribosomally incorporate an azide and alkyne into peptides (**3.2.3.1**). Furthermore, we validated the orthogonal, sequential on-resin chemistries (bisalkylation with α,α' -dibromo-*m*-xylene followed by CuAAC) to create *in vitro* translated bicyclic peptides (**3.2.3.2**). We next sought to demonstrate the compatibility of this technology with mRNA display.

4.2.1 Compatibility of CuAAC chemistry with mRNA display

To create mRNA-displayed peptide libraries, it is essential that the proposed chemistries do not degrade mRNA. The α,α' -dibromo-*m*-xylene cyclization chemistry has already been shown to be compatible with mRNA display.^{55,117,139} However, others have shown that CuAAC reagents degrade nucleic acids over time and that RNA is particularly susceptible to oxidation.^{154,155} The use of polytriazole ligands, such as tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA-**Figure 4.2a**) or the water-soluble tris(3-hydroxypropyl-triazolylmethyl)amine (THPTA-**Figure 4.2b**), have been shown to protect Cu(I) from oxidation or disproportionation.¹⁵⁶⁻¹⁵⁹ The addition of the copper coordinating ligands acts to both increase the rate of reaction and decrease the amount of reactive oxygen species (ROS) present as a result of the increased oxidation of ascorbate to dihydroascorbate resulting from copper reduction. In a reaction mixture containing trillions of peptide library members, increasing the ROS would have detrimental effects that would likely compromise phenotype fidelity. While the water solubility of THPTA would be advantageous with respect to translation yield and potential reagent compatibility, relative to the dimethylsulfoxide (DMSO) soluble TBTA, others have shown that the presence of DMSO in the click reaction increases the percentage of intact nucleic acid.¹⁵⁵ We tested the compatibility of library mRNA with each ligand and found that mRNA was highly stable up to 2 h using click

reagents in the presence of TBTA with 33% DMSO relative to the water soluble THPTA (**Figure 4.2c,d**). While the direct determination of the effect of the coordinating ligand on our peptide library is problematic, our presumption is that since the oxidative damage to the mRNA is far less with TBTA, the protective effect would be transferrable to oxidizable residues in the peptides displayed on the mRNA. Furthermore, the robust nature of CuAAC and the protective effect of polytriazole ligands in reducing oxidative byproducts are well-documented.^{158,160-163}

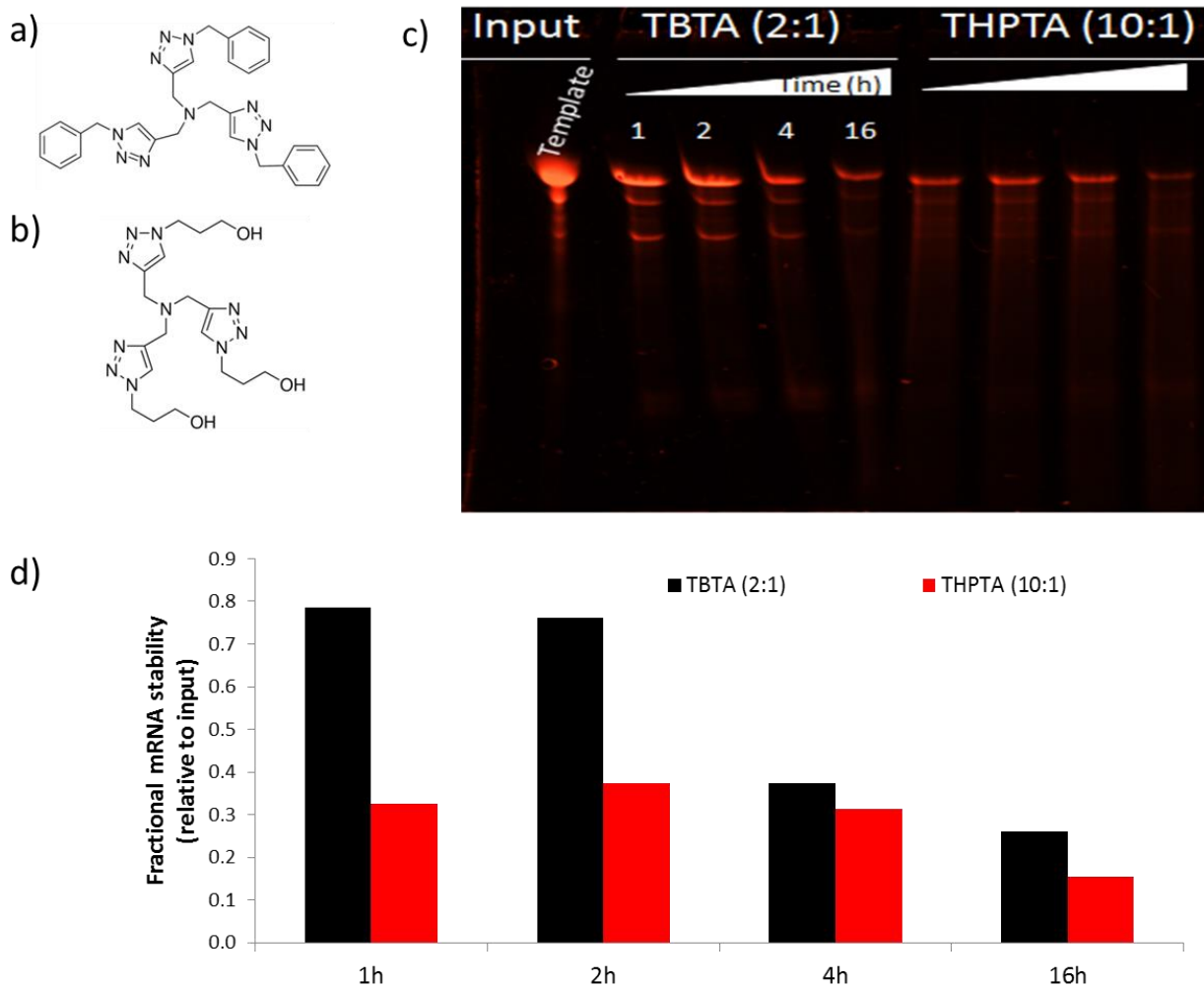


Figure 4.2 Copper coordinating ligand effect on mRNA stability. Copper coordinating ligands a) TBTA and b) THPTA. c) CuAAC reagents were added to mRNA template with protecting ligands. Aliquots were removed at the indicated time points, analyzed by 10% Urea-PAGE and d) quantified by densitometry.

We verified that these optimized click conditions can be applied to an mRNA-peptide fusion library. We first generated mRNA-displayed peptides using our mixed scaffold libraries designed for selection (**Table 1**), which encode AzHA, F-yne and two cysteines. We initiated an *in vitro* translation including F-yne with mRNA covalently attached to puromycin via a linker with a poly-(dA) region.¹⁰⁷ The mRNA-(F-yne)-peptide fusions were initially cyclized by cysteine-bisalkylation while immobilized on oligo(dT) magnetic beads,¹¹⁷ eluted, and subsequently bound to Ni-NTA resin. We then added 250 μ M Texas Red-azide, along with the optimized click reagents, and analyzed the reaction over time by SDS-PAGE. Labeling of the mRNA-(F-yne)-peptide fusion library, without mRNA degradation, occurred after 2 h under inert atmosphere (**Figure 4.3a**). We also prepared mRNA-peptide fusion libraries containing both AzHA and F-yne, and another with only F-yne. While immobilized on oligo(dT) resin, we performed CuAAC using 250 μ M Texas Red-azide on the (F-yne)-fusion library. After washing and elution, we observed fluorescent labeling of the (F-yne)-fusion library (**Figure 4.3b, left**). CuAAC was performed on the (AzHA, F-yne)-fusions in the absence of Texas Red-azide. Upon subsequent addition of the fluorophore, no labeling occurred (**Figure 4.3b, right**), indicating that the formation of the triazole-linked intramolecular bicycle had already occurred.

Lastly, we ensured that azide-containing mRNA-peptide fusions were not compromised during the most reductive step in the mRNA display process, reverse transcription, which requires 5 mM DTT and elevated temperatures for 30 min. (**Figure 4.4**). Taken together these experiments show that mRNA displayed peptide libraries are compatible with CuAAC chemistry.

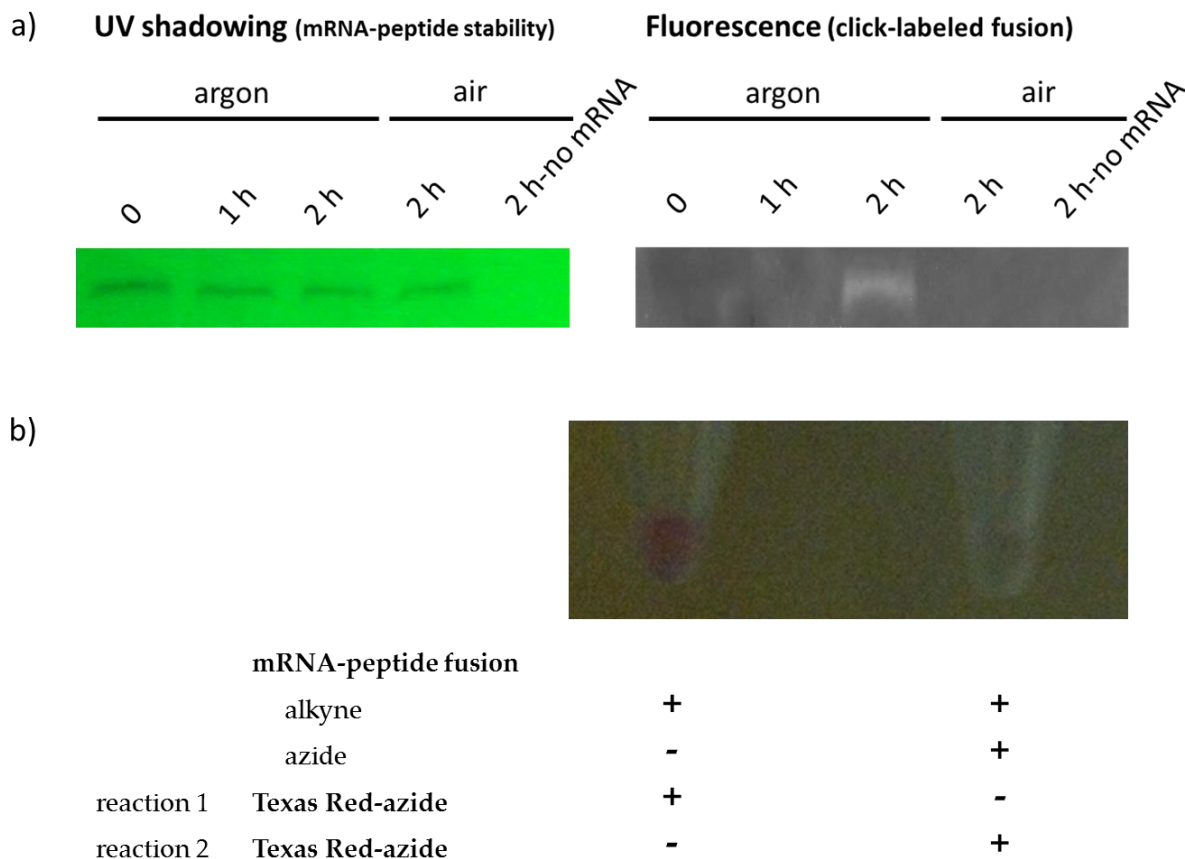


Figure 4.3 Compatibility of click reaction with peptides displayed on mRNA. a) Click reaction was performed with mRNA-(F-yne)-peptide fusions immobilized on Ni-NTA resin and 250 μ M Texas Red-azide. Aliquots of beads were removed at the noted time points, fusions were washed and eluted with imidazole, and analyzed by 10% SDS-PAGE. Left, UV shadowing of mRNA-peptide fusion libraries over time. Right, Fluorescence image (ex 532, em 610) of the same gel. b) Elution from oligo(dT) resin after CuAAC reaction of (F-yne)-fusions with 250 μ M Texas Red-azide (left tube). CuAAC was performed with (AzHA, F-yne)-fusions. CuAAC was performed a second time with 250 μ M Texas Red-azide and eluted from the resin (right tube).

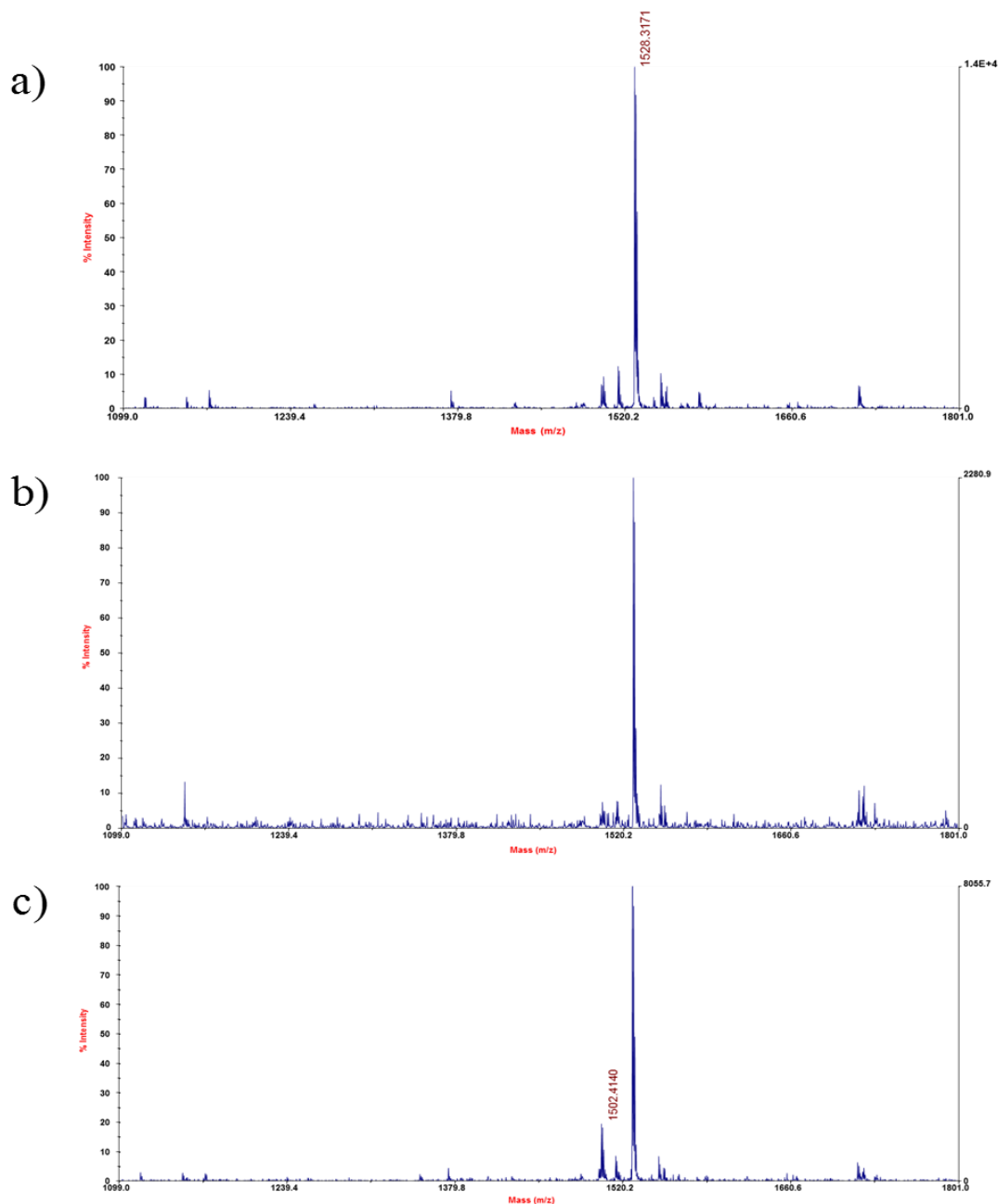


Figure 4.4 Compatibility of azide-peptide with dithiothreitol (DTT). (a) MALDI-TOF of a test peptide (MDYKMH₆) containing two AzHA residues (calc. 1527.78, obs. 1528.32) before treatment with DTT, and (b) after treatment with 5mM DTT, 55°C for (b) 30 min and (c) 1 h. The 1 h time point shows an additional peak consistent with the reduction of one of the azide residues (calc. 1501.77, obs. 1502.41).

4.2.2 Comparative scaffold-diverse *in vitro* selections

To assess the relative fitness of the bicyclic peptide library generation technology in an mRNA display selection (**Figure 4.5**), we chose to perform a side-by-side *in vitro* selection with a monocyclic library generated by cysteine-bisalkylation similar to those we generated previously.^{55,117,139} We designed five libraries, each containing two cysteine codons (one fixed at the 3' end of the random region and one that was varied), and one Phe codon that was varied in position between libraries (**Figure 4.6a-c and Table 4.1**). The initiator Met codon fixed the position of the azide to the N-terminus. This design resulted in the generation of five libraries which, when cyclized solely with α,α' -dibromo-*m*-xylene, resulted in macrocycles with 7-11 member rings. However, when subsequently cyclized a second time with CuAAC, these libraries generate a variety of highly-constrained, overlapping, θ -bridged scaffolds characteristic of knotted peptides (**Figure 4.5d**). Since target-scaffold matching is important from a ligand discovery perspective,¹⁶⁴ as the target for our selection we chose streptavidin, a highly studied protein with a well-defined binding pocket which has been shown in previous selections to prefer constrained peptides.^{165,166}

For Round 1 of the parallel selections against streptavidin, we initiated a single translation with an equimolar mix of the five library mRNAs pre-linked to the puromycin oligo and used incorporation of ³⁵S-cysteine for quantification. While immobilized on oligo(dT) resin, the mRNA-peptide fusion pool was bisalkylated with α,α' -dibromo-*m*-xylene (**Figure 4.5**). At this point, the fusions were split into monocyclic and bicyclic pools, reverse transcribed and captured onto Ni-NTA agarose. CuAAC was performed on the bicyclic pool generate the second triazole-containing cycle (**Figure 4.5**). The purified monocyclic and bicyclic fusions were independently incubated for 4 h with magnetic streptavidin beads. Bound fusions were competitively eluted with

D-biotin and separately PCR-amplified to initiate round 2. Stringency was increased by using additional washes of the beads beginning in round 2. While we noted enrichment beginning in round 4 and continuing through round 7 for the monocyclic selection (**Figure 4.7**), we observed no enrichment throughout the bicyclic selection, despite two additional rounds. The purified PCR-amplified DNA from both selections was sequenced using the Illumina MiSeq platform (**4.5.13**). The raw sequencing data was analyzed using AptaTOOLS, a comprehensive software collection designed specifically for the *in silico* processing of several types of *in vitro* selections. First, the data was preprocessed using AptaPLEX¹⁶⁷ by filtering out low quality reads and extracting the relevant sequence region. Remaining sequences in each selection cycle were then computationally analyzed using AptaCLUSTER¹⁶⁸ by elucidating and tracking the behavior of aptamers and aptamer families (clusters) throughout the sequenced portion of the selections. In conjunction with its graphical user interface AptaGUI,¹⁶⁹ the algorithm also determined global properties of the selections such as the convergence of the pool towards certain families of sequences, as well as local characteristics including, but not limited to, the abundance and enrichment rate of each individual sequence and cluster respectively.

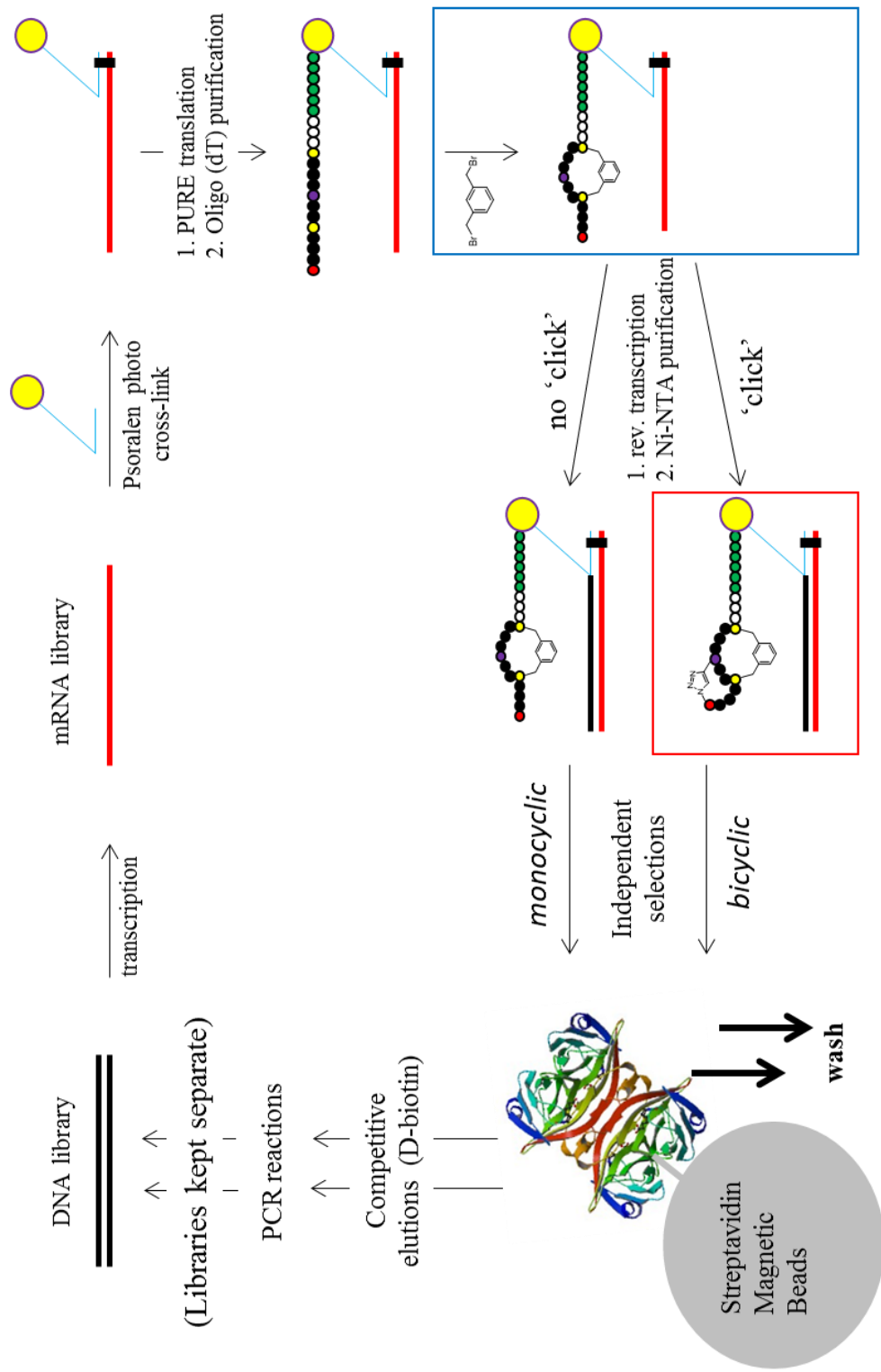


Figure 4.5 mRNA display scheme. mRNA-peptide fusion preparation highlighting the bisalkylation with α, α' -dibromo-*m*-oxylyene (blue box), followed by splitting of the pool, with half bicycled via CuAAC (red box). Purified mono- and bicyclic pools were then panned against streptavidin, binders were eluted and PCR-amplified to create the templates for round 2.

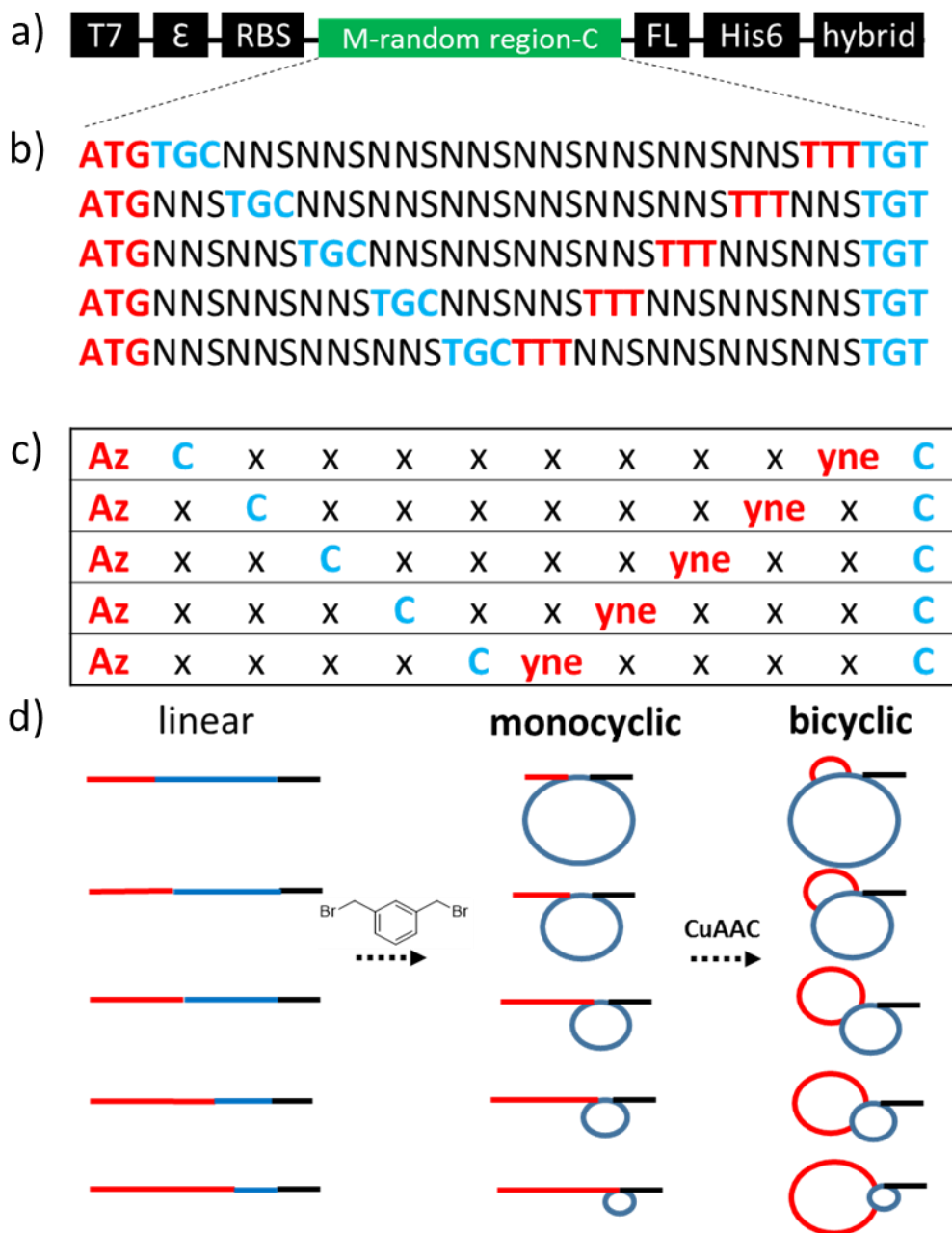


Figure 4.6 Design of mono- and bicyclic libraries. a) General DNA library elements, b) Random library region construction, highlighting fixed ‘clickable’ codons in red and bisalkylatable cysteine codons in blue. c) Peptide sequences generated from the library mRNA, **Az** = β -azido-L-homoalanine (AzHA), **yne** = *p*-ethynyl-L-phenylalanine (F-yne), **C** = cysteine. d) Representative cartoon depictions of scaffold diversity generated from the linear mRNA fusions prepared by bisalkylation (monocyclic) and bisalkylation followed by copper click (bicyclic) while fusions are immobilized on Ni-NTA.

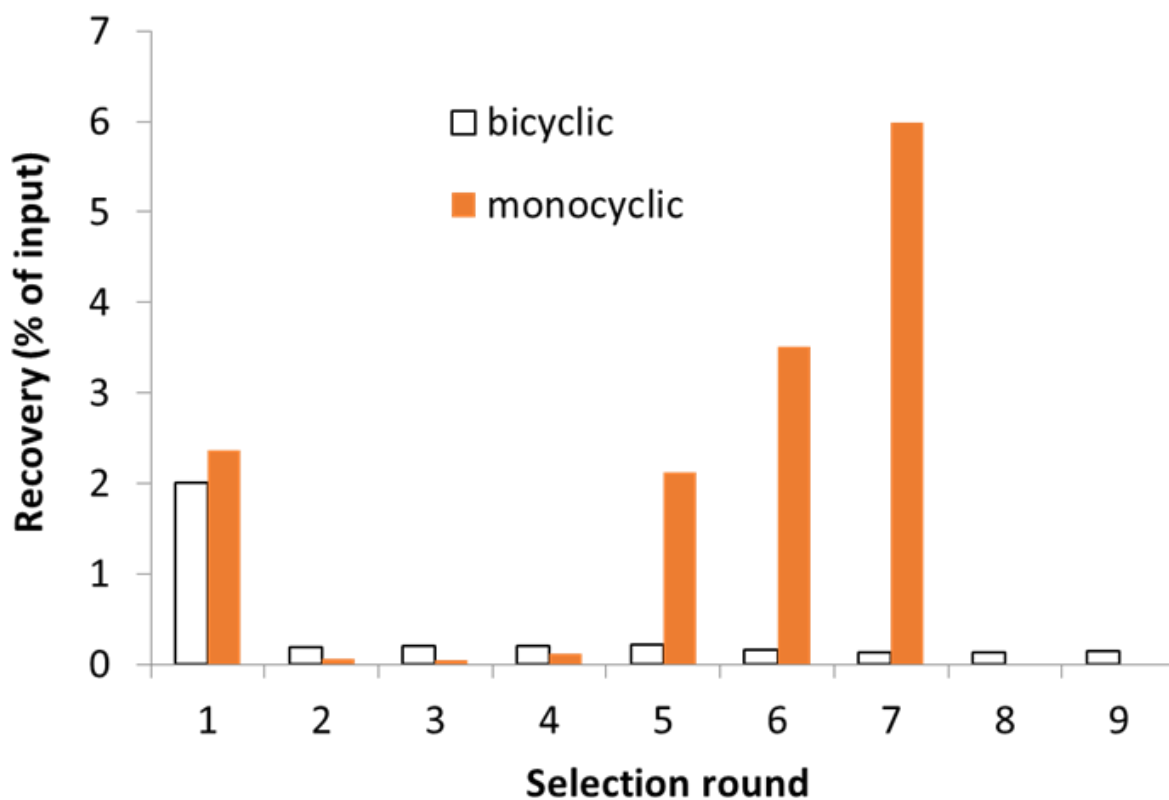
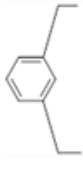


Figure 4.7 Selection enrichment. ³⁵S-cysteine quantification of monocyclic selection enrichment through 7 rounds. The bicyclization selection continued through round 9.

4.2.3 Sequence analysis

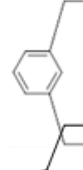
As expected based on the selection enrichment measurement (**Figure 4.7**), the monocyclic selection showed a high level of sequence convergence after 6 rounds (**Figure 4.8a**). Moreover, there was a clear scaffold preference—all of the most abundant sequences came from the MXXXXCFXXXXC library. The sequences were identified with the following nomenclature, using the most abundant peptide from the monocyclic selection as an illustrative example (7.1m): The first number (7) represents the round of selection the sequence was derived from, (1) is abundance rank within the given round, and (m) indicates the sequence was found in the monocyclic selection ('b' identifies sequences which originated from the bicyclic selection). Surprisingly, 4 of the top 8 sequences (9.1b=7.1m, 9.2b=7.2m, 9.3b=7.3m, and 9.8b=7.4m) from the bicyclic selection were also abundant in the monocyclic selection, although the percentage of these peptides in the final pool was significantly less (1-2% vs. 20-26%) (**Figure 4.8b**). The bicyclic selection also contained peptides that were not found at all in monocyclic selection winners (9.4b and 9.6b) as well as a linear mutant which did not contain a F-yne or a second cysteine (9.5lin). Based on these results, we decided to further analyze peptides 7.1m-7.4m in the linear, mono- and bicyclic configurations as well as peptide 9.4b, which was the most abundant unique bicyclic peptide. Finally, we also analyzed linear peptide 9.5lin as a point of comparison. (Note: The top 100 sequences for the monocyclic and bicyclic selections can be found in **Appendix Tables H1a,b and H2a,b**, respectively).

a)



ClusterID	Rd 7		Rd 6		Rd 5		Rd 4	
	count	freq(%)	count	freq(%)	count	freq(%)	count	freq(%)
7.1m	M T D P N C F	4188 26.75 1.494	6026 17.91 3.87	6776 4.63 93.82	161 0.049			
7.2m	M T D P N C F	3940 25.17 1.190	7114 21.14 3.01	10276 7.02 72.95	314 0.096			
7.3m	M T N P D C F	3257 20.81 1.128	6845 20.35 3.55	10240 6.99 62.59	419 0.064			
7.4m	M T N P D C F	447 2.86 1.058	908 2.70 2.23	1770 1.21 65.76	60 0.018			
7.5m	M T N P Q C F	210 1.34 0.673	671 1.99 1.77	1647 1.12 32.78	112 0.034			
7.6m	M T N P N C F	188 1.20 0.946	427 1.27 1.85	1004 0.69 50.86	44 0.013			
7.7m	M T D P D C F	112 0.72 0.652	369 1.10 1.42	1127 0.77 36.94	68 0.021			
7.8m	M T D P N C F	108 0.69 1.080	215 0.64 2.58	363 0.25 38.53	21 0.006			

b)



ClusterID	Rd 9		Rd 8		Rd 7		Rd 6		Rd 5		Rd 4	
	count	freq(%)	count	freq(%)	count	freq(%)	count	freq(%)	count	freq(%)	count	freq(%)
9.1b	M T D P N C F	2191 1.96 5.31	1232 0.370 5.12	731 0.07222 6.935	32 0.0104 9.61	6 0.00030						
9.2b	M T D P N C F	1586 1.42 4.10	1154 0.346 4.85	723 0.07145 6.28	39 0.0127 8.79	8 0.00061						
9.3b	M T N P D C F	1290 1.15 4.33	1004 0.301 5.24	735 0.07253 6.98	53 0.0173 10.12	10 0.00076						
9.4b	M H P Q N C F	120 0.11 2.80	128 0.038 2.70	144 0.01421 1.749	25 0.0081 6.44	7 0.00008						
9.5lin	M T I H Q W L Y H P Q	79 0.07 33.67	7 0.002 5.32	4 0.0004 N/A	0 0.0000 N/A	0 0.0000 N/A						
9.6b	M H P Y C F G F	62 0.06 9.25	20 0.006 3.38	18 0.00182 2.732	2 0.0007 3.61	1 0.00000						
9.7b	M T N P D C F	58 0.05 4.02	43 0.013 4.21	31 0.0031 N/A	0 0.0000 N/A	0 0.0000 N/A						
9.8b	M T N P D C F	40 0.04 3.98	30 0.009 3.26	28 0.0028 N/A	0 0.0000 0.00	1 0.0002 1.19	2 0.00015					

Figure 4.8 AptaTools round-by-round sequence analysis. Results of Illumina NextGen sequencing of a) rounds 4-7 of the monocyclic selection and b) rounds 4-9 of the bicyclic version. Round-by-round analysis includes tracking of the absolute sequence occurrence (count), percentage of sequences in the round (freq %), and fold-increase from the previous round (enrich). Cluster ID denotes pooled sequences with similar motifs. Highlighting indicates fixed flanking cyclization residues (red), intra-random region cyclization residues (aqua), and consensus residues within the random region (yellow). HPQ motif is in bold. Note: Top 100 most abundant sequences are in Appendix Tables H1a,b (monocyclic) and H2a,b (bicyclic)

4.2.4 Affinity for streptavidin

Each of the peptides was prepared with *in vitro* translation in the presence of ^{35}S -Cys, and the pertinent cyclization steps were carried out while the peptide was immobilized on Ni-NTA resin (**Figure 4.9**). As an initial screen for affinity, we chose to test each peptide in its linear, mono- and bicyclic forms for its ability to bind to 1 μM immobilized streptavidin. The four most abundant monocyclic selection winners (7.1m-7.4m) all showed enhanced binding to streptavidin in their monocyclic conformations relative to the linear and bicyclic forms (**Figure 4.10a**). The unique 9.4b knotted peptide showed binding in both the monocyclic and bicyclic configurations, although the bicyclic configuration bound to a greater extent. Since this peptide contained two F-yne residues (**Figure 4.10b**), there are two potential places for the click cyclization to occur. To test the preference, we independently substituted each of the F-yne residues with tyrosine giving 9.4b-7Y and 9.4b-10Y, to force the triazole-containing ring sizes to 10 or 7 amino acid-sized rings respectively, and performed the identical screen. The knotted peptide containing the 10AA triazole cycle (9.4b-7Y) bound to a much greater extent than the 7AA ring (**Figure 4.10b**).

We next determined the absolute affinity of our top binding peptides (7.2m, 7.3m, 9.4b, and 9.4b7Y) using a magnetic bead-based modification of the spin filter binding-inhibition assay (**Figure 4.11**).¹⁷⁰ Both 7.2m and 7.3m had K_d 's of approximately 300 nM. 9.4b as well as the 9.4b7Y mutant had K_d 's in the 500-600 nM range.

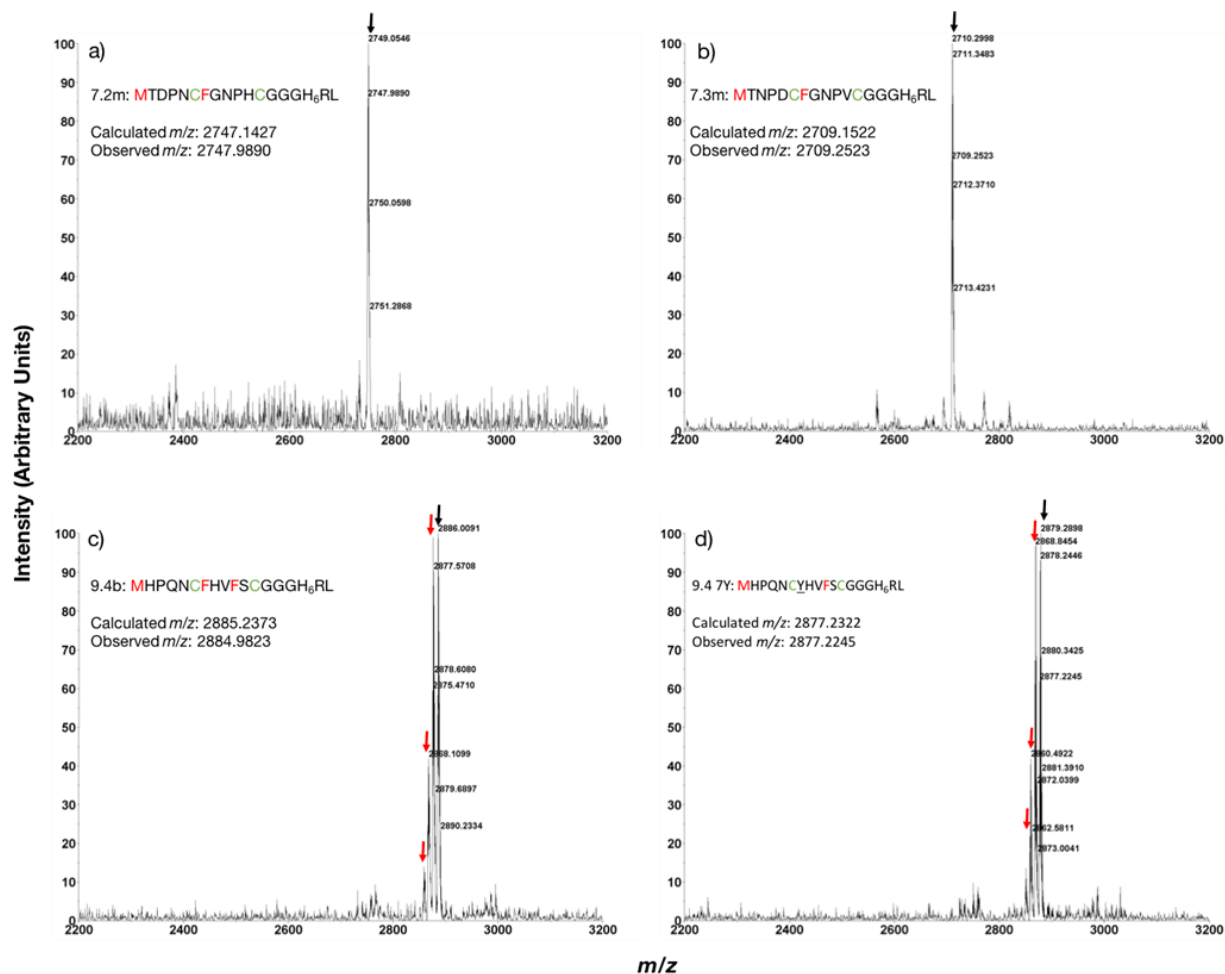


Figure 4.9 MALDI-TOF spectra of *in vitro* translated selection winners. Selection winners 7.2m (a), 7.3m (b), 9.4b (c), and 9.4b 7Y (d) were made using the PURE translation system, purified and cyclized on Ni-NTA resin. Black arrow indicates product of correct mass; red arrow indicates Q \rightarrow H misincorporation, which we demonstrated in 2.2.1 (Figure 2.4). (Experiment performed by Emil S. Iqbal)

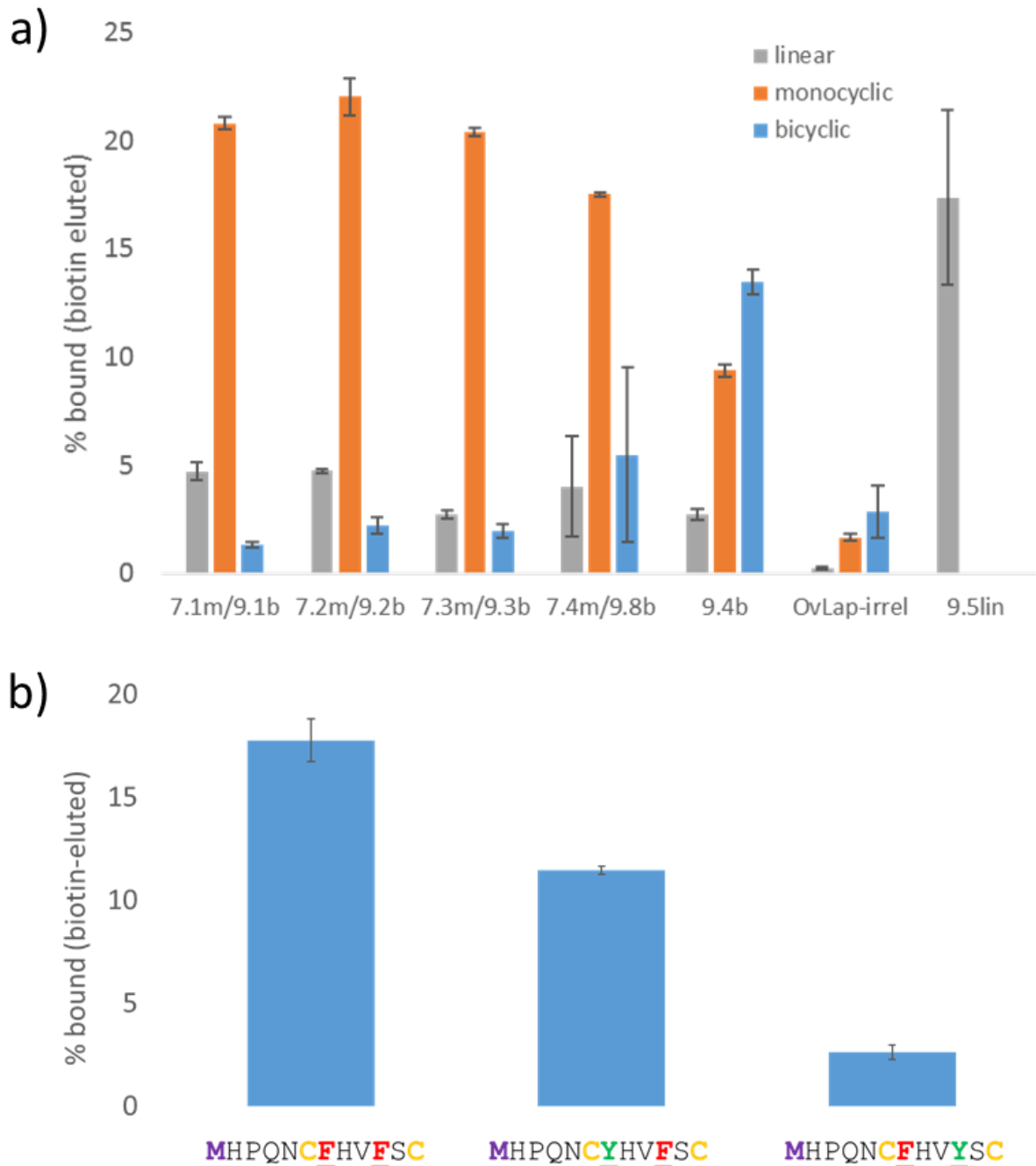


Figure 4.10 Relative binding to streptavidin. a) *In vitro* translated, ^{35}S -Cys-labeled peptides were created in linear, monocyclic, and bicyclic conformations while immobilized on NI-NTA resin, bound to $1\ \mu\text{M}$ streptavidin beads and eluted with 2.5mM D-biotin. Model peptide was used as an irrelevant control and 9.5lin was tested in only its linear form due to the absence of cyclizable residues. b) Selected bicyclic peptide 9.4b, with two F-yne residues (left). Tyrosine mutations to force the 10-member triazole cycle (middle) and the 7-member cycle (right).

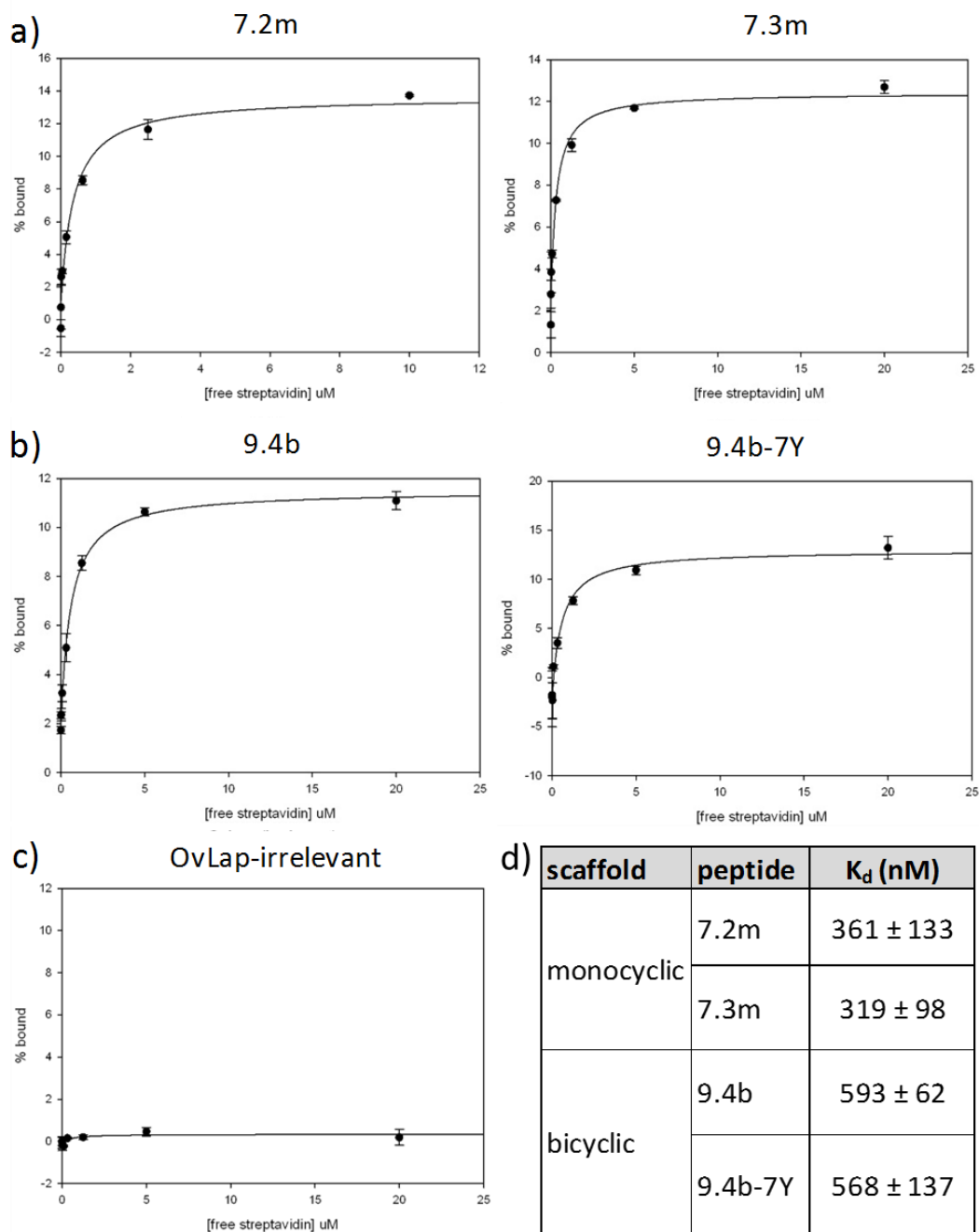


Figure 4.11 Binding curves of selection winners. Results of bead-based modification of a spin-filter binding inhibition assay to determine binding affinity.¹⁷⁰ a) monocyclic b) bicyclic c) irrelevant control (OvLap). Magnitude of fractional binding is expectedly low as a result of the presence of unincorporated ³⁵S-cysteine during the peptide purification process (Cys binds to the Ni-NTA resin along with our peptide). We have consistently observed between 75-90% of the signal represented as the unbindable fraction. Results represent experiments done at least in duplicate.

