Phage Display as a Combinatorial Chemistry Platform for Discovery of Chemical Structure-Activity Relationships

By

Vivian Triana Guzman

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

Master of Science

Department of Chemistry

University of Alberta

Abstract

Optimization of chemical reactions, discovery of optimal substrates and determination of substrate scope involve exhaustive screening of conditions as well as measurements of rates and conversions for all series of structurally similar substrates. These systematic screenings used in reaction development and reaction discovery are usually performed on a group of substrates that belong to a region of the chemical space that has been targeted using chemical intuition and empiricism. The number of tests to be performed is often high and unbiased approaches that avoid contribution of the scientist's rationalization and insight, demand for an even higher number of experiments. One-well-one-experiment approaches are the best at replicating real reaction conditions but require many work hours or expensive robotics and automation. Several technologies have been developed for one-well-multiple-experiments screenings with the aim of reducing time and cost of screening, however, sophisticated technologies for analysis of mixtures are necessary and real one-substrate conditions are not simulated.

This thesis presents a new methodology for Structure-Activity Relationship discovery that uses phage display as a combinatorial chemistry platform for one-well-multiple-experiments screenings of substrates. As a first step, conditions for Wittig reaction between an ester stabilized ylide and libraries of glyoxaldehydes displayed on the pIII protein of M13 phage were optimized. The Wittig reaction installed an N-terminal ester of maleic/fumaric acid on phage libraries that underwent both Michael addition and Diels-Alder cycloaddition with cyclopentadiene. After demonstrating its suitability to be applied on phage, Structure-Activity Relationship of the Wittig reaction between ester stabilized ylide and glyoxaldehydes *in water* was performed using focused libraries that can be entirely visualized using Illumina sequencing. Enrichment and depletion of specific residues in several positions of the sequences allowed us to find slow, medium and fast

reacting peptides in the Wittig reaction covering a range of 50-fold difference in rate. Sequences lacking hydrogen-bond donors in the two positions adjacent to the aldehyde (Ald-Pro-Pro-X-X) showed up to 16-fold decrease in rate compared to average reactivity, while sequences with two Trp residues in those two positions (Ald-Trp-Trp-X-X) showed up to 5-fold increase in reactivity. These observations demonstrated the importance of hydrogen-bond stabilization of the oxaphosphetane intermediate. Preliminary studies on using phage display as a combinatorial chemistry platform for discovery of cyclization-prone sequences were performed. First, we installed the α,β-unsaturated ester *via* the Wittig reaction and then selected for sequences able to undergo intramolecular attack by nucleophiles present in the peptide sequence. We were able to find three sequences prone to cyclization and we isolated two of the formed adducts, demonstrating the feasibility of using phage display as a tool for discovery of optimal substrates in a reaction.

Preface

Chapter 1 was written as an early stage draft for an invited review on 'Phage display methods for discovery of bioactive macrocycles" for the *Chemical Reviews* journal. I reviewed the literature that describes different combinatorial platforms for production and screening of libraries of macrocycles, making emphasis on genetically encoded techniques and more specifically, methodologies that apply phage display as a platform. Advantages and disadvantages of the phage display platform were mentioned as well as challenges for further development of this technology and potential strategies to surmount them.

Chapter 2 is based on published work and reproduced with permission of R. Derda, "Tandem Wittig/Diels–Alder diversification of genetically encoded peptide libraries", Triana, V.; Derda, R. *Org. Biomol. Chem.* **2017**, *15* (37), 7869. I was responsible for synthesis of a biotin tagged ylide used for development and optimization of the Wittig reaction on N-terminal glyoxaldehyde synthetic peptides *in water*. I applied the optimal conditions to functionalize peptides displayed on phage with 40-70% conversion. I corroborated the presence of the olefin installed after the Wittig reaction by further diversification with Michael addition and Diels-Alder transformations.

Chapter 3 is based on non-published work initiated by R. Derda. I was responsible for library functionalizations and selections. Neon green phage used for internal standard were made by Dr. Nicholas Bennet and used with his authorization. Analysis of Illumina sequencing was performed with help of Dr. Derda using scripts developed by him. Figure 3-4C and the script used for generation of 20:20 plots were reported by Derda and coworkers in He, B.; Tjhung, K. F.; Bennett, N. J.; Chou, Y.; Rau, A.; Huang, J.; Derda, R. *Sci. Rep.* **2018**, *8* (1), 1214 and used with his authorization. I was responsible for measurement of kinetics for the Wittig reaction on synthesized peptides as well as rates for hydrazine ligation and measurement of E/Z selectivities. R. Derda has entire credit for the idea of using phage display in SAR development.

Chapter 4 is based on non-published work initiated by R. Derda. I was responsible for library functionalization and selections. Analysis of Illumina data was performed using matlab scripts developed by R. Derda. I was responsible for measurements of Wittig reaction rates and

isolation of “cyclization adducts.” 2D NMR measurements of adducts were performed with great help of Mark Miskolzie.

Acknowledgements

First and foremost, I want to thank my mother Beatriz and my brother Oswaldo for their unconditional support, love and care. For implanting in me constant critical thinking and for teaching me that my decency, honesty, courage and moral boundaries are not worth sacrificing no matter the benefits or delusions of success. To them I owe everything I am and the notions of what I want to become.

I am thankful to my supervisor Ratmir Derda for accepting me in his group, for his guidance, financial support of my research and for having a very well-equipped laboratory. I thank Dr. Hall and Dr. Lundgren for being members of my committee during these three years. I also thank Dr. Julianne Gibbs for accepting to be my arm's length examiner.

I thank Dr. Nicholas Bennet for the neon green phage. I thank Pavel Kitov and Ying Chow for general training at the beginning of my research in Derda Lab. I thank Gareth Lambkin for training in several instruments at the Biological Services facility. I am very thankful to Bela Reiz for hours of training, invaluable help on methodologies for purification and characterization of compounds and for advice on purchasing columns and HPLC parts. I am also thankful to Dr. Angelica Morales for help in characterization of compounds and for stress releasing conversations. I am very thankful to Dr. Mark Miskolzie for hours of training, measurement, advise and discussion on characterization of peptide-derived molecules.

I am extremely thankful to Dr. Burcin Akgun and Radhika Chakraberty for their friendship, wise advice and unconditional support. Making it through these three years would have not been possible without their company and patience, I am infinitely grateful to both of them. I also want to thank my classmates Rochelin, Hahn and Christopher for their unconditional friendship. Finally, I want to thank my friends Sekou, Yimmie, Lady and Nasra for inspiring me to be fearless and to develop my potential in multiple directions.

Contents

List of Tables

List of Figures

List of Schemes

Abl	Abelson Murine Leukemia Viral Oncogene Homolog
ACA	Anti-Cardiolipin Antibody
ACE	Angiotensin Converting Enzyme
Aha	Azidohomoalanine
Akt2	Serine/Threonine-Protein Kinase 2
Ald	Aldehyde
AMA1	Apical Membrane Antigen 1
ARSs	Aminoacyl-tRNA Synthetases
BCL	B-Cell Lymphoma
Bio	Biotin
BPs	Biosynthetic Precursors
BSA	Bovine Serum Albumin
BSBBA	3,3′-bis (sulfonato)-4,4′-bis(bromoacetamido) azobenzene
BSBCA	3,3′-bis(sulfonato)-4,4′-bis(chloroacetamido)- azobenzene
CaM	Calmodulin
CCR5	C-C Chemokine Receptor Type 5
ClpX	ATP-Dependent Clp Protease ATP-Binding Subunit clpX-like, Mitochondrial
CMG2	Capillary Morphogenesis Gene 2
COSY	Correlation Spectroscopy
CtBP	C-Terminal-Binding Protein
DC	Deep Conversion
DCM	Dichloromethane
DE	Differential Enrichment
DECL	DNA Encoded Chemical Library
DMF	Dimethyl Formamide
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Nucleoside Triphosphate
DOS	Diversity Oriented Synthesis
DTS	DNA Templated Synthesis

EDTA	Ethylenediaminetetraacetic Acid
EGFR	Epithelial Growth Factor Receptor
EpCAM	Epithelial Cell Adhesion Molecule
eq	Equivalent (s)
ERK2	Extracellular-Regulated Protein Kinase 2
ESI	Electrospray Ionization
EtOH	Ethanol
FIT	Flexible *in-vitro* Translation
GE-SAR	Genetically Encoded Structure Activity Relationships
GLP-1	Glucose-Dependent Insulinotropic Peptide
Grb2	Growth Factor Receptor-Bound Protein 2
GSH	Glutathione
GSK	Glaxo-Smith-Klein
GyrA	DNA Gyrase (Subunit A)
h	Hours
HBsAg	Australian Antigen
HCV	Hepatitis C Virus
Hdm2	Human Double Minute 2 Homolog
Hdmx	Human Double Minute X
HECT	Homologous to the E6-AP Carboxyl Terminus
HeLa	Henrietta Lacks
HIF	Hypoxia-Inducible Factor
HILIC	Hydrophobic Interaction Liquid Chromatography
HIV	Human Immunodeficiency Virus
HMBC	Heteronuclear Multiple Bond Correlation
HOMO	Highest Occupied Molecular Orbital
HPLC	High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
HSA	Human Serum Albumin
HSP	Heat Shock Protein
HSQC	Heteronuclear Single Quantum Correlation Experiment

HTS	High Throughput Screening
IDE	Insulin Degrading Enzyme
IGF	Insulin-Like Growth Factor
IL	Interleukin
INCENP	Inner Centromere Protein
KDM4A	Lysine-Specific Demethylase 4A
LCMS	Liquid Chromatography/Mass Spectrometry
LFER	Line Free Energy Relation
LUMO	Lower Unoccupied Molecular Orbital
mAb	Monoclonal Antibody
MATE	Membrane Drug-Transporter
ML	Machine Learning
MLR	Multiple Linear Regression
MOJ	More O'Ferrall Jencks
MOPS	3-(N-morpholino)propanesulfonic acid
MOrPHs	Macrocyclic Organo-Peptide Hybrids
mRNA	Messenger RNA
MS	Mass Spectrometry
mTSAbs	Monoclonal Thyroid-Stimulating Antibodies
muPA	Murine Urokinase-Type Plasminogen Activator
NEB	New England Biolabs
NMR	Nuclear Magnetic Resonance
Notch1 NRR	Notch Homolog 1, Translocation-Associated (Drosophila) Negative Regulatory Region
O2beY	O-2-bromoethyl-tyrosine
oaTOF	Orthogonal Acceleration/Extraction Time of Flight
OBOC	One Bead One Compound
OBTC	One Bead Two Compounds
OPA	Oxaphosphetane
OpgY	O-propargyl-tyrosine
PA	Protective Antigen

PBS	Phosphate Buffer Saline
Pcns	Prochlorosins
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PgI	Propargylglycine
PPIs	Protein-Protein Interactions
ProcM	Promiscuous Marine Lanthionine Synthetase
PTM	Post-Translational Modification
PURE	Peptide-Translation by Using the Recombinant Elements
RaPID	Random non-Standard Peptide Integrated Discovery
RCM	Ring Closing Metathesis
RiPPs	Ribozomally Synthesized and Post-Translationally Modified Peptides
RNA	Ribonucleic Acid
ROESY	Rotating-Frame Overhauser Spectroscopy
RSV	Respiratory Syncytial Virus
RT	Room Temperature
SDS-PAGE	Sodium dodecyl sulfate–Polyacrylamide Gel Electrophoresis
SICLOPPS	Split-Intein Circular Ligation of Peptides and Proteins
SIRT2	Sirtuin 2
S_N2	Bimolecular Nucleophilic Substitution
S_NAr	Aromatic Nucleophilic Substitution
SPs	Synthetic Precursors
sRNA	Small RNA
SSC	Solid Supported Chemistry
ssDNA	Single Stranded DNA
TAE	Tris base, Acetic acid and EDTA buffer
TATA	1,3,5-triacryloyl-1,3,5-triazinane
TBAB	N,N′,N″-(benzene-1,3,5-triyl)-tris(2-bromoacetamide)
TBMB	1,3,5-trisbromomethylbenzene
TET1	Ten-Eleven Translocation Methylcytosine Dioxygenase 1
TEV	Tobacco Etch Virus

TFA	Trifluoroacetic Acid
THF	Tetrahydrofuran
TOCSY	Total Correlation Spectroscopy
TRAIL	Tumour Necrosis Factor-Related Apoptosis-Inducing Ligand
Tris	Trisaminomethane
TS	Transition State
TSG	Tumor Susceptibility Gene
uPA	Urokinase-Type Plasminogen Activator
UPLC	Ultra Performance Liquid Chromatography
VEGF	Vascular Endothelial Growth Factor
VHL	Von-Hippel-Lindau Tumor Suppressor
WT	Wild Type
XIAP	X-Linked Inhibitor of Apoptosis Protein
YEB	Ylide Ester Biotin

Chapter 1: Introduction

1.1 Macrocyclic Drug-Like Molecules

Lipinski's rules play critical role in reducing attrition of small molecule drug candidates due to poor pharmacokinetics and bioavailability of these molecules. The same rules hampered researchers from exploring chemical space containing pharmaceutically interesting and useful molecules that do not comply with these rules.[1,2] The group of drug-like molecules that violate the Lipinski's rules is largely represented by macrocycles,[3] a family of molecules that are usually classified as "medium-sized molecules", with higher molecular weights than most "small organic molecules" but that are not considered macromolecules such as antibodies and proteins.[4] Multiple macrocyclic-type compounds obtained from natural resources have outstanding pharmacological properties. As many as 68 macrocycles are already marketed as drugs (2014)[5] and they have been extensively used as antibiotics,[6] to block Protein-Protein Interactions (PPIs)[7–10] and to find binders of difficult targets.[11,12] Thus, macrocycles are highly promising for further exploration in development of new pharmaceuticals and bioactive ligands.[13]

Many pharmacologically important targets do not contain a small, deep binding site suitable for binding of small organic molecules. Instead, many targets such as areas of proteins engaged in PPIs, have large and shallow binding surfaces that require analogous, extended-area molecules as ligands. Proteins and antibodies thus were usually developed as drugs to block PPI.[11] Because of their topology, size and binding surfaces, macrocycles are larger than small molecules and are a promising alternative to expensive antibodies and other protein-based drugs for blocking of PPIs.[14] Cyclization often equips molecules with higher resistance to proteolysis; it also improves pharmacokinetic and biodistribution properties.[15,16] Macrocyclic peptides are known to be more rigid than linear peptide precursors. This rigidity could give rise to higher affinity towards a target because of the lower entropic penalty during interaction: linear peptides are highly flexible and adopt many more conformations in solution than macrocycles; as only one of them possess high affinity towards the target, an entropic cost must be paid for collapsing the multitude of conformations to only one spatial organization of the linear peptide during binding. Because of all these features conferred by macrocyclization, there is a growing interest in development of chemical strategies that can convert linear molecules to cyclic topologies.

Development of some macrocyclic drug-like molecules starts from the macrocycles obtained from natural resources and often only minimal modifications are performed to yield a drug candidate.[17–20] However, despite their advantages and widespread use in the pharmacological field, some naturally occurring families of macrocycles exhibit high toxicity,[21] mandating numerous chemical modifications to be performed to mitigate these unwanted properties. Efficient modification of these large, structurally and stereochemically complex molecules either through total synthesis or late stage functionalization often is a challenging task. Despite these challenges, as of 2008, most of the approved macrocyclic drugs were based on or identical to naturally occurring macrocycles.[22] One explanation is difficulty associated with *de novo* synthesis of structurally complex macrocyclic species. The development of technologies for generation of large diversity libraries of macrocycles, which, in turn, would enable High Throughput Screening (HTS) against targets is not trivial. Limited scope of HTS techniques for drug discovery in a macrocycle space as well as non-compliance to Lipinski's rules has lead to low representation of macrocyclic-type molecules in pharmaceutical companies.[23] To address this, multiple efforts have been made to provide new technologies for efficient generation of libraries of macrocycles that can be easily screened against biological targets.

The developed technologies for generation of libraries of macrocycles can be divided in two main groups: non-tagged and tagged collections of macrocycles. Non-tagged libraries require one-well-one-experiment screenings usually performed in 96-well plates. The biological activity of compounds against the target is measured individually and identification of active molecules after screening requires traditional robotic infrastructure for parallel liquid handling. In tagged libraries, members are attached to an information tag that ideally, is easily analyzed after screening. Besides facilitating identification of positive hits, some technologies use the tags as labels for separating of the positives from the rest of the population, allowing one-well-multiple-experiments screenings. Here, we review these methodologies (Figure1-1), with emphasis on genetically encoded techniques and focusing on phage display generation and screening of libraries of macrocycles.

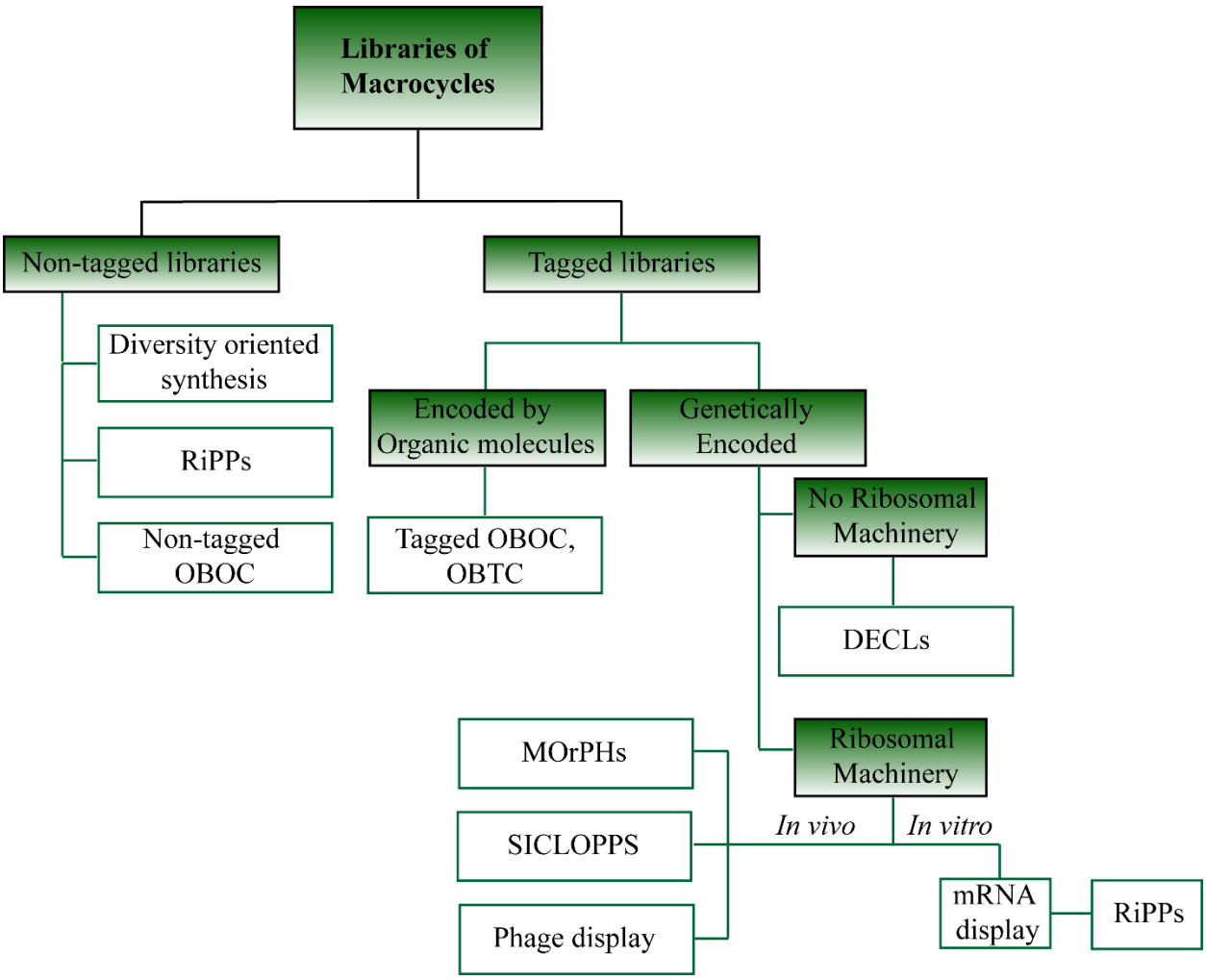


Figure 1-1. Strategies for production of libraries of macrocycles to be described here

1.2. Non-Tagged Combinatorial Libraries of Macrocycles

1.2.1. Synthetic Strategies for Generation of Libraries of Macrocycles

Examples of chemistries used to generate libraries of macrocycles include S_N2 (24 cyclic-72 linear member library[24] and 20 member library[25]), azide-alkyne cycloaddition (12-member natural product-like library of macrocycles,[26] a 12-member library of jasplakinolide analogs[27] and a 10-member library of β-turn mimetic tetrapeptides),[28] Pd catalyzed Heck reaction (20 to 24 members library),[29] Pd catalyzed carbonylation (122 members),[30] Ring Closing Metathesis (~10^2-10^3 members both in solid supported, and in solution approaches)[31–34] and multicomponent reactions (mainly Ugi 4-component, Passerini and Staudinger reactions).[35–52] Macrolactamization and macrolactonization in combination with Solid Supported Chemistry (SSC) allowed for parallel

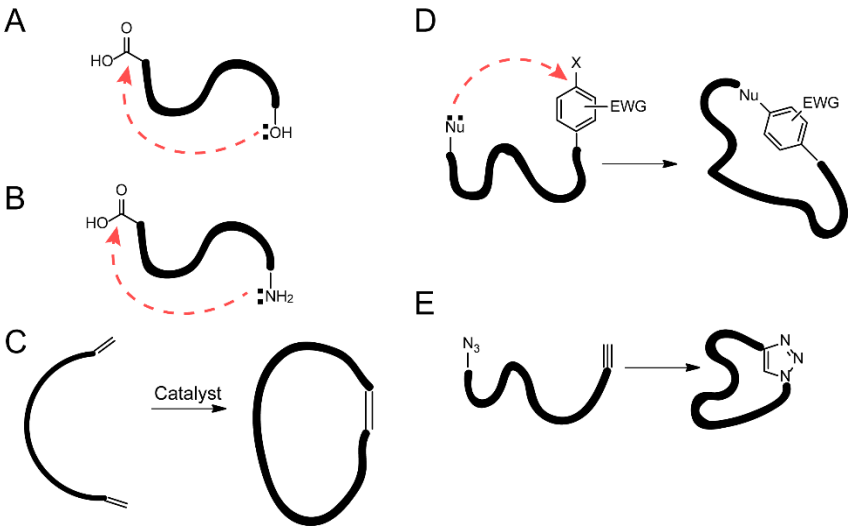


Figure 1-2. Main strategies for synthesis of macrocyclic molecules. (A) Macrolactonization, (B) macrolactamization, (C) Ring Closing Metathesis, (D) S_{NAr}, (E) Azide-alkyne cycloaddition.

syntheses of a 12-member library of macrolides,[53] 3-member, 9-member and 27-member libraries of hapalosin analogs[54] and ring-closure of unprotected tri-peptides[55] and hexapeptides.[56] This strategy was used for generation of a larger 10000-member library of peptides that was screened against human motilin receptor[57] and Ghrelin receptor.[58] S_NAr chemistry applied as the cyclization step on collections of peptide-like molecules (obtained *via* solid phase synthesis) provided mini-libraries (55 members[59] and a 39-member library of β-turn mimics[60]) but also larger 1320-member[61] and 12000-member libraries.[62]

The strategies that provided the most diverse libraries ($\geq 10^4$ members) rely on amino-acids as building blocks and SSC.[63] The One Bead One Compound (OBOC) technology invented by Kit Lam can be understood as a refinement of SSC through the application of split and pool combinatorial techniques for generation of peptide-based libraries. OBOC generated libraries can reach diversities of up to 10^8 different compounds[64,65] and collections of macrocycles have been accessed using this technique.[66–97] Despite its popularity and the large diversities that can be achieved, OBOC libraries of macrocycles have two main bottlenecks: i) difficulties in the purification of the macrocycles from resin cleavage mixtures[87] and ii) difficulties in the structural determination of binders obtained after screening. The latter is usually performed by identification of desired beads by manual picking using a microscope, followed by physical separation of

positive beads from the rest of the library, cleavage of compounds from individual beads and mass-spectrometry (MS) characterization of compounds.[10]

An important feature of libraries of compounds based on amino-acid building blocks is the possibility of using genetic engineering and the translational machinery of organisms (usually *Escherichia coli* or yeast) for *in vivo* generation of drug-like peptides. In these approaches, recombinant or synthetic DNA that encodes the target molecule is introduced into the host organism that can then use its own transcription and translation machinery to express such molecule. Production of libraries is achieved when the DNA of the target molecule is randomized and thus can express any combination of natural amino-acid in each position of the target sequence. This principle allows production of non-tagged and tagged libraries of peptide macrocycles as described in several of the following sections.

1.2.2 Ribosomally Synthesized and Post-Translationally Modified Peptide (RiPPs) Macrocycles

RiPPs designation has been introduced by a consortium of more than 60 researchers[98] to describe peptide-derived natural products that are biosynthesized using the regular ribosomal machinery and extensively modified by other enzymes present in the organism after translation. Many families of RiPPs consist of cyclic topologies[98] but only a few examples of combinatorial production of libraries have been reported. Although groups of enzymes that perform Post-Translational Modifications (PTM) are highly substrate selective, several RiPPs systems have been found to be promiscuous. Examples include: i) the pathway for synthesis of prochlorosins (Pcns), where the enzyme responsible for cyclization (ProcM) can accept up to 29 different linear peptides as precursors;[99,100] ii) the pathway for production of microviridins that allow production of artificial tricyclic depsipeptides[101] and iii) the cyanobactin biosynthesis pathway, which is highly substrate tolerant[102,103] and has allowed synthesis of a library of 12 prenylated cyanobatin-cyclic peptides (aestuaramides).[104] During the early stages of the cyanobactin pathway, Recognition Sequences (RSs) facilitate "docking" between the enzymes and the peptidic substrates.[105,106] These RSs were used to obtain peptides with non-natural amino-acids or non-proteinogenic cores,[105–110] increasing the pharmacological value of the macrocyles produced using RiPPs.

Despite the high structural complexity of macrocycles obtained with RiPPs, this system suffers from similar disadvantages to *in solution* Diversity Oriented Synthesis (DOS) and parallel synthesis: i) macrocycles need to be purified (in this case out of the cell lysate), ii) macrocycles have to be characterized using traditional techniques and iii) generation of high diversity libraries ($>10^2$ members) remained elusive until recently (Section 1.3.6.4). Importantly, the genotype-phenotype connection between the macrocycle and the DNA used to encode the linear precursor is lost after translation, lysis and purification, hindering production of genetically tagged collections of macrocycles. Other genetically encoded methodologies described later may not produce macrocycles as complex as the RiPPs system, but can generate genetically tagged libraries with higher diversities ($>10^5$)

All the strategies mentioned before allow access to libraries of macrocycles that can be either structurally complex (DOS, parallel synthesis and RiPPs) or highly diverse (SSC and OBOC). One-well-one-experiment HTS using 96-well requires the set up of thousands of experiments in parallel, demanding high cost robotics and automation.[111] Another option of screening for non-tagged high diversity libraries involve one-well-multiple-experiments methodologies, where advances in mass spectrometry facilitate the identification of hits out of mixtures.[112,113] Tandem MS/MS *de novo* peptide and peptoid sequencing[114–117] are very useful for sequencing linear peptides; however, cyclic peptides are difficult to sequence due to complex fragmentation patterns.[118] Because of this, linkages used for library cyclization are designed so that after screening, they can be cleaved to re-generate de linear peptide and allow easy sequencing by MS/MS.[119–125] A recently developed strategy by Pentelute and coworkers that blends nanoliquid chromatography with MS/MS is very promising for high throughput sequencing of peptides[126] and might be applied to selections of highly diverse libraries of macrocycles.

1.3 Encoded Libraries of Macrocycles

Tagging members of a library aims to facilitate identification of selected molecules. In some occasions, it also allows labeling and separation of an active population in one-well-multiple-experiments screenings. Tags can consist of an easy-to-characterize organic molecule (sections 1.3.1 and 1.3.2) or genetic information. Genetic tags are convenient because signals can be amplified using polymerase chain reaction (PCR) and hits can be identified using DNA-

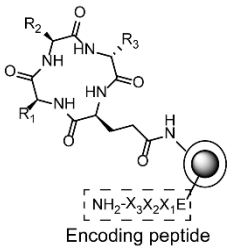


Figure 1-3. Example of one-bead-two-compounds macrocyclic peptide encoded by linear counterpart

sequencing techniques (Sanger sequencing and more recently, Deep Sequencing). This also allows for the use of directed evolution methods to discover active molecules present in libraries. Directed evolution is analogous to natural selection, where production of local variants of DNA by random mutation of individual base pairs or distal variants by recombination of extended DNA segments allows certain population of organisms to produce molecules that make them resistant to external pressures (analogous to active members in a library or "hits"), these population has a survival advantage over other organisms without the mutation that allows production of such active compounds (non-active members of the library). All of these strategies will be described in the sections below.

1.3.1 One Bead Two Compounds (OBTC) Libraries of Macrocycles

An example of libraries of macrocycles encoded by easy-to-identify organic molecules that facilitate structural determination of peptide "hits" is the One-Bead-Two-Compounds technology developed by Pei and coworkers. This strategy can be viewed as an adaptation to macrocycles of the biphasic solvent strategy developed by Lam and coworkers, that allows testing molecules to be on the surface while encoding tags are in the interior of the bead.[127] Here, the macrocyclic peptide (attached to the surface of the bead) is encoded by its linear counterpart (attached to the internal part of the bead) as shown in Figure 1-3. After screening, partial Edman degradation-mass spectrometry permits identification of the peptide binders.[128] This technology enabled the synthesis and screening of both monocyclic[129,130] and bicyclic[131–133] peptide libraries. Despite facilitating identification of hits, these strategies still require manual separation of positive beads from the rest of the population under a microscope. An interesting report described the use of cleavable linkers that allow magnetization (*via* iron based labeling of targets) of positive sequences, which then can be separated from the non-positive population, cleaved and identified

with mass spectrometry. Using this strategy, affinity quantification (i.e. validation) is also performed without the need to re-synthesize the hit molecules.[134]

1.3.2 DNA Encoded Chemical Libraries (DECL) of Macrocycles

DNA-encoding of chemical libraries is a concept first introduced by Lerner and Brenner for encoding chemically synthesized tripeptides based on Gly and Met.[135,136] This powerful idea allows the use of split and pool techniques to synthesize libraries of peptide-like compounds (that have potential diversities of millions of different sequences) with genetic tags to accompany each building block used in the construction of the library.[136,137]. Genetic encoding allows signal amplification (*via* PCR) and thus, use of nanomolar concentrations of libraries for screening. At these low concentrations, active members of the library would be difficult to detect using MS/MS methods applied to non-tagged or OBTC libraries. Also, identification of hits can be performed using DNA-sequencing techniques.

With the aim of applying DNA-encoded combinatorial chemistry for generation of soluble libraries of small organic molecules not dependant on SSC, in the early 2000s, Liu and coworkers introduced the concept of DNA-Templated Synthesis (DTS), where reactions between two reagents are accelerated due to annealing of complementary DNA strands attached to the building blocks. The annealing of complimentary strands puts building blocks in close proximity, increasing the local concentration of reactants and thus, increasing the rate of reaction. In this way, transformations using very low concentrations (that in a non-templated set up would take near infinite time) can be performed.[138–145] In 2004, Liu and coworkers applied the DTS concept to produce a 65-member library of macrocycles. The library was assembled using three consecutive amine acylation steps: template molecules attached to three coding regions (ssDNA, 3 codons) were reacted with pools of building blocks. Each building block was attached to ssDNA containing one coding region (1 codon) complementary to one of the codons of the template molecules' ssDNA. The final cyclization step was achieved using the Wittig reaction.[146] This small library was successfully screened against carbonic anhydrase. Several optimization steps during the synthesis and incorporation of a higher variety of building blocks allowed production of a library with a diversity of 13000 different macrocyclic compounds.[147] Other groups are working actively in development of further functionalization for generation of complex topologies in peptide-based libraries that use original SSC approach by Lenner and Brenner.[148,149] Paegel and coworkers have

developed a microfluidic device for selection of DNA-encoded libraries immobilized on beads that facilitates separation of positive beads from the rest of the population as well as identification by using sequencing techniques.[150]

DECLs have been successful in discovering macrocyclic binders against the intracellular X-linked inhibitor of apoptosis protein XIAP (starting from a DNA-encoded library of 160000 compounds),[151] several kinase enzymes,[152–154] and Insulin Degrading Enzyme (IDE).[155] Recently, Liu and coworkers made further improvements on the technology and synthesized a second-generation library of macrocycles with diversity of 256000 different compounds.[156] This library was screened against IDE and a binder with IC_{50}=40 nM was found. The authors propose that this new library has improved drug-like properties and will be useful in small-molecule discovery.[156] DTS libraries were used to start Ensemble Therapeutics in ~2004 and after publishing and patenting of a number of successfully discovered inhibitors, the company was closed in 2018.

Several chemical approaches were used for cyclization of molecules attached to DNA. The Wittig reaction was used by Liu and coworkers for ring closure of both, first and second generation DNA-encoded libraries (Figure 1-4A).[146,147,156] Huisgen 1,3-dipolar cycloaddition was used by Terret and coworkers at Ensemble therapeutics for the synthesis of the 160000-members library that allowed discovery of the XIAP binders.[151] This same ring closing chemistry was used by Zhu et.al. at X-Chem to make DNA-encoded library of cyclic peptides composed of $2.4x10^{12}$ members. The library was screened to find binders against Von-Hippel-Lindau tumor suppressor (VHL) and Respiratory Syncytial Virus (RSV) N-protein (Figure 1-4B).[157] Recently, a DNA-compatible, Ru catalyzed Ring Closing Metathesis (RCM) reaction was reported, although its use in production of libraries of macrocycles has not been published to date (Figure 1-4C).[158]

1.3.2.1 DECL of Macrocycles in the Industry

DECL technology has been widely embraced by the industry. Large pharmaceutical companies like Glaxo-Smith-Kline (GSK), Roche, Sanofi and Novartis have divisions for production and screening of DECLs, as well as countless partnerships with smaller biotech companies performing research based on this technology.[159–162] Recently, Nuevolution reported

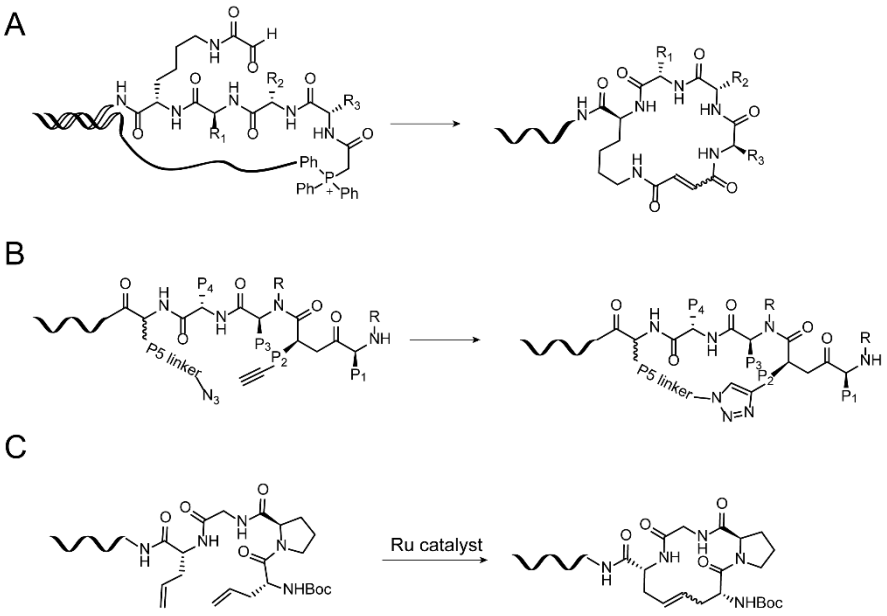


Figure 1-4. (A) Cyclization of DNA-templated libraries with Wittig reaction. (B) Cyclization using 1,3-dipolar cycloaddition. (C) Cyclization using RCM

the synthesis of 40 trillion DNA-encoded compounds, the largest DECL library of small molecules to date.[163,164] When it comes to DECL of macrocyclic species, several small companies have been founded based on this technology. Some of them are: i) Ensemble Therapeutics, founded by Dr. Liu in 2004, ii) Philochem AG, founded by Dr. Neri in 2006 and iii) DICE molecules, founded in 2013 by Harbury and iv) Praecis Pharmaceuticals founded in 1993 that developed the modern DNA ligation-based implementation of DECL used by GSK, X-Chem and Hitgen.Ensemble Threapeutics successfully discovered antagonists for XIAP inhibitor and also focused a big part of its research on the discovery of IL-17 antagonists.[165] Despite this, it was reported by the *Boston Business Journal* in November 2017 that the company had been shut down. On the other hand, Philochem and DICE Molecules have yet to report successful inhibitors. These examples illustrate that even if very promising, DECL technology has challenges that need to be addressed. The main one is the high cost of initial construction and implementation of soluble, small organic molecule-based DNA-encoded libraries of macrocycles because the cost of encoded building blocks is higher that regular non-tagged building blocks.[166,167] A researcher in academia or a founder of a small company would most likely choose a more economical technology for molecular discovery. This in turn centralizes DECL screening research on big pharmaceutical companies.[166] On the other hand, only chemistries that are compatible with DNA can be used.[168] However, efforts to optimize

screening conditions,[169] generate higher throughput and lower cost DNA-compatible chemistry[168] as well as automation,[150] might result in DECL technology being more accessible and efficient in the future.

Despite the mentioned drawbacks, this technology is popular due to the advantages of having genetic tags linked to the chemical molecules to be screened. In nature, macrocyclic peptides and macrocyclic natural products biosynthesized by an organism are encoded by its DNA. *In vivo* and *in vitro* technologies that exploit this genotype-phenotype natural connection are described in the following sections.

1.3.3 Split-Intein Circular Ligation of Peptides and Proteins (SICLOPPS)

Inteins are portions of proteins that can excise themselves from a sequence, ligating in this process the N-terminus to the C-terminus of the flanking sequences referred to as exteins (Figure 1-5A)[170,171] In some cases, it is possible to split the intein in N-terminus and C-terminus domains, after which both domains reassemble and catalyze the ligation of the corresponding N and C-terminus exteins. This is a process named *trans*-splicing and is used by biological systems for reconstitution of full-length proteins starting from two fragments (Figure 1-5B).[172–174] The *trans*-splicing ability of certain inteins was used by Benkovic and coworkers to synthesize circular proteins and peptides inside *E. coli*. This can be achieved by expressing a fusion protein (or peptide) that has the C-terminus intein domain followed by the extein (target sequence to be cyclized) and final N-terminus intein domain (Figure 1-5C). This methodology for *in vivo* production of cyclic proteins and peptides is referred to as Split-Intein Circular Ligation of Peptides and Proteins (SICLOPPS).[175–178] One of the limitations of the SICLOPPS technique for producing libraries of cyclic peptides is substrate tolerance, a problem that could be solved by the engineering of promiscuous and kinetically faster inteins.[179] On the other hand, because the library gets amplified *via* a host cell (*in vivo* technique), the diversities that can be achieved with SICLOPPS are $<10^9$, which is several orders of magnitude lower than diversities achieved with *in vitro* genetically encoded techniques described later. This deficiency, however, is offset by the ease of which libraries are generated since preparation of collections of peptides only requires the SICLOPPS plasmid accompanied by the randomized nucleotide region,[177,180] making this strategy affordable and easy to implement for many academic groups.[181] Most screening techniques are

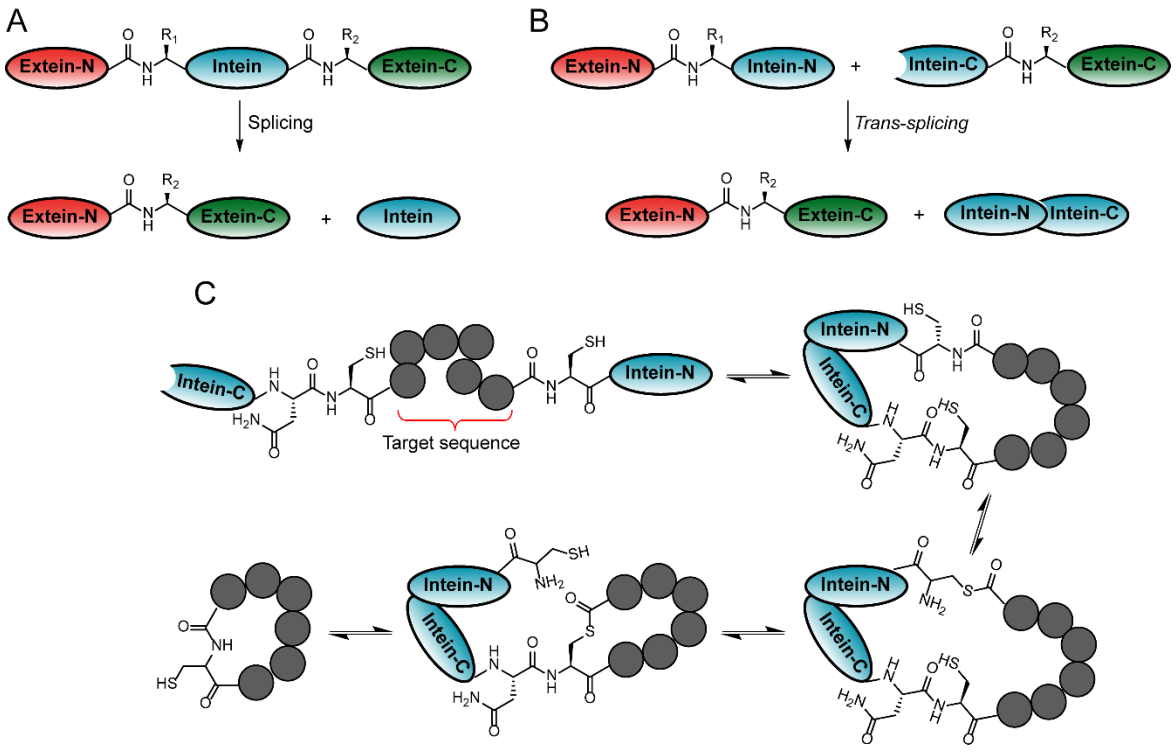


Figure 1-5. SICLOPPS system for generation of macrocycles. (A) Intein splicing mechanism. (B) *Trans*-splicing mechanism. (C) SICLOPPS cyclization showing key intein interactions

limited to extracellular surface targets because members of the libraries are attached to tags (or are themselves) not suitable for cell internalization. The members of libraries generated by SICLOPPS are screened intracellularly,[10] inside the host cell producing the macrocycle and thus, the screening is not only performed for affinity towards an intracellular target but also for actual functional effect on the organism due to interaction of the macrocycle with the target.[182] Another advantage of SICLOPPS includes the possibility of integrating non-natural amino-acids *via* genetic code expansion.[183]

Inhibitors against *E. coli* Dam methyltransferase,[184] ClpXP[185] protease, as well as inhibitors of the function of the NarZ gene[186] for generation of antibacterial peptides were reported. Multiple inhibitors of protein-protein interactions such as the HIV gag protein-TSG101 interaction,[187] the Hfq-sRNA interaction,[188] ribonucleotide reductase,[189] 5-aminoimidazole-4- carboxyamide-ribotide transformylase[190] and β-sliding clamp[191] dimerization domains were found. Hdm2 and Hdmx antagonists[192] and TEV protease inhibitors,[193] were reported. Recently, Tavassoli and coworkers used the *E. coli* system to screen for inhibitors of the B-Cell Lymphoma 6 (BCL-6)

transcription factor function, a difficult cancer related target for which few potent inhibitors are reported[194] and for which no drugs are approved for treatment. The selection was focused on inhibiting BCL-6 homodimerization, a protein-protein interaction process for which no binders had been reported before this study.[195]

The SICLOPPS platform has also been implemented in mammalian cells. Inhibitors of IL-4 signaling,[196] as well as inhibitors of hypoxia-inducible factor (HIF) heterodimerization[197,198] and C-terminal-binding protein (CtBP) transcriptional repressor dimerization[199] were found. Inhibitors of the INCENP IN-box-Aurora B interaction were discovered after screening on HeLA cells.[200] SICLOPPS inteins also work in yeast systems: a peptide that reduces toxicity of α-synuclein was found[201] and binders of LexA protein[202] as well as inhibitors of Abl kinase were discovered.[203]

In an interesting example that illustrates the robustness of screening methods using SICLOPPS, a library of 3.2 million cyclic hexapeptides (CX_5) was screened to find inhibitors of the PA/CMG2 protein-protein interaction involved in *Anthrax* infection. Here, the top three hits (CLRFT, CLRPT and CMNHFPA sequences) had stop codons in the randomized region, which means these peptides were linear instead of cyclic and thus, were underrepresented when compared with cyclic motifs in the initial library. Still, these peptides were selected and reproducibly observed after selection and when validated *in vitro*, they effectively inhibited the target interaction.[204]

Several efforts have focused on secreting SICLOPPS libraries out of cells to allow screening of extracellular receptors. Soumillion and coworkers reported a system where the C-terminal intein was fused to a co-translational translocation signal, allowing extracellular secretion. However, this strategy secretes macrocycles that are no longer attached to the genetic tag, requiring one-well-one-experiment screenings or sophisticated MS/MS for hit deconvolution as in non-tagged screenings.[205] In a recent report, a photo-switchable intein was devised that allows purification of a stable intein splicing linear precursor from cell extract. This precursor undergoes cyclization only after UV-light activation of the complex, generating an almost pure cyclic peptide that can be tested in traditional *in vitro* assays. Again, in the process, the phenotype-genotype connection is lost.[206] The SICLOPPS technology is a valuable platform complementary to other screenings techniques because it can be used on intracellular targets.

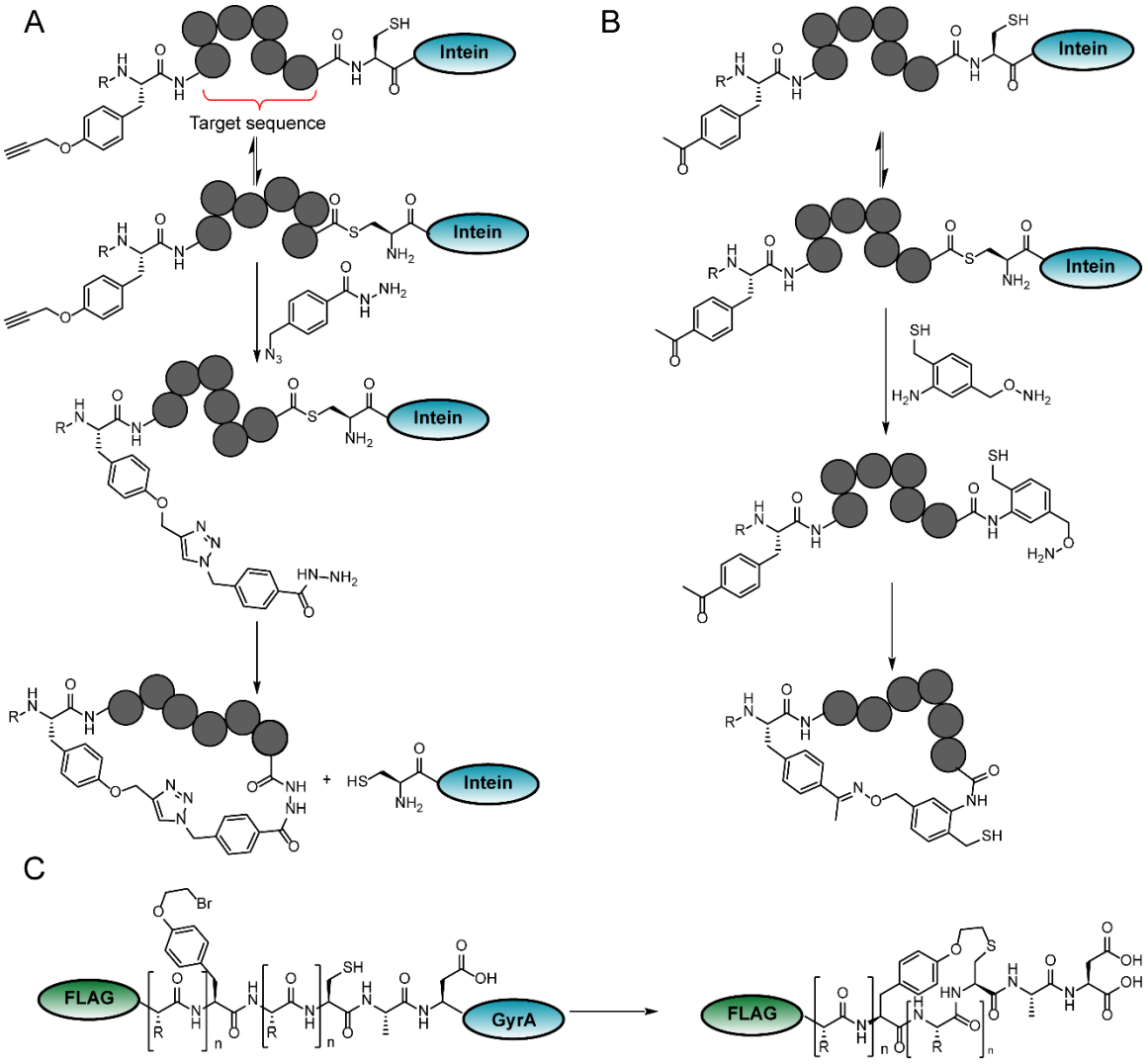


Figure 1-6. Strategies for generation of MOrPHs. (A) Cyclization using "click" chemistry and hydrazine ligation. (B) Cyclization using amine acylation and oxime ligation. (C) Cyclization without the need of synthetic precursor (SP) introduction.

1.3.4 Macrocyclic Organo-Peptide Hybrids (MOrPHs)

A technology introduced by Fasan and coworkers for the production of complex peptides *in vivo*, allows diversification beyond just the introduction of unnatural amino-acids or localized post-translational modifications by combining biosynthetic and organic synthesis strategies. In his approach, non-proteinogenic scaffolds and several non-amino-acid related moieties are introduced in peptides using recombinant protein precursors (BPs) and bifunctional Synthetic Precursors (SPs). In turn, BPs can be highly randomized by regular combinatorial mutagenesis The BPs

consist in target sequences flanked by an unnatural amino-acid and GyrA, an intein unable to perform C-terminal splicing.[207] Macrocyclic peptides have been produced using azide/hydrazide functionalized SP, and BPs that have an unnatural tyrosine residue functionalized with an alkyne moiety OpgY (Figure 1-6A).[208,209] Another methodology consists of an oxyamino/1,3-amino-thiol-aryl SP and BPs containing a *p*-acetyl-Phe, which serves as the source of a ketone moiety for oxime ligation with the SP (Figure 1-6B).[210,211] In a recent report, macrocycles were formed without the need of an SP by encoding a bromoalkyl-containing tyrosine (O2beY) residue, a cysteine and an aspartic acid. The O2beY and Cys residues are used for spontaneous intracellular cyclization while the aspartic acid residue induces intein cleavage (Figure 1-6C). A library was generated and screened to find Sonic Hedgehog/Patched Interaction inhibitors.[212]

1.3.5 mRNA Display Libraries of Macrocycles

In the previous examples, the genotype-chemotype linkage between the molecules to be screened and the DNA encoding them was achieved using the machinery of an organism and thus *in vivo* translation was required. Conversely, in the ribosome display technology *in vitro* selection of peptide libraries displayed on the surface of a ribosome is possible due to physical linkage of each peptide to the ribosome and parent mRNA.[213–215] Szostak and Roberts developed a simpler technology where the peptides are directly attached to the encoding mRNA via a hydrolytically stable puromycin linkage between the peptide C-terminus and the mRNA at the end of the translation process. In this way, the mRNA-peptide complex can be purified from the translation mixture and selection does not have to be performed in an environment where the mRNA-ribosome-peptide complex stays assembled.[216,217] Both of these *in vitro* techniques do not require amplification of the library *via* a host cell and thus the diversities that can be accessed are higher ($>10^9$) than the ones achieved by *in vivo* display techniques (phage display, yeast display, bacterial display and SICLOPPS). Further optimizations of the technique[218,219] made it robust enough to allow genetic code reprogramming strategies to be applied. The aim of genetic code reprograming is to enable the synthesis of peptides that include non-canonical amino-acids such as D and N^{α}-methylated amino-acids or amino-acids with unnatural side chains and other chemical features not achievable with regular translation machinery. Several approaches to genetic code reprogramming have been reported such as the PURE (protein synthesis using recombinant elements) system,[220–225] systems where the innate promiscuity of Aminoacyl-tRNA synthetases (ARSs) is

exploited[226,227] and the flexizyme system, where artificial, highly promiscuous ribozymes are used.[228–233] The use of flexizymes has allowed introduction of more than 300 non-canonical amino-acids in polypeptides using the Flexible *in vitro* Translation (FIT) strategy, a translation system using pre-charged, non-canonical aa-tRNAs and flexizyme ribozymes.[234]

Using these non-proteinogenic building blocks results in production of peptides with improved pharmacokinetics, cell permeability and higher protease resistance.[235] Macrocyclization of peptides helps improve both their affinity to targets with extended binding surfaces and their pharmacokinetic properties.[16] mRNA display in conjunction with the FIT system (fusion of these two techniques is referred to as the RaPID system) and macrocyclization chemistry have allowed access to many promising drug candidates reviewed below.

1.3.5.1 Synthesis of Libraries of Macrocycles Using mRNA Display

The first examples of macrocyclic libraries using mRNA display used post-translational modifications with linkers that contained reactive moieties. Macrocyclic libraries were obtained by reacting an mRNA display library of linear peptides with disuccinimidyl glutarate for head-to-chain cyclization (Figure 1-7A),[236] Michael addition using non-canonical dehydrobutyrine,[237] oxidative coupling between 5-hydroxyindole and benzylamine[238] and S_N2 using dibromoxylene (Figure 1-7B).[239,240] Bicyclic peptide libraries have been obtained by encoding two cysteines, β-azidohomoalanine and p-ethynyl phenylalanine into the sequences, performing sequential S_N2 with dibromoxylene and azide-alkyne cycloaddition (Figure 1-7C).[241]

Another methodology for production of macrocyclic libraries using mRNA display consists in exploiting the FIT system for encoding a Cysteine-Proline-Glycolic acid sequence in the C-terminus of the nascent peptide. This sequence contains a C-terminus ester moiety that accelerates a series of intramolecular rearrangements. A thioester is formed, and the N-terminal amine attacks it to perform a final head-to-tail cyclization.[242–244] Similarly, macrocyclic libraries have been made using the FIT system to charge t-RNA^{met} with N^{α}-cloroacetyl amino-acid (instead of methionine) as the initiator for peptide biosynthesis. Spontaneous cyclization occurs between the N-terminus chloroacetyl group and a C-terminus cysteine (Figure 1-8A).[245] Libraries of bicyclic peptides were produced using the N^{α}-cloroacetyl amino-acid strategy to form one of the

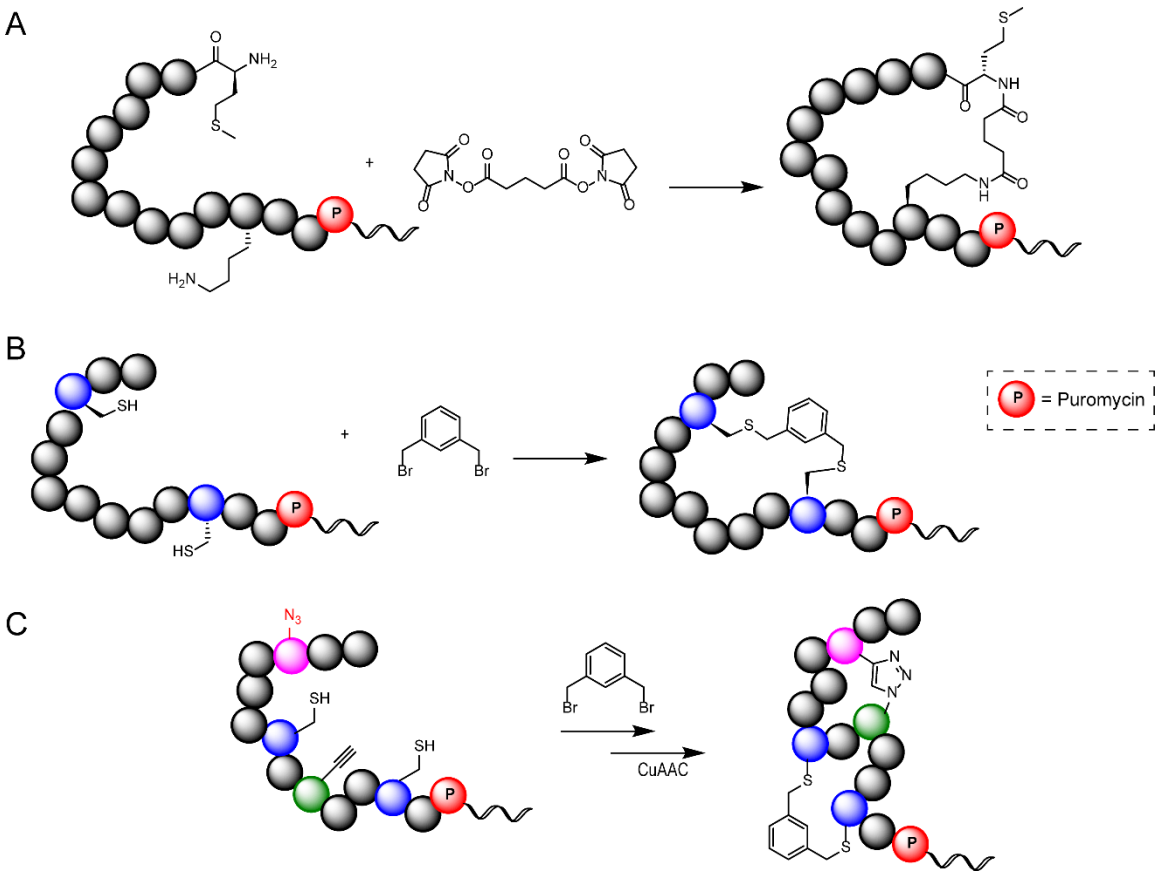


Figure 1-7. Examples of post-translation modification for cyclization of mRNA display libraries of peptides.

cycles and click chemistry between azidohomoalanine (aha) and propargylglycine (PgI) to close the second cycle (Figure 1-8B). All of these non natural amino-acids were charged into peptides using the FIT system.[246] The N^{α}-cloroacetyl amino-acid strategy approach, coupled with encoding three other cysteines inside the sequence and reaction with TBMB provided tricyclic topologies (Figure 1-8C).[247] Recently, the N^{α}-chloroacetyl amino-acid strategy was used to couple carbohydrates to macrocyclic libraries without the need of a linker by use of an extra cysteine (Figure 1-8D).[248]

1.3.5.2 Bioactive Macrocyclic Peptides Discovered Using mRNA Display

Three different research groups have been able to independently find active macrocyclic peptides against different targets using the mRNA display technology. Binders against thrombin,[236] sortase A,[249] Gαi1,[250] histone deacetylase SIRT2,[251] protein kinase Akt2,[252] the

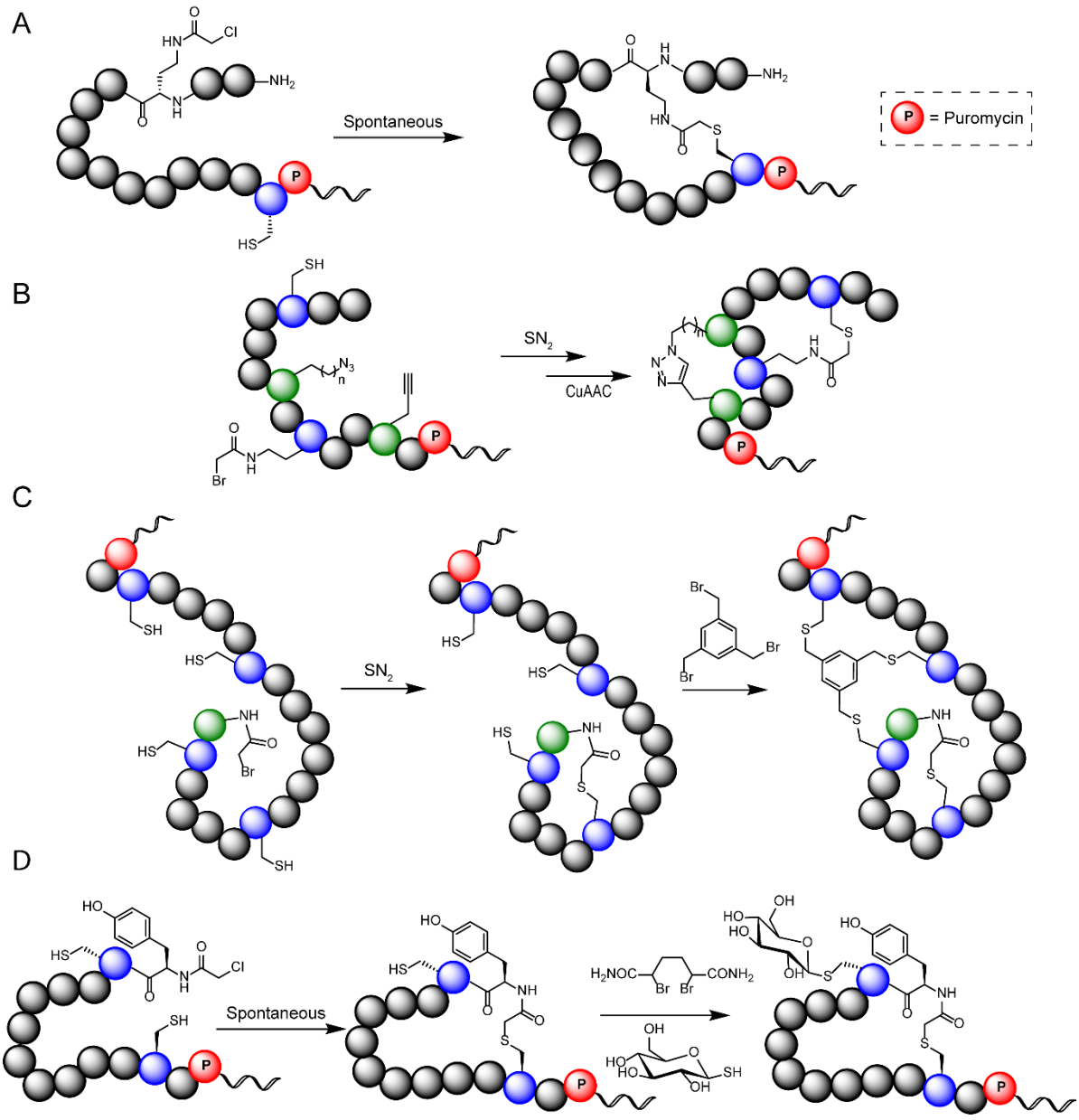


Figure 1-8. Formation of bicycles (A) and tricycles (B) using the FIT system to encode unnatural amino-acids.

bacterial membrane drug-transporter MATE,[253,254] eukaryotic ABC-drug transporter,[254] ubiquitin ligase,[255] VEGFR2,[256] TET1,[257] cell-surface receptors MET,[258,259] cell-surface receptor EpCAM,[260] plexin B1 cell-surface receptor,[261,262] KDM4A,[263] human pancreatic α-amylase,[264] nematode phosphoglycerate mutase,[265] P-glycoprotein transporters,[266] protein-protein interaction of Zaire Ebola virus[267] and hepatitis B virus cellular entry[268] have been discovered.