4.2.5 Bicyclization enhances protease stability

Others have demonstrated that knotted peptides are far more biostable and remarkably resistant to enzymatic degradation,^{127,171} qualities that increase the therapeutic potential of ligands discovered using selection methods. We selected the linear (9.5lin), monocyclic (7.2m) and bicyclic peptide (9.4b) variants and incubated them with immobilized chymotrypsin. At each time point, we removed an aliquot of peptide and measured its ability to bind immobilized streptavidin. As expected, the activity of the linear peptide was almost completely abolished after 1 h, but the monocyclic peptide 7.2m displayed far greater resistance to protease degradation, retaining 57% of its activity after 24 h relative to the control (**Figure 4.12a**). Interestingly, and in accordance with our hypothesis, 96% of bicyclic peptide 9.4b was able to bind to streptavidin after 24 h incubation with chymotrypsin. Furthermore, even after extended incubation with the protease, the bicyclic peptide showed remarkable retention of activity, despite the fact that it had an additional predicted cleavage site relative to the monocyclic peptide (**Figure 4.12b**).

While this assay demonstrates that protease stability directly correlates with constraint, there are a few caveats that bear mentioning. Since chymotrypsin is specific for peptides containing aryl amino acids, we make the assumption that F-yne is recognized by the enzyme. This assumption is supported by the fact that the F-yne in peptide 7.2m is the only putative chymotrypsin cleavage site and that this peptide is degraded over time. Second, since relative fractional binding is monitored via radioisotope binding it is possible that a distal label could be cleaved, yet the peptide still be bound to the protein. This scenario would result in artificially high ‘degradation’. Lastly, the peptides are all derived from different primary sequences. The monocyclic peptide 7.2m has a single cleavage site, while the other peptides each have two. These variables could potentially bias the stability results.

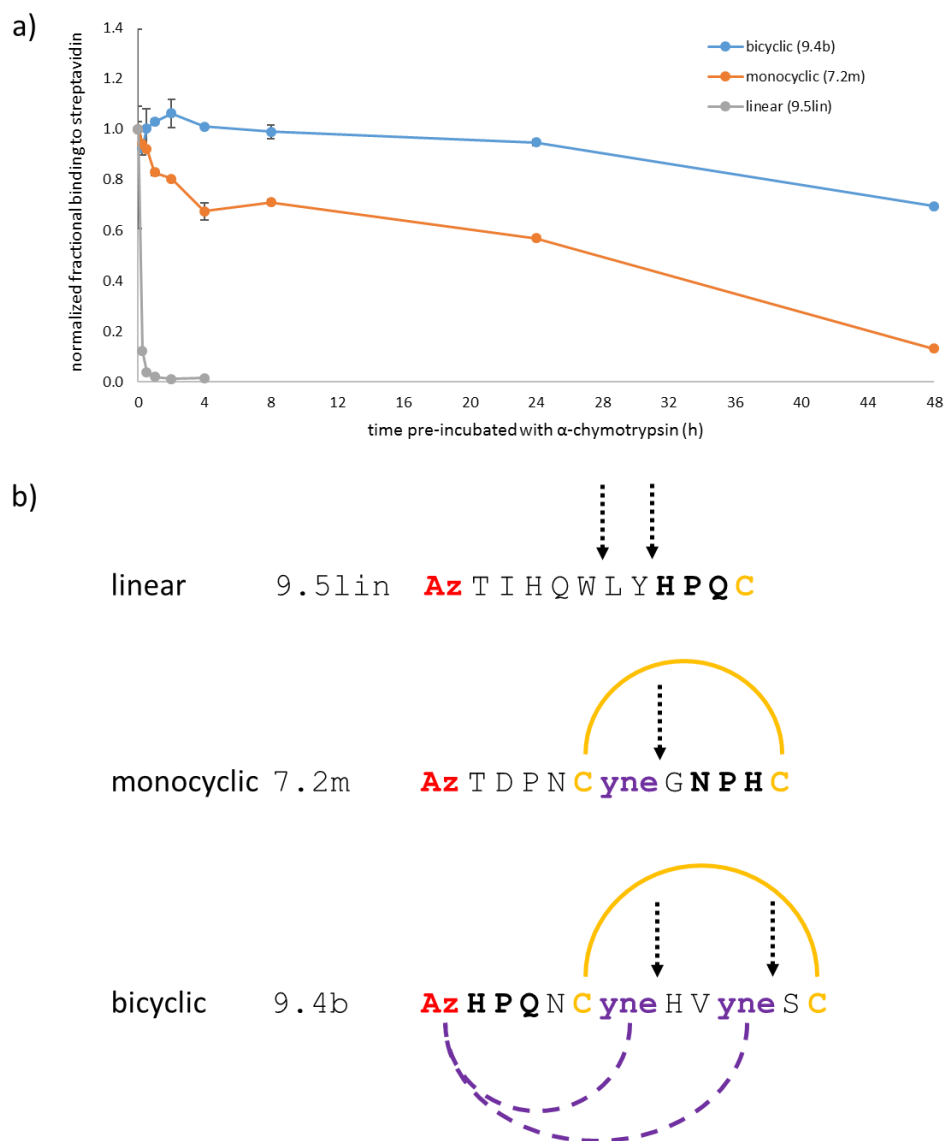


Figure 4.12 Protease stability. a) Normalized protease degradation of each of the three scaffolded peptides shown to bind streptavidin with the highest affinity (7.2m, 9.4b, 9.5lin). 5 nM peptide was added to 1.4U of immobilized α -chymotrypsin. Aliquots were removed at the indicated time points and added directly to 1 μ M streptavidin beads, incubated for 2 h, and eluted with biotin. Peptides were quantified by scintillation of 35 S-cysteine labeled peptides and normalized to the no protease time point. Results are an average of duplicate trials. b) α -chymotrypsin cleavage sites. Dotted arrows represent putative chymotrypsin substrates present in peptides used in the protease stability experiment. Yellow arc represents the α,α' -dibromo-*m*-xylene cyclization of two cysteines. Dotted purple arcs represent the two potential click cyclization products of the bicyclic peptide 9.4b. Linear peptide 9.5lin is lacking both a 2nd cysteine and an F-yne residue and as a result is inert to both chemistries performed during the mRNA-peptide fusion preparation steps in selection.

4.3 Discussion

The advent of Next Generation Sequencing (NGS) has proven to be a powerful tool in the ligand discovery/DNA template-based selection field. When combined with software such as AptaTools, which allows in-depth round-by-round analysis of enrichment rates, important ligands which may have otherwise gone unnoticed can now be uncovered. For the monocyclic selection, we observed high enrichment rates from round 4 to round 5 while the subsequent increase in the remaining rounds for each of the monocyclic peptides was much more moderate. This pairs nicely with our selection enrichment results based on radioactivity (**Figure 4.7**) and suggests that the selection had plateaued after round 5. At round 5 however, the abundance of the winner sequences was still quite low (<4%)--too low to meaningfully detect these sequences by traditional Sanger sequencing. Therefore, in principle, use of NGS and AptaTOOLS could shorten the number of rounds necessary to detect winners as compared to standard sequencing. For the bicyclic selection, use of NGS was essential for the detection of peptide 9.4b due to its low abundance.

Our streptavidin-binding peptides have a unique sequence motif as compared to previous selections. For example, the HPQ motif revealed in numerous selections against streptavidin^{165,172-175} was absent from our most abundant sequences in the monocyclic selection. Instead, sequencing revealed a consensus binding motif that included portions of the N-terminal linear and C-terminal cyclic regions (**Figure 4.13**). The only significant variability in the top 8 monocyclic sequences was in the 11th position. The fact that these peptides showed much weaker affinity in their linear forms leads us to conclude that cyclization of these sequences is essential for a high affinity interaction with streptavidin. When we added a second cycle to these peptides, the binding was significantly reduced (**Figure 4.10a**). It is expected that the second cycle will dramatically change

the conformation of these peptides, and we surmise that the bicyclic peptides are locked in a compacted conformation unproductive for binding.

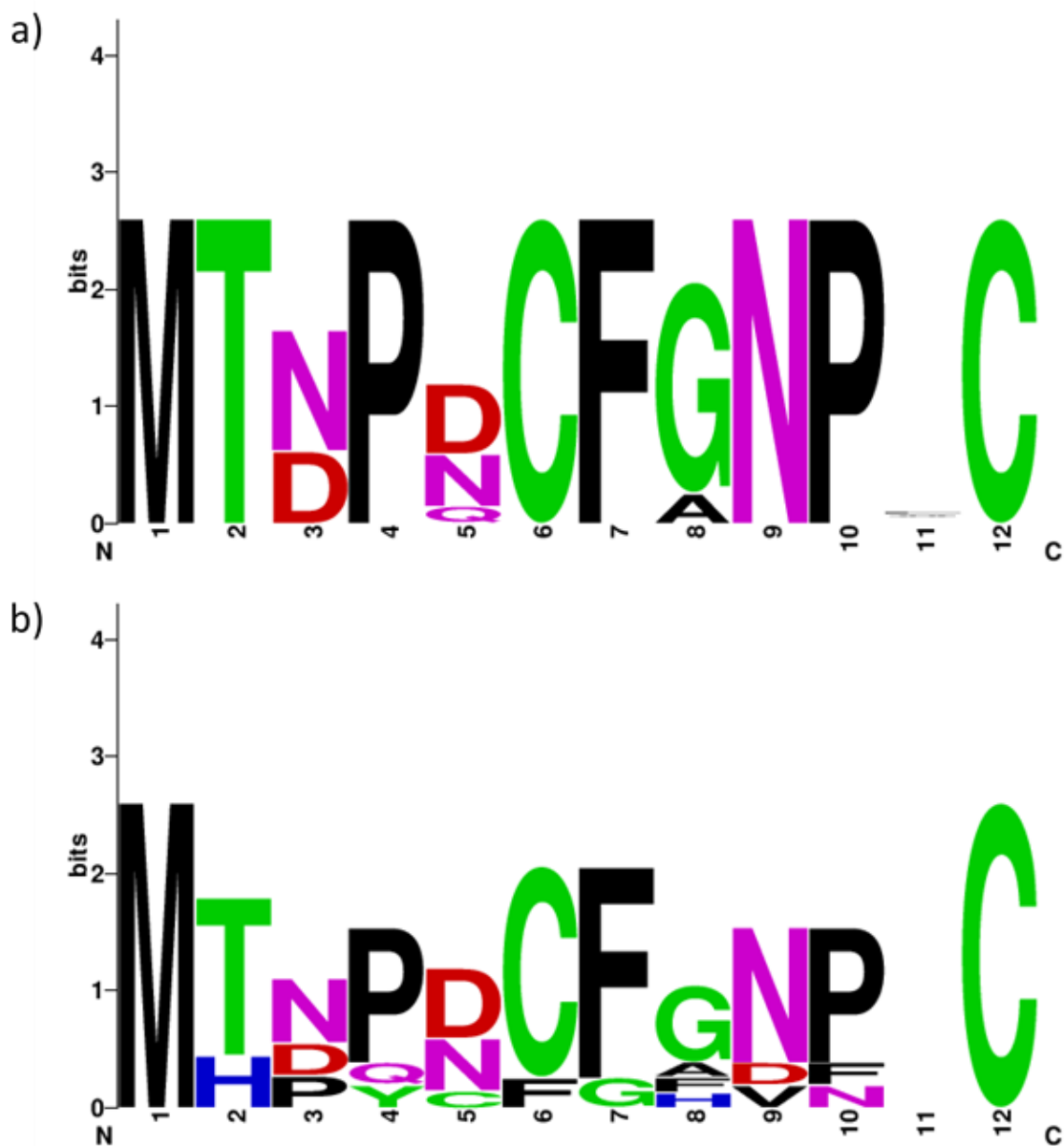


Figure 4.13 Selection winner homology. a) The monocyclic selection resulted in a high degree of homology, both in the cyclic portion (6-12, with the exception of residue 11) and the linear portion (1-5) of the top 8 most abundant sequences. While the bicyclic selection b) showed a lesser degree of homology there was a strong similarity between the consensus regions of the two selections. Residue 11 was also highly variable in the bicyclic selection, indicating that the residue was likely not critical for interaction with streptavidin. Top 100 sequence homology diagrams are in Appendix Figures H1 (monocyclic) and H2 (bicyclic).

On the other hand, our bicyclic selection-derived peptide 9.4b did bind effectively as a knotted peptide. This peptide has two F-yne residues as potential cyclization points. When we removed each of these in turn, only the bicyclic peptide with the F-yne at position 10 led to effective binding (**Figure 4.10b**). This is significant, because this F-yne was located in the random region of the library, while the F-yne at position 7 was fixed. From a library design standpoint, this is quite interesting; this peptide effectively “chose” a scaffold that was underrepresented in the library. Our low enrichment rates from the bicyclic selection could, therefore, be a consequence of our biased scaffold choice.

The peptide winners common to the two selections (**Figure 4.8a,b**) did not bind to streptavidin in their bicyclic configurations (**Figure 4.10a**). It is therefore quite surprising that these sequences were present in the bicyclic selection winners. First, we note that the enrichment rates of these peptides are quite low in comparison to the enrichment values in the monocyclic selection prior to its plateau. The low enrichment explains why we did not see significant convergence in sequences, even though we performed 2 additional rounds relative to the monocyclic selection. While we demonstrated that the click cyclization results in high conversion to the bicyclic peptide using our model peptide, it is certainly possible that the cyclization efficiencies of these sequences under selection conditions are less than 100%. If so, the low enrichment could be due to a small amount of the monocyclic sequences being present and captured during the bicyclic selection. Also, although we were very careful to avoid this, we cannot definitively rule out that cross contamination could be responsible for the presence of the monocyclic sequences in the bicyclic library since the selections were performed side-by-side.

In summary, we have developed the technology to create knotted, bicyclic peptides using a cell-free translation system, and we have optimized the method to make it compatible with

mRNA display to generate mRNA-bicyclic peptide fusions for *in vitro* selection. We designed our parallel selection strategies in order to be able to evaluate the potential for either mono- and bicyclic libraries to uncover peptide binders. We conclude that high-affinity versions of both types of ligands are present in the diverse libraries we created; however, the bicyclic peptides have the advantage of dramatically enhanced (>4-fold) protease stability. Moreover, because we have fixed the position of the cyclization residues to only a few possibilities, we have only probed a very small subset of the potential diversity of these bicyclic libraries. It should therefore now be possible to randomize these positions to create a whole host of diverse knotted peptide libraries using mRNA display. This diversity should enable the discovery of protease-stable bicyclic ligands to a wide range of therapeutic targets.

4.4 Future Directions

Based on our premise that random cyclizable codon placement will greatly increase scaffold diversity and will result in target selection of the optimal conformation, we will perform a second *in vitro* selection against streptavidin using this strategy. We will also use this method to conduct a selection against DNA repair protein XRCC4.

4.5 Experimental

Table S1. Oligonucleotides used in selection	
	Sequence
Library [MC(X ₈)FC]	5'-CTAGCTACCTATAGCCGGTGGTGATGGTGATGATGGCCA CCGCCACA AAA SNNNSNNSN NSNNSNNSNNSNNGCACA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGAG TCGTATTA A TTTCG-3'
Library [M(X)C(X ₆)F(X)C]	5'-CTAGCTACCTATAGCCGGTGGTGATGGTGATGATGGCCA CCGCCACA SNNAAA SNNSN NSNNSNNSNNSNNGCASNCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGAG TCGTATTA A TTTCG-3'
Library [M(X ₂)C(X ₄)F(X ₂)C]	5'-CTAGCTACCTATAGCCGGTGGTGATGGTGATGATGGCCA CCGCCACA SNNSNNA AASN NSNNSNNSNNGCASNNSNCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGAG TCGTATTA A TTTCG-3'
Library [M(X ₃)C(X ₂)F(X ₃)C]	5'-CTAGCTACCTATAGCCGGTGGTGATGGTGATGATGGCCA CCGCCACA SNNNSNNSNAA ASNNSNNGCASNNSNNSNCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGAG TCGTATTA A TTTCG-3'
Library [M(X ₄)CF(X ₄)C]	5'-CTAGCTACCTATAGCCGGTGGTGATGGTGATGATGGCCA CCGCCACA SNNNSNNSNNSN NAAA GCASNNSNNSNNSNCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGA GTCGTA TTA A TTTCG-3'
ExtRTP2	5'-TTTTTTTTTTTTTTTCGATGATACGAA GATTC TCAAGATTA GGGCTAGCTACCTATAGC CGGTGGTGA TGGTGA TGA TGGCCA CCGCCACA -3'
ExtT7Fwd	5'-CGA AATTAATACGACTCA CTATAGGGTTAACTTTAAGAAGGA G-3'
UniRevP2	5'-CGA TGATACGGAAGATTC TCAAGATTA GGG-3'
T7Fwd	5'-CGA AATTAATACGACTCA CTATAGGG-3'
T7Rev	5'-GCTAGTTA TTGCTCAGCGG-3'
OvLap	5'-CCTA TAGCCGGTGCTA A TGGTGA TGGTGA TGA TGGCCA CCGCCACAGACA CTCGTA A AAA CGCTGCA GTTGGTCA CCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGAG TCGTATTA A TTTCG-3'
9.4b	5'-CTAGCTA CCTA TAGCCGGTGGTGA TGGTGA TGA TGGCCA CCGCCACAGAGA ACA CG TGAAAGCA GTTCTGCGGGTGCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGA GTCGTA TTA A TTTCG-3'
9.4b-7Y	5'-CTAGCTA CCTA TAGCCGGTGGTGA TGGTGA TGA TGGCCA CCGCCACAGAGA ACA CG TGA TAGCA GTTCTGCGGGTGCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGA GTCGTA TTA A TTTCG-3'
9.4b-10y	5'-CTAGCTA CCTA TAGCCGGTGGTGA TGGTGA TGA TGGCCA CCGCCACAGAA TACAG CG TGAAA GCA GTTCTGCGGGTGCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGA GTCGTA TTA A TTTCG-3'
7.4m/9.8b	5'-CTAGCTA CCTA TAGCCGGTGGTGA TGGTGA TGA TGGCCA CCGCCACAGAGGGGGTTG GCAAA GCAGTCCGGGTTCGTCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGA GTCGTA TTA A TTTCG-3'
7.3m/9.3b	5'-CTAGCTA CCTA TAGCCGGTGGTGA TGGTGA TGA TGGCCA CCGCCACACAGGGGGTTG CCAAA GCAGTCCGGGTTCGTCA TAGTATA TCTCCTCTTAAA GTTAA ACCCTA TAGTGA GTCGTA TTA A TTTCG-3'

Table 4.1 Oligonucleotides used in mono- vs. bicyclic scaffold selection

4.5.1 Amino acids and translation reagents: For natural amino acid preparation, see **2.5.1**.

Preparation of *in vitro* translation components and reagents is described in **2.5.3**. Unnatural amino acids are described in **3.5.1**.

4.5.2 Selection Reagents: Magnetic streptavidin beads tested for selection: Dynabeads M-280 streptavidin (#11205D) and M-270 streptavidin (#65305) (Invitrogen, Life Technologies), streptavidin magnetic particles (Roche #11641778001); Oligo-(dT)₂₅ magnetic beads (NEB #S1419S); Superscript III reverse transcriptase (Invitrogen #18080093); RNase inhibitor (Bioline #65027); TurboDNase (Invitrogen #AM2238); Taq DNA polymerase (NEB #M0237L). All oligonucleotides were purchased from Integrated DNA Technologies (IDT). α,α' -dibromo-*m*-xylene was purchased from Fluka.

4.5.3 Preparation of mRNA-peptide fusion library

T7 *in vitro* transcription¹¹⁶ was performed with an equimolar mix of the five bottom strand oligo libraries (see Table 4.1), following pre-annealing to the forward primer (70° C for 5min, followed by cooling on ice for 1min.). Library mRNA was photo-crosslinked to the linker containing puromycin at the 3' end as described previously.¹⁰⁸ For round 1 of the parallel selections, a single 2 mL standard translation reaction was initiated with the addition of the puromycin-linked mRNA, incubated at 37° C for 1.25 h, supplemented with 550 mM KCl and 55 mM MgCl₂, returned to the incubator for 1.5 h and subsequently frozen overnight at -80° C. The resulting mRNA-peptide fusions were diluted 6-fold with oligo-dT binding buffer (20 mM Tris-HCl pH 7.8, 10 mM EDTA, 1 M NaCl, 0.2% Triton X-100, 0.5 mM TCEP), added to 1.5 mL of oligo-dT magnetic beads which were equilibrated thrice with 5 mL of oligo-dT binding buffer, and rotated at 4° C for 30 min. The beads were washed twice with 5 mL of oligo-dT wash buffer (20 mM Tris-HCl pH 7.8, 0.3 M NaCl, 0.1% Triton X-100, 0.5 mM TCEP), and the first cyclization was performed on resin by the

addition of 5 mL of cyclization buffer (20 mM Tris-HCl pH 7.8, 0.66 M NaCl, 3 mM α,α' -dibromo-*m*-xylene, 33% acetonitrile (v/v), 0.5 mM TCEP) and rotated at RT for 30 min. The beads were washed once with 5 mL of oligo-dT wash buffer containing 5 mM BME (in lieu of TCEP), to quench the unreacted linker. Beads were washed a second time with 5 mL of wash buffer containing TCEP and eluted in 1 mL fractions with 0.5 mM TCEP. The five elutions with the highest scintillation counts were pooled and precipitated with 4 vol of ethanol, 0.1 vol of 3 M KOAc, pH 5.2 and 0.001 vol of 5 mg/mL glycogen. The pellet was resuspended in 543.5 μ L of water and split into two portions for the monocyclic and bicyclic selections.

4.5.4 Monocyclic.

One half of this solution was reverse transcribed in a final volume of 400 μ L in the presence of RT mix (0.5 μ M RT primer, 0.5 mM dNTPs, 5 mM MgCl₂, 1 mM DTT, 2 U/ μ L RNase inhibitor, 5 U/ μ L Superscript III, 1x First Strand buffer) at 55° C for 30 min. The RT reactions were subsequently diluted 5-fold with denaturing Ni-NTA binding buffer (100 mM NaH₂PO₄, 10 mM Tris-HCl, 6 M guanidinium hydrochloride, 0.2% Triton X-100, 5 mM BME, pH 8) and added to 100 μ L of Ni-NTA agarose resin (MCLab) in a 10 mL BioRad disposable column. Columns were placed on a tumbler at 4° C for 1 h. The monocyclic fusions were washed four times with 3 mL of Ni-NTA wash buffer (100 mM NaH₂PO₄, 300 mM NaCl, 0.2% Triton X-100, 5 mM BME, pH 8) and eluted in one column volume fractions with Ni-NTA elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 350 mM imidazole, 0.2% Triton X-100, 5 mM BME, pH 8). Fractions containing significant radioactivity were pooled and dialyzed overnight against selection buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 4 mM MgCl₂, 0.25% Triton X-100).

4.5.5 Bicyclic

The second half of this solution was reverse transcribed in a final volume of 400 μ L in the presence of RT mix (0.5 μ M RT primer, 0.5 mM dNTPs, 5 mM $MgCl_2$, 1 mM DTT, 2 U/ μ L RNase inhibitor, 5 U/ μ L Superscript III, 1x First Strand buffer) at 55° C for 30 min. The RT reactions were subsequently diluted 5-fold with denaturing Ni-NTA binding buffer (100 mM NaH_2PO_4 , 10 mM Tris-HCl, 6 M guanidinium hydrochloride, 0.2% Triton X-100, 5 mM BME, pH 8) and added to 100 μ L of Ni-NTA agarose resin (MCLab) in a 10 mL BioRad disposable column. Columns were placed on a tumbler at 4° C for 1 h. After binding to the Ni-NTA agarose, bicyclic fusions were washed once with 3 mL of Ni-NTA wash buffer, and 3 mL total volume of CuAAC cyclization reagents were added (100 mM phosphate pH 8, 1 mM $CuSO_4$, 300 mM NaCl, 2 mM TBTA, 33% DMSO (v/v)). The solution was degassed with argon prior to the addition of 10 mM sodium ascorbate. An argon blanket was placed over the solution and the tube was sealed and rotated at room temperature for 2h. After incubation, the resin was washed once with 3 mL of Ni-NTA wash buffer containing 10% DMSO (v/v). Following two additional washes with 3 mL of Ni-NTA wash buffer, the purified fusions were eluted, pooled and dialyzed similarly to the monocyclic fusions. The translation volume was reduced to 500 μ L for rounds 2-4, 250 μ L for rounds 5-7. Purification reagent usage, wash and elution volumes were adjusted accordingly to coincide with translation yield.

4.5.6 Selection against streptavidin

Two tubes containing 125 μ L Dynabeads M-280 Streptavidin (Invitrogen #11205D) were equilibrated twice with 1 mL selection buffer and 0.5 pmols (3.01×10^{11} fusions, 11.75-fold above theoretical library diversity) each of monocyclic or bicyclic dialyzed fusions (125 μ L total volume) were added to the respective tubes (with 0.1% BSA) and rotated at 4° C for 1h. The beads were

then washed twice with 1 mL selection buffer and streptavidin-binding fusions were competitively eluted for 4 h (to account for slow off-rates) with 200 μ L of 2 mM D-biotin in selection buffer. Eluted fusions (2.4% and 2.0% of monocyclic and bicyclic input, respectively) were dialyzed overnight against 0.1% Triton X-100. Following PCR amplification using ExtT7fwd and UniRev2 primers (2 min. at 94°C, followed by 24 rounds of 94° C (30s), 65° C (30s), 72° C (45s), library DNA was extracted with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with 3 vol. ethanol and 0.1 vol 3 M KOAc pH 5.2. The resuspended DNA served as the templates for Round 2 of the parallel selections. The bead-to-fusion volume ratio was kept constant throughout the selection.

4.5.7 Ribosomal translation and preparation of selected labeled peptides for binding and stability studies

T7 transcriptions (500 μ L) were performed using either PCR product of selection clones (T7 Fwd and T7Rev primers-see **Table 4.1**) or bottom strand oligos (IDT) annealed to the T7 forward primer to produce mRNA used to initiate a standard translation reaction (as described in **2.5.4**). ³⁵S-cysteine (10 μ Ci) was used when the sequences contained Cys and ³⁵S-methionine (8 μ Ci) when no cysteines were encoded. Individual standard 250 μ L reactions were conducted with each selection winner mRNA, incubated for 80 min., and quenched with 750 μ L of TBS (50 mM Tris-HCl pH 8, 300 mM NaCl, 5 mM BME). Reactions were subsequently split 1:2:2 into three tubes containing 30:50:50 μ L of Ni-NTA slurry to produce the linear:monocyclic:bicyclic conformation of each peptide. Peptides were incubated at 4° C with the resin for 1 h. Linear peptides were subsequently washed six times with 500 μ L TBS. Cyclization chemistries were conducted with peptides immobilized on Ni-NTA resin. To bisalkylate the two cysteines, bound peptides targeted for mono- and bicyclization were washed once with 500 μ L TBS containing 0.5 mM TCEP (in

lieu of BME) and incubated for 45 min at RT after the addition of 500 μ L of bisalkylation buffer (20 mM Tris-HCl pH 7.8, 0.5 M NaCl, 3 mM α,α' -dibromo-*m*-xylene, 33% acetonitrile (v/v), 0.5 mM TCEP). These peptides were washed once with 500 μ L TBS (containing 5 mM BME to quench unreacted brominated linker), and monocyclic constructs were washed three additional times with 500 μ L TBS. Click buffer (100 mM potassium phosphate pH 8, 500 mM NaCl, 1 mM CuSO₄, 2 mM TBTA, 33% DMSO (v/v)) was added to the bound monocyclic peptides targeted for bicyclization. After degassing with argon, 10 mM sodium ascorbate was added to make total volume 500 μ L. A blanket of argon was added to the reactions, tubes were sealed and rotated at RT for 2 h. Bicyclic peptides were then washed once with 500 μ L TBS containing 10% DMSO, and twice more with 500 μ L TBS. To elute, all bound peptides were incubated for 10 min. with 50 μ L of 1% TFA, centrifuged at 5900 RPM to collect the filtrate, neutralized with 1 eq. 3M KOH, buffered with 2 μ L of 1 M Tris-HCl pH 7.6, aliquoted and stored at -20° C. Peptide concentrations were determined by scintillation count of the elution in duplicate and corrected for background by subtraction of a control reaction lacking mRNA.

4.5.8 Relative streptavidin binding

In vitro transcribed and radiolabeled peptides prepared above were diluted in SBB to 20 nM. 10 μ L of M-280 streptavidin beads were equilibrated three times each with 100 μ L of SBB. The beads were bound to a magnetic stand and the buffer was removed and 20 μ L of each diluted peptide was added to each tube. The beads were resuspended with diluted peptide solution, resulting in approximate streptavidin-on-bead concentration of 1 μ M, calculated based on manufacturer's reported biotin binding capacity. Tubes were placed on a rotisserie at 4° C for 2 h, then placed on the magnetic stand and the supernatant was removed. The beads were washed twice each with four CV of SBB (with mild vortex to ensure resuspension). Washes were

combined with the flow through and a portion was counted via scintillation and considered as the unbound fraction. 200 μ L of SBB with 2.5 mM D-biotin was added to each tube and rotated at RT for 2 h. The supernatant was removed and a portion was counted as the bound fraction. Fractional, or % bound to streptavidin and eluted with D-biotin, was calculated with the following equation: % bound = [bound/(unbound + bound) x 100]. % bound was determined for each peptide and the highest relative binders for each of the three conformations were selected for further study.

4.5.9 Determination of peptide K_a

In vitro translated, radiolabeled peptides prepared as described above (4.5.8) were diluted in SBB to 8 nM and added 1:3 to streptavidin in SBB (1.2 nM – 20 μ M final 4-fold serial dilution), resulting in 2 nM peptide final concentration in 20 μ L total volume. Reactions were incubated at 4° C for 2 h and added to 10 μ L washed and dried Dynabeads M-280 streptavidin magnetic beads, mixed and incubated for 1 min at RT. Tubes were then placed on a magnetic stand for 1 min and the supernatant was removed. 17 μ L of the supernatant was added to 2 mL of Econo-Safe scintillation fluid (RPI) and counted on a Beckman scintillation counter for 5 min. This was the fraction which bound the free streptavidin (B). The beads were resuspended in 50 μ L of SBB, vortexed vigorously and 40 μ L was counted as above. This was the unbound fraction which was capable of binding the matrix (U). Fractional binding was then calculated using the following equation: % bound = B/(B+U) x 100%. To account for the portion of the affinity purified peptide mixture incapable of binding streptavidin (primarily unincorporated radiolabel which bound the nickel resin), the assay containing no free streptavidin was subtracted from each streptavidin-containing assay, effectively providing a baseline for % bound calculation. Data was then plotted

using SigmaPlot and a curve was generated using a hyperbolic dynamic fit with the following equation: $y = y_0 + [ax/(b+x)]$. Experiments were done at least in duplicate.

4.5.10 Chymotrypsin stability assay

Radiolabeled linear (9.5lin), monocyclic (7.2m) and bicyclic (9.4b) peptides prepared above (4.5.8) showing the highest relative streptavidin affinity were selected for the protease stability experiment. Surface-activated magnetic beads (Dynabeads M-270 epoxy, Invitrogen #14301) were decorated with α -chymotrypsin (Sigma, C4129). To do this, we added 1 mL of 0.1 M sodium phosphate buffer pH 7.4 (buffer A) to 5 mg of beads, vortexed for 30 s, rotated for 10 min. at RT, then removed the supernatant. The wash was repeated twice, and the beads were resuspended in 100 μ L of buffer A. 100 μ g of 1 mg/mL chymotrypsin (in buffer A) was added to the tube and mixed thoroughly. Following the addition of 100 μ L of 3 M ammonium sulfate in buffer A, the tube was rotated at 37 °C for 18 h. The supernatant was removed and the chymotrypsin beads were washed a total of four times with 1 mL of PBS with 0.1% BSA (buffer B), mixing thoroughly after each wash. The beads were resuspended in 185 μ L of buffer B and the chymotrypsin activity was calculated based on the initial enzyme activity and assuming 100% conjugation to the activated beads and retention of activity during the coupling process (13.4 U/mL of suspended beads).

The peptides were diluted to 8 nM with streptavidin binding buffer (SBB, 40 mM Tris-HCl pH 7.4, 300 mM KCl, 2 mM EDTA, 5 mM BME, 0.013% Triton X-100) and 350 μ L (2.8 pmols) of each peptide was added to 1.4 U of chymotrypsin beads (unit calculation based on chymotrypsin functionalization input and assumes 100% conjugation efficiency and retention of activity) in a 1.7 mL tube and placed on a tumbler at RT. At the indicated time points the tube was placed on a magnetic stand and 25 μ L was removed from the tube and placed at -20° C. After 24 h the reaction was incubated for an additional 24 h at 37°C. Each time point was thawed on ice, mixed and 10

μL was added to 5 μL of M-280 streptavidin beads pre-equilibrated thrice with 200 μL of SBB. The solution was allowed to reach equilibrium on a rotisserie at 4° C for 2h. Tubes were bound to the magnetic stand and the supernatant removed. The beads were washed twice each with four CV of SBB. The washes were combined with the supernatant, mixed and a portion was counted via scintillation as unbound peptide. 200 μL of SBB containing 2.5 mM D-biotin was added to the beads and the tubes were rotated at RT for 2 h to selectively elute the bound labeled peptides. After binding to the magnet, the supernatant was removed, and a portion was counted as bound peptide. Fractional, or % binding to streptavidin was calculated with the following equation: % bound = [bound/(unbound + bound) x 100]. % bound for each peptide was normalized to the no chymotrypsin (0 time) control and plotted as a function of time.

4.5.11 Compatibility of azide-peptide with DTT in selection:

In vitro translation reactions (41 μL) were initiated with mRNA encoding the sequence MDYKMH₆ (**Table 2.1**), with the addition of only the encoded amino acids at the following concentrations (200 μM Asp, Lys, Tyr; 400 μM His; 2.5 mM AzHA). The reactions were incubated at 37° C for 1h to produce the peptides containing two AzHA. 100 μL of TBS, with and without DTT (0 and 5 mM final concentration) was added and the reactions were incubated at 55° C for 30 and 60 min. 359 μL of Ni-NTA denaturing binding buffer was added (total volume of 500 μL) and the contents were added to 30 μL of Ni-NTA resin in a microcentrifuge filter tube and rotated at 4° C for 1 h. Beads were washed three times each with 500 μL of TBS with 5 mM BME and the peptides were subsequently eluted with 30 μL of 1% TFA. The filtrate was desalted and concentrated by microchromatography, eluted with 5 μL of 10 mg/mL CHCl₃ in 70:29.9:0.1 ratio (by vol) of acetonitrile:water:TFA. 1 μL of each eluted peptide was spotted on a plate and analyzed by MALDI-TOF.

4.5.12 mRNA stability with click reagents (TBTA vs. THPTA):

CuAAC reagents (100 mM potassium phosphate pH 8, 500 mM NaCl, 1 mM CuSO₄) were added to two 0.6 mL microcentrifuge tubes along with 5 μM mRNA (202 bases). Tube 1 contained 2 mM TBTA ligand w/ 33% (v/v) DMSO; Tube 2, 10 mM THPTA ligand. Tubes were degassed with argon, 10 mM sodium ascorbate was added and a small septum was used to seal each tube. A blanket of argon was placed over each reaction mixture. Total volume of each tube was 100 μL. At the indicated time points, 20 μL was removed from each tube with a microsyringe, quenched with 5 μL of 200 mM EDTA and frozen at -20°C. Samples were analyzed by 10% urea-PAGE, stained with ethidium bromide and imaged on a ChemiDoc MP imaging system (BioRad).

4.5.13 Illumina NextGen Sequencing

DNA concentrations were verified using a Thermo Fisher Scientific Qubit fluorometer. Libraries were constructed using KAPABiosystems Library Preparation Kit protocol following standard protocol. In brief, DNA was submitted to end repair, 3' ends adenylation followed by adaptor ligation. DNA libraries were submitted to 5 cycles of PCR amplification and then quantified by qPCR assay using KAPA Library Quant kit. Libraries were normalized and all samples were pooled in equimolar amounts. Sequence was performed on the Illumina MiSeq instrument using 2 x 150 paired-end recipe. DNA mixing ratios are in **Appendix Table H7**.

Chapter 5. Scaffold-Diverse Peptide Libraries;

***In vitro* Selection Against Streptavidin #2**

5.1 Introduction

The ability to generate trillions of unique peptide sequences using *in vitro* display methods such as mRNA display make these ligand discovery techniques very powerful tools. The numerous cyclization strategies used to generate mono- and bicyclic peptide libraries for use with these technologies have resulted in ligands with improved target affinity.^{64-66,165,166,175} However, current cyclization techniques typically require fixation of the residue(s) necessary for cyclization within the otherwise random region of the library. In most cases, only a single library conformation, or ring size, is generated for target binding during selection.^{119,120,176,177} This results in the creation of a very limited number of peptide conformations and the ring size, or scaffold, required for optimal sequence alignment with the target binding interface may not be present among the purified fusion pool. Consequently, high affinity macrocyclic binders may go undiscovered.

We hypothesize that designing libraries with only one of the two codons required for cyclization fixed within the sequence, while allowing the random insertion of the compatible residue to determine ring size, would result in very diverse scaffolds (**Figure 5.1a**). Furthermore, the application of a second orthogonal cyclization method using this same approach would either create monocyclic peptides with different linkages, bicyclic peptides with a similar degree of diversity or even linear peptides in the absence of the necessary cyclizable residues. By combining our ability to apply two orthogonal chemistries to create knotted peptide libraries (**4.2.1**) for mRNA display with the semi-random cyclization residue design we describe herein, we can generate a fusion pool that contains an unprecedented degree of scaffold diversity to enhance the typical sequence diversity already present in mRNA display.

To test this novel semi-random cyclization library generation method, we performed a second selection against streptavidin as a point of comparison to our previous selection using the fixed cyclic residue design approach (**Figure 5.1b**).

a) Semi-random programming of cyclizable residue(s)

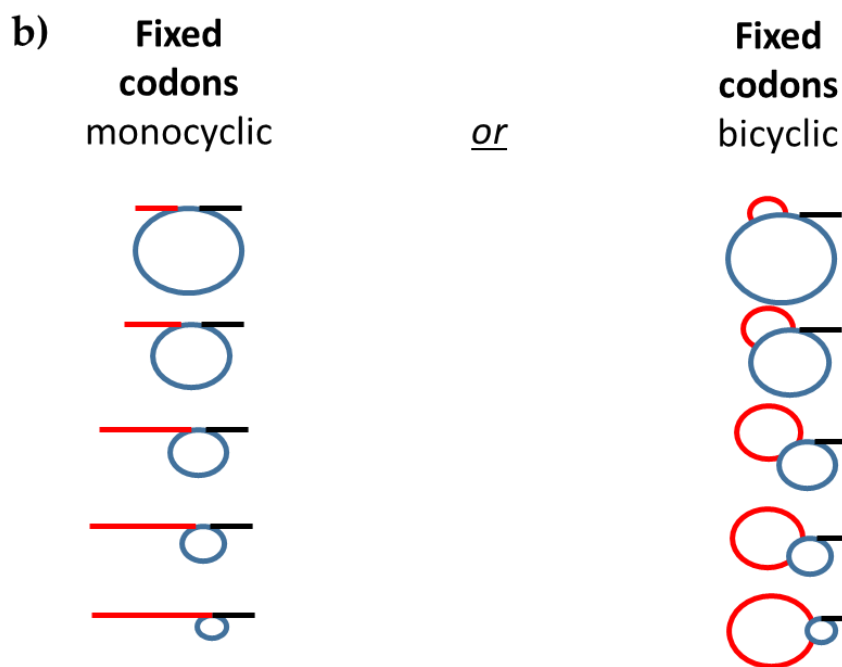
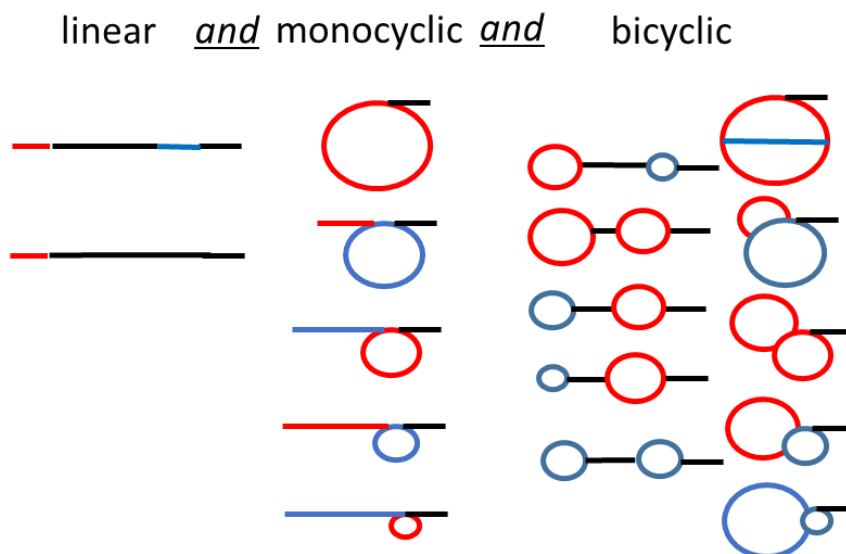


Figure 5.1 Library scaffolds for *in vitro* selections. a) Peptide conformations resulting from the semi-random scaffold generation design used in the streptavidin #2 selection. Target preference determines the optimal scaffold and hence the degree of cyclization. b) Fixed cyclizable codon design used in our previous streptavidin selection. This method is more scaffold-restrictive.

5.2 Results

5.2.1 Design of semi-random cyclizable libraries

In our previous selection against streptavidin, we used fixed cyclizable codons to limit the conformations generated by our library bicyclization method in order to exclusively create theta-bridged bicyclic peptides (**Figure 5.1b**). This enabled a direct comparison to a more traditional monocyclic scaffold. However, we considered the possibility that this limited scaffold architecture may have locked the peptides into conformations that were unproductive for binding.

To account for this limitation, we chose to design semi-random cyclizable libraries (**Figure 5.2a,b**), which had codons strategically placed that would permit the chemistry necessary for cyclization, but only in those peptides generated during fusion preparation that contained the complementary cyclization residue. One library, **M(NNY)₁₁C**, contained one each of the orthogonal residues necessary for bicyclization (**M** = β -azido-L-homoalanine or AzHA) (**Figure 5.2c**). Depending on target scaffold preference (linear, mono-, or bicyclic) and sequence selection, the second of one or both of the orthogonal residues could be enriched, effectively allowing the target to select not only the optimal sequence, but the preferred scaffold as well. The use of the NNY codon (Y = C or T) would omit 5 amino acids from the library (M, K, E, Q and W). By biasing the codons in this way, we would eliminate all stop codons, prevent the incorporation of a second AzHA (for M) and increase the probability of cyclizable amino acids (Cys or F-yne) appearing in the random region to 6.25% at each position. In our NNY library, the chance of one Cys or F-yne being encoded was 51%, and the probability of both appearing in a given sequence was 26% (cyclizable codon probability calculations for all libraries are in **Appendix Table H8**). Another library, **M(NNY)₁₁F**, would result in triazole-linked monocyclic peptides after the sequential chemistries were performed (**F** = *p*-ethynyl-L-phenylalanine or F-yne). In this library,

there was a 26% chance that two Cys would be encoded in any given sequence. If the target preferred the bicyclic conformation, random region cysteines, or a second AzHA/F-yne pair, would be enriched. Our third library, a more general **M**(NNS)₁₂, encoded only for AzHA, allowing for an even greater degree of sequence diversity and target scaffold possibilities. In this library, the ochre and opal nonsense codons are omitted and there is a 3.13% probability that Cys or F-yne would be encoded at each NNS position. Hence, there was a 32% chance that either Cys or F-yne would appear in the random region, and 9.3% that two Cys would be encoded. When the three mixed libraries are *in vitro* translated and the cyclization chemistries are subsequently performed, the result is a highly diverse set of scaffolds (**Figure 5.2d**) that will theoretically increase the likelihood of target-scaffold matching (**Figure 5.3**).

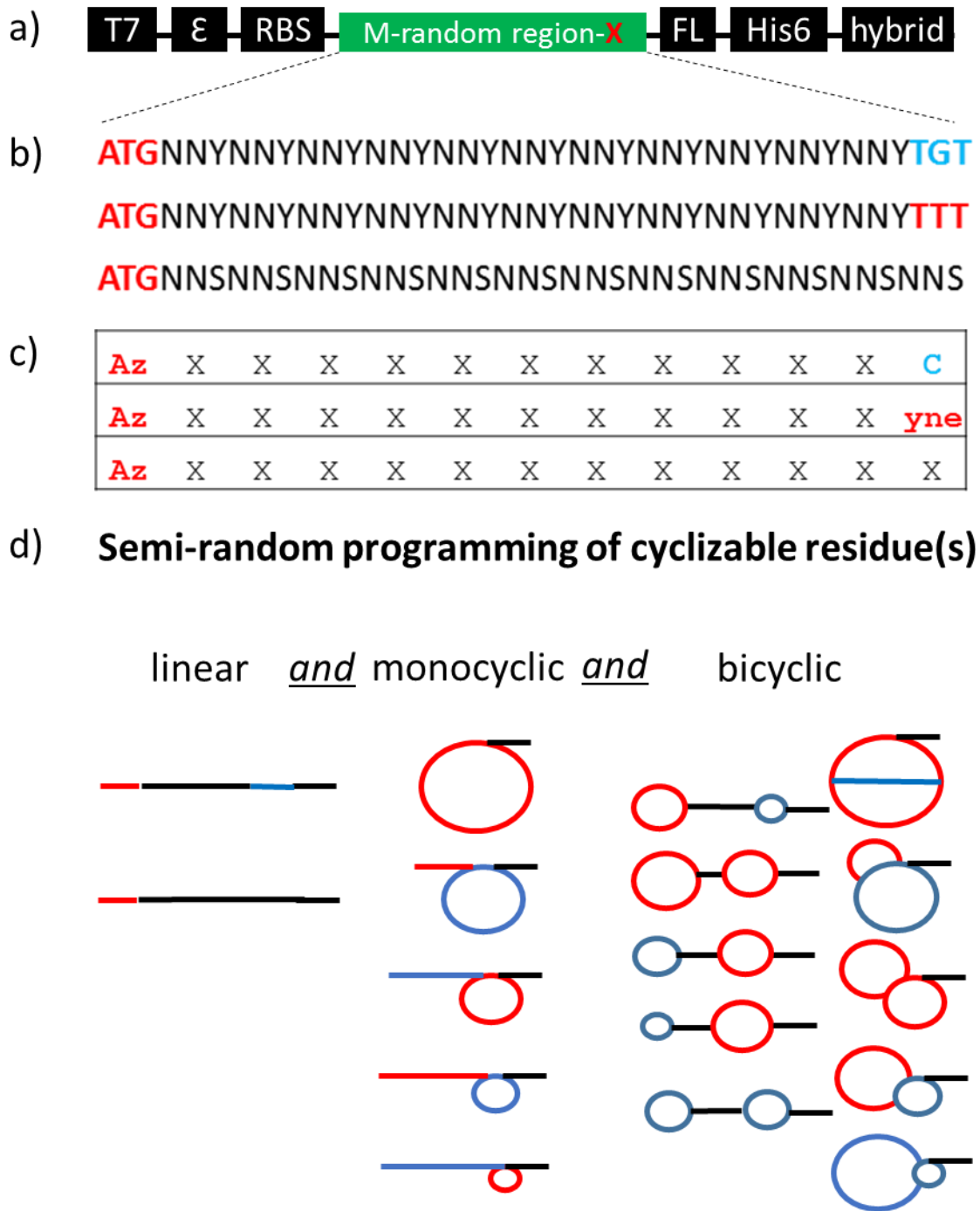


Figure 5.2 Semi-random library scaffold generation method. a) General DNA library elements, b) Random region of the libraries, highlighting fixed ‘clickable’ codons in red and bisalkylatable cysteine codons in blue. c) Peptide sequences generated from the library mRNA, **Az** = β -azidohomoalanine (AzHA), **yne** = *p*-ethynyl-phenylalanine (F-yne), **C** = cysteine. d) Highly random scaffold diversity generated by bisalkylation and CuAAC during mRNA-peptide fusion preparation.

5.2.2 *In vitro* selection against streptavidin #2

For the initial round of selection, we used an equal mix of our three UV-crosslinked mRNA libraries to initiate a translation that included AzHA and F-yne. During the fusion purification process, we performed the bisalkylation reaction with the fusions immobilized on oligo(dT) resin. The library fusions that contained two cysteines would form the macrocycle bridged by the bithioether-benzene linker, while those that included a single cysteine would likely form the peptide-BME adduct that we showed previously (3.2.1.2). Linear peptides would result from sequences that did not contain a cysteine (**Figure 5.3**). Following reverse transcription and binding to Ni-NTA resin, CuAAC was performed. Similarly, the formation of the triazole-linked cycle was predicated upon the encoding and expression of an F-yne residue (**Figure 5.3**). This technique allowed for the generation of a pool of linear, monocyclic and bicyclic scaffolds spanning a wide range of the potential topological landscape.

For round 1 target selection, the purified fusions resulting from the cyclization reactions were pre-cleared with magnetic beads containing an irrelevant protein (GST-XRCC4) and subsequently incubated with immobilized streptavidin. Functional fusions were competitively eluted with D-biotin. PCR amplification provided the DNA template for round 2. In round 7, we split the purified pool of fusions and continued with half of them in the standard streptavidin selection for rounds 7 and 8. For the other half of the fusions, we challenged them with chymotrypsin prior to selection in an effort to enrich protease-resistant sequences (**Figure 5.4**). Others have shown that chymotrypsin resistance also resulted in ~ 100-fold improved stability in human serum.¹⁷⁸ We then PCR-amplified the fusions that had been eluted with D-biotin and performed an additional round with the chymotrypsin digestion step.























Amino acids appearing in the random region					
Library	(0) Cys	(1) Cys	(2) Cys	(0) Cys	(0) Cys
aa(codon) _n aa	(0) F-yne	(0) F-yne	(0) F-yne	(1) F-yne	(2) F-yne
AzHA (NNY) ₁₁ C	—				
AzHA (NNY) ₁₁ F-yne					
AzHA (NNS) ₁₂	—	—			
Amino acids appearing in the random region					
Library	(1) Cys	(1) Cys	(2) Cys	(1) AzHA	(1) AzHA
aa(codon) _n aa	(1) F-yne	(2) F-yne	(1) F-yne	(1) F-yne	(2) F-yne
AzHA (NNY) ₁₁ C				N/A	N/A
AzHA (NNY) ₁₁ F-yne				N/A	N/A
AzHA (NNS) ₁₂					

Figure 5.3 Scaffold diversity. Possible 2D conformations from each of the three libraries after the two on-resin cyclization chemistries are performed on mRNA-peptide fusions expressing the noted amino acids. Blue represents cysteine or α,α' -dibromo-*m*-xylene-cyclized peptide. Red represents the azide, alkyne or the triazole-linked peptide. Note: The structures above are intended to represent general degree of cyclization (linear, mono-, or bicyclic) and in many cases, are one of multiple possible topologies dependent upon random incorporation. Note: Since the NNY libraries do not encode AUG, then AzHA can not appear in the random region.

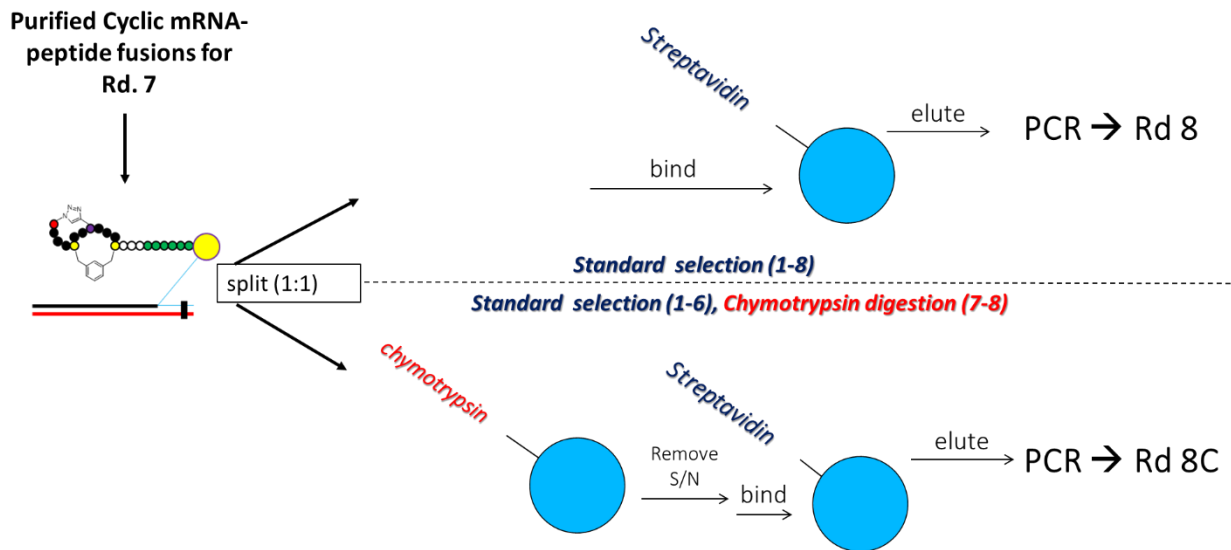


Figure 5.4 Scheme for chymotrypsin digestion in selection. Purified fusions were split prior to round 7 selection. Half continued with the same selection protocol, while the other half were incubated with immobilized chymotrypsin prior to selection against streptavidin (bottom route). After elution and PCR, the fusions were created and purified independently and a second round (rd 8) of the above scheme was performed. The purified DNA for both portions of the selection was sequenced after round 8.

We observed enrichment beginning in round 4 and continuing through round 7 and it plateaued in round 8 (**Figure 5.5**). The addition of chymotrypsin to the protocol resulted in ~ 60% recovery for round 7 compared to the standard streptavidin selection, and further decreased to ~ 8% for round 8. Following round 8, the DNA was sequenced using the Illumina MiSeq platform (**4.5.13**). The results were analyzed using AptaTOOLS software (**4.2.2**).¹⁶⁷⁻¹⁶⁹

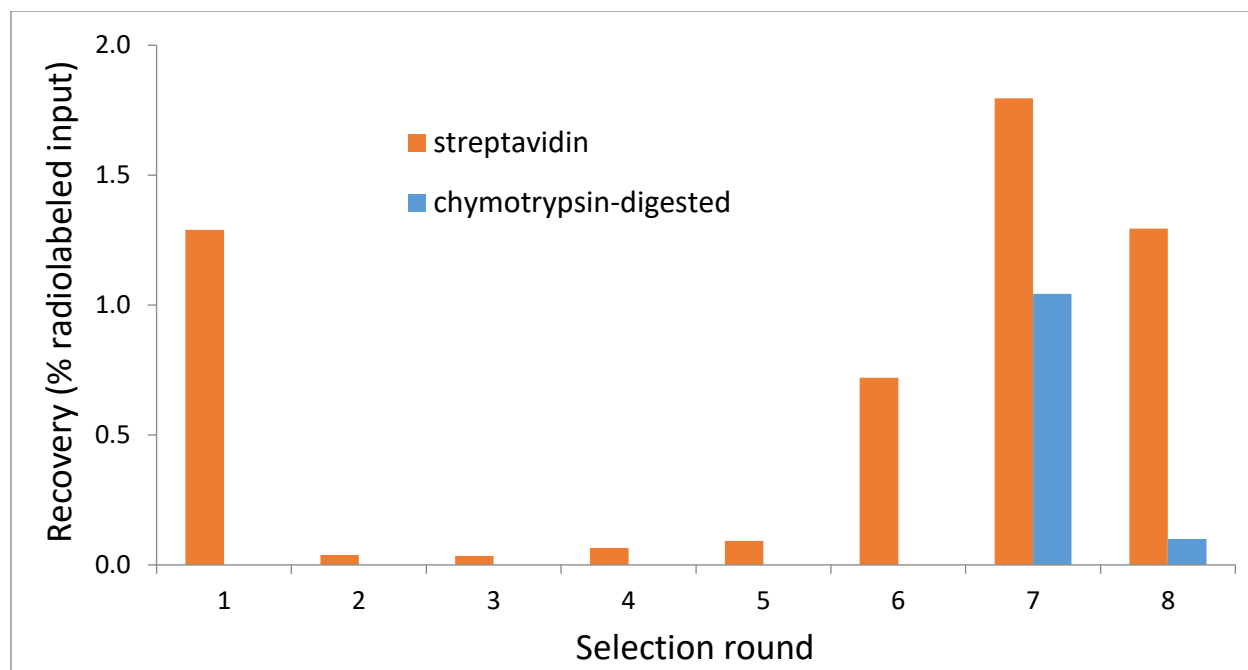


Figure 5.5 Selection enrichment. Monitoring the progress of selection by measuring the percentage of ³⁵S-Cys labeled mRNA-peptide fusions eluted from the beads. The purified fusion pool was split in round 7 and half were incubated with chymotrypsin prior to binding to streptavidin. The chymotrypsin digestion step was repeated for round 8.

5.2.3 Sequencing results: Streptavidin #2

The results from sequencing indicated an overwhelming preference for the commonly found HPQ motif (**Figure 5.6a**). The 25 most abundant peptides (and 98 of the top 100-**Appendix Tables H3a,b**) all contained this sequence. The HPQ sequence appeared almost exclusively in the 9-11 positions of the random region, with a few occurring in positions 11-13. Interestingly, all of the top 100 most abundant sequences originated from the **M(NNS)₁₂** library, and the presumed binding motif (HPQ) is not constrained within a macrocycle in any of these peptides.

5.2.4 Sequencing results: Streptavidin #2-chymotrypsin

The chymotrypsin challenge introduced in rounds 7 and 8 significantly reduced selection recovery relative to the standard selection; however, certain sequences were highly enriched (**Figure 5.6b**). Not surprisingly, the HPQ motif dominated this portion of the selection as well, with 24 of the top 25 most-enriched peptides containing this sequence. Notably, there was an increased occurrence of F-yne, which correlates with the enrichment of cyclic peptides (13 of the top 25 enriched sequences). As with the standard selection, the top 100 most enriched chymotrypsin-digested peptides (**Appendix Table H4a,b**) all originated from the **M(NNS)₁₂** library.

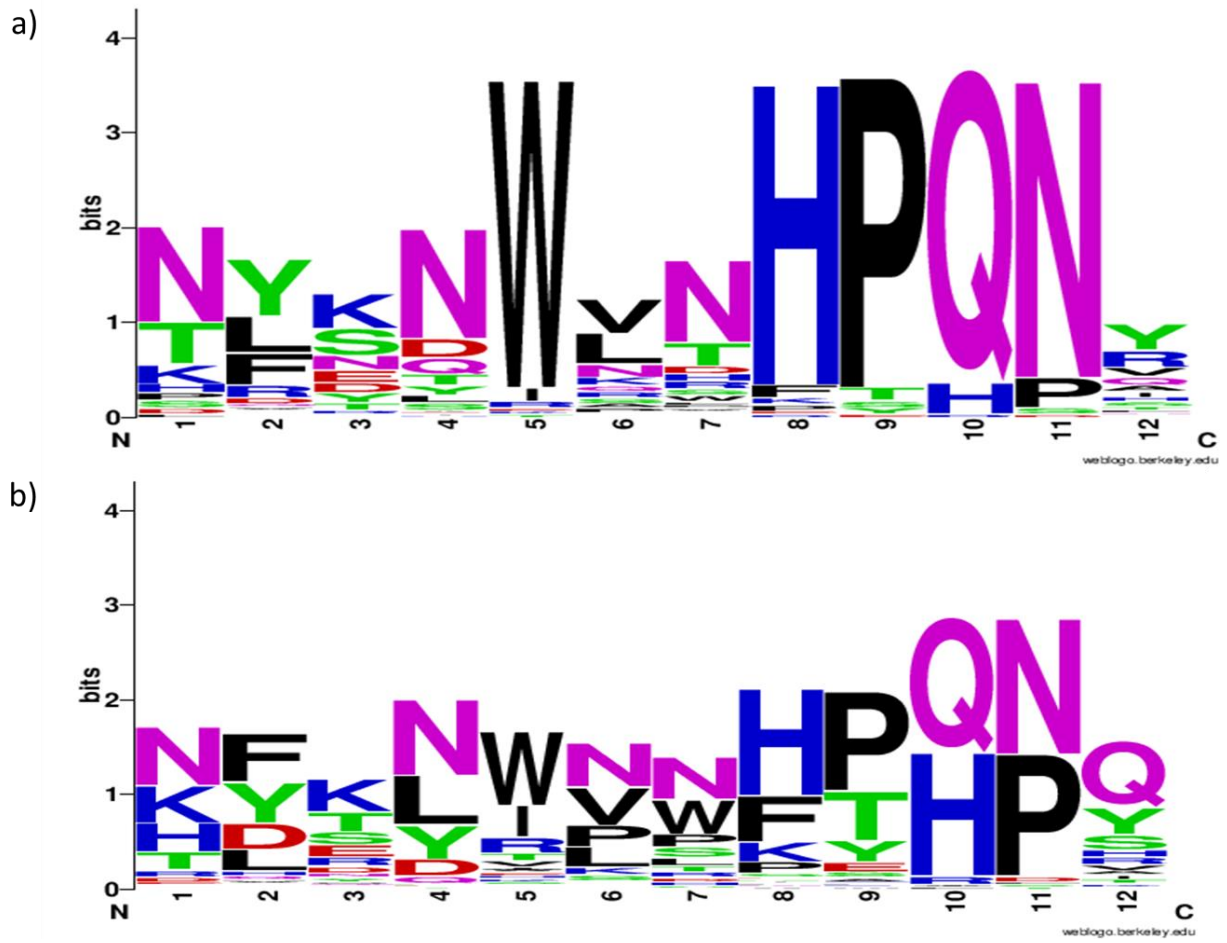


Figure 5.7 Sequence homology. Consensus sequences of a) Top 100 most abundant peptides for the standard streptavidin #2 selection and b) Top 100 most enriched sequences from the chymotrypsin pre-treated streptavidin selection (Rd. 8-Chymotrypsin fold enrich relative to Rd. 6).

5.3 Discussion

We hypothesized that the use of a semi-random cyclizable codon library design method would generate a wide variety of peptide conformations, including linear, mono- and bicyclic scaffolds. The target would then dictate the preferred conformation, rather than imposing constraint *de novo* as was the case in our first selection against streptavidin. The exclusive isolation of **M**(NNS)₁₂ library-templated peptides indicated a strong selective pressure for linear peptides by streptavidin. Moreover, the HPQ motif is found almost exclusively toward the C-terminal portion of the random region. This implies an enhancement to binding from contacts made by the linear N-terminal portion of the peptides, a finding supported by a prior comprehensive deletion analysis.¹⁷⁴ Sequence homology of the streptavidin #2 results indicate an extended WxxHPQN motif in amino acids 6-12 of the random region (**Figure 5.7a**). Previous selections have identified high affinity linear peptides with an HPQN motif,^{165,172,175} and the optimized Strep-tag II epitope (WSHPQFEK) contains an upstream Trp, albeit in the +1 position relative to our consensus sequence.¹⁷⁹ We consider the possibility that the fixed cyclizable codons in our first selection against streptavidin (**Chapter 4**) may have limited the creation of fusions with a consecutive, linear WxxHPQN sequence, and as a result this sequence was not available for target selection.

Previous selections against streptavidin have revealed both a preference for linear peptides with extended contacts outside of the biotin binding pocket,^{172,174,180,181} as well as shorter, more constrained sequences,^{165,166,182} including the mid-nM cyclic binders we identified previously in Chapter 4. Gao and Kodadek performed a side-by-side screen against streptavidin with linear and monocyclic peptoid libraries displayed on beads.¹⁸³ They found that streptavidin prefers cyclic binders, but only those containing 17 atoms, not 20 or 23. To our knowledge, this work is the first example of a direct comparison of multiple peptide library scaffold options provided to a target

for selection. Furthermore, this one-pot preparation method eliminates the need for independent peptide cyclization and selections. Given previous results, we expected to obtain both linear and constrained peptides from our selection. The fact that exclusively linear sequences resulted from this method must be qualified by the fact that the two NNY libraries omitted glutamine from the random region and as a result, the HPQ motif was not present in the purified fusions created from these templates. Moreover, based on the prevalence of the HPQ motif (encoded in the linear NNS library), the competing factors of peptide sequence and scaffold (constraint) preference appears to favor target sequence preference in this case.

To increase the stringency of our selection and obtain more protease-stable peptides, we introduced a chymotrypsin challenge to half of the purified fusion pool for the last two rounds of selection (**Figure 5.4**). We looked at sequences that were highly enriched in the chymotrypsin relative to the non-chymotrypsin selection and found that the top 3, and 8 of the top 10, most-enriched sequences contained F-yne and were therefore cyclic (**Figure 5.8**). All 8 of these peptides contained the NWFTHPQ sequence at the C-terminus. This is an extension of the WxxHPQ portion of the consensus sequence identified in the standard selection. We suspect that the presence of the F-yne cyclization within this sequence acts to protect the Trp from digestion by chymotrypsin. Furthermore, the unanimous homology of these sequences implies that the cyclization also locks the peptide into a conformation productive for binding. When we compared round 6 to round 8 enrichment of the top 25 chymotrypsin selection sequences relative to the same sequences (from the same rounds) in the non-chymotrypsin selection (ranked based on relative fold enrichment-**Figure 5.6c**), we found the same preference for HPQ (22/25) as well as an increased presence of F-yne (11/25) relative to F-yne occurrence in the top 25 sequences from the standard selection (6/25).









Rank	Peptide
C-1	 MKFKLSNWF THPQ
C-2	 MEFKLINWF THPQ
C-3	 MKFRLINWF THPQ
C-4	MNYENWVN HPQ NA
C-5	 MKFKLNNWF THPQ
C-6	MNYKNWVN HPQ NS
C-7	 MEFKLINWF THPQ
C-8	 MKFKLTNWF THPQ
C-9	 MKFKLVNWF THPQ
C-10	 MKFKLINWF THPQ

Figure 5.8 Protease-stable sequences. The Top 10 most highly enriched sequences from round 6 to round 8 of the chymotrypsin digestion portion of the selection. HPQ motif is in bold. Solid arc represents a triazole cycle between AzHA (M) and F-yne (F). Dotted arcs indicate two possible click cyclizations with peptides containing two F-yne residues.

Therefore, the addition of the chymotrypsin protocol to our selection resulted in the enrichment of cyclic peptides. We next looked at the top 100 most enriched sequences from Round 7 to 8 in the chymotrypsin selection (**Appendix Table H4a,b**). As expected, we observed a decrease in the presence of Trp and Tyr in the peptides (82% relative to the 100 most abundant non-chymotrypsin selection peptides), both of which are substrates for chymotrypsin. However, Trp appeared in the top 10 chymotrypsin selection sequences, and 8 of the 10 were followed directly by a triazole-linked F-yne. The consensus sequence from the standard selection indicates that this Trp is important for binding (**Figure 5.7a**). Therefore, the protease resistance of this chymotrypsin substrate implies a stabilizing effect of cyclization toward digestion and increases the likelihood that protease-susceptible high affinity cyclic ligands will survive selection.

In summary, our selection against streptavidin using our semi-random cyclizable codon library design resulted in the nearly unanimous appearance of the HPQ motif. The results were dramatically different than the results from our previous selection, which used fixed cyclizable codons to generate mono- and bicyclic peptides. We have not yet tested these peptides for affinity for streptavidin, so at this point the different sequencing results alone may give insight into which library design is preferable. However, the lack of glutamine in two of the three libraries may have biased the results toward linear peptides and possibly even to the prevalence of the HPQ motif. Furthermore, cysteine was rarely found in the sequences for either selection and there were no peptides that contained two cysteines. As a result, none of the top 100 sequences contained a xylene-linked macrocycle. The chymotrypsin-treated portion of the last two rounds resulted in the relative decrease in Trp and Tyr (**Figure 5.7**), along with an increase in F-yne and hence, triazole-cycle formation. While we did not test these peptides for protease stability as in our previous

selection, we postulate that the increased appearance of F-yne would act to make the peptides more protease resistant.

5.4 Future Directions

The selection winners could be tested for affinity for streptavidin and protease stability. Of particular interest is the relative affinity and stability of the highly enriched cyclic peptide that survived incubation with chymotrypsin, **MRFKLINWFTHPQ**. The lack of glutamine in the cyclization-biased NNY libraries likely skewed the results of the selection. A repeat of the selection with equivalent random regions would give more generalizable results.

5.5.1 Amino acids and translation reagents: For natural amino acid preparation, see **2.5.1**. Preparation of *in vitro* translation components and reagents is described in **2.5.3**. Unnatural amino acids are described in **3.5.1**.

5.5.2 Selection Reagents: Magnetic streptavidin beads: Dynabeads M-280 streptavidin (Invitrogen #11205D); Oligo-(dT)₂₅ magnetic beads (NEB #S1419S); Superscript III reverse transcriptase (Invitrogen #18080093); RNase inhibitor (Bioline #65027); TurboDNase (Invitrogen #AM2238); Taq DNA polymerase (NEB #M0237L); All oligonucleotides were purchased from Integrated DNA Technologies (IDT). α,α' -dibromo-*m*-xylene was purchased from Fluka.

5.5.3 Preparation of mRNA-peptide fusion library for streptavidin #2 selection

An equimolar mix of the three bottom strand oligo libraries (see **Table 5.1**) was *in vitro* transcribed and UV-crosslinked as described in **4.5.3**. For round 1 of selection, a 2 mL standard translation reaction containing AzHA and F-yne, in lieu of Met and Phe, was initiated with the addition of the puromycin-linked mRNA. The reaction was incubated at 37° C for 1.25 h, supplemented with 550 mM KCl and 55 mM MgCl₂, returned to the incubator for 1.5 h and subsequently frozen overnight at -80° C. The resulting mRNA-peptide fusions were diluted 6-fold with oligo-dT binding buffer (20 mM Tris-HCl pH 7.8, 10 mM EDTA, 1 M NaCl, 0.2% Triton X-100, 0.5 mM TCEP), added to 1.5 mL of oligo-dT magnetic beads which were pre-equilibrated thrice with 5 mL of oligo-dT binding buffer, and rotated at 4° C for 30 min. The beads were washed twice with 5 mL of oligo-dT wash buffer (20 mM Tris-HCl pH 7.8, 0.3 M NaCl, 0.1% Triton X-100, 0.5 mM TCEP), and the first cyclization was performed on resin by the addition of 5 mL of cyclization buffer (20 mM Tris-HCl pH 7.8, 0.66 M NaCl, 3 mM α,α' -dibromo-*m*-xylene, 33% acetonitrile (v/v), 0.5 mM TCEP) and rotated at RT for 30 min. The beads were washed once with 5 mL of oligo-dT wash buffer containing 5 mM BME (in lieu of TCEP), to quench the unreacted linker.

Beads were washed a second time with 5 mL of wash buffer containing TCEP and eluted in 1 mL fractions with 0.5 mM TCEP. The five elutions with the highest scintillation counts were pooled and precipitated with 4 vol of ethanol, 0.1 vol of 3 M KOAc, pH 5.2 and 0.001 vol of 5 mg/mL glycogen. The pellet was resuspended in 500 μ L of water and reverse transcribed in a final volume of 1 mL in the presence of RT mix (0.5 μ M RT primer, 0.5 mM dNTPs, 5 mM MgCl₂, 1 mM DTT, 2 U/ μ L RNase inhibitor, 5 U/ μ L Superscript III, 1x First Strand buffer) at 55° C for 30 min. The RT reaction was subsequently diluted 5-fold with denaturing Ni-NTA binding buffer (100 mM NaH₂PO₄, 10 mM Tris-HCl, 6 M guanidinium hydrochloride, 0.2% Triton X-100, 5 mM BME, pH 8) and added to 500 μ L of Ni-NTA agarose resin pre-equilibrated per the manufacturer's instructions (MCLab) in a 10 mL BioRad disposable column. The column was placed on a tumbler at 4° C for 1 h. After binding to the Ni-NTA agarose, the fusions were washed once with 3 mL of Ni-NTA wash buffer (100 mM NaH₂PO₄, 300 mM NaCl, 0.2% Triton X-100, 5 mM BME, pH 8) and 3 mL total volume of CuAAC cyclization reagents were added (100 mM phosphate pH 8, 1 mM CuSO₄, 300 mM NaCl, 2 mM TBTA, 33% DMSO (v/v)). The solution was degassed with argon prior to the addition of 10 mM sodium ascorbate. An argon blanket was placed over the solution and the tube was sealed and rotated at room temperature for 2 h. After incubation, the resin was washed once with 3 mL of Ni-NTA wash buffer containing 10% DMSO (v/v). Following two additional washes with 3 mL of Ni-NTA wash buffer, the purified fusions were eluted in fractions with Ni-NTA elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 350 mM imidazole, 0.2% Triton X-100, 5 mM BME, pH 8), pooled and precipitated with 4 volumes of ethanol and 0.1 volume of KOAc, pH 5.2. The pellet was stored at -80 °C until needed.

The translation volume was reduced to 500 μ L for rounds 2-4 and 250 μ L for rounds 5-8. Purification reagent usage, wash and elution volumes were adjusted accordingly to coincide with

translation volume. For round 7, the purified fusion pool was split in half prior to the selection step: Half were continued in the standard selection and half were targeted for chymotrypsin digestion prior to selection against streptavidin. The two fusion pools were purified independently for round 8.

5.5.4 Selection against streptavidin #2:

For round one target selection, pelleted fusions (**from 5.5.3**) were resuspended in 0.8 mL of selection buffer containing 0.1% BSA and added to a tube containing immobilized GST-XRCC4 (negative selection against an irrelevant protein) and rotated at 4 °C for 1 h. The supernatant containing the pre-cleared fusions was removed and added to a tube containing 100 µL of M-280 Streptavidin beads, which had been equilibrated thrice with 1 mL of selection buffer containing 0.1% BSA. The GST-XRCC4 beads were washed once with 200 µL of selection buffer, which was then added to the tube containing the streptavidin beads, for a total volume of 1 mL. The tube was rotated at 4° C for 1 h. After binding, the beads were washed once with 1 mL of selection buffer. 150 µL of selection buffer containing 2.5 mM D-biotin was added to the streptavidin beads and the tube was rotated at 4 °C for 2 h. The supernatant was removed and dialyzed overnight against 0.1% Triton-X 100. Following PCR amplification using SAX4FWD and SAX4REV (2 min. at 94°C, followed by 22 rounds of 94° C (30s), 65° C (30s), 72° C (45s), library DNA was extracted with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with 3 volumes of ethanol and 0.1 vol of 3 M KOAc pH 5.2. The DNA was resuspended in water and served as the template for Round 2 of selection.

5.5.5 Chymotrypsin digestion for streptavidin #2 rounds 7 and 8: The half of the fusion pool (0.5 mL) that was separated and targeted for chymotrypsin digestion before the round 7 selection step were incubated for 15 min. at room temperature with pre-equilibrated Dynabeads M-270

epoxy beads (Invitrogen, #14301) which had been decorated with 1 U of α -chymotrypsin (Sigma #C1429) per the protocol described in **4.5.10**. The supernatant was removed and added directly to 100 μ L of streptavidin beads. Beads were washed thrice with 0.5 mL of selection buffer and the fusions were eluted with 150 μ L of selection buffer containing 2.5 mM D-biotin. Functional cDNA was PCR amplified and prepared for round 8 as in **5.5.7**. This process was repeated for round 8 fusions, which had been purified independently.

5.5.6 Sequencing

Following round 8, DNA from both selections (streptavidin #2 and chymotrypsin-digested streptavidin #2) were sequenced independently using the Illumina MiSeq platform (**4.5.13**). The raw sequences were analyzed using AptaTOOLS software (**4.2.2**).¹⁶⁷⁻¹⁶⁹ The top 100 sequences from the streptavidin # 2 selection can be found in **Appendix Tables H3a,b** (non-chymotrypsin) and **H4a,b** (chymotrypsin).

Chapter 6: *In vitro* Selection Against XRCC4 Using Scaffold-Diverse Peptide Libraries

6.1 Introduction

In our previous selection against streptavidin using the semi-random cyclizable codon library design (**Chapter 5**), we observed nearly unanimous enrichment of peptides that contained an HPQ motif, a sequence commonly found in selections by others against this target.^{165,172-175} Our most abundant selection winners were almost exclusively linear. However, the addition of a chymotrypsin digestion protocol prior to the selection step resulted in preferential enrichment of cyclic peptide binders. Having validated our library design methodology, we chose to use these libraries in a selection against a biologically relevant protein.

We chose DNA repair protein XRCC4 (X-ray repair cross-complementing 4) as a target for *in vitro* selection (**Figure 5.2**). Radiotherapy is a form of cancer treatment that uses ionizing radiation to cause DNA damage (primarily in the form of double strand breaks) ultimately leading to cell death and tumor regression. XRCC4 and XLF (XRCC4-Like Factor) are involved in non-homologous end-joining (NHEJ),¹⁸⁴ the primary DNA repair pathway in mammalian cells.¹⁸⁵ NHEJ involves the recruitment of several proteins^{186,187} (**Figure 6.1**) and within this assembly, XLF interacts with XRCC4 and forms a series of filaments that likely serve as a scaffold to stabilize the NHEJ repair proteins.^{188,189} Cells deficient in XLF have been shown to be highly radiosensitive.^{190,191} Thus, by discovering a ligand that could bind the XLF site of XRCC4 and block the interaction, we could mimic XLF-deficient conditions. This would dramatically reduce the efficiency of NHEJ as a DNA repair pathway and ultimately make tumors more sensitive to ionizing radiation therapy.

XRCC4 contains a leucine lock at the XLF interface and has a shallow pocket with a hydrophobic surface¹⁸⁸ that appears amenable for probing with our bicyclic peptide library that contains hydrophobic bridging elements. However, in the event that the highly-constrained conformation is unsuitable, or perhaps too compact to bind to the interface of XRCC4, a wide

variety of alternative scaffolds will be generated to maximize the likelihood of ligand-target scaffold matching.

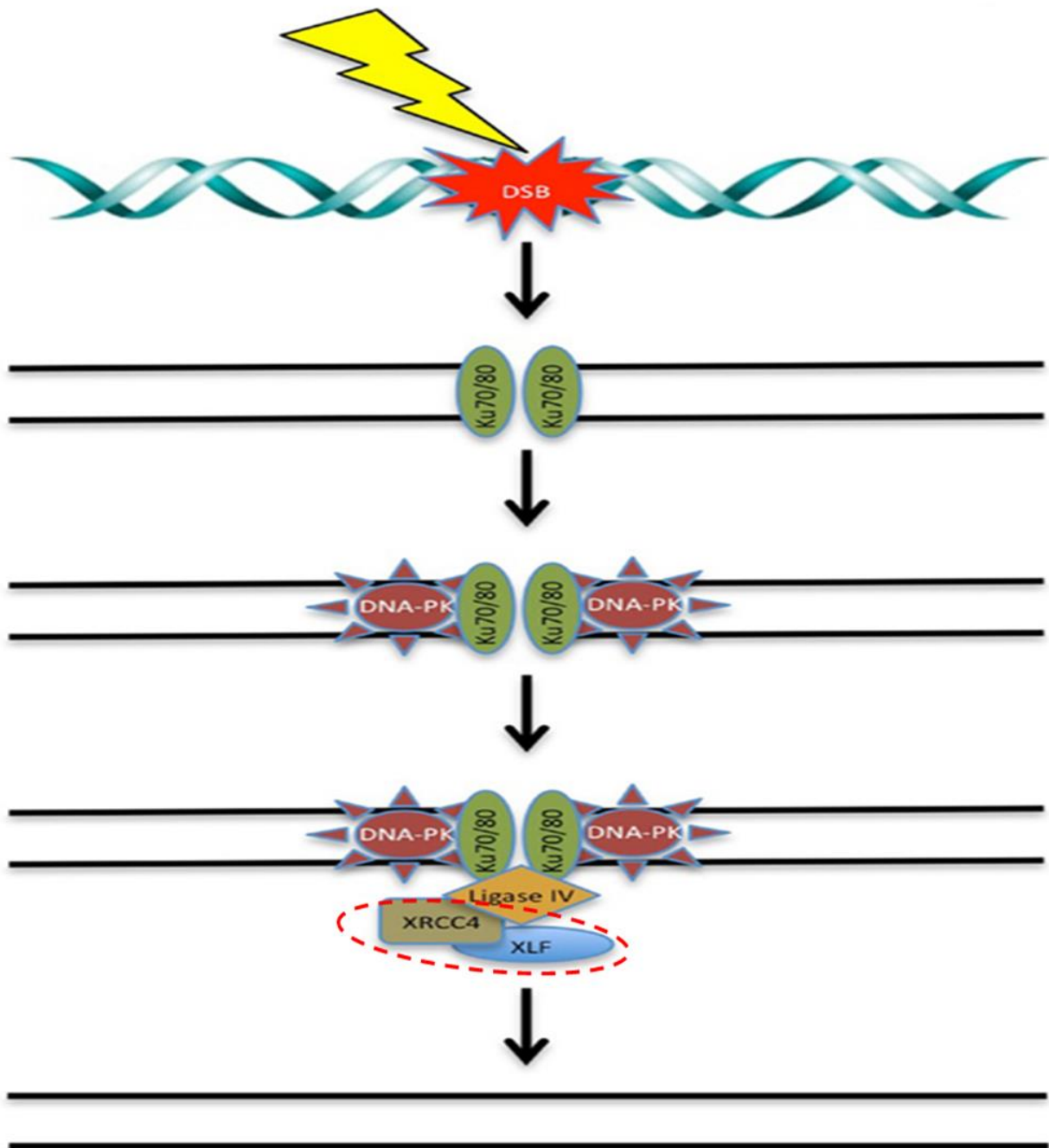


Figure 6.1 Non-homologous end-joining (NHEJ). The primary DNA repair pathway in mammalian cells depicts the recruitment of proteins in response to DNA double-strand breaks upon exposure to irradiation. XRCC4 and XLF bind head-to-head, forming a series of alternating filaments that stabilize the protein complex and lead to gap filling and ligation of the DNA.

6.2 Results

6.2.1 Design of semi-random cyclizable libraries

For our selection against XRCC4 we chose to use the identical semi-random cyclizable codon libraries previously validated against streptavidin (**M**(NNY)₁₁**C**, **M**(NNY)₁₁**F** and **M**(NNS)₁₂). Our rationale for this design is discussed thoroughly in **5.2.1** and the resulting scaffold diversity is depicted in **Figures 5.2 and 5.3**. Our lab performed a previous selection against XRCC4 (unpublished work) that did not produce a high affinity ligand. The set of 5 libraries used in that selection consisted of an N-terminal MC prior to the 13-amino acid random region, of which one was a second fixed cysteine. The second cysteine was placed incrementally from the first cysteine (i.e. $i, i + 3$; $i, i + 5$; $i, i + 7$; $i, i + 9$ and $i, i + 11$) (see **Appendix Table H6**). The libraries were subsequently cyclized using α, α' -dibromo-*m*-xylene to create monocyclic fusions with increasing ring sizes, analogous to our previous monocyclic design discussed in Chapter 4. We reasoned that the preparation of our more highly diverse set of scaffolds would increase the likelihood of the peptides adopting a productive conformation for target interaction and result in a high affinity binder to XRCC4.

6.2.2 *In vitro* selection against XRCC4

To produce our mRNA-peptide fusion library, we used an equimolar mix of our three UV-crosslinked library mRNAs to initiate a translation that included AzHA and F-ycne (and omitted methionine and phenylalanine). We performed the two orthogonal cyclization reactions with peptides immobilized on resin to produce our pool of purified mRNA-peptide fusions with highly diverse scaffolds (discussed thoroughly in **5.2.2**).

For round 1 of selection, we incubated the fusion library with 20 μ M GST-XRCC4₁₋₁₅₇ and washed the beads once after binding. We increased stringency in two ways throughout the selection: The concentration of GST-XRCC4₁₋₁₅₇ was decreased to 10 μ M in round 2, and finally

to 1 μ M for round 3 until the end of selection. Since relatively few copies of each sequence are present in the naïve fusion pool, the GST-XRCC4₁₋₁₅₇ beads were only washed once (with 2 CV of selection buffer) in round 1. Washes were increased to three (5 CV each) in round 2 and four (5 CV each) for round 3 until the end of selection.

The selection was monitored by the incorporation of ³⁵S-cysteine in mRNA-peptide fusions. After 10 rounds of selection, we did not observe enrichment and radiolabeled yield was very low (**Figure 6.2**). The purified PCR-amplified DNA from rounds 9 and 10 were sequenced using the Illumina MiSeq platform.

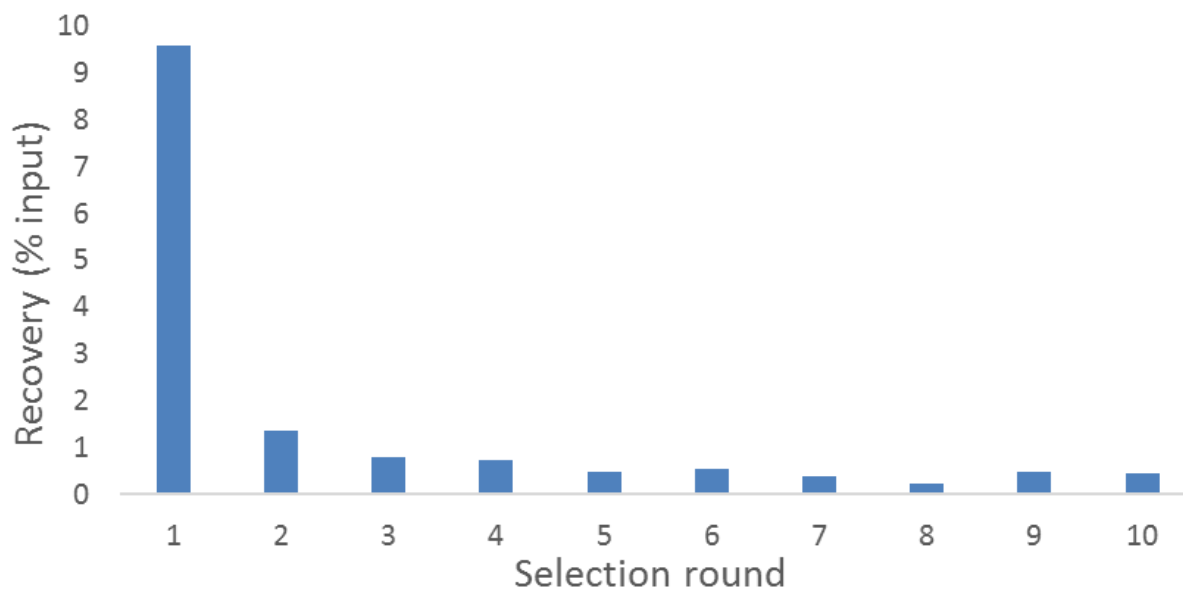


Figure 6.2 Selection enrichment. Monitoring the progress of selection by measuring the percentage of ³⁵S-Cys labeled mRNA-peptide fusions eluted from the beads.

6.2.3 Sequence analysis-XRCC4

The lack of observable radiolabel-based enrichment was reflected in the sequencing analysis using AptaTOOLS software (4.2.2). The most abundant sequence from Rd. 10 (X1 in **Figure 6.3a**) was present at only 0.033% of the final pool. Comparatively, the winner of our selection against streptavidin, using the same libraries, comprised over 17% of the final pool. The lack of frequency (individual/total counts) in the top sequences indicated a lack of convergence and hence, the homology diagram did not reveal any consensus regions within the top 100 peptides (**Figure 6.3b**).

a)

Rank	Peptide	Cluster ID	Count	Frequency	Enrich
X1	NYKNWVNHPQNY	335020	167	3.28E-04	1.21
X2	KKNKWRKDTYKN	259439	99	1.95E-04	1.91
X3	NLSNWLNHPQNR	281403	94	1.85E-04	1.75
X4	MKNHKNKHKKHD	227030	88	1.73E-04	3.15
X5	NKKHKNKKTCKM	227815	73	1.43E-04	5.04
X6	TKFKKHKNIKKT	155350	60	1.18E-04	3.22
X7	LAALHDSILLLK	319348	56	1.10E-04	1.26
X8	NNHINRKDKYNR	264058	52	1.02E-04	3.35
X9	NYENWVNHPQNY	335020	50	9.83E-05	1.21
X10	KFNDWKHHPQNI	207462	46	9.04E-05	1.14
X11	FNHKNNTKQHYI	146442	46	9.04E-05	1.65
X12	HKKHAKFRLRKT	319946	45	8.84E-05	2.07
X13	HKRKPRKYFDRK	277250	43	8.45E-05	2.77
X14	TFKIYSIKPLHT	322381	43	8.45E-05	1.26
X15	NYNHTHGRYNRF	300760	41	8.06E-05	1.65
X16	NNHFHNRNFRHF	258767	41	8.06E-05	2.33
X17	MLAALHDSISFL	202533	41	8.06E-05	0.90
X18	KKKHSKHKTFRI	198613	40	7.86E-05	2.76
X19	HNNNYNHSGHRF	169045	40	7.86E-05	2.42
X20	HNHRFNRRNRNF	231827	39	7.66E-05	3.14
X21	PFNTWQDHPQNA	326287	38	7.47E-05	2.82
X22	TRYQWNTHPQNV	146842	36	7.08E-05	2.32
X23	KKTKHKMIKRSL	219347	35	6.88E-05	3.07
X24	HNNKNFKVRHKT	188225	34	6.68E-05	1.56
X25	HNFNRNFRGNHF	345413	33	6.49E-05	1.68

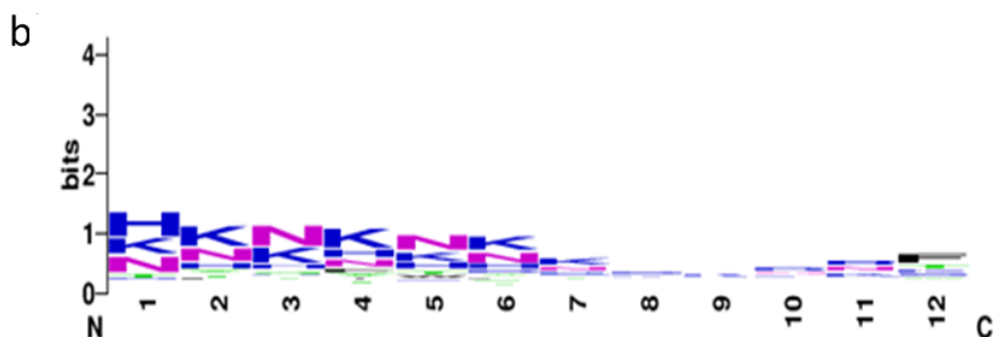


Figure 6.3 Sequence analysis. a) Top 25 most abundant sequences following Rd. 10 of selection. AptaTOOLS groups homologous sequences in clusters. Count represents abundance, frequency is a function of total reads and enrich is fold relative to the previous round. Sequences highlighted in gray contain the HPQ motif and are identical to sequences that were found in our Streptavidin #2 selection (Chapter 5). Top 100 sequences are in Appendix Tables H5a,b. b) Homology diagram of the Top 100 most abundant sequences from Rd. 10.

6.3 Discussion

Despite our demonstrated success using the semi-random cyclizable codon libraries in our selection against streptavidin, the selection against XRCC4 did not result in enriched sequences. The reason for this lack of success is unknown, but we speculate as to potential causes: It is typical to have very low stringency in round 1 of *in vitro* selections to ensure functional sequences with very low copy numbers are retained during target selection, to be subsequently amplified by PCR. Our strategy to accomplish this was to omit a negative selection (pre-clear against GST-beads) and only perform a single (2 CV) wash. However, this resulted in > 9% recovery of our purified fusions after selection (**Figure 6.2**). This unusually high percentage likely contained a very large number of non-functional sequences that were subsequently amplified prior to Rd. 2. This would generally increase the number of selection rounds necessary to observe enrichment and many of the potential hits could have been ‘lost’ amid the non-functional noise.

The increases in stringency we implemented during selection, both in the form of decreasing the concentration of target and increasing the number of washes, could have resulted in the loss of functional sequences, or at least a drastic reduction in count. For example, if a peptide has a low affinity for its target, the fractional binding at 1 μ M target concentration would be very low. As a result, the quantity of functional peptides that are eluted (and subsequently amplified) relative to the input quantity would result in decreased rate of enrichment, and likely a very low overall actual percent recovery.

AptaTOOLS analysis revealed the presence of peptides from our previous selection against streptavidin in 6 of the top 25 sequences from the XRCC4 selection. Since the libraries for the two selections were identical, the primers used for PCR were the same as well. So even a small amount of cDNA could contaminate a selection given the sensitive nature of PCR. If the XRCC4 selection

converged and abundance of the winning sequences was high, this phenomenon would be unobserved, even if present. However, given the lack of enrichment and functional sequence frequency during this selection, the streptavidin sequences appeared in our Top 25 results from Rd. 10. While this contamination is not likely a cause for lack of convergence of XRCC4 binders (given the low abundance of these HPQ-containing sequences in our selection), it does strongly suggest the need for using unique primers for every selection performed using the same instruments.

6.4 Future Directions

We have been unsuccessful in our two attempts to find a high affinity binder to XRCC4 using two very different sets of libraries. We allow for the possibility that XRCC4 is not a viable target for peptide libraries.