All the previous examples of discovery of biologically active macrocycles illustrate the success of the mRNA display platform. This success has been highly boosted by research at the start up company Peptidream, a Japanese biotech company founded by Kiichi Kubota and Hiroaki Suga that currently has licensed its RaPID technology to companies like Merck, Genentech, Bayer, Eli Lilly, Novartis, Roche among others.

1.3.6 Phage Display Libraries of Macrocycles

The idea of using the phage display platform to generate libraries of short peptides that could enable the discovery of epitopes against antibodies was introduced by Smith and coworkers in 1990.[269] Soon, this approach was applied for the discovery of epitopes of other monoclonal antibodies.[270,271] DeGrado and coworkers envisioned that these peptide libraries can not only yield epitopes against antibodies but also might provide specific binders of proteins and thus, mimics of natural ligands.

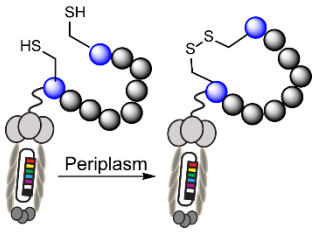


Figure 1-9. Cyclization of displayed peptides on phage *via* disulfide formation

1.3.6.1 Disulfide-Cyclized Peptides Displayed on pIII Protein of Phage

As a proof of concept, DeGrado and coworkers decided to use the platform to find ligands of the platelet glycoprotein IIb/IIIa, a member of the very well-known integrin family of proteins involved in cell adhesion. The X_6 and CX_6C libraries were displayed in the N-terminus of pIII protein of M13 phage and due to the oxidizing environment of the periplasm, the CX_6C library consists of disulfide cyclized peptides (Figure 1-9). Both libraries were screened and after the second round of selection, the CX_6C library exhibited 500-fold higher number of phage particles than X_6 library,[272] in agreement with the prediction that constrained moieties generate better binders.[273] After sequencing, cyclized peptides with the known integrin-binding RGD motif were discovered, demonstrating the applicability of phage display for discovery of peptide ligands of proteins.[272]

Libraries of linear peptides that have not been specifically constructed to have two cysteines in determinate positions generated cyclic "hits" after screening due to disulfide formation between two cysteines present in randomized positions that could have had any of the 20 natural amino-acids. This demonstrates that constrain and stable spatial arrangement are factors that generate sufficiently strong binders as to provide disulfide cyclized "hits" from libraries that initially are composed mainly by linear libraries.[269,274] These results prompted Koivunen and coworkers during the early 90's, to make the CX_7C library displayed on the pIII protein of M13 phage. Use of this library allowed the discovery of binders for α5β1 integrin[275] as well as binders for α-chymotrypsin.[276] Koivunen and coworkers also produced the CX_5C, CX_6C, CX_7C and CX_9C libraries displayed on the pIII protein of phage using the fuse5 vector and shared those libraries with different research groups that eventually found binders for α5β1, α4β3 and α4β5 integrins,[277] DNA binding peptides,[278] binders for leukocyte-specific β2 integrins,[279] binders for urokinase-type plasminogen activator (uPA) from human urine,[280] binders for Plasminogen Activator Inhibitor 1 (PAI-1)[281] and binders for murine urokinase-type plasminogen activator (muPA).[282]

Production of these cyclic libraries in the mid 1990's soon became robust enough to be adapted by several groups. For example, the library SCX_8CGS was displayed on the pIII protein of M13 phage and screened to find calmodulin (CaM) binders.[283] The CX_6C displayed on pIII protein was constructed and screened to find antagonists of Mac-1 β2 integrin.[284] The CX_9C library was made by King and coworkers, and screened against Grb2 SH2 domain to find pTyr independent peptidic binders.[285] In this last example, when the disulfide-type cyclization was replaced with a redox stable thioether linkage, the peptide retained its activity while increasing its stability.[286] On the other hand, Coutts and coworkers made a library of the form $AGPC(X)_7CPG$ and screened it to find binders of an anti-cardiolipin antibody (ACA 6501).[287]

Additional benefits emanated from libraries that introduced randomization outside the disulfide bridge. A library of the type $X_2CX_4CX_2$ was displayed on the pIII protein of M13 and screened against human angiogenin to identify the GECRENVCMG sequence. After adding four additional amino-acids in both sites based on sequence of the N-terminus region of pIII protein, the AQLAGECRENVCMGIEGR showed IC50 of 1 μM.[288] Libraries of the form $DGX_3CRGDCX_3$ and $X_3CRGDCX_3$ were generated on the pIII protein of M13 phage and panned against α_V integrins.[289] The $ADGXCX_4CXSYIDGRI$ library was synthesized in the pIII protein

and screened against streptavidin and anti-β-endorphin monoclonal antibody.[290] In a series of studies that illustrate the importance of cyclization for identification of good binders, libraries $AECX_6C$, $AECX_5C$ and $AECX_4C$ were made and screened against streptavidin. All hits showed the HPQ motif previously shown by linear binders but affinities of the cyclized versions were 2-3 orders of magnitude higher than the linear counterparts found in previous selections.[291] Explanation for the 2-3 orders of magnitude in affinity increase for the cyclized peptides is primarily entropic, since the disulfides were found not to interact with the protein.[292] Studies where the disulfide bridge was replaced by thioether linkages after the selection yielded binders with similar affinity but with higher stability.[293]

Phage display technology of short peptides is accessible and robust enough as to allow commercialization of disulfide cyclized libraries. Corvas, Dyax Corp. and New England Biolabs have commercialized several libraries, and some of them have been used in more than 100 publications. The CX_6C library obtained from Corvas was screened against 82D6A3 mAb.[294] Libraries of the form X_3CX_4C, X_3CX_8C and $X_3CX_{10}C$ displayed on pIII of M13 were made by the Dyax corp. team and screened against caprylate-HSA to find peptides with applications in peptide purification[295] and for selection of peptides with high affinity to factor VIII that would allow application in chromatographic purification of this protein.[296,297] The same libraries were panned to find selective binders of fibrin.[298] Mori and coworkers used these libraries to find binders of the cyanovirin-N.[299] Recently, Dyax Corp. was bought for 5.6 billion dollars mainly because of Ecallantide, a 60 -amino-acid peptide discovered using phage display.[300]

The ACX7C library (PhD™-C7C™) displayed on the pIII protein of M13 bacteriophage and commercialized on the early 2000s by New England Biolabs has become the most used constrained library, screened against multiple targets, including cells lines and *in vivo* panning procedures. Screenings discovery of binders of monoclonal thyroid-stimulating antibodies (mTSAbs; B6B7 and 101–2),[301] human synovium,[302] HT29 colon carcinoma cells,[303] Hepatitis B core antigen (HBcAg),[304] HCV RNA-dependent RNA polymerase (NS5B),[305] prostate specific membrane antigen,[306] anti-EGFR antibody,[307] HBsAg,[308,309] SiO_2 and TiO_2 nanoparticles,[310] Newcastle disease virus virulent strains,[311] family 18 chitinases from *Serratia Marcescens*,[312] hemaglutinin, neuramidinase and ion chanel protein M2 from AIV,[313] tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) receptor DR5,[314] amyloid beta $A\beta_{42}$,[315,316] rat lung

alveolar epithelial primary cells,[317] brain vascular receptors,[318] interleukin 13 receptor α2,[319] HSP70-PCs,[319] human squalene synthase,[320] human gastric mucin MUC5AC,[321] βB2-crystallin fibrillated protein,[322] Caco-2 cells,[323] insulin degrading enzyme,[324] and tissue-homing peptides.[325]

The consensus of most of these reports is that for the targets studied, high affinity binding was dependent on the presence of the disulfide bridge. However, in a couple of reports,[313,319] linear, reduced species showed better binding than cyclized peptides. In an example, libraries of the form X_8 and CX_6C were panned against the mAb KAA8 (anti-angiotensin II peptide AII) and at all cases, constrained peptides showed 100-fold less affinity than linear ones.[326] Similarly, libraries displaying a 34-mer with an alanine in the central position as well as a 43-mer with a cysteine in the central position were assembled on the pIII protein of M13 phage and panned against mAb 7E11-C5 raised against prostate-specific membrane antigen (PSM). The M-(H/Y/W/I/S)-X-X-L-(H/R) motif was found in both libraries and in the library with the central cysteine, the motif was found both, inside and outside of constrained cycles (which can form by presence of another cysteine), showing that constrained species did not yield stronger binders.[327] Finally, the libraries X_9, CX_6C and $CX_{10}C$ were screened against the surface human neutrophil cells (PMN). The best binders corresponded to the motifs (G/A)PNLTGRW and DLXTSK(M/L)X(V/I/L), both coming from the X_9 library.[328]

In the mentioned reports, cyclized peptides did not show improvement in affinity and even showed less affinity than linear peptides but these are rare exceptions. Converse examples exist in which the cyclized hits were found after screening linear libraries. A linear decamer peptide library displayed on the pIII protein of fd phage was constructed[329] and screened against monoclonal anti-colicin A antibody 1C11 in order to study its interaction with colicin A. Disulfide constrained binders were found.[330] Similarly, A 15-mer linear library of peptides displayed on the gIII protein of fuse5 phage was screened against 24822.111, F9, anti-CCR5 and Myastenia Gravis (MG) monoclonal antibodies. In all examples disulfide constrained hits were obtained.[331–334] A linear library of 20-mers displayed in the pIII protein of M13 phage was panned to find mimics of the folded AMA1 membrane antigen of the parasite *Plasmodium falciparum* (malaria), hits consisting in disulfide constrained peptides were found.[335] The SRX12(S/P/T/A)A(V/A/D/E/G)X12SR library displayed on the pIII protein was screened against the activated form (C3b) of C3 and a constrained peptide was found to be the best binder.[336]

In cases where linear and cyclized libraries were screened against the same target, selection was generally more efficient when using cyclized moieties. For example, the libraries X_{20}, Z_{10}, CX_4C and CX_6C (where X=amino acids encoded with NNK and Z=amino acids encoded with VNK and VVK) displayed on N-terminus of pIII protein of M13 phage, were panned to find binders to *Escherichia Coli* thioredoxin (Trx), in an attempt to find peptides mimicking protein-protein interactions; peptides containing disulfides were found.[337] Analogously, the PhD7, PhD12 and CX7C libraries displayed on pIII protein of M13 phage (NEB) were panned to find binders to FcγRIIa and RhoA GTPase. In both cases, the CX7C library provided the best binders.[338,339] Similarly, Ph.D7 and CX7C libraries displayed on pIII protein of M13 (NEB) were screened to find binders of the integrin β1, the best binders being the ones constrained by disulfides.[340] In some cases, second-generation linear libraries made after finding a specific motif can yield cyclized final hits.[341]

Longer, more complex libraries were made and screened. The advantages of longer peptide libraries are larger interaction area and possibility for stable secondary structure exhibited by peptides. The main disadvantage is the exponential drop in library coverage (i.e., only one in 10^{10} peptides is present in typical PhD12 library). For example, a library of the type $X_4CX_{10}CX_4$ was displayed on the pIII protein of filamentous phage using the fuse5 system and screened against Grb7 SH2 Domain, a protein overexpressed in breast, esophageal, and gastric cancer cells.[342] A highly selective peptide was found and a thioether analog was synthesized that showed good affinity, selectivity and increased stability. This peptide has been subjected to several optimization studies to generate a final potent and selective inhibitor of Grb7.[343–345] $X_2CX_{14}CX_2$ library was displayed on the pIII protein of fd-tet phage and screened against the mAb 5A2[346] and extracellular-regulated protein kinase 2 (ERK2).[347] A 34-mer library yielded inhibitors of the Interleukin IL-6/IL-6 receptor signaling complex.[348]

In some cases, long peptides with established activities have served as scaffolds for randomization and generation of more potent binders[349–352] and this technique has found application on diagnostics. For example, 12-mer peptide of the immunodominant loop of HIV-1 envelope glycoprotein gp41 was displayed on the pIII and pVIII proteins. When the displayed peptides had the cysteine residues responsible for constrain, phages were able to efficiently (98.4%) recognize HIV positive samples.[353] Similarly, the disulfide cyclized peptide somatostatin

has been displayed both on pIII and pVIII proteins and panned against polyclonal anti-somatostatin serum.[354]

Libraries with double functionality have also been generated and screened. For example, a peptide library displayed on the pIII protein of fuse5 phage with the form PPPVPPRGGGGCX$_6$C was screened against the *Caenorhabditis elegans* SEM-5 SH3 C-terminal domain, which is known to recognize the proline rich sequence displayed besides the GGG linker. Adding the disulfide constrained second binding loop allows for exploration of other regions in the domain.[355]

1.3.6.2.1 Disulfide-Cyclized Bicyclic Peptides Displayed on pIII Protein of Phage

Generation of bicyclic, disulfide-cyclized libraries is also feasible. For example, libraries of the $X_mCX_nCX_oCX_p$ topology were generated and cyclized either by oxidation or by TBMB treatment (section 1.3.6.3). Selections against streptavidin and uPA yielded motifs with even number of cysteines for libraries cyclized via oxidation while sequences with odd number of cysteines were obtained in TBMB selections, demonstrating feasibility of getting bicycles via disulfide formation.[356] Bicyclic peptides constrained *via* formation of disulfide bonds can form three isomers and usually, just one of those isomers is biologically active. In presence of natural thiols, isomers can interconvert and this is a drawback of this type of libraries, although different reports have tried to overcome this difficulty.[357,358] It is expected that generation and screening of these bicyclic libraries will be performed routinely in the near future, since accessible protocols and techniques were developed disregarding the target.[359]

1.3.6.3 Disulfide-Cyclized Peptides Displayed on pVIII Protein of Phage

Unlike minor protein pIII present in 3-5 copies on the tip of the virion, protein pVIII coats densely the side of the M13 phage; its copy number is 2000-3000 particle. Disulfide libraries very similar to the ones displayed on pIII have also been assembled on pVIII protein. For example, a CX$_9$C library displayed on the N-terminus pVIII protein of M13 was constructed and screened against mAb H107, which was raised against the native conformation of recombinant human H-subunit ferritin (H-Fer).[360] Similarly, a linear 12-mer library[271] and a CX$_9$C library,[360] both displayed on the N-terminus of pVIII coat protein of phage were panned to find binders of two antibodies previously raised against protein-attached glucitollysine and cyclic peptides were enriched both in the CX$_9$C, and the 12-mer library.[361] The CX$_8$C library displayed in the pVIII

protein was screened to find peptides with high affinity for the erythropoietin receptor. Affinity maturation prompted the generation of a library that had three random amino-acids at the sides of each cysteine, yielding multiple binders.[362] In an important study, RGD-like constrained peptides were displayed on pVIII to generate virions that can be internalized in mammalian cells.[363]

Libraries of disulfide cyclized peptides with different topologies have allowed the discovery of binders of 14G2a mAb, an antibody raised against GD2 ganglioside,[364] binders of IGF-binding proteins (IGFBPs),[365] Hepatocyte Growth Factor (HGF),[366] vascular endothelial growth factor (VEGF),[367–369] and mAb A7617E3C3 (which is an antibody raised against *Brucella abortus* outer membrane protein).[370] Analogous to observations on pIII selections, when linear and cyclized libraries are screened against the same target, constrained libraries usually provide the best binders.[371–374] It is important to mention that libraries displaying an odd number of cysteines on the pVIII protein might not be presenting peptides in a cyclic form. The peptide libraries $X_{15}CX$, X_8CX_8, XCX_2SDLX_3CI, XCX_4CX, $XCX_{12}CX$, X_6 and X_{15} were evaluated via SDS-PAGE to determine propensity to form homodimers of the peptides displayed on the pVIII protein. It was found that libraries expressing just one cysteine display their peptides mostly in a dimeric form, meanwhile libraries with two cysteines show small amount of dimer, being the monomer cyclic peptides the favored species.[375]

1.3.6.4 Generation of Phage Display Libraries of Macrocycles Using Post-Translational Modifications

Peptides constrained *via* formation of disulfide bonds can interconvert in presence of natural thiols[357] and suffer oligomerization under certain conditions.[293] Generating constrained peptides with redox stability[286] is also necessary to improve pharmacokinetics and distribution of potential drug leads. For these reasons, the pioneering work by Katz and coworkers[293] where the peptides were cyclized *via* thioether bonds instead of disulfides is important. Timmerman and coworkers developed a strategy for generation of macrocycles on unprotected peptides by taking advantage of the superior nucleophilicity of cysteine at pH values in which most other nucleophiles are protonated. Reacting peptides that have two, three and four cysteine residues with bis, tris and tetrakis (bromomethyl) benzene derivatives generated monocycles, bicycles and tricycles *via* S_N2 reaction.[376] This strategy was adapted by Winter and coworkers for functionalization of phage display libraries that had three constant cysteines encoded in the displayed sequences (CX_6CX_6C

where X is any amino-acid). After reaction with tris (bromomethyl) benzene (TBMB), bicyclic moieties were produced (Figure 1-10A). This library of bicyclic peptides was screened to find inhibitors of plasma kallikrein and after affinity maturation, an inhibitor K_i=1.5 nM was found. This screen was done in plasma *ex vivo* and the fact that inhibitors were found, demonstrated that the bicyclic moiety generates peptides that are stable enough to resist protease degradation during the experiment, thus allowing performance of a functional selection [377]

The stability conferred by bicyclization was further demonstrated in a study where degradation in mouse plasma (*ex vivo*) for linear, separated monocyclic loops and bicyclic peptides was compared. In order to avoid renal clearance *in vivo*,[378] the bicyclic peptide can be non-covalently bound to albumin *via* a previously discovered albumin-binding cyclic peptide.[379] This type of bicyclic peptide-albumin complexes also showed to diffuse into solid tumors.[380] Recently, conjugation of a fatty acid to an albumin binding portion of the complex and screening and optimization to find a peptide with good affinity towards albumin, generated soluble complexes.[381] Another strategy consisted in fusing linear peptides encoding three cysteines to the Fc portion of an antibody for subsequent cyclization using TBMB. The bicyclic peptide-Fc conjugate showed *in vivo* half-life of 1.5 days and was active against plasma kallikrein.[382] A screen specifically designed to find proteolytically stable hits consisted of incubating the library of bicyclic peptides in a pancreatic extract of proteases before selection against plasma kallikrein.[383]

The TBMB CX_6CX_6C functionalized library was also used to find inhibitors of human urokinase-type plasminogen activator (uPA) with K_i=53 nM.[384] The activity of this inhibitor was improved 2-fold simply by changing a natural L-glycine for its D counterpart,[385] demonstrating the importance of efforts that aim to encode unnatural amino-acids on phage display libraries. On the other hand, a bicyclic library with smaller loops (CX_3CX_3C) was reacted with TBMB to generate selective binders towards human plasma kallikrein and its orthologues but inactive towards its paralogues.[386] In an interesting application, these highly selective and active bicyclic inhibitors were used to quantify the activity of their specific targets in biological samples.[387] The authors also envisioned that screening libraries of bicycles of different sizes can be performed simultaneously by mixing all the libraries and reacting all of them with TBMB in one pot, instead of performing one by one modification and screening. After screening of 14 libraries of different

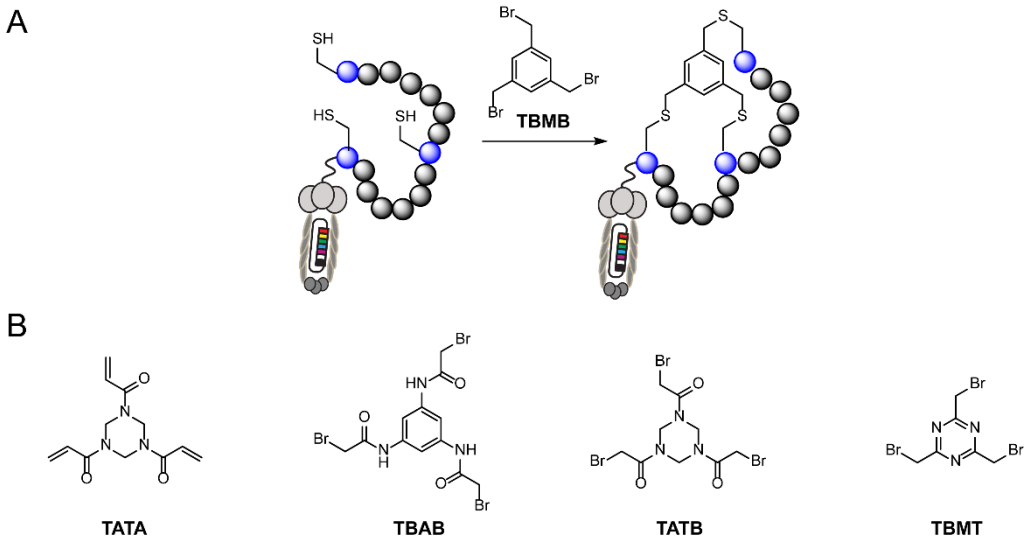


Figure 1-10. Post-translational cyclization of peptides displayed on phage via oxidation (A) and SN_2 reactions (B and C)

bicyclic topologies against serine protease, different motifs for different sizes of bicycles were found, demonstrating the validity of this hypothesis.[388] TBMB bicyclized libraries have also been used to find inhibitors of coagulation factor XII,[389] epidermal growth factor receptor Her2,[390] Notch1 NRR,[391] *S. Aureus* sortase A[392] and HECT-type ubiquitin ligases.[393] The success of these studies and the use of deep sequencing to find motifs[394] might facilitate screening of these libraries of bicycles against many targets.[395]

Other linkers with three thiol reactive functional groups and threefold rotational symmetry similar to TBMB (Figure 1-10B) were used by Heinis and coworkers for generation of bicycles in phage display libraries. These linkers react with thiols via S_N2 or Michael addition to generate bicycles with different conformations, spatial arrangements and flexibilities,[396] factors that according to a pilot selection, affect the binding efficiency towards plasma kallikrein,[397] showing that peptides cyclized with different organic templates can provide binders with diverse affinities for different targets. This hypothesis was confirmed in a study where several libraries of different loop sizes were cyclized with TBAB and TATA and only TATA bicyclic selected peptides showed both binding and inhibition against Factor XIIa.[398,399] The affinity of the binder found with TATA modification was further improved by adding atoms to the macrocyclic core.[400] In another study, libraries cyclized with either TATA, TBAB or TBMB showed binding towards different sites of

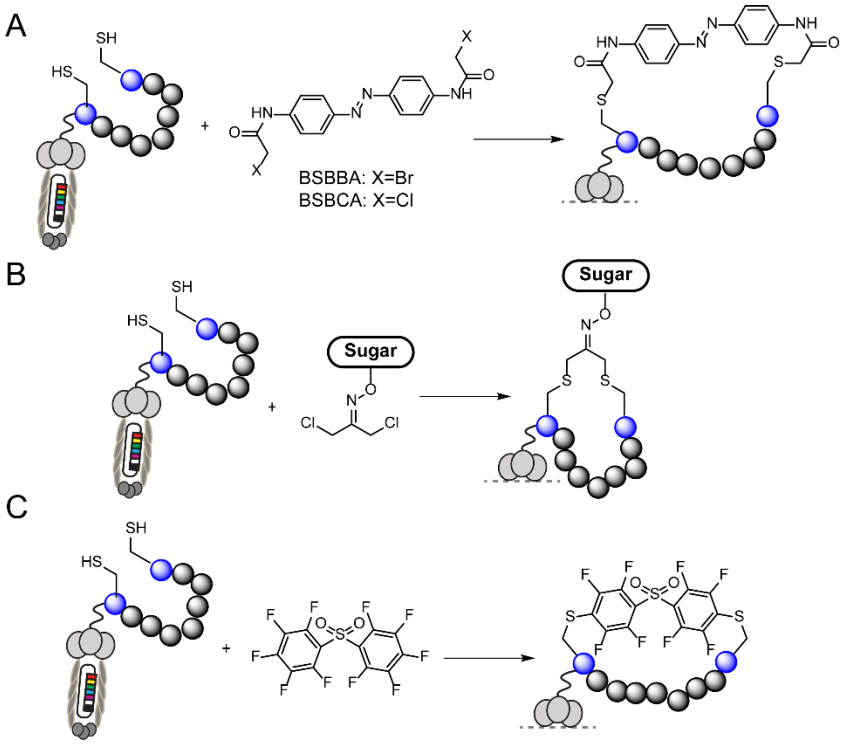


Figure 1-11. Generation of light responsive libraries (A), carbohydratre functionalized libraries of macrocycles (B) and macrocyclization *via* S_NAr (C)

the target β-catenin.[401] Recently, two new linkers (TATB and TBMT) with extra nitrogen atoms and higher dipolar moments were synthesized (Figure 10B). The peptide bicycles generated are more soluble than their TATA, TBAB and TBMB counterparts. It is expected that these new reagents will improve biological activity and will facilitate validation of hits.[402]

Recently, libraries with cysteines in three constant sites of the displayed peptide sequences were reacted with two linkers that each one contained two thiol reactive groups. The authors hypothesized that bicyclic peptides would be obtained by reaction of the three constant cysteines plus an extra cysteine coming from the randomized regions. Using this strategy, multiple cyclic scaffolds can be generated in one single reaction and libraries with a high number of topologies and thus chemical diversity can be generated. These simultaneously cyclized libraries were screened against plasma kallikrein. Nanomolar and sub-nanomolar inhibitors were discovered and three of macrocyclic peptides (specific regioisomers) showed sub-nanomolar K_{ir}>1000-fold target specificity.[403]

The azobenzene moiety has been introduced into phage display libraries encoding two cysteine residues to screen for macrocyclic light-responsive ligands. Heinis and coworkers introduced the linker BSBBA via S_N2 reaction (Figure 1-11A) and screened for streptavidin binders that can be activated with light (the *cis* isomer is the active species). Single digit micromolar binders were found, however, *cis* and *trans* isomers of the selected sequences bound with similar affinity.[404] Derda and coworkers used the linker BSBCA to cyclize a library *via* S_N2 with two cysteines encoded in constant positions (Figure 1-11A). In this case, the screening was performed so that the active *trans* isomers would be selected. The best binder had K_d of 229 μM and one of the selected peptides had a ~5-fold difference between *cis* and *trans* isomers.[405]

The S_N2 reaction has allowed Derda and coworkers to generate macrocyclic libraries of peptides displayed on phage that have additional value due to introduction of different carbohydrates in the linker used for post-translational modification (Figure 1-11B).[406] On the other hand, S_NAr is a chemistry that has been extensively used in diversity oriented synthesis of macrocycles and generation of small synthetic libraries as described in Section 1.2.1. This approach was used by Derda and coworkers for generation of libraries cyclized with decafluoro-diphenylsulfone, one of the fastest Cys reactive linkers reported to date (Figure 1-11C).[407]

Recently, scientists at MorphoSys and Lanthio Pharma developed a technology where linear precursors of lanthipeptides (a subclass of RiPPs) were encoded in the C-terminus of pIII protein of filamentous M13 phage. Heterologous co-expression of the RiPPs enzymes of the ProcM system allowed peptides to undergo post-translational modifications in the cytoplasm of the producing cell to generate a library of lanthipeptides displayed on phage that was screened to find binders or urokinase plasminogen activator (uPA) and streptavidin.[408] This important report combines the advantages of complex topologies and pharmacological potential of RiPPs with the benefits of genetically encoded techniques. Libraries of RiPPs with diversities $>10^5$ can be produced, amplified and screened using several rounds of selection as in traditional phage display protocols.

All the chemistries developed by Derda and coworkers and the new polycyclic moieties developed by Heinis and coworkers can be used for discovery of new macrocyclic drug-like molecules for difficult targets. Such screenings will benefit from recent technological developments of the phage display platform. For example, with deep sequencing it is possible to

visualize the genetic information of a million different clones before and after selection, which has allowed development of sophisticated analysis for discovery of motifs[394,409,410] and characterization of commercial libraries.[411] Also, efforts have been made to standardize methods that do not require several rounds of selection in order to avoid enrichment of parasitic sequences.[412,413] Optimization of one-round selections would also allow the implementation of the silent encoding technology developed by Derda and coworkers.[414] Finally, introduction of unnatural amino-acids in phage display libraries has been reported[415,416] and its use in production and screening of macrocyclic libraries is expected to grow in the future.

1.4 Thesis Overview

Size of a library is not the only important feature to ensure success of a selection, structural and functional diversity are also critical.[417] Phage display is a platform that can generate high diversity libraries of up to 10^9 different compounds but unless PTMs or genetic code expansion are used for topological diversification, the library is only composed of proteinogenic amino-acid based peptide macrocycles. Genetic code expansion on phage display is not as well established as in mRNA display and thus, new methodologies for diversification and cyclization of phage display libraries that do not rely on encoding of unnatural amino-acids need to be developed. Only four chemistries have been used for generation of libraries of macrocycles displayed on phage and all of them require encoding of at least two cysteines in constant positions: i) oxidation to form disulfides, ii) S_NAr, iii) S_N2 and iv) Michael addition. New chemistries for phage diversification must fulfill several conditions: compatibility with the tag, i.e., reagents added to phage libraries should react only with the displayed peptides and be orthogonal to all other functional groups present in the proteins of the bacteriophage and also, solvents and reaction conditions should not disassemble the virion in order to conserve the genotype-phenotype linkage. These conditions are not met by most traditional organic chemistry transformations and that is why only the four mentioned reactions have been applied to macrocyclization of peptides displayed on phage. In this thesis, we describe new topological diversifications on peptides displayed on phage as well as use of the platform for reaction development.

In Chapter 1, I described the main technologies for generation of libraries of macrocycles and the pros and cons of each one of them. I made emphasis on macrocyclic libraries displayed on

phage and mentioned how one of the challenges consists in increasing chemical complexity of the libraries, and how one of the strategies to circumvent this problem is development of new post-translational modifications compatible with the phage platform.

In Chapter 2 I describe my work on the Wittig reaction between an ester stabilized phosphorene ylide and glyoxaldehydes displayed on phage. The developed chemistry is compatible with the phage display platform and generates a dienophile that can undergo Diels-Alder cycloaddition in the presence of cyclopentadiene and Michael addition in the presence of thiols.

In Chapter 3 I describe my work on Genetically Encoded Structure Activity Relationships (GE-SAR) of Wittig reaction between ester stabilized ylide and glyoxaldehydes in aqueous media. Affinity pull-down and deep-sequencing uncovered substrate-dependent reactivity that spanned a factor of 50 from most unreactive to most reactive peptide aldehyde. GE-SAR uncovered a combined role of backbone amides and side-chain functionalities in the aqueous Wittig reaction. Two Trp residues adjacent to the glyoxaldehyde increase its reactivity in the Wittig reaction, whereas substituting them by Pro removes two backbone amides adjacent to the aldehyde, decreasing the rate of this reaction by a factor of 50.

In Chapter 4, the potential use of ester-functionalized libraries of peptides to find cyclization prone sequences is summarized. We found several sequences undergoing transformations that generate species with mass corresponding to cyclized peptides. We proposed a structural identity for the obtained adducts.

Chapter 2: Tandem Wittig/Diels-Alder Diversification of Genetically Encoded Peptide Libraries

2.1 Introduction

Chemically-modified peptide libraries are a valuable source for discovery of ligands for fundamental research, development of diagnostics and biomaterials as well as discovery of therapeutic leads.[418] Chemical post-translational modification of peptides made of 20 "natural" amino acids is the simplest strategy for production of such libraries because it bypasses the need of advanced constructs for incorporation of unnatural amino acids. Introduction of unnatural fragments into genetically-encoded peptide libraries makes it possible to equip these libraries with value-added functionality not encodable by conventional translational machinery.[418–421] Although there exists a rich palette of chemical transformations for site-specific modification of proteins made of natural amino acids, only a limited scope of chemical transformations have been adopted for diversification of genetically-encoded peptide libraries. These reactions include S_N2,[377,405,406,422] and S_NAr substitution,[423] Michael[396] and allenamide[424] addition reactions as well as tandem elimination-addition reaction[248] starting from sulfhydryl group of Cys; nucleophilic substitutions of selenocysteine,[425,426] addition of alpha-nucleophiles to aldehydes,[414,427–429] and cycloadditions to unnatural side chains containing azide or propargyl group.[415]

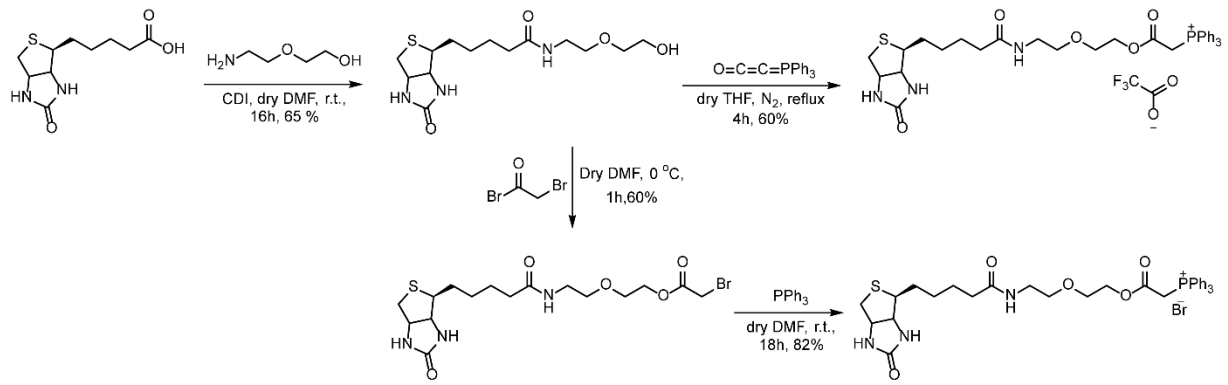
Several carbon-carbon bond forming strategies—based on cycloadditions and transition metal catalyzed reactions—have been reported for modification of amino acid side chains of individual proteins or peptides in water.[429–432] However, the only example to date of a C-C bond formation reaction performed on the phage display context has been reported by Lin and coworkers, who described an elegant optimization of the Sonogashira cross coupling in a subset of peptides on phage libraries. This coupling, however, hinged on expression of an unnatural homopropargylglycine amino acid residue *via* amber suppression.[416]

Wittig olefination is one of the attractive reactions to explore for diversification of peptide libraries on phage because it has already been employed to modify individual proteins in water as well as amino acids ligated to DNA strands.[433–435] Wittig chemistry has also been employed by Liu and coworkers for macrocyclization of DNA-templated libraries.[146] In proteins, aldehyde handles suitable for Wittig reaction are readily introduced by mild oxidative cleavage of natural

residues that contain 1,2-aminoalcohols (N-Ser/Thr)[427,436,437] or 1,2-diols (glycans);[438] it can also be incorporated via UAA-mutagenesis,[439] or encoded as a L(C/A)T(P/A)S(A/R) peptide sequence recognized by the formylglycine generating enzyme.[429,440] Specifically, in peptide libraries displayed on phage, the orthogonal aldehyde handle can be quantitatively introduced by selective oxidation of N-terminal serine.[427] Conveniently, the Wittig reaction between N-terminal glyoxy-aldehyde and ester-stabilized ylide produces an olefin within an extended conjugation framework (Figure 2-1) that enhances its performance both as Michael acceptor and dienophile. The introduced N-terminal ester of fumaric/maleic acid could enable survey of up to 10^6-10^9 Michael electrophiles to identify suitable covalent or reversible inhibitors.[441–443] Introduction of norbornene-like moieties *via* tandem Wittig-Diels Alder reaction also opens the door for robust labeling *via* bioorthogonal aziridination[444] or inverse electron demand Diels-Alder cycloaddition with tetrazine[445] in a phage display context. While individual aforementioned reactions are well-established, the tandem use of such modifications in 10^6-10^9 scale genetically encoded libraries has not been investigated or reported to date.

The Wittig reaction of stabilized ylides can yield both E and Z isomers, and the rate and selectivity of this chemistry strongly depends on the reaction conditions.[446] Similarly to reactions in methanol, most "on-water" and "in-water" Wittig reactions of stabilized ylides with substituted benzaldehydes, α,β-unsaturated (vinyl) aldehydes, heterocyclic derived aldehydes or aliphatic aldehydes exhibit an increase in rates compared to analogous reactions in aprotic solvents at the expense of lower E selectivity.[447–451] Difference in further reactivity of the conjugated olefin might be a powerful approach to separate phage populations that preferentially generate either of both isomers.

Here we developed a tandem of two carbon-carbon bond-forming reactions to chemically diversify libraries of peptides displayed on bacteriophage. Wittig reaction of a biotin-ester from a stabilized phosphorane ylide with model peptides containing N-terminal glyoxal exhibited reaction rates of 0.07 to 5 $M^{-1}s^{-1}$ in water at pH 6.5-8.0. The log(k) scaled linearly with pH from pH 6 to 8; above pH 9 the reaction was accompanied by hydrolysis of the ester functionality. Capture of the phage displaying the biotinylated product by streptavidin beads confirmed the rate of this reaction in a library of 10^8 peptides ($k=0.23$ $M^{-1}s^{-1}$ at pH=6.5) and also confirmed the regioselectivity of



Scheme 2-1. Synthetic pathways towards ylide ester biotin phosphonium salt precursor (YEB).

this modification. The olefins introduced in the Wittig reaction can act as Michael acceptors: addition of glutathione, cysteamine, and DYKDDDDKC ("FLAG-Cys") peptide occured with k=0.12-4.1 $M^{-1}s^{-1}$ at pH 7.8. Analogous reactions with DYKDDDDKC peptide take place on phage-displayed peptides modified *via* Wittig reaction. This reaction was manifested as a progressive emergence of FLAG-epitope on phage and detected by capture of this phage using anti-FLAG antibody. Olefins introduced in Wittig reaction also act as dienophiles in Diels-Alder reaction with cyclopentadiene. The conversion of the dienophile to norbornene-like adducts on phage was observed by monitoring the disappearance of the thiol-reactive olefin on phage. Described below, different isomers present different reactivity towards hydrolysis and Diels-Alder reactions.

2.2 Results and Discussion

2.2.1 Synthesis of Ylide Ester Biotin (YEB) Linker

Nucleophilic addition of biotin-PEG-alcohol to Bestmann ylide yielded a carbonyl stabilized phosphonium salt (YEB) in 65% yield. The same positively charged product was obtained in two steps by reaction of the alcohol with bromoacetyl bromide and triphenylphosphine (Scheme 2-1), although scaling up of this second strategy generated bisacetylated adducts, considerably reducing the yield and making it less practical. To optimize the Wittig reaction in the environment akin to that of the glyoxal on phage-displayed peptides, we employed a model sequence o-VEKY

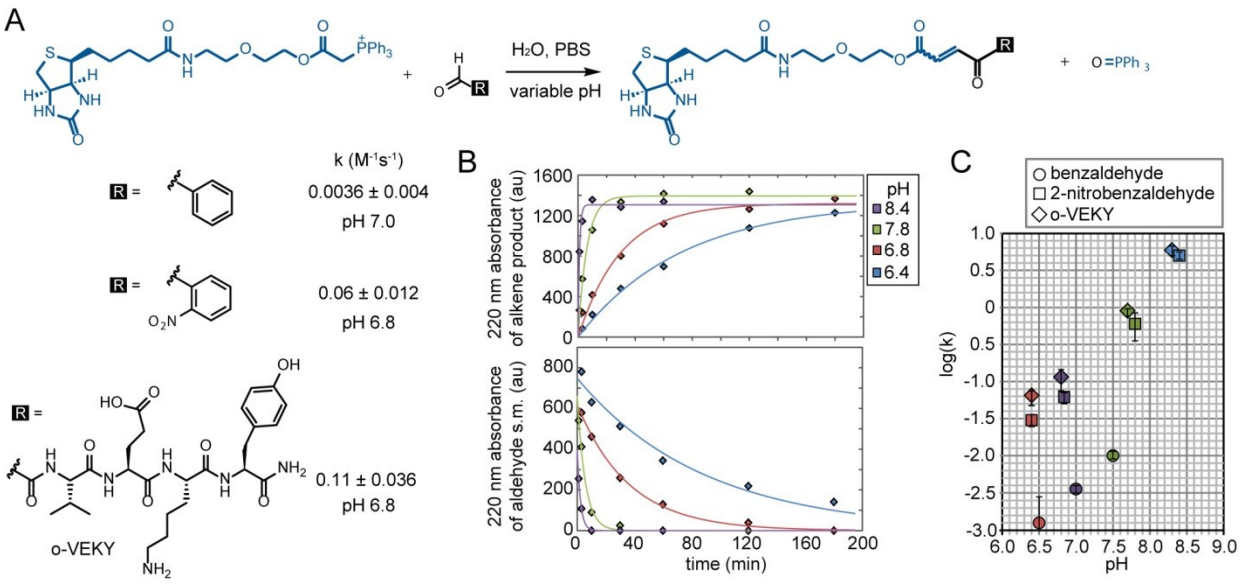


Figure 2-1. (A) Wittig reaction on model aldehydes. (B) Absorbance of reactant and product (extracted as absolute area of the peaks in HPLC trace) of reaction between YEB and o-VEKY at different times and pH values. The data was fit to a equation At = A(1–e-k[YEB]t), where k is the pseudo-first order rate constant, [YEB] – initial concentration of the ylide, and A is the maximum absorbance. (C) Rate constants of reactions of three aldehydes at different pH values.

produced by $NaIO_4$ oxidation of the peptide SVEKY.[427] The Wittig product (YEB-VEKY) was then used for evaluation of rates for Michael addition, retroMichael and Diels-Alder reactions.

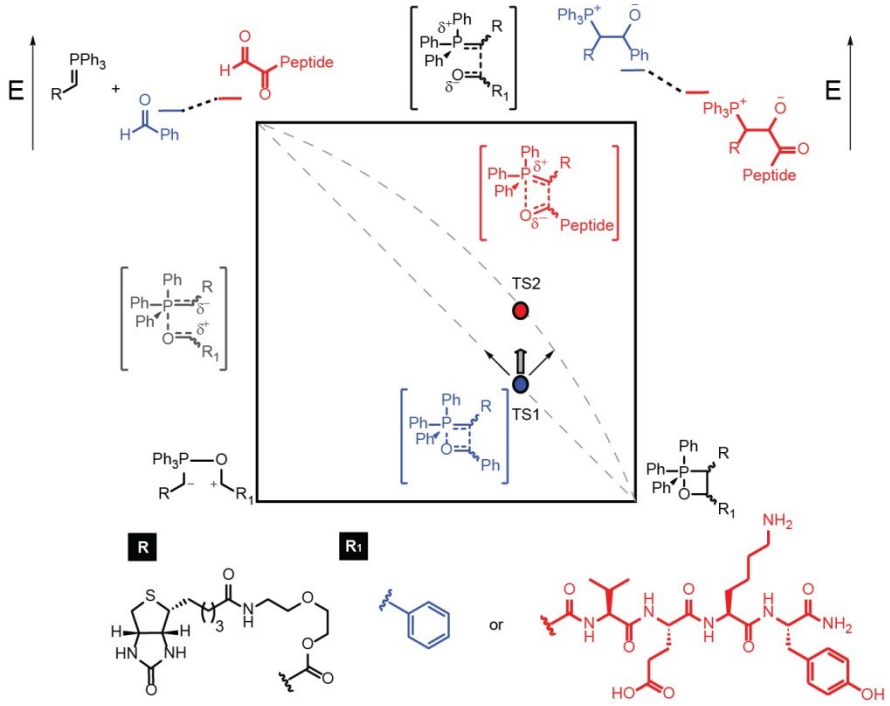
2.2.2 Wittig Reaction Between YEB and Model Aldehydes: Rate and Stereoselectivity

The YEB phosphonium salt, with expected pKa of ~10-11 in water,[452] was sufficiently deprotonated in pH 6-8 buffered solution to undergo Wittig reaction with benzaldehyde, 2-nitrobenzaldehyde and o-VEKY. Because the active species in the reaction is the phosphorane ylide which results after deprotonation of the phosphonium salt, it is expected that when the reaction is conducted at pH << pKa, the log(k) of such reaction should scale with pH and level off at pH near this pKa. In our studies, the rate increased exponentially with pH: the graph for the log(k) vs. pH for all types of aldehydes had identical slope of ~1 (Fig. 1C and Appendix A-1 to Appendix A-3) in the pH 6-8 range. The slope suggested that the rate-determining step involves an anion. Also, lack of any leveling of rate dependence at pH 8 suggested that these anionic species have pKa>9. This observation is in line with previously measured pKa of an ester stabilized phosphonium salt (10.5 in water)[452] and it differs from the reported value of 8.5 in DMSO.[453]

Analysis of substrate dependence of rate and stereoselectivity showed that benzaldehyde reacted with modest rate and produced high E/Z isomers selectivity (93:7 or 93% E) whereas 2-nitrobenzaldehyde reacted significantly faster but yielded 60-70% E isomer (Appendix A-4 to Appendix A-6). The loss in E-selectivity in *o*-nitrobenzaldehyde, when compared to benzaldehyde was in accord with a previous report by Bergdahl and coworkers.[454] The rate of reaction and E/Z isomers selectivity of oxoacetamide (50-70 % E) were similar to that of 2-nitrobenzaldehyde and significantly different from that of benzaldehyde. Similar E/Z isomers ratios were obtained *via* HPLC (Appendix A-7 and Appendix A-8) and ^{1}H NMR analysis of either the crude reaction mixtures, or the purified Wittig products (Appendix A-4 to Appendix A-6).

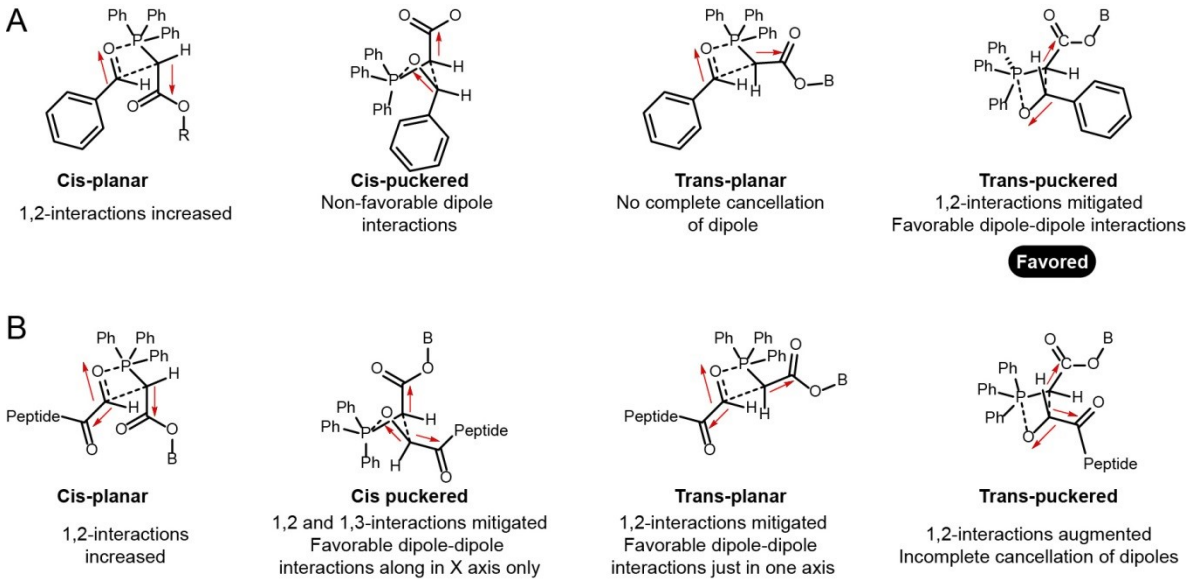
Our results are in agreement with previous published observations, the rate and stereoselectivity of the Wittig reaction are sensitive to changes in substrates (i.e. ylide and aldehyde) and reaction conditions. Many groups sought to propose a universal mechanism that encompasses all of the experimental observations collected to date.[455] In the case of the Wittig reaction *in* and *on water* with standard, electron rich aldehydes and stabilized ylides (reference[455,456] and references within), kinetically controlled formation of an O-apical oxaphosphetane *via* [2+2] cycloaddition through an asynchronous transition state is accepted as the rate determining step, followed by pseudo-rotation to an O-equatorial oxaphosphetane and final cycloreversion. However, the oxoacetamide generated after N-terminal serine oxidation is electronically different from regular benzaldehyde-like substrates and such a difference is manifested with significantly different E/Z ratio of products.

A comparison of transition states between the Wittig reaction of oxoacetamides and benzaldehyde can be visualized *via* More O'Ferrall Jencks (MOJ) diagram (Scheme 2). Lowering of the betaine corner due to the presence of the α-carbonyl that stabilizes the negative charge developed on the aldehyde oxygen results in an *anti*-Hammond movement towards the betaine. As a result, the TS2 for oxoacetamide should have more betaine character (or higher asynchronous character, where C-C bond is formed while P-O bond has significant charge character)[457] when compared to the TS1 for benzaldehyde-like aldehydes. We anticipated that such a shift in the transition state could be manifested as a different pH-dependence of rate for different aldehydes at the pH near the pKa of the betaine-like transition state (Scheme 2). Unfortunately, hydrolysis of the substrate and products above pH 10 complicated such studies.



Scheme 2-2. More O'Ferrall Jencks plot showing transition states of Wittig reactions that proceed *via* concerted oxaphosphetane (TS1) and movement of TS1 towards asynchronous betaine-like transition state (TS2) in response to changes in electronic properties of the aldehyde.

We note that the location of the transition state on MOJ plot alone does not reflect the 3D conformation of TS.[446,456,458] 1,2-interactions, 1,3-interactions and dipole-dipole interactions in *cis* and *trans* TSs towards oxaphosphetanes (OPA-TS) can influence the 3D conformation and determine the final E/Z selectivity. The latter statement however assumes irreversibility of the OPA formation, which has been observed experimentally[458,459] and suggested theoretically for benzaldehyde-like substrates.[446] The conformations of OPA-TSs were previously proposed to adopt a conformation that minimizes the overall dipole moment and 1,2 and 1,3-interactions simultaneously. A semi-puckered TS of reaction originating from the benzaldehyde substrate towards the *trans* OPA was proposed to minimize dipole-dipole interactions.[446,455,456,460] As the additional α-amide of oxoacetamide introduces an extra dipole, we hypothesize that the TS originating from oxoacetamide towards *trans* OPA cannot adopt the same geometry as benzaldehyde TS due to incomplete cancellation of dipole moment of the carbonyl group. In a similar analysis to the one done by Harvey and Aggarwal on α-alkoxy aldehydes,[457] we propose



Scheme 2-3. (A) Transition states towards oxaphosphetane of Wittig reaction between benzaldehyde and triphenylphosphine ester stabilized type ylides showing a conformer that minimizes 1,2 and 1,3-interactions as well as dipole-dipole interactions. (B) Analogous transition states towards oxaphosphetane of Wittig reaction between oxoacetamide and triphenylphosphine ester-stabilized type ylides cannot completely cancel dipole moment in any conformation.

that the additional dipole destabilizes the puckered *trans* TS, making it similar in energy to a nearly planar *cis* TS. Similar energy of the two TS decreases the expected stereoselectivity of the Wittig reaction (Scheme 2-3). Similar dipole-induced de-stabilization of the *trans*-puckered TS could potentially explain the lower E/Z selectivity of 2-nitrobenzaldehyde.

2.2.3 Assessment of Differential Reactivity of Oxoacetamide E/Z Wittig Products Towards Hydrolysis, Michael Addition and Diels-Alder Reactions

The Z-isomer produced in the Wittig reaction hydrolyzed significantly faster than the E-isomer (Appendix A-9). At pH 8, the half-life of hydrolysis of the E-ester group was $t_{1/2} = 18$ hours, while the Z-ester hydrolyzed completely in 7 hours. At pH 5, half-life of hydrolysis was >75 hours for both isomers. At these pH values, the rate of hydrolysis can be readily decoupled from the rate of Wittig reaction (Appendix A-10) and at pH 5, the product can be stored for three days without any hydrolysis. Although we have not shown it explicitly, selective hydrolysis of the Z-isomer can be used as one of the strategies to mitigate the poor E/Z-selectivity in the product and yield predominantly E-product.

The Wittig product contains an N-terminal ester of fumaric/maleic acid that could potentially form Michael addition products by attack of thiols. We tested glutathione, cysteamine and the peptide sequence DYKDDDDKC, aka "FLAG-Cys" peptide, with this Michael acceptor and rates were ranging from 0.12 to 4.1 $M^{-1}s^{-1}$ at pH 7.8 (Figure 2-2). We noticed a minor difference in reactivity in Michael addition of FLAG-Cys thiol to E and Z alkenes; (Figure 2-2C) however, both isomers were quantitatively converted to the Michael addition adducts. The Michael acceptor can form either irreversible or reversible covalent interactions with cysteine-like nucleophiles. Therefore, Wittig-diversified libraries can be potentially used for discovery of either irreversible or reversible covalent inhibitors. We evaluated the efficiency of Michael warhead using the approach of Tauton and co-workers.[441] Monitoring the retro-Michael reaction in isolated addition products of thiols with different pKa values (GSH and cysteamine) at pH $\geq$ 6.5[441,461,462] uncovered little to no retro-Michael reaction (Appendix A-11), indicating that these Michael acceptors favor irreversible additions. Lack of reversibility of analogous Michael adducts has also been reported by Bernardes and coworkers.[463]

The Wittig product can be "quenched" with cyclopentadiene *via* the Diels-Alder reaction to suppress any Michael addition to the olefin functionality. Aqueous Diels-Alder reaction in the model peptide exhibited a differential reactivity of E and Z isomers: after 48 hours the E-isomer exhibited 100% conversion but only minimal conversion for the Z isomer was observed (Appendix A-12). Identity of the residual Z olefin was confirmed by NMR (Appendix A-12D). Addition of co-solvents (acetonitrile or DMF) or Lewis acids (LiCl or $Cu(NO_3)_2$) known to catalyze the Diels-Alder reaction,[464,465] did not improve the reactivity of the Z isomer (Appendix A-13). Changing the pH had no effect on the reaction and, therefore, the Diels-Alder cycloaddition can be effectively performed at pH 5 in conditions that suppress hydrolysis of the ester. In these conditions, the E isomer is consumed after 20 hours, while conversion of the Z isomer is negligible (Appendix A-12B). Most importantly, the olefin that did not react in Diels-Alder preserved its ability to act as a Michael acceptor in reaction with FLAG-Cys thiol (Appendix A-12C). This observation was used for subsequent "pulse-chase" quantification of the Diels-Alder reaction on phage-displayed peptides.

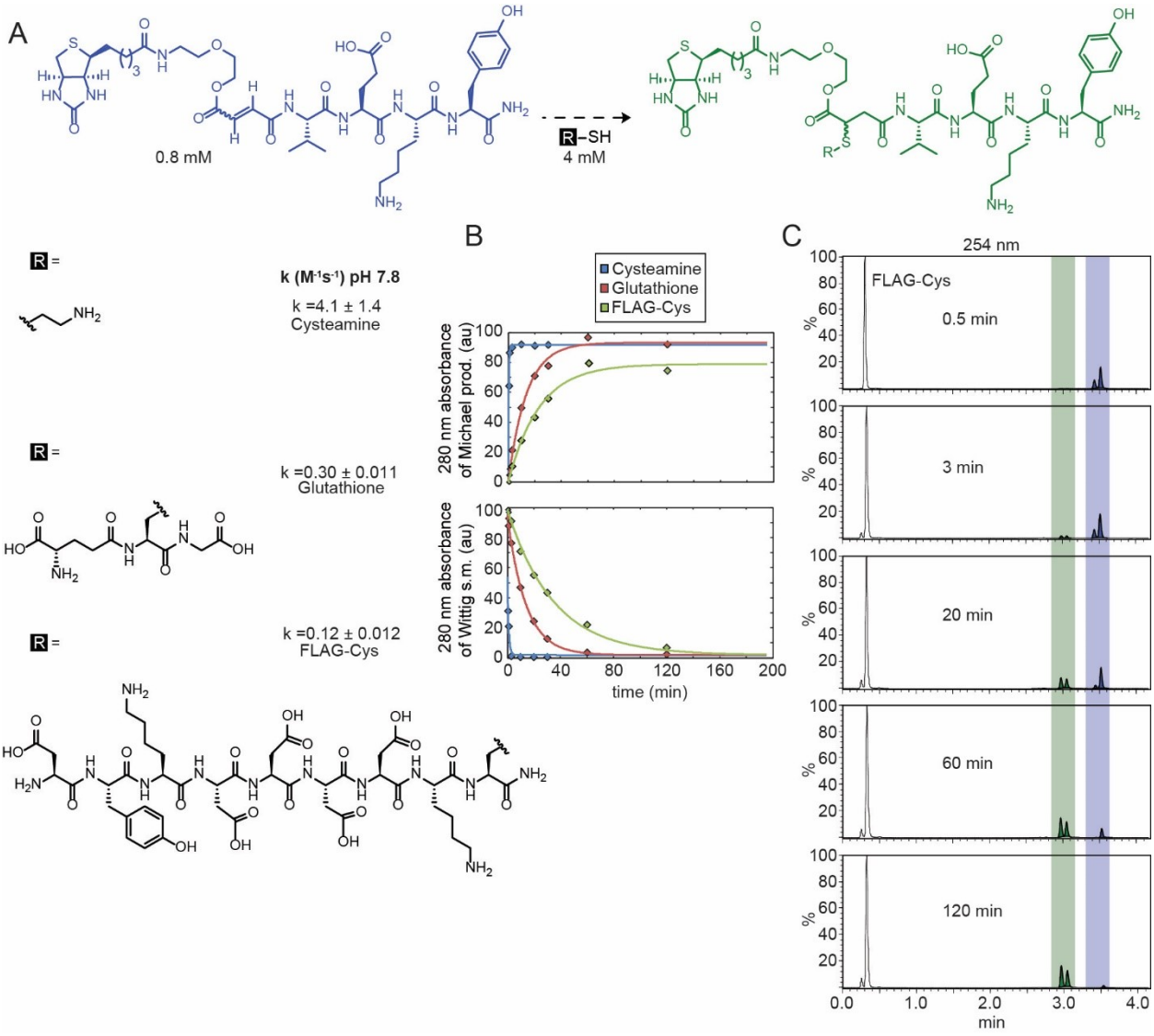


Figure 2-2. (A) Michael addition on YEB-VEKY. B) Kinetic traces for the tested three thiols. (C) LCMS traces (254 nm) for Michael addition using FLAG-Cys peptide. In the tested conditions, the cis isomer reacts faster than trans isomer

2.2.4 Generation of Wittig and Diels-Alder Functionalized Peptide Libraries on Phage

To confirm the yield and specificity of the Wittig reaction on phage displayed peptides, we employed a technology based on a streptavidin pull-down assay extensively used by our group.[427] Specifically, we use the phosphorane ylide containing a biotin tag (YEB) to modify a mixture of two phage types: (i) one displaying a SVEK sequence or SX_7 library of peptides displayed on pIII protein and *Lac*Z reporter; this phage produces blue plaques when plated on agar supplemented with X-Gal/IPTG and (ii) phage that displays N-terminal alanine and contains no *Lac*Z, forming

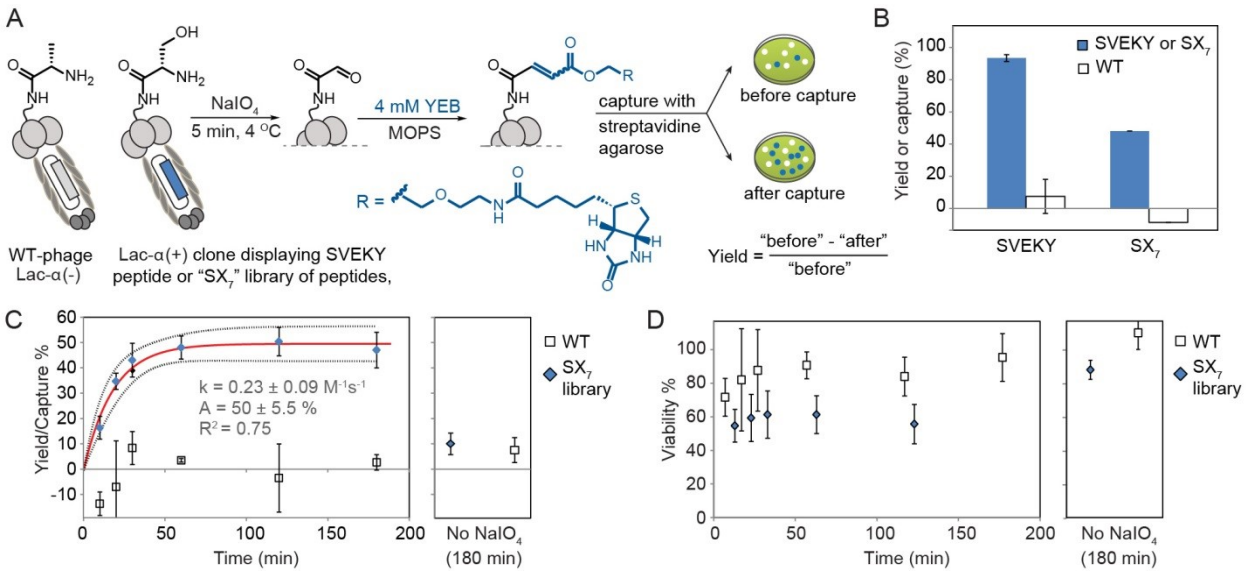


Figure. 2-3. (A) Quantification of yield and selectivity of Wittig reaction on phage using biotin-capture technique. (B) Capture % of phage clones displaying a SVEK-sequence and phage displaying the SX7 library of peptides in presence of WT phage as control at pH 7.8 after one hour reaction with 1 mM YEB. (C) Kinetics of reaction between SX7 library and YEB (4 mM) at pH 6.5. (D) Monitoring of viability of SX7 library in the reaction described in (C).

white plaques on the same plates. In regioselective functionalization only "blue" phages should acquire biotin and this biotinylation can be detected by pull-down with streptavidin beads. The biotin capture technique[427] confirmed that the reaction of YEB with clonal phage displaying SVEK peptide at pH 7.8 exhibited nearly 90% conversion after one hour, whereas wild type phage in the same reaction, which do not possess the N-terminal serine, exhibited insignificant reactivity (Figure 2-3B). The modification of 100,000,000-member SX_7 library of peptides displayed on phage reached 40-60% conversion in the same conditions (Figure 2-3B). Incomplete conversion of Wittig reaction in phage libraries was not due to low reaction rates, as shown by measurement of kinetics of modification of the library (Figure 2-3C). The half-life of library modification was ~20 min even at lower pH (6.5); the convergence exhibited a first order kinetic profile and reached saturation at ~50% after one hour. The same incomplete conversion has been observed in N-terminal modification of phage libraries.[405,406] To rule out sequence dependence of the oxidation step before Wittig reaction, SX_7 library was subjected to $NaIO_4$ treatment for different time intervals and then reacted with YEB. Increase in oxidation time did not change the percentage of

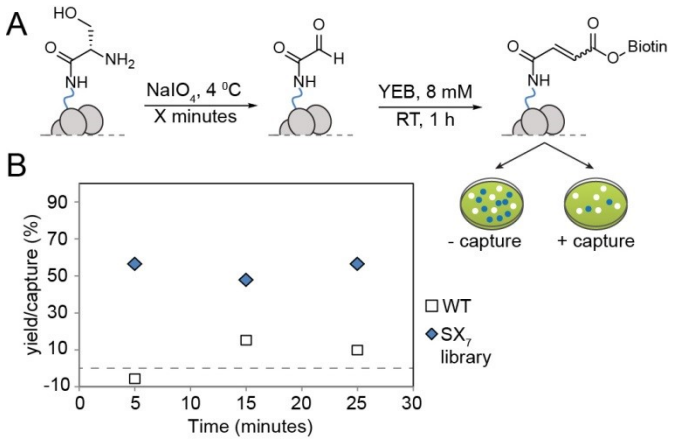


Figure 2-4. (A) Scheme of the time course of periodate oxidation followed by capping with YEB and biotin capture. (B) Conversion (%) of Wittig reaction as a function of oxidation time using biotin pull-down assay.

functionalization in the SX_7 library context (Figure 2-4). Wittig reaction, thus, effectively diversified phage displayed library of 10^8 peptides with only a modest impact on integrity of the library: phage "viability" decreased only to ~50% after three hours reaction, which is similar to viabilities observed in other reactions.[405,423]

We applied Michael addition on phage-displayed alkene-peptide obtained by Wittig derivatization of SVEK monoclonal phage to confirm the integrity of the Wittig product on phage. Removal of the excess reagent was critical for this reaction and we lowered the excess YEB from initial 4 mM to a sub-nanomolar concentration *via* dialysis of the modified phage against pH 5 buffer, which minimized possible hydrolysis. After 5-6 buffer exchanges were performed in 48 hours, we observed <20% loss of modification (Appendix A-14); the concentration of small organic molecule reference (fluorescein) present in this condition was reduced to single digit nM concentrations (Appendix A14-C). Purified Wittig product displayed on phage (YEB-VEK phage) reacted effectively with FLAG-Cys and were captured by anti-FLAG (Figure 2-5). Neither the WT phages nor purified YEB-VEK phage not exposed to FLAG-Cys peptide were captured in this condition (Figure 2-5). To rule out non-specific reactivity with thiol, we confirmed that the SVEK phage not exposed to YEB did not exhibit significant reactivity in Michael addition (Figure 2-6C).