6.5 Experimental

6.5.1 Amino acids and translation reagents: For natural amino acid preparation, see **2.5.1**. Preparation of *in vitro* translation components and reagents is described in **2.5.3**. Unnatural amino acids are described in **3.5.1**.

6.5.2 Reagents and oligonucleotides for selection: For general selection reagents see **5.5.2**. Glutathione agarose (Thermo Fisher #16100); Glutathione magnetic beads (Thermo Fisher #88822). All oligonucleotides were purchased from Integrated DNA Technologies (IDT). For oligonucleotides used for selection, see **Table 5.1**.

6.5.3 Expression and purification of GST-XRCC4₁₋₁₅₇

The head domain of XRCC4 (aa 1-157) was cloned into a pGEX-4T-1 vector digested with BamH1 and *Xho*I. The plasmid was transformed into BL-21-DE3 strain of *E. coli*, and an overnight culture was used to inoculate 200 mL of LB media. Cells were grown to log phase, induced with 0.1 mM IPTG and allowed to express overnight at 18 °C. The recombinant fusion protein, containing a C-terminal hexahistidine epitope, was purified using the manufacturer's protocol (MCLAB Ni-NTA agarose, NINTA-400) with the following exceptions. Binding and wash buffers contained 20 mM imidazole and GST-XRCC4 was eluted with 350 mM imidazole. The fusion protein was dialyzed against enzyme storage buffer (50 mM HEPES-KOH pH 7.6, 100 mM KCl, 10 mM MgCl₂, 7 mM BME, 30% glycerol) and stored at -80 °C. Protein concentration was determined by UV absorbance at 280 nM using a molar extinction coefficient (ϵ) of 62,870.

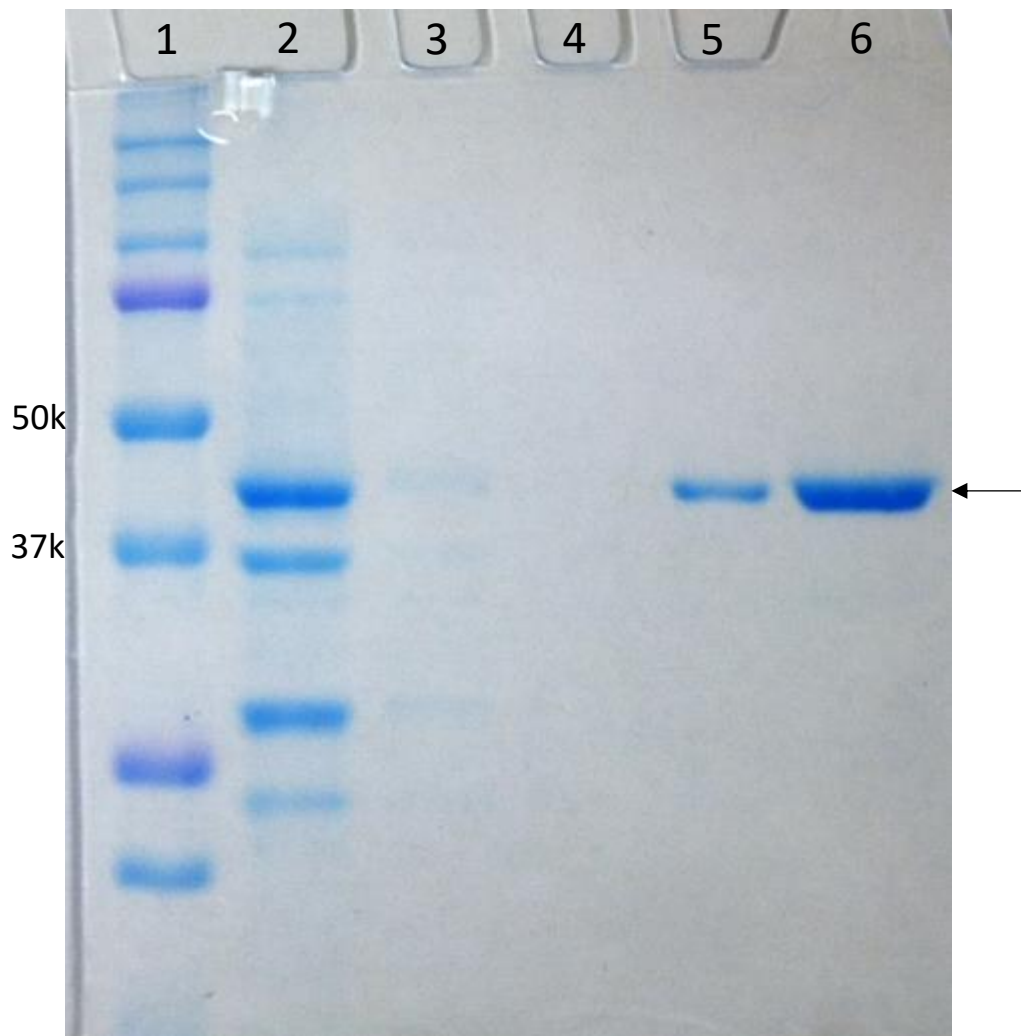


Figure 6.4 Purification of GST-XRCC4. 10% SDS-PAGE analysis of Ni-NTA purification of the bacterial expression of recombinant GST-XRCC4₁₋₁₅₇. Lane 2, Flow-through; Lanes 3 and 4, washes; Lane 5, 350 mM imidazole; Lane 6, 350 mM imidazole (2x lane 5 loading). GST-XRCC4₁₋₁₅₇ MW = 45.13 kDa.

6.5.4 Preparation of mRNA-peptide fusion library for XRCC4 selection

For mRNA-peptide fusion preparation, see **5.5.3**. The translation volume was reduced to 500 μL for rounds 2-4 and 250 μL for rounds 5-10. Purification reagent usage, wash and elution volumes were adjusted accordingly to coincide with translation volume.

6.5.5 Selection against XRCC4

For round one of selection, 1 mL of 20 μM GST-XRCC4₁₋₁₅₇ (with 0.1% BSA) was bound to 100 μL of glutathione magnetic beads equilibrated per the manufacturer's instructions, with an additional final 1 mL wash with selection buffer (**4.5.6**). Pelleted fusions (**from 6.5.4**) were resuspended in 0.8 mL of selection buffer containing 0.1% BSA and added to the tube containing immobilized GST-XRCC4₁₋₁₅₇ and rotated at 4 °C for 1 h. The beads were washed once with 200 μL of selection buffer, then 150 μL of GSH elution buffer (250 mM Tris-HCl pH 9.0, 500 mM NaCl, 100 mM reduced glutathione, 1% Triton X-100) was added to the tube and rotated at RT for 15 min. The supernatant was removed and dialyzed overnight against 0.1% Triton X-100. Following PCR amplification using SAX4FWD and SAX4REV (2 min. at 94°C, followed by 24 rounds of 94° C (30s), 65° C (30s), 72° C (45s), library DNA was extracted with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with 3 volumes of ethanol and 0.1 vol of 3 M KOAc pH 5.2. The DNA was resuspended in water and served as the template for Round 2 of selection. The stringency was increased by pre-clearing against immobilized GST beginning in round 2 and continuing until the end of selection; Washing thrice with 1 mL of selection buffer in round two and four times beginning in round 3 until the end of selection; The concentration of GST-XRCC4₁₋₁₅₇ was decreased to 10 μM in round 2 and 1 μM in round 3 until the end of selection.

6.5.6 Sequencing

Following round 10, DNA was sequenced using the Illumina Mi-Seq platform (**4.5.13**). The raw sequences were analyzed using AptaTools software (**4.2.2**).¹⁶⁷⁻¹⁶⁹

Bibliography

References

1. Costa, F. F. Non-coding RNAs: new players in eukaryotic biology. *Gene* **2005**, *357*, 83-94.
2. Hüttenhofer, A.; Schattner, P.; Polacek, N. Non-coding RNAs: hope or hype? *Trends Genet.* **2005**, *21*, 289-297.
3. Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature* **1981**, *289*, 751-758.
4. Nirenberg, M. W.; Matthaei, J. H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **1961**, *47*, 1588-1602.
5. Spirin, A. S.; Ovodov, S. Y.; Alakhov, Y. B. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **1988**, *242*, 1162-1164.
6. Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **2001**, *19*, 751-755.
7. Crick, F.; Barnett, L.; Brenner, S.; Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature* **1961**, *192*, 1227-1232.
8. Crick, F. H. On protein synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 8.
9. Barrell, B.; Bankier, A.; Drouin, J. A different genetic code in human mitochondria. *Nature* **1979**, *282*, 189-194.
10. Ambrogelly, A.; Palioura, S.; Söll, D. Natural expansion of the genetic code. *Nature* **2007**, *3*, 29-35.
11. Bock, A.; Forchhammer, K.; Heider, J.; Baron, C. Selenoprotein synthesis: an expansion of the genetic code. *Trends Biochem. Sci.* **1991**, *16*, 463-467.
12. Hao, B.; Gong, W.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **2002**, *296*, 1462-1466.
13. Srinivasan, G.; James, C. M.; Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **2002**, *296*, 1459-1462.
14. Wang, L.; Xie, J.; Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225-249.

15. Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **2010**, *79*, 413-444.
16. Quast, R. B.; Mrusek, D.; Hoffmeister, C.; Sonnabend, A.; Kubick, S. Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis. *FEBS Lett.* **2015**, *589*, 1703-1712.
17. Noren, C. I.; Anthony-Cahill, S. I.; Griffith, M. C.; Schultz, P. G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids. *Science* **1989**, *244*, 182-188.
18. Bain, J.; Diala, E. S.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **1989**, *111*, 8013-8014.
19. Mamaev, S.; Olejnik, J.; Olejnik, E. K.; Rothschild, K. J. Cell-free N-terminal protein labeling using initiator suppressor tRNA. *Anal. Biochem.* **2004**, *326*, 25-32.
20. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020-11024.
21. Xie, J.; Liu, W.; Schultz, P. G. A genetically encoded bidentate, Metal-Binding amino acid. *Angew. Chem.* **2007**, *119*, 9399-9402.
22. Lee, H. S.; Spraggon, G.; Schultz, P. G.; Wang, F. Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. *J. Am. Chem. Soc.* **2009**, *131*, 2481-2483.
23. Cusack, S. Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.* **1997**, *7*, 881-889.
24. Josephson, K.; Hartman, M. C.; Szostak, J. W. Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.* **2005**, *127*, 11727-11735.
25. Hartman, M. C.; Josephson, K.; Szostak, J. W. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 4356-4361.
26. Hartman, M. C.; Josephson, K.; Lin, C. W.; Szostak, J. W. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS One* **2007**, *2*, e972.
27. Perona, J. J.; Hadd, A. Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry* **2012**, *51*, 8705-8729.
28. Bessho, Y.; Hodgson, D. R.; Suga, H. A tRNA aminoacylation system for non-natural amino acids based on a programmable ribozyme. *Nat. Biotechnol.* **2002**, *20*, 723-728.

29. Murakami, H.; Saito, H.; Suga, H. A versatile tRNA aminoacylation catalyst based on RNA. *Chem. Biol.* **2003**, *10*, 655-662.
30. Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods* **2006**, *3*, 357-359.
31. Niwa, N.; Yamagishi, Y.; Murakami, H.; Suga, H. A flexizyme that selectively charges amino acids activated by a water-friendly leaving group. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3892-3894.
32. Goto, Y.; Murakami, H.; Suga, H. Initiating translation with D-amino acids. *RNA* **2008**, *14*, 1390-1398.
33. Kawakami, T.; Murakami, H.; Suga, H. Messenger RNA-programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol.* **2008**, *15*, 32-42.
34. Kawakami, T.; Murakami, H.; Suga, H. Ribosomal Synthesis of Polypeptoids and Peptoid–Peptide Hybrids. *J. Am. Chem. Soc.* **2008**, *130*, 16861-16863.
35. Puente, X. S.; Lopez-Otin, C. A genomic analysis of rat proteases and protease inhibitors. *Genome Res.* **2004**, *14*, 609-622.
36. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2012**, *64*, 4-17.
37. Hewitt, W. M.; Leung, S. S.; Pye, C. R.; Ponkey, A. R.; Bednarek, M.; Jacobson, M. P.; Lokey, R. S. Cell-permeable cyclic peptides from synthetic libraries inspired by natural products. *J. Am. Chem. Soc.* **2015**, *137*, 715-721.
38. Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. N-methylation of peptides: a new perspective in medicinal chemistry. *Acc. Chem. Res.* **2008**, *41*, 1331-1342.
39. Chatterjee, J.; Rechenmacher, F.; Kessler, H. N-Methylation of Peptides and Proteins: An Important Element for Modulating Biological Functions. *Angew. Chem., Int. Ed.* **2013**, *52*, 254-269.
40. Chatterjee, J.; Ovadia, O.; Zahn, G.; Marinelli, L.; Hoffman, A.; Gilon, C.; Kessler, H. Multiple N-methylation by a designed approach enhances receptor selectivity. *J. Med. Chem.* **2007**, *50*, 5878-5881.
41. Biron, E.; Chatterjee, J.; Ovadia, O.; Langenegger, D.; Brueggen, J.; Hoyer, D.; Schmid, H. A.; Jelinek, R.; Gilon, C.; Hoffman, A. Improving Oral Bioavailability of Peptides by Multiple N-Methylation: Somatostatin Analogues. *Angew. Chem., Int. Ed.* **2008**, *47*, 2595-2599.

42. Molhoek, E. M.; Van Dijk, A.; Veldhuizen, E. J.; Haagsman, H. P.; Bikker, F. J. Improved proteolytic stability of chicken cathelicidin-2 derived peptides by D-amino acid substitutions and cyclization. *Peptides* **2011**, *32*, 875-880.
43. Ovadia, O.; Greenberg, S.; Chatterjee, J.; Laufer, B.; Opperer, F.; Kessler, H.; Gilon, C.; Hoffman, A. The effect of multiple N-methylation on intestinal permeability of cyclic hexapeptides. *Mol. Pharmaceutics* **2011**, *8*, 479-487.
44. Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H. Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library. *Chem. Biol.* **2011**, *18*, 1562-1570.
45. Sako, Y.; Goto, Y.; Murakami, H.; Suga, H. Ribosomal synthesis of peptidase-resistant peptides closed by a nonreducible inter-side-chain bond. *ACS Chem. Biol.* **2008**, *3*, 241-249.
46. Dinsmore, C. J.; Bogusky, M. J.; Culberson, J. C.; Bergman, J. M.; Homnick, C. F.; Zartman, C. B.; Mosser, S. D.; Schaber, M. D.; Robinson, R. G.; Koblan, K. S. Conformational restriction of flexible ligands guided by the transferred noe experiment: potent macrocyclic inhibitors of farnesyltransferase. *J. Am. Chem. Soc.* **2001**, *123*, 2107-2108.
47. Khan, A. R.; Parrish, J. C.; Fraser, M. E.; Smith, W. W.; Bartlett, P. A.; James, M. N. Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry* **1998**, *37*, 16839-16845.
48. Le Nguyen, D.; Heitz, A.; Chiche, L.; Castro, B.; Boigegrain, R.; Favel, A.; Coletti-Previero, M. Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. *Biochimie* **1990**, *72*, 431-435.
49. Clark, R. J.; Fischer, H.; Dempster, L.; Daly, N. L.; Rosengren, K. J.; Nevin, S. T.; Meunier, F. A.; Adams, D. J.; Craik, D. J. Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13767-13772.
50. Rozek, A.; Powers, J. S.; Friedrich, C. L.; Hancock, R. E. Structure-based design of an indolicidin peptide analogue with increased protease stability. *Biochemistry* **2003**, *42*, 14130-14138.
51. Rezai, T.; Bock, J. E.; Zhou, M. V.; Kalyanaraman, C.; Lokey, R. S.; Jacobson, M. P. Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: successful in silico prediction of the relative permeabilities of cyclic peptides. *J. Am. Chem. Soc.* **2006**, *128*, 14073-14080.
52. Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J. Am. Chem. Soc.* **2006**, *128*, 2510-2511.

53. Millward, S. W.; Takahashi, T. T.; Roberts, R. W. A general route for post-translational cyclization of mRNA display libraries. *J. Am. Chem. Soc.* **2005**, *127*, 14142-14143.
54. Seebeck, F. P.; Szostak, J. W. Ribosomal synthesis of dehydroalanine-containing peptides. *J. Am. Chem. Soc.* **2006**, *128*, 7150-7151.
55. Guillen Schlippe, Y. V.; Hartman, M. C.; Josephson, K.; Szostak, J. W. In vitro selection of highly modified cyclic peptides that act as tight binding inhibitors. *J. Am. Chem. Soc.* **2012**, *134*, 10469-10477.
56. Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H. Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* **2008**, *3*, 120-129.
57. Yamagishi, Y.; Ashigai, H.; Goto, Y.; Murakami, H.; Suga, H. Ribosomal synthesis of cyclic peptides with a fluorogenic oxidative coupling reaction. *ChemBioChem* **2009**, *10*, 1469-1472.
58. Goto, Y.; Iwasaki, K.; Torikai, K.; Murakami, H.; Suga, H. Ribosomal synthesis of dehydrobutyrine- and methyllanthionine-containing peptides. *Chem. Commun.* **2009**, *23*, 3419-3421.
59. Sako, Y.; Morimoto, J.; Murakami, H.; Suga, H. Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions. *J. Am. Chem. Soc.* **2008**, *130*, 7232-7234.
60. Kawakami, T.; Ohta, A.; Ohuchi, M.; Ashigai, H.; Murakami, H.; Suga, H. Diverse backbone-cyclized peptides via codon reprogramming. *Nat. Chem. Biol.* **2009**, *5*, 888-890.
61. Luo, H.; Hallen-Adams, H. E.; Scott-Craig, J. S.; Walton, J. D. Ribosomal biosynthesis of α -amanitin in *Galerina marginata*. *Fungal Genet. Biol.* **2012**, *49*, 123-129.
62. Tezuka, Y.; Oike, H. Topological polymer chemistry: systematic classification of nonlinear polymer topologies. *J. Am. Chem. Soc.* **2001**, *123*, 11570-11576.
63. Chafin, D. R.; Guo, H.; Price, D. H. Action of alpha-amanitin during pyrophosphorolysis and elongation by RNA polymerase II. *J. Biol. Chem.* **1995**, *270*, 19114-19119.
64. Angelini, A.; Cendron, L.; Chen, S.; Touati, J.; Winter, G.; Zanotti, G.; Heinis, C. Bicyclic peptide inhibitor reveals large contact interface with a protease target. *ACS Chem. Biol.* **2012**, *7*, 817-821.
65. Bionda, N.; Fasan, R. Ribosomal Synthesis of Natural-Product-Like Bicyclic Peptides in *Escherichia coli*. *ChemBioChem* **2015**, *16*, 2011-2016.
66. Getz, J. A.; Rice, J. J.; Daugherty, P. S. Protease-resistant peptide ligands from a knottin scaffold library. *ACS Chem. Biol.* **2011**, *6*, 837-844.

67. Bartoloni, M.; Jin, X.; Marcaida, M. J.; Banha, J.; Dibonaventura, I.; Bongoni, S.; Bartho, K.; Gräbner, O.; Sefkow, M.; Darbre, T. Bridged bicyclic peptides as potential drug scaffolds. *Chem. Sci.* **2015**, *6*, 5473-5490.
68. Lian, W.; Jiang, B.; Qian, Z.; Pei, D. Cell-permeable bicyclic peptide inhibitors against intracellular proteins. *J. Am. Chem. Soc.* **2014**, *136*, 9830-9833.
69. Trinh, T. B.; Upadhyaya, P.; Qian, Z.; Pei, D. Discovery of a Direct Ras Inhibitor by Screening a Combinatorial Library of Cell-Permeable Bicyclic Peptides. *ACS Comb. Sci.* **2015**, *18*, 75-85.
70. Quartararo, J. S.; Eshelman, M. R.; Peraro, L.; Yu, H.; Baleja, J. D.; Lin, Y.; Kritzer, J. A. A bicyclic peptide scaffold promotes phosphotyrosine mimicry and cellular uptake. *Bioorg. Med. Chem.* **2014**, *22*, 6387-6391.
71. Chen, S.; Rentero Rebollo, I.; Buth, S. A.; Morales-Sanfrutos, J.; Touati, J.; Leiman, P. G.; Heinis, C. Bicyclic peptide ligands pulled out of cysteine-rich peptide libraries. *J. Am. Chem. Soc.* **2013**, *135*, 6562-6569.
72. Feener, E. P.; Shen, W. C.; Ryser, H. J. Cleavage of disulfide bonds in endocytosed macromolecules. A processing not associated with lysosomes or endosomes. *J. Biol. Chem.* **1990**, *265*, 18780-18785.
73. Yang, J.; Chen, H.; Vlahov, I. R.; Cheng, J. X.; Low, P. S. Evaluation of disulfide reduction during receptor-mediated endocytosis by using FRET imaging. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13872-13877.
74. Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **2009**, *5*, 502-507.
75. Chen, S.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C. Peptide ligands stabilized by small molecules. *Angew. Chem., Int. Ed.* **2014**, *53*, 1602-1606.
76. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. The future of peptide-based drugs. *Chem. Biol. Drug Des.* **2013**, *81*, 136-147.
77. Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Last 25 Years. *J. Nat. Prod.* **2007**, *70*, 461-477.
78. Dolle, R. E. In *In Comprehensive survey of combinatorial library synthesis*; J. Comb. Chem; Citeseer: 1999; .
79. Hermkens, P. H.; Ottenheijm, H. C.; Rees, D. Solid-phase organic reactions: a review of the recent literature. *Tetrahedron* **1996**, *52*, 4527-4554.

80. Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **2000**, *287*, 1964-1969.
81. Kaiser, M.; Wetzel, S.; Kumar, K.; Waldmann, H. Biology-inspired synthesis of compound libraries. *Cell. Mol. Life Sci.* **2008**, *65*, 1186-1201.
82. Noren-Muller, A.; Reis-Correa, I., Jr; Prinz, H.; Rosenbaum, C.; Saxena, K.; Schwalbe, H. J.; Vestweber, D.; Cagna, G.; Schunk, S.; Schwarz, O.; Schiewe, H.; Waldmann, H. Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10606-10611.
83. Ponomarenko, E. A.; Poverennaya, E. V.; Ilgisonis, E. V.; Pyatnitskiy, M. A.; Kopylov, A. T.; Zgoda, V. G.; Lisitsa, A. V.; Archakov, A. I. The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* **2016**, *2016*, 1-7.
84. Smith, L. M.; Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186-187.
85. Landry, Y.; Gies, J. Drugs and their molecular targets: an updated overview. *Fundam. Clin. Pharmacol.* **2008**, *22*, 1-18.
86. Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727-730.
87. Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731-1737.
88. Conte, L. L.; Chothia, C.; Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **1999**, *285*, 2177-2198.
89. Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C. Binding of small molecules to an adaptive protein-protein interface. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 1603-1608.
90. Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. Site-directed ligand discovery. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9367-9372.
91. Du, L.; Sánchez, C.; Shen, B. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metab. Eng.* **2001**, *3*, 78-95.
92. Bowers, A. A. Biochemical and biosynthetic preparation of natural product-like cyclic peptide libraries. *MedChemComm* **2012**, *3*, 905-915.

93. Fischbach, M. A.; Lai, J. R.; Roche, E. D.; Walsh, C. T.; Liu, D. R. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11951-11956.
94. Hoyer, K. M.; Mahlert, C.; Marahiel, M. A. The iterative gramicidin S thioesterase catalyzes peptide ligation and cyclization. *Chem. Biol.* **2007**, *14*, 13-22.
95. Sergeeva, A.; Kolonin, M. G.; Mollrem, J. J.; Pasqualini, R.; Arap, W. Display technologies: application for the discovery of drug and gene delivery agents. *Adv. Drug Delivery Rev.* **2006**, *58*, 1622-1654.
96. Murray, C. J.; Baliga, R. Cell-free translation of peptides and proteins: from high throughput screening to clinical production. *Curr. Opin. Chem. Biol.* **2013**, *17*, 420-426.
97. Ullman, C. G.; Frigotto, L.; Cooley, R. N. In vitro methods for peptide display and their applications. *Briefings Funct. Genomics* **2011**, *10*, 125-134.
98. Wada, A. Development of next-generation peptide binders using in vitro display technologies and their potential applications. *Front. Immunol.* **2013**, *4*, 224.
99. Smith, G. P. Filamentous phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985**, *228*, 1315-1317.
100. Sidhu, S. S. Phage display in pharmaceutical biotechnology. *Curr. Opin. Biotechnol.* **2000**, *11*, 610-616.
101. Sidhu, S. S.; Lowman, H. B.; Cunningham, B. C.; Wells, J. A. [21] Phage display for selection of novel binding peptides. *Meth. Enzymol.* **2000**, *328*, 333-363.
102. Rothe, A.; Hosse, R. J.; Power, B. E. In vitro display technologies reveal novel biopharmaceutics. *FASEB J.* **2006**, *20*, 1599-1610.
103. Mattheakis, L. C.; Bhatt, R. R.; Dower, W. J. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 9022-9026.
104. Bonetto, S.; Spadola, L.; Buchanan, A. G.; Jermutus, L.; Lund, J. Identification of cyclic peptides able to mimic the functional epitope of IgG1-Fc for human Fc gammaRI. *FASEB J.* **2009**, *23*, 575-585.
105. Thom, G.; Cockroft, A. C.; Buchanan, A. G.; Candotti, C. J.; Cohen, E. S.; Lowne, D.; Monk, P.; Shorrock-Hart, C. P.; Jermutus, L.; Minter, R. R. Probing a protein-protein interaction by in vitro evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 7619-7624.
106. Fennell, B.; Darmanin-Sheehan, A.; Hufton, S.; Calabro, V.; Wu, L.; Müller, M.; Cao, W.; Gill, D.; Cunningham, O.; Finlay, W. Dissection of the IgNAR V domain: molecular

- scanning and orthologue database mining define novel IgNAR hallmarks and affinity maturation mechanisms. *J. Mol. Biol.* **2010**, *400*, 155-170.
107. Roberts, R. W.; Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 12297-12302.
108. Kurz, M.; Gu, K.; Lohse, P. A. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. *Nucleic Acids Res.* **2000**, *28*, E83.
109. Hayashi, Y.; Morimoto, J.; Suga, H. In vitro selection of anti-Akt2 thioether-macrocyclic peptides leading to isoform-selective inhibitors. *ACS Chem. Biol.* **2012**, *7*, 607-613.
110. Ishizawa, T.; Kawakami, T.; Reid, P. C.; Murakami, H. TRAP display: a high-speed selection method for the generation of functional polypeptides. *J. Am. Chem. Soc.* **2013**, *135*, 5433-5440.
111. Kawakami, T.; Ishizawa, T.; Fujino, T.; Reid, P. C.; Suga, H.; Murakami, H. In vitro selection of multiple libraries created by genetic code reprogramming to discover macrocyclic peptides that antagonize VEGFR2 activity in living cells. *ACS Chem. Biol.* **2013**, *8*, 1205-1214.
112. Fujino, T.; Goto, Y.; Suga, H.; Murakami, H. Ribosomal Synthesis of Peptides with Multiple β -Amino Acids. *J. Am. Chem. Soc.* **2016**, *138*, 1962-1969.
113. Maini, R.; Nguyen, D. T.; Chen, S.; Dedkova, L. M.; Chowdhury, S. R.; Alcalá-Torano, R.; Hecht, S. M. Incorporation of β -amino acids into dihydrofolate reductase by ribosomes having modifications in the peptidyltransferase center. *Bioorg. Med. Chem.* **2013**, *21*, 1088-1096.
114. White, E. R.; Reed, T. M.; Ma, Z.; Hartman, M. C. Replacing amino acids in translation: Expanding chemical diversity with non-natural variants. *Methods* **2013**, *60*, 70-74.
115. Pedersen, J.; Lauritzen, C.; Madsen, M. T.; Dahl, S. W. Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases. *Protein Expression Purif.* **1999**, *15*, 389-400.
116. Milligan, J. F.; Uhlenbeck, O. C. [5] Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **1989**, *180*, 51-62.
117. Ma, Z.; Hartman, M. C. In vitro selection of unnatural cyclic peptide libraries via mRNA display. *Ribosome Display and Related Technologies: Methods and Protocols* **2012**, 367-390.
118. Soutourina, J.; Plateau, P.; Blanquet, S. Metabolism of D-aminoacyl-tRNAs in *Escherichia coli* and *Saccharomyces cerevisiae* cells. *J. Biol. Chem.* **2000**, *275*, 32535-32542.

119. Passioura, T.; Katoh, T.; Goto, Y.; Suga, H. Selection-based discovery of druglike macrocyclic peptides. *Annu. Rev. Biochem.* **2014**, *83*, 727-752.
120. Bashiruddin, N. K.; Suga, H. Construction and screening of vast libraries of natural product-like macrocyclic peptides using in vitro display technologies. *Curr. Opin. Chem. Biol.* **2015**, *24*, 131-138.
121. Frost, J. R.; Smith, J. M.; Fasan, R. Design, synthesis, and diversification of ribosomally derived peptide macrocycles. *Curr. Opin. Struct. Biol.* **2013**, *23*, 571-580.
122. White, C. J.; Yudin, A. K. Contemporary strategies for peptide macrocyclization. *Nat. Chem.* **2011**, *3*, 509-524.
123. Timmerman, P.; Beld, J.; Puijk, W. C.; Meloen, R. H. Rapid and quantitative cyclization of multiple peptide loops onto synthetic scaffolds for structural mimicry of protein surfaces. *ChemBioChem* **2005**, *6*, 821-824.
124. Dewkar, G. K.; Carneiro, P. B.; Hartman, M. C. Synthesis of novel peptide linkers: Simultaneous cyclization and labeling. *Org. Lett.* **2009**, *11*, 4708-4711.
125. Wang, C. K.; Hu, S. H.; Martin, J. L.; Sjogren, T.; Hajdu, J.; Bohlin, L.; Claeson, P.; Goransson, U.; Rosengren, K. J.; Tang, J.; Tan, N. H.; Craik, D. J. Combined X-ray and NMR analysis of the stability of the cyclotide cystine knot fold that underpins its insecticidal activity and potential use as a drug scaffold. *J. Biol. Chem.* **2009**, *284*, 10672-10683.
126. Saez, N. J.; Senff, S.; Jensen, J. E.; Er, S. Y.; Herzig, V.; Rash, L. D.; King, G. F. Spider-venom peptides as therapeutics. *Toxins* **2010**, *2*, 2851-2871.
127. Colgrave, M. L.; Craik, D. J. Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry* **2004**, *43*, 5965-5975.
128. Tam, J. P.; Lu, Y. A.; Yang, J. L.; Chiu, K. W. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 8913-8918.
129. Tang, Y. Q.; Yuan, J.; Osapay, G.; Osapay, K.; Tran, D.; Miller, C. J.; Ouellette, A. J.; Selsted, M. E. A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science* **1999**, *286*, 498-502.
130. Kim, J.; Park, S.; Hwang, I.; Cheong, H.; Nah, J.; Hahm, K.; Park, Y. Protease inhibitors from plants with antimicrobial activity. *Int. J. Mol. Sci.* **2009**, *10*, 2860-2872.
131. Daly, N. L.; Ekberg, J. A.; Thomas, L.; Adams, D. J.; Lewis, R. J.; Craik, D. J. Structures of muO-conotoxins from *Conus marmoreus*. Inhibitors of tetrodotoxin (TTX)-sensitive and

- TTX-resistant sodium channels in mammalian sensory neurons. *J. Biol. Chem.* **2004**, *279*, 25774-25782.
132. Olivera, B. M.; Rivier, J.; Clark, C.; Ramilo, C. A.; Corpuz, G. P.; Abogadie, F. C.; Mena, E. E.; Woodward, S. R.; Hillyard, D. R.; Cruz, L. J. Diversity of *Conus* neuropeptides. *Science* **1990**, *249*, 257-263.
133. Olivera, B. M. *Conus* peptides: biodiversity-based discovery and exogenomics. *J. Biol. Chem.* **2006**, *281*, 31173-31177.
134. Moore, S. J.; Hayden Gephart, M. G.; Bergen, J. M.; Su, Y. S.; Rayburn, H.; Scott, M. P.; Cochran, J. R. Engineered knottin peptide enables noninvasive optical imaging of intracranial medulloblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14598-14603.
135. Kimura, R. H.; Teed, R.; Hackel, B. J.; Pysz, M. A.; Chuang, C. Z.; Sathirachinda, A.; Willmann, J. K.; Gambhir, S. S. Pharmacokinetically stabilized cystine knot peptides that bind alpha-v-beta-6 integrin with single-digit nanomolar affinities for detection of pancreatic cancer. *Clin. Cancer Res.* **2012**, *18*, 839-849.
136. Cox, N.; Kintzing, J. R.; Smith, M.; Grant, G. A.; Cochran, J. R. Integrin-Targeting Knottin Peptide-Drug Conjugates Are Potent Inhibitors of Tumor Cell Proliferation. *Angew. Chem., Int. Ed.* **2016**, *55*, 9894-9897.
137. Kolb, H. C.; Finn, M.; Sharpless, K. B. Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004-2021.
138. Hacker, D. E.; Almohaini, M.; Anbazhagan, A.; Ma, Z.; Hartman, M. C. Peptide and Peptide Library Cyclization via Bromomethylbenzene Derivatives. *Peptide Libraries: Methods and Protocols* **2015**, *1248*, 105-117.
139. White, E. R.; Sun, L.; Ma, Z.; Beckta, J. M.; Danzig, B. A.; Hacker, D. E.; Huie, M.; Williams, D. C.; Edwards, R. A.; Valerie, K. Peptide Library Approach to Uncover Phosphomimetic Inhibitors of the BRCA1 C-Terminal Domain. *ACS Chem. Biol.* **2015**, *10*, 1198-1208.
140. Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. Lasso peptides from proteobacteria: genome mining employing heterologous expression and mass spectrometry. *Pept. Sci.* **2013**, *100*, 527-542.
141. Hashempour, H.; Koehbach, J.; Daly, N. L.; Ghassempour, A.; Gruber, C. W. Characterizing circular peptides in mixtures: sequence fragment assembly of cyclotides from a violet plant by MALDI-TOF/TOF mass spectrometry. *Amino Acids* **2013**, *44*, 581-595.

142. Krug, D.; Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **2014**, *31*, 768-783.
143. King, G. J.; Jones, A.; Kobe, B.; Huber, T.; Mouradov, D.; Hume, D. A.; Ross, I. L. Identification of disulfide-containing chemical cross-links in proteins using MALDI-TOF/TOF-mass spectrometry. *Anal. Chem.* **2008**, *80*, 5036-5043.
144. Peraro, L.; Siegert, T.; Kritzer, J. Chapter Fourteen-Conformational Restriction of Peptides Using Dithiol Bis-Alkylation. *Methods Enzymol.* **2016**, *580*, 303-332.
145. Sarver, A.; Scheffler, N. K.; Shetlar, M. D.; Gibson, B. W. Analysis of peptides and proteins containing nitrotyrosine by matrix-assisted laser desorption/ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 439-448.
146. Kuhn, D. M.; Sadidi, M.; Liu, X.; Kreipke, C.; Geddes, T.; Borges, C.; Watson, J. T. Peroxynitrite-induced nitration of tyrosine hydroxylase: identification of tyrosines 423, 428, and 432 as sites of modification by matrix-assisted laser desorption ionization time-of-flight mass spectrometry and tyrosine-scanning mutagenesis. *J. Biol. Chem.* **2002**, *277*, 14336-14342.
147. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues. *ChemBioChem* **2002**, *3*, 235-237.
148. Link, A. J.; Vink, M. K.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10180-10185.
149. 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. U. S. A.* **2002**, *99*, 19-24.
150. Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chem., Int. Ed.* **2009**, *48*, 9879-9883.
151. Schägger, H. Tricine-SDS-PAGE. *Nat. Protoc.* **2006**, *1*, 16-22.
152. Rentero Rebollo, I.; Sabisz, M.; Baeriswyl, V.; Heinis, C. Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic Acids Res.* **2014**, *42*, e169.
153. Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett.* **1997**, *414*, 405-408.

154. Paredes, E.; Das, S. R. Click chemistry for rapid labeling and ligation of RNA. *ChemBioChem* **2011**, *12*, 125-131.
155. Abel Jr, G. R.; Calabrese, Z. A.; Ayco, J.; Hein, J. E.; Ye, T. Measuring and Suppressing the Oxidative Damage to DNA During Cu (I)-Catalyzed Azide–Alkyne Cycloaddition. *Bioconjugate Chem.* **2016**, *27*, 698-704.
156. Chen, X.; Khairallah, G. N.; Richard, A.; Williams, S. J. Fixed-charge labels for simplified reaction analysis: 5-hydroxy-1, 2, 3-triazoles as byproducts of a copper (I)-catalyzed click reaction. *Tetrahedron Lett.* **2011**, *52*, 2750-2753.
157. Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. Labeling live cells by copper-catalyzed alkyne–azide click chemistry. *Bioconjugate Chem.* **2010**, *21*, 1912-1916.
158. Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. Polytriazoles as copper (I)-stabilizing ligands in catalysis. *Org. Lett.* **2004**, *6*, 2853-2855.
159. Kennedy, D. C.; McKay, C. S.; Legault, M. C.; Danielson, D. C.; Blake, J. A.; Pegoraro, A. F.; Stolow, A.; Mester, Z.; Pezacki, J. P. Cellular consequences of copper complexes used to catalyze bioorthogonal click reactions. *J. Am. Chem. Soc.* **2011**, *133*, 17993-18001.
160. Li, S.; Cai, H.; He, J.; Chen, H.; Lam, S.; Cai, T.; Zhu, Z.; Bark, S. J.; Cai, C. Extent of the oxidative side reactions to peptides and proteins during the CuAAC reaction. *Bioconjugate Chem.* **2016**, *27*, 2315-2322.
161. Haldon, E.; Nicasio, M. C.; Pérez, P. J. Copper-catalysed azide–alkyne cycloadditions (CuAAC): an update. *Org. Biomol. Chem.* **2015**, *13*, 9528-9550.
162. Soriano del Amo, D.; Wang, W.; Jiang, H.; Besanceney, C.; Yan, A. C.; Levy, M.; Liu, Y.; Marlow, F. L.; Wu, P. Biocompatible copper (I) catalysts for in vivo imaging of glycans. *J. Am. Chem. Soc.* **2010**, *132*, 16893-16899.
163. Lallana, E.; Riguera, R.; Fernandez-Megia, E. Reliable and efficient procedures for the conjugation of biomolecules through Huisgen azide–alkyne cycloadditions. *Angew. Chem., Int. Ed.* **2011**, *50*, 8794-8804.
164. Mendes, K.; Ndungu, J. M.; Clark, L. F.; Kodadek, T. Optimization of the magnetic recovery of hits from one-bead–one-compound library screens. *ACS Comb. Sci.* **2015**, *17*, 506-517.
165. Giebel, L. B.; Cass, R.; Milligan, D. L.; Young, D.; Arze, R.; Johnson, C. Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities. *Biochemistry* **1995**, *34*, 15430-15435.
166. Meyer, S. C.; Gaj, T.; Ghosh, I. Highly selective cyclic peptide ligands for NeutrAvidin and avidin identified by phage display. *Chem. Biol. Drug Des.* **2006**, *68*, 3-10.

167. Hoinka, J.; Przytycka, T. AptaPLEX—A dedicated, multithreaded demultiplexer for HT-SELEX data. *Methods* **2016**, *106*, 82-85.
168. Hoinka, J.; Berezhnoy, A.; Dao, P.; Sauna, Z. E.; Gilboa, E.; Przytycka, T. M. Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery. *Nucleic Acids Res.* **2015**, *43*, 5699-5707.
169. Hoinka, J.; Dao, P.; Przytycka, T. M. AptaGUI-A Graphical User Interface for the Efficient Analysis of HT-SELEX Data. *Mol. Ther. --Nucleic Acids* **2015**, *4*, e257.
170. Keefe, A. D.; Wilson, D. S.; Seelig, B.; Szostak, J. W. One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. *Protein Expression Purif.* **2001**, *23*, 440-446.
171. E Garcia, A.; A Camarero, J. Biological activities of natural and engineered cyclotides, a novel molecular scaffold for peptide-based therapeutics. *Curr. Mol. Pharmacol.* **2010**, *3*, 153-163.
172. Devlin, J. J.; Panganiban, L. C.; Devlin, P. E. Random peptide libraries: a source of specific protein binding molecules. *Science* **1990**, *249*, 404-406.
173. Gissel, B.; Jensen, M. R.; Gregorius, K.; Elsner, H. I.; Svendsen, I.; Mouritsen, S. Identification of avidin and streptavidin binding motifs among peptides selected from a synthetic peptide library consisting solely of D-amino acids. *J. Pept. Sci.* **1995**, *1*, 217-226.
174. Wilson, D. S.; Keefe, A. D.; Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3750-3755.
175. Katz, B. A. Binding to protein targets of peptidic leads discovered by phage display: crystal structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence. *Biochemistry* **1995**, *34*, 15421-15429.
176. Heinis, C.; Winter, G. Encoded libraries of chemically modified peptides. *Curr. Opin. Chem. Biol.* **2015**, *26*, 89-98.
177. Ito, K.; Passioura, T.; Suga, H. Technologies for the synthesis of mRNA-encoding libraries and discovery of bioactive natural product-inspired non-traditional macrocyclic peptides. *Molecules* **2013**, *18*, 3502-3528.
178. Howell, S. M.; Fiacco, S. V.; Takahashi, T. T.; Jalali-Yazdi, F.; Millward, S. W.; Hu, B.; Wang, P.; Roberts, R. W. Serum stable natural peptides designed by mRNA display. *Sci. Rep.* **2014**, *4*, 6008.
179. Korndörfer, I. P.; Skerra, A. Improved affinity of engineered streptavidin for the Strep-tag II peptide is due to a fixed open conformation of the lid-like loop at the binding site. *Protein Sci.* **2002**, *11*, 883-893.

180. McLafferty, M.; Kent, R.; Ladner, R.; Markland, W. M13 bacteriophage displaying disulfide-constrained microproteins. *Gene* **1993**, *128*, 29-36.
181. Schatz, P. J. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Nat. Biotechnol.* **1993**, *11*, 1138-1143.
182. Bessette, P. H.; Rice, J. J.; Daugherty, P. S. Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng., Des. Sel.* **2004**, *17*, 731-739.
183. Gao, Y.; Kodadek, T. Direct comparison of linear and macrocyclic compound libraries as a source of protein ligands. *ACS Comb. Sci.* **2015**, *17*, 190-195.
184. Ahnesorg, P.; Smith, P.; Jackson, S. P. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **2006**, *124*, 301-313.
185. Guirouilh-Barbat, J.; Huck, S.; Bertrand, P.; Pirzio, L.; Desmaze, C.; Sabatier, L.; Lopez, B. S. Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. *Mol. Cell* **2004**, *14*, 611-623.
186. Gu, J.; Lu, H.; Tsai, A. G.; Schwarz, K.; Lieber, M. R. Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res.* **2007**, *35*, 5755-5762.
187. Yano, K.; Morotomi-Yano, K.; Wang, S. Y.; Uematsu, N.; Lee, K. J.; Asaithamby, A.; Weterings, E.; Chen, D. J. Ku recruits XLF to DNA double-strand breaks. *EMBO Rep.* **2008**, *9*, 91-96.
188. Andres, S. N.; Modesti, M.; Tsai, C. J.; Chu, G.; Junop, M. S. Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Mol. Cell* **2007**, *28*, 1093-1101.
189. Hammel, M.; Yu, Y.; Fang, S.; Lees-Miller, S. P.; Tainer, J. A. XLF regulates filament architecture of the XRCC4· ligase IV complex. *Structure* **2010**, *18*, 1431-1442.
190. Dai, Y.; Kysela, B.; Hanakahi, L. A.; Manolis, K.; Riballo, E.; Stumm, M.; Harville, T. O.; West, S. C.; Oettinger, M. A.; Jeggo, P. A. Nonhomologous end joining and V(D)J recombination require an additional factor. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2462-2467.
191. Buck, D.; Malivert, L.; de Chasseval, R.; Barraud, A.; Fondanèche, M.; Sanal, O.; Plebani, A.; Stéphan, J.; Hufnagel, M.; le Deist, F. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **2006**, *124*, 287-299.
192. Martin, L. J.; Sculimbrene, B. R.; Nitz, M.; Imperiali, B. Rapid Combinatorial Screening of Peptide Libraries for the Selection of Lanthanide-Binding Tags (LBTs). *QSAR Comb. Sci.* **2005**, *24*, 1149-1157.

193. Richardson, F. S. Terbium (III) and europium (III) ions as luminescent probes and stains for biomolecular systems. *Chem. Rev.* **1982**, *82*, 541-552.
194. Selvin, P. R. Principles and biophysical applications of lanthanide-based probes. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 275-302.
195. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning*; Cold spring harbor laboratory press New York: 1989; Vol. 2.
196. Ren, G.; Okerberg, C. K.; Mathews, S. T. Ultrasensitive Protein Detection and Imaging: Comparison of Lumitein™, ProteoSilver™, SYPRO® Ruby, and Coomassie® Brilliant Blue Gel Stains. *Protein Electrophoresis: Methods and Protocols* **2012**, *869*, 621-632.
197. Goto, Y.; Katoh, T.; Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* **2011**, *6*, 779-790.
198. Ederth, J.; Mandava, C. S.; Dasgupta, S.; Sanyal, S. A single-step method for purification of active His-tagged ribosomes from a genetically engineered Escherichia coli. *Nucleic Acids Res.* **2009**, *37*, e15.

Appendices

Appendix A: Lanthanide Binding Peptide for Real-Time Translation Monitoring

A1 Introduction

The ability to monitor the yield of *in vitro* ribosomal peptide or protein synthesis is useful in many respects. For example, assessing the relative incorporation of non-proteinogenic amino acids into peptides, optimization or troubleshooting of the many translation components, testing the efficiency of mRNA or DNA templates or determining the effect of stressors or effectors on ribosomal synthesis. Methods currently used for this purpose include radioisotope incorporation, fluorescent labeling and SDS-PAGE. However, these methods can be costly, laborious and, in the case of SDS-PAGE, lacking in sensitivity.

Imperiali and coworkers evolved a short synthetic peptide, termed a Lanthanide Binding Tag (LBT), using a series of combinatorial screens.¹⁹² This peptide bound to terbium with high affinity (30 nM) and contained the spectroscopic properties of lanthanides.^{193,194} The transfer of energy from a tryptophan residue (ex. 280 nm) in the peptide to the coordinating terbium atom, results in luminescence emission at 544 nm.

We hypothesized that if terbium was present in an *in vitro* translation reaction initiated by an mRNA encoding for the LBT, the expressed peptide would rapidly coordinate with terbium and the resulting luminescent signal would be proportional to LBT yield (**Figure A1a**). In this way, we could potentially monitor peptide yield quickly and easily, simply by the addition of terbium to a standard translation reaction.

A2 Results

We designed a set of initial experiments to test the ability of the LBT to bind to terbium and also to determine the effect of terbium on *in vitro* translation. We first produced the peptide

(FIDTNNDGWIEGDELLA-**Figure A1b**) using Fmoc-based microwave-assisted synthesis to conduct our luminescence tests. We then tested the impact of terbium on *in vitro* translated peptide yield by titrating terbium(III) acetate (0.001-100 μM) in standard translation reactions (**2.5.4**). We found that while the addition of terbium negatively impacted translation, particularly at higher concentrations (**Figure A1c**), this decrease was not prohibitive since translation volumes can be increased to compensate for the lower yield.

We then tested the ability of the peptide to transfer energy to terbium by conducting time-resolved luminescence experiments with synthetic LBT. We performed a series of titrations of both terbium and peptide and observed a concentration dependent increase in luminescence (**Figure A2a**). We next quantified the sensitivity of the method (using 11.1 μM terbium) and determined that in order to be above the limit of detection (LOD) of the assay, the minimum quantity of peptide produced during a standard translation reaction (50 μL) would be 12.5 picomoles (**Figure A2b**).