The Wittig modified phage can also be diversified *via* Diels-Alder reaction with cyclopentadiene. To evaluate the efficiency of the Diels Alder reaction on phage, we employed

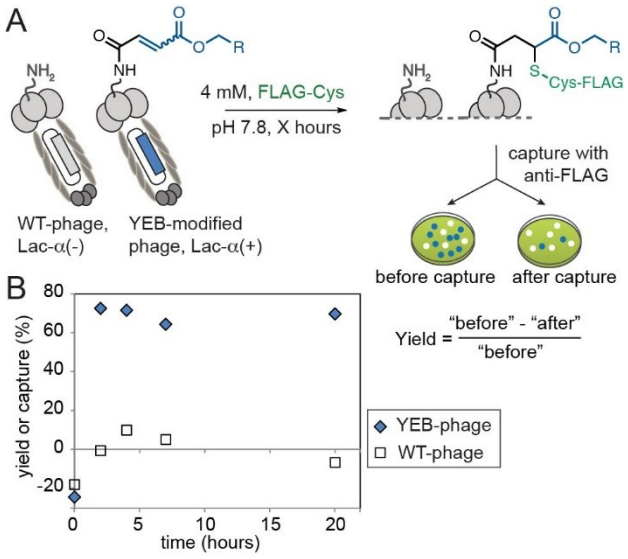


Figure 2-5. (A) Reaction of Wittig functionalized phage and control WT phage with DYKDDDDKC peptide (FLAG-Cys peptide). (B)Pull down assay with anti-FLAG immobilized on protein G magnetic beads quantified the yield of incorporation of FLAG and confirmed that FLAG was not incorporated into WT-phage

the pulse-chase technique developed in our group.[405,427] Specifically, exposure of purified YEB-modified SVEK phage (Figure 2-6B) to FLAG-Cys (pH 7.8, 2 h, RT) renders phage tagged only if it has accessible Michael acceptor. As phage was "pulsed" with cyclopentadiene at pH 5 for 48 hours, the alkene was consumed in the Diels-Alder reaction and the labeling in the "chase" reaction with FLAG-Cys peptide decreased. The pulse-chase approach suggested that 70% of the population modified *via* Wittig reaction was converted to a norbornene product (Figure 2-6D left bar). We recognized that the loss of Michael reactivity in "pulse chase" approach can be due to: i) blocking of the Michael acceptor due to Diels-Alder reaction, ii) hydrolysis of the ester in α-position to the double bond, iii) reaction with other nucleophiles present in the reaction buffer. To rule out possibilities ii) and iii), we demonstrated that Wittig-modified phage preserved its ability to react with FLAG-Cys peptide even after incubation with pH 5 buffer for 48 hours (Figure 2-6D middle bar). We note that incomplete (70%) conversion in Diels-Alder reaction mirrors the lack of reactivity of the Z Wittig adduct observed in reaction on the synthetic peptide model. While direct structural analysis of E/Z isomers on phage-displayed peptides is presently impossible, the results of pulse-chase study strongly suggest that: i) Wittig product is formed on phage in a similar E/Z ratio as was observed with the model synthetic peptide (~70% E), ii) Diels-Alder reaction

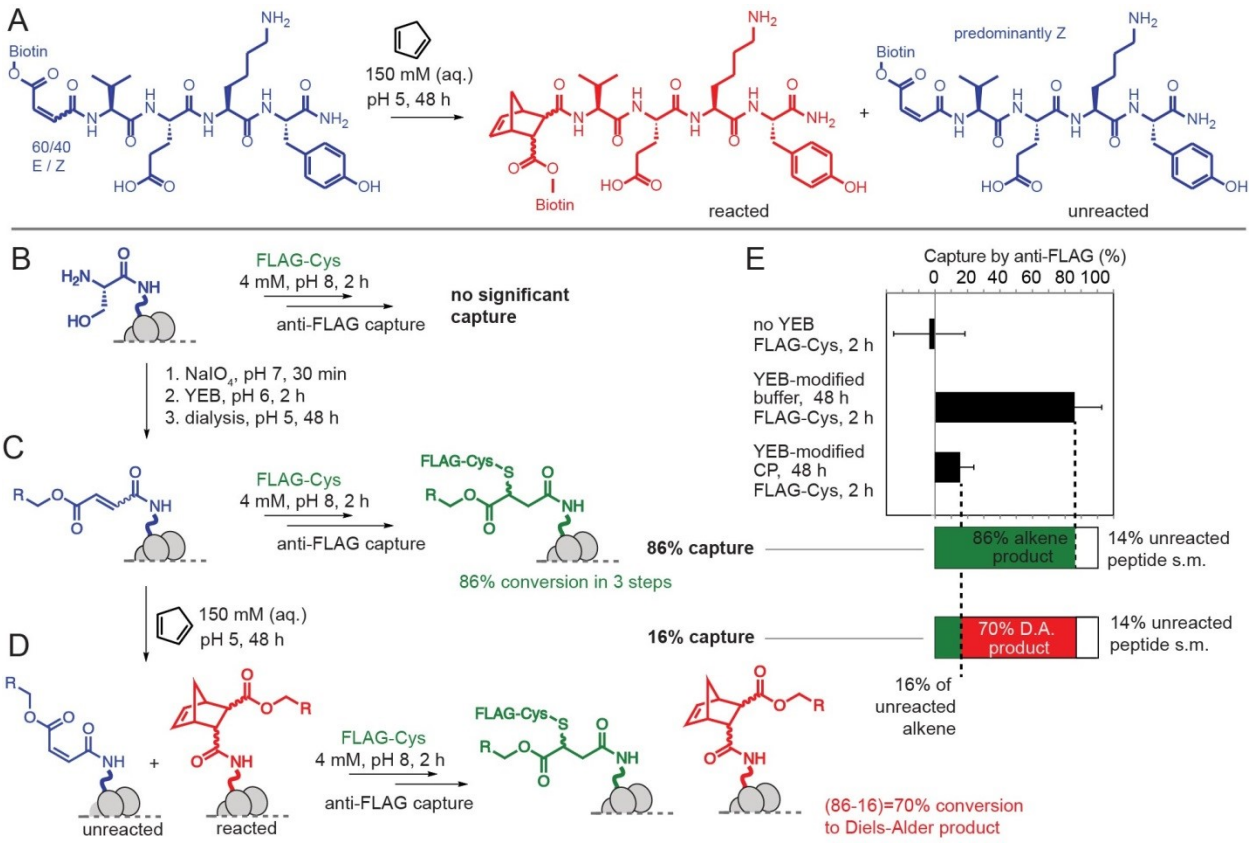


Figure 2-6. (A) Diels-Alder reaction on Wittig product of model peptide. E selectivity was confirmed by analyzing the recovered unreacted starting material by NMR (Appendix A-12D). (B) Wittig reaction on phage clone displaying SVEK peptide. Phage not subjected to Wittig reaction did not exhibit any detectable capture whereas (C) alkene-functionalized phage, when reacted with FLAG-Cys peptide captured by anti-FLAG yields 86% capture. (D) Exposure of phage for 48 h with 250 mM cyclopentadiene after which an aliquot was taken and reacted with FLAG-Cys peptide quantified the efficiency of Diels-Alder modification for 2 h. (E) Bar graph summary: 86% of the phage population reacted with FLAG-Cys and was captured by anti-FLAG when no cyclopentadiene was added. After exposure to cyclopentadiene for 48 h, only 16% of phage remained reactive towards FLAG-Cys. These observations extrapolate a 70% conversion for Diels-Alder reaction.

selectively converts phage-population displaying the E-isomer to *trans*-norborene products; iii) subset of library remaining after Diels-Alder reaction is significantly enriched in Z-isomer.

2.3 Conclusions

In conclusion, we have successfully employed the Wittig reaction to diversify a library of peptides displayed on phage. The modification is highly regio- and chemo-selective and it exhibits

only a minimal effect on the infectivity of phage. We further diversified the Wittig-products displayed on phage *via* Michael addition and Diels-Alder reactions. The efficiency of both transformations can be monitored in phage-displayed context; neither of these reactions have a significantly impact on the infectivity of the virions. These transformations broaden the scope of reactions compatible with genetically-encoded libraries of phage-displayed peptides and increase the potential of this platform for identification of diverse ligands for important protein targets.

The mechanistic hypothesis underpinning the E/Z selectivity in peptide aldehydes can be further attacked by profiling of reactions in various peptide substrates that contain amino-acid side chains capable of screening or enhancing dipole-dipole interactions. The most important conclusion emanating from the pH profile is the invariance of E/Z selectivity at all tested pH values. Increase in proton concentration, thus, does not significantly change the location of the transition state nor the mechanism of the reaction; the location of the TS is only substrate dependent (Appendix A-4). In peptide-aldehyde substrates, neighboring amino acid side chains can exhibit measurable effects on the transition state of the Wittig reaction.

Implementation of the Wittig reaction in a population of 100 million diverse peptides open doors to numerous mechanistic studies that involve selection of substrates of higher reactivity and even selection of subsets of library that yield exclusively E or Z products (e.g., using a differential reactivity of these isomers in a Diels-Alder or ester hydrolysis reactions). We believe that fundamental mechanistic investigations of Wittig reaction and a downstream Michael addition and Diels-Alder reactions will serve as a robust foundation for future substrate profiling studies within a library of genetically-encoded substrates. This proposal encouraged us to develop a technology for profiling libraries of peptide aldehydes displayed on phage to discover sequences with privileged reactivity for the Wittig reaction. These results are described in Chapter 3 by measuring rates in ~10^5 peptides at once.

2.4 Experimental Procedures

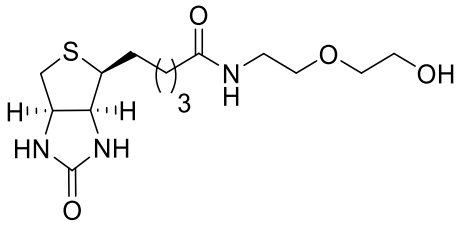
2.4.1 Materials and General Information

PBS 10X (0.1 M) buffer was prepared by mixing NaCl (80 g), KCl (2 g), Na_2HPO_4 (14.4 g), KH_2PO_4 (2.4 g) in 1 L of milliQ water (pH is adjusted to 7.4 or the needed values); Tris buffer was

prepared as a solution of 50 mM Tris and adjusted with NaOH and HCl to the required pH values. MOPS buffer was prepared as a solution of 100 mM MOPS and adjusted to pH 6.5. All solutions used for phage work were prepared with milliQ water. The peptide (SVEKY) was synthesized using the Prelude X peptide synthesizer (Gyros Protein Technologies) using 200 mg Rink Amide AM resin. (Triphenylphosphoranylidene)ketene (Bestmann ylide), was purchased from TCI America (#15596-07-3). D-(+)-biotin was purchased from AK Scientific Inc (#58-85-5). Carbonyldiimidazole was purchased from ChemPep (CDI, #530-62-1). 2-(2'-aminoethoxy) ethanol was purchased from Sigma Aldrich. THF was dried with Na/Benzophenone and distilled. Product purification was accomplished with automated chromatography machine (CombiFlash® Rf, Teledyne Isco, Inc.) and preparative HPLC (Waters 2489). ^{1}H and ^{13}C NMR spectra were acquired on a 600 MHz four channel Agilent VNMRS spectrometer, equipped with a z gradient HCN probe and using VNMRJ 4.2A as the acquisition software. When specified, samples were dissolved/recorded in 90% H_2O/10% D_2O mixtures and suppression of the H_2O signal was performed using either presaturation or excitation sculpting.[466] Chemical shifts are reported in ppm and J couplins in Hz. The following abbreviations classify the multiplet peaks in the ^{1}H NMR: s = singlet, d = doublet, m = multiplet or unresolved. HRMS (ESI) spectra were recorded on Agilent 6220 oaTOF mass spectrometer using either positive or negative ionization mode. Characterization of reaction crude was performed with UPLC-MS using a C18 column (Phenomenex Kinetex 1.7 μm EVO C18, 2.1×50 mm) running with a gradient of water/acetonitrile with 0.1% formic acid from 98/2 at 0 min to 40/60 at 5 min under a flow rate of 0.5 mL/min. HPLC kinetics were performed in an Agilent 1100 Series using a Thermo Scientific C18 column (Hypersil GOLD, 3 μm, 50×2.1mm). The phage library (SX_7) was supplied by New England Biolabs and prepared as detailed in our previous publication.[467] Phage displaying SVEK sequence was a generous gift of Chris Noren and Beth Pascal (New England Biolabs).

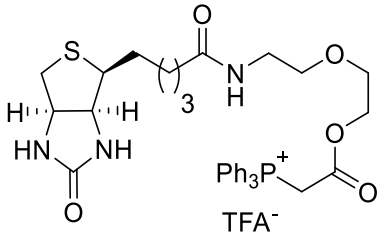
2.4.2 Synthesized Compounds

2.4.2.1 Synthesis of Biotin-PEG



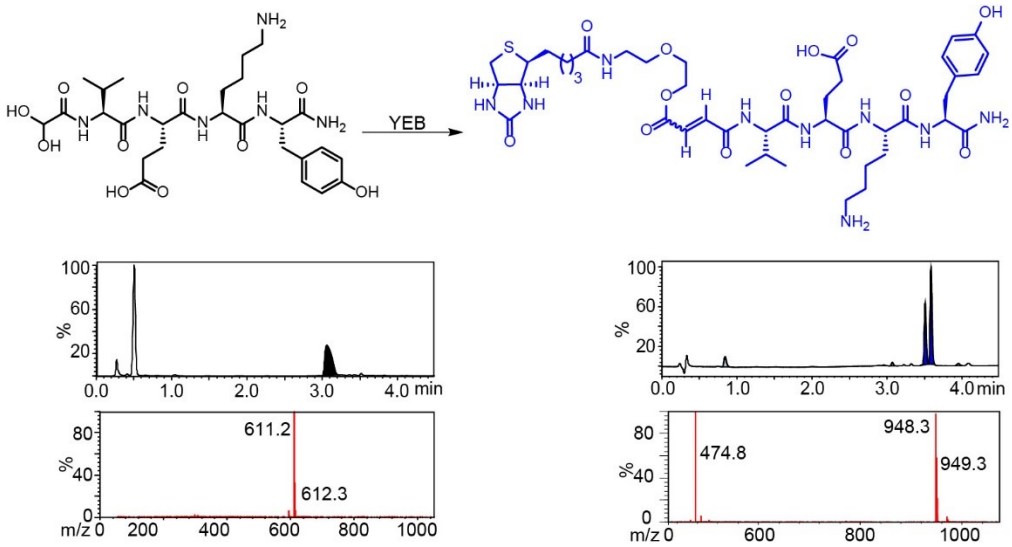
D-(+)-Biotin (1.0 eq, 0.30 g) was dissolved in DMF (60 mL) at 55 °C. The solution was then brought to room temperature. Carbonyldiimidazole (1.7 eq, 0.34 g.) was added in DMF (1.5 mL) to the reaction mixture, which was stirred for 4 hours at room temperature. 2-(2'-Aminoethoxy) ethanol (3.9 eq, 0.50 g, 0.48 mL.) in DMF (3.0 mL) was added to the reaction mixture which was stirred for another 16 hours. The solvent was removed using rotary evaporator and the crude residue was purified on silica gel (80 g) with a gradient of 0–7.5% methanol in DCM using CombiFlash® Rf to yield the product as a white solid (0.29 g, 72% yield). ^{1}H NMR (400 MHz, DMSO) δ = 7.80 (t, *J* = 5.6 Hz, 1 H), 6.41 (s, 1 H), 6.34 (s, 1 H), 4.56 (app. t, *J* = 5.2 Hz, 1 H), 4.30 (dd, *J* = 7.6, 5.2 Hz, 1 H), 4.14-4.08 (m, 1 H), 3.51-3.47 (m, 2 H), 3.42-3.38 (m, 4 H), 3.21-3.16 (m, 2 H), 3.12-3.07 (m, 1 H), 2.82 (dd, *J* = 12.4, 4.8 Hz 1 H), 2.57 (d, *J* = 12.4, 1H), 2.06 (t, *J* = 7.4, 2H), 1.61-1.27 (m, 6H). ^{13}C NMR (100 MHz, DMSO) δ = 172.6, 163.2, 72.6, 69.6, 61.5, 60.6, 59.7, 55.9, 39.4, 38.9, 35.6, 28.6, 28.4, 25.7.

2.4.2.2 Synthesis of Ylide-Ester-Biotin (YEB)



Biotin-PEG (1.0 eq, 0.10 mg,) was charged in a two-neck round bottom flask. The flask was evacuated and filled with N_2 three times. A mixture of dry THF (50 mL) and dry DMF (2.0 mL) were added into the flask and the solid was dissolved under reflux. Bestmann ylide (1.2 eq, 0.11 g) was weighed and added to a vial that was then evacuated and filled with N_2 three times. Dry THF (5.0 mL) was used to dissolve and rinse resulting solution into the round bottom flask *via* cannula. The solution was refluxed for 4 hours, concentrated and purified by HPLC using an amide column (Waters XBridge BEH prep OBD Amide 5 μm 19×250 mm) running with a gradient of

acetonitrile/water both with 0.1% TFA from 98/2 at 0 min to 50/50 at 30 min under a flow rate of 8 mL/min. The THF salt ylide was obtained as a golden oil (140 mg, 62% yield). ^{1}H NMR (400 MHz, DMSO) δ = 7.88-7.69 (m, 16 H), 6.38 (b, 2 H), 5.38 (s, 1 H), 5.34 (s, 1 H), 4.27 (dd, *J* = 7.4, 4.6 Hz ,1 H), 4.15-4.11 (m, 3 H), 3.38-3.35 (m, 2 H), 3.26 (t, *J* = 6.2 Hz, 2 H), 3.14-3.06 (m, 3 H), 2.81 (dd, *J* = 12.6, 4.8 Hz, 1 H), 2.57 (dd, *J* = 12.4, 4.8 Hz, 1 H), 2.06 (t, *J* = 7.2 Hz 2 H), 1.62-1.26 (m, 6 H). ^{13}C NMR (100 MHz, DMSO) δ = 172.2, 164.6, 162.7, 135.1 [3], 135.0 [6], 133.7 [6], 119.3, 118.5, 117.7, 68.9, 67.5, 65.4, 61.1, 59.2, 55.4, 38.2 [2], 35.08, 29.6, 29.2, 28.1, 25.2.



2.4.2.3 Synthesis of o-VEKY-Wittig product

HPLC purified o-VEKY peptide (1.0 eq, 2.0 mg) was dissolved in MOPS buffer (0.30 mL, pH 6.5, 0.10 M). The solution was added into a 1.7 mL epi-tube containing YEB (2.0 eq, 4.9 mg). The mixture was incubated for 3 hours at RT, diluted to 1.0 mL with milliQ H_2O and purified by HPLC using a preparative C18 column (Waters SymmetryPrep C18 5 μm 19×50 mm) running a gradient of water/acetonitrile both with 0.1% TFA from 98/2 at 0 min to 50/50 at 30 min under a flow rate of 8 mL/min. The acetonitrile was removed using speed-vac and the sample was lyophilized overnight to yield the adduct (o-VEKY-Wittig) as a white solid (1.2 mg, 38% yield).

2.4.2.4 Synthesis of o-VEKY Michael addition adducts

o-VEKY-Wittig (1.0 eq, 1.0 mg) was dissolved in of 10X PBS (0.10 mL). Each thiol was added as a solution in 10X PBS (0.10 mL, 2.0 eq) and the reactions were incubated at RT for 3

hours. The mixtures were diluted to 1.0 mL with H_2O and purified by HPLC using a C18 column (Waters SymmetryPrep C18 5 μm 19×50 mm) running a gradient of water/acetonitrile both with 0.1% TFA from 98/2 at 0 min to 50/50 at 30 min under a flow rate of 8 mL/min. The acetonitrile was removed using speed-vac and the sample was lyophilized overnight to yield the Michael addition adduct as a white solid (0.70 mg), 63% yield for cysteamine; 1.2 mg, 54% yield for DYKDDDDKC).

2.4.2.5 Diels Alder Reaction on YEB-VEKY

A solution of YEB-VEKY (2.0 μL, 11 mM in water) was mixed with cyclopentadiene (0.25 μL) and the buffer, co-solvent and Lewis acid as described in Table 1. The mixtures were incubated for 48 hours at RT in a Labquake tube rotator (Barnstead/Thermolyne) after which an aliquot (2.0 μL) was diluted 10-fold with water and injected into HPLC to monitor the reaction. Another aliquot (2.0 μL) from one of the reactions was also injected into LCMS in order to obtain the mass values for the peaks.

2.4.3 Kinetic Studies

2.4.3.1 Kinetic Studies for the Wittig Reaction of Aldehydes with YEB at Different pH Values

Aqueous YEB solution (5.00 eq, 13.8 μL, 34.8 mM) was added to solutions of aldehydes in 100 mM buffer (1.00 eq, 106 μL, 0.900 mM). The reactions were incubated at RT. At each time interval, an aliquot of the mixture (10 μL) was quenched by mixing with HCl (1 μL, 1 M). The quenched mixtures were then analyzed by UPLC-MS or HPLC to characterize the progress of the reaction. See Appendix-7 and Appendix-8 for the spectra and Appendix-1 to Appendix-3 for kinetic traces. All reactions were performed under pseudo-first-order condition. Fitting of the kinetic curve to the equation $A_t = 1 - e^{-k*[YEB]*t}$ yielded the second-order rate constant k, where A_t is the fraction of the product at time t and [YEB] is 0.004 M. MATLAB script used to fit the kinetic curve is outlined in Appendix A-16.

Table 1. Screening of conditions for Diels Alder reaction between o-VEKY Wittig and cyclopentadiene

Entry	pH buffer	μL buffer	μL DMF	μL acetonitrile	μL LiCl	μL $Cu(NO_3)_2$
1	5	17.5	-	-	-	-
2	6	17.5	-	-	-	-
3	5	15.5	2.00	-	-	-
4	6	15.5	2.00	-	-	-
5	5	13.5	4.00	-	-	-
6	6	13.5	4.00	-	-	-
7	5	15.5	-	2.00	-	-
8	6	15.5	-	2.00	-	-
9	5	13.5	-	4.00	-	-
10	6	13.5	-	4.00	-	-
11	5	9.00	-	-	8.80	-
12	6	9.00	-	-	8.80	-
13	5	16.5	-	-	-	1.20
14	6	16.5	-	-	-	1.20

2.4.3.2 Kinetic Studies for Wittig Reaction of o-VEKY with YEB at Different Buffer Concentrations

Aqueous YEB solutions (5.00 eq, 13.8 μL, 34.8 mM) were added to solutions of aldehydes in PBS buffer with pH 7.50 at different concentrations (1.00 eq, 106 μL, 0.900 mM). The reactions were incubated at RT. At each time interval, an aliquot of the mixture (10 μL) was removed from the reaction and quenched by mixing with HCl (5 μL, 1 M). The quenched mixtures were then analyzed by HPLC to characterize the progress of the reaction. See Appendix A-7 and A-8 for model chromatograms and Appendix A-1 to A-3 for kinetic traces. All reactions were performed under pseudo-first-order conditions. Fitting of the kinetic curve to the equation $A_t = 1 - e^{-k*[YEB]*t}$

yielded the second-order rate constant k, where A_t is the fraction of the product at time t and [YEB] is 0.004 M. MATLAB script used to fit the kinetic curve is outlined in Appendix A-16.

2.4.3.3 Kinetic Studies for Michael Addition of YEB-VEKY with Different Thiols

A solution of YEB-VEKY product (82.8 μL, 1.00 mM, 1.00 eq) in MOPS (pH 6.50, 100 mM) was mixed with an aliquot of the thiol stock solutions (7.2 μL, 50 mM, 5.0 eq). The reactions were incubated at RT. At each time interval, an aliquot of the mixture (5 μL) was quenched by mixing with HCl (1 μL, 1 M). The quenched mixtures were then analyzed by UPLC-MS to characterize the progress of the reaction (see Figure 2-2 for kinetic traces). All reactions were performed under pseudo-first-order condition. Fitting of the kinetic curve to the equation $A_t = 1 - e^{-k*[THIOL]*t}$ yielded the second-order rate constant k, where A_t is the fraction of the product at time t and [Thiol] is 0.004 M.

2.4.4 E/Z Ratio Determination

2.4.4.1 E/Z Selectivity Determination Following Purification

o-VEKY (1 eq, 1 mg) was dissolved in 10X PBS (1 mL) at the indicated pH values (6.5, 7.5, 8.5) and mixed with YEB (2.2 eq, 3.3 mg). The reaction mixture was incubated for 1.5 hours and purified by HPLC with a C18 preparative column using a 30 minute gradient with 5-50% acetonitrile in water both with 0.1% TFA. The acetonitrile was removed using speed-vac for 40 minutes at 40 °C and the peptide derivative was obtained as white solid after lyophilizing the aqueous residue overnight. All the obtained solid was dissolved in D_2O (750 μL) for NMR analysis.

2.4.4.2 *In situ* E/Z Selectivity Determination

A solution of aldehyde (12.3 μL, 16.3 mM, 1.00 eq, 0.200 mmol) was mixed with 175 μL of 10X PBS at the specified pH values. A solution of YEB (37.4 μL, 26.7 mM, 5.00 eq, 1.00 mmol) was added. The mixture was incubated for 3 hours (pH 6.3) or 1.5 hours (pH 7.4) after which 25 μL of D_2O were added. Samples were transferred into 3 mm diameter NMR tubes and the ^1H NMR spectra were collected after solvent suppression was performed as indicated in Section 2.4.1.

2.4.5 Hydrolysis Monitoring

2.4.5.1 Hydrolysis Monitoring of YEB-VEKY by NMR

Wittig product (1 mg) as an E/Z isomer mixture was dissolved in 1X PBS/D_2O (675/75 μL) and loaded into an NMR tube. Solvent suppression was performed and the ^1H NMR was monitored at the specified time intervals (Appendix A-9).

2.4.5.2 Hydrolysis Monitoring of YEB-VEKY LCMS at Different pH Values

Solutions of Wittig product (1.0 mM) in the specified buffers (10 mM) were prepared. The hydrolysis rate was monitored using UPLC at the specified time intervals (Appendix A-10) by injecting an aliquot of 2.0 μL.

2.4.6 Phage Modifications

2.4.6.1 Modification of SX_7 Library and SVEK Phage with YEB

A mixture of WT phage and SX_7 library or SVEK phage (1:1 ratio) in PBS buffer (pH 7.8, 0.10 M) was prepared to yield a final titer of 2×10^{12} pfu mL^{-1}. To the mixture of phage (98 μL), $NaIO_4$ (1.0 μL, 6.0 mM solution in water) was added and the reaction was incubated on ice for 5 minutes. YEB (0.1 mL, 2 mM in water) was added and the reaction mixture was incubated at RT for 1 h. The efficiency of the modification was quantified by biotin-capture assay as described previously.[427]

2.4.6.2 Kinetic Studies of Modification of SX_7 Phage Library with YEB

A mixture of phage SX_7 library and WT phage (1:1 ratio) in MOPS (pH 6.5, 0.10 M) was prepared to yield a final titer of 2×10^{12} pfu mL^{-1}. To the mixture of phage (98 μL), $NaIO_4$ (1.0 μL, 6.0 mM solution in water) was added and the reaction was incubated on ice for 5 minutes. YEB (0.1 mL, 8 mM in water) was added to the reaction mixture, which was incubated at RT. At each time interval, a 2 μL aliquot was removed from the reaction and quenched by diluting 10^5 times. The efficiency of the modification was quantified by biotin-capture assay as described previously.[427] The studies—modification and quantification—were repeated three times on separate days to validate the reproducibility of the experiments. The reactions were performed

under pseudo-first-order condition (i.e., modifying reagent was present in a large excess when compared to the concentration of peptide substrate displayed on phage).

2.4.6.3 Modification of SVEK Phage with YEB Followed by Reaction with DYKDDDDKC.

A mixture of SVEK phage and WT phage in MOPS buffer (pH 6.5, 0.10 M) was prepared to yield a final titer of 2×10^{12} pfu mL^{-1}. To the mixture of phage (98 μL), YEB (0.10 mL, 8.0 mM in water) was added. The reaction mixture was incubated at RT for 3 h, injected into a dialysis cassette and dialyzed for 48 h against acetate buffer (10 mM at pH 5.0, at least 6 buffer exchanges). To an aliquot (10 μL) of purified, Wittig functionalized phage (1×10^{11} pfu mL^{-1}), a solution of DYKDDDDKC peptide was added (10 μL, 8.0 mM in PBS pH 8.0, 0.10 mM). The reaction was incubated at RT for 2 hours. An aliquot (2.0 μL) was diluted 10^5 times at the time intervals described in Figure 2-5. The diluted phage solution (0.30 mL) was incubated with 30 μL of protein G beads for 1 hour at RT in a Labquake tube rotator. The protein G beads (Promega) had been previously coated with anti-FLAG antibody (10 μg antibody in 300 μL 2% BSA in PBS 1X, overnight in Labquake tube rotator at 4 °C). The efficiency of the modification was quantified by FLAG-capture assay, by using steps analogous to those of biotin-capture assay described previously.[427]

2.4.6.4 Modification of SVEK Phage with YEB Followed by Reaction with Cyclopentadiene and Chase with DYKDDDDKC.

Cyclopentadiene was prepared by thermal cracking of dicyclopentadiene.[468] To an aliquot of purified, Wittig functionalized phage (1×10^{11} pfu mL^{-1}) in a total volume of 0.10 mL in acetate buffer (10 mM, pH 5.0), freshly cracked cyclopentadiene (1.3 μL) was added to reach a final concentration of 0.25 M. The reaction was incubated at RT in a Labquake tube rotator for 48 h. An aliquot of the cyclopentadiene modified phage (10 μL), was then mixed with a solution of DYKDDDDKC peptide (10 μL, 4.0 mM in PBS pH 8.0 10X). The reaction was incubated at RT for 2 hours. An aliquot (2.0 μL) was diluted 10^5 times, and the phage dilution (0.30 mL) was incubated with 30 μL of protein G beads previously coated with anti-FLAG antibody (10 μg antibody in 0.30 mL 2.0% BSA in 1X PBS, overnight coating in Labquake tube rotator at 4 °C) for 1 hour at RT in a Labquake tube rotator. The efficiency of the modification was quantified by

FLAG-capture assay, using steps analogous to those of biotin-capture assay described previously.[427] The same capture methodology was performed in the Wittig functionalized phage before cyclopentadiene treatment. Diels-Alder reaction efficiency was calculated by subtraction of Michael addition efficiency before and after incubation with cyclopentadiene.

Chapter 3: Genetically-Encoded Substrate Profiling of Peptide Sequences for the Wittig Reaction

3.1 Introduction

Profiling of multiple substrates in the same reaction conditions is a cornerstone of mechanistic organic chemistry. Optimization of chemical reactions, discovery of catalytic systems, and mechanistic studies hinge on measurements of rates/conversions of multiple substrates and conditions. Although measurement of the rates of plurality of substrates in different reaction conditions one by one could be time consuming,[469] the data collected in such Structure Activity Relationship (SAR) studies serve as critical input for development of mechanistic hypotheses and decision making in discovery of new reactions. Starting from pioneering work of Hammet and co-workers on linear free energy relationships (LFER)[470] to modern approaches that employ Multiple Linear Regression (MLR)[471] and other Machine Learning (ML) methods,[472] these methodologies permit agglomeration of observations from SAR studies and building of qualitative models that relate reactivity to observable physical properties such as pKa or theoretically calculated parameters such as HOMO/LUMO energies. These models combined with "qualitative chemical intuition" allow predicting optimal conditions and substrates for a particular reaction and provide critical insight in reaction mechanisms. The most valuable input into LFRE/MLR/ML models consists of both, "positive" (i.e. fast reactions) and "negative" data (i.e. slow or unsuccessful reactions). Same requirement exist in other machine learning fields: only sets with positive and negative observations can serve as appropriate input for training of models.[473–475] Effective mechanistic understanding of reactions critically depends on development of methods that allow collecting a large unbiased set of reactivity data; in general, increase in the size of data allows minimizing the bias that could originate from human decision making (i.e. conscious selection of substrates with anticipated reactivity).

High-throughput screening (HTS) techniques developed originally for discovery of pharmaceutically valuable molecules[476] have later been repurposed for high throughput acquisition of SAR information and facilitated not only optimization of chemical reactions[477] but also discovery of catalytic systems,[478–481] and unexpected chemical transformations.[476] Both one-well-one-reaction[111] and one-well-multiple-reactions[482] HTS strategies have been used for reaction discovery. In the former approach, the size of collected data set scales proportionally with the

number of different reactions set up per unit time. The one-well-multiple-reactions approach strives to make the screening of multiple experiments faster and more practical. The disadvantage of conducting multiple reactions in one well is the potential for synergistic and interfering effects of multiple substrates that do not exist in the same reactions that contain only one substrate.[111] Performing reaction between one reagent R and multiple substrates S under pseudo first-order kinetics offers a fundamental way for overcoming cross-reactivity in mixed solution.

Desired rate:

$R + S_i \rightarrow P_i$ $\quad d[P_i]/dt = k[R]_0[S_i] = k'[S_i]$ $\quad t_{1/2} = \ln 2/k'$

Cross substrate rate:

$S_i + S_j \rightarrow CP_{ij}$ $\quad d[CP_{ij}]/dt = k[S_i][S_j]$ $\quad t_{1/2} = \frac{1}{k([Si^0]-[Sj^0])} \ln \frac{[Sj^0]}{2[Sj^0]-[Si^0]}$

Multi substrate rate:

$R + S_i + S_j \rightarrow MP$ $\quad d[MP_i]/dt = k[R]_0[S_i][S_j] = k'[S_i][S_j]$

By maintaining the concentration of reagent $[R]_0$ high and decreasing the concentration of all substrates $[S_i]$ in mixture, the characteristic time (e.g. $t_{1/2}$) of conversion of desired transformations remains constant, while the characteristic times of all reactions that involve cross-reactivity between screened substrates becomes disproportionally large. The only problem with decreasing concentrations of substrates is decreased amount of product generated in the reaction which, in turn, leads to challenges in detection of product by traditional spectroscopic tools. Genetic Encoding (GE) of substrates by DNA or RNA tags that can be amplified from single copy number molecules helps overcoming this drawback. Both genetically-encoded peptide libraries[476,483] and DNA-encoded small molecule libraries[484–488] offer added advantages of screening at low concentrations that minimize undesired reactions while conveniently increasing the number of substrates that can be interrogated in one solution to 10^6-10^{12} scale.

To date, interrogation of peptide libraries and DNA-encoded small-molecule libraries was limited to screens that aim to find singular sequences that exhibit enhanced reactivity when

compared to an "average" or random peptide sequence. Examples are reports that have used GE-peptide libraries of polypeptides displayed on phage to discover proteins[489] and peptides[490] that exhibit faster aldolase activity than an average sequence, peptides that form stable enaminones[491] or react with acyl hydrazides more effectively than random peptides[492,493] and peptide sequences flanking an unnatural propargyl glycine residue that reacted 5x faster in the Sonogashira cross-coupling than the same residue surrounded by random amino-acids.[416] While such findings have important utility in areas like biorthogonal chemistry and production of antibody-drug conjugates, where fast and efficient labeling of species is beneficial,[431] there are presently no examples showing that findings from GE-approaches can fuel an investigation of SAR of reactions. The latter, in turn, rely on availability of data about fast, medium and slow substrates. Here we describe a functional SAR pipeline that starts from a mixture of 10^5-10^6 peptide sequences displayed on phage and yields a complete profile of a given reaction by extrapolating the rate of the reaction from copy number of tagged and pulled-down substrates measured by Illumina sequencing. This SAR illuminated the role of functional groups in stabilizing or destabilizing the transition state *via* short or long-range non-covalent interactions.

Design of genetically-encoded SAR profiling was based on several criteria: (I) we repurposed classical tagging strategies commonly employed for physical separation of the reacted sub-set of library members from the remaining unreacted population.[427,467,494] The most common approach is introduction of biotinylation handle followed by separation *via* affinity pull-down with streptavidin. (II) As library of substrates, we employed phage library due to its simplicity in handling, commercial availability, and ease of production of custom DNA-tagged collections of 10^6-10^{11} substrates.[495] Chemical transformations can be performed on phage display libraries to introduce chemical moieties impossible to obtain with the natural expression machinery of *E. coli*.[377,405–407,416,427,428,467,491–494,496,497] (III) We focused on the Wittig reaction because it can be performed on phage displayed libraries[497] and because the mechanism of this reaction in protic solvents for stabilized ylides is not well understood.[446,455,457] We hypothesized that different residues might either stabilize or destabilize the oxaphosphetane (OPA) transition state in this reaction (Figure 3-1A). (IV) We used libraries of two topologies, linear (SX_4) and cyclic (SXC_3C)

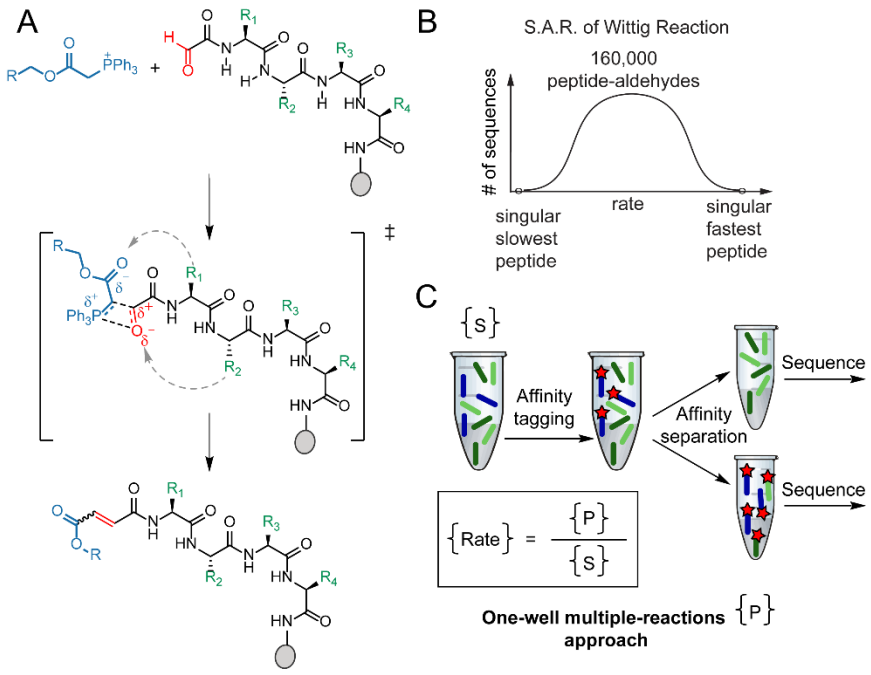


Figure 3-1. (A) Illustration of distribution of rates on focused library population. B) Rates found for tested sequences from SX_4 1% and 0.1% sets. C) Diagram for one-well-multiple-reactions approach for SAR using phage display.

to test this hypothesis. Both libraries can be entirely visualized by Illumina sequencing[411] and thus, enrichments and depletions of substrates in these libraries can be easily quantified. A critical aspect of data used for SAR studies is the breath of physical properties of interrogated substrates. Conveniently, even a natural repertoire of amino-acid side chains provides valuable chemical diversity that could influence the stability of the transition states or intermediates: examples are hydroxyls (OH) with pKa of 7-9 and 13-15; R-NH_3^+ with pKa 7-12; carboxylates, four types of aromatic rings with different "π-basicity" and four distinct H-N bonds (side chain of Trp, Asn and backbone).

Here, we employed Wittig functionalization of mixtures of ~10^5 genetically-encoded (GE) N-terminal glyoxaldehyde peptides displayed on phage under kinetically controlled conditions to yield SAR of Wittig reaction. Pulsing an aldehyde peptide library with a biotin labeled, ester-stabilized ylide (YEB) followed by affinity pull-down and deep-sequencing uncovered substrate-dependent reactivity that spanned a factor of 50-fold from most unreactive to most reactive. Conversion of the i^{th} substrate in the mix was assessed as $DC_i=(A_i-B_i)/N_i$, where A_i, B_i and N_i are

normalized frequencies of the i^{th} peptide from deep sequencing measurements for test (A_i), control (B_i) and input (N_i) sets. DC determined by Illumina sequencing correlates with the conversion of the purified peptides by HPLC. GE-SAR uncovered a combined role of backbone amides and side-chain functionalities in the Wittig reaction *in water* between stabilized ylides and glyoxaldehydes. Two Trp residues adjacent to the glyoxaldehyde moiety accelerate the reactivity, whereas removal of two backbone hydrogen bond donors adjacent to the aldehyde, slowed the reaction down.

3.2 Results and Discussion

3.2.1 0.1-1% Biotinylation of Peptide Libraries

The libraries of peptide aldehydes were generated by $NaIO_4$ oxidation as described in our previous reports,[427] and exposed to biotin-tagged phosphorane ylide (YEB) under kinetically controlled conditions. Based on previously reported rates of reaction on phage,[497] 10 minutes reaction time and 0.04-0.4 mM [YEB] would allow 0.1-1% of the population to become biotinylated (Figure 3-2). Adding these pre-determined concentrations of YEB to 10^9 phage virions and quenching with acidic buffer at the specified time intervals, ensured reproducible control over conversion of populations of peptides. Based on the prior knowledge of the composition and diversity of the library, we hypothesized that population conversion of 1% should provide an effective range of 3-4 orders of magnitude to interrogate both 100x increase and 100x decrease in the rate of reaction in the libraries: in 10^9 clones of a library with diversity of 10^5 (each clone is present at copy number of ~10^4), a conversion of 1% tags 10^2 out of 10^4 clones with average reactivity. Up to a 100x increase ($10^4/10^4$ clones tagged) and 100x decrease ($<1/10^4$ clones tagged) could be potentially observed and quantified under these conditions.

3.2.2 Pull-Down Methodology for Isolation of Wittig Reacted Virions

The pull-down process selects not only the biotinylated population but also peptide sequences that have affinity for other components in the system such as protein streptavidin conjugated to agarose beads. A set of two controls (Figure 3-3A-B) assessed the magnitude of such non-specific binding to be less than 1% of the specific biotin-streptavidin interaction. Control C tested the intrinsic binding of the aldehyde-peptide population not biotinylated whereas control B tested for binding of biotinylated peptides to streptavidin with blocked biotin-binding site.

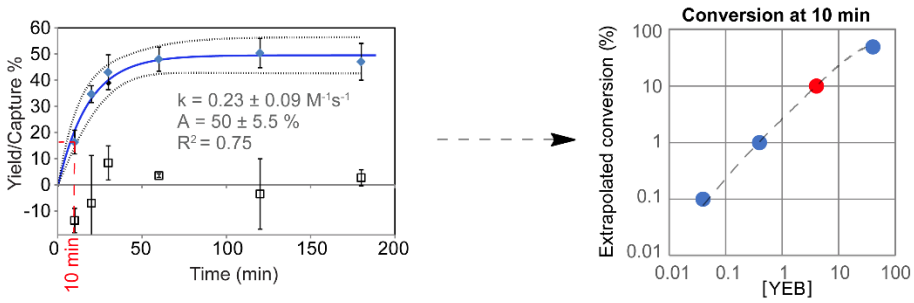


Figure 3-2. Extrapolation of conversion for phage display experiments.

To further test the specificity of the pull-down, we spiked the library with prospectively biotinylated, silently-encoded clones[414] and tracked the behavior of these sequences during pull-down, PCR and Illumina sequencing steps (Appendix B-1). Biotinylated clones were recovered in sets A and B and they decreased significantly in set C. Applying a normalization and sampling correction to raw Illumina counts, generated values denominated "Deep Conversions" or DC. Figures 3-3D and Appendix B-1 summarize the calculation of the absolute number of biotinylated particles from sequencing data accounting for factors like sequencing depth and amount of phage particles in each specific experiment.

3.2.3 Selection of "Fast", "Medium" and "Slow" Reacting Sequences

20x20 plots[411] allow a library-wide visualization of DC for every peptide sequence in the library. For the SX_4 1% selection, the plot clearly showed enhanced conversion of peptides with SWWXX motif and decreased conversion of peptides with SPPXX motif (Figure 3-4). According to this analysis, SWWXX-type sequences are predicted to react faster while SPPXX sequences are predicted to react slower than an average tetrapeptide-aldehyde.

Division by "N_i" was critical to normalize the data and to account for sequences that were present in a high copy number in naïve libraries but did not react fast in the Wittig reaction. For example, biotin-tagged SPPXX sequences were observed in high copy in captured population and they can be mistakenly interpreted as "most reactive" substrates (Appendix B-2). On the other hand, these sequences were also present in high copy number in naïve library (Appendix B-2) and

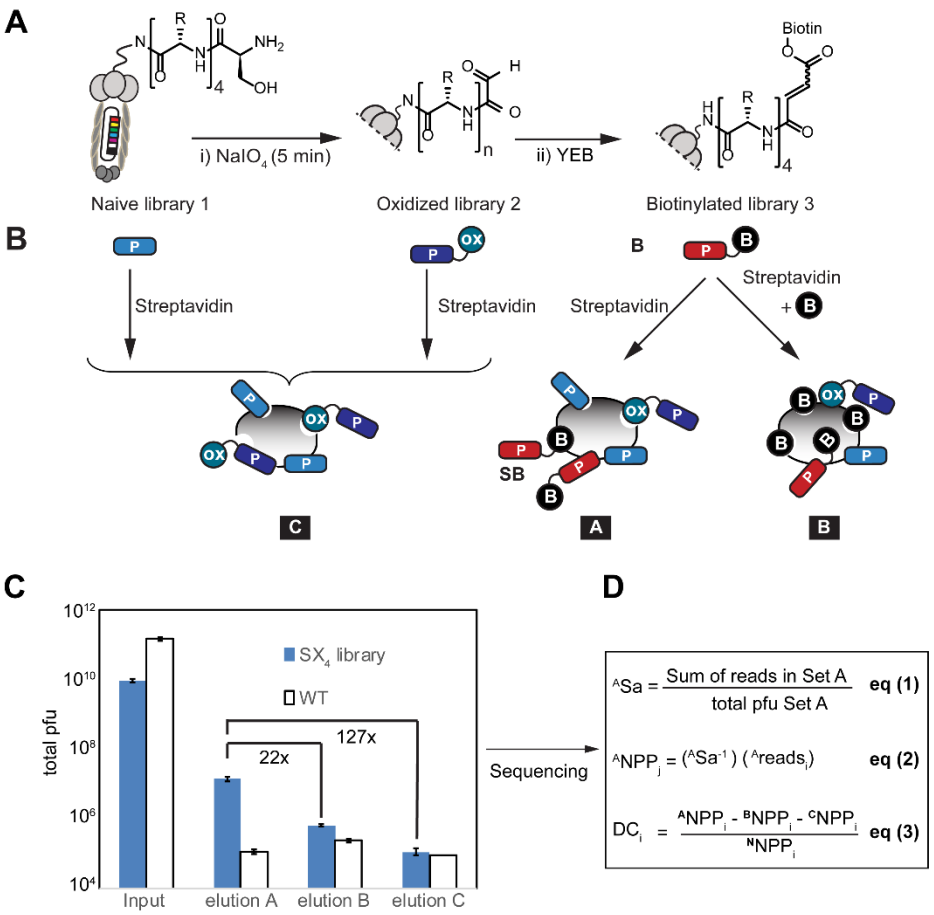


Figure 3-3. (A) Workflow of chemical modification on SX_4 and SXC_3C libraries. (B) We envisioned two classes of false positives: i) Members of the library with innate streptavidin/agarose affinity, ii) members of the library with streptavidin/agarose affinity when oxidized. (C) Pull-down tittering data showing higher capture of phage expressing peptide (blue bars) vs wild type phage (white bars) and higher capture in 1% biotinylated set (A) than in controls (B and C). (D) We used sampling factor SA to convert number of reads observed in deep sequencing to number of phage particles and used their ratio in biotinylated and original set to estimate the "Deep Conversion" (DC) for each sequence.

analysis after normalization predicted that SPPXX sequences in fact have the lowest conversion in the Wittig reaction (Figure 3-4).

3.2.4 Validation of Selected Peptides

To check the validity of observations predicted by Deep Conversion (DC), we selected a series of peptides predicted to be "fast" ($DC \geq 0.35$), "medium" ($0.15 \leq DC \leq 0.35$) and "slow" ($DC \leq 0.15$), synthesized them and validated them (Appendix B3-B6). We found an agreement

between the experimentally determined conversion measured by HPLC and the DC (Figure 3-5B) in a range of two orders of magnitude for both values. Previously,[497] we found that the average rate for all peptides present in the SX_7 library was ~0.23 $M^{-1}s^{-1}$. This rate constant was indistinguishable (within the error of measurement) of the rate constant for "average" sequences that did not contain Trp or Pro residues in the first two positions (entries 4 and 5 in Figure 3-5A).

The 20x20 plot (Figure 3-4A) predicted further decrease in conversion for SPPXX sequences. Indeed, synthesized aldehydes Ald-PPLA and Ald-PPPL sequences showed rates of 0.017 ± 0.005 and 0.014 ± 0.03 $M^{-1}s^{-1}$ respectively. These sequences were up to 13-16-fold slower than the average population. To test the effect of introduction of proline in a specific position (or concurrent removal of a backbone H-bond donor from that position) we systematically evaluated the reactivity of Ald-PPPA, Ald-PPAA, Ald-PAAA, Ald-APAA and Ald-AAAP sequences (Figure 3-6). Removal of either 1st or 2nd H-bond donor lead to modest 2-3 fold decrease in rate whereas concurrent removal of 1st and 2nd H-bond donors resulted in 10-fold decrease (~1.4 kcal). Subsequent removal of 3rd H-bond donor has little additional effect. To further support this hypothesis, we observed that the rates of Ald-PPXX and Ald-PPPP were similar (i.e. removal of amides from Ald-PPXX sequences had no observable effect on reactivity). These results suggested that only concurrent removal of two penultimate hydrogen bond donors from backbone amides leads to significant decrease of rate. Similar energetic additive effects were observed for the Ald-WWXX family. For example, Ald-QWLH, Ald-WIVR, Ald-HWFP, Ald-LWYR and Ald-WLPR (entries 3, 6, 13, 16 17 in Figure 3-5A), show no statistically significant increase from average reactivity, whereas sequences Ald-WWPQ and Ald-WWGL with tryptophan residues in both positions, exhibited up to 5-fold increase from average rate and more than 50-fold increase from lowest rate in the population.

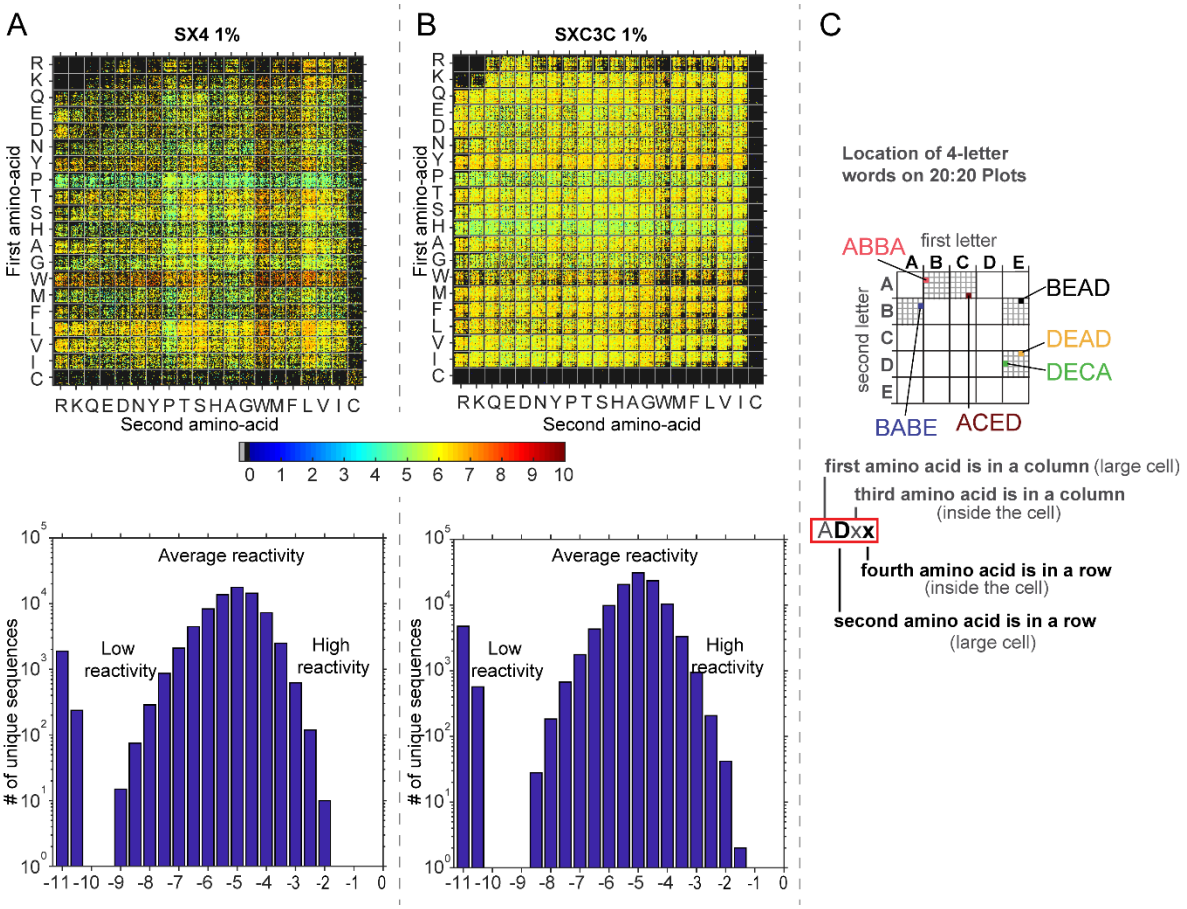


Figure 3-4. 20x20 plots displaying library-wide DC values for SX_4 and SXC_3C selections. (C) How to read a 20x20 plot (copied with permission of Ratmir Derda).[411]

The GE-SAR uncovers some decreases in conversion, which were not linked directly to Wittig reaction. For example, we uncovered that the predicted 10X decrease in conversion of Ald-PXXX sequences to Bio-PXXX is due to a combination of reactivity of Ald-PXXX and unique side reaction that consumes Ald-PXXX aldehydes *via* proline-assisted 6-exo-trig attack of amide, leading to cyclization[498] (Appendix B-7 and B-8). Measuring the rates of both cyclization and Wittig reaction independently, found them to be in a range of 0.12-0.17 $M^{-1}s^{-1}$; for a given sequence, the rates of cyclization and Wittig reaction were similar. Therefore, while SPXXX sequences reacted only 2-fold slower than the average population, the conversion to a biotinylated substrate deviated by 5-fold from an average.

We note that aforementioned Ald-PPXX sequences did not exhibit any side cyclization reaction due to lack of second-residue secondary amide and thus, the observed 10x decrease in conversion od Ald-PPXX substrates is exclusively due to low reactivity in the Wittig reaction.

3.2.5 SX_4 0.1% Selection Analysis

Repeating the same reaction using the same time (10 minutes) with $1/10^{th}$ of the substrate (0.04 mM [YEB] instead of 0.4 mM [YEB]) converted 0.1% of SX_4 library to biotinylated substrates. Capture and analysis of this population yielded observations similar to the one for 1% selection (Appendix B-8), but identification of slow-reacting sequences in these conditions was less efficient: depletion of the SPPXX sequences was less pronounced in this selection. The analysis emphasized new features in sequences with higher than average conversions, like modest enrichment of Tyr residues in the second position and potential rate enhancement for Ald-WYXX and Ald-YWXX aldehydes. Depletion of proline residues in the first position was observed only for SPXXX but this observation is expected because the loss of aldehyde moiety *via* side cyclization is independent of substrate concentration.

3.2.6 SXC_3C Selections

Applying the optimal, 1% conversion (10 minutes reaction with 0.4 mM YEB) to SXCXXXC library, we found that the effect of amino-acid residues on conversion was significantly attenuated when compared to the SX_4 library (Figure 3-4B). These results agree with the previous finding that emphasize the additive effect of the two adjacent residues in position 1 and 2 downstream the aldehyde moiety. This synergistic effect between position 1 and 2 is not possible in SXCXXXC library because the second position is a constant cysteine. As in SX_4, presence of proline in the first position but not the third position lead to decrease of DC due to intramolecular cyclization. A new observation was antagonistic effect of His residue in the first position with the same predicted decrease in conversion as that caused by introduction of single penultimate Pro residue. Tryptophan and tyrosine in the first (but not the third) position lead to a modest increase in conversion. Synthesis and testing of several sequences suggested that average rates of these peptides from SXCXXXC set is similar to that average peptides from SX_4 series, and placement of Y in the first but not the third position led to modest but detectable increase in rate (Appendix B-9 and B-10).

A

Entry	Sequence	(DC)	k ($M^{-1}s^{-1}$)	HPLC Conversion**
1	SWWPQ	0.70 ± 0.04	0.56 ± 0.01	0.74 ± 0.01
2	SWWGL	0.83 ± 0.05	0.5 ± 0.2	0.7 ± 0.2
3	SQWLH	1.03 ± 0.07	0.29 ± 0.06	0.5 ± 0.1
4	SAAAP	0.25 ± 0.03	0.29 ± 0.06	0.5 ± 0.1
5	SALRV	0.24 ± 0.06	0.22 ± 0.04	0.41 ± 0.08
6	SWIVR	0.36 ± 0.09	0.17 ± 0.04	0.34 ± 0.08
7	SPQPL	0.11 ± 0.02	0.15 ± 0.03	0.30 ± 0.06*
8	SPYPA	0.05 ± 0.02	0.13 ± 0.03	0.27 ± 0.06*
9	SPPLA	0.13 ± 0.01	0.017 ± 0.005	0.04 ± 0.01
10	SPPPL	0.104 ± 0.002	0.014 ± 0.03	0.033 ± 0.007
11	SPPPP	0.100 ± 0.001	0.020 ± 0.003	0.05 ± 0.03
12	SPRLP	0.121 ± 0.008	0.17 ± 0.07	0.3 ± 0.1*
13	SHWFP	0.250 ±0.005	0.25 ± 0.04	0.45 ± 0.07
14	SPPAA	0.118 ± 0.001	0.024 ± 0.002	0.056 ± 0.005
15	SPPPA	0.116 ± 0.003	0.020 ± 0.002	0.047 ± 0.005
16	SLWYR	0.55 ± 0.02	0.30 ± 0.02	0.51 ± 0.03
17	SWLPR	0.632 ± 0.007	0.21 ± 0.01	0.40± 0.02
18	SAPAA	0.18 ± 0.02	0.10 ± 0.03	0.21± 0.06
19	SPAAA	0.073 ± 0.006	0.12 ± 0.03	0.25 ± 0.06*

B

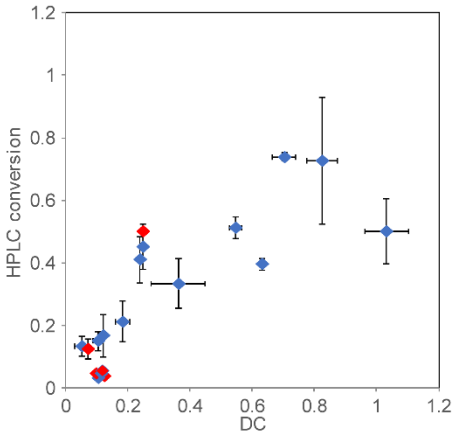


Figure 3-5. Comparison between "DC" and HPLC conversion in peptides. A) DC, rate (k) and HPLC conversion** for synthesized sequences. B) DC vs "HPLC conversion". **HPLC conversion = 1-EXP(-k x 0.004 x 600)] or *[1-EXP(-k1 x 0.004 x 600)]/(2 x (k2/k1))

3.2.7 Self-Catalysis and Cross-Catalysis Tests

To test whether peptides are self-catalytic or cross-catalytic towards the Wittig reaction, mixtures of o-VEKY and the sequence of interest at equal concentrations (0.8 mM) were reacted with excess YEB (4 mM). The obtained rates in the mixture of substrates were similar to the ones found in individual experiments with one exception (Appendix B-11 and B-12): in a mixture of Ald-VEKY and Ald-WWRR, the "slow" Ald-VEKY peptide exhibited a 2-fold increase when compared to kinetics of pure Ald-VEKY. As Ald-WWRR peptide aldehyde was prone to aggregation under reaction conditions, the possible explanation to this observation is interaction of Ald-WWRR and Ald-VEKY, leading to potential stabilization or destabilization of TS of one peptide by another in the aggregate.