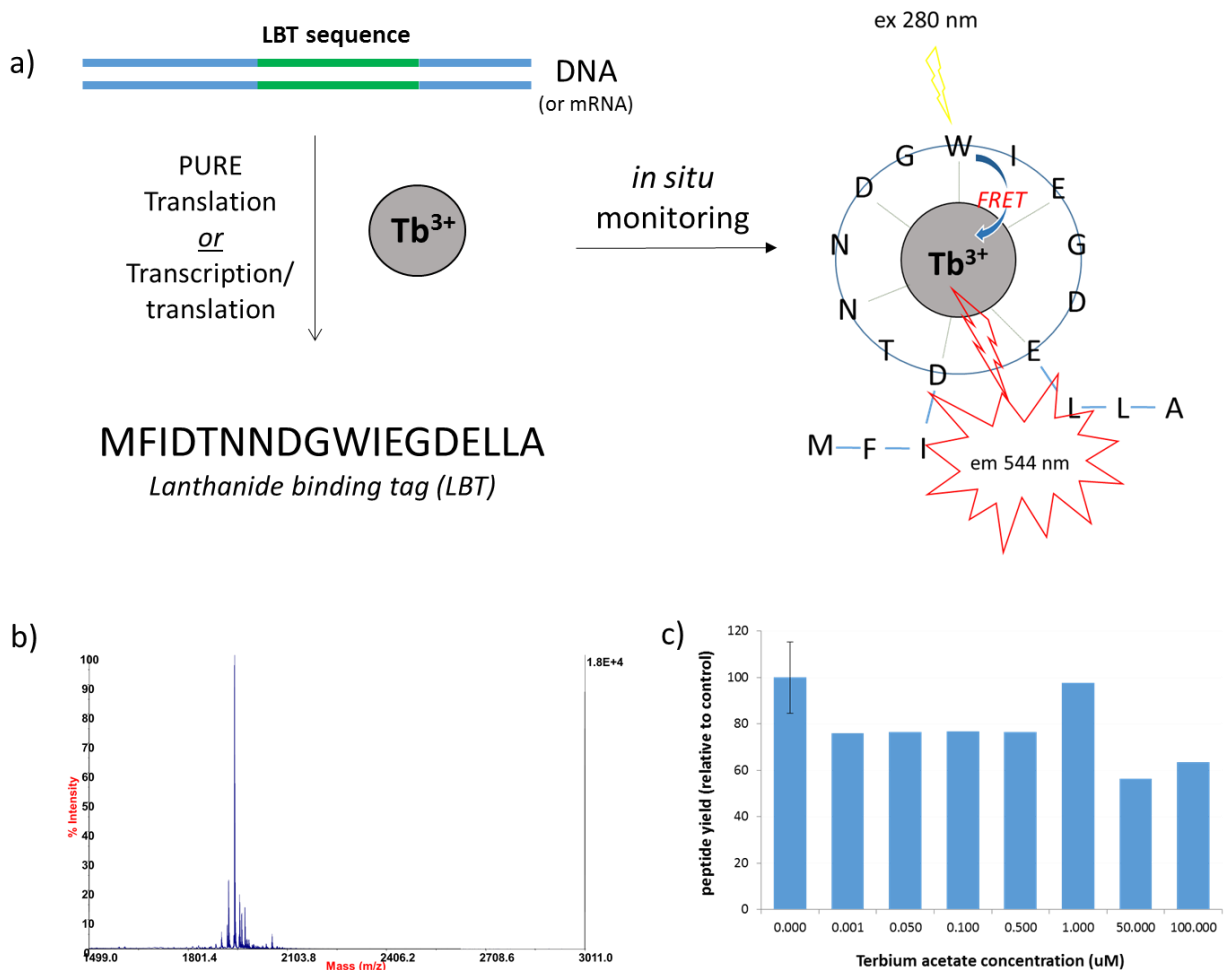


Figure A1 Lanthanide Binding Tag for monitoring peptide synthesis. a) DNA or mRNA encoding the LBT is used to initiate PURE translation containing terbium. Synthesized peptides coordinate Tb^{3+} *in situ*. Transfer of energy (FRET) between tryptophan (ex. 280 nm) and the terbium ion (em. 544 nm) is then used to indirectly determine the yield of peptide in translation. b) MALDI-TOF spectrum of synthetic LBT peptide, Calc (M+Na): 1943.86, Obs: 1942.50. c) Effect of terbium acetate titration on *in vitro* translated ^{35}S -Met peptide yield.

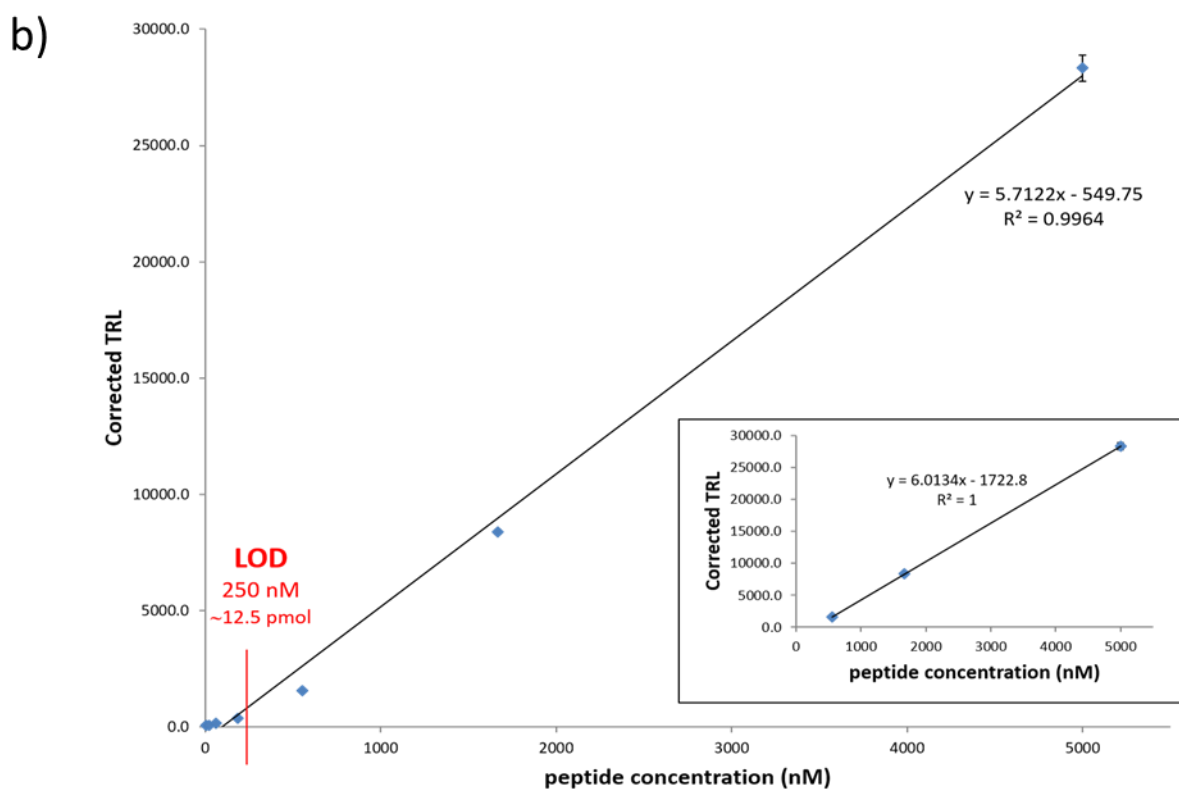
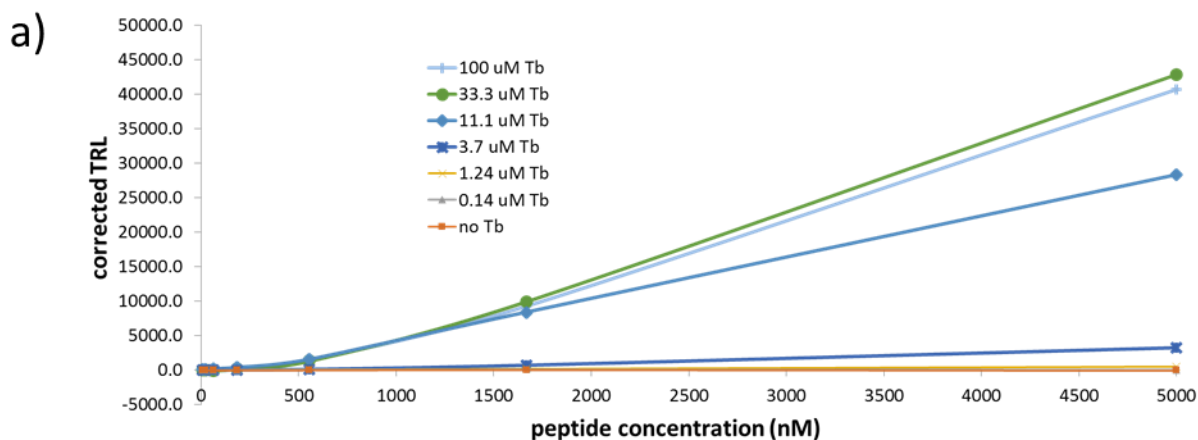


Figure A2 Sensitivity of LBT to terbium energy transfer a) Time-resolved luminescence of peptide and terbium acetate titrations. b) Plot of peptide titration against 11.1 μM terbium. Red line indicates the LOD ($S/N=3$) of the assay, (inset) plot containing a regression analysis of only the points that are above the LOD.

A3 Discussion

We observed a concentration-dependent increase in luminescence when terbium was added to the LBT. When terbium was added to translation, we observed up to a 45% decrease in radiolabeled peptide yield. The sensitivity of the method requires the ribosomal production of 12.5 pmols of LBT. During a standard 50 μ L translation to produce a similarly short peptide, we typically obtain between 5-10 pmols. Moreover, the exploratory experiments were conducted with purified synthetic peptide in buffer. It is likely that the presence of all *in vitro* translation components in the TRL experiment would result in lower sensitivity and increase the LOD. Given the negative impact of terbium on translation, we determined that this method would require translation reaction volumes of 150-300 μ L to synthesize sufficient peptide to make this technique practical. Assays used for our intended purpose often require titrations over a wide range of concentrations and multiple (10-25) parallel individual translation reactions. The reagent volumes required to perform the translations would be very high and the preparation of individual components for the compilation of the PURE system is laborious and time-consuming. While our exploratory experiments with the synthetic peptide demonstrated that the method was valid, the sensitivity was too low for our practical purposes and we chose not to pursue the project further.

A4 Future Directions

The cause of the negative impact of terbium on *in vitro* translation yield is unknown, but this effect could potentially be compensated for by the sequestration of terbium via chelation by the LBT, which would decrease the effective concentration of free terbium ion in translation over time. However, we did not test this hypothesis.

While we assumed the *in vitro* translation yield of the LBT would be proportional to our typical yields, we have seen a wide variety of ‘translatability’ of mRNAs of similar lengths and it is possible that the LBT mRNA would yield more peptide than we calculated. This would decrease translation volumes and could potentially make the method more practical.

A5 Experimental

A5.1 Peptide synthesis. The LBT peptide was synthesized using a Liberty Automated Peptide Synthesizer with a Discover microwave module (CEM). Fmoc-PAL-PEG-PS resin (Applied Biosystems) was chosen as the solid support for a standard 0.1 mmol scale synthesis using N- α -Fmoc-protected amino acids (CEM). The peptide was cleaved from the resin using a cocktail containing trifluoroacetic acid (TFA-Chem Impex)/triisopropylsilane (TIS-Sigma)/3,6-dioxa-1,8-octanedithiol (DODT-Sigma)/water in a ratio of 92.5/2.5/2.5/2.5 and incubated at RT for 4 h. The TFA was bubbled off of the filtrate with argon, the peptide was precipitated with 10 volumes of cold diethyl ether and collected by centrifugation. The supernatant was discarded and the pellet was dissolved in 25% acetonitrile (Fisher Optima grade), followed by freezing and lyophilization. The peptide was purified by reverse phase HPLC using a Shimadzu Prominence instrument with a Vydac Protein & Peptide (P/N 218TP52210) semi-preparative column with a 0.1% TFA in water (A) and 0.1% TFA in acetonitrile (B) as the mobile phase monitoring at 215 nm and 280 nm. Mobile phase flow rate was 10 mL/min. from 20-40% B and t_R was 23.5 min. Mass spectrometry data was collected using a Voyager DE-Pro MALDI instrument with a TOF detector operating in reflectron mode.

A5.2 Time-resolved photoluminescence experiments. TRL measurements were recorded on a Tecan Infinity M1000 plate reader. Tryptophan-sensitized emission spectra were collected using a 280 nm excitation wavelength and emission was detected at 544 nm. Excitation and emission slit widths were 5 nm. Delay and gate times were set at 100 μ s and 2 ms, respectively. Integration time was tested in the range of 400-2000 μ s and lag time at 60 and 100 μ s. Maximal S/N was achieved using 2000 μ s integration time and 100 μ s lag time, and these instrument settings were used for the luminescence experiments. Lyophilized peptide was suspended in 10 mM HEPES buffer containing 100 mM NaCl and concentration was determined using the method described by Gill and von Hippel.¹¹⁰ Tb(III) acetate (Sigma) was dissolved in 1 mM HCl and diluted in the same solvent for the titration experiment.

A5.3 Effect of Terbium on *in vitro* translation. Terbium (III) acetate was dissolved in 1 mM HCl and titrated over the range of 0.001-100 μ M in standard 50 μ L PURE translation reactions (2.5.4) containing ³⁵S-methionine initiated by the addition of 1 μ M mRNA encoding the peptide MHFSWDYKDDDDK (Table 2.1). Radiolabeled flag-tagged peptides were affinity purified with M2 anti-flag agarose (Sigma) using the manufacturer's protocol and eluted with 1% TFA. Eluted peptides were quantified by scintillation.

Appendix B TRAP: Streamlining mRNA display

B1 Introduction

The ability to generate trillion-member peptide libraries for target selection using mRNA display makes it a very powerful tool for ligand discovery. However, the mRNA display process is labor intensive. The covalent linking of the mRNA to the puromycin linker, two mRNA-peptide fusion purification steps, as well as precipitation and dialysis during the fusion preparation are time-consuming. A single round of mRNA display takes 4-5 days. Most selections take between 7 and 10 rounds before enrichment plateaus and DNA can be sequenced. Hence, to perform a selection using mRNA display can take as long as two to three months. When library design, testing and sequencing are considered, the total time could be up to four months. Moreover, there is no easy way to check the progress of the selection process until the end.

To address these limitations, others have designed a streamlined version of mRNA display. TRAP display (Transcription-translation with association of puromycin linker)¹¹⁰ uses DNA, instead of covalently associated mRNA-puromycin linker, as the template for fusion formation (**Figure B1**-route 1). By including T7 RNA polymerase in the reaction, the mRNA is transcribed and hybridizes *in situ* with a long G-C rich region at the 5' end of puromycin linker via a complementary region at the 3' end of the mRNA. The mRNA-puromycin hybrid is then decoded by the ribosome and the translated peptide is covalently linked to the associated puromycin to form the fusions. In TRAP display, the mRNA-peptide fusions formed during the coupled transcription-translation are added directly to the target in the selection step. This eliminates the need for the purification steps present in the mRNA display scheme. These modifications reduce the time significantly and a single round of TRAP display takes only 2-3 hours.

Given the streamlined nature and time savings of TRAP, we designed a system to validate the method for future use in our lab.

TRAP scheme (~3 hrs/round)

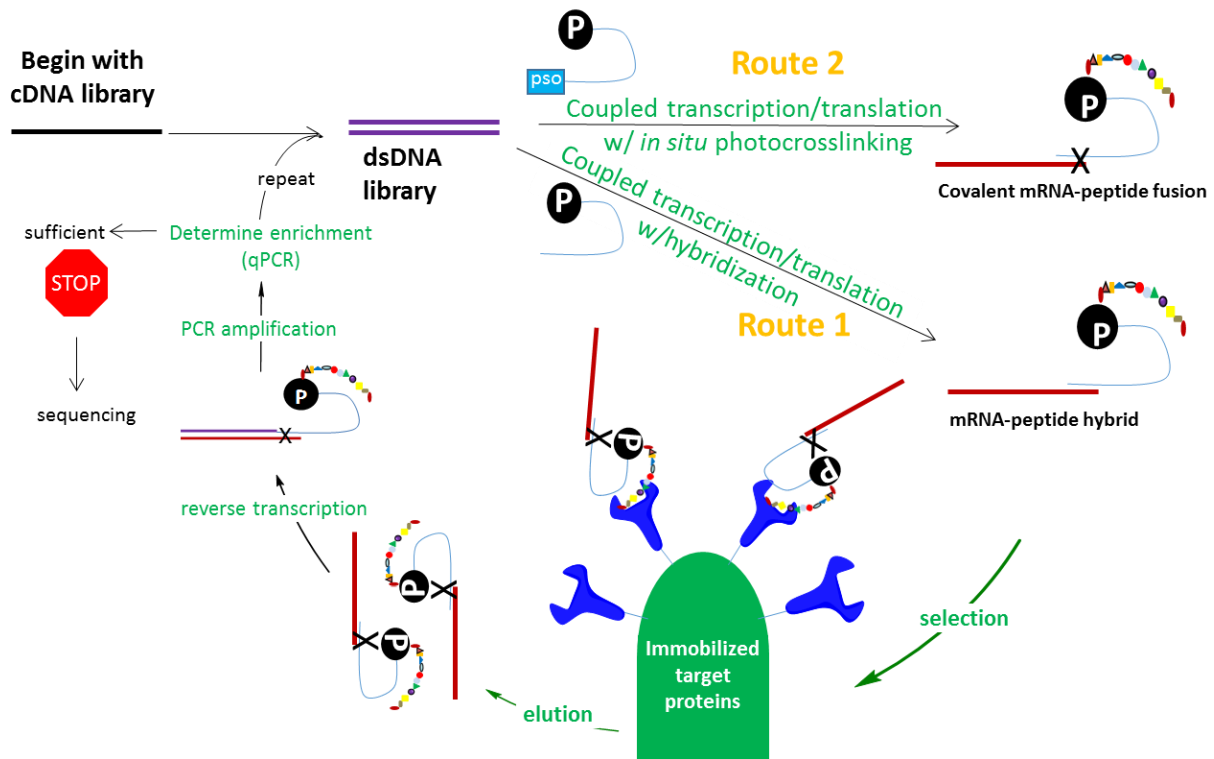


Figure B1 TRAP and modified mRNA display. These display systems use DNA as the template for coupled transcription-translation. In TRAP display (Route 1), the puromycin linker hybridizes *in situ* to generate the non-covalent association between the phenotype and genotype. In the proposed modified mRNA display (Route 2), the conventional UV-induced covalent cross-linking of the mRNA to the puromycin linker is performed *in situ*. Both sets of fusions are then bound directly to the target without purification and reverse transcription is then performed on the functional fusions prior to PCR amplification to generate the template for the next round.

B2 Results

B2.1 TRAP display

We designed a system to test the performance of TRAP display relative to the ability of conventional mRNA display to enrich a known binder among a random library (**Figure B2**). We encoded the Flag epitope sequence (DYKDDDK) within the open reading frame of the DNA containing the constant region elements necessary for conventional mRNA display, which includes a 9 bp region at the 3' end of the mRNA that hybridizes with the puromycin linker and is subsequently linked covalently via the psoralen molecule attached to the 5' end of the linker (**Figure B3a-top**). We created a similar sequence with (NNS)₈ in place of the Flag coding region. We used the identical strategy for the TRAP display construct, which includes a 21 bp G-C rich region at the 3' end of the mRNA that forms a non-covalent hybridization with a complementary region at the 5' end of the TRAP puromycin linker (**Figure B3a-bottom**).

In order to measure enrichment, we strategically placed restriction sites for *PsiI* and *FokI* within the Flag DNA sequence. In this system, if the FLAG peptide remains associated with its mRNA, these mRNAs will be captured selectively and will be amplified by RT-PCR. This increase would be measured by an increase in the intensity of the *PsiI*/*FokI* digestible band on an agarose gel. However, if the FLAG peptide does not associate with its mRNA, no enrichment would be observed.

Our initial control digest confirmed the ability of the restriction enzymes to selectively cut the Flag DNA (**Figure B3b**). Since *PsiI* digested fragments for both DNA libraries (lanes 5 and 11) were more visible on the gel than the equivalent *FokI* fragments (lanes 6 and 12) and given that *FokI* appeared to partially digest the random DNA (lanes 3 and 9), we chose *PsiI* for our test.

With the analysis method in place, we created two sets of doped libraries of Flag:random sequence at ratios of 1:1 and 1:10 for both display methods. The conventional mRNA-peptide fusions (with covalently attached linker) were prepared similarly to our previous selections (4.5.3). The Trap fusions are typically created starting with a DNA template that is transcribed *in situ* to form mRNA, which then forms a hybrid association with the complementary GC-rich region of the puromycin linker. However, since we wanted to test only the ability of the TRAP construct to maintain association with the peptide (through the hybrid linker), we chose to initiate translation with an equal ratio of TRAP mRNA and linker and allow the association to occur *in situ*. While this strategy was not an identical representation of TRAP display, we did not want *in situ* transcription as an uncontrolled variable. Following selection of the two sets of fusions against anti-Flag affinity resin, only TRAP fusions required reverse transcription, after which the cDNA from both methods was PCR amplified, digested with *PsiI* and analyzed on a 4% agarose gel relative to the input templates (Figure B3c). We were surprised to find that both of the methods at each of the ratios resulted in a lower percentage of digested DNA after selection than was present beforehand. The inability to even enrich the conventional template suggests that something was wrong with our selection strategy itself.

B2.2 Modified mRNA display (*in situ* covalent association)

Having been unable to enrich the Flag sequence in either of the templates, we took a different approach to streamlining the mRNA display process. Standard mRNA display utilizes mRNAs photo-crosslinked to the puromycin linker *in vitro*. To eliminate the transcription and crosslinking steps, we decided to attempt to perform transcription and crosslinking *in situ*.

We planned to add the light-activatable puromycin linker to our coupled transcription/translation mixture with the hope that it would hybridize to the mRNA as it was being synthesized (Figure

B1, Route 2). Concurrent UV irradiation would then facilitate *in situ* covalent crosslinking. This would potentially generate a more stable covalent linkage between the genotype and phenotype and, if the resulting fusions were added directly to the target, would eliminate several steps in the cycle and result in a similar time/round as TRAP display.

To test this method, we first labeled the 5' end of the conventional library mRNA with ³²P. We then initiated translation in two ways. In one experiment, we annealed the mRNA with the conventional linker prior to adding them to the reaction; and in the second experiment we added the two components independently, thus simulating the *in situ* association that would exist in the proposed coupled reaction. After addition of the template, we placed a 365 nm handheld lamp directly above the reaction for 20 min. The reactions were done at 37 °C as in the conventional method, and at 4 °C to account for any potential heat provided by the lamp. Since the puromycin linker contains an oligo-A stretch, we also analyzed the capturability of the mRNAs on an oligo(dT) resin.

There was very little cross-linking detected in the reaction mixture (**Figure B4a**) or the oligo(dT) elutions (**Figure B4b**) from the reactions performed at either temperature, for both the annealed and coupled reaction simulation. Analysis via scintillation revealed less than 30% covalent association relative to the control that had been 'pre'cross-linked conventionally. Furthermore, the control reaction that was kept in the dark showed a higher degree of cross-linking than all of the *in situ* experiments.

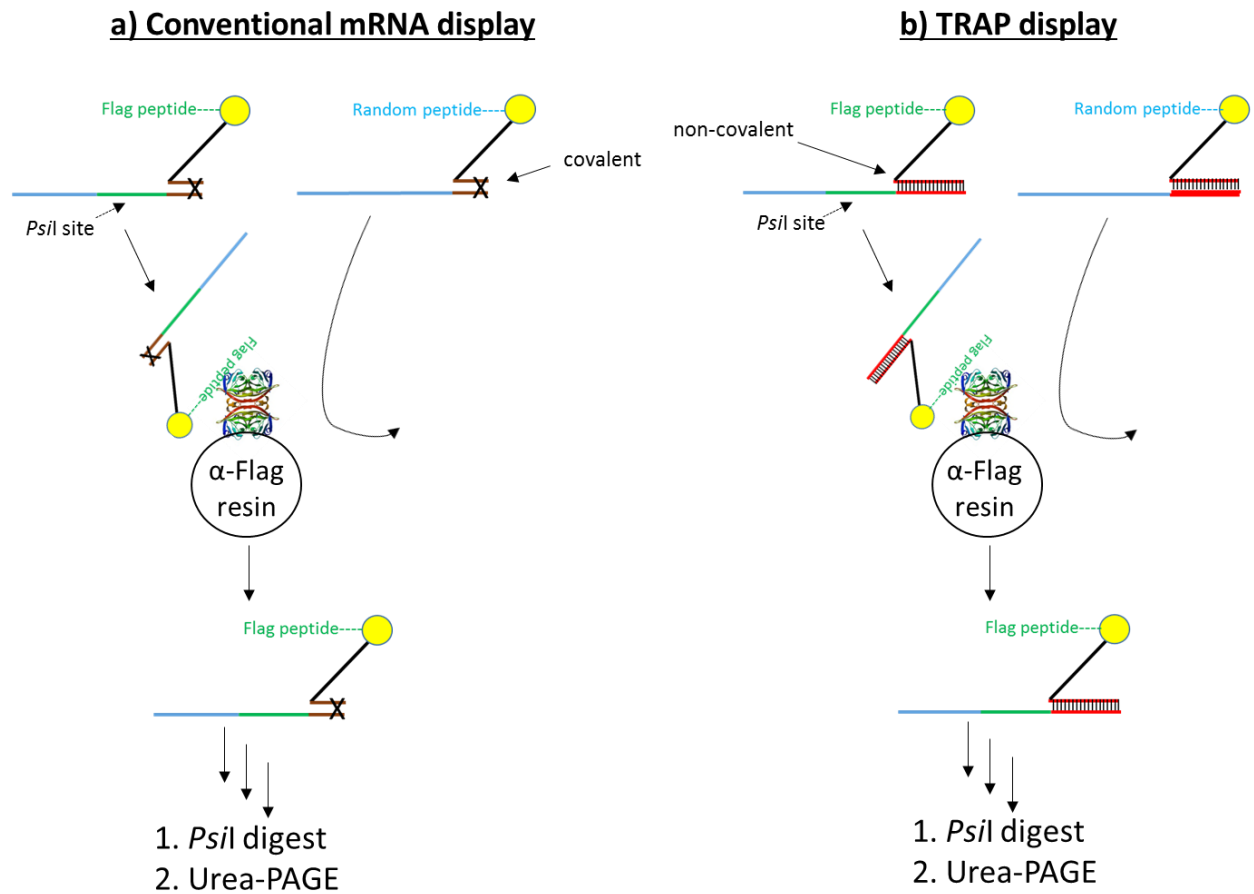


Figure B2 TRAP display validation scheme. mRNA-peptide fusions encoding for Flag peptide were mixed with a random fusion library and incubated with α -Flag resin for a) mRNA display (with covalently associated linker) and b) TRAP display (with hybridized linker). After one round of selection, functional fusions were reverse transcribed and PCR-amplified. cDNA was digested with *Psil* and analyzed by Urea-PAGE to determine the extent of Flag sequence enrichment for each method.

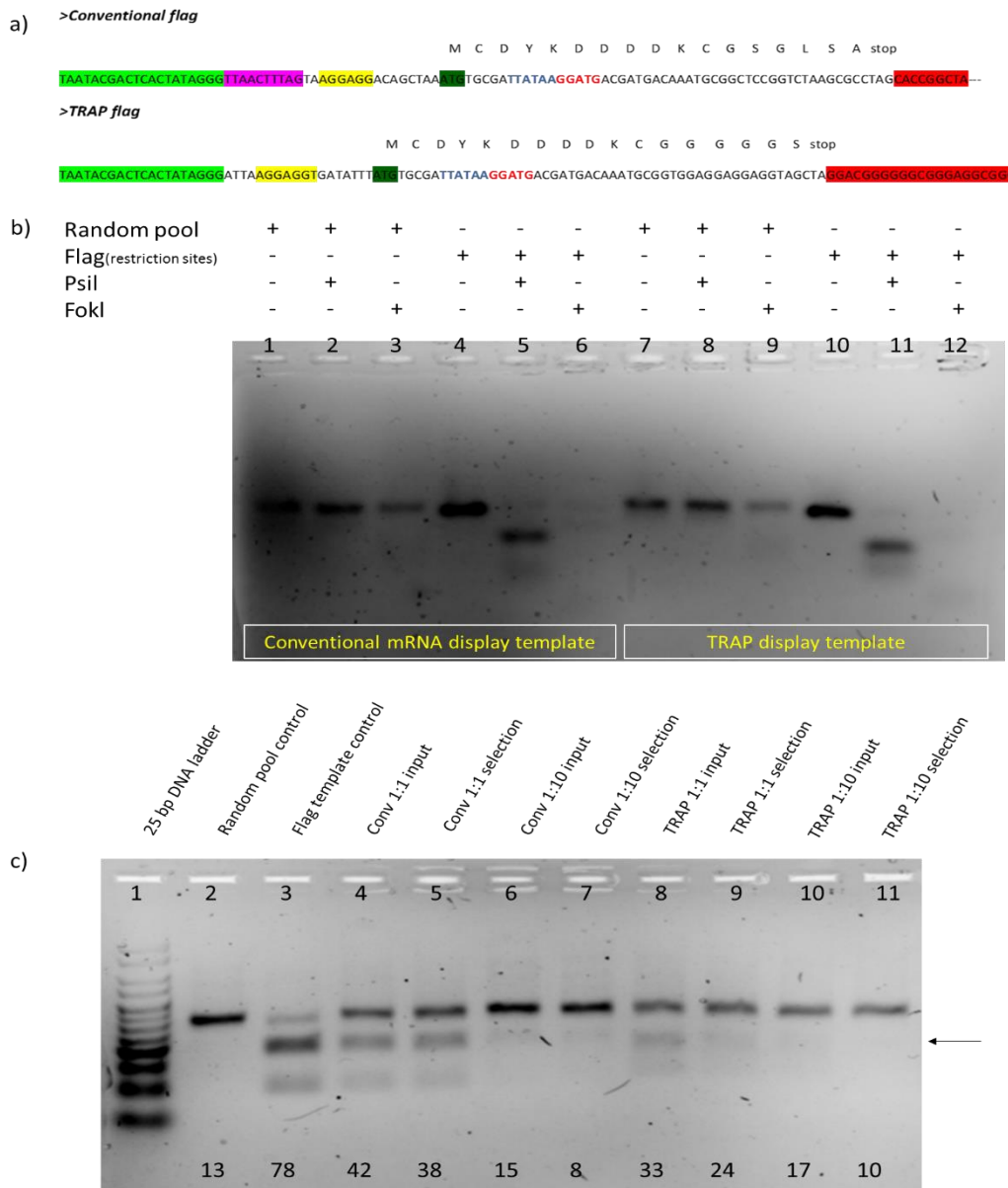


Figure B3 Comparative Flag-tag enrichment of TRAP and conventional selections. a) Flag sequences. Red box = hybrid regions, blue letters = *PsiI* site, red letters = *FokI* site, light green box = T7 promoter, purple = epsilon translation enhancer, yellow = RBS b) 4% agarose gel of control digest indicating selectivity of *PsiI* for Flag sequence in conventional (lane 5) and TRAP (lane 11) DNA. c) *PsiI*-digested samples analyzed on a 4% agarose gel. Conventional DNA (lanes 4-7) and TRAP DNA (lanes 8-11). Selected DNA was compared to the input DNA based on the percentage digested (noted at the bottom of the gel) and quantified by densitometry using the larger of the two fragments (black arrow).

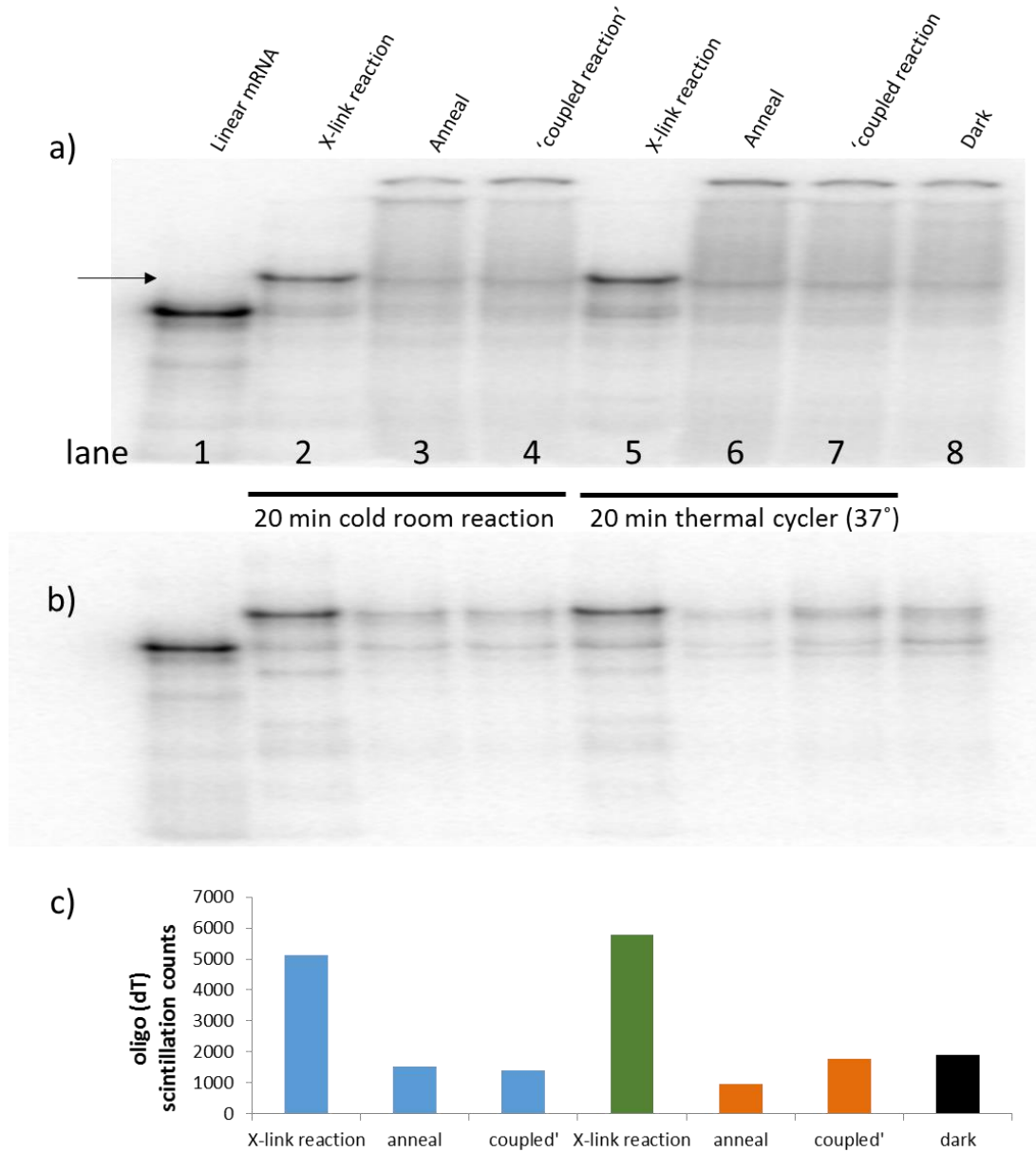


Figure B4 Modified mRNA display: *in situ* covalent association. a) Phosphorimage of 12% Urea-PAGE of translation reactions containing 5' ³²P-labeled linear mRNA. mRNA and linker were pre-annealed (lanes 3 and 6) or added separately to hybridize *in situ* (lanes 4 and 7). Reactions were performed at 4 °C (left) and 37 °C (right). Reactions were irradiated at 365 nm for the first 20 min. of the 60 min. reaction. Lanes 2 and 5 are a pre-crosslinked control. Lane 8 control reaction was protected from light throughout the experiment. Black arrow represents covalently attached mRNA-linker. b) The crude reactions from a) were purified on oligo(dT) beads prior to loading on the gel. c) Scintillation-based quantification of the crude reactions.

B3 Discussion

We showed that our restriction digest method for identifying Flag sequences among a library pool was functional based on the results of our control experiment. However, we were unable to enrich the Flag DNA after one round of selection against anti-Flag affinity resin using either TRAP display or even the conventional mRNA display which served as a positive control. We considered the potential causes for lack of Flag sequence enrichment, particularly in the covalently-linked conventional mRNA display control. While we did not spend a significant amount of time troubleshooting these experiments, we had difficulty detecting full length peptide via MALDI-TOF for both schemes. We frequently observed truncations of the peptide at and beyond the Flag sequence and there were typically abundant truncations prior to reaching the linker region. This stalling or truncation phenomenon was unresolved and may have negatively impacted fusion formation in both schemes. The lack of appropriate linker length may have prevented puromycin from reaching the A-site of the ribosome and attaching to the stalled peptide. Or it may have caused an issue with capture by the antibody resin (M2 resin is designed for capturing Flag epitopes at all regions of the peptide or protein). Any of these potential ribosome processivity issues could have resulted in lack of Flag enrichment. We also consider the possibility that PCR-amplification of the cDNA directly from the beads may have negatively affected Flag enrichment due to non-specific amplification of bead-binders.

Our attempt at creating covalently-linked mRNA and puromycin *in situ* by UV irradiation for the first 20 min of the translation reaction was unsuccessful. While there was a small amount of cross-linking detected on the denaturing gel in all of the reactions, the quantities were below the level of the control reaction that was kept in the dark. The bands and streaking observed near the top of the gel lanes in the crude reaction tests were absent in the gel of purified cross-linked

mRNA. Since oligo(dT) binds to the poly(dA) portion of the linker, we conclude that the bands and streaks are unrelated to cross-linked mRNA and could be a result of UV-induced damage, aggregation or non-specific covalent association.

B4 Future Directions

The time and labor savings associated with TRAP display make it an attractive method. While we feel the design to test the method was sound, we were unable to successfully validate TRAP. If the Flag-tag processivity and capture issues can be corrected, this scheme could potentially be successful. Otherwise, an alternative design would be necessary.

The lack of covalent *in situ* cross-linking in the modified mRNA display experiment, combined with the by-products likely associated with UV exposure makes the future success of this method unlikely.

B5 Experimental

oligos	sequences
conv-Flag	TAGCCGGTGCTAGGCACTTAGACCGGAGCCGCATTGTGCATCGTCATCCTTGTAATCGCA CATTAGCTGTCCTCCTTACTAAAGTTAACCCCTATAGTGAGTCGTATTA
conv-Library	TAGCCGGTGCTAGGCACTTAGACCGGAGCCGCASNNSNNSNNSNNSNNSNNSNNSN GCACATTTAGCTGTCCTCCTTACTAAAGTTAACCCCTATAGTGAGTCGTATTA
conv-FWD	TAATACGACTCACTATAGGGTTAACTTTAGTAAGGAGGACAGCTAA
conv-REV	TAGCCGGTGCTAGGCACTTAGACCGGAGCCGCA
conv-RT	TTTTTTTTTTTTTCTAGGCACTTAGACCGGAGCC
conv-puromycin linker	PsoC ₆ -(UAGCCGGUG) ₂ '-OMe-15xA-2xSpacer9-dAdCdC-Puro
TRAP-Flag	CCCGCCTCCCGCCCCCGTCCTAGCTACCTCCTCCTCCACCGCATTGTGCATCGTCGTCTT TATAATCGCACATAAATATCACCTCCTTAATCCCTATAGTGAGTCGTATTA
TRAP-Library	CCCGCCTCCCGCCCCCGTCCTAGCTACCTCCTCCTCCACCGCASNNSNNSNNSNNSN SNNSNNSNNGCACATAAATATCACCTCCTTAATCCCTATAGTGAGTCGTATTA
TRAP-FWD	TAATACGACTCACTATAGGGATTAAGGAGGTGATATTATGTGC
TRAP-REV	CCCGCCTCCCGCCCCCGTCCTAGCTACCTCCTCCTCCACC
TRAP-RT	TAGCTACCTCCTCCTCCACC
TRAP-puromycin linker	CCCGCCTCCCGCCCCCGTCC-(SPC18) ₅ -CC-Puro

Table B1 Oligonucleotides used for TRAP display validation

B5.1 Amino acids, reagents and enzymes: For natural amino acid preparation, see **2.5.1**. Preparation of *in vitro* translation components and reagents is described in **2.5.3**. Restriction enzymes *PsiI* (R0657S) and *FokI* (R0109S) were purchased from New England Biolabs (NEB). ATP, [γ -³²P]- 3000Ci/mmol 10mCi/ml (NEG002A250UC) was purchased from Perkin Elmer.

B5.2 Selection Reagents: Obtained or prepared as in **4.5.2**; M2 anti-Flag agarose (Sigma # A2220)

B5.3 Preparation of mRNA: T7-mediated bottom strand synthesis of the oligos (IDT) was performed as described in **4.5.3**.

B5.4 Preparation of conventional mRNA-peptide fusions: Fusions were prepared as in **4.5.3** with the following exceptions: Translation volume was 0.2 mL and fusion purification reagents were scaled accordingly; Cyclization steps were omitted.

B5.5 Preparation of TRAP mRNA-peptide fusions: Fusions were prepared as described by others,¹¹⁰ and adjusted for 0.2 mL scale.

B5.6 Selection against anti-Flag: One round of selection was performed with conventional fusions as described previously (**4.5.6**) and TRAP fusions as described by others,¹¹⁰ with the following exceptions: Purified conventional fusions from **B5.4** (200 μ L) and the crude translation reaction that contained the hybrid TRAP linker (**B5.5**) were each added to 20 μ L anti-Flag resin and rotated at 4 °C for 1 h; Bound TRAP fusion were washed 3x with 0.25 mL HBST (50 mM HEPES-KOH pH 7.5, 300 mM NaCl, 0.05% Tween-20); Reverse transcription was performed on the beads for TRAP fusions; PCR for both reactions (2 min. at 94°C, followed by 20 rounds of 94° C (30s), 65° C (30s), 72° C (45s)) was also performed on the beads using the appropriate primers from **Table B1**. The DNA was extracted with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with 3 vol. ethanol and 0.1 vol 3 M KOAc pH 5.2.

B5.7 Control digest with *FokI* and *PsiI*: Restriction digests were performed per manufacturer's instructions (New England Biolabs). Reactions were analyzed on a 4% agarose gel and visualized on a Chemi-Doc imager (BioRad).

B5.8 Digestion of the selection DNA with *PsiI*: DNA from **B5.6** was resuspended in 50 μ L of water and digested as in **B5.7** with *PsiI*. Reactions were analyzed on a 4% agarose gel and

visualized on a Chemi-Doc imager (BioRad). Densitometry was performed using the larger of the two digest fragments as a relative quantification of the Flag sequence in each reaction.

B5.9 Modified mRNA display: Conventional mRNA was 5' end labeled with [γ - 32 P] ATP as described previously.¹⁹⁵ Standard 100 μ L translation reactions were prepared (2.5.4) and initiated with the following three template preparations: Covalently cross-linked mRNA (4.5.3); mRNA pre-annealed with the linker (2.5.2); and mRNA and linker added independently to simulate *in situ* hybridization. Reactions were done in 96-well plates and allowed to proceed for 1 h (at 4 °C and 37 °C). For the first 20 minutes of the reaction, a UV (365 nm) handheld lamp was placed 5 cm above the plate for the light reactions. Half of each reaction was purified on oligo(dT) resin as described previously (4.5.3) with the following exceptions: Purification reagents were scaled to 50 μ L translation volumes; Cyclization steps were omitted and eluted cross-linked mRNA was not precipitated. The crude reactions and oligo(dT) elutions were analyzed by 10% SDS-PAGE, and phosphorimaged using a Typhoon 9410 Variable Mode Imager. 1 μ L of each reaction was added to 2 mL of scintillation fluid (EconoSafe) and counted on a Beckman Coulter scintillation counter for 5 min.

Appendix C HiStrep: Dual Affinity Epitope Tag

C1 Introduction

We isolated the peptide MTIHQWLYHPQVWRW serendipitously from a peptide library during an *in vitro* selection against streptavidin. Unlike the other peptides in the library, this peptide had a mutation in its encoding sequence so that it did not contain a hexahistidine tag (**Figure C1a**). Yet somehow it survived the selection process which included a Ni-NTA purification step. However, the peptide does have two histidines and two glutamines which likely enabled binding to the nickel resin. The HPQ motif in the sequence has been commonly identified in selections against streptavidin.^{165,172-174}

Most protein purification methods are based on a single step, which oftentimes does not result in protein of high purity. Our peptide (HiStrep) was found to bind both nickel (since it survived mRNA-peptide fusion purification) and streptavidin (it was highly enriched during selection) and, along with its short sequence, makes its possible application as a dual affinity tag intriguing. Dual tag approaches designed to increase protein purity often require long sequential tags which may impede protein function in downstream assays. However, HiStrep-tag is interdigitated, with the amino acids required for binding existing within the same short peptide. Given these advantages, we chose to investigate the ability of HiStrep to act as a dual affinity tag (**Figure C1b**).

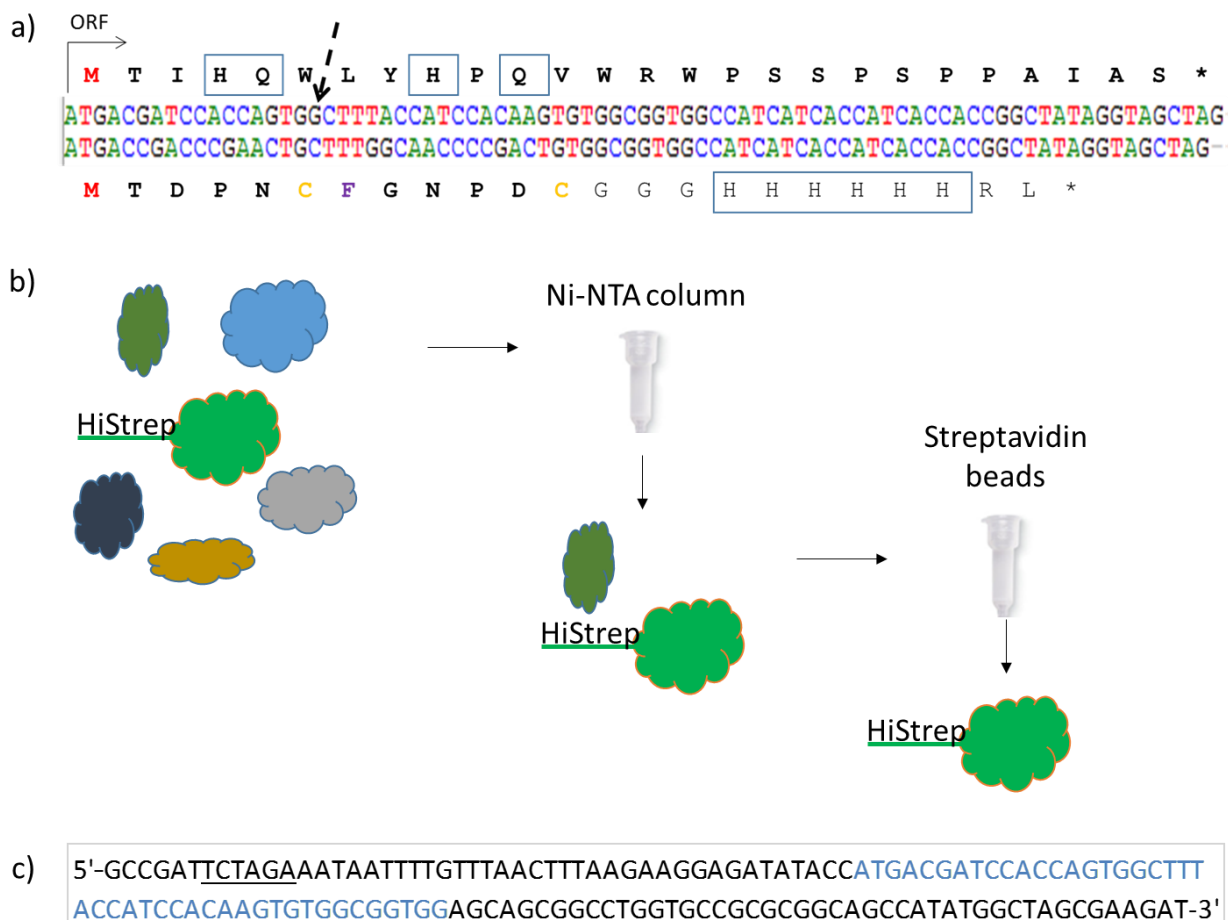


Figure C1 HiStrep-tag discovery and scheme for dual affinity purification. a) Library cDNA of HiStrep mutant (top sequence) discovered during an *in vitro* selection against streptavidin. Bottom sequence represents a selection winner that contains the designed library elements. Dotted arrow represents a gain-of-function insertion. Blue boxes indicate residues that bind to nickel resin. b) Scheme for dual affinity purification with HiStrep-tag, c) Gene fragment encoding the HiStrep epitope. The sequence was cloned into the sfGFP-pRSET vector digested with *Xba*I and *Nhe*I and replaced the N-terminal His-tag sequence.

C2 Results

C2.1 Ni-NTA affinity

We first inserted the HiStrep sequence onto the N-terminus of sfGFP (**Figure C1c**) and compared the ability of HiStrep-sfGFP to bind to Ni-NTA resin relative to the N-terminal hexahistidine-tagged sfGFP (His₆-sfGFP). Our initial purification protocol included 10 mM imidazole in the binding step and 20 mM imidazole in the 15 CV wash step. Under these typical conditions, we found that almost 80% of the total HiStrep-sfGFP was stripped from the resin prior to the imidazole step gradient elutions, implying that HiStrep had a lower affinity than His₆ for the nickel resin. We next performed an identical purification with the exception of omitting imidazole from the binding and wash steps. While His₆-sfGFP was eluted at higher imidazole concentrations (**Figure C2a**), HiStrep-sfGFP was retained on the resin following the column washes (**Figure C2b**) and ~70% of the total expressed protein was eluted between 50 and 100 mM imidazole (**Figure C2c**).

C2.2 Streptavidin affinity

Dual affinity purification can often be laborious and time-consuming. Ideally, the protein eluted from one affinity column is added directly to the second column, without the need for the removal of interfering reagents. To test the ability of HiStrep-sfGFP to bind to streptavidin in the presence of imidazole (100 mM) used to elute the proteins from nickel, we dialyzed a sample of HiStrep-sfGFP and tested the comparative ability to bind immobilized streptavidin. The undialyzed sample yielded a similar amount of HiStrep-sfGFP after elution with biotin (**Figure C2d-Lane 5**), relative to the dialyzed HiStrep-sfGFP (**Figure C2d-Lane 7**) suggesting that the purification methods could be used sequentially. Lastly, we determined the affinity of HiStrep for streptavidin. We titrated streptavidin against *in vitro* translated ³⁵S-Met-labeled HiStrep fused to mRNA and determined the K_d to be 840 nM (**Figure C2e**).

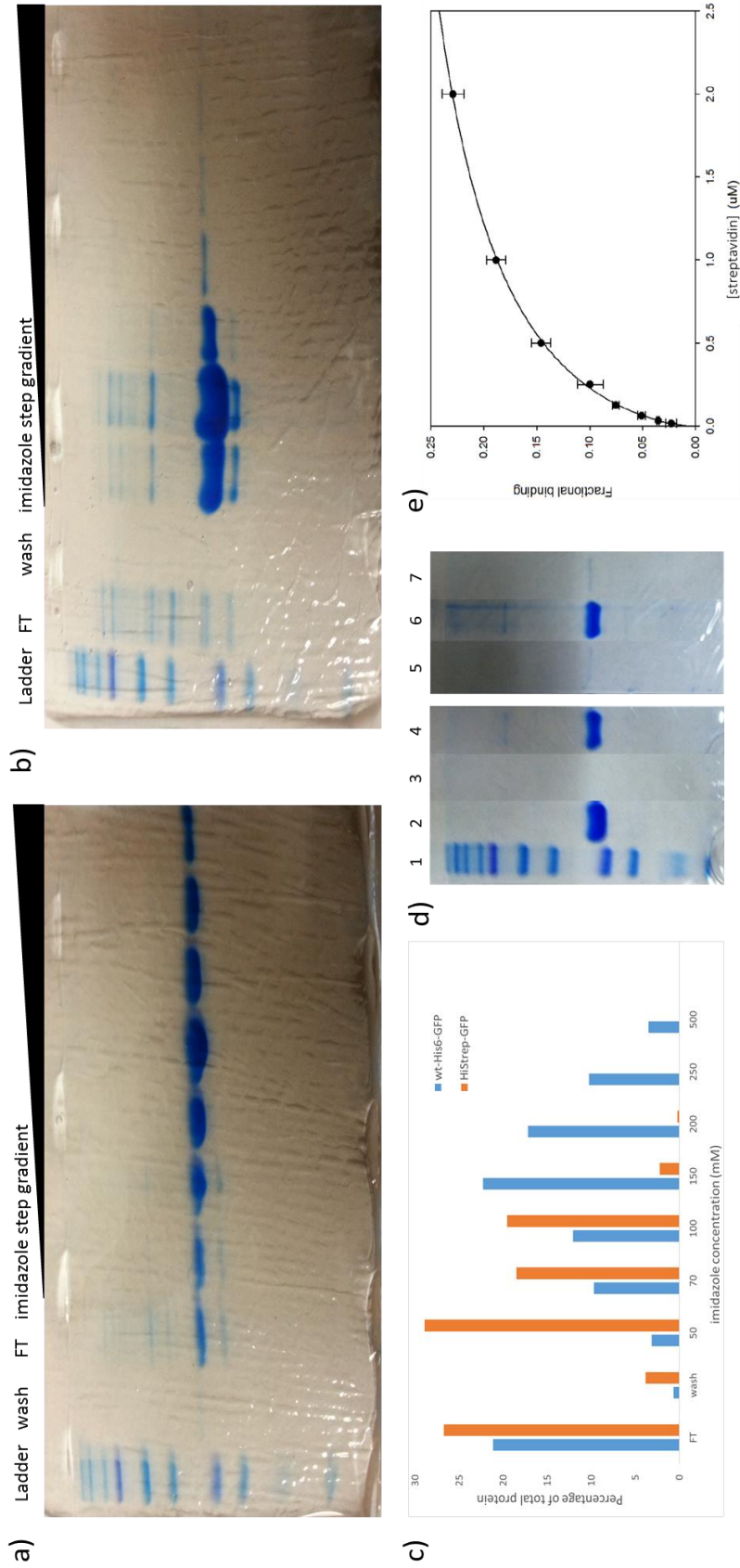


Figure C2 Dual affinity purification of HiStrep-sfGFP. 10% SDS-PAGE of Ni-NTA batch purification of a) His6-sfGFP and b) HiStrep-sfGFP. Binding and wash steps were performed in the absence of imidazole and eluted with an imidazole gradient. c) Densitometric analysis of the gels in a and b. Elutions were performed with an imidazole step gradient (50-500 mM-1 CV each). d) Proteins eluted from Ni-NTA (100mM imidazole in 50mM phosphate buffer pH 8, 500mM NaCl) were incubated for 1h at 4°C with 20µL of streptavidin magnetic beads and subsequently eluted after 1h RT incubation with 2.5mM biotin in PBS. Lane 1: protein ladder, Lanes 2,3: His6-sfGFP input (dialyzed against PBS), biotin elution, Lane 4,5: undialyzed HiStrep-sfGFP input, biotin elution, Lanes 6,7: dialyzed HiStrep-sfGFP input, biotin elution. The indicated experiments were performed simultaneously and the pertinent lanes from two 12% SDS-PAGE gels run concurrently are shown. e) Binding curve derived from ³⁵S-Cys labeled HiStrep mRNA-peptide fusion titrated with streptavidin. Data points were fitted to a dynamic curve using SigmaPlot and the EC₅₀ was determined to be 840 ± 280 nM. The experiment was done in duplicate.

C3 Discussion

We demonstrated the ability of HiStrep to bind to nickel and optimized the purification conditions to maximize the retention of HiStrep-sfGFP on the resin prior to elution. We also showed that HiStrep-sfGFP binds to streptavidin, and the presence of 100 mM imidazole did not interfere with the capture by streptavidin beads. This would streamline the dual purification, since the imidazole would not need to be removed prior to binding to streptavidin.

While HiStrep binds to streptavidin, the intensity of the bands on SDS-PAGE was low (Figure C2e, Lanes 5 and 7). We calculated that a maximum of 0.35 μ g of biotin-eluted HiStrep-sfGFP was loaded on SDS-PAGE, based on biotin binding capacity of the beads used in the assay and assuming 100% ligand binding and elution. This quantity is close to the limit of detection of coomassie staining (0.1-0.5 ng).¹⁹⁶

HiStrep (fused to mRNA) showed a high affinity for streptavidin (840 nM). In our experience, peptides fused to mRNA typically show weaker affinity than peptides alone, so we'd expect the actual affinity to be higher.

Aside from the omission of imidazole from the Ni-NTA purification of HiStrep-sfGFP, we did not optimize conditions for binding. We considered that sequence optimization, likely via an alanine scan, would give insight into the residues that were important for binding to either of the two affinity resins. We could then use this information to rationally design an improved HiStrep sequence. Alternatively, the residues not required for binding could be randomized in an affinity maturity selection designed with both affinity resins as targets. Although we did not pursue this project further, we believe that HiStrep, if optimized, has the potential to serve as a practical, short, interdigitated affinity tag used to obtain proteins of high purity.

C4 Future Directions

Further optimization of HiStrep is necessary to enable its use as a dual affinity tag. Sequence optimization, through rational or irrational design, could lead to improved affinity for one or both resins, although the low concentration of imidazole needed to elute it from the nickel column may be advantageous since higher imidazole may negatively impact binding to streptavidin.

Binding optimization would likely increase affinity, particularly to streptavidin. We also considered that the sequence of the dual purification steps could impact recovery and/or purity. Since immobilized streptavidin is costly, we chose to purify with Ni-NTA first, which would clear most of the non-specific proteins and presumably increase capture on streptavidin beads. However, reversing the sequence could lead to higher yield and/or purity.

C5 Experimental

C5.1 Cloning of HiStrep-sfGFP

An extended gene block fragment containing the HiStrep sequence (**Figure C1c**) and pRSET-sfGFP plasmid were each doubly digested with *Xba*I and *Nhe*I, which removed the N-terminal hexahistidine tag from the plasmid. The HiStrep gene fragment was then ligated into the vector, forming HiStrep-tagged sfGFP.

C5.2 Expression of HiStrep-sfGFP and His₆-sfGFP

The pRSET-sfGFP vectors that contained the N-terminal HiStrep or His₆ tags were transformed into *E. Coli* strain BL-21(DE3) cells and grown overnight in LB-media. 100 mL of ZYM-5052

auto-inducing media was inoculated 1:200 from the starter culture and allowed to incubate for 24 h at 18 °C. Cells were harvested by centrifugation at 4000 rpm.

C5.3 Ni-NTA purification

Harvested cells containing overexpressed HiStrep-sfGFP or His₆-sfGFP were treated similarly. The cultures were lysed with 5 mL of B-per (ThermoFisher #78248) and the cleared lysate, supplemented with 10 mM imidazole, was added to 1 mL of Ni-NTA resin (MCLAB) in a 10 mL disposable column (Bio-Rad) and rotated at 4 °C for 1 h. The column was drained and washed three times with 5 CV of Ni-NTA wash buffer (all buffers prepared per MCLAB) containing 20 mM imidazole. Bound protein was eluted with an imidazole step gradient in 1 CV each (50, 100, 150, 200, 250, 500 mM imidazole). The elutions were analyzed by 12% SDS-PAGE and quantified by densitometry using a ChemiDoc imager (BioRad). For the optimized purifications, all steps were identical except imidazole was omitted from the binding and wash steps. Half of each protein eluted with 100 mM imidazole was dialyzed overnight against manufacturer's recommended streptavidin binding buffer A (20 mM potassium phosphate, 150 mM NaCl, pH 7.5).

C5.4 Streptavidin purification

20 µL of magnetic streptavidin beads (Roche #11641778001) was added to each of three 1.7 mL tubes. Beads were washed three times with 0.5 mL buffer A. 50 µL each of undialyzed His₆-sfGFP, undialyzed HiStrep-sfGFP and dialyzed HiStrep-sfGFP were added to the beads along with 50 µL of buffer A and rotated at 4 °C for 2 h. Beads were washed three times with 0.2 mL of buffer A, 50 µL of buffer A containing 2 mM D-biotin was added and the tubes were rotated at 4 °C for 1.5 h. The purifications steps were analyzed by 12% SDS-PAGE and visualized using a Chemi-Doc imager (BioRad).

C5.5 Affinity of HiStrep-mRNA for immobilized streptavidin

HiStrep mRNA was obtained via *in vitro* transcription as previously described (2.5.2) from a streptavidin selection clone described in Chapter 4. Radioisotope-labeled HiStrep-mRNA fusions were created as previously described (4.5.3) with the following exceptions: Translation volume was 0.2 mL and purification reagents were scaled accordingly; all natural amino acids were used; 0.11 μCi ^{35}S -Met was used to radiolabel the peptide; and both on-resin cyclization reactions were omitted. The pelleted fusions were resuspended in 100 μL of selection buffer. The binding assay was performed and analyzed as described previously (4.5.8) with the following exceptions: Streptavidin beads were diluted 2-fold (0.17-2.0 μM); radiolabeled HiStrep-mRNA concentration was 5 nM.

Appendix D Coupled *In vitro* Transcription/Translation

D1 Introduction

In prokaryotic organisms, mRNA transcription and protein translation are coupled. Binding of the ribosome to the mRNA typically occurs while the RNA polymerase is synthesizing the RNA from the DNA template. The PURE system, an *E. coli*-based cell-free technology, is designed to synthesize peptides or proteins with reconstituted components.⁶ To initiate the PURE reaction, a messenger RNA (mRNA) template encoding the gene of interest is added. Since protein expression is coupled in *E. coli*, DNA can also serve as the template. In this case, RNA polymerase can be added to transcribe the DNA *in situ*, followed directly by ribosome binding and translation of the encoded peptide or protein.

Many *in vitro* systems utilize DNA as the template to produce peptides for various purposes (such as TRAP display-Appendix B). Our lab uses mRNA to produce peptides as well as peptide libraries for *in vitro* selection. However, the synthesis and purification of mRNA is laborious and time-consuming. Herein, we describe our efforts to adapt and optimize our version of the PURE system for initiation of *in vitro* protein and peptide coupled synthesis using template DNA.

D2 Results

To adapt PURE translation to coupled transcription-translation, we supplemented the reaction with the necessary reagents (magnesium and NTPs for mRNA synthesis),¹⁹⁷ added 0.1 μ M T7 RNA polymerase and initiated a 50 μ L coupled reaction with 2 pmols of His₆-sfGFP-pRSET plasmid. After 1 h incubation at 37 °C and Ni-NTA purification, we were unable to detect sfGFP, either by fluorescence emission (**Figure D1a**) or SDS-PAGE (**Figure D1b**). We analyzed the components of the system, and the SDS-PAGE analysis of the translation factor mix used in

the experiment was lacking RF-1 (**Figure D1c-lane 2**). Both RF-1 and RF-2 recognize the UAA codon found in the sfGFP mRNA and catalyze the hydrolysis of the peptidyl-tRNA peptide bond, resulting in the release of the protein from the ribosome. Since we do not add RF-2 to our translations, the presence of RF-1 is critical for protein production. We then prepared a new mixture of 9 of the 10 translation factors (EF-Tu is added separately) and analyzed the mix using SDS-PAGE. This confirmed the presence of RF-1 in our new mix. We then attempted to express sfGFP again, this time using the newly created factor mix containing RF-1. Again, we were unable to detect sfGFP emission from the reaction (**Figure D1d**).

Having been unsuccessful in our effort to express sfGFP using the plasmid to initiate the coupled reaction, we tried to use DNA that encoded a small peptide. To do this we used one of our small translation templates for incorporating unnatural amino acids in translation, MDYKMH₆. We used MDYKMH₆ mRNA for translation, pET12b containing the MDYKMH₆ insert and linearized MDYKMH₆ DNA (**2.5.2**) in a side-by-side comparison. We determined peptide yield by the incorporation of ³⁵S-methionine and found that both forms of DNA used to initiate the coupled reaction produced similar amounts of peptide as the translation reaction initiated by the addition of mRNA, with addition of linear DNA yielding 23% more peptide than the translation alone (**Figure D2a**). The reactions performed in the presence of methionine resulted in fidelity of the coupled reaction products that were similar to that of the peptide produced by the translation reaction (**Figure D2b-d**).

We then performed further optimizations of the coupled reaction using the MDYKMH₆ plasmid. We tested the components that were either supplemented to our translation components (magnesium ion), or added specifically for synthesis of the transcript (plasmid and T7 RNA polymerase). The addition of 9 mM magnesium acetate (**Figure D3a**), 12 nM MDYKMH₆ plasmid

(Figure D3b) and 1 μ M T7 RNA polymerase **(Figure D3c)** resulted in the highest peptide yield from the coupled transcription-translation reactions. However, even with these additional optimizations, we were still unable to express sfGFP using the plasmid to initiate the coupled reaction.

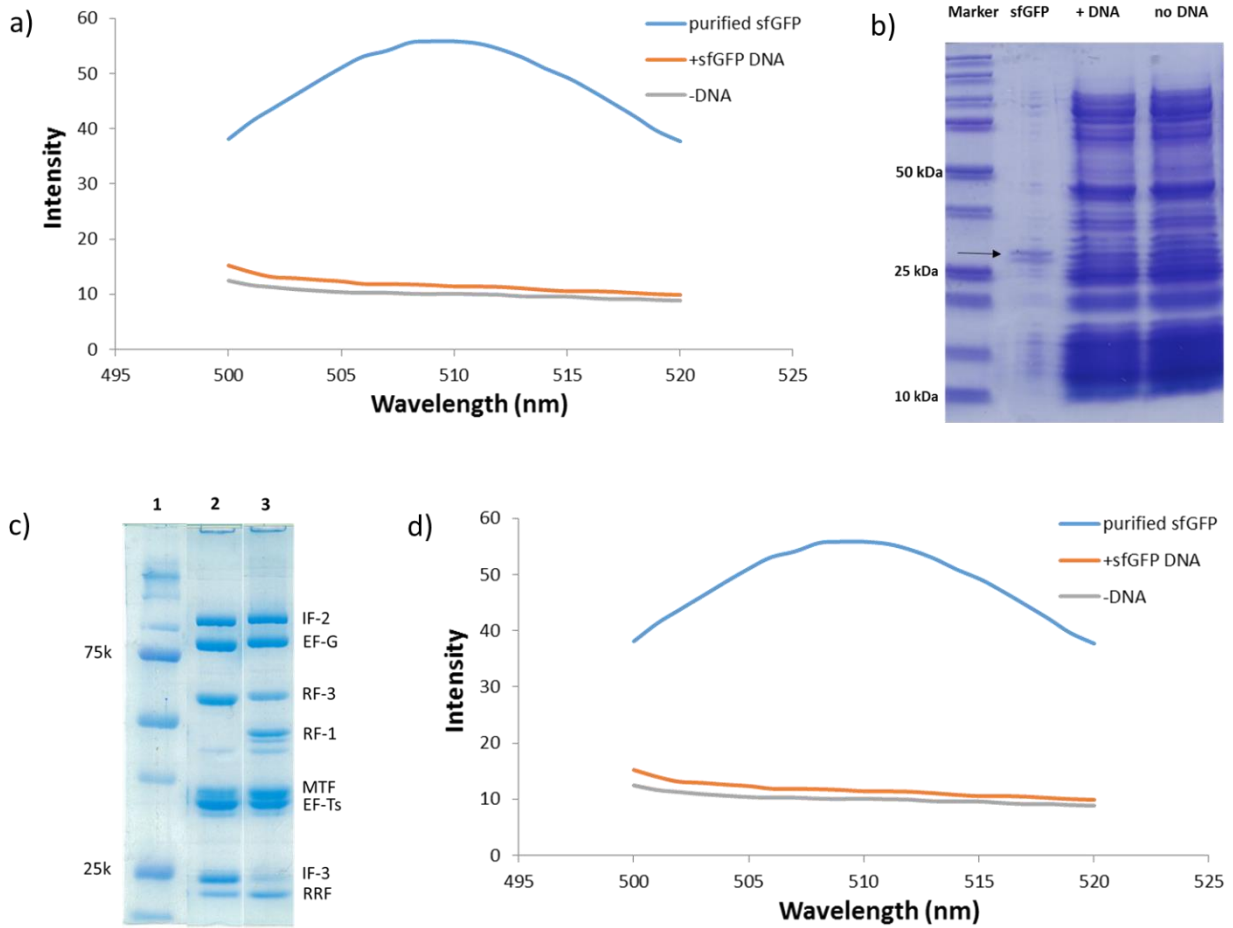


Figure D1 Coupled transcription-translation reactions with sfGFP plasmid DNA. The results of the initial 50 μ L coupled reaction initiated with sfGFP-pRSET characterized by a) fluorescence emission of sfGFP and b) 10% SDS-PAGE, sfGFP lane represents the purified protein control c) 10% SDS-PAGE analysis of 9 translation factors (IF-1 ran off the bottom of the gel) that were used in the initial experiment (lane 2) and the newly prepared mix containing RF-1 (lane 3). d) Fluorescence spectrum of the reaction containing RF-1.

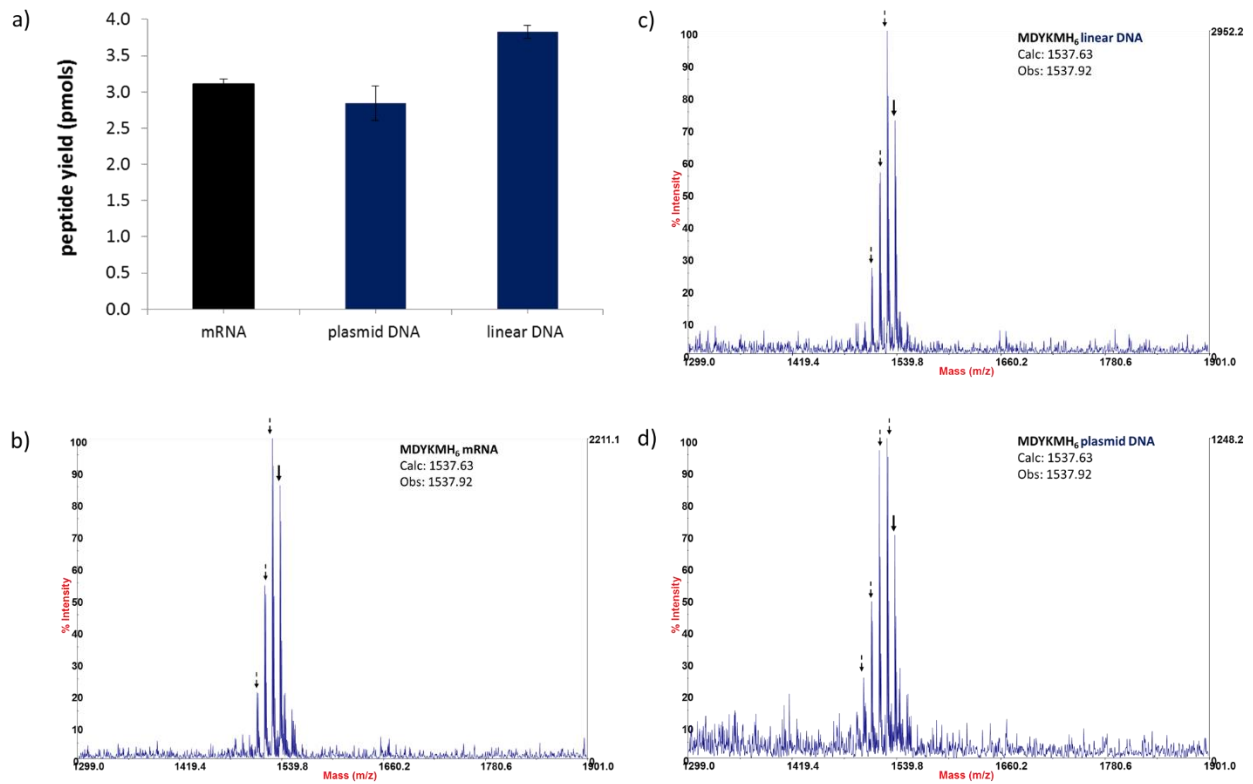


Figure D2 Comparison of translation and coupled reactions. a) 50 μ L reactions were initiated by the coding sequence MDYKMH₆ in the form of mRNA for translation and plasmid DNA or linear DNA for coupled reaction. All reactions contained ³⁵S-Met for quantification of peptide incorporation. MALDI spectra of the reactions containing methionine and initiated by b) mRNA in translation; and coupled reactions with c) linear DNA and d) plasmid DNA.

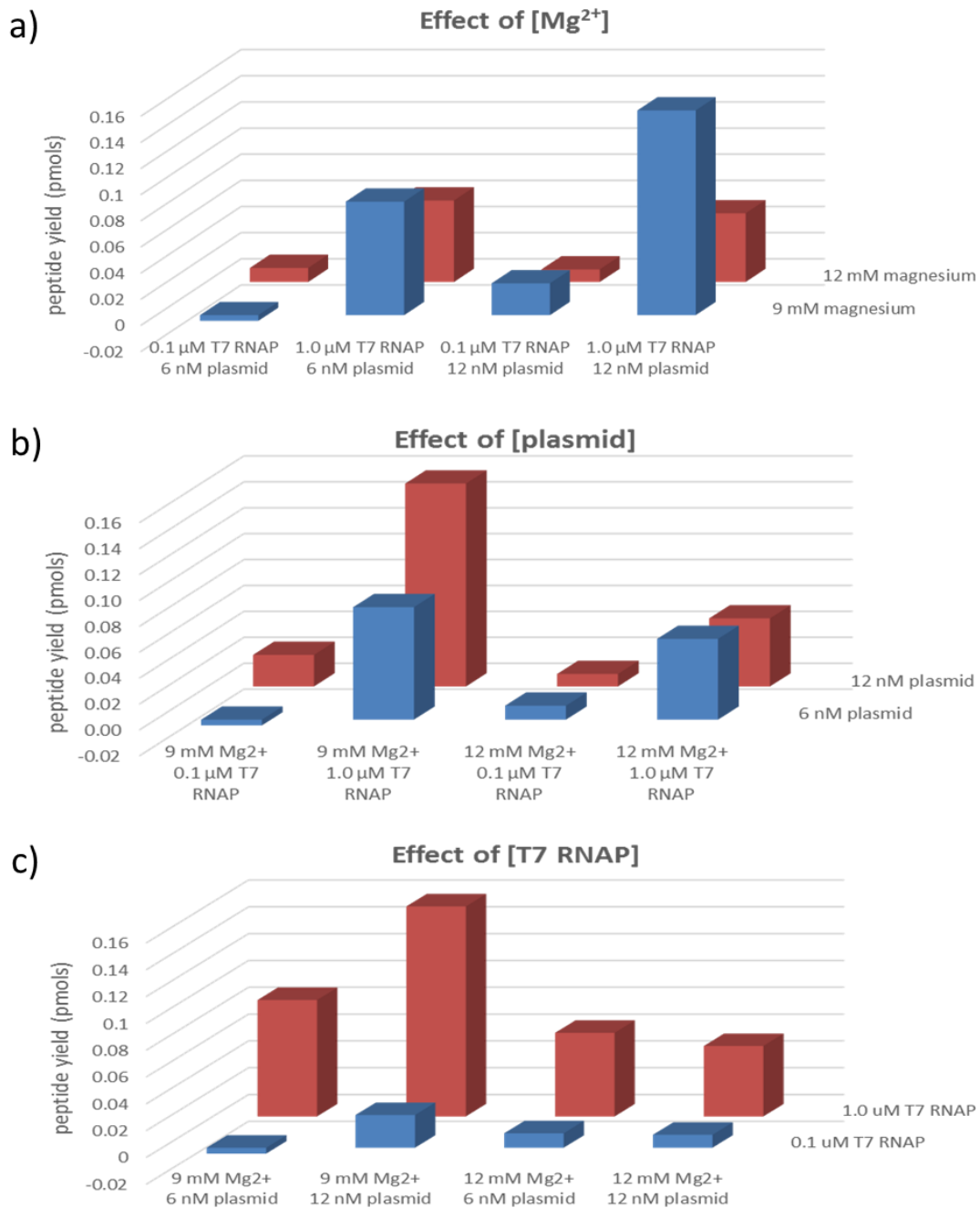


Figure D3 Optimization of coupled transcription-translation. Coupled reactions were initiated by the addition of MDYKMH₆ plasmid and quantified by ³⁵S-Met. The effects of a) magnesium acetate concentration, b) plasmid concentration, and c) T7 RNA polymerase concentration on radiolabeled yield of MDYKMH₆ peptide.

D3 Discussion

We were unable to synthesize sfGFP with our PURE coupled transcription-translation reaction initiated with plasmid DNA, despite our efforts to troubleshoot and optimize the reaction. While the reason is unknown, we speculate as to the possible causes: The PURE system contains over 80 individually prepared components that are mixed together to perform each reaction. While we painstakingly purify, prepare, store and ultimately mix these components together, the possibility certainly exists that one or more of the components are of insufficient purity or imbalanced in the reaction (or even inadvertently omitted as in our translation factor mix lacking RF-1). These problems could potentially lead to stalling, truncation or lack of fidelity that could adversely affect protein production and/or function.

We were able to process through a short template in coupled reactions, using plasmid and linear DNA encoding MDYKMH₆ in these experiments. Both yield and fidelity of the peptide produced in the coupled reaction compared favorably to that of the peptide produced in a translation reaction. We also have performed numerous translations with this and several other mRNA templates (shown in Chapter 2), typically with sufficient yield and fidelity. But these templates are short and each contains only a small percentage of the proteinogenic amino acids and hence is not an accurate gauge of overall system fidelity and ribosomal processivity. That being the case, we still observed instances of natural amino acid misincorporation in these short sequences (eg. R→W in MHFSW-Flag, and E→D, Q→H in the epitope tags of the templates). This phenomenon observed in synthesis of short templates would be amplified in proteins that contain 5-fold or more amino acids, potentially leading to lack of detectable protein.

D4 Future Directions

The ribosomal synthesis of proteins will require additional troubleshooting and optimization. A more effective tool to monitor yield in real-time, such as the Lanthanide Binding Tag (**Appendix A**) would permit simpler, faster and more comprehensive analysis (and subsequent optimization) of PURE system components.

D5 Experimental

D5.1 Amino acids

Only natural amino acids were used in the experiments. Preparation information can be found in **2.5.1**

D5.2 mRNA preparation

In vitro transcription and purification of MDYKMH₆ mRNA was performed as previously described (**2.5.2**).

D5.3 Enzymes and ribosome purification (see 2.5.3)

D5.4 *In vitro* translation

50 μ L reactions were performed and peptides were purified and analyzed for yield and/or fidelity as described previously (**2.5.4**).

D5.5 *in vitro* coupled transcription-translation

50 μ L reactions containing standard translation reagents (**2.5.4**) were supplemented with magnesium acetate (9 mM final concentration-unless varied), 1 μ M T7 RNA polymerase (unless varied), 1 mM each of ATP, GTP, UTP, CTP and 12 nM plasmid or linear DNA (unless varied). The peptides were purified and analyzed for yield and/or fidelity as described in **2.5.4**

Appendix E Optimization of *in vitro* Translation

E1 Introduction The PURE translation system contains over 80 individual components that are prepared individually prior to combining them in an *in vitro* translation reaction to create peptides in the presence of the ribosomal machinery. There is an understandably high degree of synergy involved among the components and even a small disparity between actual and expected activity, or calculated vs. effective concentration can lead to a decrease in peptide yield and/or fidelity of translation. The following figures represent the effects of varying the translation input of multiple reagents (>50), from small molecules and buffers to recombinant proteins and isolated ribozymes. These results were obtained during either reagent preparation or PURE system troubleshooting.

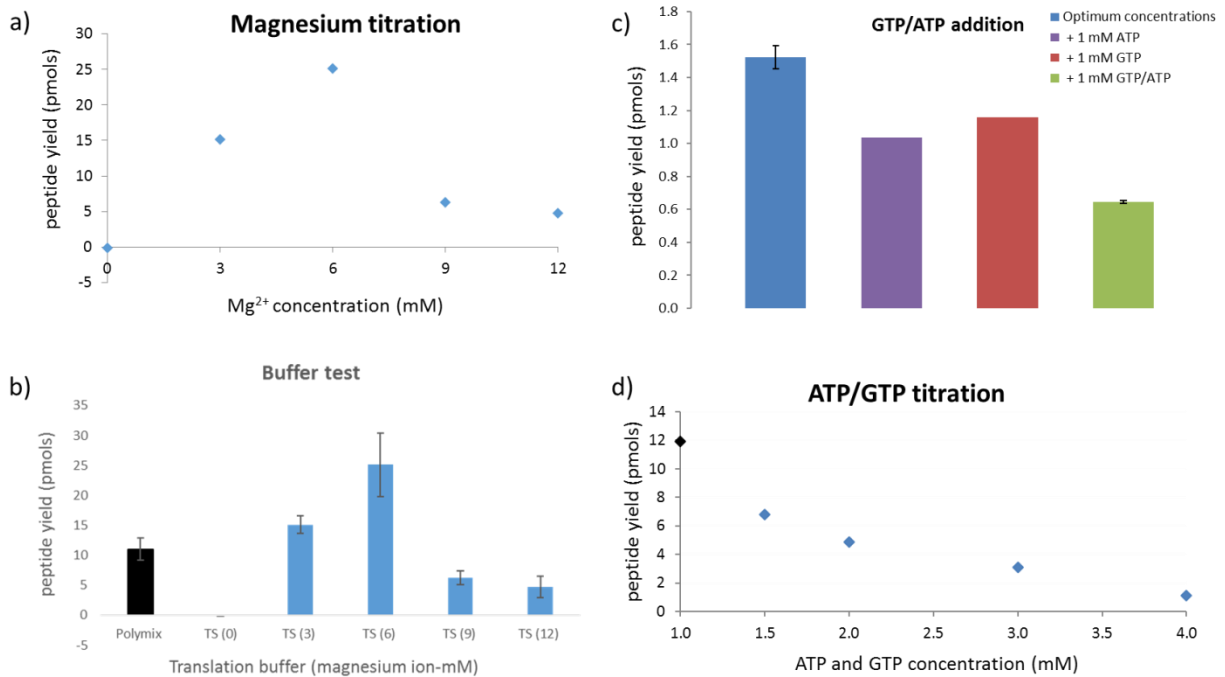


Figure E1 Buffers and small molecules. Optimization of translation components in 50 μ L translation reactions. a) Magnesium acetate titration; b) Comparative buffer test with Polymix (needs to be prepared fresh) against a mix of tRNA and small molecules (TS) that is prepared in bulk and stored in aliquots at -80 $^{\circ}$ C until ready for use (magnesium concentration is indicated parenthetically); c) Addition of 1 mM ATP and/or GTP to the optimum translation concentrations (1 mM each); d) Titration of ATP/GTP. X-axis represents mM each added to translation (black point represents the concentration of each of ATP/GTP in a standard translation).

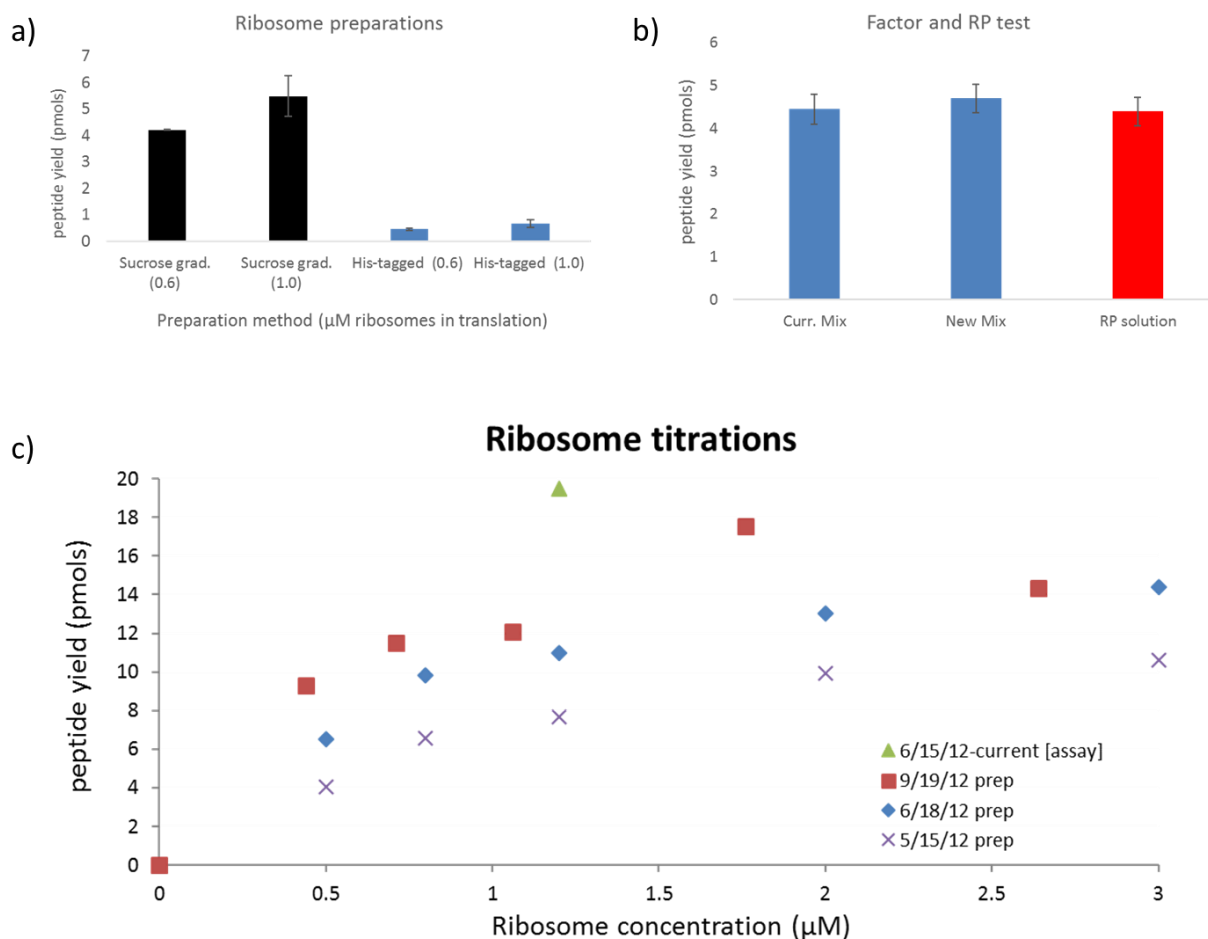


Figure E2 Ribosomes and RP mix. a) Ribosomes prepared using blender lysis and sucrose gradient centrifugation (2.5.3) or using a His-tagged ribosomal protein purification protocol¹⁹⁸ were added to translation reactions; b) Relative translation yields of two preparations of PURE translation factor mixes (current and new). The new mix of factors was used to make a mix containing all factors, energy recycling enzymes and ribosomes (ribosomes and essential protein factors or RP solution).¹⁹⁷ The RP mix was flash frozen and stored at $-80\text{ }^{\circ}\text{C}$. The RP solution was then thawed and used in translation; c) Titrations of three preparations of ribosomes using the sucrose gradient protocol. The ‘current’ ribosomes were added at the standard translation concentration ($1.2\text{ }\mu\text{M}$) as a point of comparison.

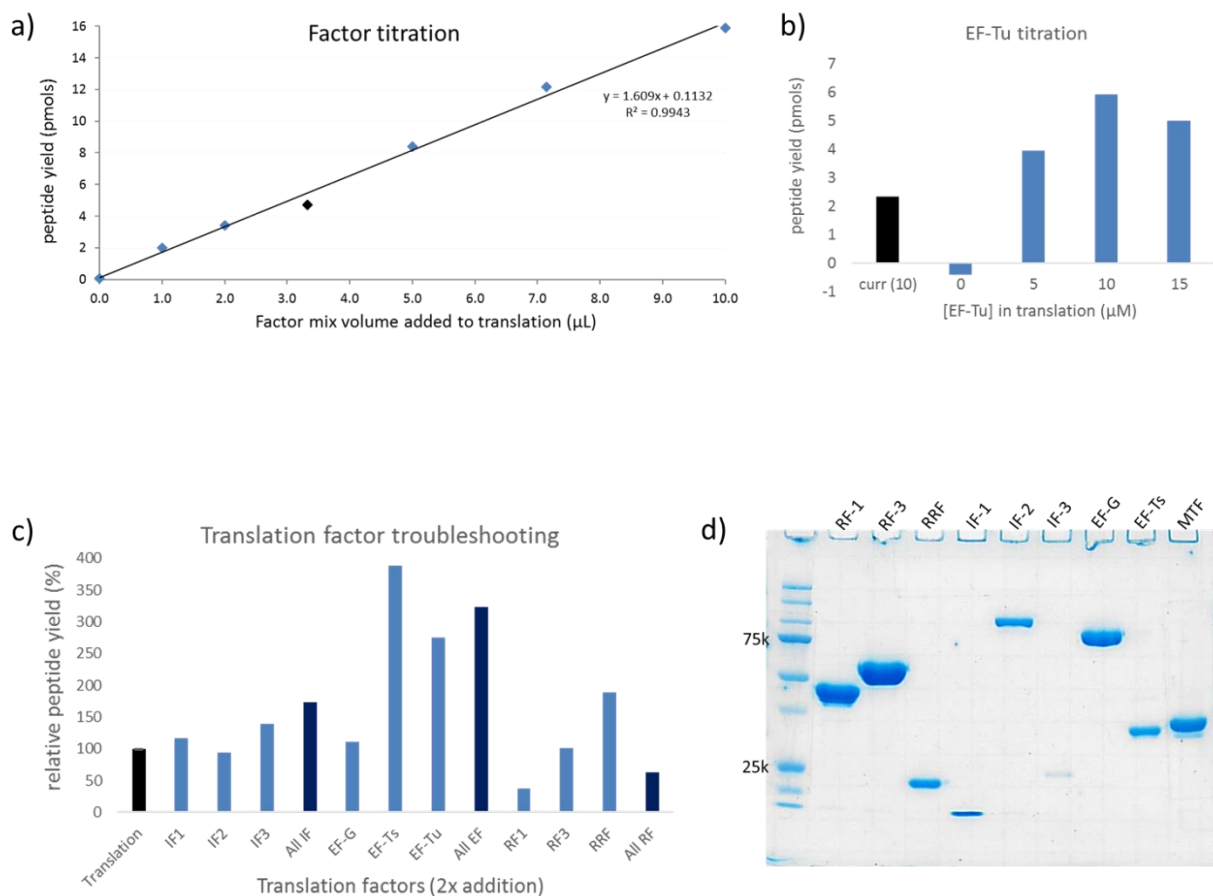


Figure E3 PURE Translation Protein Factors. a) Radiolabeled peptide yield as a function of factor concentration in translation (black point represents the concentration in standard translation). b) Effect of EF-Tu concentration on peptide yield relative to a previous preparation (current). Standard concentration of EF-Tu is 10 μM . c) Effect of individual supplementation at 2x concentration of each of the translation factors or subset of factors (all IF, all EF or all RF) on peptide yield relative to a standard reaction (translation). d) 10% SDS-PAGE of a typical PURE translation factor purification. Left lane is protein ladder (BioRad #1610374).

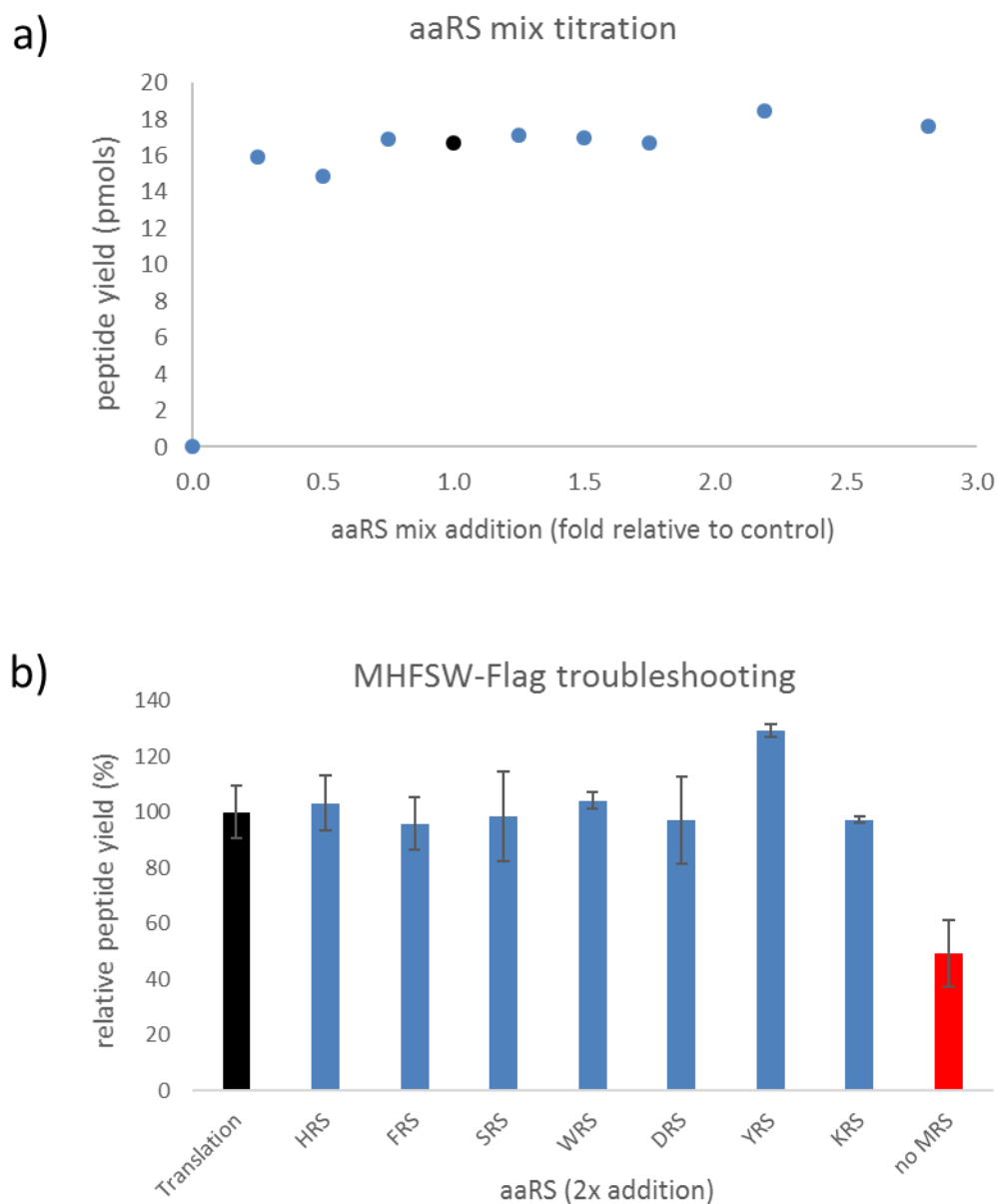


Figure E4 Aminoacyl-tRNA synthetases. a) Effect of the titration of a mix containing all synthetases on peptide yield. Black point represents standard translation concentration (or relative control); b) *in vitro* translation reaction supplementation of 2x of each of the individual synthetases necessary to produce the aminoacyl-tRNAs required for translation of the MHFSW-Flag template. This template demonstrated poor fidelity in previous translations (2.2.2). Relative radiolabeled peptide yield is expressed relative to a translation reaction containing the standard aaRS concentrations.