3.2.8 Hydrazone Ligation Experiments

To determine whether the increase and decrease of rate of the Wittig reaction was due to amino-acid influences on the OPA TSs or due to inherent reactivity of substrates towards attacks by nucleophiles, we measured the rate of reaction of SPPAA and SWWRR sequences with 2,4-dinitrophenylhydrazine. To our surprise, the reactivity for hydrazine ligation was reversed (Appendix B-13 and B-14) and Ald-PPAA reacted at least ~2x faster than Ald-WWRR both at pH

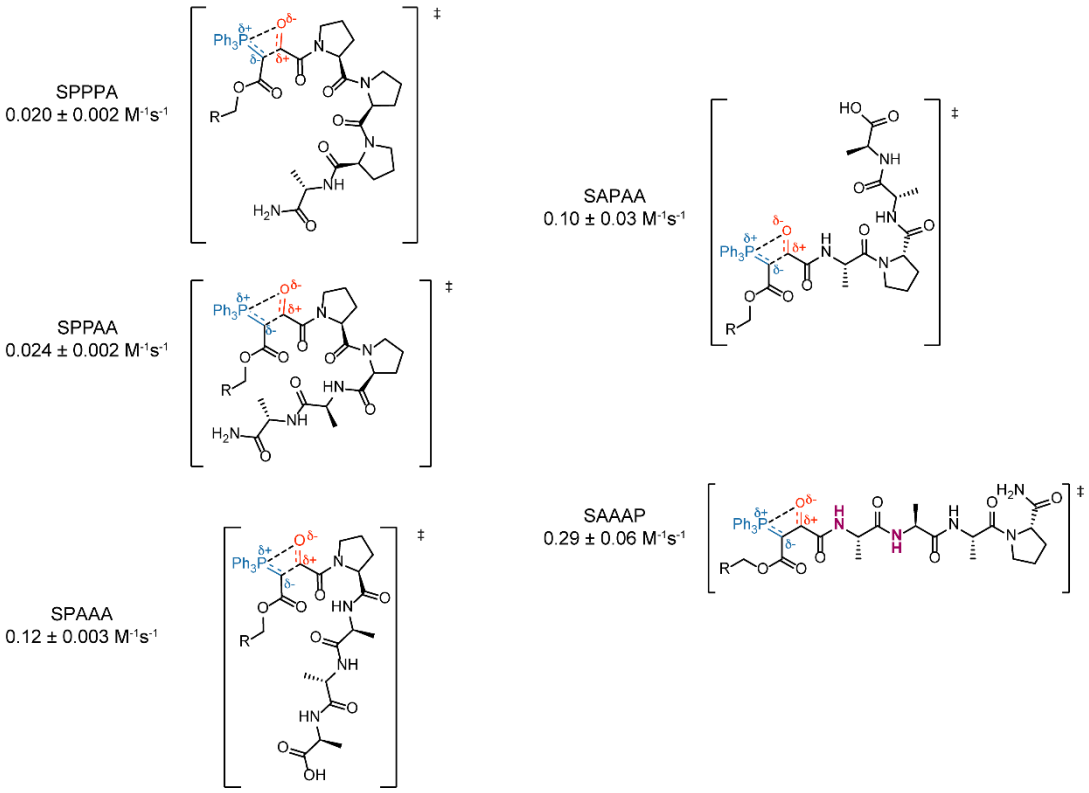


Figure 3-6. Alanine scan to determine position and quantity of hydrogen-bond donors for stabilization of OPA transition state.

0 and pH 5 (Appendix B-15 and B-16). These results demonstrate that the increase in reactivity is not due to the inherent increase of the electrophilicity of the aldehyde substrates but to reaction specific effects.

3.2.9 E/Z Selectivity Experiments

Although the selection pipeline was not designed to select sequences that react stereo selectively, we noticed curious changes in E/Z selectivity in the Wittig reaction of "fast" (Ald-WWRR 50/50 E/Z), "medium" (Ald-HWFP 1/9 E/Z) and "slow" (Ald-PPAA 1/9 E/Z) reacting sequences (Appendix B-17-B-19). These results suggest side chains of amino-acids and backbone amides might exhibit different influence on E-OPA vs Z-OPA transition state. Future development of selection strategies that can separate *E* vs *Z* products from a mixture could illuminate the origin of these differences.

3.3 Conclusions

Genetically encoded, complete, fully sequenced library of ~10^6 members gave rise to a rich dataset guiding SAR studies for the Wittig reaction in water between stabilized ylide (YEB) and glyoxaldehydes and highlighted the cooperative effect of the two residues adjacent to the aldehyde on the rate of the reaction. The 50-fold range for the rates of the Wittig reaction suggest 3 kcal/mol cumulative effect from backbone H-bond donors and side chains, with backbone amides providing the strongest effects. Loss of reactivity upon introduction of PP motif highlights the role of backbone amides on stabilization of TS of Wittig reaction with an average peptide aldehyde. Further gain of reactivity upon introduction of WW residues (but not H, F or Y), suggest a unique stabilizing role of indole ring (either via aromatic interaction or H-bond with indole nitrogen) on TS of Wittig reaction in water.

In principle, One-Bead-One-Compound libraries combined with calibrated mass spectrometry analysis,[499] could permit analogous analysis of positive and negative populations and yield SAR analysis. Still, the advantage of sequencing is "digital" accounting of single copies of DNA molecules, which is not currently possible for mass-spectrometry techniques.

We anticipate that coupling of deep sequencing methodology to investigation of chemical reactivity and SAR studies with only minor changes can be applied to investigate other reactions already known to be compatible with M13 virion (thiol S_N2,[397,405,406,421,496] thiol S_NAr,[407] oxime ligation,[427,500] copper catalyzed azide-alkyne cycloaddition,[501] copper-free azide alkyne cycloaddition,[502] Diels Alder cycloaddition,[497]). It can also illuminate SAR of other reactions that have potential to be applied on phage such as multicomponent reactions compatible with aqueous conditions[503] or chemistry that has been already explored on proteins, unprotected peptides and other display platforms.[207,418,429,431,494,504]

3.4 Experimental Procedures

3.4.1 Materials, Methods and General Information

Acetate buffer 200 mM was prepared by mixing sodium acetate (16.4 g), NaCl (80 g) and adding 800 mL of milliQ water. The pH was adjusted to 5.0 with glacial acetic acid and the volume was completed to 1L with milliQ water. MOPS buffer for kinetic measurements was prepared as

a solution of 200 mM MOPS and adjusted to pH 6.5 with NaOH (1M). All solutions used for phage work were prepared with milliQ water. Peptides were synthesized using the Prelude X peptide synthesizer (Gyros Protein Technologies) using 200 mg Rink Amide AM resin. Ylide Ester Biotin (YEB) was synthesized as already reported.[497] HRMS (ESI) spectra were recorded on Agilent 6220 oaTOF mass spectrometer using either positive or negative ionization mode. Characterization of reaction crude was performed with UPLC-MS using a C18 column (Phenomenex Kinetex 1.7 µm EVO C18, 2.1×50 mm) running with a gradient of water/acetonitrile with 0.1% formic acid from 98/2 at 0 min to 40/60 at 5 min under a flow rate of 0.5 mL/min. HPLC kinetics were performed in an Agilent 1100 Series using a Thermo Scientific C18 column (Hypersil GOLD, 3 µm, 50×2.1mm). The SX_4 library was bulk-amplified at home from a stock library synthesized by Katrina Tjhung. The SXC_3C library was synthesized by Nicholas Bennet as well as the silently encoded SWYD neon green clones for internal control. ^1H and ^{13}C NMR spectra were acquired on a 600 MHz four channel Agilent VNMRS spectrometer, equipped with a z gradient HCN probe and using VNMRJ 4.2A as the acquisition software. Suppression of the H_2O signal was performed using either presaturation or excitation sculpting.[466]

3.4.2 Phage Functionalizations

3.4.2.1 SXC_3C and SX_4 1% Biotinylation and Purification

10^{12} pfu of library were taken to a final volume of 0.10 mL with MOPS 0.20 M. To the solution, 1.0 µL of freshly prepared $NaIO_4$ (6.0 mM in water) was added and the mixture was incubated in ice for 8 minutes. Then, 1.0 µL of methionine (50 mM in water) was added and the mixture was incubated for 20 minutes at room temperature. 0.10 mL of YEB (0.80 mM in MOPS 0.20 M) were added and the reaction was incubated for 10 minutes at room temperature. After this, 0.80 mL of acetate buffer (0.20 M) and 0.20 mL of PEG NaCl were added and the mixture was kept at 4 °C for at least 12 hours. After this, the sample was centrifuged at 21100 x g for 30 minutes and the supernatant was discarded. The sample was further centrifuged for 5 minutes and remaining supernatant was removed. The phage pellet was re-suspended in 0.20 mL of acetate buffer (10 mM) and incubated at room temperature for 15 minutes to allow phage to completely suspend. After this, remaining solid debris was removed by centrifugation at 21100 x g for 5

minutes. The supernatant was transferred to a clean epi-tube and maintained at 4 °C for use in pull-down screenings.

3.4.2.2 SX_4 0.1% Biotinylation and Purification

10^{12} pfu of library were taken to a final volume of 0.10 mL with MOPS 0.20 M. To the solution, 1.0 µL of freshly prepared $NaIO_4$ (6.0 mM in water) was added and the mixture was incubated in ice for 8 minutes. Then, 1.0 µL of methionine (50 mM in water) was added and the mixture was incubated for 20 minutes. 0.10 mL of YEB (0.80 mM in MOPS 0.20 M) were added and the reaction was incubated for 10 minutes at room temperature. After this, PEG purification was performed as indicated in Section 3.4.2.3 and the phage was stored at 4 °C, ready to be used in screenings.

3.4.2.3 SXC_3C and SX_4 Oxidations and Purification

10^{12} pfu of library were taken to a final volume of 0.10 mL with MOPS 0.20 M. To the solution, 1.0 µL of freshly prepared $NaIO_4$ (6.0 mM in water) was added and the mixture was incubated in ice for 8 minutes. Then, 1.0 µL of methionine (50 mM in water) was added and the mixture was incubated for 20 minutes. After this, PEG purification was performed as indicated in section 3.4.2.3 and the phage was stored at 4 °C, ready to be used in screenings.

3.4.3 Phage Pull-Down Experiments

Set A: ~10^9 pfu of biotinylated, purified library were mixed with ~10^5 pfu of internal control mixture neon green phage (in the case of 1% selections) and the volume was taken to 1 mL with BSA 2.0 % in acetate buffer in a 1.5 mL microcentrifuge tube.

Set B: ~10^9 pfu of biotinylated, purified library were mixed with ~10^5 pfu of internal control mixture (in the case of 1% selections) and the volume was taken to 1 mL with BSA 2.0 % + 0.90 mM biotin in acetate buffer in a 1.5 mL microcentrifuge tube; this would make

Set C: ~10^9 pfu of oxidized, purified library were mixed with ~10^5 pfu of internal control mixture neon green phage (in the case of 1.0 % selections) and the volume was taken to 1.0 mL with BSA 2.0 % in acetate buffer in a 1.5 mL microcentrifuge tube.

All mixtures were prepared in triplicates. 0.10 mL of streptavidin magnetic beads per each sample (9 in total) were washed three times with 1.0 mL of acetate buffer (10 mM) and suspended in 1.0 mL of BSA 2.0 % in acetate buffer.

All mixtures and bead suspensions were incubated for 30 minutes at room temperature in a Labquake Tube Rotator. After this, 10 µL aliquots of each phage mixture (9 in total) were taken for tittering and PCR of inputs. The beads suspensions (9 in total) were placed in a magnet and the BSA supernatants were discarded. Each phage mixture (0.99 mL each one) was added to one corresponding tube of beads and incubated at room temperature in a Labquake Tube Rotator for 1 hour. The unbound phage was washed from the beads with Tween 0.10 % in acetate buffer (10 washes of 1.0 mL each one) by using a Kingfisher magnetic bead washer (ThermoFisher Scientific).

The washed beads were suspended in acetate buffer (10 mM), placed in a magnet and the supernatant was discarded (this step was necessary to remove any remaining Tween that might interfere with PCR). The rinsed beads were suspended in 20 µL of NaOH solution (10 mM) and incubated at room temperature for one hour in a Labquake Tube Rotator. After that, suspensions were placed in a magnet and the supernatant was transferred to a 0.60 mL microcentrifuge tube containing 20 µL of 1X Phusion® high fidelity (HF) buffer to give a final volume of 40 µL of phage elution. 2.0 µL are taken for tittering of this elution and the rest is stored at 4 °C for use in direct PCR.

3.4.4 Direct PCR of Phage

3.4.4.1 Direct PCR of NaOH Elution

Each elution consists of 40 µL of phage as specified in Section 3.4.3.1. This elution was split in two aliquots of 20 µL and each one was mixed with 1.0 µL of dNTPs (10 mM each), 1.0 µL of Phusion® High Fidelity Polymerase, 2.5 µL of reverse primer (10 mM), 2.5 µL of forward primer (10 mM) and 23 µL of RNAase free water. PCR was performed like this: first, the sample was held at 95 °C for 30 seconds and then 25 cycles of 95 °C (10 seconds), 60.5 °C (15 seconds) and 72 °C (30 seconds). Finally, the sample was maintained at 72 °C for 5 minutes and held at 4 °C for storage. The two PCR products for each elution were mixed before gel electrophoresis characterization.

3.4.4.2 Direct PCR of Inputs

5.0 µL of the input aliquots (9 samples in total, from Section 3.4.3.1) were mixed with 10 µL of Phusion® High Fidelity buffer, 1.0 µL of dNTPs (10 mM each), 1.0 µL of Phusion® High Fidelity Polymerase, 2.5 µL of reverse primer (10 mM), 2.5 µL of forward primer (10 mM) and 28 µL of RNAase free water. The temperature cycles used to PCR these samples were the same as the ones described in Section 3.4.4.1.

3.4.5 Gel Electrophoresis Characterization of PCR Products

10 µL of the PCR products obtained in Sections 3.4.4.1 and 3.4.4.2 were mixed with 2.0 µL of gel dye and loaded into a 2.0 % agarose gel in Tris Acetate EDTA buffer (TAE) containing SYBR-safe dye. 44 ng of DNA low molecular weight ladder (New England Biolabs) were also loaded in one line for quantification purposes. The gel was developed in TAE at 500 mV and 100 A for 40 minutes. Pictures were taken with an ImageQuant RT ECL machine using 520 nm wavelength filter. After quantification, all DNA samples (18 in total for each screening) were mixed together by adding 20 ng of each one. Purification with E-Gel® SizeSelect™ 2.0 % agarose gel (Invitrogen, #G6610-02) provided the sample that was sequenced using the Illumina NextSeq platform at MBSU, University of Alberta.

3.4.6 Illumina Sequencing Data Analysis

Illumina data was processed following already reported strategies.[410,414,505] Briefly, raw FASTQ data was processed using Matlab Scripts developed by this lab. These scripts identified the barcodes and constant flanking regions and extracted the reads of correct length for each type of library. Each sequence and copy numbers at which they appear were arranged in a NxM matrix, where N is the total number of unique sequences and M is the number of experiments and replicates. After this, the data arranged in the matrix was used to find sequences significantly enriched in the positive sets. This was done with the Differential Enrichment matlab script already reported by this lab[410] and attached below (Appendix B-21). Basically, average normalized frequency of appearance was calculated for each sequence in each set (positive, controls and inputs). Then, the average frequency in the positive set was divided by the average frequency in control sets to obtain a final ratio. A p-value between replicates of positive and control sets was calculated using a two-sided unequal variance t-test. Sequences that were considered “hits” had p-

value < 0.05 and R ≥ 3. For further information on all the math behind these calculations go to previous report.[414] The 20:20 plots were obtained by using techniques already reported by this lab[411,414] and the script is attached below (Appendix B-22). Logo plots were obtained by transforming the txt data table after Differential Enrichment analysis back to fastq and uploading it into the https://weblogo.berkeley.edu/logo.cgi website.

3.4.7 Oxidation of N-terminal Serine Peptides

Unless specified differently, ~20 mg of peptide were diluted into 1.0 mL of 1X PBS and required amount of acetonitrile (when needed). The peptide was incubated in ice for 5 minutes and solubility was checked, in case of precipitation, more acetonitrile was added. 2.5 equivalents of $NaIO_4$ were added as a saturated solution in water. The mixture was incubated in ice for 20 minutes and injected into C18 or HILIC column for quenching and purification. The acetonitrile was evaporated *via* speedvac and the oxidized peptide was dried using lyophilization. When the peptide contained a methionine, a 2.0 mM solution of peptide (in 1X PBS) was kept in ice while 1.5 equivalents of NaOH were dissolved in water. The NaOH was added and after 10 seconds, the reaction was quenched with 12 equivalents of glutathione. The mixture was incubated in ice for ten minutes and then injected on C18 or HILIC column for purification. Acetonitrile was removed with speedvac and the oxidized peptides were lyophilized.

3.4.8 Kinetics of Selected Peptides

Solutions of oxidized peptides (7.70 mM) were prepared in water. A stock solution of YEB (38.7 mM) was prepared in water. An aliquot of peptide solution (10.4 µL, final concentration 800 µM) was dissolved in 79.3 µL of MOPS buffer (200 mM, pH 6.50). YEB solution (10.3 µL, final concentration 4.00 mM) was added. An aliquot of the reaction (10.0 µL) was taken out and quenched with HCl (1.0 µL, 1.0 M). Quenching was performed after 3 min, 10 min, 30 min, 60 min, 120 min and 180 reaction time. An aliquot (3.0 µL) of each quenched solution was injected into UHPLC-MS to obtain traces. Area percentages were extracted and the pseudo first order rate constant and kinetic traces were obtained by using the matlab script attached in Appendix A-16. It is important to mention that several oxidized peptides required certain amount of co-solvent (acetonitrile) to improve solubility. At all measurements, the final percentage of acetonitrile was ≤ 2%, since rate and stereoselectivity of the Wittig reaction is highly dependent on the type of

solvent.[457] Also, the volumes of aliquots were kept the same at all times to ensure all peptides were measured at the same pH.

3.4.9 Self-Catalysis and Cross-Catalysis Tests

Solutions of oxidized peptides (7.70 mM) were prepared in water. A stock solution of YEB (38.7 mM) was prepared in water. An aliquot of peptide solution (10.4 μL, final concentration 800 μM) was dissolved in 69.3 μL of MOPS buffer (200 mM, pH 6.50). 10.4 μL of the second peptide solution were added. YEB solution (10.3 μL, final concentration 4.00 mM) was added. An aliquot of the reaction (10.0 μL) was taken out and quenched with HCl (1.0 μL, 1.0 M). Quenching was performed after 3 min, 10 min, 30 min, 60 min, 120 min and 180 reaction time. An aliquot (3.0 μL) of each quenched solution was injected into UHPLC-MS to obtain traces. Area percentages were extracted and the pseudo first order rate constant and kinetic traces were obtained by using the matlab script attached in Appendix A-16. It is important to mention that several oxidized peptides required certain amount of co-solvent (acetonitrile) to improve solubility. At all measurements, the final percentage of acetonitrile was ≤ 2%, since rate and stereoselectivity of the Wittig reaction is highly dependent on the type of solvent.[457] Also, the volumes of aliquots were kept the same at all times to ensure all peptides were measured at the same pH.

3.4.10 Hydrazone Ligation Experiments

0.10 mg of 2,4-Dinitrophenylhydrazone were dissolved in 1.1mL of concentrated H_2SO_4. This solution was slowly added to a mixture of 5.7 mL EtOH (95%)/1.6 mL H_2O to generate a 61 mM solution of hydrazine. 10.4 μL of aldehyde solution (7.70 mM) were mixed with 73.5 μL of H_2O, 10.0 μL of EtOH were added and finally, 6.50 μL of hydrazine solution were added. An aliquot of the reaction (2.0 μL) was taken out and quenched by dilution with 198 μL of water. Quenching was performed at the times specified in Appendix B-13 and Apppendix B-14. An aliquot (25.0 μL) of each quenched solution was injected into UHPLC-MS to obtain traces. Area percentages were extracted and the pseudo first order rate constant and kinetic traces were obtained by using the matlab script attached in Appendix A-16.

3.4.11 In Situ E/Z Selectivity Determination

Aliquots (usually 20-50 μL, 1.00 eq, final concentration 0.8 mM) of stock solutions of aldehydes (usually 7.7 mM) were mixed with PB 200 mM at pH 6.5 in an NMR tube. 55 μL of D_2O were added. A solution of YEB (51.7 μL, 38.7 mM, 5.00 eq, final concentration 4.00 mM) was added. The final volume was 550 μL (D_2O 10%). Solvent suppresion was performed and ^1H NMR spectra were collected at the specified time intervals.

Chapter 4: Selection-Based Discovery of Privileged Peptide Sequences as Substrates for Macrocyclization

4.1 Introduction

There is an increasing interest in developing macrocycle-based drugs due to the possibility of generating drug candidates with molecular area significantly higher than that of small molecules, yet, with pharmacokinetics and cell permeability properties similar to those of small molecules. Among 3254 approved drugs (DrugBank), there are 68 macrocycle-based drugs already in the market.[506] The increased molecular area of macrocycles when compared to that of small molecules is a critical property that give rise to potent and selective non-covalent interaction with targets that possess an extended binding surface (i.e., interfaces of protein-protein interactions).[22] It is estimated that 30%-40% of targets in human genome are not druggable with small molecules, but significant fraction of these targets might be possible to block with large area macrocyclic ligands. While the strategies for diversification of small drug-like molecules are growing explosively, one of the main limitations for development of macrocyclic drug-like molecules is lack of equally diverse and flexible synthetic strategies that give rise to large number of diverse macrocycles (i.e., libraries). Such macrocyclic compound "libraries" serve as starting point for high-throughput screening or genetically-encoded screening of macrocycles with desired biological activity. The purity and identity of the cyclized species is one of the key requirements for a successful screen.[23]

Peptide derived macrocycles are a high-value subset of cyclic scaffolds. Useful attributes of peptide macrocycles are: i) availability of building blocks, ii) trivial, robust steps for the assembly of the backbone prior to macrocyclization, iii) natural origin and low toxicity of building blocks, iv) availability of genetically-encoded technologies such as phage, mRNA and SICLOPPS that allow generation and screening of libraries with diversities of 10^8-10^{13} compounds.[507] Many of the ligands emanating from such screens can serve as valuable starting points for drug-discovery and "hit-to-lead" optimization. To aid such optimization, there is a rich 50^+-year history of converting natural peptide sequences to drug-like molecules (e.g., Angiotensin Converting Enzyme (ACE)-inhibitor). Although linear peptides in some cases can be used as precursors for drug design (e.g., GLP-1), cyclization is as a powerful tool that maintains high molecular area and

molecular footprint but minimizes unwanted conformational flexibility. Cyclization also improves pharmacokinetic properties of linear peptides by increasing proteolytic stability.[15,16]

Intramolecular macrocyclization of functionality-rich molecules is one of the most challenging problems in organic chemistry. While rules are defined for stereoelectronic preferences in reactions that form small-to-medium ring sizes (Baldwin rules and alike), there are no general rules for prospective identification of successful or unsuccessful reaction manifolds that close large cycles. Drawing inspiration from the concept of "attack trajectories" that led to rules for closing of small macrocycles, one can propose that similar notion of "probable attack trajectories" applies in closure of larger cycles starting from a subset of conformers of linear precursors that favor intramolecular reaction between two groups in the linear precursor. Unfortunately, conformational preferences of the linear precursor and intramolecular non-covalent interactions that stabilize the attack conformation could be very difficult to predict. As complimentary approach to rational prediction of cyclization preference, we aimed to devise a process that can start from a large collection of linear molecules and select linear precursors predisposed to undergo intramolecular cyclization. The analysis of results of such selection in turn, will provide insight into the structural characteristics of peptides that promote rapid and efficient macrocyclization, starting from a broad pool of unbiased sequences. In the methodology described below, the only biasing feature in the library is presence of mandatory N-terminal serine. It serves as anchor point for installation of the functionality that contains a tagged leaving group and an electrophile that engages in a cyclization process. As the goal of this process is discovery of new cyclization reactions, we do not pre-install any specific nucleophiles into any specific location on the peptide chain. Rather, we permit presence of any amino-acid at any location with equal probability. This is in contrast to most reported techniques that install cysteine residues in predefined locations to allow for attachment of handles that cross-link Cys residues.[359,377,396,406]

In a report by Roberts and coworkers, promiscuous amidation strategy in mRNA-displayed peptides that contain activated esters generated cyclic peptides *via* attack of ester by amino groups present at the N-terminus or at the side chain lysine. As optimization or selection of cyclization conditions in this setup is not feasible, the library is likely to contain both linear and cyclic peptides.[236] We envisioned that a subtle change in this strategy that introduces affinity purification tag into the leaving group can permit not only separation of cyclic and linear unreactive species

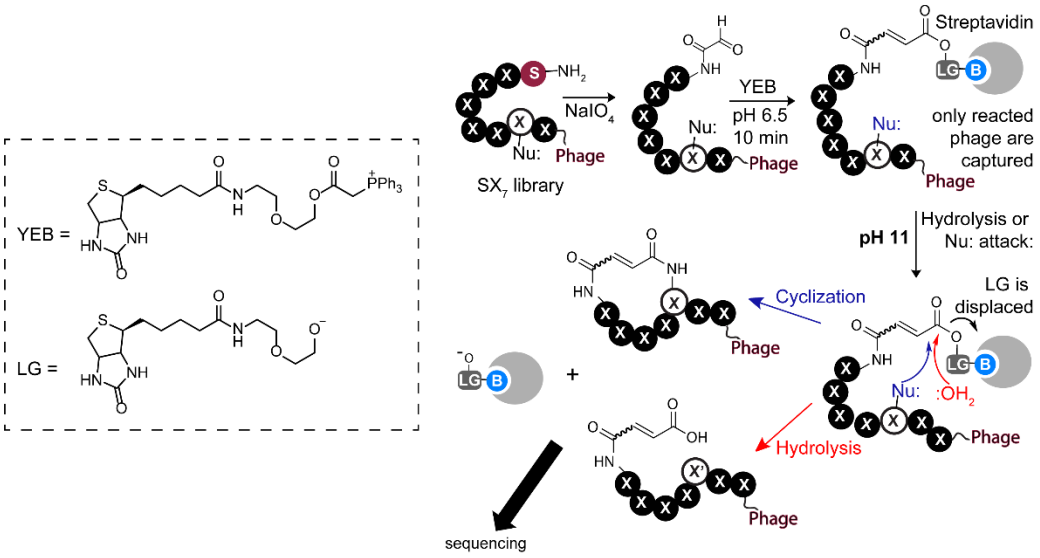


Figure 4-1. Selection of privileged substrates for peptide cyclization

but also selecting rare cyclized members of the library (Figure 4-1). Specifically, we installed the α,β-unsaturated ester functionality on phage libraries using Wittig reaction and hypothesized that it might be possible to screen for sequences prone to undergo nucleophilic attack by functional groups present in peptide sequences using a two-step capture and release process. We hypothesize that unbiased selection-based approach will provide an important starting point that can complement rational design of cyclization pathways and provide new strategies for efficient generation of peptide macrocycles.

4.2 Preliminary Evidence

Feasibility of intramolecular cyclization came from observation of a minor byproduct in the stability tests of the Wittig adduct at basic pH in our published report (Chapter 2),[497] The main

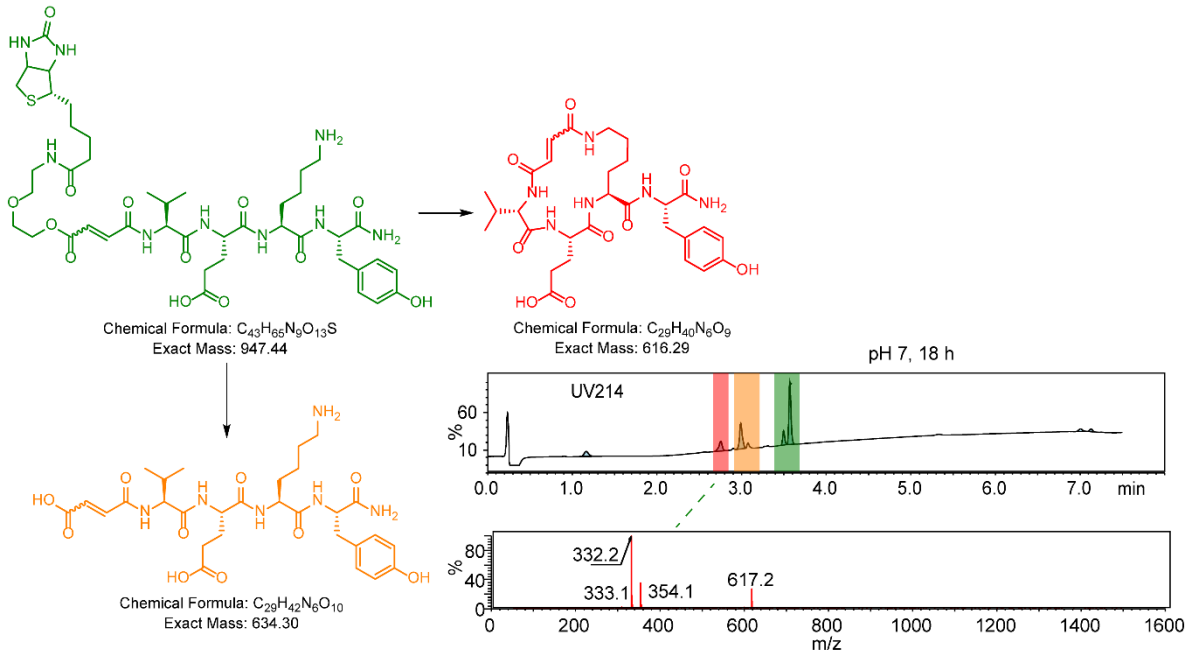


Figure 4-2. Proposed identity of hydrolysis-H_2O adduct for Wittig product of Ald-VEKY

product of hydrolysis lost the leaving group and gained water, meanwhile the biproduct lost leaving group and water concurrently (Figure 4-2 and Appendix A-10), indicating that the attack on acyl might have originated from a nucleophile in the peptide itself. We did not determine the structure of this byproduct explicitly, but a feasible explanation to the observed mass=M^+-H_2O-LG was formation of cyclic amide *via* nucleophilic attack of Lys side on conjugated alkyl ester (Figure 4-2). More intriguing observation that did not fit the direct amidation proposal was formation of the byproduct with the same mass difference (hydrolysis product-H_2O) while performing Wittig reaction between YEB and Ald-WWRR, an aldehyde that does not contain a Lys residue (Figure 4-3). Unfortunately, separation of the WWRR adduct was not achieved due to time constrains but we hypothesize that arginine might perform either acylation or Michael addition+acylation on the Wittig product in a mechanism similar to that in the reaction between glyoxals and arginine (Figure 4-3C).[508–510] These observations prompted us to use phage display libraries of peptides to find sequences prone to undergo "cyclization".

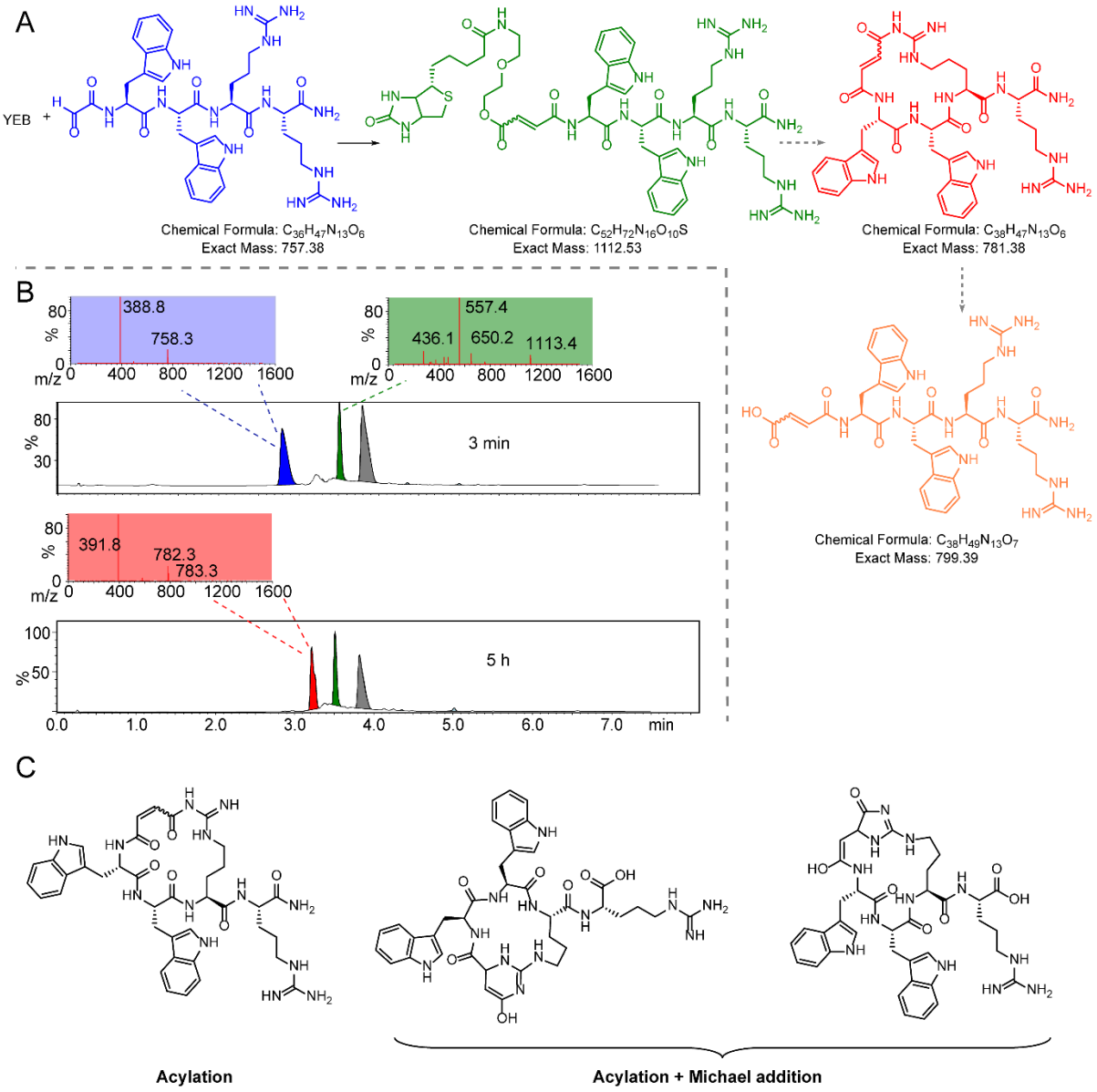


Figure 4-3. Observation of cyclization of o-WWRR while collecting traces for kinetics. (C) Proposed structures for Hydrolysis-H_2O adduct.

4.3 Results and Discussion

Wittig reaction between ~10^8 genetically encoded N-terminal glyoxaldehydes Ald-X_7 (where X denotes one of 20 natural amino acids displayed on phage) and a biotin labeled, ester stabilized phosphorene ylide, generated a library of peptides with electrophilic ester that contained a biotin tagged leaving group. Affinity capture on streptavidin beads separated the ester-containing members denoted Bio-COOR-X_7 from the rest of the peptide library without an ester moiety.

Exposure of the captured population to pH 11 can lead to either hydrolysis of the ester or an intramolecular attack of the ester by one of the nucleophilic functionalities present in the adjacent heptapeptide (Figure 4-1). We hypothesize that in certain cases, nucleophiles in privileged peptide sequences will favor nucleophilic acyl substitution at the ester carbonyl, releasing the biotin-tagged group and yielding a population enriched not only in hydrolyzed but also in cyclic motifs. Although a secondary screen could be employed to distinguish cyclic vs. hydrolyzed species, we sought to examine the entire eluted population and check whether in a small lead population, sequences exhibited a cyclization preference. Indeed, we show below that after testing of 6 synthetic peptides by LCMS we identified three that exhibited formation of species that lacked both the leaving group and H_2O moiety, suggesting intramolecular attack by side chains and possible formation of cyclic products.

4.3.1 0.1-1% Biotinylation of SX_7 Library with YEB to Introduce Alkyl Ester Tagged with Biotin

The SX_7 library was oxidized with $NaIO_4$ as described in our previous reports,[427] and exposed to biotin-tagged phosphorane ylide (YEB) under kinetically controlled conditions for 10 minutes reaction time and 0.04-0.4 mM [YEB] for 0.1-1% biotinylation respectively. We used SX_7 instead of focused, small libraries because higher diversity and longer sequences increase the chance for secondary interactions between side chains. We used specific conversions (1% and 0.1%) because during this time, a defined subset of library underwent Wittig reaction and most library members exhibited similar conversion (Appendix C-1). Specifically, in a library of 10^{10} pfu and 10^8 diverse sequences, we anticipate that 10^8 (1%) or 10^7 (0.1%) of them will be biotinylated. At these conversion percentages, removal of excess YEB is straightforward and capture of the entirety of the biotinylated library is possible. Performing the purification (removal of YEB by dialysis) and capture at pH 5 minimizes nucleophilic activity due to protonation of most side chains and unwanted hydrolysis of ester groups. We continued to employ YEB as ylide source because the ester functionality generated in YEB is sufficiently electrophilic to show potential reactivity with peptide side chains in preliminary studies (Figures 4-2, 4-3). Using esters derived from alcohols with lower pKa, while possible, would unnecessarily over activate the ester and make it difficult to handle in aqueous media.[511]

4.3.2 Screening Methodology to Find Cyclization Prone Sequences Using the SX_7 Library

Two rounds of selection were performed. In the first round, pull-down of the biotinylated populations was performed. After pull-down, phage attached to beads were eluted with buffer at pH 11 to favor hydrolysis of the ester moiety. Eluted population was amplified and used in a second round of selection**.** Because the pull-down process selects not only the biotinylated population but also peptide sequences that have affinity for other components in the system (non-specific binders), a set of two controls (Figure 4-4A) were used in the second round. Control C tested the intrinsic binding of the aldehyde-peptide population not biotinylated whereas control B tested for binding of biotinylated peptides to streptavidin with blocked biotin-binding site.

The Differential Enrichment (DE) matlab script (section 3.4.6.1) was used to find enriched sequences in the positive set. DE-analysis has two steps: (i) dividing the average normalized frequency in set A by the average normalized frequency of control sets to yield a ratio R for every sequence. The hits were defined as sequences with $R>3$ and $p<0.05$ (student t-test of normalized frequencies in 3 replicates of set A and 3 replicates in control sets yielded a significance estimate p). We observed that several enriched peptides in the hit set contained a positively charged amino-acid in the second position (Figure 4-4C). Probability of finding this amino-acid in this position in naïve SX_7 library is reduced since traditional libraries are depleted on positively charged amino-acids close to the N-terminus.[411] We hypothesized that appearance of such sequences was one of the indications of the selection pressure.

4.3.3 Validation of Selected Peptides

We picked six peptides that contained R or K (SRYVSAPL, SRLIDSPW, SQEDLTKA, SRLLPVHP, SRMLIHAV and SNLAPYRT), synthesized them and subjected them to the Wittig reaction. We observed that three of those sequences showed formation of the “cyclized” adduct (Figure 4-5).

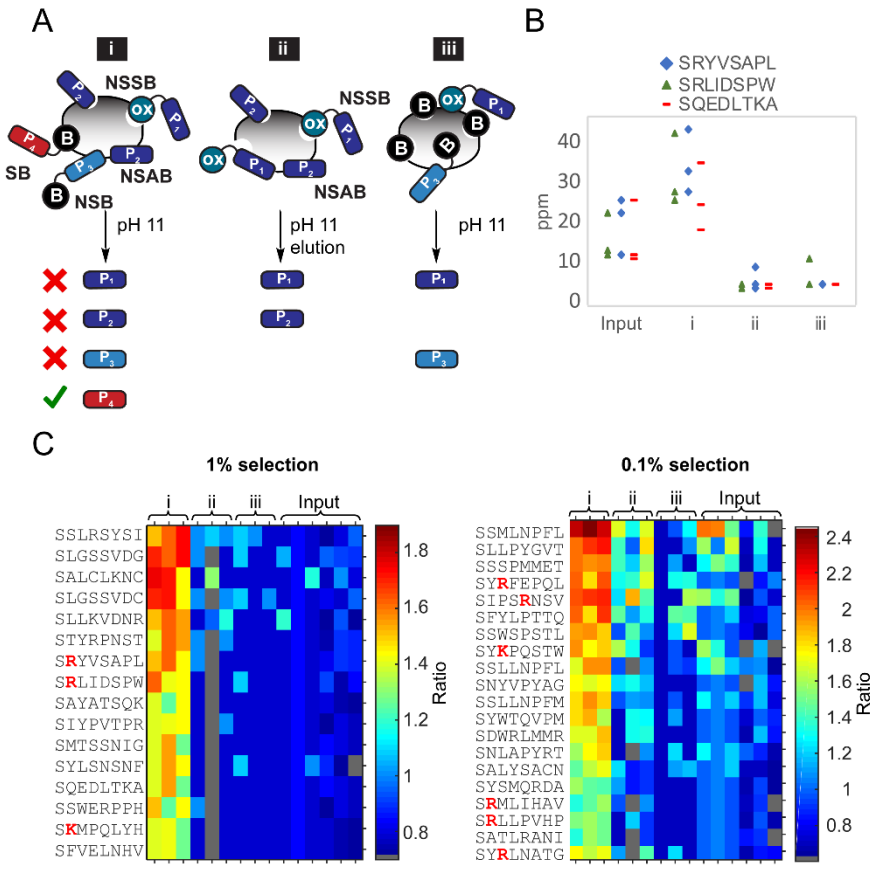


Figure 4-4. (A) Diagram of selection methodology depicting test and control sets. (B) Observed ppm counts for synthesized and validated sequences from 1% selection. (C) Heat map showing sequences for DE analysis with $p<0.05$ and $R=3$.

4.3.4 Characterization of Adducts by NMR

When attempting to isolate the adducts, we noticed that rate of hydrolysis competes with rate of "cyclization" (Appendix C-2). Despite this, the adducts for SRYVSAPL and SNLAPYRT sequences were isolated. Different 1D and 2D NMR (Appendix C-4) were performed in an attempt to characterize the products. In SRYVSAPL adduct, we observed that there was a mixture consisting in a major (A) and a minor (B) component. However, the signals from the major product were easily identifiable from the impurity. After identifying all signals from amino-acids with help of TOCSY, the only peculiar signal was a singlet at 7 ppm with integration for two hydrogens. This singlet is attached to a carbon at 135 ppm (HSQC data), a chemical shift in the region of

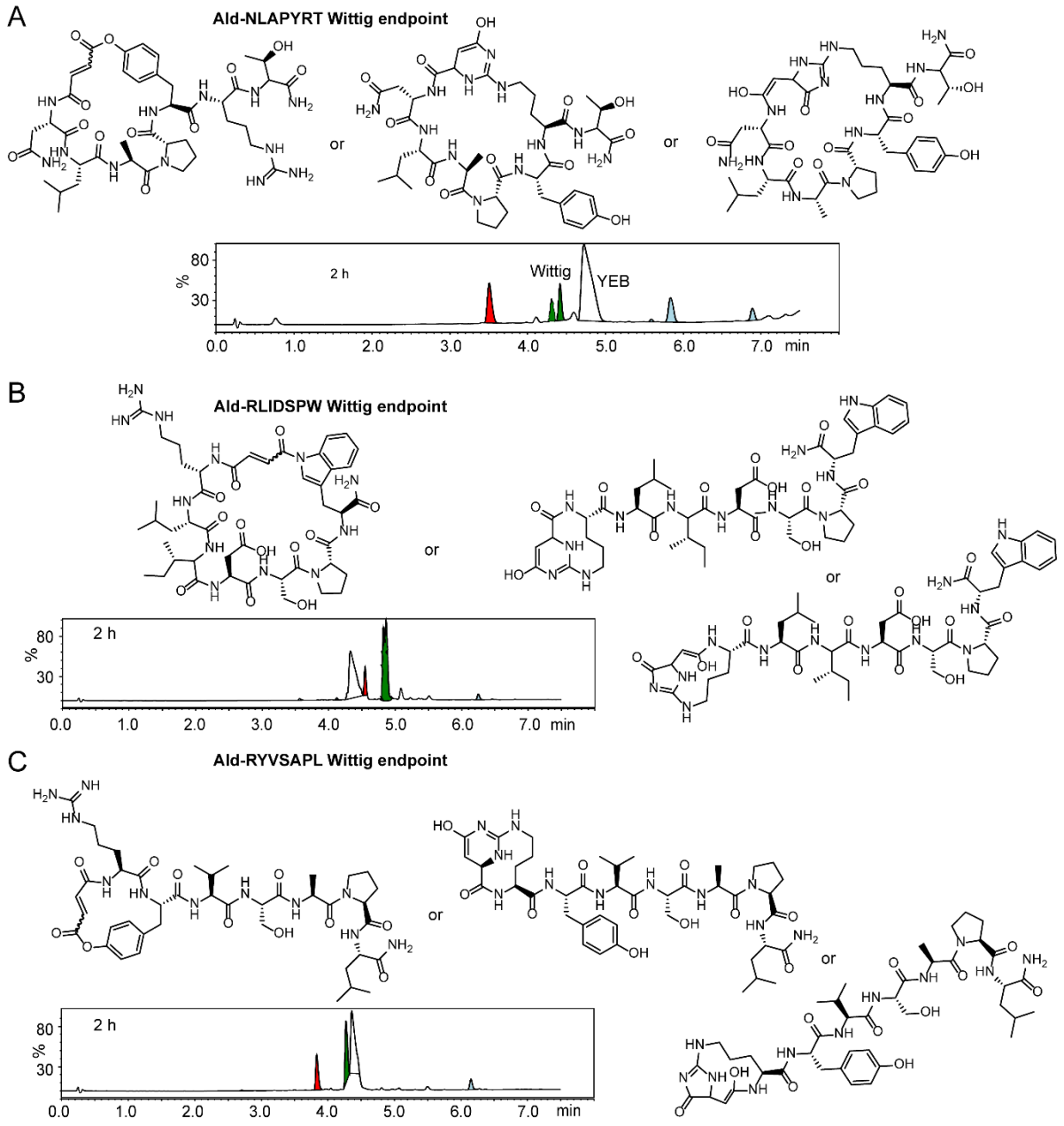


Figure 4-5. Observation of nucleophilic attack product during kinetic studies of (A) Ald-NLAPYRT, (B) Ald-RLIDSPW and (C) Ald-RYVSAPL

olefins and aromatic functional groups. The signal has a 2-bond coupling to a carbon at ~170 ppm (HMBC data), which has chemical shift for carbonyl functional group. The signal did not show spatial H-H coupling in the ROESY or 3-bond coupling in gmcCOSY experiments. Nucleophilic attack of the ester by an -NH_2 group produces a conjugated olefin that agrees with the chemical shifts for the olefin signal (135 ppm) and the HMBC coupling to carbonyl group. We propose that

the chemical environment of the resulting protons is similar enough to generate a singlet instead of the expected two doublets. Because we can observe both protons for the C-terminal -NH_2, we propose that the arginine acts as the nucleophile (Figure 4-6). In an attempt to determine if other sequence would show a similar signal, we isolated the adduct for Ald-NLAPYRT. The signal at the exact same chemical shift with same splitting and coupling behavior was observed. We can observe Asn and C-terminal NH_2 signals and therefore, we propose Arg nucleophilic attack again. We added D_2O to determine if the signal at 7 ppm was exchangeable, we did not observe any shift or broadening of the signal, meaning that the protons are not acidic (Appendix C-4). Comparing the nature of the 6 tested sequences, and knowing that only RYVSAPL, NLAPYRT and RLIDSPW (not isolated) undergo macrocyclization at the conditions used, we postulate that the sequences require Arg attacking group, Pro to pre-organize and generate a loop in the peptide and finally, an aromatic residue. Sequences QEDLTKA, RLLPVHP and RMLIHAV do not satisfy all of these conditions.

4.4 Conclusions

The phage display platform can and has been used to find peptide sequences prone to present certain reactivity. Here, we have demonstrated that the platform can be applied to find cyclization prone sequences. The selection methodology used here serves as a standpoint for future, more successful selection. The methodology can be tuned as to improve the ratio between cyclization prone and hydrolysis prone sequences. We propose that eluting at lower pH might diminish the percentage of false positives in the selection. We also propose that a chemistry able to differentiate between cyclized adducts and hydrolyzed ones might help in the selection process. We hypothesize that cyclized species should have higher Michael addition rate constants than the hydrolyzed ones due to presence of the carboxylate group in the latter that makes approach of thiolate group unfavorable. Thus, after elution, the population could be reacted with a tagged thiol and only the pulled-down set would be considered to contain “hits”.

4.5 Experimental Procedures

4.5.1 Materials, Methods and General Information

Acetate buffer 200 mM was prepared by mixing sodium acetate (16.4 g), NaCl (80.0 g) and adding 800 mL of milliQ water. The pH was adjusted to 5.00 with glacial acetic acid and the volume was completed to 1.00 L with milliQ water. MOPS buffer for kinetic measurements was prepared as a solution of 200 mM MOPS and adjusted to pH 6.50 with NaOH (1.00 M). All solutions used for phage work were prepared with milliQ water. Peptides were synthesized using the Prelude X peptide synthesizer (Gyros Protein Technologies) using 200 mg Rink Amide AM resin. Ylide Ester Biotin (YEB) was synthesized as already reported.[497] HRMS (ESI) spectra were recorded on Agilent 6220 oaTOF mass spectrometer using either positive or negative ionization mode. Characterization of reaction crude was performed with UPLC-MS using a C18 column (Phenomenex Kinetex 1.7 μm EVO C18, 2.1×50 mm) running with a gradient of water/acetonitrile with 0.1% formic acid from 98/2 at 0 min to 40/60 at 5 min under a flow rate of 0.5 mL/min. HPLC kinetics were performed in an Agilent 1100 Series using a Thermo Scientific C18 column (Hypersil GOLD, 3 μm, 50×2.1mm). The SX_7 library was bulk-amplified at home from a stock library kindly provided by New England Biolabs. ^{1}H and ^{1}H-^{1}H 2D NMR spectra were acquired on a 600 MHz four channel Agilent VNMRS spectrometer, equipped with a z gradient HCN probe and using VNMRJ 4.2A as the acquisition software. ^{1}H-^{13}C 2D NMR spectra were acquired on a 700 MHz spectrometer. Suppression of the H_2O signal was performed using either presaturation or excitation sculpting.[466]

4.5.2 Phage Functionalizations

4.5.2.1 SX_7 1% Biotinylation and Purification

10^{12} pfu of library were taken to a final volume of 0.10 mL with MOPS 0.20 M. To the solution, 1.0 μL of freshly prepared $NaIO_4$ (6.0 mM in water) was added and the mixture was incubated in ice for 8 minutes. Then, 1.0 μL of methionine (50 mM in water) was added and the mixture was incubated for 20 minutes at room temperature. 0.10 mL of YEB (0.80 mM in MOPS 0.20 M) were added and the reaction was incubated for 10 minutes at room temperature. After this, 0.80 mL of acetate buffer (0.20 M) and 0.20 mL of PEG NaCl were added and the mixture was kept at 4 °C for at least 12 hours. After this, the sample was centrifuged at 21100 x g for 30 minutes and the

supernatant was discarded. The sample was further centrifuged for 5 minutes and remaining supernatant was removed. The phage pellet was re-suspended in 0.20 mL of acetate buffer (10 mM) and incubated at room temperature for 15 minutes to allow phage to completely dissolve. After this, remaining solid debris was removed by centrifugation at 21100 x g for 5 minutes. The supernatant was transferred to a clean epi-tube and maintained at 4 °C for use in pull-down screenings.

4.5.2.2 SX_7 0.1% Biotinylation and Purification

10^{12} pfu of library were taken to a final volume of 0.10 mL with MOPS 0.20 M. To the solution, 1.0 μL of freshly prepared $NaIO_4$ (6.0 mM in water) was added and the mixture was incubated in ice for 8 minutes. Then, 1.0 μL of methionine (50 mM in water) was added and the mixture was incubated for 20 minutes. 0.10 mL of YEB (0.80 mM in MOPS 0.20 M) were added and the reaction was incubated for 10 minutes at room temperature. After this, PEG purification was performed as indicated in Section 3.4.2.3 and the phage was stored at 4 °C, ready to be used in screenings.

4.5.3 Pull-Down Experiments, PCR and Illumina Sequencing

Pull-down of 0.1-1% biotinylated libraries was performed following the methodologies described in section 3.4.3. PCR of samples was performed as described in section 3.4.4. PCR product characterization and Illumina sequencing was done as in Sections 3.4.5 and 3.4.6.

4.5.4 Peptide Oxidations and Wittig Rate Measurements

Peptides were oxidized following indications in Section 3.4.7. Wittig reaction rates were determined using the protocol in Section 3.4.8.

4.5.5 Isolation of Cyclized Adduct for SRYVSAPL

5.0 mg of oxidized peptide were dissolved in 2.0 mL of MOPPS 0.20 M at pH 6.5. 8.5 mg of YEB were dissolved in 0.80 mL of MOPPS 0.20 M at pH 6.5. Both solutions were mixed and incubated at room temperature for 3 h. The mixture was injected in C18 column, acetonitrile was evaporated using speedvac and the peptides were lyophilized.

4.5.6 Isolation of Cyclized Adduct for SNLAPYRT

4.0 mg of oxidized peptide were dissolved in 1.2 mL of MOPPS 0.20 M at pH 6.5. 6.6 mg of YEB were dissolver in 1.0 mL of MOPPS 0.20 mM at pH 6.5. Both solutions were mixed and incubated at room temperature for 3 h. The mixture was injected in C18 column, acetonitrile was evaporated using speedvac and the peptides were lyophilized.

Chapter 5: Conclusions and Outlook

5.1 Conclusions

Modern strategies for high-throughput optimization of reaction conditions and reaction discovery can be classified as one-well-one-experiment and one-well-multiple-experiments strategies. Unlike reaction that contains a mixture of tested reagents, traditional one-well-one-experiment strategies can be considered ideal because they can provide unambiguous understanding of reaction outcome in the absence of interfering and competing species. Amount of information acquired in this approach, however, scales linearly, with the number of experiments; throughout beyond 100 experiments often calls for significant optimization of workflow, specialized liquid handling and automation.[111] Molecular libraries that contain information tags allows tracing multiple molecules in a mixture and offer a potential to make selections of optimal reaction conditions or new reaction types faster and scalable by interrogating the reactivity of multiple reactions at once in a one-well-multiple-experiment set up. Genetically encoded libraries are a sub-type of tagged libraries that facilitate isolation and identification of reactive species; the amplifiable nature of the genetic tag permits tracking of molecules present in low copy number; in some implementation, use of GE-libraries also open the door to application of directed evolution techniques.

Phage display is one genetically encoded platform that has been successfully used for discovery of proteins and peptides that exhibit unique reactivity. Traditional phage display platform can incorporate only natural amino acids as building blocks in the library. Despite this limitation, these libraries attracted signification attention because they are commercially available or easy to produce, amplify and analyze. In the past decade, phage display platform has been combined with site-specific protein modification strategies to allow production of peptide libraries that contain moieties that are not encoded by translational machinery. Increased scope of reaction conditions compatible with phage-displayed peptide libraries opened the door to performing multi-step chemical transformation on these libraries. This capability allows applying selection and evolution techniques to investigate the reactivity of the “unnatural” functional groups (aldehydes, esters, conjugated alkenes) grafted into the peptide library. Specifically, we demonstrate that these libraries are well poised for investigating the effect of the functional groups present in peptides on

carbon-carbon bond forming reaction (Wittig reaction) and carbon-heteroatom reaction (intramolecular cyclization) that occur with these peptides.

Chapter 1 of this thesis was focused on development of effective multi-step functionalization of phage display libraries. Specifically, we developed the first example of the Wittig reaction that modifies a library phage displayed peptide-aldehydes produced by oxidation of naturel N-terminal serine. Optimal conditions allowed conversion of 40-70% peptide library with N-terminal serine to the peptide library with built-in N-terminal ester of maleic/fumaric acid. The latter functionality introduced by Wittig reaction serves either as a dienophile or Michael acceptor and efficiently underwent Diels-Alder cycloaddition and Michael addition reactions when installed on phage. In Chapter 2, we used the developed phage display platform for SAR studies of the Wittig reaction between ester stabilized ylide and glyoxaldehydes *in water*. We uncovered the importance of hydrogen bond between OPA transition state and N-H groups of backbone amide residues in the two amino-acid positions adjacent to the glyoxaldehyde. These groups, as well as tryptophan residues in the same positions accelerated Wittig reaction. Finally, in Chapter 4, we assessed whether the ester moiety installed on phage *via* the Wittig reaction can be attacked intromolecularly by side chain in specific sequences. We identified two sequences prone to cyclization, albeit the formed putative cyclic product is prone to hydrolysis and identification of their exact structures was not trivial. We propose Arginine as the attacking amino-acid and we propose the importance of proline and aromatic residues in these cyclization prone sequences.

5.2 Future studies

There is a wide range of water-compatible chemical reactions that have not been explored on phage and that have the potential to diversify these peptide libraries and increase their potential pharmacological value. In Chapter 2, we demonstrated that traditional organic chemistry reactions can be applied to phage and optimized with relative easiness. The SAR profiling in Chapter 3 was successful, however, to demonstrate the versatility of the developed platform, screenings with other reactions should be performed. We hypothesize that the methodology can be used for substrate profiling of other reactions that are already performed on the phage platform. This includes S_N2, S_NAr, Michael addition, oxime ligation, click chemistry, among others. These

studies would not only help to find optimal substrates with applications in biorthogonal chemistry but would also improve the understanding of chemical reactions *in water*.

Finally, the screenings performed in Chapter 4 can be improved to separate the hydrolysis-prone population from the cyclization-prone sequences. Because the selections were performed at pH 6.5, nucleophiles with higher pKa (Lysine) that would generate stable macrocycles were not deprotonated and only hydrolysis-prone adducts were discovered. The methodology can be modified as to increase the opportunity for Lysine residues to perform amidation. Also, alanine scan is needed to confirm the proposed importance of proline and aromatic residues in cyclization prone sequences. We also envision that determination of structural factors that favor macrocyclization of peptides can help in construction of libraries of linear precursors prone to undergo efficient and selective cyclization on phage. This would allow easy access to libraries of macrocycles that do not require encoding of two cysteines in constant positions and thus, libraries with different topologies to the traditional side-chain to side-chain cyclized libraries.

References

(1) Abad-Zapatero, C. *Drug Discov. Today* **2007**, *12* (23–24), 995.
(2) Zhang, M.-Q.; Wilkinson, B. *Curr. Opin. Biotechnol.* **2007**, *18* (6), 478.
(3) Mallinson, J.; Collins, I. *Future Med. Chem.* **2012**, *4* (11), 1409.
(4) Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J. *Chem. Biol.* **2014**, *21* (9), 1115.
(5) Giordanetto, F.; Kihlberg, J. *J. Med. Chem.* **2014**, *57* (2), 278.
(6) Luther, A.; Bisang, C.; Obrecht, D. *Bioorg. Med. Chem.* **2017**.
(7) Qian, Z.; Dougherty, P. G.; Pei, D. *Curr. Opin. Chem. Biol.* **2017**, *38*, 80.
(8) Cunningham, A. D.; Qvit, N.; Mochly-Rosen, D. *Curr. Opin. Struct. Biol.* **2017**, *44*, 59.
(9) Cardote, T. A. F.; Ciulli, A. *ChemMedChem* **2016**, *11* (8), 787.
(10) Dougherty, P. G.; Qian, Z.; Pei, D. *Biochem. J.* **2017**, *474* (7), 1109.
(11) Valeur, E.; Guéret, S. M.; Adihou, H.; Gopalakrishnan, R.; Lemurell, M.; Waldmann, H.; Grossmann, T. N.; Plowright, A. T. *Angew. Chemie Int. Ed.* **2017**, *56* (35), 10294.
(12) Naylor, M. R.; Bockus, A. T.; Blanco, M.-J.; Lokey, R. S. *Curr. Opin. Chem. Biol.* **2017**, *38*, 141.
(13) Zorzi, A.; Deyle, K.; Heinis, C. *Curr. Opin. Chem. Biol.* **2017**, *38*, 24.
(14) Surade, S.; Blundell, T. L. *Chem. Biol.* **2012**, *19* (1), 42.
(15) Bhat, A.; Roberts, L. R.; Dwyer, J. J. *Eur. J. Med. Chem.* **2015**, *94*, 471.
(16) Nielsen, D. S.; Shepherd, N. E.; Xu, W.; Lucke, A. J.; Stoermer, M. J.; Fairlie, D. P. *Chem. Rev.* **2017**, *117* (12), 8094.
(17) Saravolatz, L. D.; Deresinski, S. C.; Stevens, D. A. *Clin. Infect. Dis.* **2003**, *36* (11), 1445.
(18) Bright, G. M. N-Methyl 11-aza-10-deoxo-10-dihydro-erytromycin A, intermediates therefor. US4474768A, **1984**.
(19) Kobrehel, G.; Djokic, S. 11-Methyl-11-aza-4-0-cladinosyl-6-0-desosaminyl-15-ethyl-7,13,14-trihydroxy-3,5,7,9,12,14-hexamethyl-oxacyclopentadecane-2-one and derivatives thereof. US4517359A, **1985**.
(20) Lee, K. C.; Lee, T. H.; Tung, Y. S.; Kao, C. C.; Lee, T. A. Process for preparation of temsirolimus. US20100249415A1, **2010**.
(21) Nghiem, P.; Pearson, G.; Langley, R. G. *J. Am. Acad. Dermatol.* **2002**, *46* (2), 228.
(22) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. *Nat. Rev. Drug Discov.* **2008**, *7* (7), 608.
(23) Terrett, N. K. *Drug Discov. Today Technol.* **2010**, *7* (2), e97.
(24) Ramaseshan, M.; Dory, Y. L.; Deslongchamps, P. *J. Comb. Chem.* **2000**, *2* (6), 615.
(25) Roberts, K. D.; Ede, N. J. *J. Pept. Sci.* **2007**, *13* (12), 811.
(26) Chouhan, G.; James, K. *Org. Lett.* **2011**, *13* (10), 2754.
(27) Hu, T.-S.; Tannert, R.; Arndt, H.-D.; Waldmann, H. *Chem. Commun.* **2007**, No. 38, 3942.
(28) Chouhan, G.; James, K. *Org. Lett.* **2013**, *15* (6), 1206.
(29) Hiroshige, M.; Hauske, J. R.; Zhou, P. *J. Am. Chem. Soc.* **1995**, *117* (46), 11590.
(30) Takahashi, T.; Kusaka, S.; Doi, T.; Sunazuka, T.; Ōmura, S. *Angew. Chemie Int. Ed.* **2003**, *42* (42), 5230.
(31) Moulin, E.; Barluenga, S.; Totzke, F.; Winssinger, N. *Chem. - A Eur. J.* **2006**, *12* (34), 8819.
(32) Lee, D.; Sello, J. K.; Schreiber, S. L. *J. Am. Chem. Soc.* **1999**, *121* (45), 10648.
(33) Peng, L. F.; Stanton, B. Z.; Maloof, N.; Wang, X.; Schreiber, S. L. *Bioorg. Med. Chem. Lett.* **2009**, *19* (22), 6319.