E2 Summary

We have observed a synergistic effect of many of the PURE components on translation yield. We have found that the purity of the components, particularly translation factors and ribosomes, is critical to the performance of the system. Careful attention to detail during reagent preparation and mixing of components is required to maintain high levels of function.

While it is often the case that increasing the concentration of a single component (eg. ribosomes or EF-Tu) results in increased yield, this effect is not generalizable (eg. NTPs or magnesium) and is often unpredictable, if not counterintuitive. While it may be tempting to increase the concentration of certain components of the system to obtain a higher yield, we have oftentimes observed a negative impact on fidelity via MALDI-TOF when this is attempted. The PURE *in vitro* system was designed to essentially recreate biological translation in a test tube and natural selection has acted to evolve and optimize gene expression, so we recommend to ‘use as directed’.

Appendix F Peptide Binding to BRCA 1

During our selection against BRCA1 we identified peptide 8.6 (3.2.1.2) that contained three arginine analogs (L-canavanine).¹³⁹ This peptide was present in high abundance based on the Sanger sequencing results. However, we were unsuccessful in our attempts to synthesize the Fmoc-protected version of L-canavanine and as a result we were unable to create the version of the peptide selected by the BRCT domain of BRCA1 during the selection. We chose to determine binding affinity of peptide 8.6 using isothermal titration calorimetry (ITC) with arginine substituted for canavanine.

As an alternative to testing the synthetic peptide containing a residue that was not present in the selection, we created *in vitro* translated versions of peptide 8.6, with either arginine or canavanine, in the presence of ³⁵S-Met. We used these radiolabeled peptides in a scintillation-based binding assay (4.5.8) to determine the affinity of the peptides for GST-BRCT and found that the binding affinity of the peptide containing canavanine was the same as for that containing arginine. Furthermore, the scintillation-based affinity of peptide 8.6 containing arginine agreed very closely with the ITC results for the synthetic peptide (1.41 ± 0.12 and 0.80 ± 0.01 , respectively). As a result, we concluded that the arginine substitution for canavanine in the synthetic peptides used for our FP experiments did not affect the binding to BRCT.

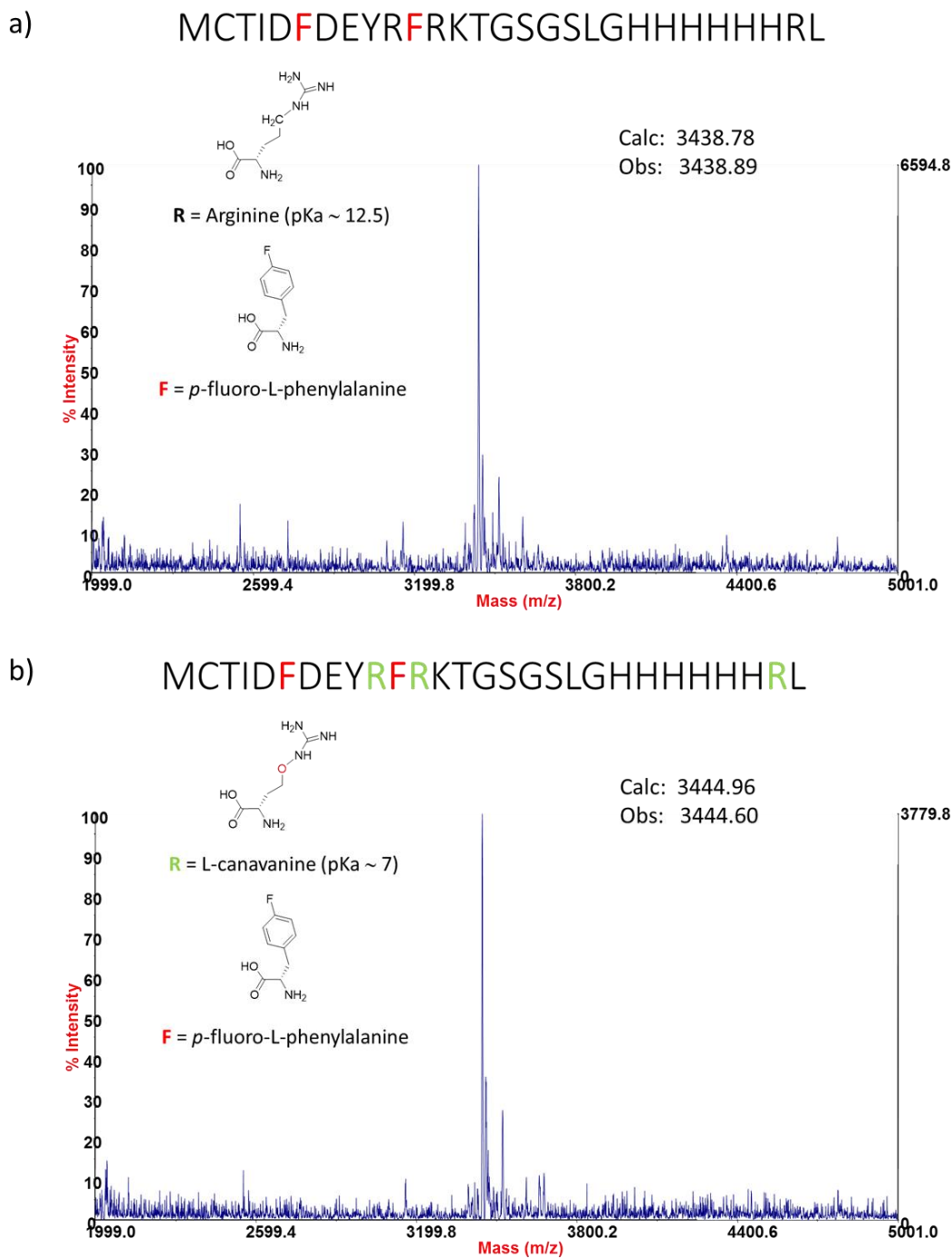


Figure F1 BRCA1-binding peptide 8.6 with and without canavanine (arginine analog). MALDI-TOF spectra of *in vitro* translated peptide 8.6 identified in the selection against BRCA1. a) Peptide 8.6 containing the Phe analog *p*-fluoro-L-phenylalanine (**F**), along with a) *wt* arginines (R) or b) L-canavanine (**R**) substituted for arginine in translation.

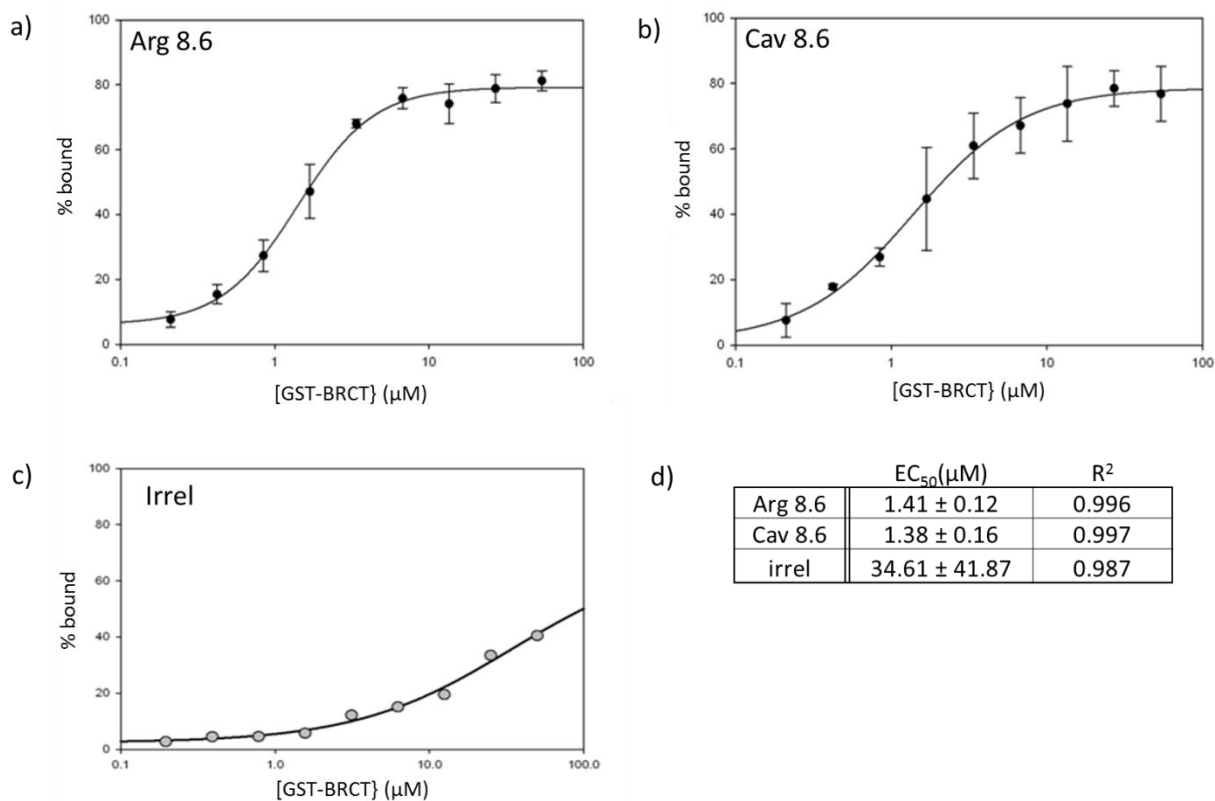


Figure F2 Binding of peptide 8.6 to BRCA1. Binding affinity was determined by the titration of GST-BRCT against peptide 8.6 with a) arginine or b) canavanine. c) Binding of an irrelevant peptide (model peptide OvLap). Peptides were produced by *in vitro* translation containing ³⁵S-Met and bound to immobilized GST-BRCT. The beads were washed twice and binders were eluted with free GSH. % bound was calculated using the equation in 4.5.8. A binding curve was generated by a dynamic fit four-point logistics equation using SigmaPlot 12.5. Experiments were performed at least in duplicate. d) EC₅₀ values and curve fit variances obtained from SigmaPlot analysis. The substitution of canavanine for arginine in peptide 8.6 had no effect on binding affinity to BRCT.

Appendix G Miscellaneous Experiments

G1 Resolving mRNA-peptide fusions

Resolution of mRNA-peptide fusions and free peptides via electrophoresis is challenging, due to both the differential monomer content (nucleic acids vs. amino acids) as well as the large difference in molecular weight of the fusion and the free peptide. We found 10-20% Tris-tricine gradient gels were ideally suited for this purpose.

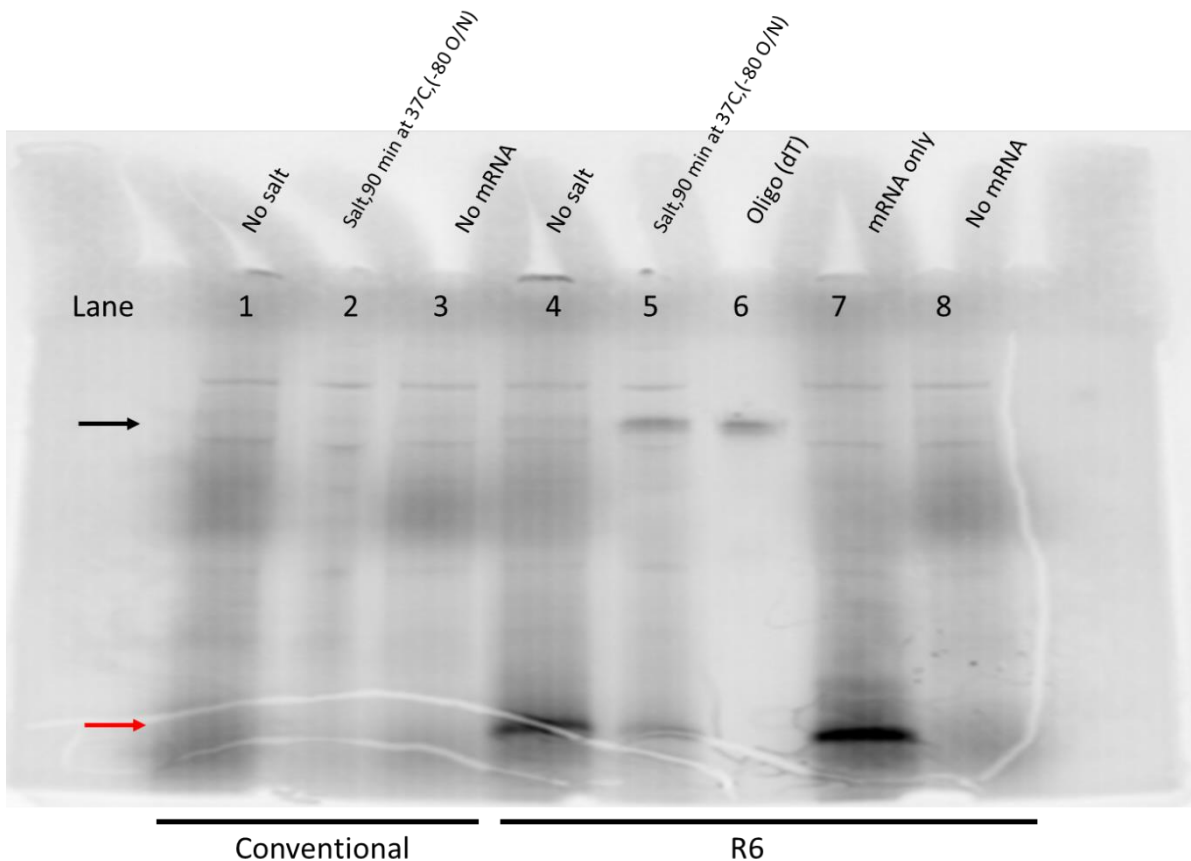


Figure G1 mRNA-peptide fusion formation. 10-20% gradient SDS-PAGE of translation reactions (with ^{35}S -Met) to generate mRNA peptide fusions using two different mRNA display library templates (Conventional-lanes 1-3 and R6-lanes 4-8). Lanes 2 and 5 were fusion formation reactions with the addition of 550 mM NaCl and 55 mM MgCl_2 , following incubation at 37 °C for an additional 90 min. Lane 6 is elution from oligo(dT) affinity resin. Lanes 1 and 4 did not include the salt addition. Lane 7 reaction was initiated by linear mRNA and lanes 3 and 8 contained no mRNA. Black arrow indicates mRNA-peptide fusions. Red arrow is the peptide alone.

G2 mRNA-puromycin linker UV-crosslinking

We quantified the efficiency of formation of covalently-linked mRNA-puromycin linker after psoralen-mediated UV crosslinking via 8% denaturing Urea-PAGE.

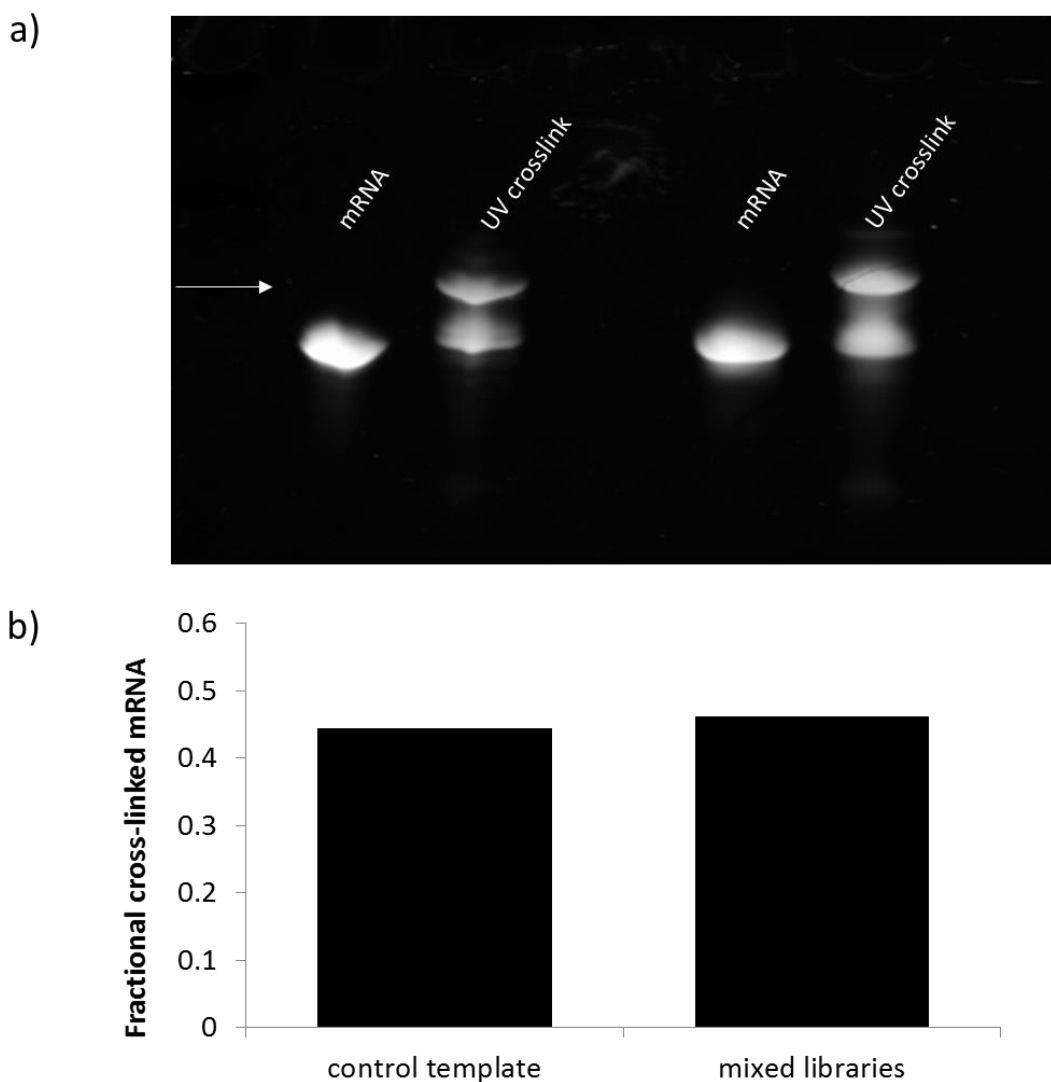


Figure G2 UV crosslinking of mRNA-puromycin linker. a) 8% Urea-PAGE of the UV-crosslinking step in mRNA display. mRNA is from the conventional library used in our first streptavidin selection (4.5.3). White arrow indicates the covalently linked mRNA and puromycin linker. b) Analysis of the efficiency of the cross-linking reaction using the model peptide (OvLap) control template (Table 3.1) and the five libraries in an equimolar mix (Table 4.1). Gel was stained with EtBr and visualized on a CHemiDoc imager (BioRad). Quantification was performed by densitometry.

G3 Characterization of cyclic peptides via SDS-PAGE

Peptides or proteins in different conformations have been shown to migrate at different rates using gel electrophoresis.³² We analyzed the same peptide in linear, mono- and bicyclic conformation by 16% Tris-tricine SDS-PAGE.¹⁵¹ We were unable to resolve the conformers using this method (see ‘mix’ lanes).

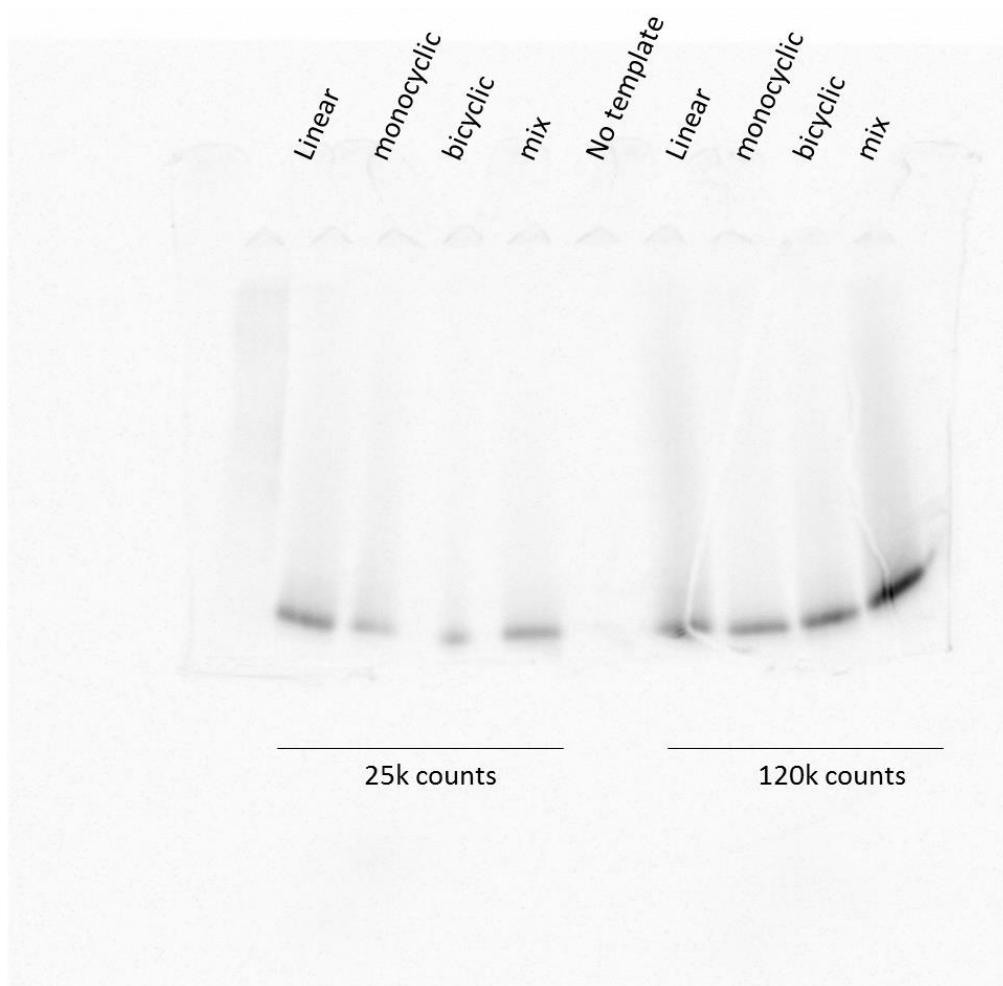


Figure G3 Tris-tricine SDS-PAGE for cyclization characterization. 16% acrylamide peptide gel¹⁵¹ of *in vitro* translated (with ³⁵S-Cyc) OvLap peptide (Table 3.1) in the linear monocyclic (bisalkylated) and bicyclic (bisalkylated and clicked) conformations. 25k and 120k counts represent the scintillation-based quantity of peptide loaded in the noted wells.

Appendix H: Miscellaneous Tables

Table H1a: Streptavidin (monocyclic) Top 100 DNA/peptide sequences from AptaTOOLS

Rank	ID	Aptamer sequence	Peptide	Cluster ID	Count	Frequency	Enrich
1	1UGtSCID	ACCGACCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPD	1404	4188	0.268	1.494
2	1UGtWMH	ACCGACCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPH	1404	3940	0.252	1.190
3	yGFICGGC	ACGAACCCCGACTGCTTTGGCAACCCCGTG	TNPDCFGNPV	1402	2888	0.184	0.999
4	1hHIiJAG	ACGAACCCGGACTGCTTTGGCAACCCCGTC	TNPDCFANPL	1402	447	0.029	1.058
5	CHJejR	ACGAACCCGACTGCTTTGGCAACCCCGTG	TNPDCFGNPV	1402	369	0.024	1.257
6	QzUAyJ	ACCAACCCCGAGTGTCTTTGGCAACCCCGAG	TNPQCFGNPQ	1218	210	0.013	0.673
7	DRjJuA	ACCAACCCCAACTGCTTTGGCAACCCCGTG	TNPNCFGNPL	1218	188	0.012	0.946
8	1IzUCiA	ATTAATACGACTCACTATAGGGTTAACTTT	INTTHYRVNF	221	128	0.008	1.397
9	1BXpWCf	ACCGACCCGACTGCTTTGGCAACCCCGAG	TDPDCFGNPE	1402	112	0.007	0.652
10	1UGtWLA	ACCGACCCGAACTGCTTTGGCAACCCCGAT	TDPNCFGNPH	1404	108	0.007	1.080
11	PIxCIIk	ACCAACCCCGACTGCTTTGGCAACCCCGTG	TNPDCFGNPV	1218	90	0.006	1.316
12	VfIGHHv	ACCAACCCGGACTGCTTTGGCAACGCGAG	TNPDCFGNAT	796	86	0.005	0.439
13	BrLAKFJ	ACGAACCCGGACTGCTTTGGCAACTCGATC	TNPDCFGNSI	1402	54	0.003	0.866
14	1HABjAAH	ACGAACCCCAACTGCTTTGGCAACCCGAAG	TNPNCFGNPK	796	39	0.002	0.482
15	1HAsHAIj	ACGATCCACCAGTGGCTTTACCATCCACAA	TIHQWLYHPQ	1275	38	0.002	1.602
16	kDLzGC	ACGGACCCGAACTGCTTTGGGAACAGATG	TDPNCFGNQM	796	34	0.002	0.430
17	KHGAIQy	ACGAACCCGAGTGTCTTTGGCAACCCGATG	TNPQCFGNPM	796	34	0.002	0.604
18	yGFIVd	ACGAACCCGACTGCTTTGGCAACCCCGCG	TNPDCFGNPA	1402	32	0.002	1.058
19	RDqHGKC	ACCGACCCCAACTGCTTTGGCAACACCATG	TDPNCFGNTM	1218	28	0.002	0.453
20	1VAAGHWu	ACTGACCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPD	1404	27	0.002	2.232
21	1TDLAIGV	ACCGACACGAACTGCTTTGGCAACCCCGAC	TDTNCFGNPH	1404	24	0.002	1.012
22	DRjHAGH	ACCAACCCCAACTGCTTTGGCAACCCCGAC	TNPNCFGNPT	1402	24	0.002	0.543
23	1UGtVJj	ACCGACCCGAACTGCTTTGGCAACCCCGCG	TDPNCFGNPR	1404	24	0.002	1.720
24	1wiFJIC	ACGGACCCCAACTGCTTTGGGAACATCTTG	TDPNCFGNIL	1402	23	0.001	0.415
25	MHHACixA	ACCGATCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPD	1404	19	0.001	1.945
26	1UGWUHP	ACCGACCCGACTGCTTTGGGAACATGATG	TDPDCFGNMM	796	19	0.001	0.530
27	PJBaBn	ACGAACCCGACTGCTTTGGGAACAACGAG	TNPDCFGNNE	1402	18	0.001	0.586
28	1SHIJJmB	ACCGCCCCGAACTGCTTTGGCAACCCCGAC	TGNPCFGNPD	1404	16	0.001	2.646
29	1TDLApHH	ACCGACACGAACTGCTTTGGCAACCCCGAC	TDTNCFGNPD	1404	15	0.001	1.112
30	1TJHWBuB	GCCGACCCGAACTGCTTTGGCAACCCCGAC	ADPNCFGNPD	1404	14	0.001	1.433
31	1VAAGHGAJ	ACTGACCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPH	1404	14	0.001	1.254
32	MHHACpAG	ACCGATCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPH	1404	13	0.001	2.540
33	1IiWJpj	ACGGACCCGAACTGCTTTGACAACCCCGAG	TDPNCFDNPQ	1402	13	0.001	0.998
34	1NsGGuAJ	ACGGACCCGAACTGCTTTTACGACGGGTGC	TDPNCFYDGC	395	13	0.001	0.621
35	1TITpAJ	ACCGACCCGAACTGCTTTGGCAGCCCGAC	TDPNCFGSPH	1404	12	0.001	1.228
36	yGFICGy	ACGAACCCGACTGCTTTGGCAACCCCGTA	TNPDCFGNPV	1402	11	0.001	2.149
37	1TITAIGF	ACCGACCCGAACTGCTTTGGCAGCCCGAC	TDPNCFGSPD	1404	11	0.001	0.909
38	HAAFJGQI	ACGAACACCGACTGCTTTGGCAACCCCGTG	TNTDCFGNPV	1402	11	0.001	0.622
39	1BfWgDAE	ACCGACCCGAACTGCTTCGGCAACCCCGAC	TDPNCFGNPH	1404	11	0.001	2.149
40	1CXAJGdf	ACCGACCCGAGTGTCTTTGGCAACCCCGAC	TDPSCFGNPD	1404	10	0.001	1.433
41	BRIAHCGV	ACCAACCCGGACTGCTTTGGGAACAACACG	TNPDCFGNNT	857	9	0.001	0.645
42	1TCeJuGA	ACCGACCCGAACTGCCTTGGCAACCCCGAC	TDPNCLGNPD	1404	9	0.001	2.764
43	JIDshAK	ACCGACCCGAACTGCTTTGGCAGCCCGAC	TDPNCFGIPD	1404	9	0.001	6.448
44	BkAsCgH	ACCGACCCGAACTGCTTTGGCAGCCCGAC	TDPNCFGTPD	1404	8	0.001	1.563
45	1UGoByIA	ACCGACCCGAACTGCTTTGGCAACCCCTCAC	TDPNCFGNPH	1404	8	0.001	1.146
46	1CXKASB	ACCGACCCGAGTGTCTTTGGCAACCCCGAC	TDPSCFGNPH	1404	8	0.001	1.433
47	1BuyDjAB	ACGAACCCGGAGTGTCTTTCAGAACATGTAC	TNPECFQNYM	49	8	0.001	0.506
48	1LzIJjHF	ACGAACCCGGACTGCTTTGACAACAAGAAC	TNPDCFDNKN	1402	8	0.001	0.593
49	UACsGXA	ACGATCCACCAGTGGCTTACCATCCACAAG	TIHQWLTiHK	887	8	0.001	1.720
50	1UGtSAJH	ACCGACCCGAACTGCTTTGGCAACCCCGGC	TDPNCFGNPG	1404	8	0.001	8.598

Table H1b: Streptavidin (monocyclic) Top 100 DNA/peptide sequences from AptaTOOLS

Rank	ID	Aptamer sequence	Peptide	Cluster ID	Count	Frequency	Enrich
51	UjCJJoI	ACGGACCCCGACTGCTTTGGCAACCCCGTG	TDPDCFVGNPV	1402	7	0.000	3.762
52	1UGtCir	ACGGACCCCGAAGTGTCTTGGCAACCCCAAC	TDPNCFVGNPN	1404	7	0.000	1.368
53	1LKIImQ	ACGGACCCCGAAGTGTCTTGGTAACCCCCAC	TDPNCFVGNPH	1404	7	0.000	3.762
54	1zCvVh	ACGAACTCCGACTGCTTTGGCAACCCCGTG	TNSDCFVGNPV	1402	7	0.000	1.505
55	UVgGIAD	ACCAACCCCGACTGCTTTGGCAACCCACCAG	TNPDCFVGNHQ	1197	7	0.000	0.752
56	1HJHAMIAF	ACCGACCCCGACTGCTTTGGCAACCCCGAC	TDPDCFVGNPD	1404	7	0.000	5.015
57	1UGtAzJA	ACCGACCCCGAAGTGTCTTGGCAACCCCTAC	TDPNCFVGNPY	1404	7	0.000	0.885
58	BVAJACIb	ACGGACCCCGAAGTGTCTTGGCAACCCCGAC	TDPNCFVGNPD	1402	7	0.000	2.508
59	ViGJIHu	ACCGACCCCGAAGTGTCTTGGCAACCCCCAC	TDPNCFVGNPH	1404	7	0.000	1.157
60	1TWJoDAA	ACGAACCCCGACTGCTTTGGCAACCCCGTG	TNSDCFVGNPV	1402	7	0.000	2.149
61	yGzGIJG	ACGAACCCCGACTGCTTTGGCAACCCCATG	TNPDCFVGNPM	1402	7	0.000	1.505
62	HuHABsE	ACGAGCCCGACTGCTTTGGCAACCCCGTG	TSPDCFVGNPV	1402	6	0.000	0.496
63	1UGtWBX	ACCGACCCCGAAGTGTCTTGGCAACCCCCAG	TDPNCFVGNPQ	1218	6	0.000	2.149
64	1UGtSCGG	ACCGACCCCGAAGTGTCTTGGCAACCCCGAT	TDPNCFVGNPD	1404	6	0.000	1.612
65	HAIAAyIF	ACGAACCCCGACTGCCTTGGCAACCCCGTG	TNPDCLVGNPV	1402	6	0.000	2.149
66	1TJHwvGF	GCCGACCCCGAAGTGTCTTGGCAACCCCCAC	ADPNCFVGNPH	1404	6	0.000	0.921
67	KJAHdKM	ACCGACTCGAAGTGTCTTGGCAACCCCGAC	TDSNCFVGNPD	1404	6	0.000	1.612
68	ViHAZJG	ACGGACCCCGAAGTGTCTTGGCAACCCCGAC	TDPNCFVGNPD	1404	6	0.000	0.921
69	pHlGcn	ACGAACCCCGACTGTCTTGGCAACCCCGTG	TNPDCFVGNPV	1402	6	0.000	1.612
70	BVAIJJIJ	ACGGACCCCGAAGTGTCTTGGCAACCCCCAC	TDPNCFVGNPH	1402	6	0.000	1.433
71	1ZJGHJjr	ACCGACCAGAAGTGTCTTGGCAACCCCGAC	TDQNCVGNPD	1404	6	0.000	6.448
72	1UGtWAGF	ACCGACCCCGAAGTGTCTTGGCAACCCCCCC	TDPNCFVGNPP	1404	6	0.000	4.299
73	1SHJAAaIF	ACCGGCCCGAAGTGTCTTGGCAACCCCCAC	TGPNCFVGNPH	1404	5	0.000	0.632
74	JIDsdQG	ACCGACCCCGAAGTGTCTTGGCATCCCCAC	TDPNCFVGNPH	1404	5	0.000	1.791
75	CJIXys	ACGAACCCCGACTGCTTTGGCAACCCCGTG	TNADCFVGNPV	1402	5	0.000	3.582
76	BzaIqFJ	ACGAACCCCAAGTGTCTTGGCAACCCCGAC	TNPDCFVGNPD	1404	5	0.000	2.149
77	1gGcfIAB	ACGAATCCCGACTGCTTTGGCAACCCCGTG	TNPDCFVGNPV	1402	5	0.000	1.535
78	1UGtVvI	ACCGACCCCGAAGTGTCTTGGCAACCCCTC	TDPNCFVGNPL	1404	5	0.000	1.791
79	1GGFJJCHH	ACCGACCCCGAGTGTCTTGGGAACCCCAAC	TDPECFVGNPN	1404	5	0.000	0.189
80	hAAWJO	ACCGACCCCAAGTGTCTTGGCAACCCCTAC	TDPNCFVGNPY	1404	5	0.000	0.896
81	1LKI IAzC	ACCGACCCCGAAGTGTCTTGGTAACCCCGAC	TDPNCFVGNPD	1404	5	0.000	1.535
82	MGGt1F	ACCGACCCCGAAGTGTCTTGGCAACCCCCAC	TDPNCFVGNPH	1404	5	0.000	1.535
83	KJAHDAeI	ACCGACTCGAAGTGTCTTGGCAACCCCCAC	TDSNCFVGNPH	1404	5	0.000	0.597
84	1UGtRGJA	ACCGACCCCGAAGTGTCTTGGCAACCCCGTG	TDPNCFVGNPV	1402	5	0.000	1.343
85	BZJoGqF	ACGAACCCCGACTGCTTCGGCAACCCCGTG	TNPDCFVGNPV	1402	5	0.000	0.977
86	1BuJHpIGD	ACGAACCCCGACTGCTTTGGGCGCCCCAG	TNPDCFVGNPQ	651	5	0.000	0.827
87	1SGOzVG	ACGAACCCCGACTGCTTTGGCAGCCCCGTG	TNPDCFVGNPV	1402	5	0.000	0.512
88	IpGHJIIC	ACGAAGCCCGACTGCTTTGGCAACCCCGTG	TKPDCFVGNPV	1402	5	0.000	1.791
89	OuTABm	TCCAACCCCGACTGCTTTGGGAACCCCGCG	SNPDCFVGNPA	359	5	0.000	0.256
90	MsGHaJA	ACGAACCCCGACTGCTTTGGCAACCCCGTG	TNRDCFVGNPV	1402	4	0.000	-1.000
91	BfvIyC	ACCGACCCCGAAGTGTCTTGGCAATCCCCAC	TDPNCFVGNPH	1404	4	0.000	1.433
92	CHJedJA	ACGAACCCCGACTGCTTTGGCAACCCCGCG	TNPDCFVGNPA	1402	4	0.000	0.782
93	1PJzWHHF	ACGAACCCCGACTGCTTTGGCAACAACAAC	TNPDCFVGNPN	1402	4	0.000	0.269
94	yHAIJKJ	ACGAACCCCGACTGCTTTGGCAACCCCTGTG	TNPDCFVGNPV	1402	4	0.000	0.860
95	1HABuCIQ	ACGAACCCCAAGTGTCTTGGCAACCCCGTG	TNPDCFVGNPV	1402	4	0.000	0.860
96	HIPIJAHF	TCGAACCCCGACTGCTTTGGCAACCCCGTG	SNPDCFVGNPV	1402	4	0.000	8.598
97	yGzICZ	ACGAACCCCGACTGCTTTGGCAACCCCCAC	TNPDCFVGNPH	1402	4	0.000	2.866
98	BmfJrJD	ACCGACCCCGAAGTGTCTTGGGAACCCCGAC	TDPNCFVGNPD	1404	4	0.000	8.598
99	1BfWcoA	ACCGACCCCGAAGTGTCTTCGGCAACCCCGAC	TDPNCFVGNPD	1404	4	0.000	0.955
100	MGGxCHJ	ACCGACCCCGAAGTGTCTTGGCAACCCCGAC	TDPNCFVGNPD	1404	4	0.000	2.866

Figure H1: Streptavidin (monocyclic) Top 100 sequence homology (10 random AAs)

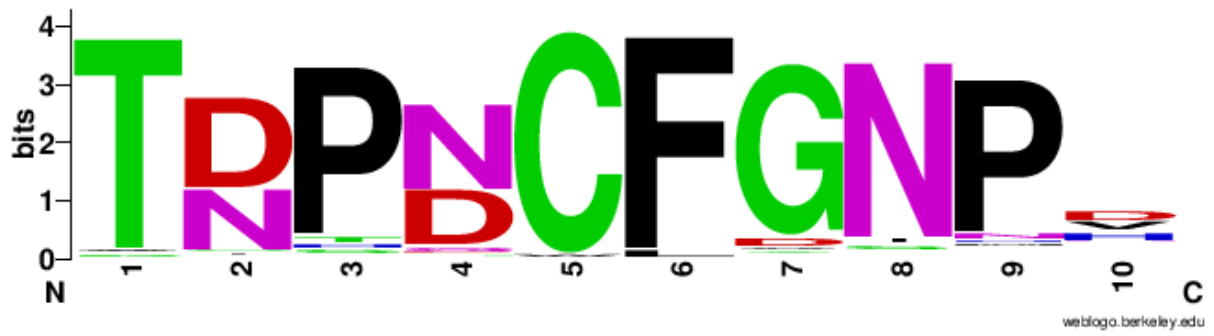


Table H2a: Streptavidin (bicyclic) Top 100 DNA/peptide sequences from AptaTOOLS

Rank	ID	Aptamer sequence	Peptide	Cluster ID	Count	Frequency	Enrich
1	1UGtSCID	ACCGACCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPD	47984	2191	0.019605	5.3052
2	1UGtWMH	ACCGACCCGAACTGCTTTGGCAACCCCCAC	TDPNCFGNPH	47984	1586	0.014192	4.0998
3	yGFICGGC	ACGAACCCCGACTGCTTTGGCAACCCCGTG	TNPDCFGNPV	49379	1074	0.00961	3.6615
4	1IzUCiA	ATTAATACGACTCACTATAGGGTTAACTTT	INTTHYRVNF	14268	382	0.003418	1.3813
5	CHJejR	ACGAACCCCTGACTGCTTTGGCAACCCCGTG	TNPDCFGNPV	49379	216	0.001933	4.9949
6	BhtCIRH	CACCCGAGAAGTGTTCACGTGTTCTCG	HPQNCFHVFS	39899	120	0.001074	2.7966
7	1HAsHAIj	ACGATCCACCAGTGGCTTTACCATCCACAA	TIHQWLYHPQ	42373	79	7.07E-04	33.666
8	1OpHJwJG	CACCCGTACTGCTTCGGCTTTGACAACAA	HPYCFGFDNN	48929	62	5.55E-04	9.2476
9	yGFIVd	ACGAACCCCGACTGCTTTGGCAACCCCGCG	TNPDCFGNPA	49379	58	5.19E-04	4.0237
10	1hHIiJAG	ACGAACCCGACTGCTTTGCCAACCCCTC	TNPDCFANPL	49379	40	3.58E-04	3.9775
11	PIxCIik	ACCAACCCCGACTGCTTTGGCAACCCCGTG	TNPDCFGNPV	49379	35	3.13E-04	2.9831
12	KJQoDp	CTGTACAAGTGAACCACTTTGACCACCAC	LYNCNHFDHH	49426	23	2.06E-04	4.9008
13	JTkAmC	CACCCGTACTGCTTCGGCTTTGAGGGCCAC	HPYCFGFQGH	21647	23	2.06E-04	7.6234
14	DRjJuA	ACCAACCCCAACTGCTTTGGCAACCCCTG	TNPNCFGNPL	47984	23	2.06E-04	2.5411
15	QzUAYj	ACCAACCCCACTGCTTTGGCAACCCCCAG	TNPQCFGNPQ	49379	21	1.88E-04	2.3202
16	DpHCHJm	ACGAACCCGACTGCTTTGGCAACCCCGTG	TNPDCFGNPV	49379	19	1.70E-04	4.7232
17	1UGtWLA	ACCGACCCGAACTGCTTTGGCAACCCCAT	TDPNCFGNPH	47984	18	1.61E-04	2.2373
18	ViHAZJG	ACCGACCCGAACTGTTTGGCAACCCCGAC	TDPNCFGNPD	47984	18	1.61E-04	4.4746
19	1JFIhCyB	CACAACTACAAGTGTTTGCGACCACCAC	HNYNCFDHH	41923	16	1.43E-04	3.6715
20	1UGtCir	ACCGACCCGAACTGCTTTGGCAACCCCAAC	TDPNCFGNPN	47984	15	1.34E-04	4.0678
21	lgzBsHE	CACGCCACAAGTGTTTACGACCACCAC	HAHNCFYDHH	41923	15	1.34E-04	1.8644
22	1VAAGHWu	ACTGACCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPD	47984	15	1.34E-04	5.5933
23	BmUlCJH	CACCACAAGTGCACAAGTTCGCCACCAC	HHNCDNFAHH	37034	14	1.25E-04	2.7842
24	1UGtVJj	ACCGACCCGAACTGCTTTGGCAACCCCGCG	TDPNCFYXPR	47984	14	1.25E-04	10.441
25	UCIrJGGC	ACGACCCCACTACTTTCGACTGCCACTTC	THYFYDCHF	43393	12	1.07E-04	5.1139
26	1VAAGHGAJ	ACTGACCCGAACTGCTTTGGCAACCCCCAC	TDPNCFGNPH	47984	12	1.07E-04	7.1594
27	1BxyEtJD	ACAGACCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGNPD	47984	11	9.84E-05	8.2035
28	UACsGXA	ACGATCCACCAGTGGCTTACCATCCACAAG	TIHQWLTlHK	28316	11	9.84E-05	5.469
29	1SJBuIeA	CACTACCCCACTGCTTTGACGGCCACTTC	HYTHCFDGHF	33393	11	9.84E-05	3.646
30	1JtAJBpB	CACAACAAGGACTGCTTTGGCCACTTCTTC	HNNDCFGHFF	45281	11	9.84E-05	2.5242
31	BmgCGoA	CACTACTAGACTGCTTTCCACCACCAC	HYDCFSTHH	46976	11	9.84E-05	4.6877
32	HxBaUH	ACCAACCCGAACTGCTTTGGCAACCCCGAC	TNPNCFGNPD	47984	11	9.84E-05	6.5628
33	KHCyJAGC	TTCACCACGGGTGCTTTCGCCACAAGTAC	FNHGCFRHHY	54209	10	8.95E-05	4.2616
34	1lISmM	CACTACAACCCGTGCTTTAACCAGCACTTC	HYNPCFNQHF	39298	10	8.95E-05	9.9436
35	1UboHR	CACCCGTACTGCTTCGGCTTTACTGACCAC	HPYCFGFYH	22766	10	8.95E-05	7.4577
36	1BxJhAUE	CACCACATCAAGTGTTTGGCGACCACCAC	HHINCFGDHH	41923	10	8.95E-05	3.3145
37	1TDLApHH	ACCGACACGAACTGCTTTGGCAACCCCGAC	TDNCFGNPD	47984	10	8.95E-05	9.9436
38	1TaDgjH	CACCACTACTACCCCAAGTACTTTAACGAC	HHYYPKYFND	24113	10	8.95E-05	4.9718
39	BsxIwJB	CACCAGACCACTGCTTTTCAACCACCAC	HQTHCFNHH	30485	10	8.95E-05	3.7289
40	zJHJgGF	AACAACCAGACTGCTTTGACCACCACTTC	NNHDCFHHE	54427	10	8.95E-05	3.7289
41	1IIDvJIU	CACCACGAGTCTGCTTTGGCTACCACCAC	HHEFCFGYHH	22327	9	8.05E-05	2.2373
42	1TITAIGF	ACCGACCCGAACTGCTTTGGCAACCCCGAC	TDPNCFGSPD	47984	9	8.05E-05	5.3696
43	1HCHHaHJD	CACCACATCAAGTGTTTGACCACCACCTC	HHINCFDHH	41923	9	8.05E-05	4.4746
44	1LAgFJBZ	ACCTACGCGAACTGCTTTACGACCACCAC	TYANCFYDHH	35969	9	8.05E-05	5.3696
45	LAHALGg	CACCACATCTGCGACCCTTTAACACGAC	HHICDFNHD	37034	9	8.05E-05	2.9831
46	1BlAHAMp	AACTGCGACACCCACAAGTCCACTTTTAC	NCDTHNCFY	51614	9	8.05E-05	6.7119
47	1BqpJfID	CACAACACCACTGCTTTTGAACCACCAC	HNTHCFLNHH	30485	9	8.05E-05	2.0652
48	1TJHWBuB	GCCGACCCGAACTGCTTTGGCAACCCCGAC	ADPNCFGNPD	47984	9	8.05E-05	13.424
49	1cqIHIH	AACAACCCTACTGCTTTTACCAGCACTTC	NNHYCFYQHF	54427	9	8.05E-05	2.4407
50	1PIJwRCI	CACGACCTCTACTGCTTTGACCACCACTAC	HDLYCFDHHY	35550	9	8.05E-05	2.4407