(34) Morton, D.; Leach, S.; Cordier, C.; Warriner, S.; Nelson, A. *Angew. Chemie Int. Ed.* **2009**, *48* (1), 104.
(35) Wessjohann, L. A.; Rivera, D. G.; Vercillo, O. E. *Chem. Rev.* **2009**, *109* (2), 796.
(36) Ricardo, M. G.; Morales, F. E.; Garay, H.; Reyes, O.; Vasilev, D.; Wessjohann, L. A.; Rivera, D. G. *Org. Biomol. Chem.* **2015**, *13* (2), 438.
(37) Wessjohann, L. A.; Ruijter, E. *Mol. Divers.* **2005**, *9* (1–3), 159.
(38) Wessjohann, L. A.; Voigt, B.; Rivera, D. G. *Angew. Chemie Int. Ed.* **2005**, *44* (30), 4785.
(39) de Greef, M.; Abeln, S.; Belkasmi, K.; Dömling, A.; Orru, R.; Wessjohann, L. *Synthesis (Stuttg).* **2006**, *2006* (23), 3997.
(40) Rivera, D.; Vercillo, O.; Wessjohann, L. *Synlett* **2007**, *2007* (2), 308.
(41) Vercillo, O. E.; Andrade, C. K. Z.; Wessjohann, L. A. *Org. Lett.* **2008**, *10* (2), 205.
(42) Leon, F.; Rivera, D. G.; Wessjohann, L. A. *J. Org. Chem.* **2008**, *73* (5), 1762.
(43) Morejón, M. C.; Laub, A.; Kaluđerović, G. N.; Puentes, A. R.; Hmedat, A. N.; Otero-González, A. J.; Rivera, D. G.; Wessjohann, L. A. *Org. Biomol. Chem.* **2017**, *15* (17), 3628.
(44) Michalik, D.; Schaks, A.; Wessjohann, L. A. *European J. Org. Chem.* **2007**, *2007* (1), 149.
(45) Rivera, D. G.; Wessjohann, L. A. *J. Am. Chem. Soc.* **2009**, *131* (10), 3721.
(46) Lee, D.; Sello, J. K.; Schreiber, S. L. *Org. Lett.* **2000**, *2* (5), 709.
(47) Hebach, C.; Kazmaier, U. *Chem. Commun.* **2003**, No. 5, 596.
(48) Beck, B.; Larbig, G.; Mejat, B.; Magnin-Lachaux, M.; Picard, A.; Herdtweck, E.; Dömling, A. *Org. Lett.* **2003**, *5* (7), 1047.
(49) Cristau, P.; Vors, J.-P.; Zhu, J. *QSAR Comb. Sci.* **2006**, *25* (5–6), 519.
(50) Beck, B.; Magnin-Lachaux, M.; Herdtweck, E.; Dömling, A. *Org. Lett.* **2001**, *3* (18), 2875.
(51) Morejón, M. C.; Laub, A.; Westermann, B.; Rivera, D. G.; Wessjohann, L. A. *Org. Lett.* **2016**, *18* (16), 4096.
(52) Salvador, C. E. M.; Pieber, B.; Neu, P. M.; Torvisco, A.; Kleber Z. Andrade, C.; Kappe, C. O. *J. Org. Chem.* **2015**, *80* (9), 4590.
(53) Schmidt, D. R.; Kwon, O.; Schreiber, S. L. *J. Comb. Chem.* **2004**, *6* (2), 286.
(54) Dai, C.-F.; Cheng, F.; Xu, H.-C.; Ruan, Y.-P.; Huang, P.-Q. *J. Comb. Chem.* **2007**, *9* (3), 386.
(55) Marsault, E.; Hoveyda, H. R.; Gagnon, R.; Peterson, M. L.; Vézina, M.; Saint-Louis, C.; Landry, A.; Pinault, J.-F.; Ouellet, L.; Beauchemin, S.; Beaubien, S.; Mathieu, A.; Benakli, K.; Wang, Z.; Brassard, M.; Lonergan, D.; Bilodeau, F.; Ramaseshan, M.; Fortin, N.; Lan, R.; Li, S.; Galaud, F.; Plourde, V.; Champagne, M.; Doucet, A.; Bhérer, P.; Gauthier, M.; Olsen, G.; Villeneuve, G.; Bhat, S.; Foucher, L.; Fortin, D.; Peng, X.; Bernard, S.; Drouin, A.; Déziel, R.; Berthiaume, G.; Dory, Y. L.; Fraser, G. L.; Deslongchamps, P. *Bioorg. Med. Chem. Lett.* **2008**, *18* (16), 4731.
(56) Zhang, L.; Tam, J. P. *J. Am. Chem. Soc.* **1999**, *121* (14), 3311.
(57) Marsault, E.; Hoveyda, H. R.; Peterson, M. L.; Saint-Louis, C.; Landry, A.; Vézina, M.; Ouellet, L.; Wang, Z.; Ramaseshan, M.; Beaubien, S.; Benakli, K.; Beauchemin, S.; Déziel, R.; Peeters, T.; Fraser, G. L. *J. Med. Chem.* **2006**, *49* (24), 7190.
(58) Hoveyda, H. R.; Marsault, E.; Gagnon, R.; Mathieu, A. P.; Vézina, M.; Landry, A.; Wang, Z.; Benakli, K.; Beaubien, S.; Saint-Louis, C.; Brassard, M.; Pinault, J.-F.; Ouellet, L.; Bhat, S.; Ramaseshan, M.; Peng, X.; Foucher, L.; Beauchemin, S.; Bhérer, P.; Veber, D.

F.; Peterson, M. L.; Fraser, G. L. *J. Med. Chem.* **2011**, *54* (24), 8305.
(59) Zaccaro, M. C.; Lee, H. B.; Pattarawarapan, M.; Xia, Z.; Caron, A.; L'Heureux, P.-J.; Bengio, Y.; Burgess, K.; Saragovi, H. U. *Chem. Biol.* **2005**, *12* (9), 1015.
(60) Feng, Y.; Burgess, K. *Chem. - A Eur. J.* **1999**, *5* (11), 3261.
(61) Jefferson, E. A.; Swayze, E. E.; Osgood, S. A.; Miyaji, A.; Risen, L. M.; Blyn, L. B. *Bioorg. Med. Chem. Lett.* **2003**, *13* (10), 1635.
(62) Jefferson, E. A.; Arakawa, S.; Blyn, L. B.; Miyaji, A.; Osgood, S. A.; Ranken, R.; Risen, L. M.; Swayze, E. E. *J. Med. Chem.* **2002**, *45* (16), 3430.
(63) Marsault, E.; Peterson, M. L. *J. Med. Chem.* **2011**, *54* (7), 1961.
(64) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354* (6348), 82.
(65) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354* (6348), 84.
(66) Cho, C. Y.; Youngquist, R. S.; Paikoff, S. J.; Beresini, M. H.; Hebert, A. R.; Berleau, L. T.; Liu, C. W.; Wemmer, D. E.; Keough, T.; Schultz, P. G. *J. Am. Chem. Soc.* **1998**, *120* (31), 7706.
(67) Cho, C. Y.; Liu, C. W.; Wemmer, D. E.; Schultz, P. G. *Bioorg. Med. Chem.* **1999**, *7* (6), 1171.
(68) Lam, K. S.; Lebl, M.; Krchňák, V. *Chem. Rev.* **1997**, *97* (2), 411.
(69) Eichler, J.; Lucka, A. W.; Houghten, R. A. *Pept. Res.* **1994**, *7* (6), 300.
(70) Chen, J. J.; Teesch, L. M.; Spatola, A. F. *Lett. Pept. Sci.* **1996**, *3* (1), 17.
(71) Spatola, A. F.; Crozet, Y.; DeWit, D.; Yanagisawa, M. *J. Med. Chem.* **1996**, *39* (19), 3842.
(72) Spatola, A. F.; Darlak, K.; Romanovskis, P. *Tetrahedron Lett.* **1996**, *37* (5), 591.
(73) Virgilio, A. A.; Ellman, J. A. *J. Am. Chem. Soc.* **1994**, *116* (25), 11580.
(74) Blondelle, S. E.; Takashi, E.; Houghten, R. A.; Pérez-Payá, E. *Biochem. J.* **1996**, *313* (1), 141.
(75) Nefzi, A.; Ostresh, J. M.; Meyer, J.-P.; Houghten, R. A. *Tetrahedron Lett.* **1997**, *38* (6), 931.
(76) Bitan, G.; Sukhotinsky, I.; Mashriki, Y.; Hanani, M.; Selinger, Z.; Gilon, C. *J. Pept. Res.* **1997**, *49* (5), 421.
(77) Wen, J. J.; Spatola, A. F. *J. Pept. Res.* **1997**, *49* (1), 3.
(78) Crozet, Y.; Wen, J. J.; Loo, R. O.; Andrews, P. C.; Spatola, A. F. *Mol. Divers. 3* (4), 261.
(79) Nefzi, A.; Dooley, C.; Ostresh, J. M.; Houghten, R. A. *Bioorg. Med. Chem. Lett.* **1998**, *8* (17), 2273.
(80) Romanovskis, P.; Spatola, A. F. *J. Pept. Res.* **1998**, *52* (5), 356.
(81) Alsina, J.; Jensen, K. J.; Albericio, F.; Barany, G. *Chem. - A Eur. J.* **1999**, *5* (10), 2787.
(82) Bourne, G. T.; Golding, S. W.; Meutermans, W. D. F.; Smythe, M. L. *Lett. Pept. Sci.* **2000**, *7* (6), 311.
(83) Horton, D. A.; Bourne, G. T.; Smythe, M. L. *Mol. Divers.* **2000**, *5* (4), 289.
(84) Akaji, K.; Teruya, K.; Akaji, M.; Aimoto, S. *Tetrahedron* **2001**, *57* (12), 2293.
(85) Sulyok, G. A. G.; Gibson, C.; Goodman, S. L.; Hölzemann, G.; Wiesner, M.; Kessler, H. *J. Med. Chem.* **2001**, *44* (12), 1938.
(86) Basso, A.; Ernst, B. *Tetrahedron Lett.* **2001**, *42* (38), 6687.
(87) Bourne, G. T.; Golding, S. W.; McGeary, R. P.; Meutermans, W. D. F.; Jones, A.; Marshall, G. R.; Alewood, P. F.; Smythe, M. L. *J. Org. Chem.* **2001**, *66* (23), 7706.

(88) Royo, M.; Farrera-Sinfreu, J.; Solé, L.; Albericio, F. *Tetrahedron Lett.* **2002**, *43* (11), 2029.
(89) Horton, D. A.; Bourne, G. T.; Smythe, M. L. *J. Comput. Aided. Mol. Des.* **2002**, *16* (5/6), 415.
(90) Kohli, R. M.; Walsh, C. T.; Burkart, M. D. *Nature* **2002**, *418* (6898), 658.
(91) Jiménez, J. C.; Chavarría, B.; López-Macià, À.; Royo, M.; Giralt, E.; Albericio, F. *Org. Lett.* **2003**, *5* (12), 2115.
(92) Takahashi, T.; Nagamiya, H.; Doi, T.; Griffiths, P. G.; Bray, A. M. *J. Comb. Chem.* **2003**, *5* (4), 414.
(93) Roberts, K. D.; Lambert, J. N.; Ede, N. J.; Bray, A. M. *J. Pept. Sci.* **2004**, *10* (11), 659.
(94) Wels, B.; Kruijtzer, J. A. W.; Garner, K. M.; Adan, R. A. H.; Liskamp, R. M. J. *Bioorg. Med. Chem. Lett.* **2005**, *15* (2), 287.
(95) Roberts, K. D.; Lambert, J. N.; Ede, N. J.; Bray, A. M. *J. Pept. Sci.* **2006**, *12* (8), 525.
(96) Aina, O. H. *Mol. Cancer Ther.* **2005**, *4* (5), 806.
(97) Lau, D. H.; Guo, L.; Liu, R.; Song, A.; Shao, C.; Lam, K. S. *Biotechnol. Lett.* **2002**, *24* (6), 497.
(98) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013**, *30* (1), 108.
(99) Zhang, Q.; Yang, X.; Wang, H.; van der Donk, W. A. *ACS Chem. Biol.* **2014**, *9* (11), 2686.
(100) Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P. J.; Joewono, I.; Rusch, D.; Chisholm, S. W.; van der Donk, W. A. *Proc. Natl. Acad. Sci.* **2010**, *107* (23), 10430.
(101) Ziemert, N.; Ishida, K.; Weiz, A.; Hertweck, C.; Dittmann, E. *Appl. Environ. Microbiol.* **2010**, *76* (11), 3568.
(102) Ruffner, D. E.; Schmidt, E. W.; Heemstra, J. R. *ACS Synth. Biol.* **2015**, *4* (4), 482.
(103) Sardar, D.; Schmidt, E. W. *Curr. Opin. Chem. Biol.* **2016**, *31*, 15.
(104) McIntosh, J. A.; Lin, Z.; Tianero, M. D. B.; Schmidt, E. W. *ACS Chem. Biol.* **2013**, *8* (5), 877.
(105) McIntosh, J. A.; Robertson, C. R.; Agarwal, V.; Nair, S. K.; Bulaj, G. W.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132* (44), 15499.
(106) Sardar, D.; Lin, Z.; Schmidt, E. W. *Chem. Biol.* **2015**, *22* (7), 907.
(107) Li, C.; Zhang, F.; Kelly, W. L. *Mol. BioSyst.* **2011**, *7* (1), 82.
(108) Kluskens, L. D.; Kuipers, A.; Rink, R.; de Boef, E.; Fekken, S.; Driessen, A. J. M.; Kuipers, O. P.; Moll, G. N. *Biochemistry* **2005**, *44* (38), 12827.
(109) Levengood, M. R.; Knerr, P. J.; Oman, T. J.; van der Donk, W. A. *J. Am. Chem. Soc.* **2009**, *131* (34), 12024.
(110) Zhang, Q.; Yu, Y.; Velasquez, J. E.; van der Donk, W. A. *Proc. Natl. Acad. Sci.* **2012**, *109*

(45), 18361.
(111) McNally, A.; Prier, C. K.; MacMillan, D. W. C. *Science.* **2011**, *334* (6059), 1114.
(112) Palmblad, M.; Drijfhout, J. W.; Deelder, A. M. *J. Comb. Chem.* **2010**, *12* (1), 65.
(113) Brown, J. M.; Hoffmann, W. D.; Alvey, C. M.; Wood, A. R.; Verbeck, G. F.; Petros, R. A. *Anal. Biochem.* **2010**, *398* (1), 7.
(114) Semmler, A.; Weber, R.; Przybylski, M.; Wittmann, V. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (2), 215.
(115) Lee, S. S.; Lim, J.; Tan, S.; Cha, J.; Yeo, S. Y.; Agnew, H. D.; Heath, J. R. *Anal. Chem.* **2010**, *82* (2), 672.
(116) Cha, J.; Lim, J.; Zheng, Y.; Tan, S.; Ang, Y. L.; Oon, J.; Ang, M. W.; Ling, J.; Bode, M.; Lee, S. S. *J. Lab. Autom.* **2012**, *17* (3), 186.
(117) Ren, J.; Mann, Y. S.; Zhang, Y.; Browne, M. D. *J. Vis. Exp.* **2018**, No. 132.
(118) Redman, J. E.; Wilcoxen, K. M.; Ghadiri, M. R. *J. Comb. Chem.* **2003**, *5* (1), 33.
(119) Lee, J. H.; Meyer, A. M.; Lim, H.-S. *Chem. Commun.* **2010**, *46* (45), 8615.
(120) Bryson, D. I.; Zhang, W.; Ray, W. K.; Santos, W. L. *Mol. Biosyst.* **2009**, *5* (9), 1070.
(121) Lee, K. J.; Lim, H.-S. *Org. Lett.* **2014**, *16* (21), 5710.
(122) Simpson, L. S.; Kodadek, T. *Tetrahedron Lett.* **2012**, *53* (18), 2341.
(123) Liang, X.; Girard, A.; Biron, E. *ACS Comb. Sci.* **2013**, *15* (10), 535.
(124) Gurevich-Messina, J. M.; Giudicessi, S. L.; Martínez-Ceron, M. C.; Acosta, G.; Erra-Balsells, R.; Cascone, O.; Albericio, F.; Camperi, S. A. *J. Pept. Sci.* **2015**, *21* (1), 40.
(125) Marani, M. M.; Oliveira, E.; Côte, S.; Camperi, S. A.; Albericio, F.; Cascone, O. *Anal. Biochem.* **2007**, *370* (2), 215.
(126) Vinogradov, A. A.; Gates, Z. P.; Zhang, C.; Quartararo, A. J.; Halloran, K. H.; Pentelute, B. L. *ACS Comb. Sci.* **2017**, *19* (11), 694.
(127) Liu, R.; Marik, J.; Lam, K. S. *J. Am. Chem. Soc.* **2002**, *124* (26), 7678.
(128) Joo, S. H.; Xiao, Q.; Ling, Y.; Gopishetty, B.; Pei, D. *J. Am. Chem. Soc.* **2006**, *128* (39), 13000.
(129) Dewan, V.; Liu, T.; Chen, K.-M.; Qian, Z.; Xiao, Y.; Kleiman, L.; Mahasenan, K. V.; Li, C.; Matsuo, H.; Pei, D.; Musier-Forsyth, K. *ACS Chem. Biol.* **2012**, *7* (4), 761.
(130) Wu, X.; Upadhyaya, P.; Villalona-Calero, M. A.; Briesewitz, R.; Pei, D. *Med. Chem. Commun.* **2013**, *4* (2), 378.
(131) Lian, W.; Upadhyaya, P.; Rhodes, C. A.; Liu, Y.; Pei, D. *J. Am. Chem. Soc.* **2013**, *135* (32), 11990.
(132) Upadhyaya, P.; Qian, Z.; Selner, N. G.; Clippinger, S. R.; Wu, Z.; Briesewitz, R.; Pei, D. *Angew. Chemie Int. Ed.* **2015**, *54* (26), 7602.
(133) Upadhyaya, P.; Qian, Z.; Habir, N. A. A.; Pei, D. *Tetrahedron* **2014**, *70* (42), 7714.
(134) Astle, J. M.; Simpson, L. S.; Huang, Y.; Reddy, M. M.; Wilson, R.; Connell, S.; Wilson, J.; Kodadek, T. *Chem. Biol.* **2010**, *17* (1), 38.
(135) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci.* **1992**, *89* (12), 5381.
(136) Nielsen, J.; Brenner, S.; Janda, K. D. *J. Am. Chem. Soc.* **1993**, *115* (21), 9812.
(137) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci.* **1993**, *90* (22), 10700.
(138) Gartner, Z. J.; Liu, D. R. *J. Am. Chem. Soc.* **2001**, *123* (28), 6961.
(139) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *Angew. Chemie Int. Ed.* **2002**, *41* (10), 1796.
(140) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2002**, *124* (35), 10304.
(141) Calderone, C. T.; Puckett, J. W.; Gartner, Z. J.; Liu, D. R. *Angew. Chemie Int. Ed.* **2002**,

41 (21), 4104.
(142) Gartner, Z. J.; Grubina, R.; Calderone, C. T.; Liu, D. R. *Angew. Chemie Int. Ed.* **2003**, *42* (12), 1370.
(143) Li, X.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125* (34), 10188.
(144) Rosenbaum, D. M.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125* (46), 13924.
(145) Li, X.; Gartner, Z. J.; Tse, B. N.; Liu, D. R. *J. Am. Chem. Soc.* **2004**, *126* (16), 5090.
(146) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. *Science.* **2004**, *305* (5690), 1601.
(147) Tse, B. N.; Snyder, T. M.; Shen, Y.; Liu, D. R. *J. Am. Chem. Soc.* **2008**, *130* (46), 15611.
(148) Kodadek, T.; McEnaney, P. J. *Chem. Commun.* **2016**, *52* (36), 6038.
(149) MacConnell, A. B.; McEnaney, P. J.; Cavett, V. J.; Paegel, B. M. *ACS Comb. Sci.* **2015**, *17* (9), 518.
(150) MacConnell, A. B.; Price, A. K.; Paegel, B. M. *ACS Comb. Sci.* **2017**, *19* (3), 181.
(151) Seigal, B. A.; Connors, W. H.; Fraley, A.; Borzilleri, R. M.; Carter, P. H.; Emanuel, S. L.; Fargnoli, J.; Kim, K.; Lei, M.; Naglich, J. G.; Pokross, M. E.; Posy, S. L.; Shen, H.; Surti, N.; Talbott, R.; Zhang, Y.; Terrett, N. K. *J. Med. Chem.* **2015**, *58* (6), 2855.
(152) Kleiner, R. E.; Dumelin, C. E.; Tiu, G. C.; Sakurai, K.; Liu, D. R. *J. Am. Chem. Soc.* **2010**, *132* (33), 11779.
(153) Georghiou, G.; Kleiner, R. E.; Pulkoski-Gross, M.; Liu, D. R.; Seeliger, M. A. *Nat. Chem. Biol.* **2012**, *8* (4), 366.
(154) Kleiner, R. E.; Dumelin, C. E.; Liu, D. R. *Chem. Soc. Rev.* **2011**, *40* (12), 5707.
(155) Maianti, J. P.; McFedries, A.; Foda, Z. H.; Kleiner, R. E.; Du, X. Q.; Leissring, M. A.; Tang, W.-J.; Charron, M. J.; Seeliger, M. A.; Saghatelian, A.; Liu, D. R. *Nature* **2014**, *511* (7507), 94.
(156) Usanov, D. L.; Chan, A. I.; Maianti, J. P.; Liu, D. R. *Nat. Chem.* **2018**, *10* (7), 704.
(157) Zhu, Z.; Shaginian, A.; Grady, L. C.; O'Keeffe, T.; Shi, X. E.; Davie, C. P.; Simpson, G. L.; Messer, J. A.; Evindar, G.; Bream, R. N.; Thansandote, P. P.; Prentice, N. R.; Mason, A. M.; Pal, S. *ACS Chem. Biol.* **2017**, acschembio.7b00852.
(158) Lu, X.; Fan, L.; Phelps, C. B.; Davie, C. P.; Donahue, C. P. *Bioconjug. Chem.* **2017**, *28* (6), 1625.
(159) Arico-Muendel, C. C. *Medchemcomm* **2016**, *7* (10), 1898.
(160) Mullard, A. *Nat. Biotechnol.* **2016**, *34* (5), 450.
(161) Gura, T. *Science.* **2015**, *350* (6265), 1139.
(162) Satz, A. L. *ACS Med. Chem. Lett.* **2018**, *9* (5), 408.
(163) Kontijevskis, A. *J. Chem. Inf. Model.* **2017**, *57* (4), 680.
(164) Kuai, L.; O'Keeffe, T.; Arico-Muendel, C. *SLAS Discov. Adv. Life Sci. R&D* **2018**, 405.
(165) Connors, W. H.; Hale, S. P.; Terrett, N. K. *Curr. Opin. Chem. Biol.* **2015**, *26*, 42.
(166) Yuen, L. H.; Franzini, R. M. *ChemBioChem* **2017**, *18* (9), 829.
(167) Shi, B.; Zhou, Y.; Huang, Y.; Zhang, J.; Li, X. *Bioorg. Med. Chem. Lett.* **2017**, *27* (3), 361.
(168) Malone, M. L.; Paegel, B. M. *ACS Comb. Sci.* **2016**, *18* (4), 182.
(169) Satz, A. L.; Hochstrasser, R.; Petersen, A. C. *ACS Comb. Sci.* **2017**, *19* (4), 234.
(170) Kane, P.; Yamashiro, C.; Wolczyk, D.; Neff, N.; Goebl, M.; Stevens, T. *Science.* **1990**, *250* (4981), 651.
(171) Clarke, N. D. *Proc. Natl. Acad. Sci.* **1994**, *91* (23), 11084.
(172) Southworth, M. W. *EMBO J.* **1998**, *17* (4), 918.

(173) Wu, H.; Hu, Z.; Liu, X. Q. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (16), 9226.
(174) Volkmann, G.; Iwaï, H. *Mol. Biosyst.* **2010**, *6* (11), 2110.
(175) Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D. C.; Benkovic, S. J. *Proc. Natl. Acad. Sci.* **1999**, *96* (24), 13638.
(176) Scott, C. P.; Abel-Santos, E.; Jones, A. D.; Benkovic, S. J. *Chem. Biol.* **2001**, *8* (8), 801.
(177) Tavassoli, A.; Benkovic, S. J. *Nat. Protoc.* **2007**, *2* (5), 1126.
(178) Evans, T. C.; Martin, D.; Kolly, R.; Panne, D.; Sun, L.; Ghosh, I.; Chen, L.; Benner, J.; Liu, X. Q.; Xu, M. Q. *J. Biol. Chem.* **2000**, *275* (13), 9091.
(179) Townend, J. E.; Tavassoli, A. *ACS Chem. Biol.* **2016**, *11* (6), 1624.
(180) Osher, E. L.; Tavassoli, A. In *Split Inteins*; Mootz, H. D., Ed.; Springer New York: New York, NY, 2017; pp 27–39.
(181) Lennard, K. R.; Tavassoli, A. *Chem. - A Eur. J.* **2014**, *20* (34), 10608.
(182) Tavassoli, A. *Curr. Opin. Chem. Biol.* **2017**, *38*, 30.
(183) Young, T. S.; Young, D. D.; Ahmad, I.; Louis, J. M.; Benkovic, S. J.; Schultz, P. G. *Proc. Natl. Acad. Sci.* **2011**, *108* (27), 11052.
(184) Naumann, T. A.; Tavassoli, A.; Benkovic, S. J. *ChemBioChem* **2008**, *9* (2), 194.
(185) Cheng, L.; Naumann, T. A.; Horswill, A. R.; Hong, S.-J.; Venters, B. J.; Tomsho, J. W.; Benkovic, S. J.; Keiler, K. C. *Protein Sci.* **2007**, *16* (8), 1535.
(186) Nilsson, L.; Louassini, M.; Abel-Santos, E. *Protein Pept. Lett.* **2005**, *12* (8), 795.
(187) Tavassoli, A.; Lu, Q.; Gam, J.; Pan, H.; Benkovic, S. J.; Cohen, S. N. *ACS Chem. Biol.* **2008**, *3* (12), 757.
(188) El-Mowafi, S. A.; Alumasa, J. N.; Ades, S. E.; Keiler, K. C. *Antimicrob. Agents Chemother.* **2014**, *58* (9), 5500.
(189) Horswill, A. R.; Benkovic, S. J. *Curr. Protoc. protein Sci.* **2006**, *Chapter 19*, Unit 19.15.
(190) Tavassoli, A.; Benkovic, S. J. *Angew. Chemie Int. Ed.* **2005**, *44* (18), 2760.
(191) Kjelstrup, S.; Hansen, P. M. P.; Thomsen, L. E.; Hansen, P. R.; Løbner-Olesen, A. *PLoS One* **2013**, *8* (9), e72273.
(192) Datta, S.; Bucks, M. E.; Koley, D.; Lim, P. X.; Savinov, S. N. *Bioorg. Med. Chem.* **2010**, *18* (16), 6099.
(193) Verhoeven, K. D.; Altstadt, O. C.; Savinov, S. N. *Appl. Biochem. Biotechnol.* **2012**, *166* (5), 1340.
(194) Cheng, H.; Linhares, B. M.; Yu, W.; Cardenas, M. G.; Ai, Y.; Jiang, W.; Winkler, A.; Cohen, S.; Melnick, A.; MacKerell, A.; Cierpicki, T.; Xue, F. *J. Med. Chem.* **2018**, *In Press.*
(195) Osher, E. L.; Castillo, F.; Elumalai, N.; Waring, M. J.; Pairaudeau, G.; Tavassoli, A. *Bioorg. Med. Chem.* **2018**, *26* (11), 3034.
(196) Kinsella, T. M.; Ohashi, C. T.; Harder, A. G.; Yam, G. C.; Li, W.; Peelle, B.; Pali, E. S.; Bennett, M. K.; Molineaux, S. M.; Anderson, D. A.; Masuda, E. S.; Payan, D. G. *J. Biol. Chem.* **2002**, *277* (40), 37512.
(197) Miranda, E.; Nordgren, I. K.; Male, A. L.; Lawrence, C. E.; Hoakwie, F.; Cuda, F.; Court, W.; Fox, K. R.; Townsend, P. A.; Packham, G. K.; Eccles, S. A.; Tavassoli, A. *J. Am. Chem. Soc.* **2013**, *135* (28), 10418.
(198) Mistry, I. N.; Tavassoli, A. *ACS Synth. Biol.* **2017**, *6* (3), 518.
(199) Birts, C. N.; Nijjar, S. K.; Mardle, C. A.; Hoakwie, F.; Duriez, P. J.; Blaydes, J. P.; Tavassoli, A. *Chem. Sci.* **2013**, *4* (8), 3046.
(200) Gohard, F. H.; St-Cyr, D. J.; Tyers, M.; Earnshaw, W. C. *Open Biol.* **2014**, *4* (11),

140163.
(201) Kritzer, J. A.; Hamamichi, S.; McCaffery, J. M.; Santagata, S.; Naumann, T. A.; Caldwell, K. A.; Caldwell, G. A.; Lindquist, S. *Nat. Chem. Biol.* **2009**, *5* (9), 655.
(202) Barreto, K.; Bharathikumar, V. M.; Ricardo, A.; DeCoteau, J. F.; Luo, Y.; Geyer, C. R. *Chem. Biol.* **2009**, *16* (11), 1148.
(203) Bharathikumar, V. M.; Barreto, K.; DeCoteau, J. F.; Geyer, C. R. *ChemBioChem* **2013**, *14* (16), 2119.
(204) Male, A. L.; Forafonov, F.; Cuda, F.; Zhang, G.; Zheng, S.; Oyston, P. C. F.; Chen, P. R.; Williamson, E. D.; Tavassoli, A. *Sci. Rep.* **2017**, *7* (1), 3104.
(205) Deschuyteneer, G.; Garcia, S.; Michiels, B.; Baudoux, B.; Degand, H.; Morsomme, P.; Soumillion, P. *ACS Chem. Biol.* **2010**, *5* (7), 691.
(206) Böcker, J. K.; Friedel, K.; Matern, J. C. J.; Bachmann, A.-L.; Mootz, H. D. *Angew. Chemie Int. Ed.* **2015**, *54* (7), 2116.
(207) Frost, J. R.; Smith, J. M.; Fasan, R. *Curr. Opin. Struct. Biol.* **2013**, *23* (4), 571.
(208) Smith, J. M.; Vitali, F.; Archer, S. A.; Fasan, R. *Angew. Chemie Int. Ed.* **2011**, *50* (22), 5075.
(209) Smith, J. M.; Fasan, R. In *Peptide Libraries*; **2015**; pp 23–38.
(210) Satyanarayana, M.; Vitali, F.; Frost, J. R.; Fasan, R. *Chem. Commun.* **2012**, *48* (10), 1461.
(211) Frost, J. R.; Vitali, F.; Jacob, N. T.; Brown, M. D.; Fasan, R. *ChemBioChem* **2013**, *14* (1), 147.
(212) Owens, A. E.; de Paola, I.; Hansen, W. A.; Liu, Y.-W.; Khare, S. D.; Fasan, R. *J. Am. Chem. Soc.* **2017**, *139* (36), 12559.
(213) Mattheakis, L. C.; Bhatt, R. R.; Dower, W. J. *Proc. Natl. Acad. Sci.* **1994**, *91* (19), 9022.
(214) Mattheakis, L. C.; Dias, J. M.; Dower, W. J. In *Methods in Enzymology*; **1996**; pp 195–207.
(215) Hanes, J.; Pluckthun, A. *Proc. Natl. Acad. Sci.* **1997**, *94* (10), 4937.
(216) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (23), 12297.
(217) Roberts, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3* (3), 268.
(218) Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. *FEBS Lett.* **1997**, *414* (2), 405.
(219) Liu, R.; Barrick, J. E.; Szostak, J. W.; Roberts, R. W. In *Methods in Enzymology*; **2000**; pp 268–293.
(220) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19* (8), 751.
(221) Forster, A. C.; Tan, Z.; Nalam, M. N. L.; Lin, H.; Qu, H.; Cornish, V. W.; Blacklow, S. C. *Proc. Natl. Acad. Sci.* **2003**, *100* (11), 6353.
(222) Pavlov, M. Y.; Watts, R. E.; Tan, Z.; Cornish, V. W.; Ehrenberg, M.; Forster, A. C. *Proc. Natl. Acad. Sci.* **2009**, *106* (1), 50.
(223) Tan, Z.; Blacklow, S. C.; Cornish, V. W.; Forster, A. C. *Methods* **2005**, *36* (3), 279.
(224) Zhang, B.; Tan, Z.; Dickson, L. G.; Nalam, M. N. L.; Cornish, V. W.; Forster, A. C. *J. Am. Chem. Soc.* **2007**, *129* (37), 11316.
(225) Kang, T. J.; Suga, H. *Biochem. Cell Biol.* **2008**, *86* (2), 92.
(226) Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *Proc. Natl. Acad. Sci.* **2006**, *103* (12), 4356.
(227) Hartman, M. C. T.; Josephson, K.; Lin, C.-W.; Szostak, J. W. *PLoS One* **2007**, *2* (10), e972.

(228) Morimoto, J.; Hayashi, Y.; Iwasaki, K.; Suga, H. *Acc. Chem. Res.* **2011**, *44* (12), 1359.
(229) Passioura, T.; Suga, H. *Chem. - A Eur. J.* **2013**, *19* (21), 6530.
(230) Goto, Y.; Suga, H. In *Ribozymes*; **2012**; pp 465–478.
(231) Goto, Y.; Katoh, T.; Suga, H. *Nat. Protoc.* **2011**, *6* (6), 779.
(232) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. *Nat. Methods* **2006**, *3* (5), 357.
(233) Niwa, N.; Yamagishi, Y.; Murakami, H.; Suga, H. *Bioorg. Med. Chem. Lett.* **2009**, *19* (14), 3892.
(234) Passioura, T.; Katoh, T.; Goto, Y.; Suga, H. *Annu. Rev. Biochem.* **2014**, *83* (1), 727.
(235) Josephson, K.; Hartman, M. C. T.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127* (33), 11727.
(236) Millward, S. W.; Takahashi, T. T.; Roberts, R. W. *J. Am. Chem. Soc.* **2005**, *127* (41), 14142.
(237) Goto, Y.; Iwasaki, K.; Torikai, K.; Murakami, H.; Suga, H. *Chem. Commun.* **2009**, No. 23, 3419.
(238) Yamagishi, Y.; Ashigai, H.; Goto, Y.; Murakami, H.; Suga, H. *ChemBioChem* **2009**, *10* (9), 1469.
(239) Guillen Schlippe, Y. V.; Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134* (25), 10469.
(240) Seebeck, F. P.; Szostak, J. W. *J. Am. Chem. Soc.* **2006**, *128* (22), 7150.
(241) Hacker, D. E.; Hoinka, J.; Iqbal, E. S.; Przytycka, T. M.; Hartman, M. C. T. *ACS Chem. Biol.* **2017**, *12* (3), 795.
(242) Kawakami, T.; Ohta, A.; Ohuchi, M.; Ashigai, H.; Murakami, H.; Suga, H. *Nat. Chem. Biol.* **2009**, *5* (12), 888.
(243) Ohshiro, Y.; Nakajima, E.; Goto, Y.; Fuse, S.; Takahashi, T.; Doi, T.; Suga, H. *ChemBioChem* **2011**, *12* (8), 1183.
(244) Kang, T. J.; Hayashi, Y.; Suga, H. *Angew. Chemie Int. Ed.* **2011**, *50* (9), 2159.
(245) Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H. *ACS Chem. Biol.* **2008**, *3* (2), 120.
(246) Sako, Y.; Morimoto, J.; Murakami, H.; Suga, H. *J. Am. Chem. Soc.* **2008**, *130* (23), 7232.
(247) Bashiruddin, N. K.; Nagano, M.; Suga, H. *Bioorg. Chem.* **2015**, *61*, 45.
(248) Jongkees, S. A. K.; Umemoto, S.; Suga, H. *Chem. Sci.* **2017**, *8* (2), 1474.
(249) Hofmann, F. T.; Szostak, J. W.; Seebeck, F. P. *J. Am. Chem. Soc.* **2012**, *134* (19), 8038.
(250) Millward, S. W.; Fiacco, S.; Austin, R. J.; Roberts, R. W. *ACS Chem. Biol.* **2007**, *2* (9), 625.
(251) Morimoto, J.; Hayashi, Y.; Suga, H. *Angew. Chemie Int. Ed.* **2012**, *51* (14), 3423.
(252) Hayashi, Y.; Morimoto, J.; Suga, H. *ACS Chem. Biol.* **2012**, *7* (3), 607.
(253) Hipolito, C.; Tanaka, Y.; Katoh, T.; Nureki, O.; Suga, H. *Molecules* **2013**, *18* (9), 10514.
(254) Tanaka, Y.; Hipolito, C. J.; Maturana, A. D.; Ito, K.; Kuroda, T.; Higuchi, T.; Katoh, T.; Kato, H. E.; Hattori, M.; Kumazaki, K.; Tsukazaki, T.; Ishitani, R.; Suga, H.; Nureki, O. *Nature* **2013**, *496* (7444), 247.
(255) Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H. *Chem. Biol.* **2011**, *18* (12), 1562.
(256) Kawakami, T.; Ishizawa, T.; Fujino, T.; Reid, P. C.; Suga, H.; Murakami, H. *ACS Chem. Biol.* **2013**, *8* (6), 1205.
(257) Nishio, K.; Belle, R.; Katoh, T.; Kawamura, A.; Sengoku, T.; Hanada, K.; Ohsawa, N.; Shirouzu, M.; Yokoyama, S.; Suga, H. *ChemBioChem* **2018**, *19* (9), 979.

(258) Hwang, D. W.; Bahng, N.; Ito, K.; Ha, S.; Kim, M. Y.; Lee, E.; Suga, H.; Lee, D. S. *Cancer Lett.* **2017**, *385*, 144.
(259) Ito, K.; Sakai, K.; Suzuki, Y.; Ozawa, N.; Hatta, T.; Natsume, T.; Matsumoto, K.; Suga, H. *Nat. Commun.* **2015**, *6* (1), 6373.
(260) Iwasaki, K.; Goto, Y.; Katoh, T.; Yamashita, T.; Kaneko, S.; Suga, H. *J. Mol. Evol.* **2015**, *81* (5–6), 210.
(261) Matsunaga, Y.; Bashiruddin, N. K.; Kitago, Y.; Takagi, J.; Suga, H. *Cell Chem. Biol.* **2016**, *23* (11), 1341.
(262) Bashiruddin, N. K.; Matsunaga, Y.; Nagano, M.; Takagi, J.; Suga, H. *Bioconjug. Chem.* **2018**, *29* (6), 1847.
(263) Kawamura, A.; Münzel, M.; Kojima, T.; Yapp, C.; Bhushan, B.; Goto, Y.; Tumber, A.; Katoh, T.; King, O. N. F.; Passioura, T.; Walport, L. J.; Hatch, S. B.; Madden, S.; Müller, S.; Brennan, P. E.; Chowdhury, R.; Hopkinson, R. J.; Suga, H.; Schofield, C. J. *Nat. Commun.* **2017**, *8*, 14773.
(264) Jongkees, S. A. K.; Caner, S.; Tysoe, C.; Brayer, G. D.; Withers, S. G.; Suga, H. *Cell Chem. Biol.* **2017**, *24* (3), 381.
(265) Yu, H.; Dranchak, P.; Li, Z.; MacArthur, R.; Munson, M. S.; Mehzabeen, N.; Baird, N. J.; Battalie, K. P.; Ross, D.; Lovell, S.; Carlow, C. K. S.; Suga, H.; Inglese, J. *Nat. Commun.* **2017**, *8*, 14932.
(266) Kodan, A.; Yamaguchi, T.; Nakatsu, T.; Sakiyama, K.; Hipolito, C. J.; Fujioka, A.; Hirokane, R.; Ikeguchi, K.; Watanabe, B.; Hiratake, J.; Kimura, Y.; Suga, H.; Ueda, K.; Kato, H. *Proc. Natl. Acad. Sci.* **2014**, *111* (11), 4049.
(267) Song, X.; Lu, L.; Passioura, T.; Suga, H. *Org. Biomol. Chem.* **2017**, *15* (24), 5155.
(268) Passioura, T.; Watashi, K.; Fukano, K.; Shimura, S.; Saso, W.; Morishita, R.; Ogasawara, Y.; Tanaka, Y.; Mizokami, M.; Sureau, C.; Suga, H.; Wakita, T. *Cell Chem. Biol.* **2018**, *25* (7), 906.
(269) Scott, J.; Smith, G. *Science.* **1990**, *249* (4967), 386.
(270) Cwirla, S. E.; Peters, E. A.; Barrett, R. W.; Dower, W. J. *Proc. Natl. Acad. Sci.* **1990**, *87* (16), 6378.
(271) Felici, F.; Castagnoli, L.; Musacchio, A.; Jappelli, R.; Cesareni, G. *J. Mol. Biol.* **1991**, *222* (2), 301.
(272) O'Neil, K. T.; Hoess, R. H.; Jackson, S. A.; Ramachandran, N. S.; Mousa, S. A.; DeGrado, W. F. *Proteins Struct. Funct. Genet.* **1992**, *14* (4), 509.
(273) Pierschbacher, M. D.; Ruoslahti, E. *J. Biol. Chem.* **1987**, *262* (36), 17294.
(274) Koivunen, E.; Gay, D. A.; Ruoslahti, E. *J. Biol. Chem.* **1993**, *268* (27), 20205.
(275) Koivunen, E. *J. Cell Biol.* **1994**, *124* (3), 373.
(276) Krook, M.; Lindbladh, C.; Eriksen, J. A.; Mosbach, K. *Mol. Divers.* **1997**, *3* (3), 149.
(277) Koivunen, E.; Wang, B.; Ruoslahti, E. *Nat. Biotechnol.* **1995**, *13* (3), 265.
(278) Wang, B.; Dickinson, L. A.; Koivunen, E.; Ruoslahti, E.; Kohwi-Shigematsu, T. *J. Biol. Chem.* **1995**, *270* (40), 23239.
(279) Koivunen, E.; Ranta, T.-M.; Annila, A.; Taube, S.; Uppala, A.; Jokinen, M.; van Willigen, G.; Ihanus, E.; Gahmberg, C. G. *J. Cell Biol.* **2001**, *153* (5), 905.
(280) Hansen, M.; Wind, T.; Blouse, G. E.; Christensen, A.; Petersen, H. H.; Kjelgaard, S.; Mathiasen, L.; Holtet, T. L.; Andreasen, P. A. *J. Biol. Chem.* **2005**, *280* (46), 38424.
(281) Jensen, J. K.; Malmendal, A.; Schiøtt, B.; Skeldal, S.; Pedersen, K. E.; Celik, L.; Nielsen, N. C.; Andreasen, P. A.; Wind, T. *Biochem. J.* **2006**, *399* (3), 387.

(282) Andersen, L. M.; Wind, T.; Hansen, H. D.; Andreasen, P. A. *Biochem. J.* **2008**, *412* (3), 447.
(283) Pierce, H. H.; Adey, N.; Kay, B. K. *Mol. Divers.* **1996**, *1* (4), 259.
(284) Houimel, M.; Mazzucchelli, L. *Matrix Biol.* **2012**, *31* (1), 66.
(285) Oligino, L.; Lung, F.-D. T.; Sastry, L.; Bigelow, J.; Cao, T.; Curran, M.; Burke, T. R.; Wang, S.; Krag, D.; Roller, P. P.; King, C. R. *J. Biol. Chem.* **1997**, *272* (46), 29046.
(286) Lung, F.-D. T.; King, C. R.; Roller, P. P. *Lett. Pept. Sci.* **1999**, *6* (1), 45.
(287) Sem, D. S.; Baker, B. L.; Victoria, E. J.; Jones, D. S.; Marquis, D.; Yu, L.; Parks, J.; Coutts, S. M. *Biochemistry* **1998**, *37* (46), 16069.
(288) Gho, Y. S.; Lee, J. E.; Oh, K. S.; Bae, D. G.; Chae, C. B. *Cancer Res.* **1997**, *57* (17), 3733.
(289) Holig, P.; Bach, M.; Volkel, T.; Nahde, T.; Hoffmann, S.; Muller, R.; Kontermann, R. E. *Protein Eng. Des. Sel.* **2004**, *17* (5), 433.
(290) McLafferty, M. A.; Kent, R. B.; Ladner, R. C.; Markland, W. *Gene* **1993**, *128* (1), 29.
(291) Giebel, L. B.; Cass, R.; Milligan, D. L.; Young, D.; Arze, R.; Johnson, C. *Biochemistry* **1995**, *34* (47), 15430.
(292) Katz, B. A. *Biochemistry* **1995**, *34* (47), 15421.
(293) Katz, B. A.; Johnson, C.; Cass, R. T. *J. Am. Chem. Soc.* **1995**, *117* (33), 8541.
(294) Depraetere, H.; Viaene, A.; Deroo, S.; Vauterin, S.; Deckmyn, H. *Blood* **1998**, *92* (11), 4207.
(295) Sato, A. K.; Sexton, D. J.; Morganelli, L. A.; Cohen, E. H.; Wu, Q. L.; Conley, G. P.; Streltsova, Z.; Lee, S. W.; Devlin, M.; DeOliveira, D. B.; Enright, J.; Kent, R. B.; Wescott, C. R.; Ransohoff, T. C.; Ley, A. C.; Ladner, R. C. *Biotechnol. Prog.* **2002**, *18* (2), 182.
(296) Kelley, B. D.; Booth, J.; Tannatt, M.; Wu, Q.-L.; Ladner, R.; Yu, J.; Potter, D.; Ley, A. *J. Chromatogr. A* **2004**, *1038* (1–2), 121.
(297) Kelley, B. D.; Tannatt, M.; Magnusson, R.; Hagelberg, S.; Booth, J. *Biotechnol. Bioeng.* **2004**, *87* (3), 400.
(298) Kolodziej, A. F.; Nair, S. A.; Graham, P.; McMurry, T. J.; Ladner, R. C.; Wescott, C.; Sexton, D. J.; Caravan, P. *Bioconjug. Chem.* **2012**, *23* (3), 548.
(299) Han, Z.; Simpson, J. T.; Fivash, M. J.; Fisher, R.; Mori, T. *Peptides* **2004**, *25* (4), 551.
(300) Lehmann, A. *Expert Opin. Biol. Ther.* **2008**, *8* (8), 1187.
(301) Byun, C. H.; Park, J. Y.; Akamizu, T.; Chae, C.-B. *J. Clin. Endocrinol. Metab.* **2001**, *86* (7), 3311.
(302) Lee, L.; Buckley, C.; Blades, M. C.; Panayi, G.; George, A. J. T.; Pitzalis, C. *Arthritis Rheum.* **2002**, *46* (8), 2109.
(303) Kelly, K. A.; Jones, D. A. *Neoplasia* **2003**, *5* (5), 437.
(304) Ho, K. L.; Yusoff, K.; Seow, H. F.; Tan, W. S. *J. Med. Virol.* **2003**, *69* (1), 27.
(305) Amin, A.; Zaccardi, J.; Mullen, S.; Olland, S.; Orlowski, M.; Feld, B.; Labonte, P.; Mak, P. *Virology* **2003**, *313* (1), 158.
(306) Lupold, S. E.; Rodriguez, R. *Mol. Cancer Ther.* **2004**, *3* (5), 597.
(307) Nakamura, T.; Takasugi, H.; Aizawa, T.; Yoshida, M.; Mizuguchi, M.; Mori, Y.; Shinoda, H.; Hayakawa, Y.; Kawano, K. *J. Biotechnol.* **2005**, *116* (3), 211.
(308) Tan, W. S.; Tan, G. H.; Yusoff, K.; Seow, H. F. *J. Clin. Virol.* **2005**, *34* (1), 35.
(309) Muhamad, A.; Ho, K. L.; A. Rahman, M. B.; Uhrín, D.; Tan, W. S. *Chem. Biol. Drug Des.* **2013**, *81* (6), 784.
(310) Chen, H.; Su, X.; Neoh, K.-G.; Choe, W.-S. *Anal. Chem.* **2006**, *78* (14), 4872.

(311) Lee, T. C.; Yusoff, K.; Nathan, S.; Tan, W. S. *J. Virol. Methods* **2006**, *136* (1–2), 224.
(312) Petter, C.; Scholz, C.; Wessner, H.; Hansen, G.; Henklein, P.; Watanabe, T.; Höhne, W. *J. Mol. Recognit.* **2008**, *21* (6), 401.
(313) Rajik, M.; Jahanshiri, F.; Omar, A.; Ideris, A.; Hassan, S.; Yusoff, K. *Virol. J.* **2009**, *6* (1), 74.
(314) Vrielink, J.; Heins, M. S.; Setroikromo, R.; Szegezdi, E.; Mullally, M. M.; Samali, A.; Quax, W. J. *FEBS J.* **2010**, *277* (7), 1653.
(315) Larbanoix, L.; Burtea, C.; Laurent, S.; Van Leuven, F.; Toubeau, G.; Elst, L. Vander; Muller, R. N. *Neurobiol. Aging* **2010**, *31* (10), 1679.
(316) André, S.; Ansciaux, E.; Saidi, E.; Larbanoix, L.; Stanicki, D.; Nonclercq, D.; Vander Elst, L.; Laurent, S.; Muller, R. N.; Burtea, C. *J. Alzheimer's Dis.* **2017**, *60* (4), 1547.
(317) Morris, C. J.; Smith, M. W.; Griffiths, P. C.; McKeown, N. B.; Gumbleton, M. *J. Control. Release* **2011**, *151* (1), 83.
(318) Li, J.; Zhang, Q.; Pang, Z.; Wang, Y.; Liu, Q.; Guo, L.; Jiang, X. *Amino Acids* **2012**, *42* (6), 2373.
(319) Pandya, H.; Gibo, D. M.; Garg, S.; Kridel, S.; Debinski, W. *Neuro. Oncol.* **2012**, *14* (1), 6.
(320) Shiuan, D.; Chen, Y.-H.; Lin, H.-K.; Huang, K.-J.; Tai, D.-F.; Chang, D.-K. *Appl. Biochem. Biotechnol.* **2016**, *179* (4), 597.
(321) Rossez, Y.; Burtea, C.; Laurent, S.; Gosset, P.; Léonard, R.; Gonzalez, W.; Ballet, S.; Raynal, I.; Rousseaux, O.; Dugué, T.; Vander Elst, L.; Michalski, J.-C.; Muller, R. N.; Robbe-Masselot, C. *Contrast Media Mol. Imaging* **2016**, *11* (3), 211.
(322) Ghaffari Sharaf, M.; Cetinel, S.; Semenchenko, V.; Damji, K. F.; Unsworth, L. D.; Montemagno, C. *Exp. Eye Res.* **2017**, *165*, 109.
(323) Yamaguchi, S.; Ito, S.; Kurogi-Hirayama, M.; Ohtsuki, S. *J. Control. Release* **2017**, *262*, 232.
(324) Suire, C. N.; Nainar, S.; Fazio, M.; Kreutzer, A. G.; Paymozd-Yazdi, T.; Topper, C. L.; Thompson, C. R.; Leissring, M. A. *PLoS One* **2018**, *13* (2), e0193101.
(325) Jirka, S. M. G.; 't Hoen, P. A. C.; Diaz Parillas, V.; Tanganyika-de Winter, C. L.; Verheul, R. C.; Aguilera, B.; de Visser, P. C.; Aartsma-Rus, A. M. *Mol. Ther.* **2018**, *26* (1), 132.
(326) McConnell, S. J.; Kendall, M. L.; Reilly, T. M.; Hoess, R. H. *Gene* **1994**, *151* (1–2), 115.
(327) McConnell, S. I.; Uveges, A. J.; Fowlkes, D. M.; Spinella, D. G. *Mol. Divers.* **1996**, *1* (3), 165.
(328) Mazzucchelli, L.; Burritt, J. B.; Jesaitis, A. J.; Nusrat, A.; Liang, T. W.; Gewirtz, A. T.; Schnell, F. J.; Parkos, C. A. *Blood* **1999**, *93* (5), 1738.
(329) Valadon, P.; Nussbaum, G.; Boyd, L. F.; Margulies, D. H.; Scharff, M. D. *J. Mol. Biol.* **1996**, *261* (1), 11.
(330) Coulon, S.; Metais, J.-Y.; Chartier, M.; Briand, J.-P.; Baty, D. *J. Pept. Sci.* **2004**, *10* (11), 648.
(331) Kaji, M.; Ikari, M.; Hashiguchi, S.; Ito, Y.; Matsumoto, R.; Yoshimura, T.; Kuratsu, J. -i.; Sugimura, K. *J. Biochem.* **2001**, *129* (4), 577.
(332) Meta, A.; Torigoe, N.; Ito, Y.; Arakaki, R.; Nakashima, H.; Sugimura, K. *Mol. Immunol.* **1999**, *36* (18), 1249.
(333) Venkatesh, N.; Im, S.-H.; Balass, M.; Fuchs, S.; Katchalski-Katzir, E. *Proc. Natl. Acad. Sci.* **2000**, *97* (2), 761.
(334) Jung, H. H.; Yi, H. J.; Lee, S. K.; Lee, J. Y.; Jung, H. J.; Yang, S. T.; Eu, Y.-J.; Im, S.-H.; Kim, J. Il. *Biochemistry* **2007**, *46* (51), 14987.

(335) Sabo, J. K.; Keizer, D. W.; Feng, Z.-P.; Casey, J. L.; Parisi, K.; Coley, A. M.; Foley, M.; Norton, R. S. *Infect. Immun.* **2007**, *75* (1), 61.
(336) Sahu, A.; Kay, B. K.; Lambris, J. D. *J. Immunol.* **1996**, *157* (2), 884.
(337) Scholle, M. D.; Banach, B. S.; Hamdan, S. M.; Richardson, C. C.; Kay, B. K. *Biochim. Biophys. Acta - Proteins Proteomics* **2008**, *1784* (11), 1735.
(338) Cendron, A. C.; Wines, B. D.; Brownlee, R. T. C.; Ramsland, P. A.; Pietersz, G. A.; Hogarth, P. M. *Mol. Immunol.* **2008**, *45* (2), 307.
(339) Drulis-Fajdasz, D.; Jelen, F.; Oleksy, A.; Otlewski, J. *Acta Biochim. Pol.* **2006**, *53* (3), 515.
(340) Rao, W. H.; Camp, R. D. . *Int. Immunopharmacol.* **2003**, *3* (3), 435.
(341) McConnell, S. J.; Dinh, T.; Le, M.-H.; Brown, S. J.; Becherer, K.; Blumeyer, K.; Kautzer, C.; Axelrod, F.; Spinella, D. G. *Biol. Chem.* **1998**, *379* (10).
(342) Pero, S. C.; Oligino, L.; Daly, R. J.; Soden, A. L.; Liu, C.; Roller, P. P.; Li, P.; Krag, D. N. *J. Biol. Chem.* **2002**, *277* (14), 11918.
(343) Porter, C. J.; Matthews, J. M.; Mackay, J. P.; Pursglove, S. E.; Schmidberger, J. W.; Leedman, P. J.; Pero, S. C.; Krag, D. N.; Wilce, M. C.; Wilce, J. A. *BMC Struct. Biol.* **2007**, *7* (1), 58.
(344) Watson, G. M.; Gunzburg, M. J.; Ambaye, N. D.; Lucas, W. A. H.; Traore, D. A.; Kulkarni, K.; Cergol, K. M.; Payne, R. J.; Panjikar, S.; Pero, S. C.; Perlmutter, P.; Wilce, M. C. J.; Wilce, J. A. *J. Med. Chem.* **2015**, *58* (19), 7707.
(345) Watson, G. M.; Kulkarni, K.; Sang, J.; Ma, X.; Gunzburg, M. J.; Perlmutter, P.; Wilce, M. C. J.; Wilce, J. A. *J. Med. Chem.* **2017**, *60* (22), 9349.
(346) Heiskanen, T.; Lundkvist, Å.; Soliymani, R.; Koivunen, E.; Vaheri, A.; Lankinen, H. *Virology* **1999**, *262* (2), 321.
(347) Rainey, M. A.; Linse, K. D.; Dalby, K. N. *J. Phys. Org. Chem.* **2004**, *17* (67), 461.
(348) Ranganath, S.; Bhandari, A.; Avitahl-Curtis, N.; McMahon, J.; Wachtel, D.; Zhang, J.; Leitheiser, C.; Bernier, S. G.; Liu, G.; Tran, T. T.; Celino, H.; Tobin, J.; Jung, J.; Zhao, H.; Glen, K. E.; Graul, C.; Griffin, A.; Schairer, W. C.; Higgins, C.; Reza, T. L.; Mowe, E.; Rivers, S.; Scott, S.; Monreal, A.; Shea, C.; Bourne, G.; Coons, C.; Smith, A.; Tang, K.; Mandyam, R. A.; Masferrer, J.; Liu, D.; Patel, D. V.; Fretzen, A.; Murphy, C. A.; Milne, G. T.; Smythe, M. L.; Carlson, K. E. *PLoS One* **2015**, *10* (11), e0141330.
(349) Li, B.; Tom, J. Y.; Oare, D.; Yen, R.; Fairbrother, W. J.; Wells, J. A.; Cunningham, B. C. *Science* **1995**, *270* (5242), 1657.
(350) Schoenfeld, J. R.; Tom, J. Y. K.; Lowe, D. G. *FEBS Lett.* **1997**, *414* (2), 263.
(351) Souriau, C.; Chiche, L.; Irving, R.; Hudson, P. *Biochemistry* **2005**, *44* (19), 7143.
(352) Zhao, R.; Dai, H.; Mendelman, N.; Cuello, L. G.; Chill, J. H.; Goldstein, S. A. N. *Proc. Natl. Acad. Sci.* **2015**, *112* (50), E7013.
(353) Cano, A.; Viveros, M.; Acero, G.; Govezensky, T.; Munguia, M. E.; Gonzalez, E.; Soto, L.; Gevorkian, G.; Manoutcharian, K. *Immunol. Lett.* **2004**, *95* (2), 207.
(354) Rousch, M.; Lutgerink, J. T.; Coote, J.; Bruïne, A.; Arends, J.-W.; Hoogenboom, H. R. *Br. J. Pharmacol.* **1998**, *125* (1), 5.
(355) Ferguson, M. R. *Protein Sci.* **2004**, *13* (3), 626.
(356) Chen, S.; Rentero Rebollo, I.; Buth, S. A.; Morales-Sanfrutos, J.; Touati, J.; Leiman, P. G.; Heinis, C. *J. Am. Chem. Soc.* **2013**, *135* (17), 6562.
(357) Zha, M.; Lin, P.; Yao, H.; Zhao, Y.; Wu, C. *Chem. Commun.* **2018**, *54* (32), 4029.
(358) Chen, S.; Gopalakrishnan, R.; Schaer, T.; Marger, F.; Hovius, R.; Bertrand, D.; Pojer, F.;

Heinis, C. *Nat. Chem.* **2014**, *6* (11), 1009.
(359) Chen, S.; Heinis, C. *Methods Mol. Biol.* **2015**, *1248*, 119.
(360) Luzzago, A.; Felici, F.; Tramontano, A.; Pessi, A.; Cortese, R. *Gene* **1993**, *128* (1), 51.
(361) Rojas, G.; Pupo, A.; Aleman, M. D. R.; Vispo, N. S. *J. Pept. Sci.* **2008**, *14* (11), 1216.
(362) Wrighton, N. C.; Farrell, F. X.; Chang, R.; Kashyap, A. K.; Barbone, F. P.; Mulcahy, L. S.; Johnson, D. L.; Barrett, R. W.; Jolliffe, L. K.; Dower, W. J. *Science* **1996**, *273* (5274), 458.
(363) Hart, S. L.; Knight, A. M.; Harbottle, R. P.; Mistry, A.; Hunger, H. D.; Cutler, D. F.; Williamson, R.; Coutelle, C. *J. Biol. Chem.* **1994**, *269* (17), 12468.
(364) Horwacik, I.; Czaplicki, D.; Talarek, K.; Kowalczyk, A.; Bolesta, E.; Kozbor, D.; Rokita, H. *Int. J. Mol. Med.* **2007**, *19*, 829.
(365) Lowman, H. B.; Chen, Y. M.; Skelton, N. J.; Mortensen, D. L.; Tomlinson, E. E.; Sadick, M. D.; Robinson, I. C. A. F.; Clark, R. G. *Biochemistry* **1998**, *37* (25), 8870.
(366) Tam, E. M.; Runyon, S. T.; Santell, L.; Quan, C.; Yao, X.; Kirchhofer, D.; Skelton, N. J.; Lazarus, R. A. *J. Mol. Biol.* **2009**, *385* (1), 79.
(367) Fairbrother, W. J.; Christinger, H. W.; Cochran, A. G.; Fuh, G.; Keenan, C. J.; Quan, C.; Shriver, S. K.; Tom, J. Y. K.; Wells, J. A.; Cunningham, B. C. *Biochemistry* **1998**, *37* (51), 17754.
(368) Pan, B.; Li, B.; Russell, S. J.; Tom, J. Y. .; Cochran, A. G.; Fairbrother, W. J. *J. Mol. Biol.* **2002**, *316* (3), 769.
(369) Pan, B.; Fairbrother, W. J. *Spectroscopy* **2003**, *17* (2–3), 169.
(370) Lichtfouse, B.; Mertens, P.; Clavareau, C.; Godfroid, F.; Weynants, V.; Denoel, P.; Tibor, A.; Letesson, J.-J. *Lett. Pept. Sci.* **1995**, *2* (3–4), 261.
(371) Arita, Y.; Allen, S.; Chen, G.; Zhang, W.; Wang, Y.; Owen, A. J.; Dentinger, P.; Sidhu, S. S. *Biochem. Biophys. Res. Commun.* **2016**, *473* (2), 370.
(372) Deshayes, K.; Schaffer, M. L.; Skelton, N. J.; Nakamura, G. R.; Kadkhodayan, S.; Sidhu, S. S. *Chem. Biol.* **2002**, *9* (4), 495.
(373) Mertens, P.; Matheise, J.-P.; Lichtfouse, B.; Clavareau, C.; Letesson, J.-J. *Lett. Pept. Sci.* **1995**, *2* (3–4), 220.
(374) Fairlie, W. D.; Spurck, T. P.; McCoubrie, J. E.; Gilson, P. R.; Miller, S. K.; McFadden, G. I.; Malby, R.; Crabb, B. S.; Hodder, A. N. *Infect. Immun.* **2008**, *76* (9), 4332.
(375) Zwick, M. B.; Shen, J.; Scott, J. K. *J. Mol. Biol.* **2000**, *300* (2), 307.
(376) Timmerman, P.; Beld, J.; Puijk, W. C.; Meloen, R. H. *ChemBioChem* **2005**, *6* (5), 821.
(377) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. *Nat. Chem. Biol.* **2009**, *5* (7), 502.
(378) Pollaro, L.; Heinis, C. *Medchemcomm* **2010**, *1* (5), 319.
(379) Angelini, A.; Morales-Sanfrutos, J.; Diderich, P.; Chen, S.; Heinis, C. *J. Med. Chem.* **2012**, *55* (22), 10187.
(380) Pollaro, L.; Raghunathan, S.; Morales-Sanfrutos, J.; Angelini, A.; Kontos, S.; Heinis, C. *Mol. Cancer Ther.* **2015**, *14* (1), 151.
(381) Zorzi, A.; Middendorp, S. J.; Wilbs, J.; Deyle, K.; Heinis, C. *Nat. Commun.* **2017**, *8*, 16092.
(382) Angelini, A.; Diderich, P.; Morales-Sanfrutos, J.; Thurnheer, S.; Hacker, D.; Menin, L.; Heinis, C. *Bioconjug. Chem.* **2012**, *23* (9), 1856.
(383) Baeriswyl, V.; Heinis, C. *Protein Eng. Des. Sel.* **2013**, *26* (1), 81.
(384) Angelini, A.; Cendron, L.; Chen, S.; Touati, J.; Winter, G.; Zanotti, G.; Heinis, C. *ACS Chem. Biol.* **2012**, *7* (5), 817.

(385) Chen, S.; Gfeller, D.; Buth, S. A.; Michielin, O.; Leiman, P. G.; Heinis, C. *ChemBioChem* **2013**, *14* (11), 1316.
(386) Baeriswyl, V.; Rapley, H.; Pollaro, L.; Stace, C.; Teufel, D.; Walker, E.; Chen, S.; Winter, G.; Tite, J.; Heinis, C. *ChemMedChem* **2012**, *7* (7), 1173.
(387) Pollaro, L.; Diderich, P.; Angelini, A.; Bellotto, S.; Wegner, H.; Heinis, C. *Anal. Biochem.* **2012**, *427* (1), 18.
(388) Rebollo, I. R.; Angelini, A.; Heinis, C. *Med. Chem. Commun.* **2013**, *4* (1), 145.
(389) Baeriswyl, V.; Calzavarini, S.; Gerschheimer, C.; Diderich, P.; Angelillo-Scherrer, A.; Heinis, C. *J. Med. Chem.* **2013**, *56* (9), 3742.
(390) Diderich, P.; Heinis, C. *Tetrahedron* **2014**, *70* (42), 7733.
(391) Urech-Varenne, C.; Radtke, F.; Heinis, C. *ChemMedChem* **2015**, *10* (10), 1754.
(392) Rentero Rebollo, I.; McCallin, S.; Bertoldo, D.; Entenza, J. M.; Moreillon, P.; Heinis, C. *ACS Med. Chem. Lett.* **2016**, *7* (6), 606.
(393) Mund, T.; Lewis, M. J.; Maslen, S.; Pelham, H. R. *Proc. Natl. Acad. Sci.* **2014**, *111* (47), 16736.
(394) Rentero Rebollo, I.; Sabisz, M.; Baeriswyl, V.; Heinis, C. *Nucleic Acids Res.* **2014**, *42* (22), e169.
(395) Rentero Rebollo, I.; Heinis, C. *Methods* **2013**, *60* (1), 46.
(396) Chen, S.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C. *Angew. Chemie Int. Ed.* **2014**, *53* (6), 1602.
(397) Chen, S.; Morales-Sanfrutos, J.; Angelini, A.; Cutting, B.; Heinis, C. *ChemBioChem* **2012**, *13* (7), 1032.
(398) Baeriswyl, V.; Calzavarini, S.; Chen, S.; Zorzi, A.; Bologna, L.; Angelillo-Scherrer, A.; Heinis, C. *ACS Chem. Biol.* **2015**, *10* (8), 1861.
(399) Middendorp, S. J.; Wilbs, J.; Quarroz, C.; Calzavarini, S.; Angelillo-Scherrer, A.; Heinis, C. *J. Med. Chem.* **2017**, *60* (3), 1151.
(400) Wilbs, J.; Middendorp, S. J.; Heinis, C. *ChemBioChem* **2016**, *17* (24), 2299.
(401) Bertoldo, D.; Khan, M. M. G.; Dessen, P.; Held, W.; Huelsken, J.; Heinis, C. *ChemMedChem* **2016**, *11* (8), 834.
(402) van de Langemheen, H.; Korotkovs, V.; Bijl, J.; Wilson, C.; Kale, S. S.; Heinis, C.; Liskamp, R. M. J. *ChemBioChem* **2017**, *18* (4), 387.
(403) Kale, S. S.; Villequey, C.; Kong, X.-D.; Zorzi, A.; Deyle, K.; Heinis, C. *Nat. Chem.* **2018**, *10* (7), 715.
(404) Bellotto, S.; Chen, S.; Rentero Rebollo, I.; Wegner, H. A.; Heinis, C. *J. Am. Chem. Soc.* **2014**, *136* (16), 5880.
(405) Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R. *ACS Chem. Biol.* **2014**, *9* (2), 443.
(406) Ng, S.; Derda, R. *Org. Biomol. Chem.* **2016**, *14* (24), 5539.
(407) Kalhor-Monfared, S.; Jafari, M. R.; Patterson, J. T.; Kitov, P. I.; Dwyer, J. J.; Nuss, J. M.; Derda, R. *Chem. Sci.* **2016**, *7* (6), 3785.
(408) Urban, J. H.; Moosmeier, M. A.; Aumüller, T.; Thein, M.; Bosma, T.; Rink, R.; Groth, K.; Zulley, M.; Siegers, K.; Tissot, K.; Moll, G. N.; Prassler, J. *Nat. Commun.* **2017**, *8* (1), 1500.
(409) Matochko, W. L.; Chu, K.; Jin, B.; Lee, S. W.; Whitesides, G. M.; Derda, R. *Methods* **2012**, *58* (1), 47.
(410) Ng, S.; Lin, E.; Kitov, P. I.; Tjhung, K. F.; Gerlits, O. O.; Deng, L.; Kasper, B.; Sood, A.;

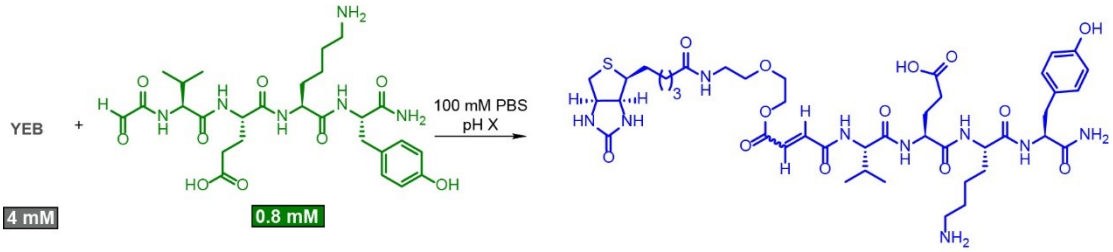
Paschal, B. M.; Zhang, P.; Ling, C.-C.; Klassen, J. S.; Noren, C. J.; Mahal, L. K.; Woods, R. J.; Coates, L.; Derda, R. *J. Am. Chem. Soc.* **2015**, *137* (16), 5248.
(411) He, B.; Tjhung, K. F.; Bennett, N. J.; Chou, Y.; Rau, A.; Huang, J.; Derda, R. *Sci. Rep.* **2018**, *8* (1), 1214.
(412) 't Hoen, P. A. C.; Jirka, S. M. G.; ten Broeke, B. R.; Schultes, E. A.; Aguilera, B.; Pang, K. H.; Heemskerk, H.; Aartsma-Rus, A.; van Ommen, G. J.; den Dunnen, J. T. *Anal. Biochem.* **2012**, *421* (2), 622.
(413) Villequey, C.; Kong, X.-D.; Heinis, C. *Protein Eng. Des. Sel.* **2017**, *30* (11), 761.
(414) Tjhung, K. F.; Kitov, P. I.; Ng, S.; Kitova, E. N.; Deng, L.; Klassen, J. S.; Derda, R. *J. Am. Chem. Soc.* **2016**, *138* (1), 32.
(415) Tian, F.; Tsao, M.-L.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126* (49), 15962.
(416) Lim, R. K. V.; Li, N.; Ramil, C. P.; Lin, Q. *ACS Chem. Biol.* **2014**, *9* (9), 2139.
(417) Galloway, W. R. J. D.; Isidro-Llobet, A.; Spring, D. R. *Nat. Commun.* **2010**, *1* (6), 1.
(418) Heinis, C.; Winter, G. *Curr. Opin. Chem. Biol.* **2015**, *26*, 89.
(419) Li, S.; Millward, S.; Roberts, R. *J. Am. Chem. Soc.* **2002**, *124* (34), 9972.
(420) Takahashi, T. T.; Austin, R. J.; Roberts, R. W. *Trends Biochem. Sci.* **2003**, *28* (3), 159.
(421) Li, S.; Roberts, R. W. *Chem. Biol.* **2003**, *10* (3), 233.
(422) Iwasaki, K.; Goto, Y.; Katoh, T.; Suga, H. *Org. Biomol. Chem.* **2012**, *10* (30), 5783.
(423) Kalhor-Monfared, S.; Jafari, M. R.; Patterson, J. T.; Kitov, P. I.; Dwyer, J. J.; Nuss, J. M.; Derda, R. *Chem. Sci.* **2016**, *7* (6), 3785.
(424) Jafari, M. R.; Lakusta, J.; Lundgren, R. J.; Derda, R. *Bioconjug. Chem.* **2016**, *27* (3), 509.
(425) Sandman, K. E. *Nucleic Acids Res.* **2003**, *31* (8), 2234.
(426) Beech, J.; Saleh, L.; Frentzel, J.; Figler, H.; Corrêa, I. R.; Baker, B.; Ramspacher, C.; Marshall, M.; Dasa, S.; Linden, J.; Noren, C. J.; Kelly, K. A. *Bioconjug. Chem.* **2015**, *26* (3), 529.
(427) Ng, S.; Jafari, M. R.; Matochko, W. L.; Derda, R. *ACS Chem. Biol.* **2012**, *7* (9), 1482.
(428) Kitov, P. I.; Vinals, D. F.; Ng, S.; Tjhung, K. F.; Derda, R. *J. Am. Chem. Soc.* **2014**, *136* (23), 8149.
(429) Spears, R. J.; Fascione, M. A. *Org. Biomol. Chem.* **2016**, *14* (32), 7622.
(430) Boutureira, O.; Bernardes, G. J. L. *Chem. Rev.* **2015**, *115* (5), 2174.
(431) Ramil, C. P.; Lin, Q. *Chem. Commun.* **2013**, *49* (94), 11007.
(432) Spicer, C. D.; Davis, B. G. *Nat. Commun.* **2014**, *5*, 4740.
(433) Lum, K. M.; Xavier, V. J.; Ong, M. J.-H.; Johannes, C. W.; Chan, K.-P. *Chem. Commun.* **2013**, *49* (95), 11188.
(434) Han, M.-J.; Xiong, D.-C.; Ye, X.-S. *Chem. Commun.* **2012**, *48* (90), 11079.
(435) McLeod, D.; McNulty, J. *European J. Org. Chem.* **2012**, *2012* (31), 6127.
(436) Gaertner, H. F.; Rose, K.; Cotton, R.; Timms, D.; Camble, R.; Offord, R. E. *Bioconjug. Chem.* **1992**, *3* (3), 262.
(437) Vilaseca, L. A.; Rose, K.; Werlen, R.; Meunier, A.; Offord, R. E.; Nichols, C. L.; Scott, W. L. *Bioconjug. Chem.* **1993**, *4* (6), 515.
(438) Agten, S. M.; Dawson, P. E.; Hackeng, T. M. *J. Pept. Sci.* **2016**, *22* (5), 271.
(439) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci.* **2003**, *100* (1), 56.
(440) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Biol.* **2007**, *3* (6), 321.
(441) Krishnan, S.; Miller, R. M.; Tian, B.; Mullins, R. D.; Jacobson, M. P.; Taunton, J. *J. Am. Chem. Soc.* **2014**, *136* (36), 12624.
(442) Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.;

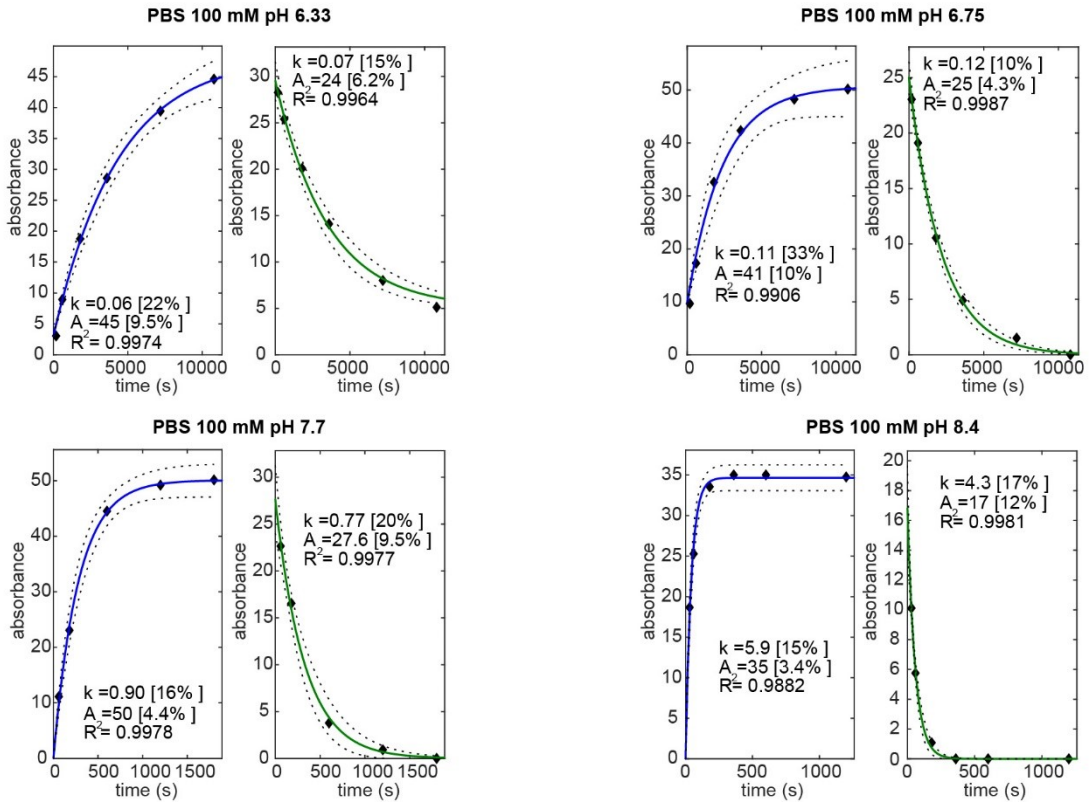
McFarland, J. M.; Miller, R. M.; Frödin, M.; Taunton, J. *Nat. Chem. Biol.* **2012**, *8* (5), 471.
(443) Miller, R. M.; Paavilainen, V. O.; Krishnan, S.; Serafimova, I. M.; Taunton, J. *J. Am. Chem. Soc.* **2013**, *135* (14), 5298.
(444) Gattner, M. J.; Ehrlich, M.; Vrabel, M. *Chem. Commun.* **2014**, *50* (83), 12568.
(445) Knall, A.-C.; Slugovc, C. *Chem. Soc. Rev.* **2013**, *42* (12), 5131.
(446) Robiette, R.; Richardson, J.; Aggarwal, V. K.; Harvey, J. N. *J. Am. Chem. Soc.* **2005**, *127* (39), 13468.
(447) Tronchet, J. M. J.; Gentile, B. *Helv. Chim. Acta* **1979**, *62* (7), 2091.
(448) Dambacher, J.; Zhao, W.; El-Batta, A.; Anness, R.; Jiang, C.; Bergdahl, M. *Tetrahedron Lett.* **2005**, *46* (26), 4473.
(449) Tiwari, S.; Kumar, A. *Chem. Commun.* **2008**, No. 37, 4445.
(450) McNulty, J.; Das, P. *Tetrahedron Lett.* **2009**, *50* (41), 5737.
(451) Das, P.; McNulty, J. *Tetrahedron Lett.* **2010**, *51* (24), 3197.
(452) Aksnes, G.; Songstad, J. *Acta Chem. Scand.* **1964**, *18*, 655.
(453) Zhang, X. M.; Bordwell, F. G. *J. Am. Chem. Soc.* **1994**, *116* (3), 968.
(454) El-Batta, A.; Jiang, C.; Zhao, W.; Anness, R.; Cooksy, A. L.; Bergdahl, M. *J. Org. Chem.* **2007**, *72* (14), 5244.
(455) Byrne, P. A.; Gilheany, D. G. *Chem. Soc. Rev.* **2013**, *42* (16), 6670.
(456) Ayub, K.; Ludwig, R. *RSC Adv.* **2016**, *6* (28), 23448.
(457) Robiette, R.; Richardson, J.; Aggarwal, V. K.; Harvey, J. N. *J. Am. Chem. Soc.* **2006**, *128* (7), 2394.
(458) Vedejs, E.; Fleck, T. J. *J. Am. Chem. Soc.* **1989**, *111* (15), 5861.
(459) Vedejs, E.; Fleck, T.; Hara, S. *J. Org. Chem.* **1987**, *52* (20), 4637.
(460) Byrne, P. A.; Ortin, Y.; G. Gilheany, D. *Chem. Commun.* **2015**, *51* (6), 1147.
(461) Kohyama, A.; Fukuda, M.; Sugiyama, S.; Yamakoshi, H.; Kanoh, N.; Ishioka, C.; Shibata, H.; Iwabuchi, Y. *Org. Biomol. Chem.* **2016**, *14* (45), 10683.
(462) Avonto, C.; Taglialatela-Scafati, O.; Pollastro, F.; Minassi, A.; Di Marzo, V.; De Petrocellis, L.; Appendino, G. *Angew. Chemie Int. Ed.* **2011**, *50* (2), 467.
(463) Bernardim, B.; Cal, P. M. S. D.; Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Albuquerque, I. S.; Perkins, E.; Corzana, F.; Burtoloso, A. C. B.; Jiménez-Osés, G.; Bernardes, G. J. L. *Nat. Commun.* **2016**, *7*, 13128.
(464) Rideout, D. C.; Breslow, R. *J. Am. Chem. Soc.* **1980**, *102* (26), 7816.
(465) Otto, S.; Engberts, J. B. F. N. *Tetrahedron Lett.* **1995**, *36* (15), 2645.
(466) Hwang, T. L.; Shaka, A. J. *J. Magn. Reson. Ser. A* **1995**, *112* (2), 275.
(467) Ng, S.; Tjhung, K. F.; Paschal, B. M.; Noren, C. J.; Derda, R. 2015; pp 155–172.
(468) Moffett, R. B. *Org. Synth.* **1952**, *32*, 41.
(469) Reetz, M. T. *Angew. Chemie Int. Ed.* **2002**, *41* (8), 1335.
(470) Hammett, L. P. *Chem. Rev.* **1935**, *17* (1), 125.
(471) Wang, T.; Yuan, X.; Wu, M.-B.; Lin, J.-P.; Yang, L.-R. *Expert Opin. Drug Discov.* **2017**, 1.
(472) Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. *Chem. Sci.* **2018**, *9* (9), 2398.
(473) Kutchukian, P. S.; Dropinski, J. F.; Dykstra, K. D.; Li, B.; DiRocco, D. A.; Streckfuss, E. C.; Campeau, L.-C.; Cernak, T.; Vachal, P.; Davies, I. W.; Krska, S. W.; Dreher, S. D. *Chem. Sci.* **2016**, *7* (4), 2604.
(474) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, *24* (4), 279.

(475) Harper, K. C.; Bess, E. N.; Sigman, M. S. *Nat. Chem.* **2012**, *4* (5), 366.
(476) Collins, K. D.; Gensch, T.; Glorius, F. *Nat. Chem.* **2014**, *6* (10), 859.
(477) Selekman, J. A.; Qiu, J.; Tran, K.; Stevens, J.; Rosso, V.; Simmons, E.; Xiao, Y.; Janey, J. *Annu. Rev. Chem. Biomol. Eng.* **2017**, *8* (1), 525.
(478) Reetz, M. T. *Angew. Chemie Int. Ed.* **2001**, *40* (2), 284.
(479) Reetz, M. T. *Angew. Chemie Int. Ed.* **2008**, *47* (14), 2556.
(480) Gasparini, G.; Dal Molin, M.; Prins, L. J. *European J. Org. Chem.* **2010**, *2010* (13), 2429.
(481) Devore, D. D.; Jenkins, R. M. *Comments Inorg. Chem.* **2014**, *34* (1–2), 17.
(482) Robbins, D. W.; Hartwig, J. F. *Science.* **2011**, *333* (6048), 1423.
(483) Maeda, Y.; Makhlynets, O. V.; Matsui, H.; Korendovych, I. V. *Annu. Rev. Biomed. Eng.* **2016**, *18* (1), 311.
(484) Kanan, M. W.; Rozenman, M. M.; Sakurai, K.; Snyder, T. M.; Liu, D. R. *Nature* **2004**, *431* (7008), 545.
(485) Rozenman, M. M.; Kanan, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2007**, *129* (48), 14933.
(486) Chen, Y.; Kamlet, A. S.; Steinman, J. B.; Liu, D. R. *Nat. Chem.* **2011**, *3* (2), 146.
(487) Hook, K. D.; Chambers, J. T.; Hili, R. *Chem. Sci.* **2017**, *8*, 7072.
(488) Chan, A. I.; McGregor, L. M.; Liu, D. R. *Curr. Opin. Chem. Biol.* **2015**, *26*, 55.
(489) Tanaka, F.; Fuller, R.; Shim, H.; Lerner, R. A.; Barbas, C. F. *J. Mol. Biol.* **2004**, *335* (4), 1007.
(490) Tanaka, F.; Barbas III, C. F. *Chem. Commun.* **2001**, No. 8, 769.
(491) Tanaka, F.; Fuller, R.; Asawapornmongkol, L.; Warsinke, A.; Gobuty, S.; Barbas III, C. F. *Bioconjug. Chem.* **2007**, *18* (4), 1318.
(492) Eldridge, G. M.; Weiss, G. A. *Bioconjug. Chem.* **2011**, *22* (10), 2143.
(493) Eldridge, G. M.; Weiss, G. A. *Methods Mol. Biol.* **2015**, 189.
(494) Ng, S.; Jafari, M. R.; Derda, R. *ACS Chem. Biol.* **2012**, *7* (1), 123.
(495) Kehoe, J. W.; Kay, B. K. *Chem. Rev.* **2005**, *105* (11), 4056.
(496) Angelini, A.; Heinis, C. *Curr. Opin. Chem. Biol.* **2011**, *15* (3), 355.
(497) Triana, V.; Derda, R. *Org. Biomol. Chem.* **2017**, *15* (37), 7869.
(498) Rose, K.; Chen, J.; Dragovic, M.; Zeng, W.; Jeannerat, D.; Kamalaprija, P.; Burger, U. *Bioconjug. Chem.* **1999**, *10* (6), 1038.
(499) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. *Nat. Chem.* **2015**, *8*, 120.
(500) Carrico, Z. M.; Farkas, M. E.; Zhou, Y.; Hsiao, S. C.; Marks, J. D.; Chokhawala, H.; Clark, D. S.; Francis, M. B. *ACS Nano* **2012**, *6* (8), 6675.
(501) Tian, F.; Tsao, M.-L.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126* (49), 15962.
(502) Urquhart, T.; Daub, E.; Honek, J. F. *Bioconjug. Chem.* **2016**, *27* (10), 2276.
(503) Pirrung, M. C.; Sarma, K. Das. *J. Am. Chem. Soc.* **2004**, *126* (2), 444.
(504) Goodnow, R. A.; Dumelin, C. E.; Keefe, A. D. *Nat. Rev. Drug Discov.* **2017**, *16* (2), 131.
(505) Matochko, W. L.; Cory Li, S.; Tang, S. K. Y.; Derda, R. *Nucleic Acids Res.* **2014**, *42* (3), 1784.
(506) Giordanetto, F.; Kihlberg, J. *J. Med. Chem.* **2014**, *57* (2), 278.
(507) Tsomaia, N. *Eur. J. Med. Chem.* **2015**, *94*, 459.
(508) Thompson, D. A.; Ng, R.; Dawson, P. E. *J. Pept. Sci.* **2016**, *22* (5), 311.
(509) Glomb, M. A.; Lang, G. *J. Agric. Food Chem.* **2001**, *49* (3), 1493.
(510) Takahashi, K. *J. Biol. Chem.* **1968**, *243* (23), 6171.
(511) Kim, B.; Lee, H.-G.; Kang, S.-B.; Sung, G.; Kim, J.-J.; Park, J.; Lee, S.-G.; Yoon, Y.-J.

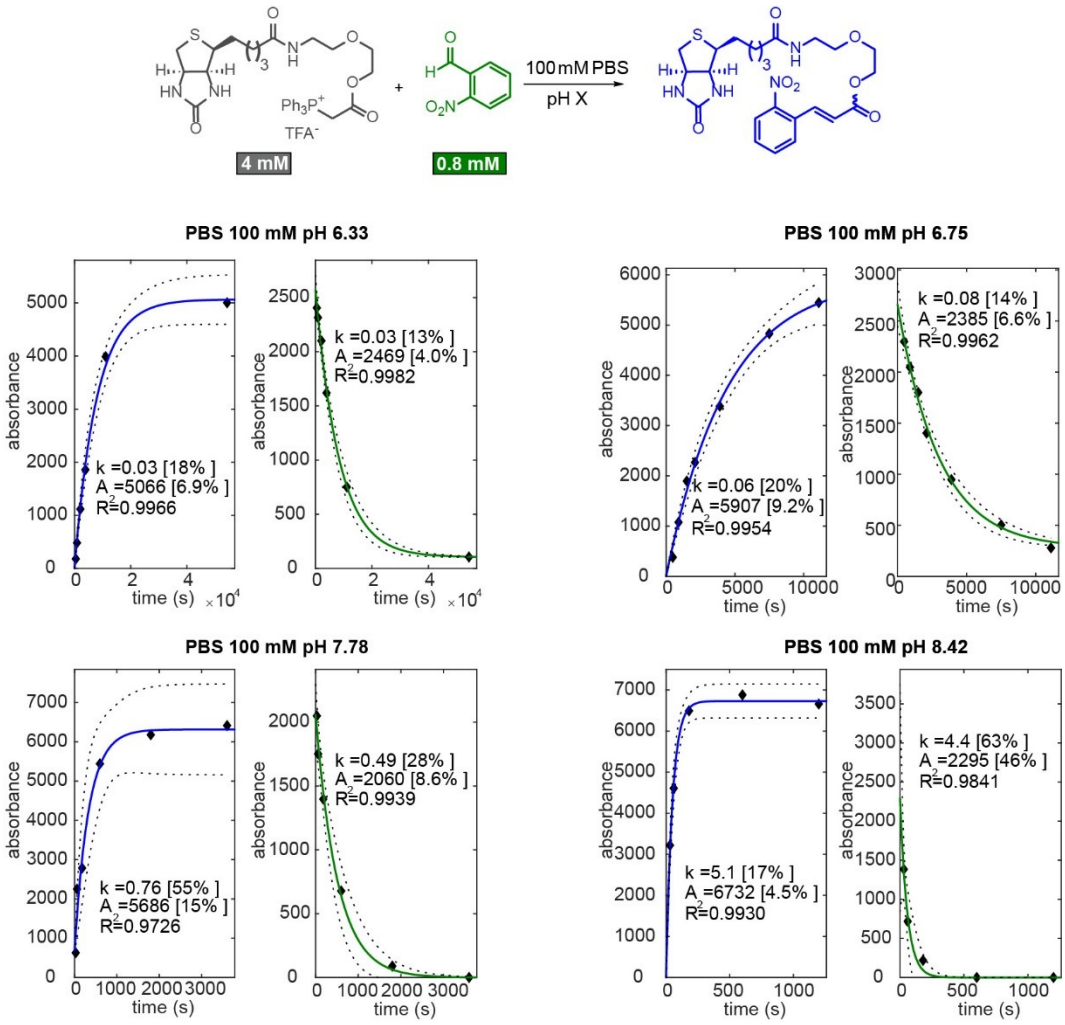
Synthesis (Stuttg). **2012**, *44* (1), 42.
(512) Appel, R.; Loos, R.; Mayr, H. *J. Am. Chem. Soc.* **2009**, *131* (2), 704.

Appendix A: Supporting information for Chapter 2

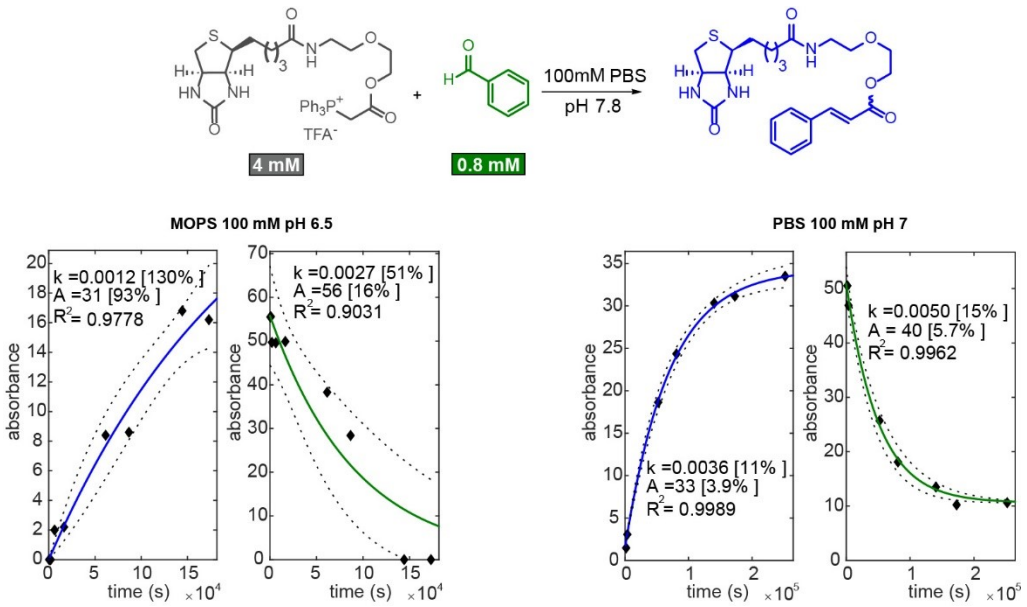




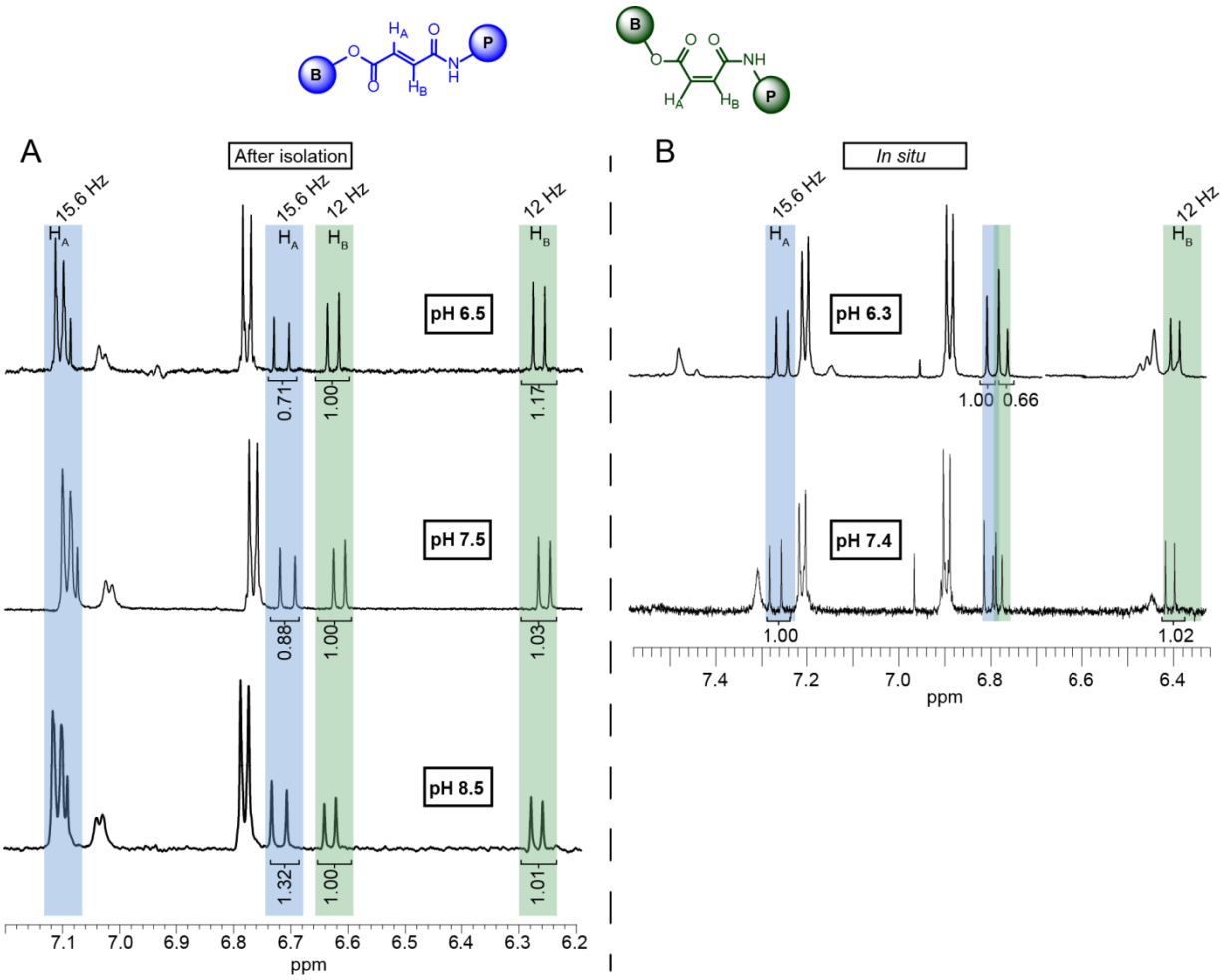
Appendix A-1. Kinetic traces for Wittig reaction on model peptide o-VEKY at different pH values



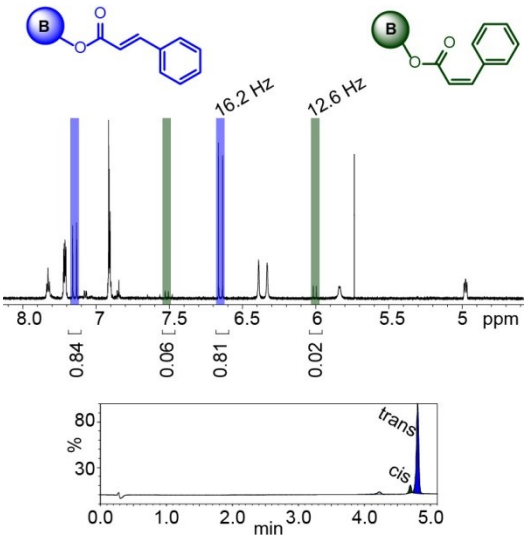
Appendix A-2. Kinetic traces for Wittig reaction on 2-nitrobenzaldehyde at different pH values



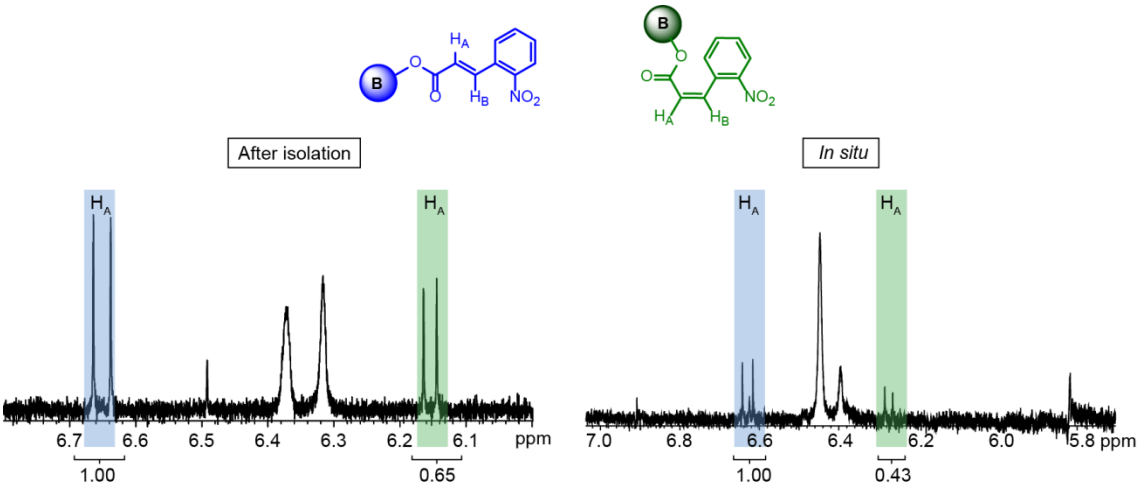
Appendix A-3 Kinetic traces for Wittig on benzaldehyde at different pH values. Acquisition of kinetic traces at pH > 7 was complicated due to the rate of ester hydrolysis being equal or higher than the rate of the Wittig reaction.



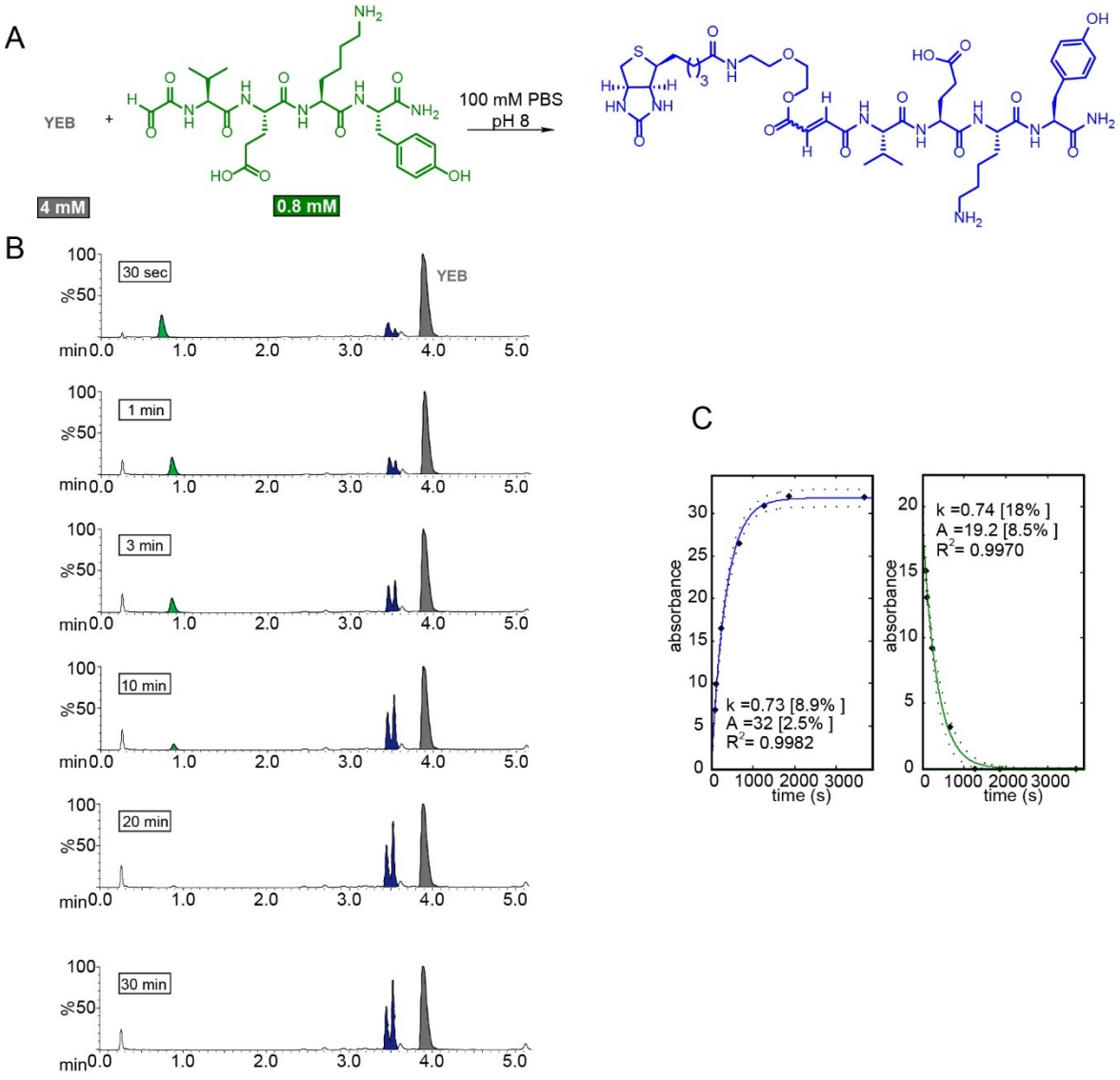
Appendix A-4 E/Z ratio for Wittig reaction with o-VEKY model peptide at different pH values. A) For measurements performed after isolation of Wittig product, the reactions were run at the indicated pH values. The Wittig adduct was subsequently purified by HPLC, dissolved in D_2O and the E/Z ratio was measured by integration of ^1H NMR spectrum. B) To confirm that the work-up and purification were not affecting E/Z ratios,[512] the reactions were run again in buffers at the pH values shown. At the endpoint of the reaction, 10% D_2O was added and the crude E/Z ratio was measured. We avoided performing the reaction directly in D_2O directly to prevent H to D exchange in YEB ylide.



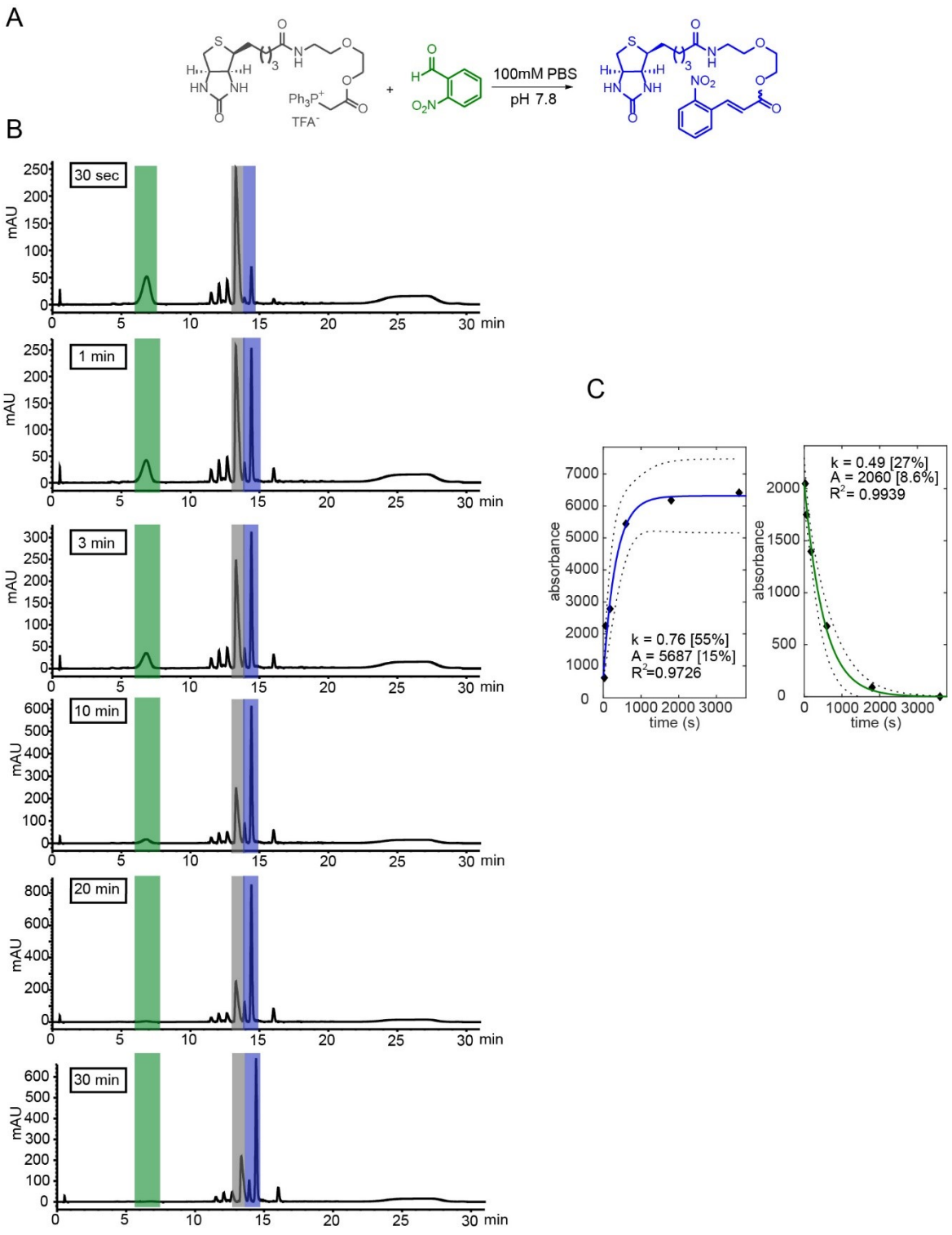
Appendix A-5 E/Z ratio for Wittig reaction with benzaldehyde. The reaction was run in 100 mM PBS at pH 7.5. Following the reaction we added 10% D_2O was added and measured the crude E/Z ratio by ^1H NMR integration. We avoided performing the reaction in D_2O to prevent H to D exchange in the YEB ylide Shown above is the in situ NMR measurement and the after purification LCMS trace (254 nm).



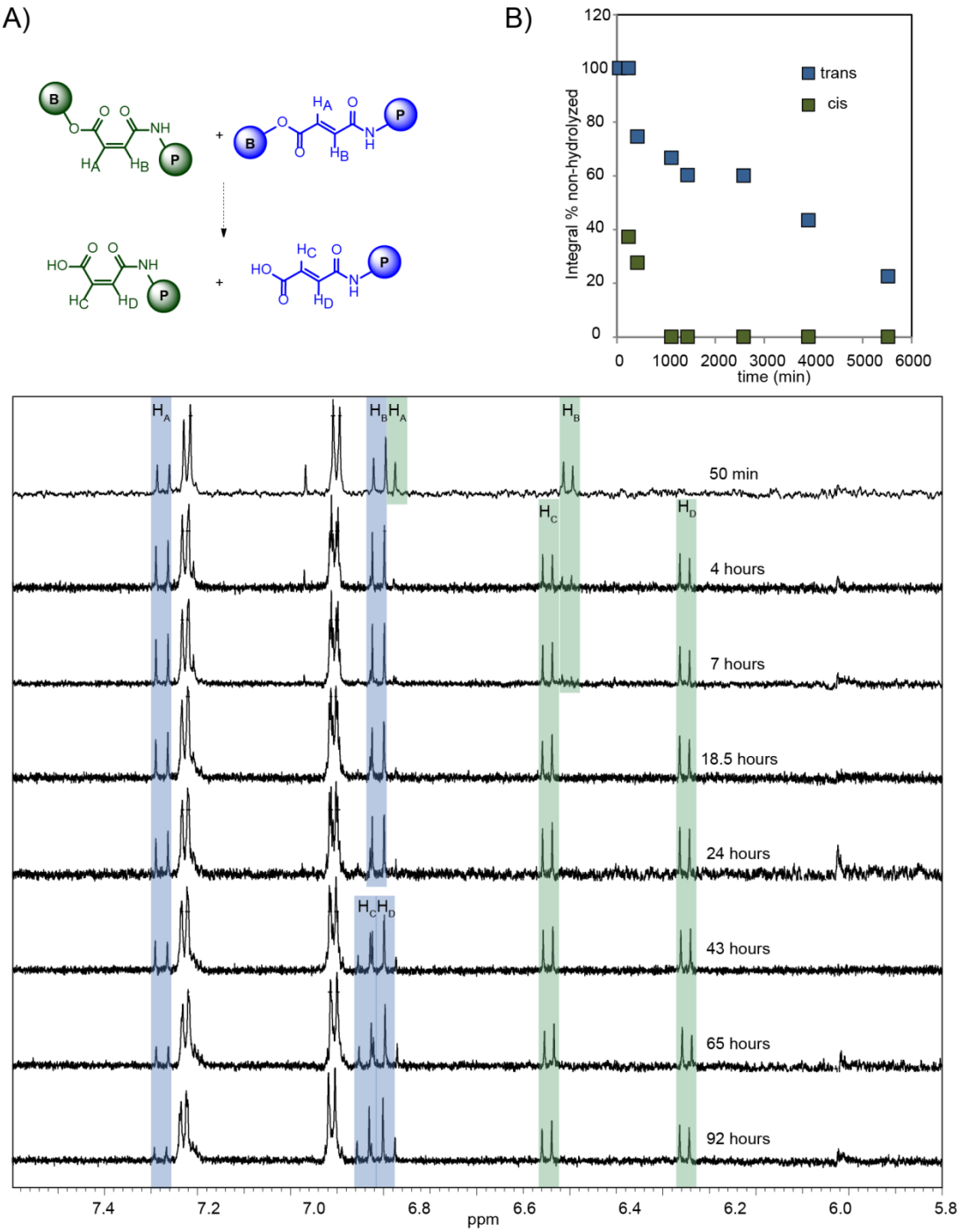
Appendix A-6 E/Z selectivity of 2-nitrobenzaldehyde Wittig reaction. A) 1H NMR-expanded spectrum of Wittig adduct in D_2O after HPLC purification. B) ^1H NMR-expanded spectrum of the crude mixture in 10% D_2O performed to confirm that the work-up and purification were not affecting E/Z ratios.[512]



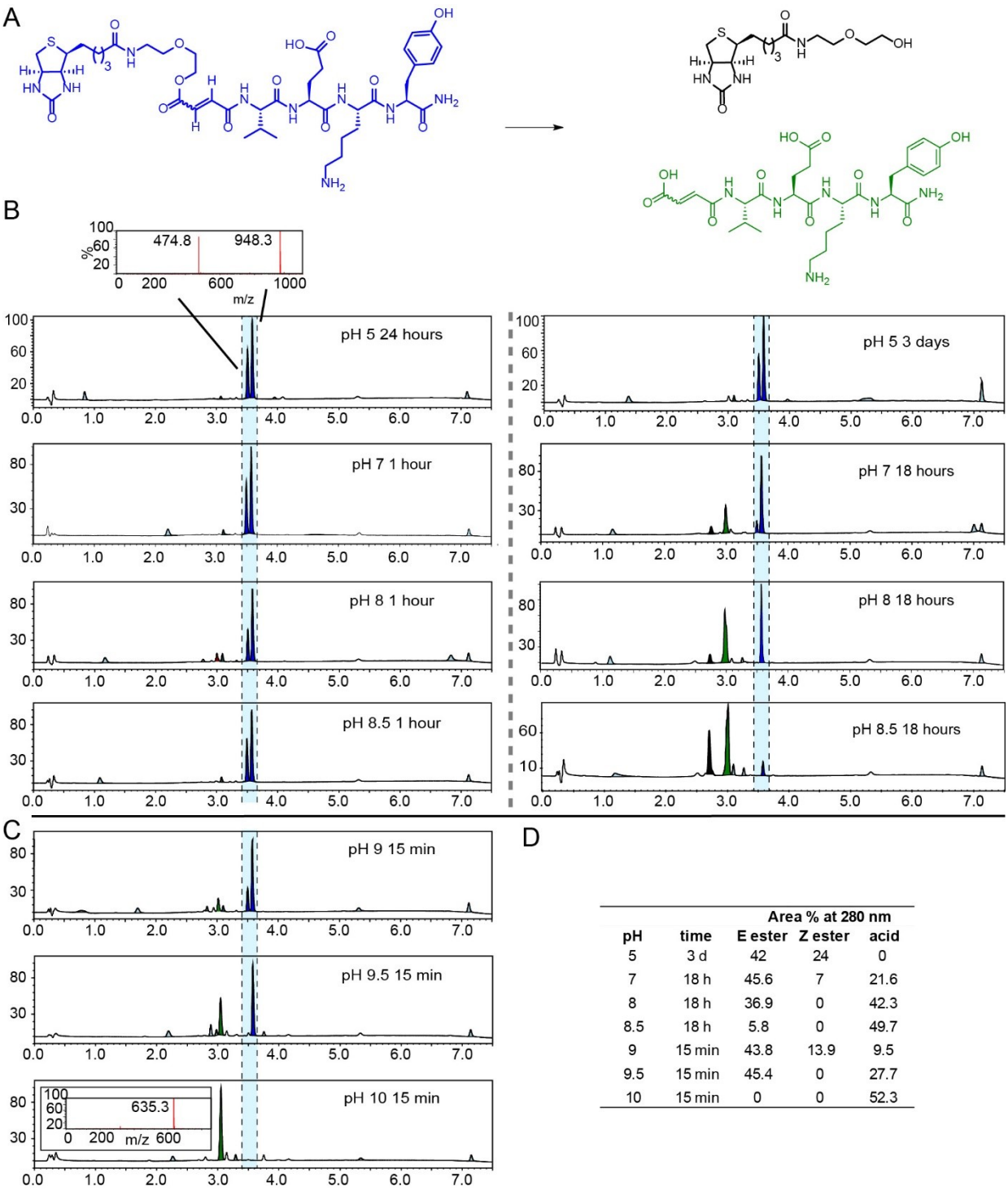
Appendix A-7 Example of LCMS traces used to obtain rate constants with SVEKY. A) Scheme of Wittig reaction on the model peptide SVEKY. B) Example of LCMS traces from which the area % were extracted for kinetic traces. C) Curve fitting was performed to pseudo first order kinetics using Matlab (Section 2.4.3.1)).



Appendix A-8 Example of HPLC traces used to obtain rate constants with 2-nitrobenzaldehyde. A) Scheme of Wittig reaction on model aldehyde 2-nitrobenzaldehyde. B) Example of HPLC trace from which absolute areas were extracted in order to obtain kinetic traces. C) Curve fitting was performed to pseudo first order kinetics using Matlab (Section 2.4.3.1).

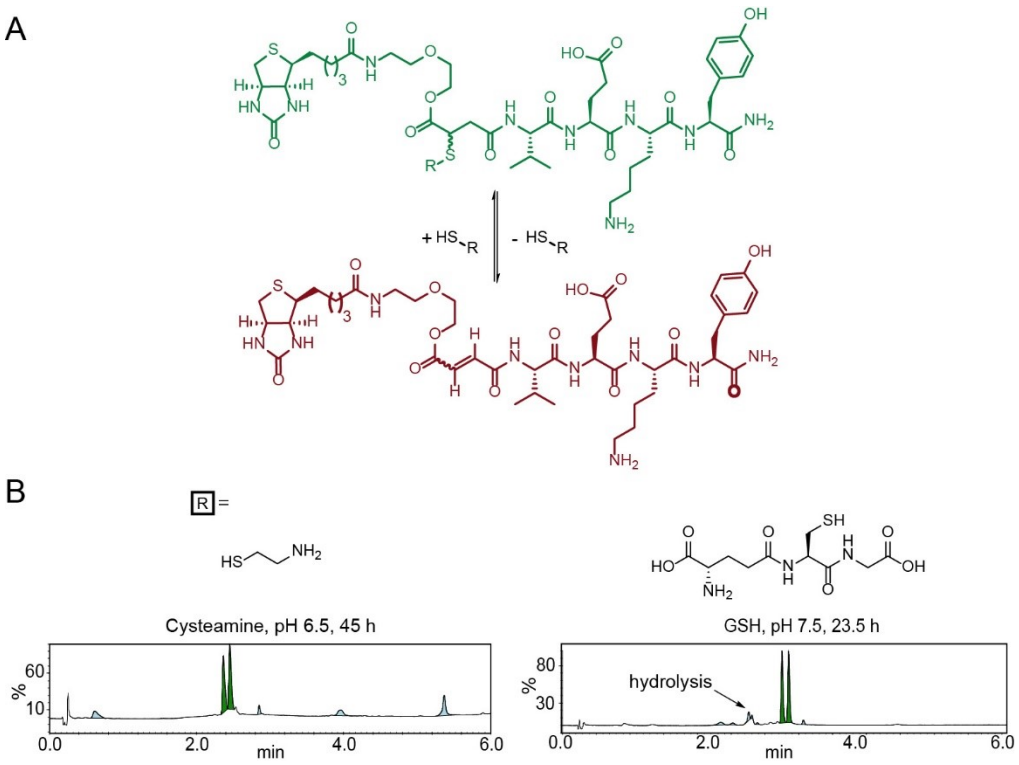


Appendix A-9 Rate of hydrolysis for cis and trans isomers of YEB-VEKY as measured by 1H NMR at pH 6.5.