Table H2b: Streptavidin (bicyclic) Top 100 DNA/peptide sequences from AptaTOOLS

Rank	ID	Aptamer sequence	Peptide	Cluster ID	Count	Frequency	Enrich
51	CqHFIHI	CACCACACGTACTGCTTTGACCACTTCCAG	HHTYCFDHFQ	40180	9	8.05E-05	1.5793
52	1BXpWCf	ACCGACCCGACTGCTTTGGCAACCCCGAG	TDPDCFGNPE	49379	9	8.05E-05	3.356
53	BgGCHuL	AACACCCACTACTGCTTTTGCAACCCACCAC	NTHYCFCTHH	46976	9	8.05E-05	2.9831
54	hKI IJGH	TTCGACCACAACGTGCTTTGCGCCCCACCAC	FDHNCFAHH	54427	9	8.05E-05	4.4746
55	1BTCHGHHn	CACTACAACAGCTGCTTTGCCACCACATC	HYNSCFAHHI	41923	9	8.05E-05	6.7119
56	1SCfBSO	CACTACTACTACTGCTTTAGGACCGCCAC	HYYYCFRTAH	21336	9	8.05E-05	-1
57	1GFIJGJII	TTCAACCACACGTGCTTTTGCCACACCCAG	FNHTCFCHTQ	43290	9	8.05E-05	1.7899
58	QGExMGC	CACAACCCCACTGCTTTAACGTCACGTTT	HNTHCFNVT	30756	9	8.05E-05	4.4746
59	BWACwAm	CACACCCACTACTGCTTTGGCCACCACCTG	HTHYCFGHH	35550	9	8.05E-05	2.2373
60	TLmAHa	CACCCGACAGACTGCTTTGACGTGTTCAAG	HPQNCFDVFK	39899	9	8.05E-05	4.4746
61	1FIHCGCHS	CACTACTAGACTGCTTTACCGCCACCTG	HYDCTFGHL	16361	9	8.05E-05	5.3696
62	LNKJAJA	TTCACAACGACTGCTTTATGAACCCACCAC	FHNDCFMNHH	43505	8	7.16E-05	2.9831
63	TGHFIGCID	CACGCCAACTGCAACGACTTTAACCACCAC	HANCNDFNHH	32428	8	7.16E-05	5.9662
64	1TITpAJ	ACCGACCCGAACTGCTTTGGCAGCCCCAC	TDPNCFGSPH	47984	8	7.16E-05	11.932
65	QIIEoHD	CACTACAACACTACTGCTTTGTGACCCACATC	HYNYCFVSHI	46976	8	7.16E-05	5.9662
66	1IGGGDAJO	TTCAACCACGTGTGCTTTTACGGCCACTAC	FNHVCFYGHY	54209	8	7.16E-05	5.9662
67	BzaIqFJ	ACCGACCCAACTGCTTTGGCAACCCCGAC	TDPNCFGNPD	47984	8	7.16E-05	7.9549
68	1HABUCIQ	ACGAACCCCACTGCTTTGGCAACCCCGTG	TNPNCFGNPV	47984	8	7.16E-05	3.4092
69	RJJJqEs	CACACCCACTGCGACGGCTTTAGCCACTAC	HTHCDGFSHY	12456	8	7.16E-05	5.9662
70	MwCJDT	AACATGCACTGCACCATCTTTGACCACCAC	NMHCTIFDHH	41672	8	7.16E-05	4.7729
71	GkfAIL	CACAACCAGTTCTGCTTTACGAGCCACCAC	HNQFCFTSHH	12840	8	7.16E-05	4.7729
72	QFJHNAJ	TTCAACACCAACTGCTTTAACCACCACCTC	FNTNCFNHHF	47579	8	7.16E-05	7.9549
73	1KaJhaE	CACTGCGAGTACGACTACGACCACTTTATG	HCEYDYDFHM	49284	8	7.16E-05	5.9662
74	JFJHJtG	CACTACACGAGCTGCTTTACCCACCACGAC	HYTSCFTHHD	41923	8	7.16E-05	4.7729
75	BeIghAF	CACCACAACCTGCGACCGCTTTGCCACCCAC	HHNCDRFATH	37034	8	7.16E-05	7.9549
76	1HGokACH	ATCAACCCGACTGCTTTACCGACCCACCAC	INPYCFDHH	23570	8	7.16E-05	3.9775
77	1RIHIUAF	CACTACACGACTGCTTTAACCACCACTAC	HYTSCFNHHY	41923	7	6.26E-05	3.4803
78	VExIBiE	AGTACCACGAGTGTGCTTTGACCCCACTC	TYHECFDPHF	39298	7	6.26E-05	6.9605
79	1BjCjrJe	ATCCACACGACTGCTTTAGCCACCACCAG	IHTDCFSHQ	43505	7	6.26E-05	-1
80	1NAjHkc	ACGTTCTACGAGTGTGCTTTAACGACCACCAC	TFYECFNDHH	32689	7	6.26E-05	2.6102
81	1JHHiKR	CACCACATCTGCAGTACTTTAACCACCAC	HHICTYFNHH	48362	7	6.26E-05	1.3921
82	JIDSDQG	ACCGACCCGAACTGCTTTGGCATCCCCAC	TDPNCFGIPH	47984	7	6.26E-05	20.882
83	pAICIIHF	AACCACGACGACTGCTTTGTGATCACCAC	NHDDCFVITH	34311	7	6.26E-05	2.3202
84	1HqJCGHHH	CACACCACTACTGCTTTGACCACCACTAC	HTNYCFDHHY	40180	7	6.26E-05	2.6102
85	1UwCHHIA	GACAACACTAGACTGCTTTACCACCACCAC	DNYDCFTTHH	46976	7	6.26E-05	5.2204
86	RHJJGoID	CACCACATCTGCAACTGCTTTGACGGCCAC	HHICNCFDGH	32428	7	6.26E-05	1.6063
87	1IAHARJGE	ACCTGACCATGCCGGCCACCACCTTTTTC	TCTMPGHFF	54333	7	6.26E-05	-1
88	BwHVoGG	ACCTGACCACTCCATCGTCCGCTTTTAC	T*HHSIVRFH	21683	7	6.26E-05	2.9831
89	HCJvbaE	AACCACGAGACGTGCTTTAACCACCACACTAC	NHETCFNHHY	39298	7	6.26E-05	1.2283
90	1BdICIAHG	ACCCACACCTGCAACGTCTTTGACCACCAC	THTCNVFDHH	42907	7	6.26E-05	1.7401
91	TyIoJQ	CACCACAACCACTGCTTTAACCACCACACTAC	HHNQCFNHHY	30485	7	6.26E-05	2.0882
92	yGzGIJG	ACGAACCCGACTGCTTTGGCAACCCCATG	TNPDCFGNPM	49379	7	6.26E-05	4.1763
93	LAQGgR	CACTACATCTGCAACAACCTTGACGACCAC	HYICNFFDDH	32428	7	6.26E-05	0.8031
94	1TDLAIGV	ACCGACACGAACTGCTTTGGCAACCCCCAC	TDNCFGNPH	47984	7	6.26E-05	5.2204
95	1BkBVjw	CACCACACCACTGCTTTTACAGCCACCAC	HHTHCFYSHH	30485	7	6.26E-05	0.9079
96	1CHHUAG	AACAACCCTGTGCTTTAACCACCACATC	NNHLCFNHHI	54427	7	6.26E-05	1.8983
97	QlaeGE	ACGTTCTGCGACCACCACCTGTTTGTGTTT	TFCDHHSFEF	25032	7	6.26E-05	6.9605
98	1QaxxIG	CACCAGCACCCTGCTTTGCCACCACACTAC	HQHHCFAHHY	39298	7	6.26E-05	4.1763
99	BtAGWAHF	CACAACCCTGCGAGTTCTTTCCGGGCCAC	HNHCEFFPGH	35202	7	6.26E-05	0.7734
100	VCGKfCJ	AACCATCACCACTGCTTTGACCACCTCCTG	NHHQCFDHF	30498	7	6.26E-05	6.9605

Figure H2: Streptavidin (bicyclic) Top 100 sequence homology (10 random AAs)

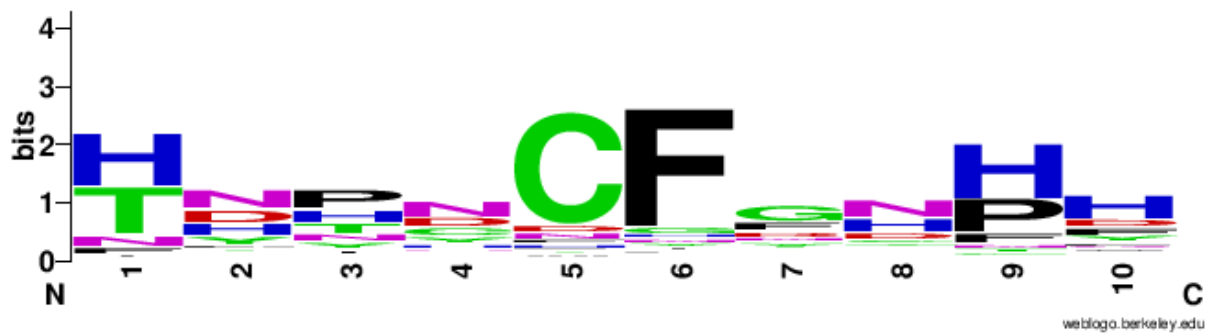


Table H3a: Streptavidin #2 Top 100 DNA/peptide sequences from AptaTOOLS

Rank	ID	Aptamer sequence	Peptide	Cluster ID	Count	Frequency	Enrich
1	lQZIBeAJ	AACTACAAGAACTGGGTGAACCACCCCCAGAACTAC	NYKNWVNHPQNY	15801	40156	0.173332	1.3079
2	JJFJAGJK	AACTTGTGCGAACTGGCTGAACCACCCCCAGAACAGG	NLSNWLNHPQNR	16012	29961	0.129326	1.2481
3	lPzJQIp	AACTACGAGAACTGGGTGAACCACCCCCAGAACTAC	NYENWVNHPQNY	15801	12612	0.054439	1.5134
4	BQJIIIAAE	ACGCGGTACCAGTGAACACCCACCCCCAGAACGTC	TRYQWNTHPQNV	15817	10301	0.044464	0.6892
5	NqAAHIIH	AAGTTCACGACTGGAAGCACCACCCCCAGAATC	KFNDWKHHPQNI	15872	9661	0.041701	0.8941
6	lzeyGm	CCGTTCACAACTGGCAGGACCACCCCCAGAACGCC	PFNTWQDHPQNA	15057	6613	0.028545	0.7561
7	RGHHHAi	AACTTGTGCGAACTGGCTAAACCACCCCCAGAACAGG	NLSNWLNHPQNR	16012	4274	0.018449	1.3668
8	lQZIBdJC	AACTACAAGAACTGGGTGAACCACCCCCAGAACTAT	NYKNWVNHPQNY	15801	3500	0.015108	1.6685
9	lBSFJBwHD	ACGGTCAACAACCTGGAGGACCCACCCGAGAACGCC	TVNNWRTHPQNA	14953	3349	0.014456	1.1847
10	JJFJAGJX	AACTTGTGCGAACTGGCTGAACCACCCCCAGAACAGT	NLSNWLNHPQNS	16012	2950	0.012734	1.5906
11	KAMGGHGJ	CACTACGAGTACGAGCTCTTGCCCTCGCACCCGCGAG	HYEYELLPSPHQ	15706	2652	0.011447	0.4815
12	lGhFYGIA	CACGACACCTACACGCCCTTGCCGGAGCACCCGCGAG	HDTYTPLPPEHPQ	14413	2549	0.011003	0.539
13	JJFJAGJAG	AACTTGTGCGAACTGGCTGAACCACCCCCAGAACAGC	NLSNWLNHPQNS	16012	2320	0.010014	1.2765
14	lrfJkGAE	ACGTTTCGACAACCTGGAGCAGGCACCCCCAGAACCAC	TFDNWSRHPQNH	15743	2135	0.009216	0.6873
15	BQJIIIAV	ACGCGGTACCAGTGAACACCCACCCCCAGAACGTT	TRYQWNTHPQNV	15817	1605	0.006928	0.9271
16	lLAIIJCHM	AACTTATCGAACTGGCTGAACCACCCCCAGAACAGG	NLSNWLNHPQNR	16012	1378	0.005948	1.1599
17	MlIIsCI	ACGTACGACGAGTGAAGTCGCACCCCCAGAACAGC	TYDEWKSHPQNT	14860	1229	0.005305	0.6633
18	lPzJQIDA	AACTACGAGAACTGGGTGAACCACCCCCAGAATAT	NYENWVNHPQNY	15801	1171	0.005055	1.8033
19	BRCGIJACJ	ACGTTTGACAACCTGGAGCAGGCACCCCCAGAACCAC	TFDNWSRHPQNH	15743	1118	0.004826	1.2895
20	lVAAlIaB	CACGACACCTACCGCCCGCAGTCCCACCCCGAG	HDTYRPPKSHPO	16176	1021	0.004407	0.3815
21	BqGAvaJ	GACTACAAGAACTGGGTGAACCACCCCCAGAACTAC	DYKNWVNHPQNY	15801	986	0.004256	1.4287
22	cyAJQC	ACGCAGTACCAGTGAACACCCACCCCCAGAACGTC	TQYQWNTHPQNV	15817	970	0.004187	1.4242
23	NqAAHJAE	AAGTTCACGACTGGAAGCACCACCCCCAGAACATT	KFNDWKHHPQNI	15872	969	0.004183	1.1489
24	KjAGqh	CTGTTGAGCAGTGGCGGACCCACCCGAGAACAGC	LFEQWRTHPQNT	15233	929	0.00401	0.5369
25	BsJEqzD	ACGTACGACAGGTGGATCGTGACCCCCAGAACAAC	TYDRWIVHPQNN	15743	807	0.003483	0.5816
26	lJCxJJGAF	CTGTTGAGATCTGGCGCGCACCCCGAGAATC	LFEIWRHPQNI	14503	786	0.003393	0.586
27	DtCHfd	AAGTTCAGTTGATCAACTGGTTTACCGCACCCGCGAG	KFKLINWFTHPO	15377	771	0.003328	0.8666
28	TAAGUHAF	CACGACACCTACAAGTCTGCGAGTACCACCCCGAG	HDTYKVEYHPQ	15823	729	0.003147	0.4184
29	lzeyGCH	CCGTTCACAACTGGCAGGACCACCCCCAGAACGCT	PFNTWQDHPQNA	15057	710	0.003065	0.9309
30	KCJXAJU	AACTACAAGAACTGGGCGAACCACCCCCAGAACTAC	NYKNWANHPQNY	15801	681	0.00294	1.6467
31	lUJFIJAAT	AACTACAAGAACTGGGTGAGCCACCCCCAGAACTAC	NYKNWVSHPQNY	15801	679	0.002931	1.3788
32	JJFJAGJAE	AACTTGTGCGAACTGGCTGAACCACCCCCAGAACAGA	NLSNWLNHPQNR	16012	624	0.002693	1.5393
33	lRHAGjBQ	AACTTGTCAAACCTGGCTGAACCACCCCCAGAACAGG	NLSNWLNHPQNR	16012	592	0.002555	1.7595
34	UwRJQE	ACGAAGGCGAACTGGGCGAGCACCCCGAGAACACC	TKANWATHPQNT	15696	554	0.002391	0.8102
35	uFIDAsA	AACTTGTGCGAACTGGCTGAGCCACCCCCAGAACAGG	NLSNWLNHPQNR	16012	553	0.002387	1.3703
36	lUJIAVGIB	ACGCGCACGTCGTGGTTGAACCACCCCCAGAACC	TRTSWLNHPQNR	15566	519	0.00224	0.581
37	lBvGKJFJG	AGCTTGTGCGAACTGGCTGAACCACCCCCAGAACAGG	SLSNWLNHPQNR	16012	501	0.002163	1.3507
38	lHFJNji	ACGTACACGAGTGGCGGAACCACCCCCAGAACTCG	TYTTWRNHPQNS	11957	487	0.002102	0.6208
39	JlQIGIGF	AACTACAAGAACTGGGTGAACCACCCCCAGAACTAC	NYKNWVDHPQNY	15801	398	0.001718	1.5843
40	lSHsGIFJH	ACGTACGAGGCGTGGCTCAGCACCCCGAGAATGG	TYEAWLTHPQNW	11588	386	0.001666	0.4434
41	lrfJkFIH	ACGTTTCGACAACCTGGAGCAGGCACCCCCAGAACCAT	TFDNWSRHPQNH	15743	381	0.001645	0.8789
42	lScllFIA	AACTTGTGCGAACTGGCTGAACCACCCCCAGAACAGG	NLSDWLNHPQNR	16012	374	0.001614	1.1988
43	lIvJrGHJ	AACTACAAGAACTGGGTAAACCACCCCCAGAACTAC	NYKNWVNHPQNY	15801	372	0.001606	1.695
44	lBlJGJpGH	AACTACAAGAGCTGGGTGAACCACCCCCAGAACTAC	NYKSWVNHPQNY	15801	368	0.001588	1.2841
45	lRGVHAHX	AACTACAGGAACCTGGGTGAACCACCCCCAGAACTAC	NYRNWVNHPQNY	15801	363	0.001567	1.0561
46	kISAY	AACTATAAGAACTGGGTGAACCACCCCCAGAACTAC	NYKNWVNHPQNY	15801	350	0.001511	1.9105
47	lBtAJGIAD	AACTACAAGGACTGGGTGAACCACCCCCAGAACTAC	NYKDWVNHPQNY	15801	349	0.001506	1.0902
48	lJJAGADxE	TCGTACGACAGTGGCAGAACCCCGAGAATTTG	SYDTWQNHPQNL	11957	337	0.001455	0.4048
49	lJHhIh	AACTACAAAACCTGGGTGAACCACCCCCAGAACTAC	NYKNWVNHPQNY	15801	331	0.001429	1.4426
50	lQZINKD	AACTACAAGAACTGGGTGAACCACCCCCAGAACTGC	NYKNWVNHPQNC	15801	314	0.001355	1.3115

Table H3b: Streptavidin #2 Top 100 DNA/peptide sequences from AptaTOOLS

Rank	ID	Aptamer sequence	Peptide	Cluster ID	Count	Frequency	Enrich
51	BQJIIIAAC	ACGCGGTACCAGTGGAAACACCCACCCCAAGACGTA	TRYQWNTHPQNV	15817	311	0.001342	0.7706
52	BxhsAJB	GACTACGAGAAGTGGGTGAACCACCCCAAGACTAC	DYENWVNHPQNY	15801	308	0.001329	1.3062
53	JJFJMGHG	AACTTGTGCGAACTGGGTGAACCACCCCAAGACGGG	NLSNWLNHPQNG	16012	308	0.001329	1.2129
54	1MaHIGFJJ	AACTACAAGAAGTGGGTGAACCACCCCAAGACTAC	NYKNWVNHPQNY	15801	290	0.001252	1.6483
55	VrQGv	AAGTTCAGGTGATCAACTGGTTCACGCACCCGACAG	KFRLINWFTHPQ	15377	290	0.001252	0.9461
56	1SHTcq	GACTTGTGCGAACTGGGTGAACCACCCCAAGACAGG	DLSNWLNHPQNR	16012	286	0.001235	1.2319
57	1BSFJBwy	ACGGTCAACAAGTGGAGGACCCACCCCAAGACGCT	TVNNWRTHPQNA	14953	279	0.001204	1.594
58	1JLBUlJJ	CACGACACCTACCGCCCGCGAAGTACCACCCCAAG	HDTYRPPKYHPQ	16176	271	0.00117	0.7183
59	BSfAoDA	AAGTTCAGCTGATCAACTGGTTCACGCACCCGACAG	KFKLINWFTHPQ	15377	268	0.001157	1.9441
60	1QZICJJo	AACTACAAGAAGTGGGTGAACCACCCCAAGACAC	NYKNWVNHPQNH	15801	268	0.001157	1.4207
61	1GJBxJJM	AACTTGTGCGAACTGGGTGAACCACCCCAAGACAGG	NLSNWLNHPQNR	16012	241	0.00104	1.3913
62	BSUCzu	AACTTGTGCGAGTGGGTGAACCACCCCAAGACAGG	NLSSWLNHPQNR	16012	241	0.00104	1.2654
63	JHhtgn	ACGCGGACCAAGTGGAAACACCCACCCCAAGACGTC	TRHQWNTHPQNV	15817	236	0.001019	1.0125
64	BfBfSB	AGCTACAAGAAGTGGGTGAACCACCCCAAGACTAC	SYKNWVNHPQNY	15801	234	0.00101	1.0618
65	1HHIaCHDB	AACTGGTACAAGTGGGCCATCCACCCCAAGACACG	NWYNWAIHPQNT	15291	229	9.88E-04	0.7629
66	CjLgAJ	AACTACAAGAAGTGGGTGAACCACCCCAAGACTAC	NYKNWVNHPQNY	15801	227	9.80E-04	1.4385
67	1ADvCHt	AACTTGCAGAACTGGGTGAACCACCCCAAGACAGG	NLFPNWLNHPQNR	16012	224	9.67E-04	1.4034
68	BgIJDhU	AACTTGTGCGAACTGGGTGAACCACCCCAAGACAGG	NLSNWLNHPQNR	16012	215	9.28E-04	1.4544
69	1QUCHCqD	AACTACAAGAAGTGGGTGAACCACCCCAAGACTAC	NYKNWVNHPQSY	15801	212	9.15E-04	1.2568
70	BQJIIHphD	ACGCGGTACCAGTGGAAACACCCACCCCAAGACGCC	TRYQWNTHPQNA	15817	211	9.11E-04	0.8163
71	1ejJIAGD	AACTACAAGAAGTGGGTGAACCACCCCAAGACTAC	NYKNWVNHPQNY	15801	210	9.06E-04	1.4472
72	1GHBzFIJI	AACTACAAGAAGTGGGTGAACCACCCCAAGACTAC	NYKNWVNHPQNY	15801	208	8.98E-04	1.9772
73	BiKqUJ	AACTACAAGAAGTGGATGAACCACCCCAAGACTAC	NYKNWVNHPQNY	15801	203	8.76E-04	1.4441
74	LBgKRI	CCGTTCAACACCTGGCAGAACCCCAAGACGCC	PFNTWQNHHPQNA	15057	201	8.68E-04	0.9764
75	LeJFzIC	AACCAACAAGAAGTGGGTGAACCACCCCAAGACTAC	NHKNWVNHPQNY	15801	198	8.55E-04	1.4085
76	RGHHHAiv	AACTTGTGCGAACTGGCTAAACCACCCCAAGACAGT	NLSNWLNHPQNS	16012	197	8.50E-04	1.8889
77	1BvCHitc	ACGCTCAGCAGTGGGAGACCCCAAGACAGCTG	TLSDWETHPQNL	15278	193	8.33E-04	0.8616
78	NqABev	AAGTTCAGAACTGGAGACCCCAAGACAGCTC	NYKNWVNHPQNV	15872	191	8.24E-04	1.1832
79	BRCGIJAo	ACGTTTGACAAGTGGAGCAGGCACCCCAAGACCAT	TDPNWSRHPQNH	15743	187	8.07E-04	1.574
80	1OwAgHAD	AACTACAAGAAGTGGGTGAACCACCCCAAGACTAC	NYKNWVNHPQDY	15801	186	8.03E-04	1.4243
81	BjFJAor	AAGTTCGACGACTGGAAGCACCACCCCAAGACATC	KFDDWKHHPQNI	15872	177	7.64E-04	1.0724
82	GixTA	AAGTTCAGTGTGATTAAGTGGTTCACGCACCCGACAG	KFKLINWFTHPQ	15377	175	7.55E-04	1.7384
83	1RHoFpC	AACTTGAACAGCTGGCTGACCCACCCCAAGACCTG	NLNSWLTHPQNL	12473	175	7.55E-04	0.8576
84	SGbZJw	CCATTCAACACCTGGCAGGACCCCAAGACGCC	PFNTWQDHPQNA	15057	174	7.51E-04	0.6526
85	1BkjHrFJ	AGCTACGAGAAGTGGACCATGCACCCCAAGACGTC	SYENWVNHPQNV	10746	172	7.42E-04	0.6079
86	RGHHHAie	AACTTGTGCGAACTGGCTAAACCACCCCAAGACAGC	NLSNWLNHPQNS	16012	171	7.38E-04	2.1426
87	1JDrBQS	ACGCGATACCAGTGGAAACACCCACCCCAAGACGTC	TRYQWNTHPQNV	15817	170	7.34E-04	1.2414
88	1vGCJAJJF	AACCTGTGCGAACTGGGTGAACCACCCCAAGACAGG	NLSNWLNHPQNR	16012	169	7.29E-04	0.9965
89	1HHJDBpH	AAGTTCAGCAGTGGAAAGCACCACCCCAAGACATC	KFNDWKHHPQNI	15872	158	6.82E-04	1.1462
90	1IABvpGH	AACTGCAAGAAGTGGGTGAACCACCCCAAGACTAC	NCKNWVNHPQNY	15801	158	6.82E-04	1.3299
91	1UCHJjAz	AACTACGAGAAGTGGGTGAGCCACCCCAAGACTAC	NYENWVNHPQNY	15801	157	6.78E-04	1.4426
92	KJHMHGIC	AACTACGAGAAGTGGGCGAACCCCAAGACTAC	NYENWVNHPQNY	15801	157	6.78E-04	1.6646
93	KAQqAe	AACTTGTGCGAACTGGGTGAACCACCCCAAGACAGG	NLSNWLNHPQSR	16012	156	6.73E-04	1.2199
94	1MIDAZDo	ACGCGGTACCAGTGGAGCACCACCCCAAGACGTC	TRYQWSTHPQNV	15817	155	6.69E-04	0.6018
95	NqAAHIF	AAGTTCAGCAGTGGAAAGCACCACCCCAAGACATA	KFNDWKHHPQNI	15872	149	6.43E-04	1.1489
96	1BfXSIK	AATTACAAGAAGTGGGTGAACCACCCCAAGACTAC	NYKNWVNHPQNY	15801	149	6.43E-04	1.5499
97	1GmAPIz	AACTTGTGCGAACTGGCCGAACCACCCCAAGACAGG	NLSNWLNHPQNR	16012	148	6.39E-04	1.3714
98	1AHJdgh	AACTCGTCAAGTGGGTGAACCACCCCAAGACAGG	NSSNWLNHPQNR	16012	147	6.35E-04	1.3854
99	1tRBav	ACGACGCGGACTGGGTGGTCCACCCCAAGACGTC	TQRDLVHPQNV	12507	147	6.35E-04	0.4942
100	1QZIDSGI	AACTACAAGAAGTGGGTGAACCACCCCAAGACAA	NYKNWVNHPQNN	15801	146	6.30E-04	1.731

Table H4a: Streptavidin #2-chymotrypsin Top 100 DNA/peptide(enrich) from AptaTOOLS

Rank	ID	Aptamer sequence	Peptide	Cluster ID	Count	Frequency	Enrich
1	1NqDBSHG	AGGTTCAAGCTGATCAACTGGTTCACGCACCCCGCAG	RFKLNWFTHPQ	15377	14	6.12E-05	15.638
2	BhhJHGt	AAGTTCAGGTTGATCAACTGGTTCACACACCCCGCAG	KFKLNWFTHPQ	15377	11	4.81E-05	12.287
3	1QZIDQGF	AACTACAAGAACTGGGTGAACCACCCCGAGAAGCAGT	NYKNWVNHPQNS	15801	11	4.81E-05	12.287
4	1PzJQHIF	AACTACGAGAAGCTGGGTGAACCACCCCGAGAAGTCC	NYENWVNHPQNS	15801	22	9.62E-05	12.287
5	1TRJDiC	CTAGTTCGACTGCATGATAGCATCTCTTCTTAA	LAALHDSILLK	7657	10	4.37E-05	11.17
6	GJGFJCGp	GAGTTCAAATTGATCAACTGGTTCACGCACCCCGCAG	EFKLNWFTHPQ	15377	10	4.37E-05	11.17
7	VATqAH	CACGACACCTACCACCCCGCGAAGTACCACCCCGAG	HDTYRPPKYHPQ	16176	10	4.37E-05	11.17
8	1JLVCqB	CACGACACCTACCACCCCGCGAAGTACCACCCCGAG	HDTYRPPKYHPQ	16176	10	4.37E-05	11.17
9	BrHwrb	AAGTTCAGCTGACCAACTGGTTCACGCACCCCGCAG	KFKLNWFTHPQ	15377	9	3.94E-05	10.053
10	UuJFJIV	AACTACGAGAAGCTGGGTGAACCACCCCGAGAAGTAC	NYENSVNHPQNY	15801	9	3.94E-05	10.053
11	HIQGjM	AAGTTCAGCTGGTCAACTGGTTCACGCACCCCGCAG	KFKLVNWFTHPQ	15377	9	3.94E-05	10.053
12	UISmAh	AAGTTCAGGTTGATCAACTGGTTCACGCACCCCGCAG	KFKLNWFTRPQ	15377	8	3.50E-05	8.9363
13	DxIDzGI	AAGTTCAGTGTGATCAACTGGTTCACGCACCCCGCAG	KFKLINWSTHPQ	15377	15	6.56E-05	8.3777
14	1QGAHAuHD	CACGACACCTACCACCCCGCGAAGTACCACCCCGAG	HDTYRPPKSHHPQ	16176	7	3.06E-05	7.8192
15	JHIHII	ACGTTTCGAGACTGGAGCAGGCACCCCGAGAAGCCG	TFDDWSRHPQNR	15743	7	3.06E-05	7.8192
16	LAHBqAD	AACTATGAGAAGCTGGGTGAACCACCCCGAGAAGTAT	NYENWVNHPQNY	15801	7	3.06E-05	7.8192
17	1IJvAfA	CACTACGAGAAGCAGCTCTTGCCCTCGCACCCCGCAG	HYENELLPSPHPQ	15706	7	3.06E-05	7.8192
18	NrUoGG	AAGTTCAGTGTGAGCAACTGGTTCACGCACCCCGCAG	KFKLSNWFTHPQ	15377	34	1.49E-04	7.5958
19	GakHC	GAGTTCAGCTGATCAACTGGTTCACGCACCCCGCAG	EFKLNWFTHPQ	15377	20	8.75E-05	7.4469
20	RlDdIJB	AAGTTCAGATTGATCAACTGGTTCACGCACCCCGCAG	KFKLNWFTHPQ	15377	53	2.32E-04	7.4003
21	1BMxKHZ	ACGCAGTACCAGTGAATACCCACCCCGAGAAGCTC	TOYQWNTHPQNV	15817	6	2.62E-05	6.7022
22	1JGIjBUB	CACGACACCTACCACCCCGCGAAGTACCACCCCGAG	HDTYRPPKYHTQ	-1	6	2.62E-05	6.7022
23	1DKJwJJC	ACGCAGTACCAGTGAACACCCACCCCGAGAAGCTC	TOYQ*NTHPQNV	15817	6	2.62E-05	6.7022
24	DyDhGm	AAGTTCAGGTTGATCAACTGGTTCACGCACCCCGCAG	KFKLNWLTHPQ	15377	6	2.62E-05	6.7022
25	1yHpAhr	AACTACAAAAGCTGGGTGAGCCACCCCGAGAAGTAC	NYKNWVSHPQNY	15801	6	2.62E-05	6.7022
26	BQFIIQw	CACGACACCTACAAGCCCTTGCCGGAGCACCCCGCAG	HDTYKPLPEHPQ	14413	6	2.62E-05	6.7022
27	1rAiBoB	ACGTTTCGACAAGCTGGAGCAGGCACCCCGAGAGCCAT	TFDNWSRHPQSH	15743	6	2.62E-05	6.7022
28	soHDn	ACGTACGACGATTGGAAGTCGCACCCCGAGAAGCAG	TYDDWKSHPQNT	14860	6	2.62E-05	6.7022
29	1jIIAUCJ	AACTACGAGAAGCTGGGTGAACCACCCCGAGAAGTAT	NYENWVNHPQNY	15801	6	2.62E-05	6.7022
30	1HCsDiz	CACGACACCTGCCCGCCCGCGAAGTACCACCCCGAG	HDTYRPPKYHPQ	16176	6	2.62E-05	6.7022
31	1SQAIWGI	AACTACGAGAAGCTGGGTGAACCACCCCGAGAAGTAT	NYENWVNHTQNY	15801	6	2.62E-05	6.7022
32	1JKGoEu	CACGACACCTACCACCCCGCGAAGTACCACCCCTCAG	HDTYRPPKYHPQ	-1	6	2.62E-05	6.7022
33	1QZINjJ	AACTACAAGAACTGGGTGAACCACCCCGAGAAGTGG	NYKNWVNHPQNW	15801	6	2.62E-05	6.7022
34	1UCIBiRc	ACGCTCGACAAGCTGGAGCAGGCACCCCGAGAAGCAT	TLDNWSRHPQNH	15743	6	2.62E-05	6.7022
35	SJjAcXj	AGATTCGTCACAGAGAAGCTGGTTCGAGCACCCCGAG	RFVNNWFEPHPQ	-1	6	2.62E-05	6.7022
36	1SmTAHN	TACGACACCTACACGCCCTTGCCGGAGCACCCCGCAG	YDYTPPLPEHPQ	14413	6	2.62E-05	6.7022
37	SJvGIJAH	AGGTTCAAGTTGATAAACTGGTTCACGCACCCCGCAG	RFKLNWFTHPQ	15377	6	2.62E-05	6.7022
38	CwbACg	ACGTACGACGAGTGAAGTCGCACCCCGAAAACACG	TYDEWKSHPQNT	14860	6	2.62E-05	6.7022
39	1BVGgJKF	AAGTTCAGGACTGGAAGCACCACCCCGGAACATC	KFSDWKHHPRNI	-1	6	2.62E-05	6.7022
40	BoIJErII	AACTACGAGAAGCTGGATGAACCACCCCGAGAAGTAT	NYENWVNHPQNY	15801	6	2.62E-05	6.7022
41	1KupJGf	AACTTGTTCGACTGGCTAAACCACCCCGAGAAGCAGT	NLSDWLNHPQNS	16012	6	2.62E-05	6.7022
42	1BTIfBmE	CACGACACCTACCCTCCGCCGGAAGTACCACCCCGAG	HDTYRPPKYHPQ	16176	6	2.62E-05	6.7022
43	1QADSNg	AACTACCAGAAGCTGGGTGAACCACCCCGAGAAGTAT	NYQNWVNHPQNY	15801	6	2.62E-05	6.7022
44	1HpGHCHAI	CACGACACCTACGTGACCAACTGGCACCCCGCAG	HDTYVTSNWHHPQ	-1	6	2.62E-05	6.7022
45	1PzJCJCHI	AACTACGAGAAGCTGGGTGAACCACCCCGAGAAGGCC	NYENWVNHPQNA	15801	24	1.05E-04	6.7022
46	KpDptG	AAGTTCAGGCTGATCAACTGGTTCACGCACCCCGCAG	KFKLNWFTHPQ	15377	30	1.31E-04	6.7022
47	VpfiHF	AAGTTCAGGTTGATCAACTGGTTCACGCACCCCGCAG	KFKLNWFTHPQ	15377	11	4.81E-05	6.1437
48	BSDuFIE	AAGTTCAGCTGATCAACTGGTTCACGCACCCCGCAG	KFKLNWFTHPQ	15377	11	4.81E-05	6.1437
49	1BtCfxe	ACACAGTACCAGTGAACACCCACCCCGAGAAGCTC	TOYQWNTHPQNV	15817	11	4.81E-05	6.1437
50	DhDkHu	AACTTGTTCGAAGCTGATGAACCACCCCGAGAAGCAG	NLSN*LNHPQNS	16012	11	4.81E-05	6.1437

Table H4b: Streptavidin #2-chymotrypsin Top 100 DNA/peptide(enrich) from AptaTOOLS

Rank	ID	Aptamer sequence	Peptide	Cluster ID	Count	Frequency	Enrich
51	SHJGJDAAF	AAGTTCAAATTGATCAACTGGTTCACGCACCCGCAG	KFKLINWFTHPQ	15377	363	0.001587	6.052
52	lPzJDrAD	AACTACGAGAAGTGGGTGAACCACCCCGAGAAGT	NYENWVNHPQNS	15801	5	2.19E-05	5.5852
53	BqGxJHHG	GACTACAAGAAGTGGGTGAACCACCCCGAGAATTAC	DYKNWVNHPQNY	15801	5	2.19E-05	5.5852
54	lQIaJpAI	AACCACAAGGACTGGGTGAACCACCCCGAGAATAC	NHKDWNHPQNY	15801	5	2.19E-05	5.5852
55	JlQHauI	AACTACAAGAAGTGGGTGAACCACCCCGAGAACAC	NYKNWVDHPQNH	15801	5	2.19E-05	5.5852
56	GgHGmE	AAGTTCAGTTGATTAAGTGGTTCACGCACCCACAG	KFKLINWFTHPQ	15377	5	2.19E-05	5.5852
57	JJFJCrJG	AACTTGTGCGAAGTGGGTGAACCACCCCGAAGTAT	NLSNWLNHPQNY	16012	5	2.19E-05	5.5852
58	BwCtDBX	AACTACAAGAAGTGGGTGAACCACCCCGAAGTAT	NYKN*VSHPQNY	15801	5	2.19E-05	5.5852
59	lGuaDsB	AGGTTCAAATTGATCAACTGGTTCACGCACCCGCAG	RFKLINWFTHPQ	15377	5	2.19E-05	5.5852
60	lHGaBep	ACGGTCAACAAGTGGGTGAACCACCCCGAGAAGCC	TVNNWRTHPQNA	14953	5	2.19E-05	5.5852
61	RGGGAHTJ	CACGACACCTACCAGGTCTGCGAGTACCACCCCGAG	HDTYQVCEYHPQ	-1	5	2.19E-05	5.5852
62	SIkxAg	AAGTTCAAATTGATCAACTGGTTCACGCACCCGCAG	KFKLINWSTHPQ	15377	5	2.19E-05	5.5852
63	lLTIOSE	AAGTTCAGTTGATCAATTGGTTCACGCACCCGCAG	KFKLINWFTHPQ	-1	5	2.19E-05	5.5852
64	lTcjBtN	AACTACAAGAAGTGGGTGAACCACCCCGAAGTAT	NYKNWVSHPQDY	15801	5	2.19E-05	5.5852
65	HCfcJvB	AAGTTCAGTTGGTCAACTGGTTCGCGCACCCGCAG	KFKLVNWFAPHQ	-1	5	2.19E-05	5.5852
66	BQAADqb	AAGTCCAAGGACTGGAAGCACCCCGAGAATCATT	KNSNDWKHHPQNI	15872	5	2.19E-05	5.5852
67	BzvJGGJE	AACTTGCAGAACTGGCTAAACCACCCCGAGAAGT	NLPNWLNHPQNS	16012	5	2.19E-05	5.5852
68	JJFJAICI	AACTTGTGCGAAGTGGGTGAACCACCCCGAAGCCG	NLSNWLNHPQNR	16012	5	2.19E-05	5.5852
69	oIIXmD	ACGCGGTACCAGTGGAAACACCCACCCAGAACGTT	TRYQWNTHHQNV	15817	5	2.19E-05	5.5852
70	lBjAIAHRJ	AACTACTAGAAGTGGGTGAACCACCCCGAAGTAT	NY*NWVNHPQNY	15801	5	2.19E-05	5.5852
71	lOErHAAI	CACGACACTTACCAGCCCGGAAGTCCCACCCCGAG	HDTYRPPKSHPQ	-1	5	2.19E-05	5.5852
72	QzICICd	GACTACAAGAAGTGGGTGAACCACCCCGAAGTAT	DYKNWVNHPQDY	15801	5	2.19E-05	5.5852
73	lTGHIGBN	AACGACACCTACCAGCCCGCGAAGTACCACCCCGAG	NDTYRPPKYHPQ	-1	5	2.19E-05	5.5852
74	bhuKH	AACTTGTGCGAAGTGGGTGAACCACCCCTAGAACAGC	NLSNWLNHP*NS	16012	5	2.19E-05	5.5852
75	lKupJGw	AACTTGTGCGAAGTGGGTAAACCACCCCGAGAAGC	NLSDWLNHPQNS	-1	5	2.19E-05	5.5852
76	TeIFJdI	AACTTGTCAAGTGGGTGGACCACCCCGAGAAGCAG	NLSNWLNHPQNR	16012	5	2.19E-05	5.5852
77	lNRCIMIE	AAGCTCAGTTGATCAACTGGTTCACGCACCCGCAG	KLRLINWFTHPQ	15377	5	2.19E-05	5.5852
78	GBsHIJI	CACGACACCTACAGCCCTTGCCGGCGCACCCGCAG	HDTYTPLPAHPQ	14413	5	2.19E-05	5.5852
79	lMxDsIa	ACGTTTGACAAGTGGAGCAGCCACCCCGGAACCAT	TFDNWSRHPRNH	15743	5	2.19E-05	5.5852
80	lCJCxJkE	CCGTTCAACACAGTGGGTGAACCACCCCGAGAAGCC	PFNTWQDHPQNA	15057	5	2.19E-05	5.5852
81	JKCJIET	AAGTTCACGACTGCAAGCACCCCGAGAATCATC	KFNDCKHHPQNI	15872	5	2.19E-05	5.5852
82	lGhxIGAC	CACGACACCTACAGCCCTTGCCGGAGCACCCGAAG	HDTYTPLPEHPK	-1	5	2.19E-05	5.5852
83	lPzIHGJB	AACTACGAGAAGTGGGTGAACCACCCCGAGAATAC	NYENWVNHRQNY	15801	5	2.19E-05	5.5852
84	GUHGHHM	AACTACAAGGACTGGGTAAACCACCCCGAGAATAC	NYKDWVNHPQNY	15801	5	2.19E-05	5.5852
85	JAHHJIT	AACGTTGTGCGAAGTGGGTAAACCACCCCGAGAAGC	NVSNWLNHPQNR	16012	5	2.19E-05	5.5852
86	lKgJIJGD	AACTTGTGCGAATGGGTGAGCCACCCCGAGAAGCAG	NLSNWLNHPQNR	16012	5	2.19E-05	5.5852
87	lIrkjC	CACGACACCTACAAGGTCTGCGTGTACCACCCCGAG	HDTYKVCVYHPQ	15823	5	2.19E-05	5.5852
88	SIDwICw	AAGTTTAGTTGATCAACTGGTTCACGCACCCGCAG	KFRLINWFTHPQ	15377	5	2.19E-05	5.5852
89	lLiCIXc	GACTATAAGAAGTGGGTGAACCACCCCGAGAATAC	DYKNWVNHPQNY	15801	5	2.19E-05	5.5852
90	lAIGoA	AACTTGTGCGAATGGGTGAACCACCCCGAGAAGT	NLSNWLNHPQNS	16012	10	4.37E-05	5.5852
91	lQZINAAD	AACTACAAGAAGTGGGTGAACCACCCCGAGAATTT	NYKNWVNHPQNF	15801	10	4.37E-05	5.5852
92	JJFJAGHIC	AACTTGTGCGAAGTGGGTGAACCACCCCGAGAAGC	NLSNWLNHPQNT	16012	10	4.37E-05	5.5852
93	lrpIDyE	CACCACACCTACAGCCCTTGCCGGAGCACCCGCAG	HHTYTPLPEHPQ	-1	10	4.37E-05	5.5852
94	GixTA	AAGTTCAGTTGATTAAGTGGTTCACGCACCCGCAG	KFKLINWFTHPQ	15377	543	0.002374	5.4644
95	lScmBrY	AAGTTCAGTTGAACAAGTGGTTCACGCACCCGCAG	KFKLNNWFTHPQ	15377	80	3.50E-04	5.2566
96	lPzJbQH	AACTACGAGAAGTGGGTGAACCACCCCGAGAAGC	NYENWVNHPQNH	15801	14	6.12E-05	5.2128
97	lBNAAAHGR	CACGACACCTACCAGCCCGGAAGTACCACCCCGAG	HDTYRPPKYHPQ	16176	55	2.41E-04	5.1197
98	lBMrHJGW	CACCACAGTACGTATCATCGGTACCACCCCGAG	HQYVYIGYHPQ	8225	9	3.94E-05	5.0266
99	DtAJrIJ	AAGTTCAGTTGATCAACTGGTTCACGCACCCGCAG	KFKLINWFTHPQ	15377	27	1.18E-04	5.0266
100	lHACJpB	AACTTGTAGAAGTGGGTGAACCACCCCGAGAAGT	NL*NWLNHPQNS	16012	9	3.94E-05	5.0266

Library	Random region	Linker	His ₆
MC (NNS ₂) C (NNS ₁₀)	MCXXCXXXXXXXXXX	GSGSLG	HHHHHH
MC (NNS ₄) C (NNS ₈)	MCXXXXCXXXXXXXX	GSGSLG	HHHHHH
MC (NNS ₆) C (NNS ₆)	MCXXXXXXCXXXXXX	GSGSLG	HHHHHH
MC (NNS ₈) C (NNS ₄)	MCXXXXXXXXXCXXXX	GSGSLG	HHHHHH
MC (NNS ₁₀) C (NNS ₂)	MCXXXXXXXXXXCXX	GSGSLG	HHHHHH

Table H6 Fixed-cysteine monocyclic libraries used in our previous selection against XRCC4 (unpublished results).

a)

Tube #	Title	Conc (ng/uL)	vol (uL)	% of total reads
1	A-4	485	10	31
2	B-4	457	10	8
3	A-5	554	10	16
4	B-5	562	10	4
5	A-6	526	10	10
6	B-6	526	10	2
7	A-7	308	10	8
8	B-7	477	10	1
9	A-8	627	10	6
10	A-9	545	10	4

A Bicyclic
B Monocyclic

b)

Tube #	Title	Conc (ng/uL)	vol (uL)	% of total reads
2	SA-5	1995	10	43
3	SA-6	1765	10	9
4	SA-7	1945	10	6
5	SA-8	2100	10	4
6	SA-9	2365	10	4
7	SA-C-8	2245	10	4
8	SA-C-9	2260	10	4
9	X4-10	1890	10	10
10	X4-11	2025	10	10

SA streptavidin #2 94
SA-C streptavidin #2-chymotrypsin
X4 XRCC4

Table H7 DNA mixing ratios for sequencing. a) DNA from scaffold-based selection against streptavidin (Chapter 4). Title represents each of the selections (A=bicyclic, B=monocyclic), followed by the round the specified DNA was used as the template (eg. A-4 represents the bicyclic selection, round 4 template = round 3 results). B) Ratios for the streptavidin # 2 (SA), streptavidin #2-chymotrypsin (SA-C), from Chapter 5. XRCC4 (X4) selection (Chapter 6) was sequenced concurrently.

a)

	A	B	C	D	E	F
1	codon	AAs omitted	# random AAs	% Cys/codon	% Phe/codon	calc. diversity
2	NNS	none	12	3.125	3.125	1.02E+13
3	NNY	M,E,K,N,W	11	6.25	6.25	5.77E+11
4						1.14E+13
5						unique molecules
6		not C or F	not C or F			
7	1st C or F	per codon	per peptide	one C or F	% with one C or F	
8	NNS	0.96875	0.6832	0.3168	32	
9	NNY	0.9375	0.4917	0.5083	51	
10		not C or F				
11	2nd C	per codon	not another C	another C	two C	% with 2 C
12	NNS	0.96875	0.7052	0.2948	0.0934	9.3
13	NNY	0.9375	0.5245	0.4755	0.2417	24

b)

codon	AAs omitted	# random AAs	% Cys/codon	% Phe/codon	calc. diversity
NNS	none	12	3.125	3.125	=20^10
NNY	M,E,K,N,W	11	6.25	6.25	=15^10
					=(F3*2)+F2 unique molecules
1st C or F	not C or F per codon	not C or F per peptide	one C or F	% with one C or F	
NNS	=(100-D2)/100	=B8^C2	=(1-C8)	=D8*100	
NNY	=(100-D3)/100	=B9^C3	=(1-C9)	=D9*100	
2nd C	not C or F per codon	not another C	another C	two C	% with 2 C
NNS	=(100-D2)/100	=B12^11	=(1-C12)	=D8*D12	=E12*100
NNY	=(100-D3)/100	=B13^10	=(1-C13)	=D9*D13	=E13*100

Table H8 Streptavidin # 2 and XRCC4 libraries: Cyclizable codon probability calculations. a) Calculations of the probability that the indicated cyclizable residues will appear in the random region of the libraries. b) Formulas with coinciding cell references (column-row) used to calculate probabilities reported in a). Data was generated using Microsoft Excel 2016.