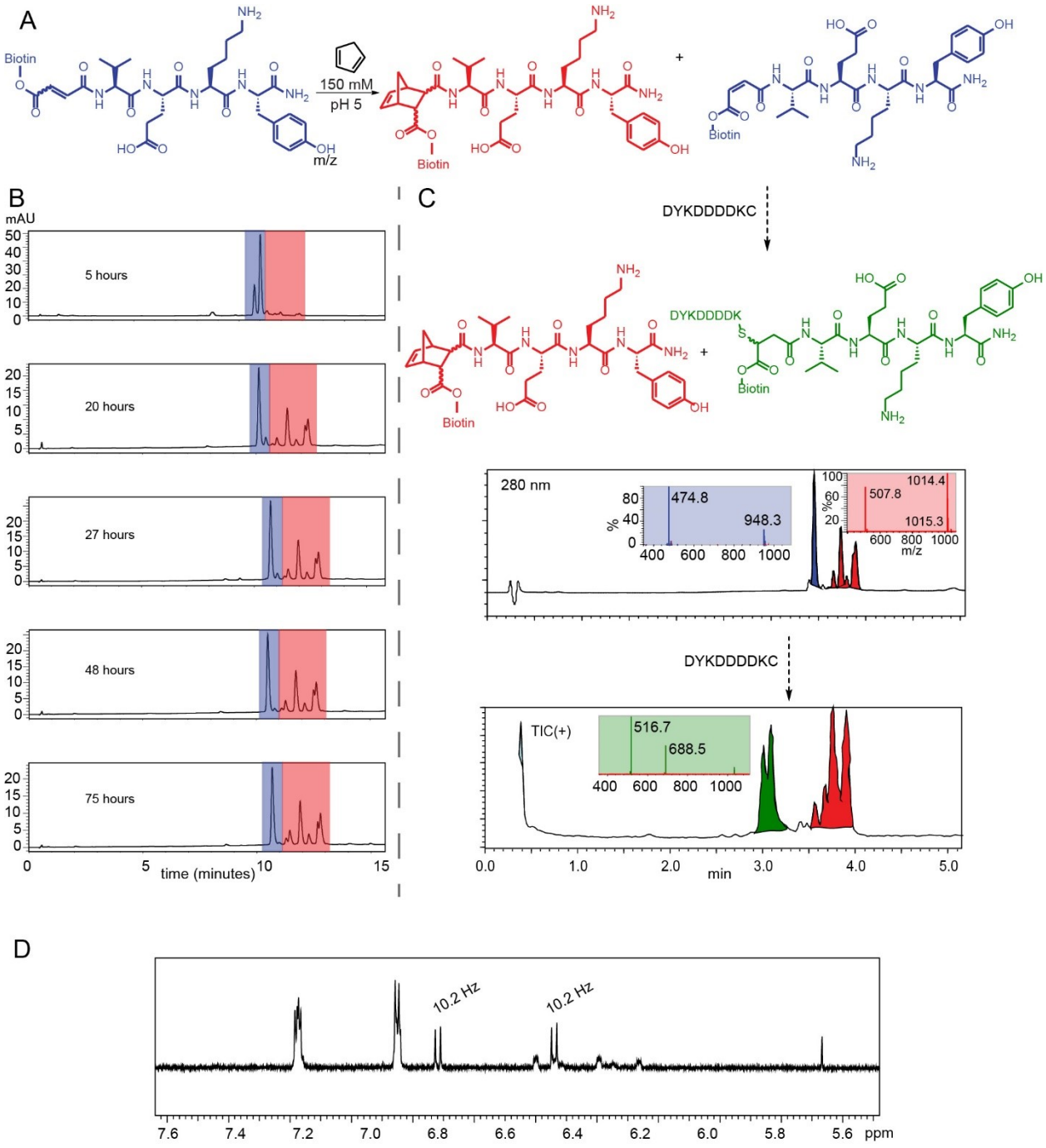


		Area % at 280 nm		
pH	time	E ester	Z ester	acid
5	3 d	42	24	0
7	18 h	45.6	7	21.6
8	18 h	36.9	0	42.3
8.5	18 h	5.8	0	49.7
9	15 min	43.8	13.9	9.5
9.5	15 min	45.4	0	27.7
10	15 min	0	0	52.3

Appendix A-10 Hydrolysis of YEB-VEKY at different pH values. A) Hydrolysis products for YEB-VEKY. B) pH values when product has a half-life >8 hours. C) pH values where product has a half-life <8 hours. D) Area percentages obtained by LCMS at 280 nm for Wittig products and after hydrolysis products**.**

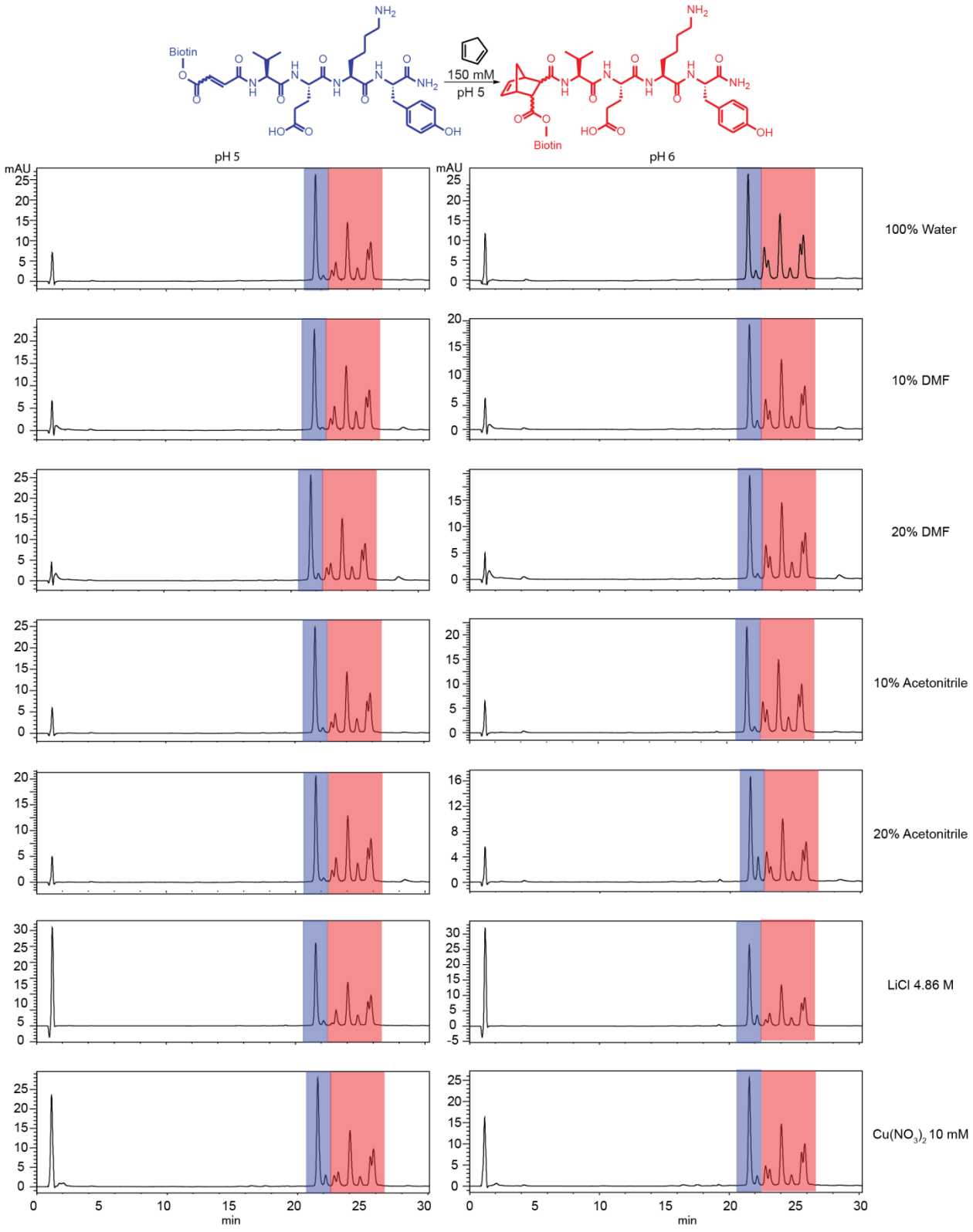


Appendix-A11 Retro-Michael tests on isolated Michael addition adducts. A solution (1 mM) of o-VEKY-Michael cysteamine adduct was prepared in (MOPS pH 6.5 100 mM). An aliquot (2 µL) of the solution was injected into the UPLC to determine the feasibility of retro-Michael reactivity. As no retro-Michael product was detected, higher pH (PBS 100 mM pH 7.5) solution of o-VEKY-Michael GSH adduct was prepared due to lower pKa of GSH when compared to cysteamine. No retro-Michael product was detected.

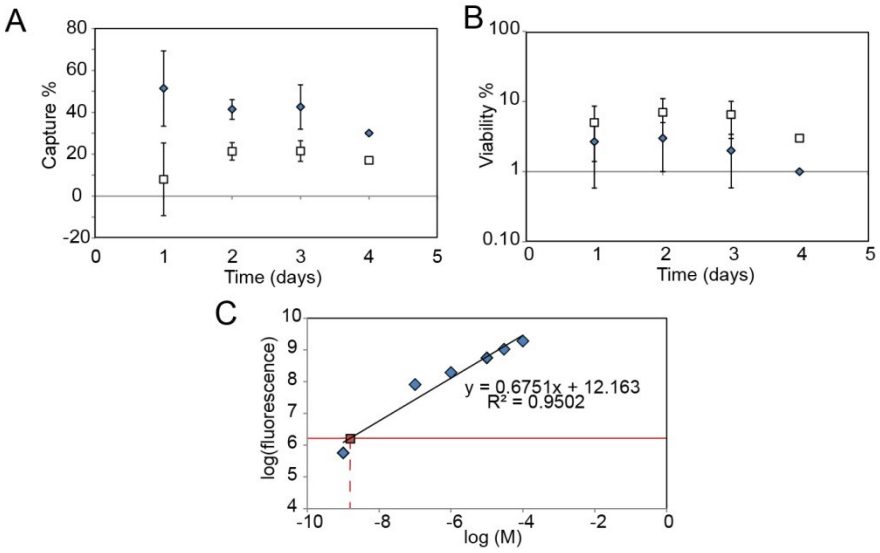


Appendix A-12 Diels-Alder on YEB-VEKY and identification of the non-reactive isomer. A) Scheme illustrating the Diels-Alder reaction on YEB-VEKY. B) Monitoring of Diels-Alder reaction at different time intervals on YEB-VEKY shows reactivity of one of the isomers only. C) Michael addition reaction performed on the crude mixture following Diels-Alder reaction on YEB-VEKY. LCMS traces of the reaction mixtures show: i) Non-reacted dienophile YEB-VEKY + Diels-Alder adduct before addition of DYKDDDDKC, ii) Michael addition adduct + Diels-Alder adduct after addition of DYKDDDDKC (TIC+ is shown for post-reaction LCMS due to excess

thiol diminishing the *relative* absorbance of products). D) ^{1}H-NMR identification of Z olefin as remaining starting material after Diels-Alder reaction. The Diels-Alder adducts appear as four separate peaks, likely due to *trans* endo/exo mixture and traces of *cis* endo/exo mixture (~6.22 and ~6.42 ppm). We considered characterization of the norbornene-like moieties by ^{1}H-NMR, however such characterization is complicated due to overlap of the characteristic norbornene signals with the aliphatic proton signals of the biotin linker and peptide.

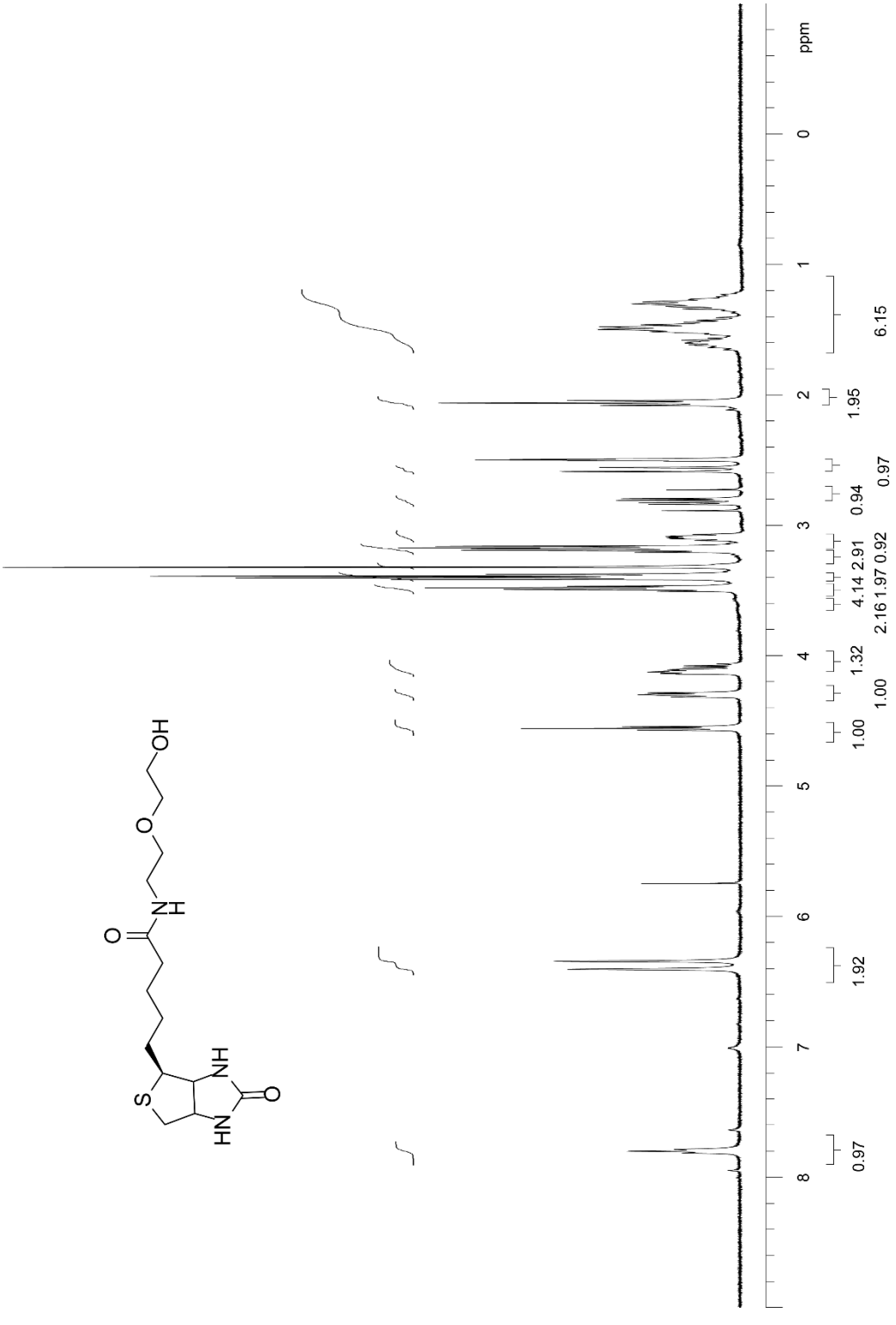


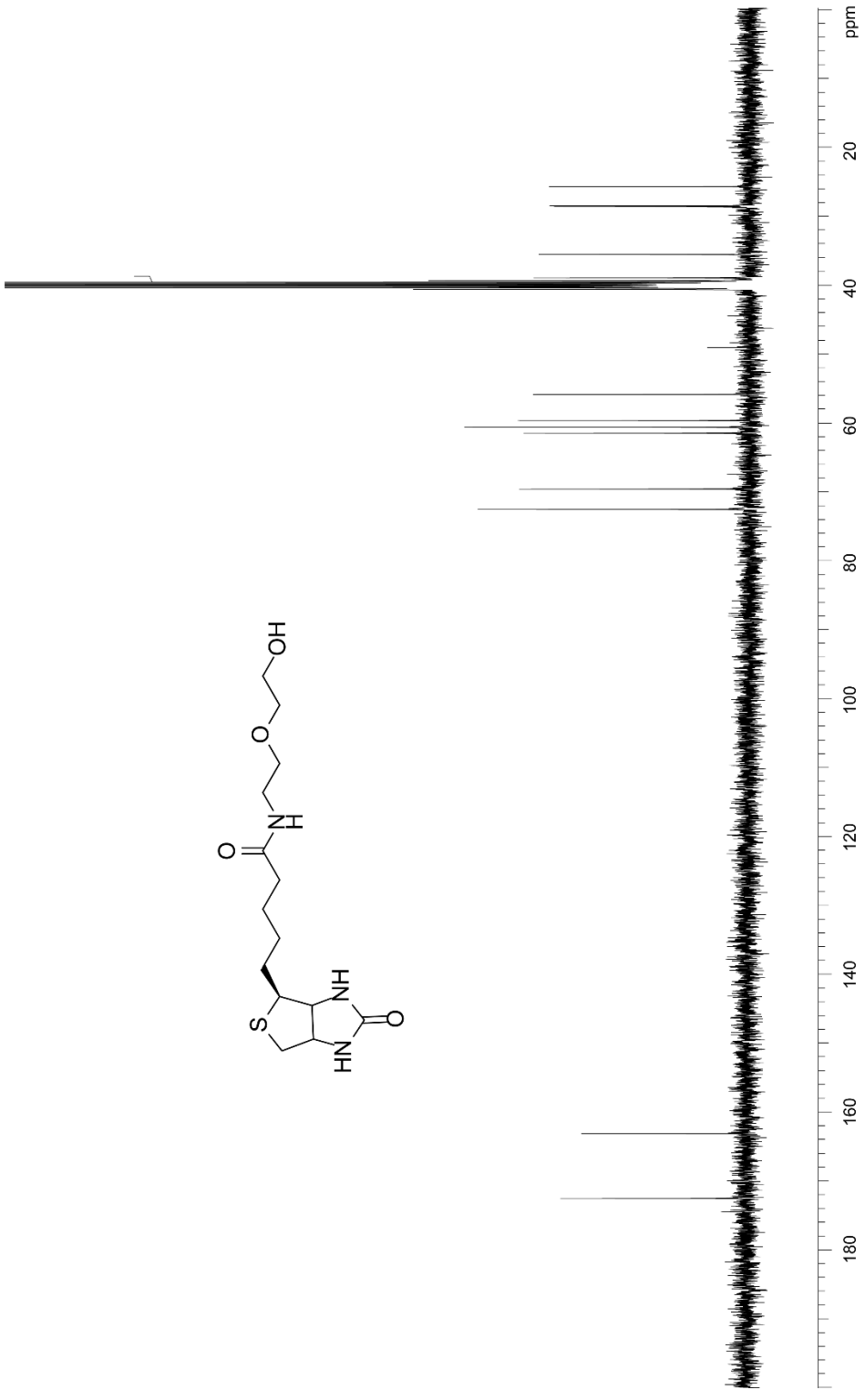
Appendix A-13 Diels-Alder reaction on YEB-VEKY (0.8 mM) in different reaction conditions.

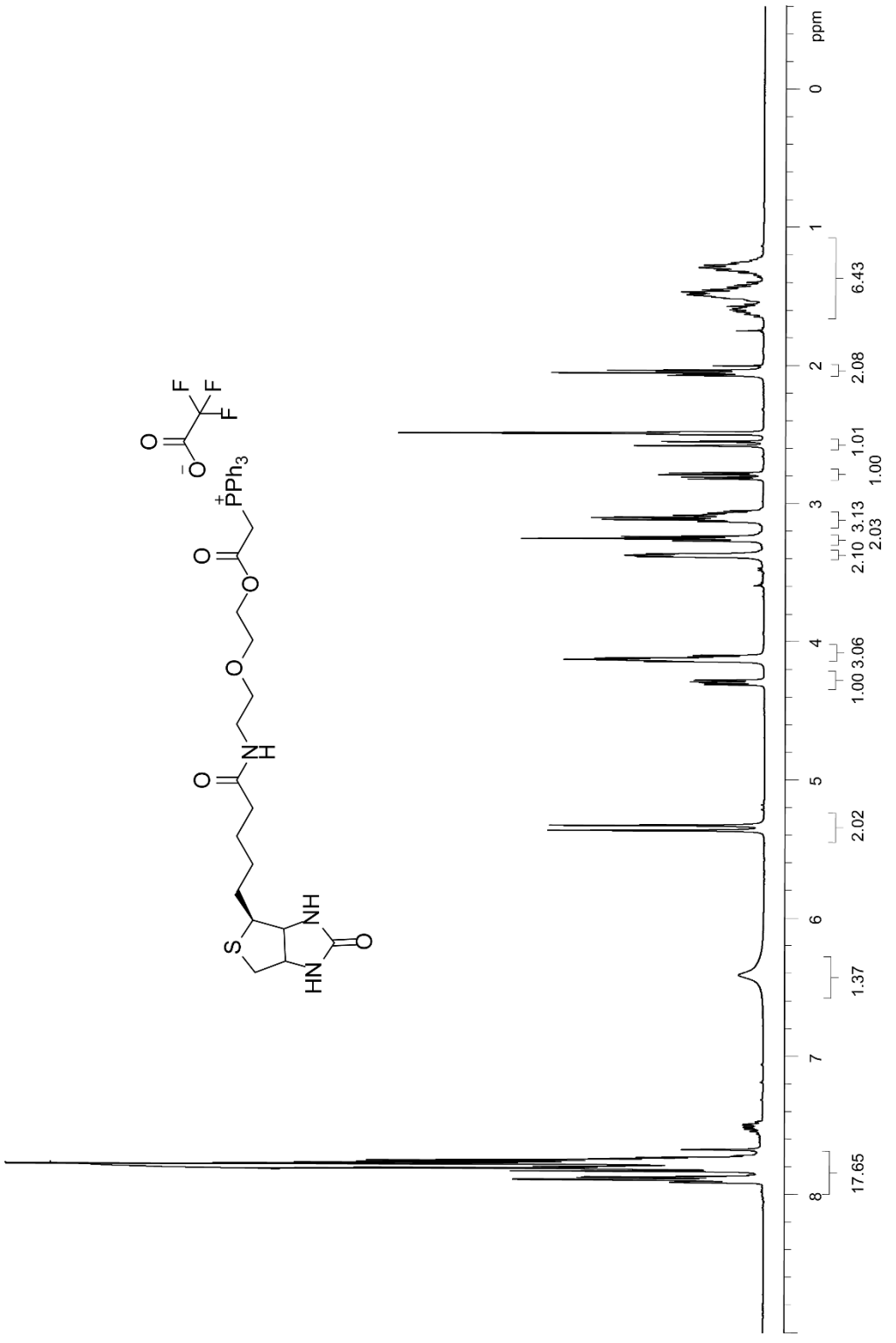


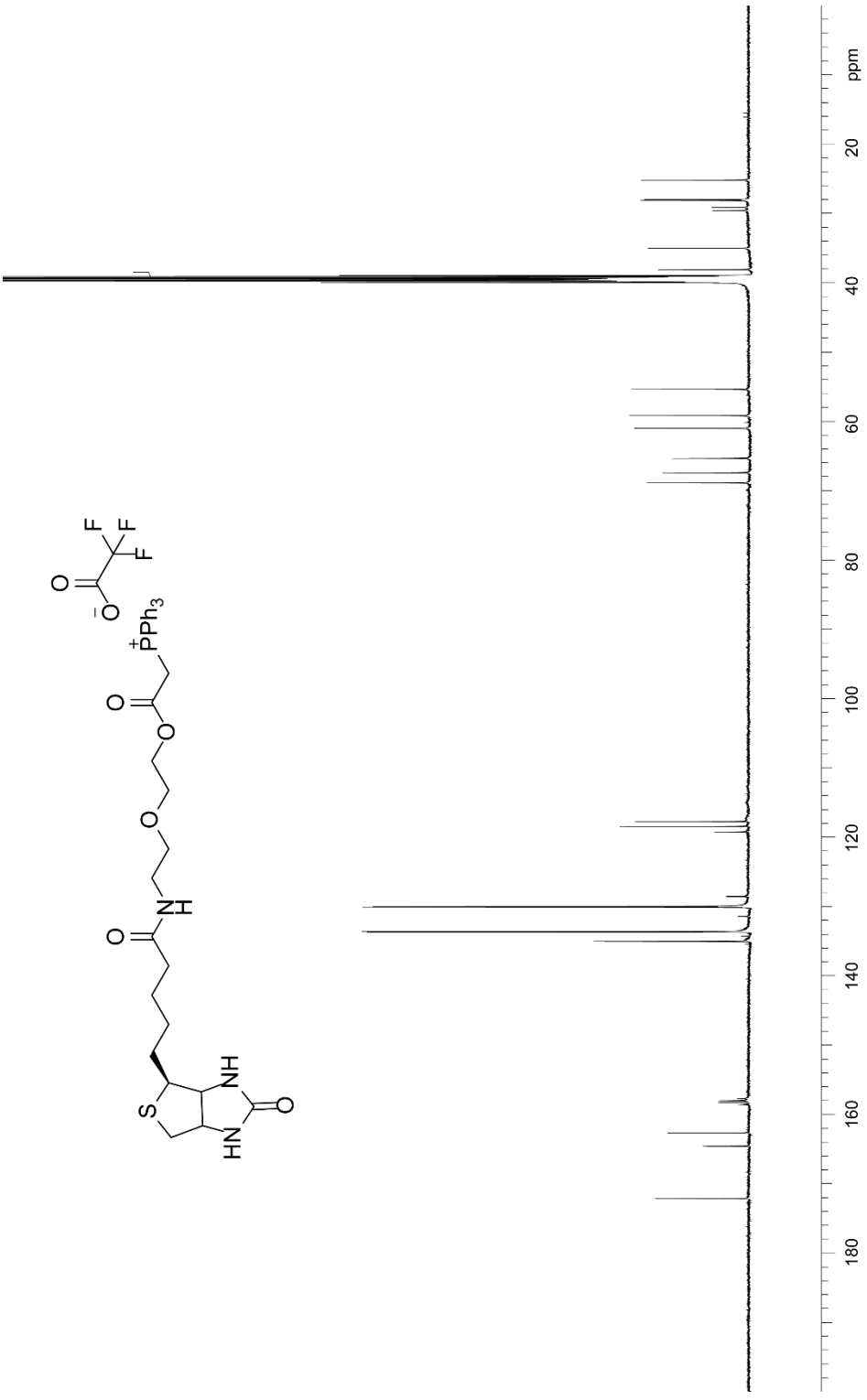
Appendix A-14 Storage and purification studies of biotinylated phage. A) Percentage of biotinylation of phage library vs. time when stored at pH 5. B) Viability of phage library vs. time at pH 5 following 24 hours of dialysis. C) Fluorescein calibration curve showing that final concentration (red line) is in the single digit nM range (1.5 nM in this specific study).

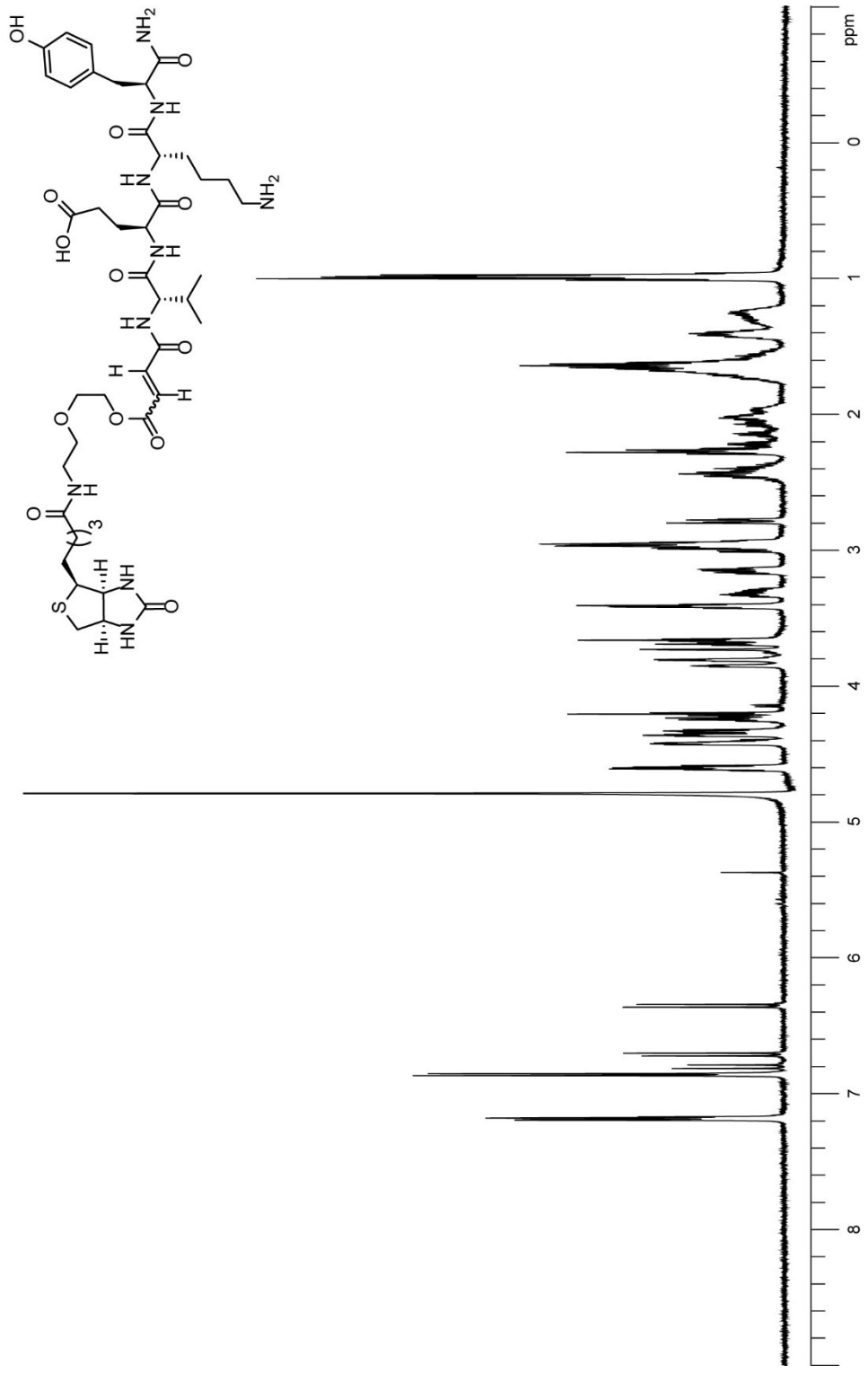
Appendix A-15: NMR of Synthesized Compounds




OH
O
NH
O
S
NH
HN
O
ppm
20
40
60
80
100
120
140
160
180


F
F
F
O
O
PPh3
O
O
O
O
NH
O
S
NH
HN
O
ppm
0
1
2
3
4
5
6
7
8
6.43
2.08
1.01
1.00
2.10 3.13
2.03
1.00 3.06
2.02
1.37
17.65


F
F
F
O
O
PPh3
O
O
O
O
NH
O
S
NH
HN
O
ppm
20
40
60
80
100
120
140
160
180


ppm
0
1
2
3
4
5
6
7
8

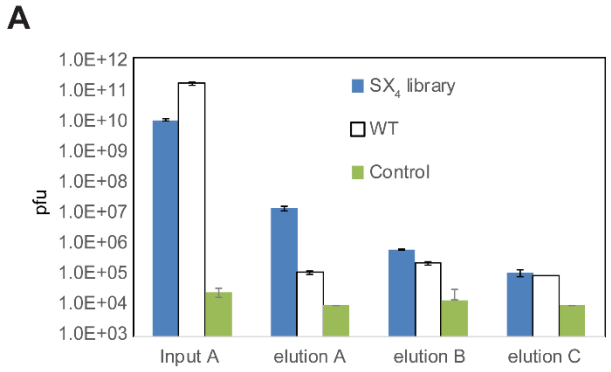
Appendix A-16: MATLAB script for calculation of rate constants

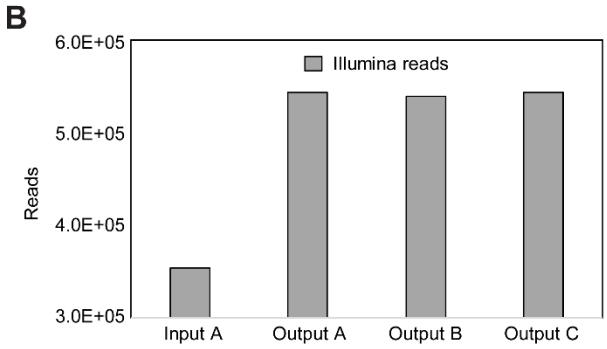
```
clear all;      % clear all variables from the workplace
close all;      % close all opened windows
indir = '';     % define the directory which holds the file (present dir)
name = 'VT-XIII-15.xlsx';    % define the file name
%name = 'pH8 area percentage.xls';
path = fullfile(indir,name);    % define a full path: directory/name
raw = xlsread(path);     % read the raw data from Excel file
rows = 1:7;      % define in which rows your data is stored
%rows = 36:39;
delay = 0;   % introduce a fixed delay to improve the fit
x = raw(rows,1) - delay;  % extract the time
x = x*60;
column = [3 2];   % define in which columns the product and reactants are
% define two types of 1st order equations for product and reactant
equation = {'A*(1-exp(-k*x))',  'A*(exp(-k*x))'};
% define the concentration of excess reagent in Molar units
conc  = 0.004;
%%%%% make your best guess about k and A parameters
%%%%%%%%%%%%%%%%%%%%
A  = 100;
k  = 0.01;
% run the same fit twice, plot the results in two subplots
for i = 1:2


y = raw(rows,column(i));  % extract the adsorbance from a current column
%%%%% define the fit options
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
s = fitoptions('Method','NonlinearLeastSquares',...
           'Lower',[0,   0],...
           'Upper',[100,   10],...
```

```
        'Startpoint',[A k],...
        'TolFun', 1e-10 );
%%%%% define the fit equation and options
%%%%%%%%%%%%%%%%%%%%%%%%%%%
ft = fittype( equation{i},'options',s );
%%%%% fit the data
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%
[c2,gof2,output] = fit(x,y,ft);
%%%%% find the 95% confidence bounds and confidence interval  %%%%%%%%%%%%%%%
CON = confint(c2); % confidence interval
x2=0:0.1:max(x);
p22 = predint(c2,x2,0.95,'functional','on');
%%%%% plot the raw data as black diamonds
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
figure(1);
subplot(1,2,i);
plot(x,y,'dk',...
     'MarkerEdgeColor','k',...
     'MarkerFaceColor','k',...
     'MarkerSize',5);
hold on;
%%%%% plot the fit data as red line and 95% confidence bounds as dash %%
plot( c2,'r');
plot(x2,p22,'k:');
legend off;
drawnow;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%  define the axis, re-scale them, label them
%%%%%%%%%%%%%%%%%%%%%%%
```

```
xmax = max(x)*1.05;
ymax = max(max(p22))*1.05;
xlim([0 xmax])
ylim([0 ymax])
xlabel('time (s)');
ylabel('absorbance');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%  display the results of the fit on the plot
% extract the fit value of k, divide it by concentration to yield real k
FIRSTorderk = c2.k / conc;
% calculate the % standard deviations for k and A
STD(1) = 100*abs(CON(1,1) - CON(2,1))/2 / c2.A;
STD(2) = 100*abs(CON(1,2) - CON(2,2))/2 / c2.k;
% create a 3-line text string that will be displayed on the chart
TL{1} = [' k =' num2str(FIRSTorderk,'%0.2f') ...
        ' [' num2str(STD(2),'%0.2f')  '% ]'];
TL{2} = [' A =' num2str(c2.A,'%0.2f') ...
         ' [' num2str(STD(1),'%0.2f')  '% ]'];
TL{3} = [' R^2=' num2str(gof2.rsquare,'%0.4f') ];
% plate the text string on the chart into a predefined location
text(0.15*xmax,0.75*ymax, char(TL));
end
```

Appendix B. Supporting information for chapter 3





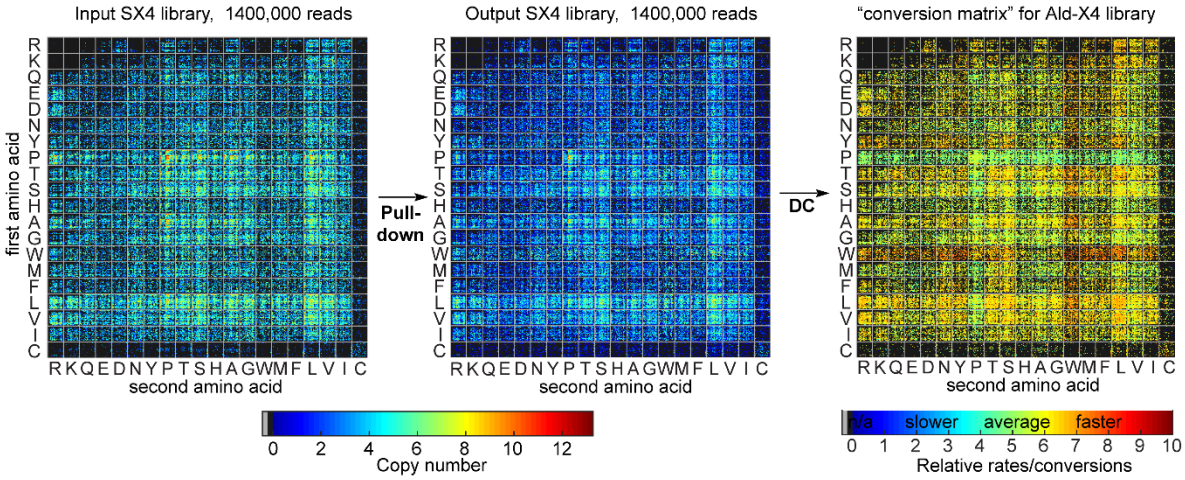
C

	Set A			Set B			Set C		
	522K	610K	516K	617K	488K	498K	644K	490K	613K
Depth	0	54	40	18119	5495	3543	284	167	690
Clone 1	0	49	33	14801	4918	3849	98	42	128
Clone 2	0	0	0	87	132	22	2	2	0
Clone 3	0	0	0	71	43	25	0	1	0
Clone 4	0	0	0	3	3	2	11	7	34
Clone 5	0	0	0	1	1	0	0	0	0
Clone 6	0	0	0	1	0	0	0	0	0
Clone 7									

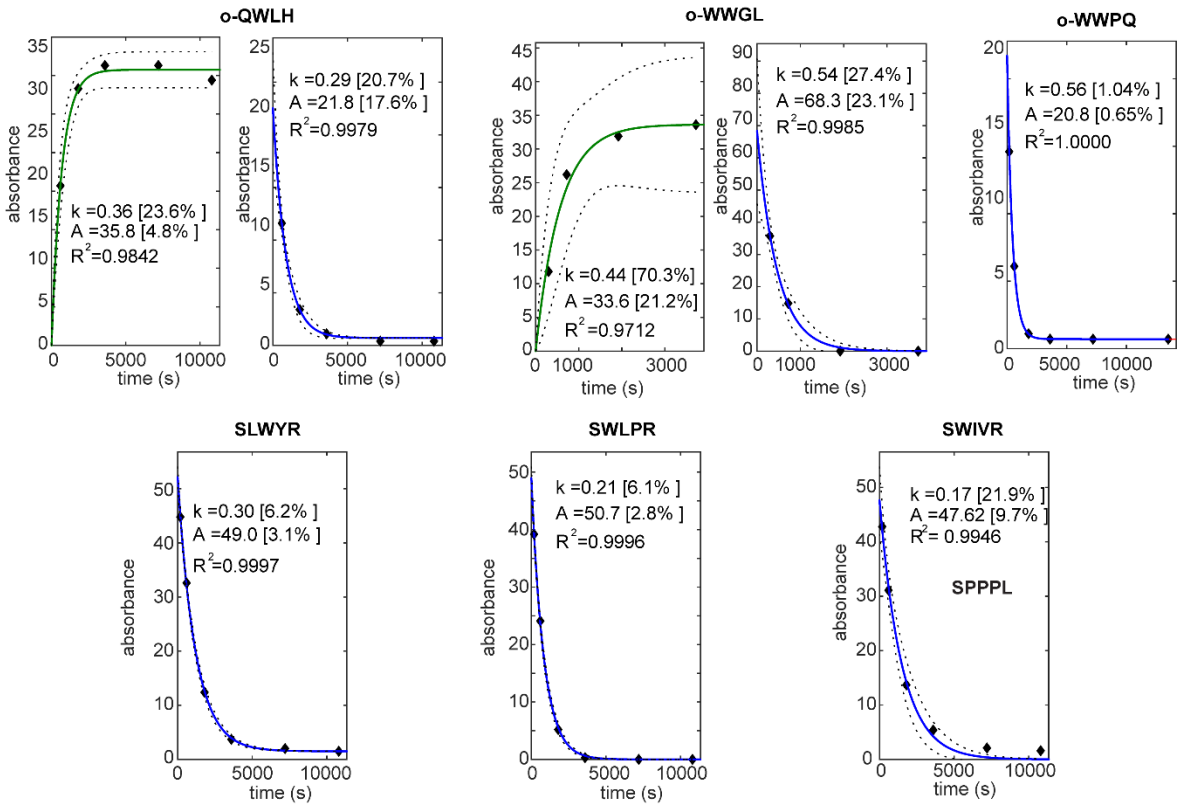
[matrix] . [1/Sa]

Clone 1	0	4426	3876	587	225	142	1	1	2
Clone 2	0	4016	3198	480	202	155	0	0	0
Clone 3	0	0	0	0	0	0	0	0	0
Clone 4	0	0	0	3	5	1	0	0	0
Clone 5	0	0	0	2	2	1	0	0	0
Clone 6	0	0	0	0	0	0	0	0	0
Clone 7	0	0	0	0	0	0	0	0	0

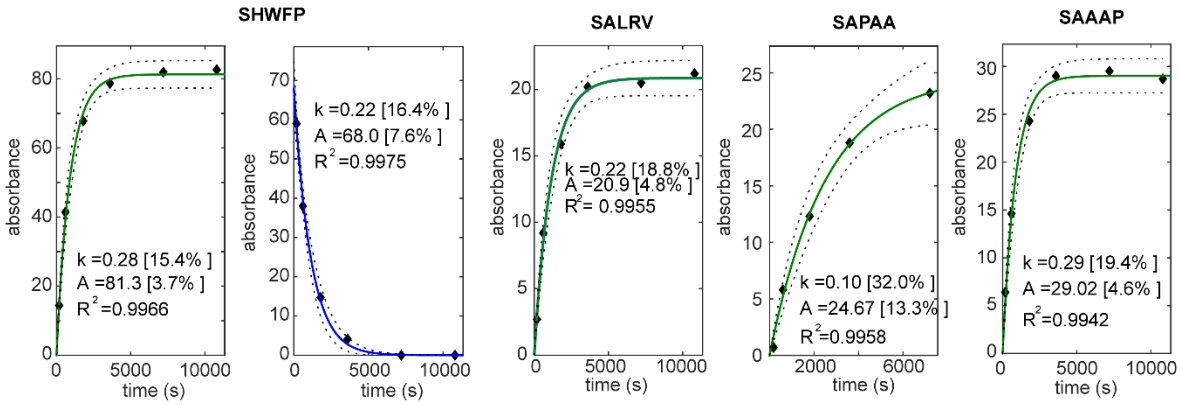
Appendix B-1. Deep sequencing analysis for 1% SX4 selection control. A) Tittering. B) Illumina reads. C) Illumina counts for control mixture and Ci using equations 1-3. D) 20:20 plots for SX_4 1% selection.



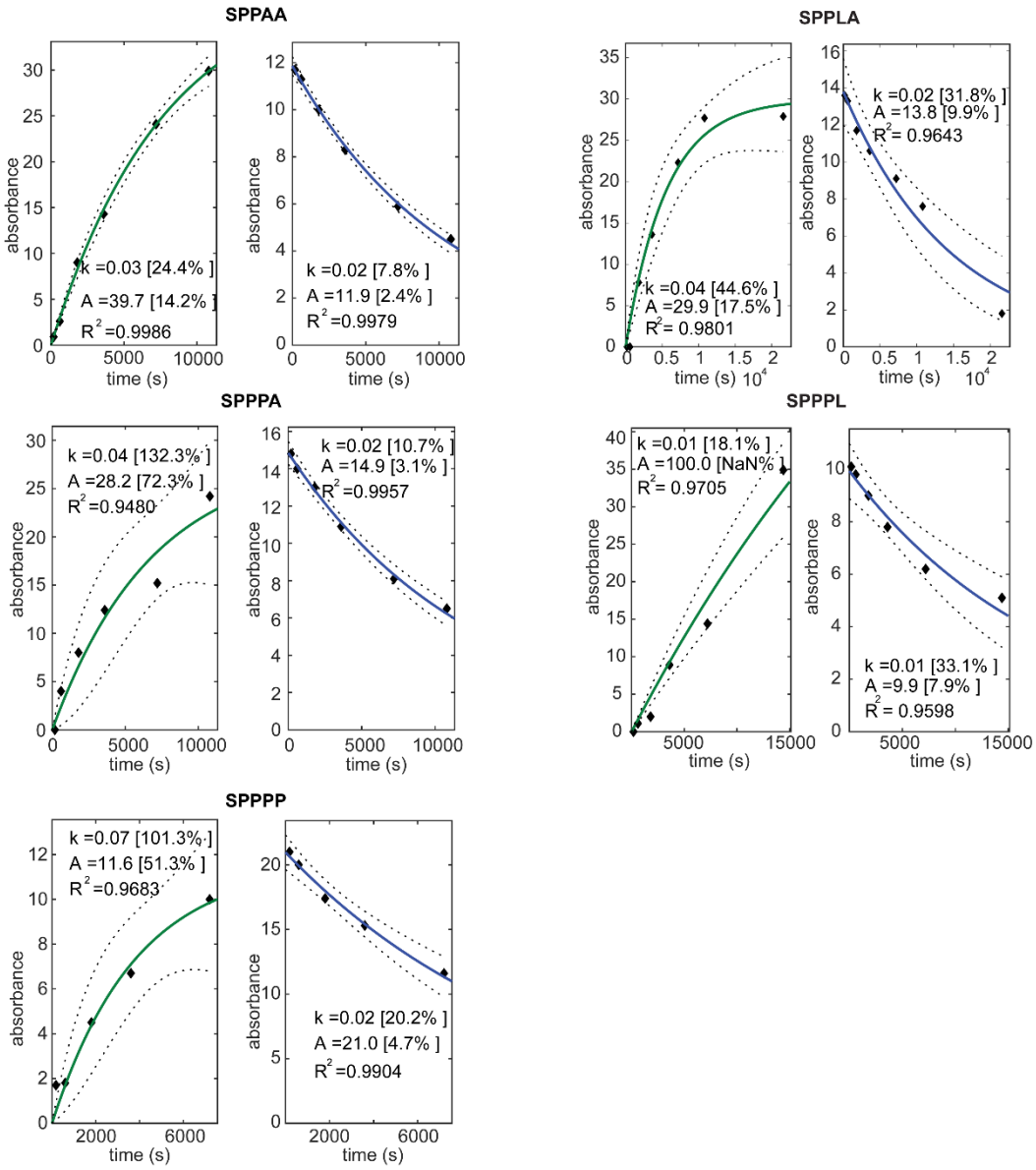
Appendix B-2. 20x20 plots for SX_4 1% selection



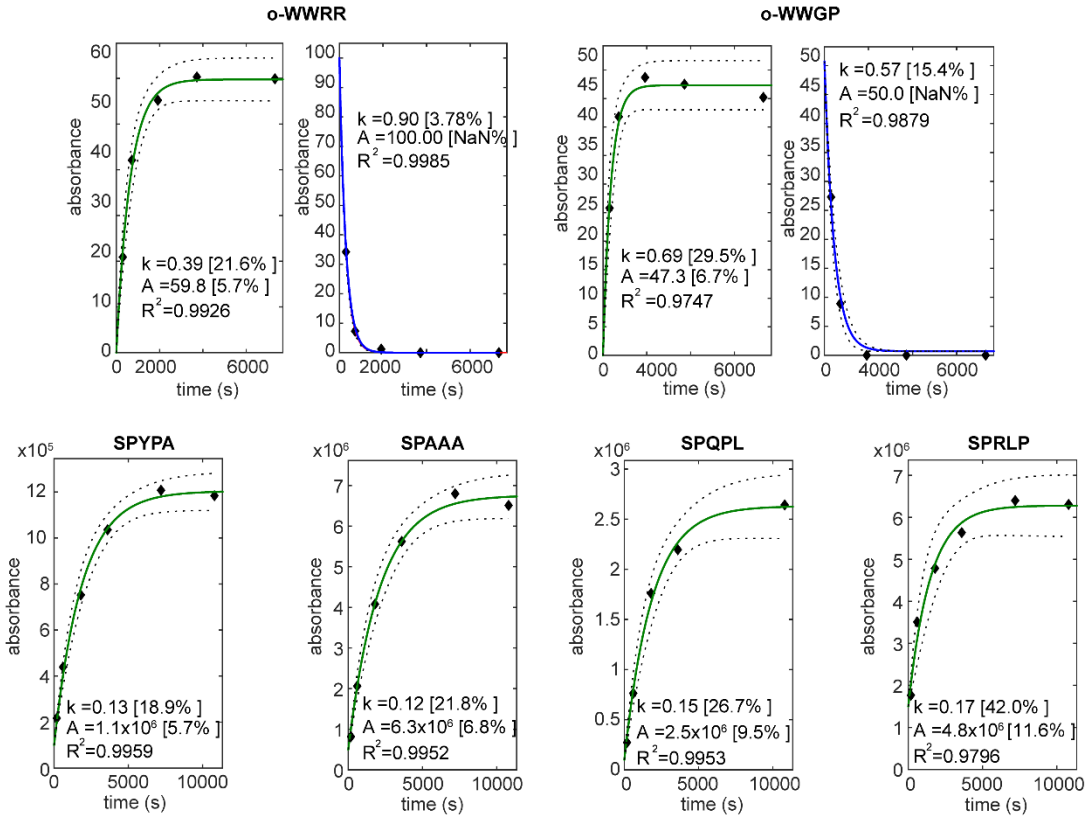
Appendix B-3. Kinetic traces for 1% SX4 DC≥3.5



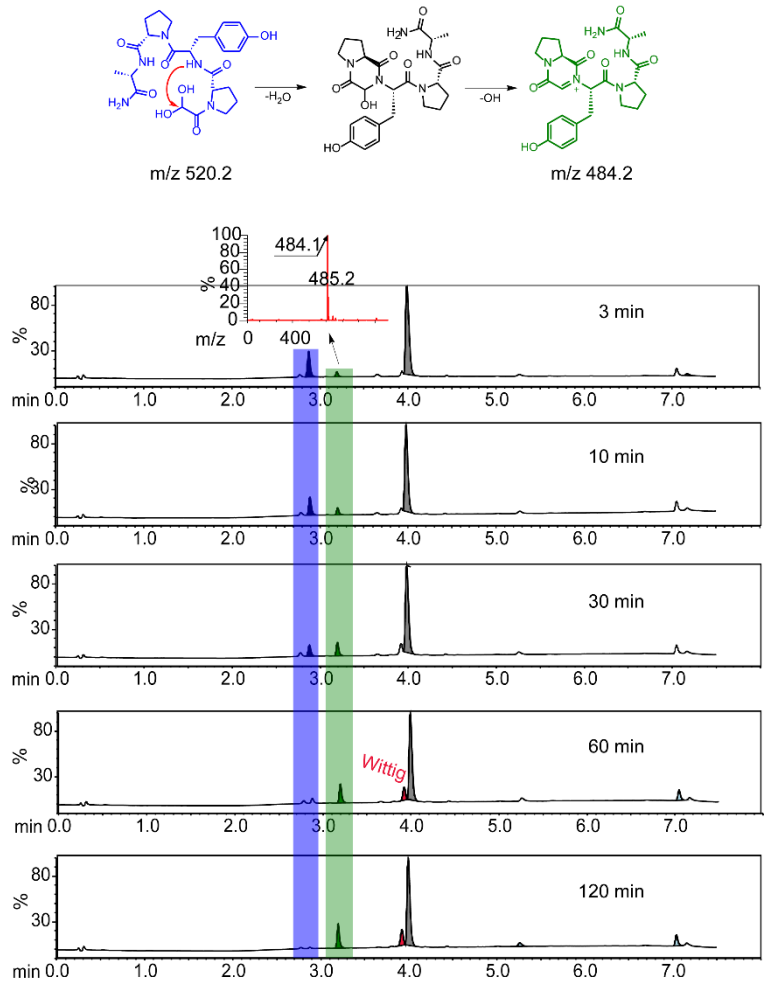
Appendix B-4. Kinetic traces for 1% SX4 and $0.15 \leq DC \leq 3.5$



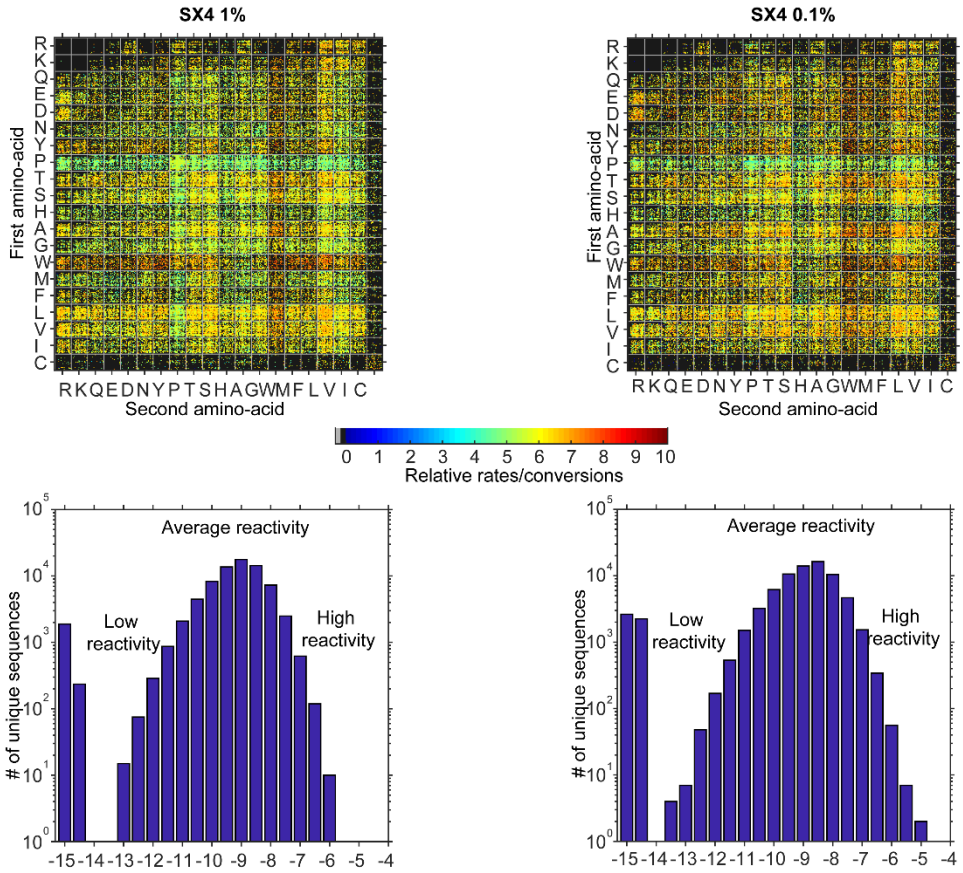
Appendix B-5. Kinetic traces for 1% SX4 and 0.15 ≥ “Illumina rate” (not including SPXXX sequences).



Appendix B-6. Kinetics for sequences with no Illumina counts and SPXXX sequences.



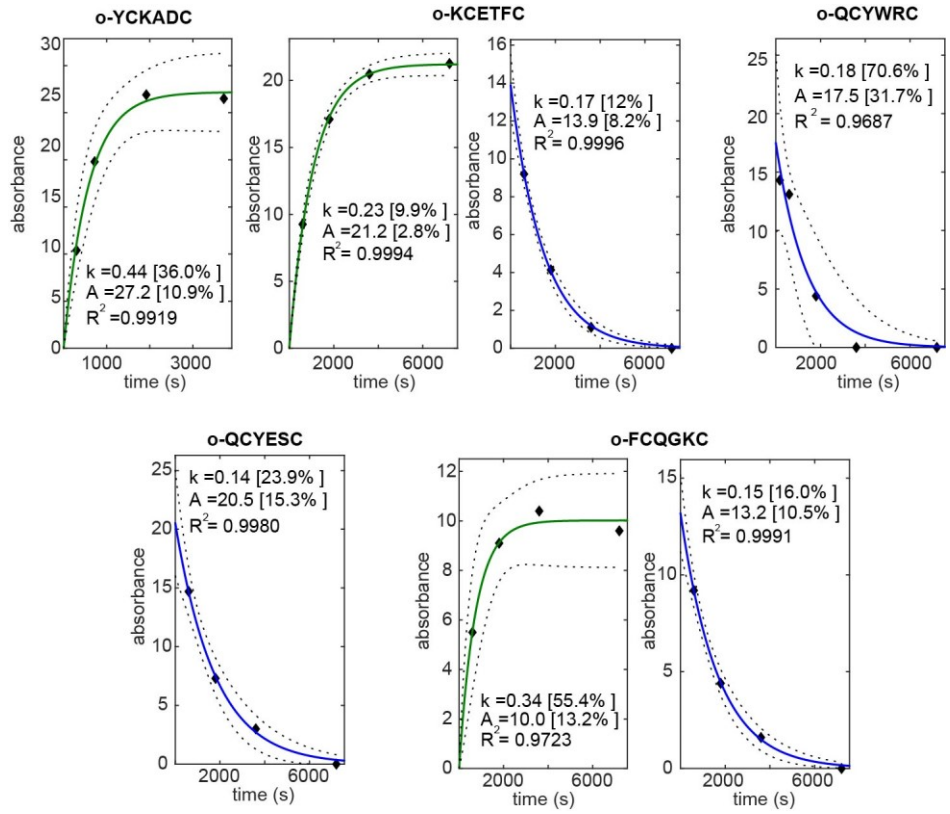
Appendix B-7. Example of SPXXX LCMS traces for determination of rate.



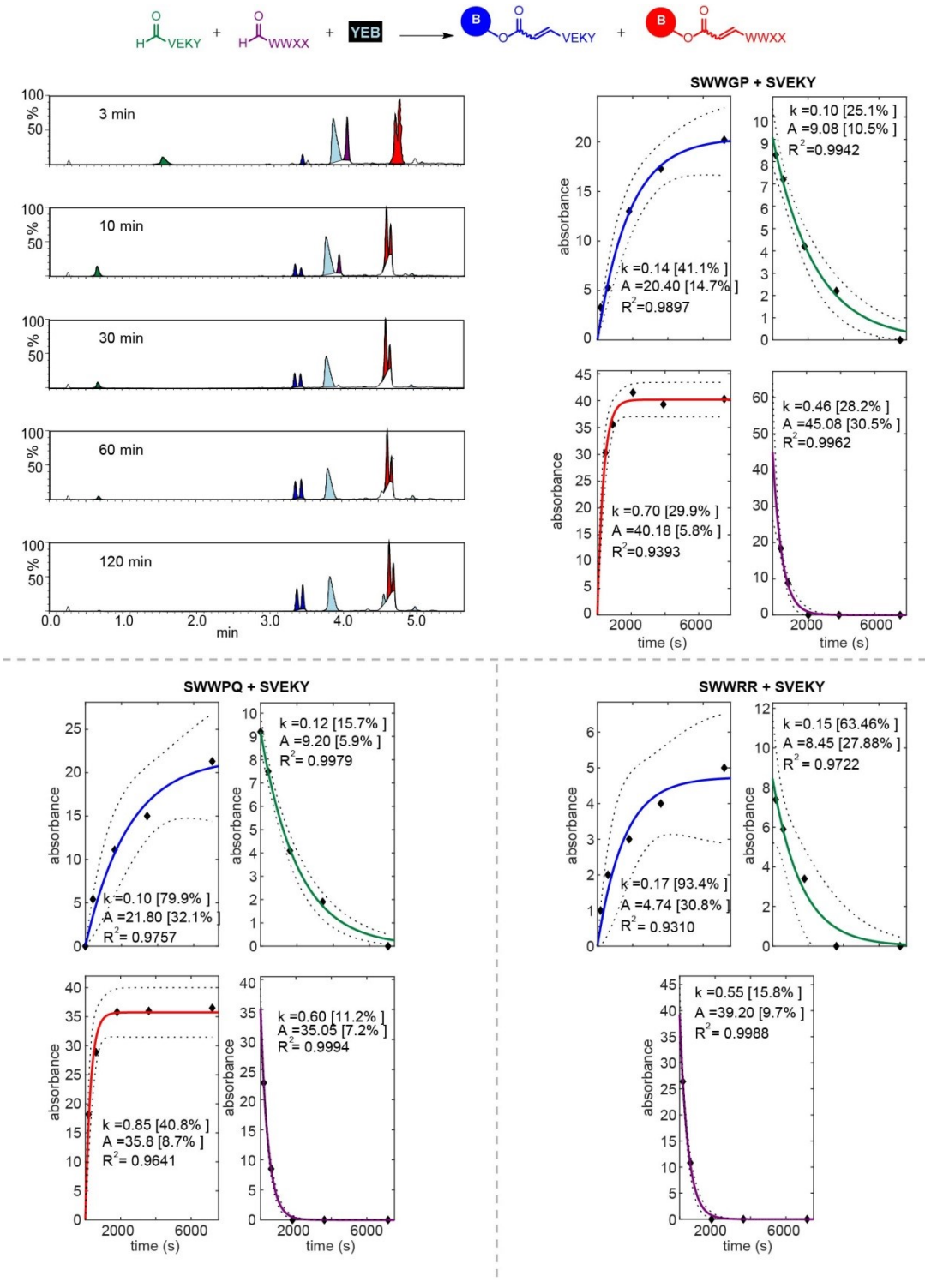
Appendix B-8. 20:20 plot for SX_4 1% vs SX_4 0.1% selection.

Entry	Sequence	DC	Rate
1	SYCKADC	0.28	0.44
2	SKCETFC	0.65	0.23
3	SQCYWRC	0.45	0.18
4	SQCYESC	-	0.14
5	SFCQGKC	0.73	0.15

Appendix B-9. Comparison between DC and rate constants in isolated peptides for SXC3C 1% selection

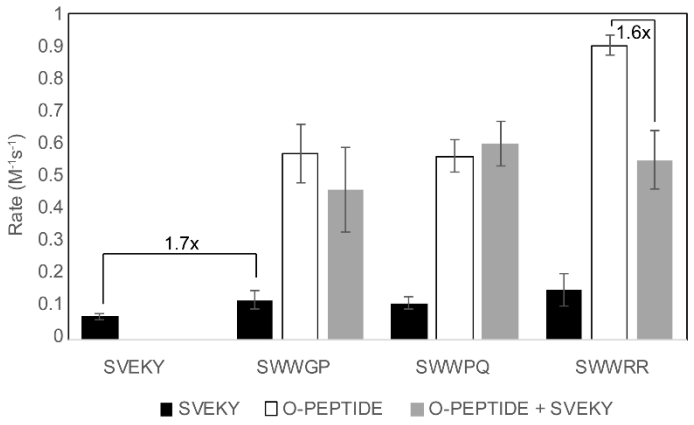


Appendix B-10. Kinetic traces for sequences of 1% SXC_3C selection

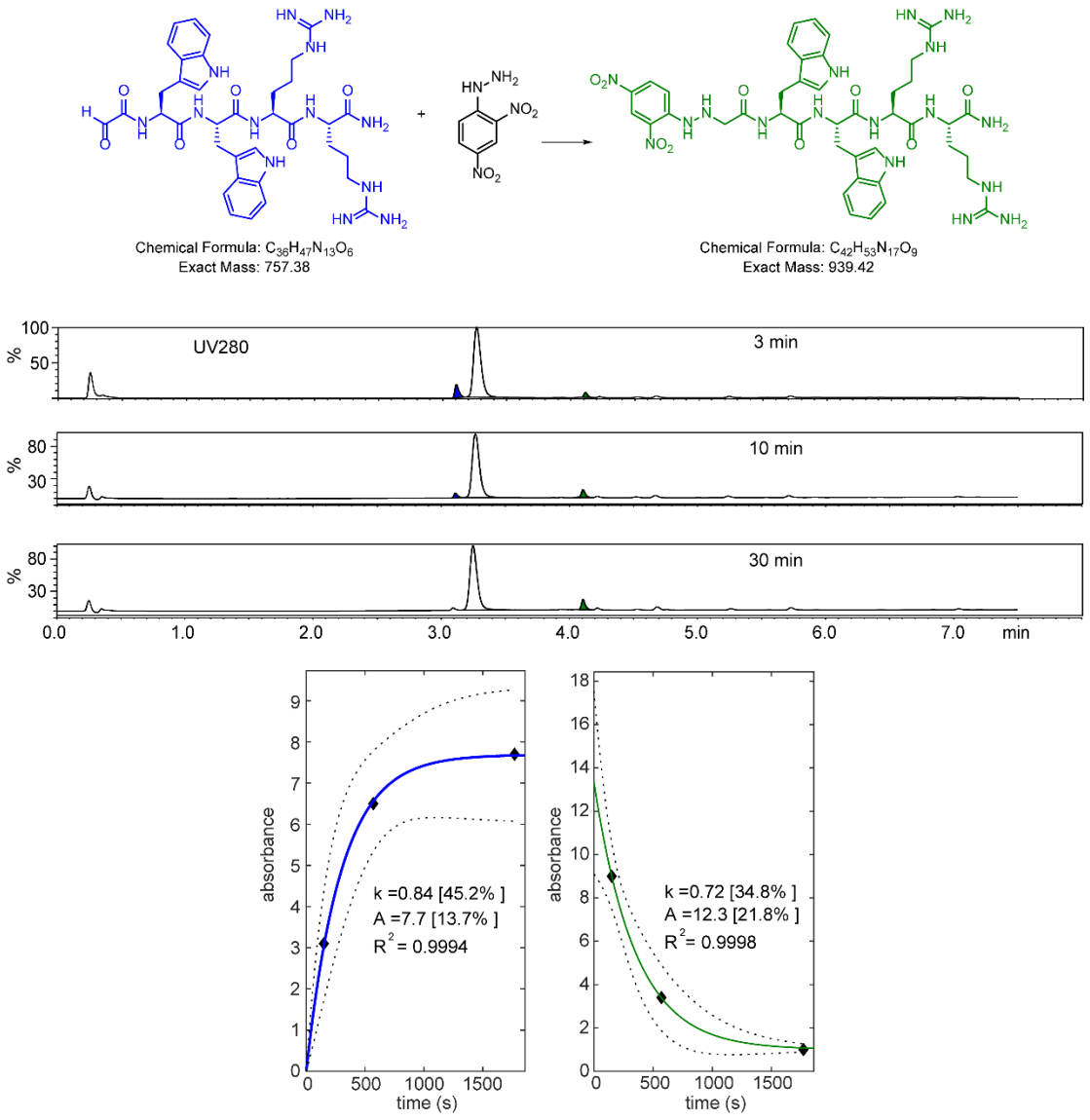


Appendix B-11. One well kinetics for control sequence (SVEKY) and several SX_4 1% “hits”

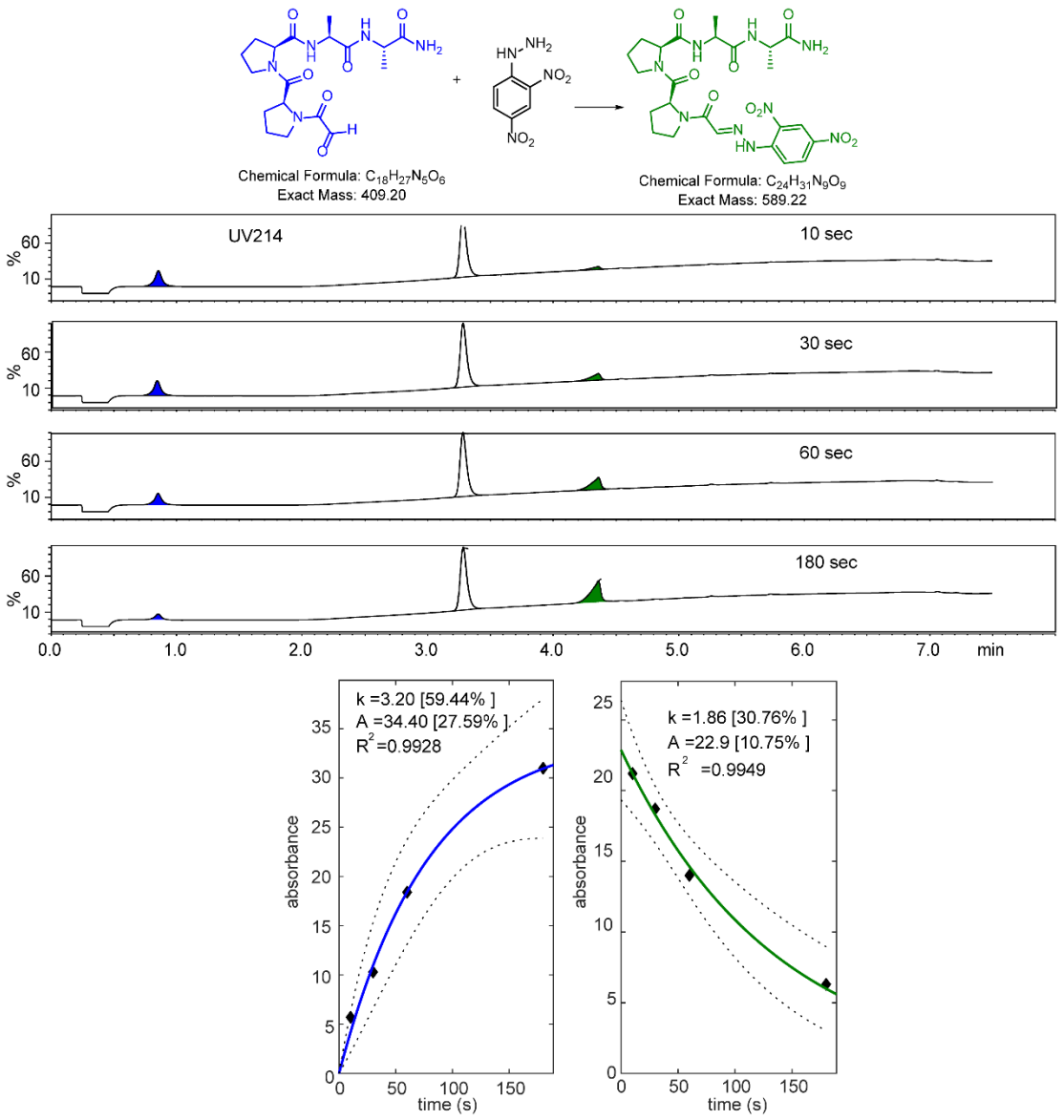
Sequence	Individual Rate	Rate in mixture	Control rate	Control sequence
SWWGP	0.57 ± 0.09	0.46 ± 0.13	0.12 ± 0.03	SVEKY
SWWPQ	0.56 ± 0.05	0.6 ± 0.07	0.11 ± 0.02	SVEKY
SWWRR	0.9 ± 0.03	0.55 ± 0.09	0.15 ± 0.05	SVEKY
SVEKY	0.07 ± 0.01			



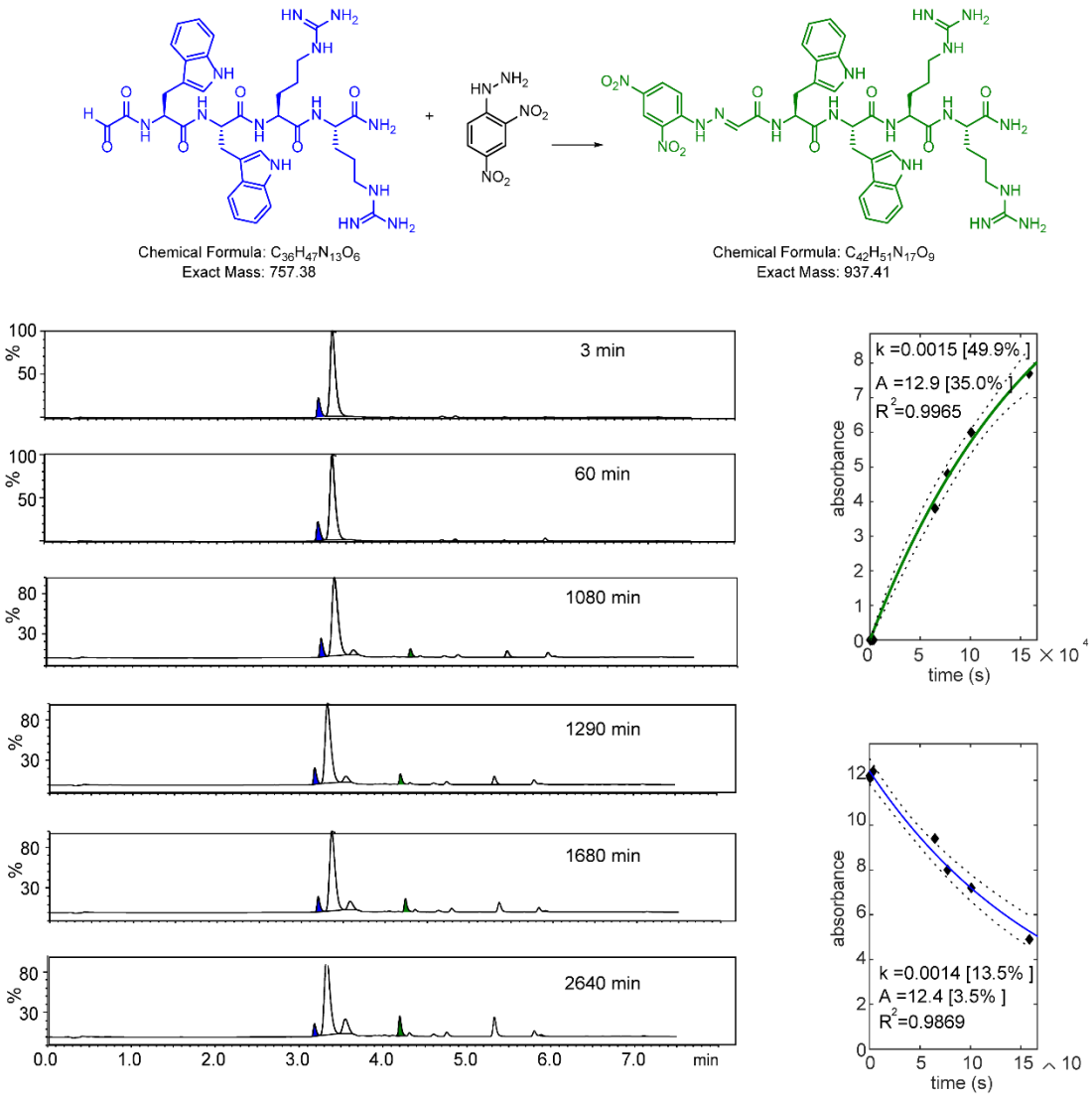
Appendix B-12. Comparison of Wittig rate of SX4-type sequences in individual and mixed aldehyde experiments.



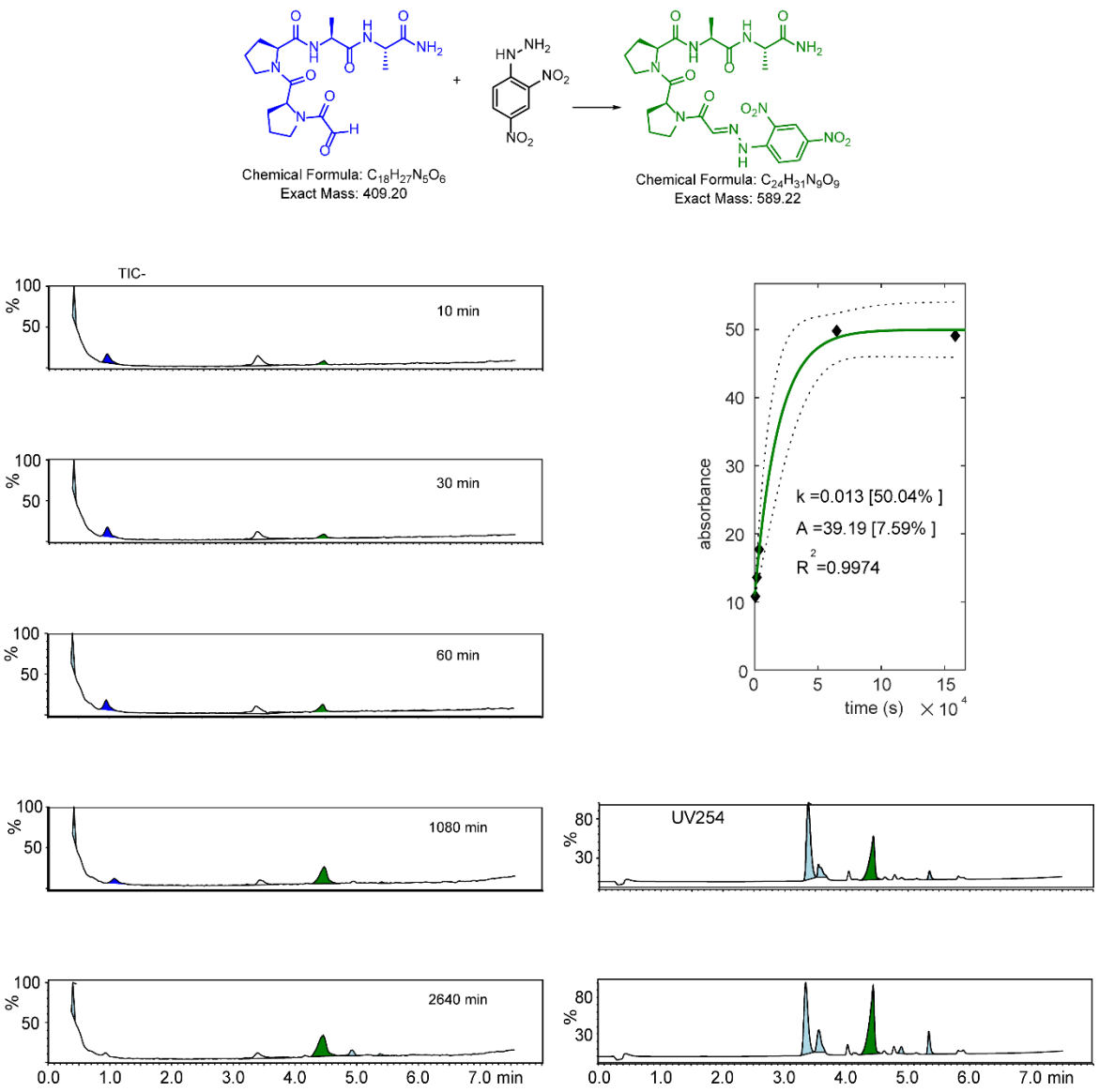
Appendix B-13. Kinetics of hydrazone ligation for Ald-WWRR at [H_2SO_4] = 65 mM.



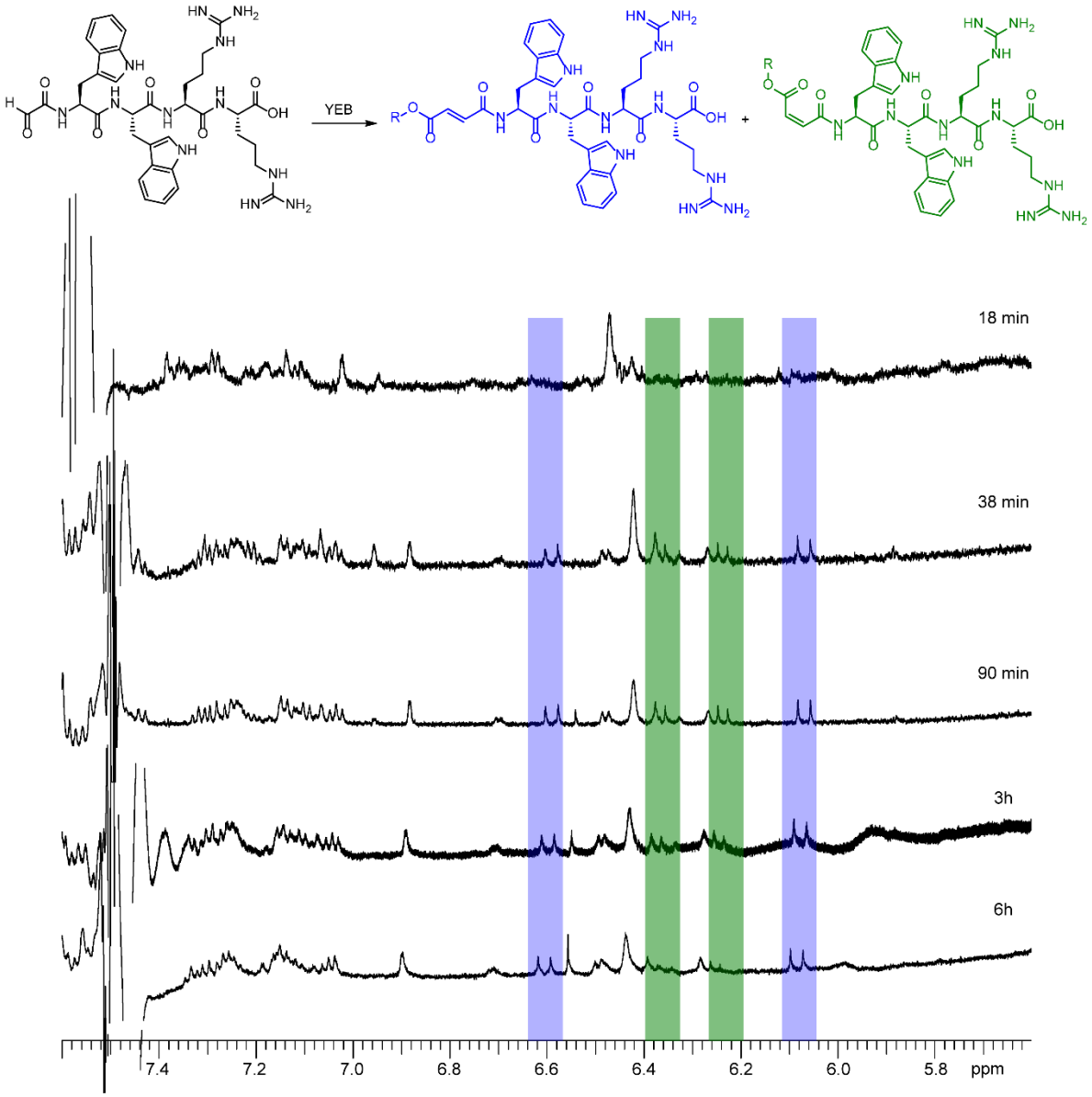
Appendix B-14. Kinetics of hydrazone ligation for Ald-PPAA at [H_2SO_4] = 65 mM.



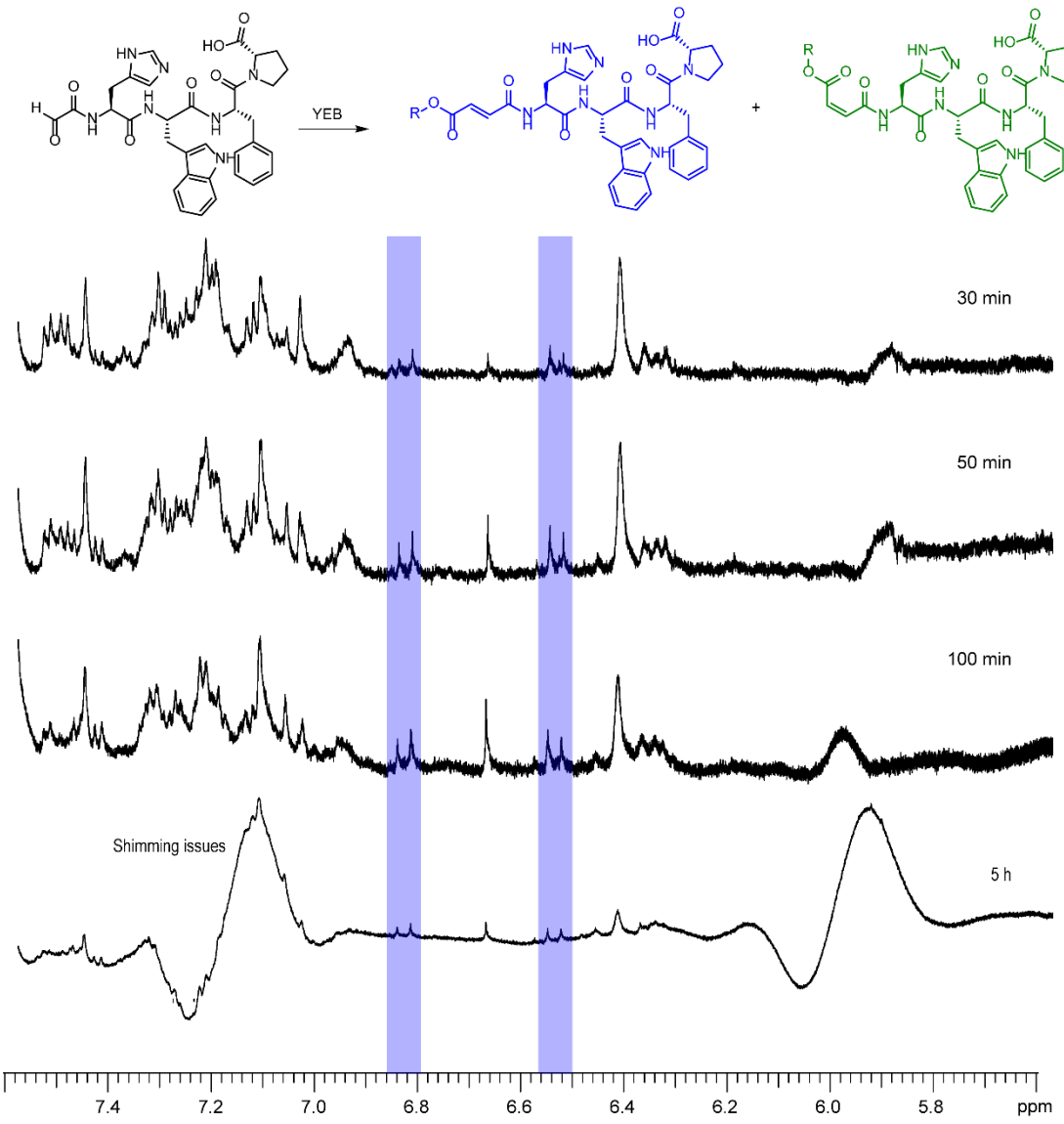
Appendix B-15. Kinetics of hydrazone ligation for Ald-WWRR at pH 5



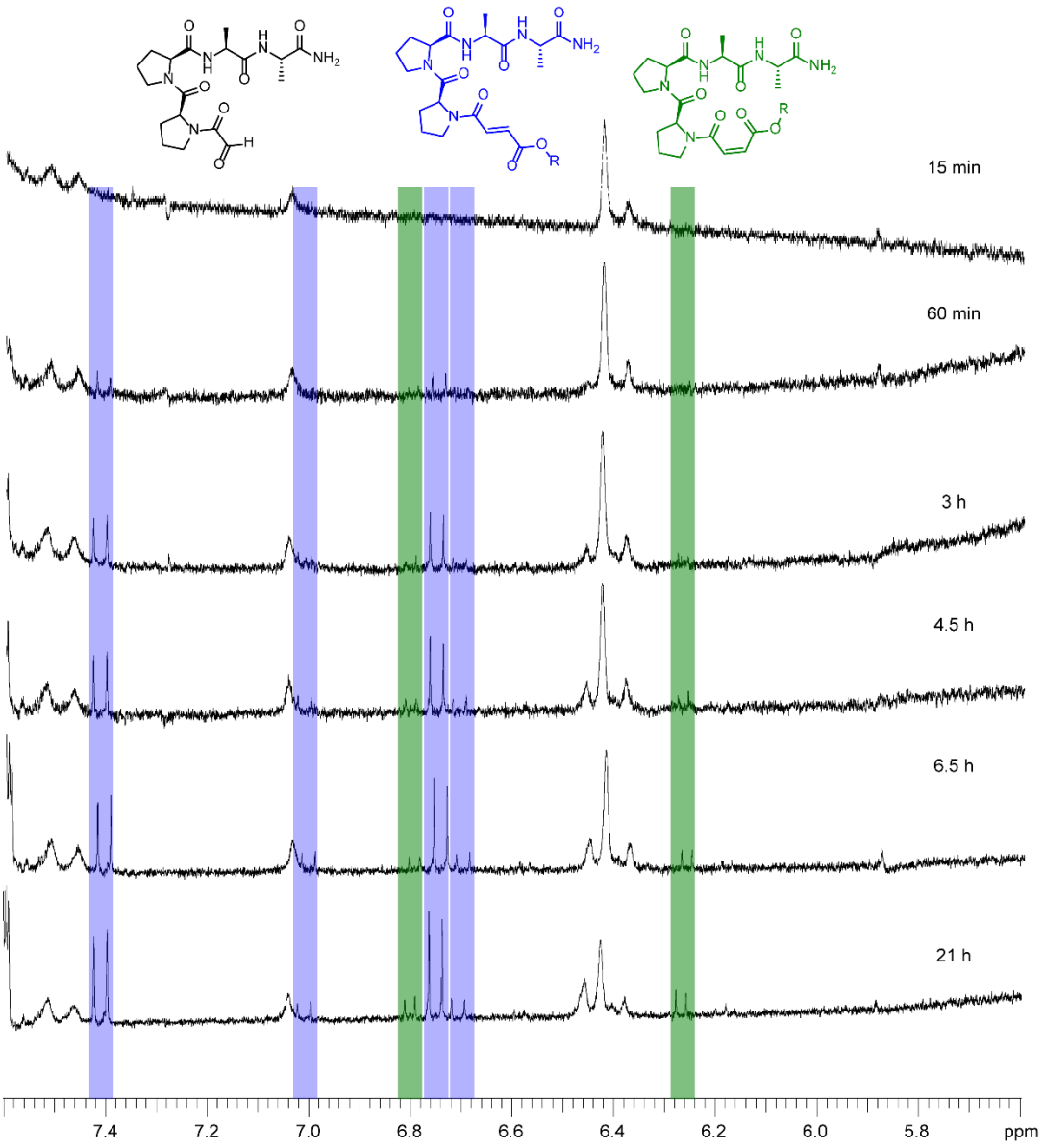
Appendix B-16. Kinetics of hydrazone ligation for Ald-PPAA at pH 5



Appendix B-17. E/Z selectivity for Ald-WWRR



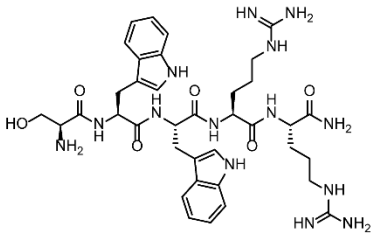
Appendix B-18. E/Z selectivity for Ald-HWFP



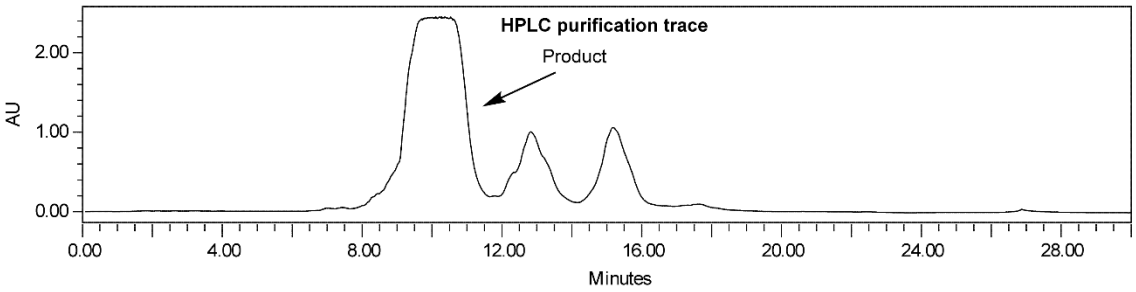
Appendix B-19. E/Z selectivity for Ald-PPAA

Appendix B-20. HPLC purity and LCMS traces of synthesized peptides

SWWRR



Chemical Formula: $C_{37}H_{52}N_{14}O_6$
Exact Mass: 788.42

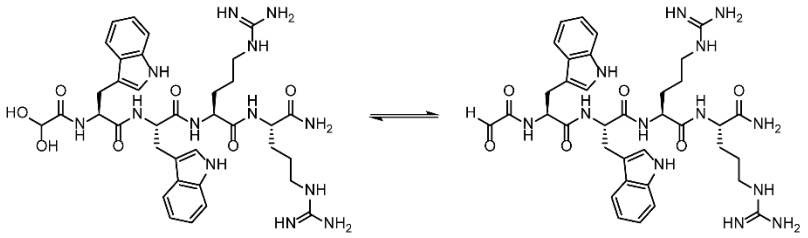


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

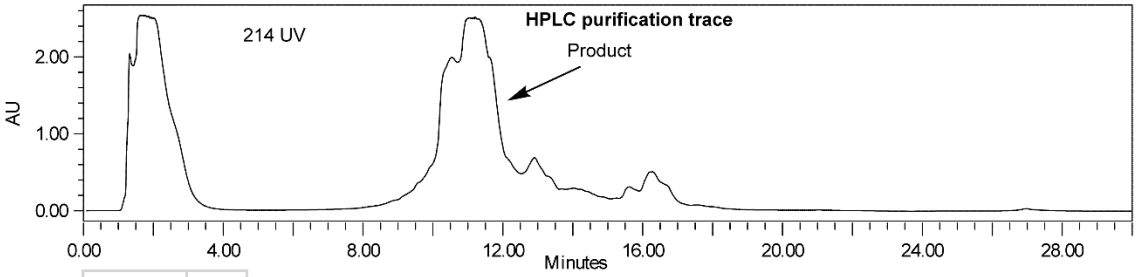
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

o-WWRR



Chemical Formula: $C_{36}H_{49}N_{13}O_7$
Exact Mass: 775.39

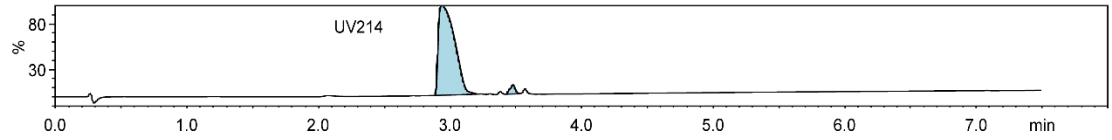
Chemical Formula: $C_{36}H_{47}N_{13}O_6$
Exact Mass: 757.38

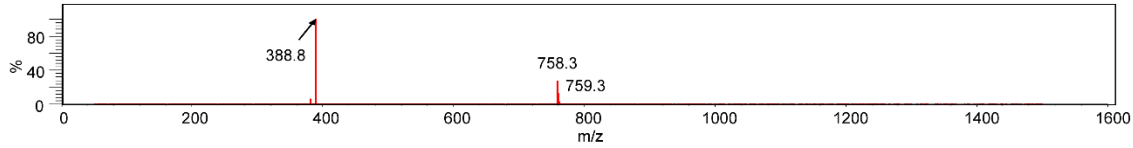


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

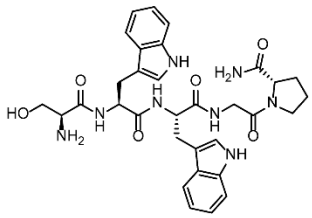
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

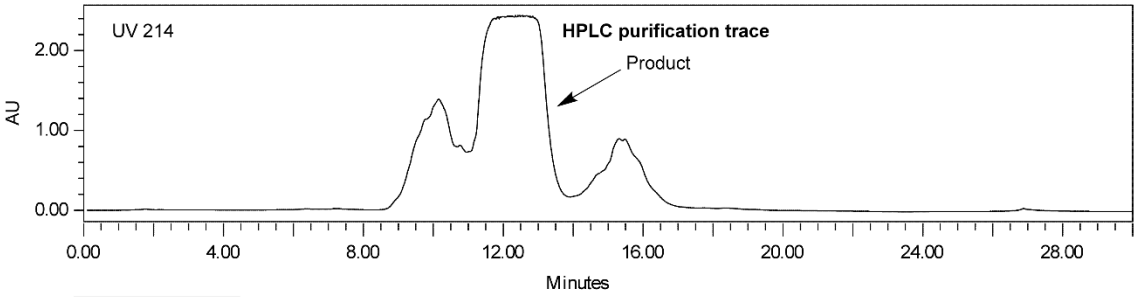




SWWGP



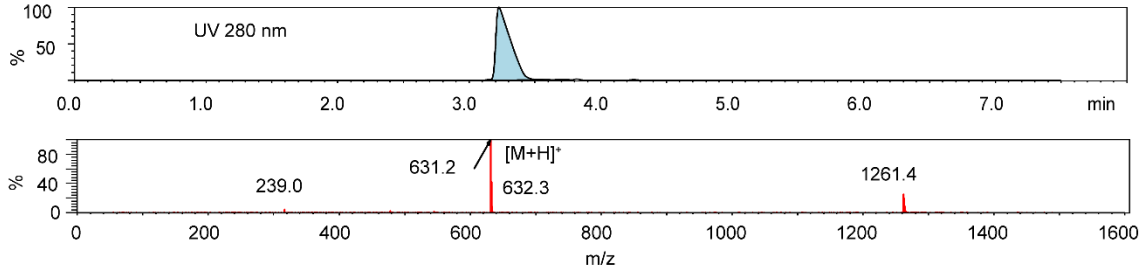
Chemical Formula: $C_{32}H_{38}N_8O_6$
Exact Mass: 630.29



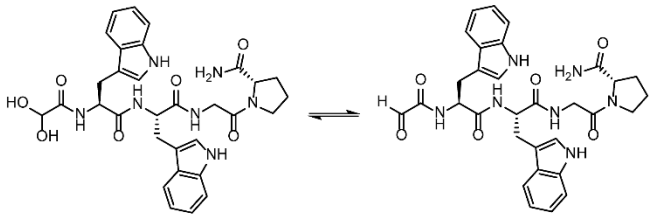
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

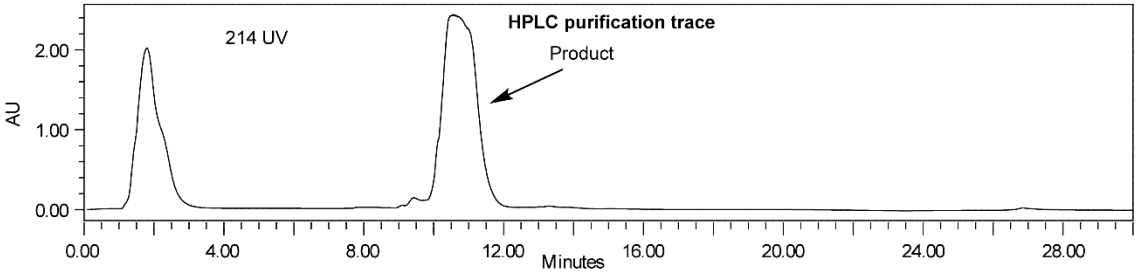


o-WWGP



Chemical Formula: $C_{31}H_{35}N_7O_7$
Exact Mass: 617.26

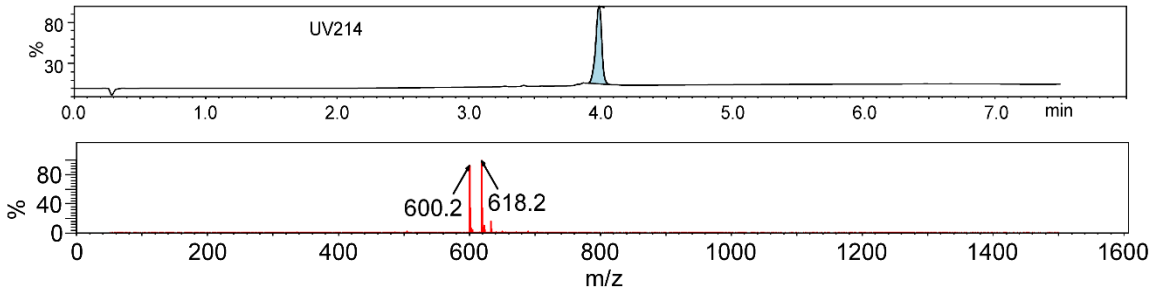
Chemical Formula: $C_{31}H_{33}N_7O_6$
Exact Mass: 599.25



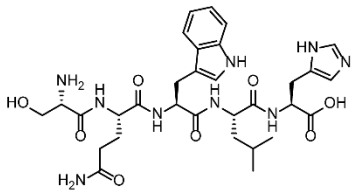
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

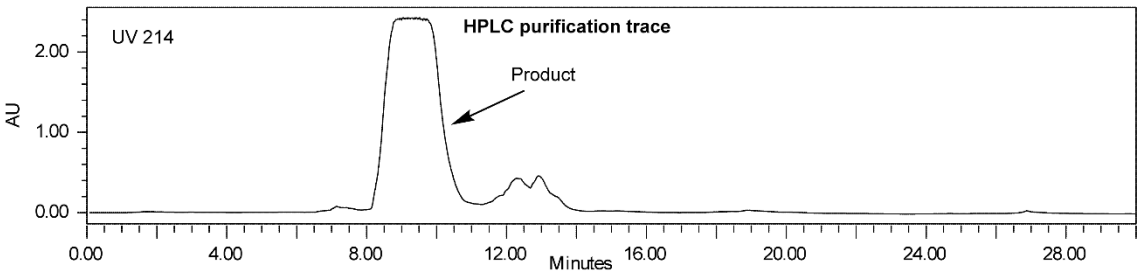
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



SQWLH



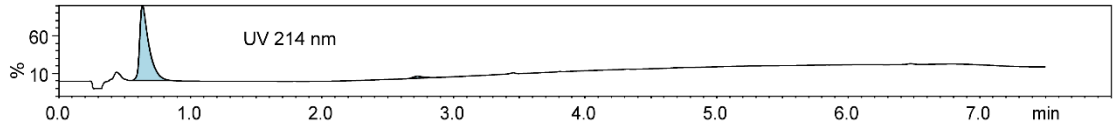
Chemical Formula: $C_{31}H_{43}N_9O_8$
Exact Mass: 669.32

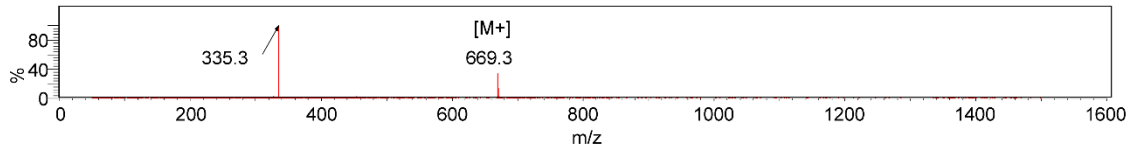


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

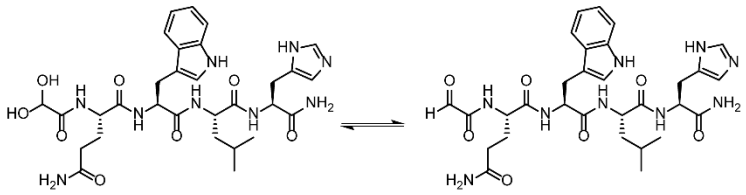
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



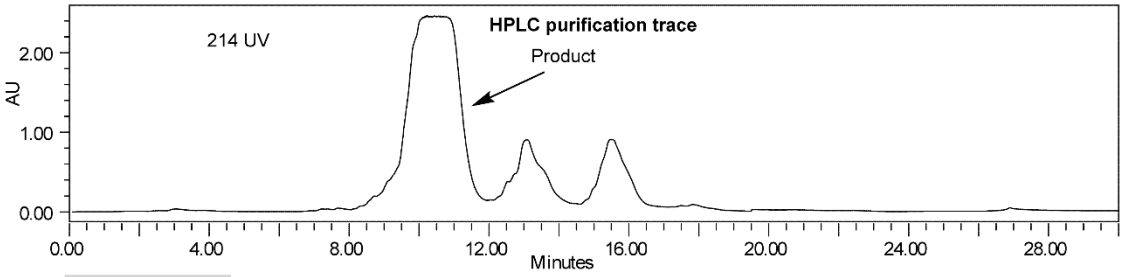


o-QWLH



Chemical Formula: $C_{30}H_{41}N_9O_8$
Exact Mass: 655.31

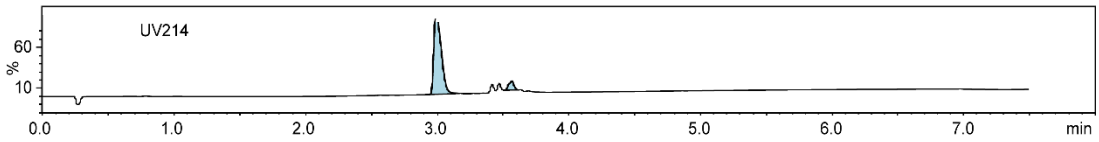
Chemical Formula: $C_{30}H_{39}N_9O_7$
Exact Mass: 637.30

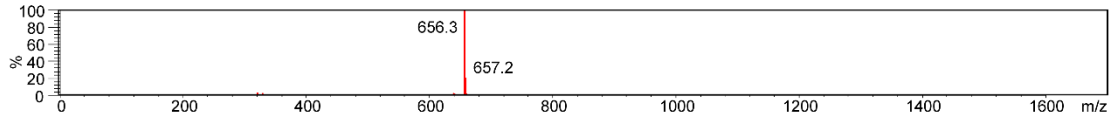


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

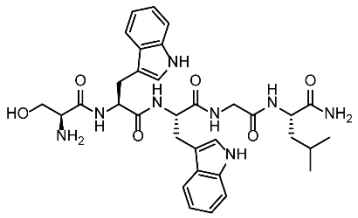
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

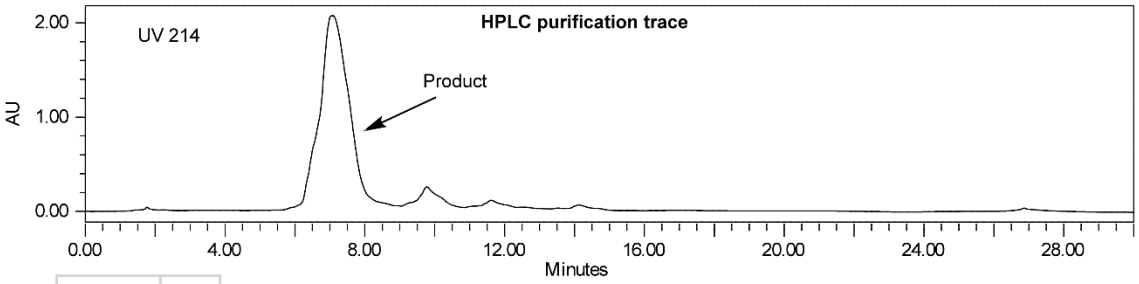




SWWGL



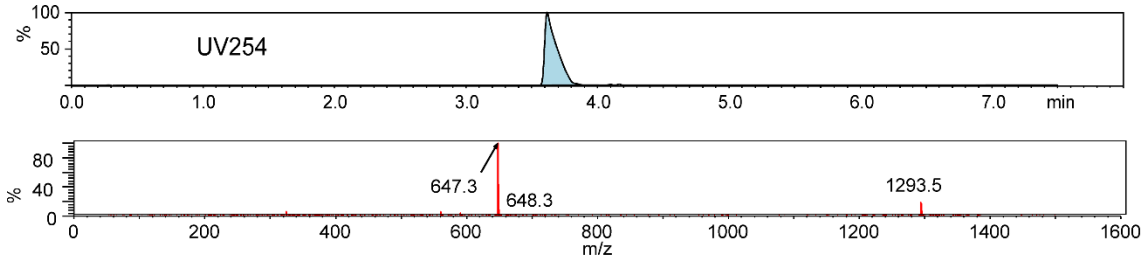
Chemical Formula: $C_{33}H_{42}N_8O_6$
Exact Mass: 646.32



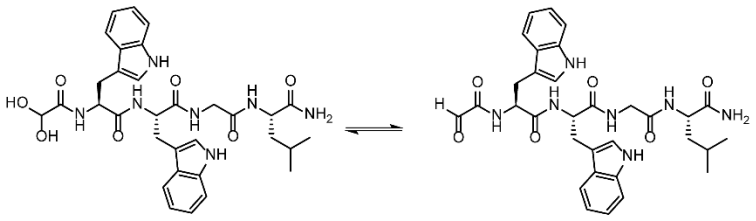
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

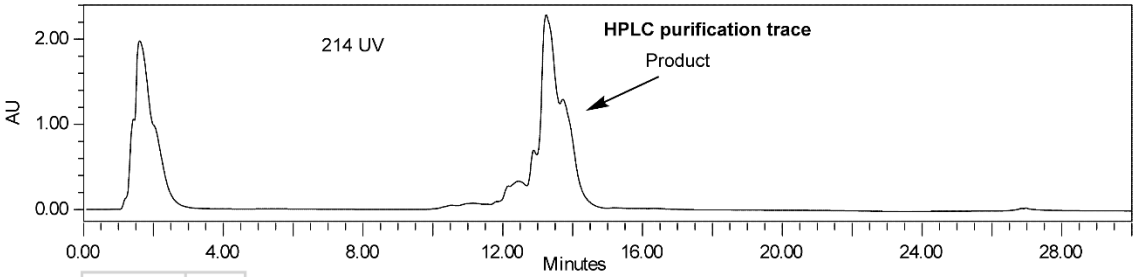


o-WWGL



Chemical Formula: $C_{32}H_{39}N_7O_7$
Exact Mass: 633.29

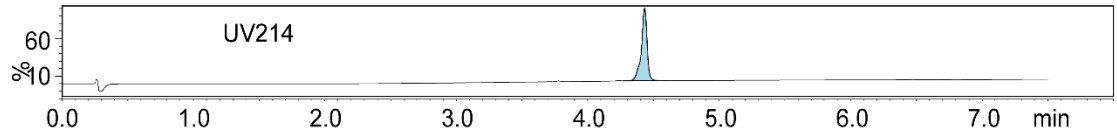
Chemical Formula: $C_{32}H_{37}N_7O_6$
Exact Mass: 615.28

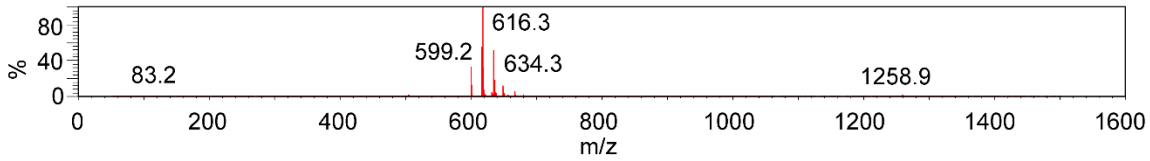


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

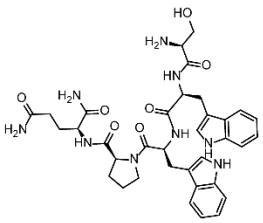
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

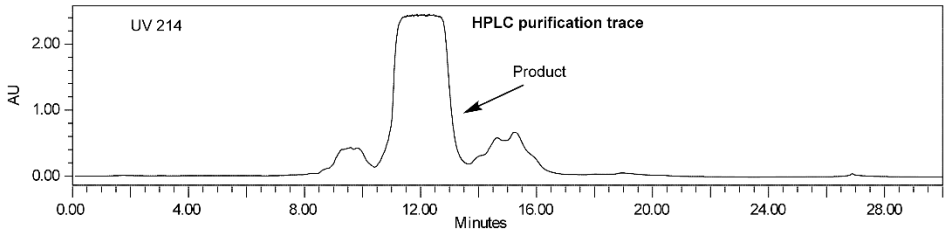




SWWPQ



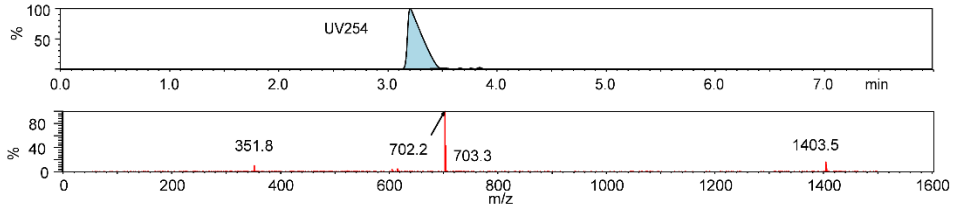
Exact Mass: 701.33
Molecular Weight: 701.79



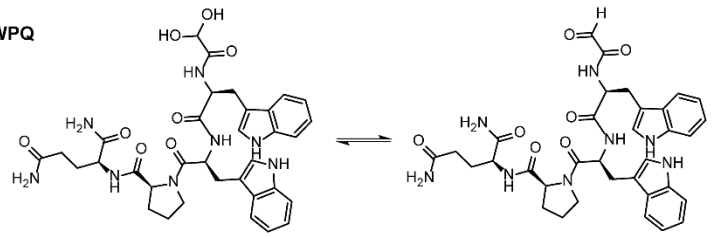
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

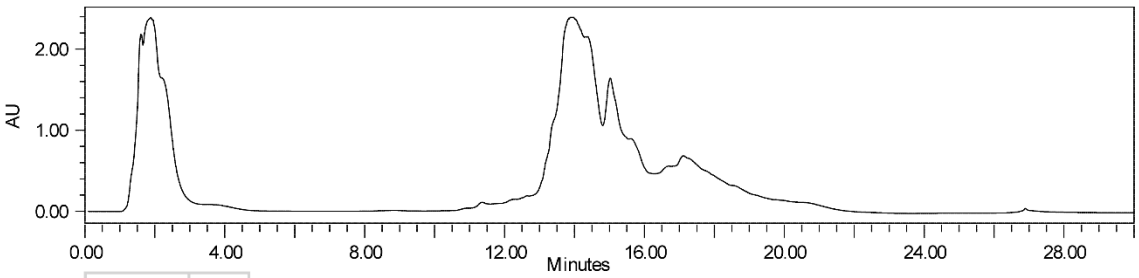


o-WWPQ



Chemical Formula: $C_{34}H_{40}N_8O_8$
Exact Mass: 688.30

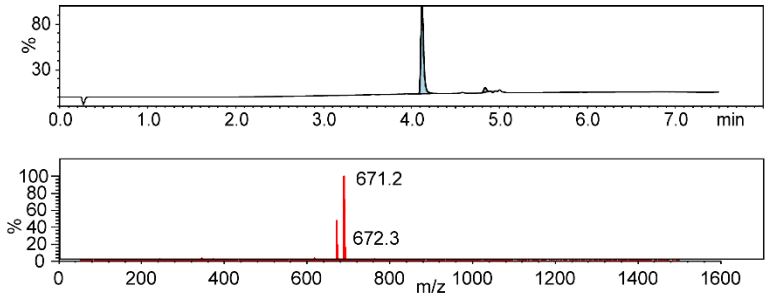
Chemical Formula: $C_{34}H_{38}N_8O_7$
Exact Mass: 670.29



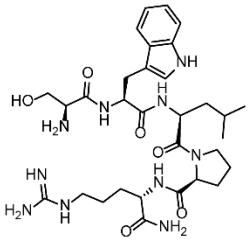
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

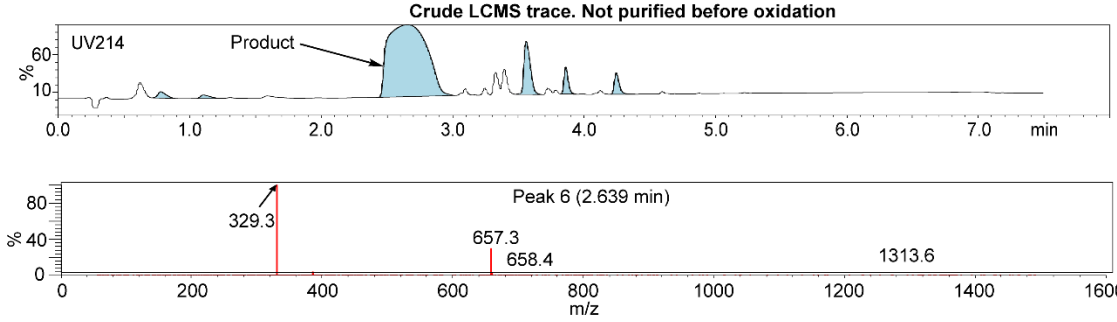


SWLPR

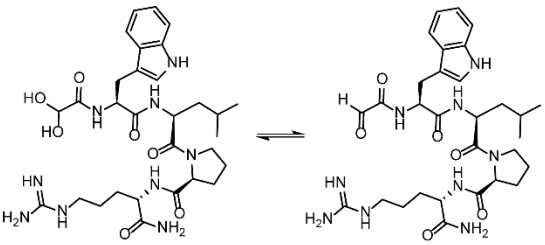


Chemical Formula: C31H48N10O6
Molecular Weight: 656.79

Crude LCMS trace. Not purified before oxidation

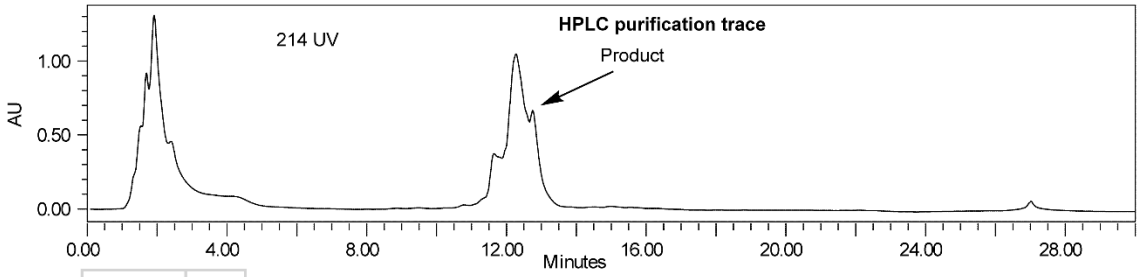


o-WLPR



Chemical Formula: $C_{30}H_{45}N_9O_7$
Exact Mass: 643.34

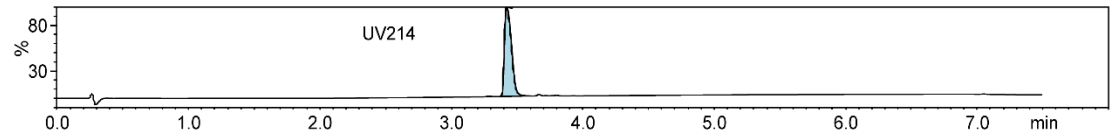
Chemical Formula: $C_{30}H_{43}N_9O_6$
Exact Mass: 625.33

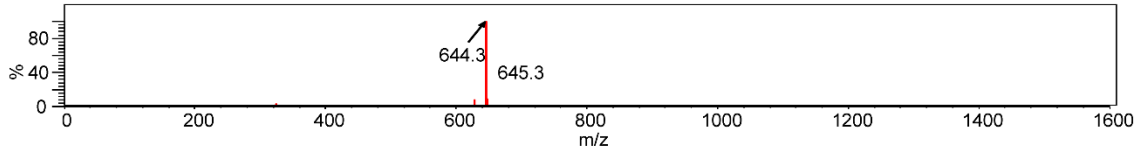


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

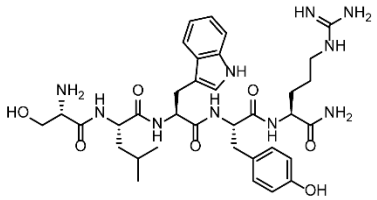
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



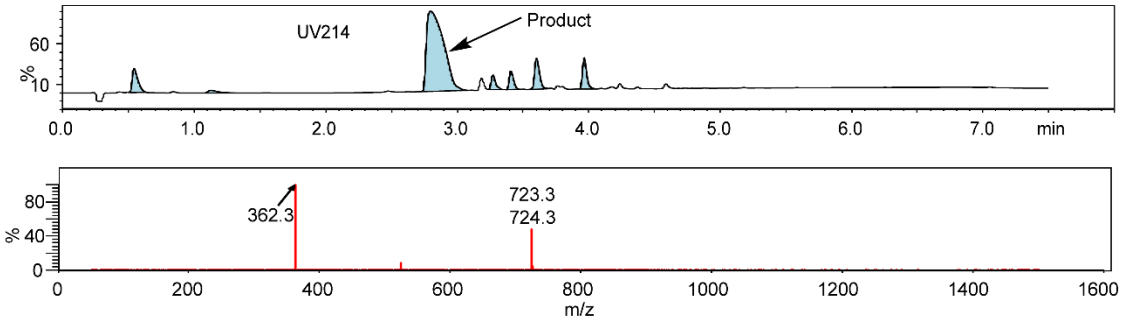


SLWYR

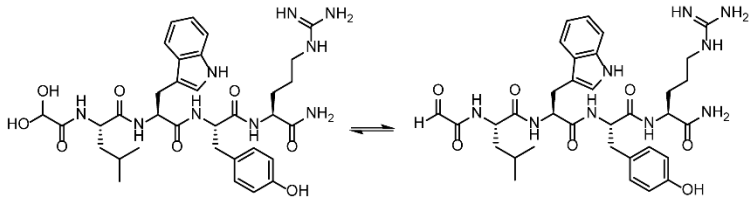


Chemical Formula: $C_{35}H_{50}N_{10}O_7$
Exact Mass: 722.39

Crude LCMS trace. Not purified before oxidation

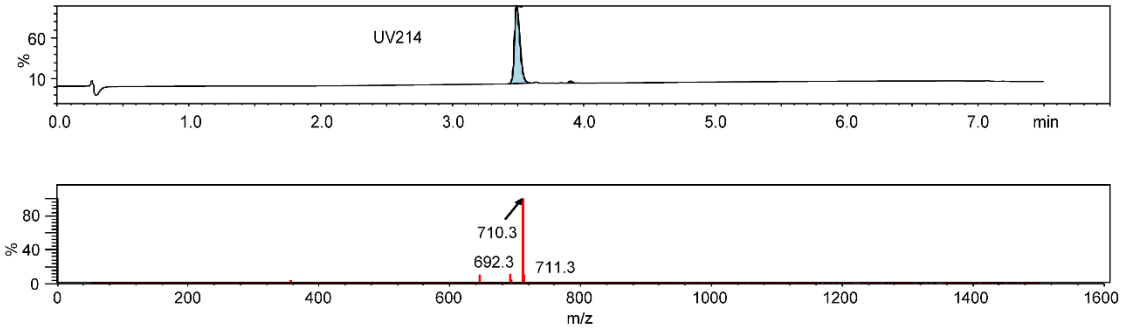


o-LWYR

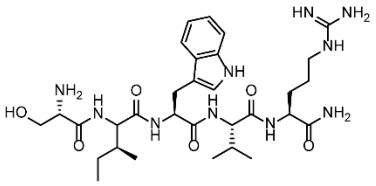


Chemical Formula: $C_{34}H_{47}N_9O_8$
Exact Mass: 709.35

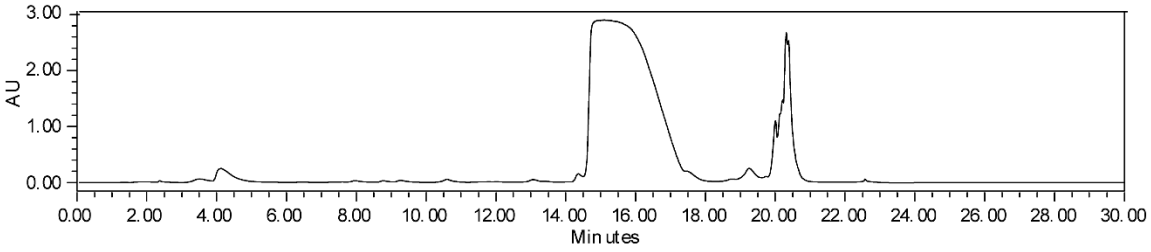
Chemical Formula: $C_{34}H_{45}N_9O_7$
Exact Mass: 691.34



SIWVR



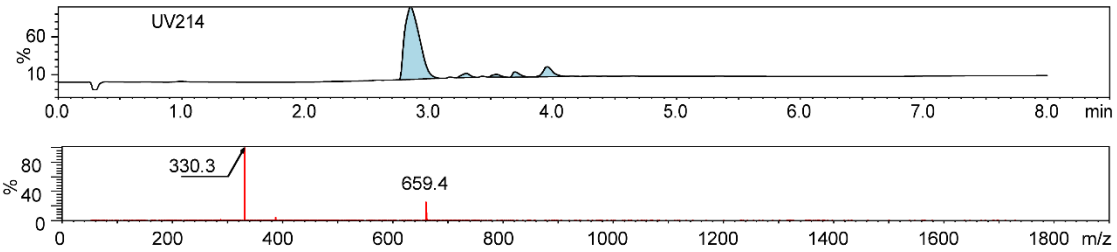
Chemical Formula: $C_{31}H_{50}N_{10}O_6$
Exact Mass: 658.39



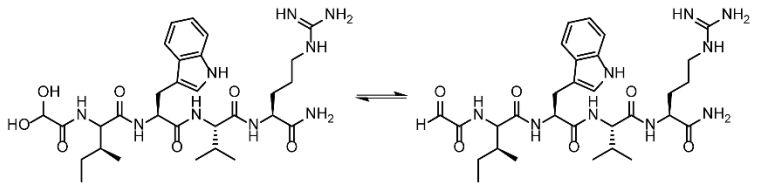
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

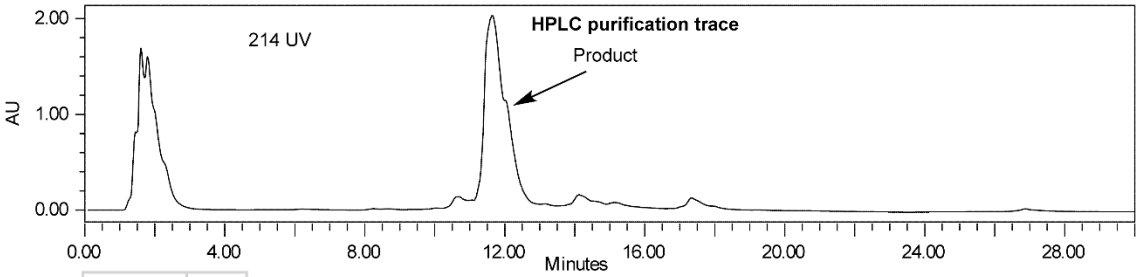


o-IWVR



Chemical Formula: $C_{30}H_{47}N_9O_7$
Exact Mass: 645.36

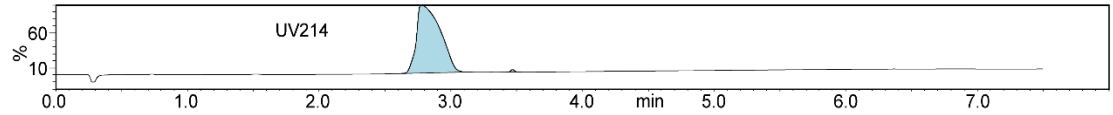
Chemical Formula: $C_{30}H_{45}N_9O_6$
Exact Mass: 627.35

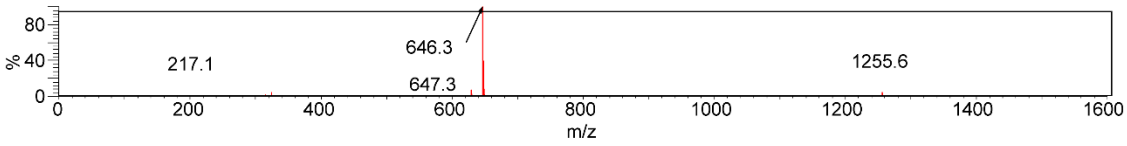


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

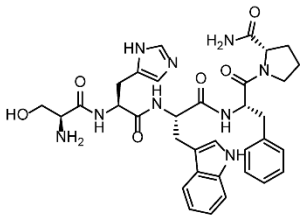
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

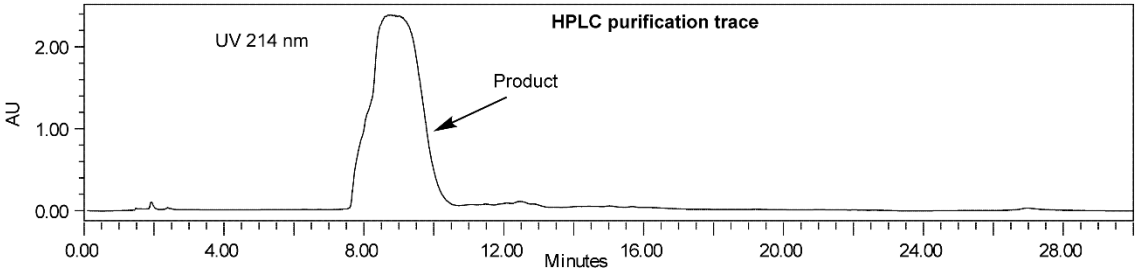




SHWFP



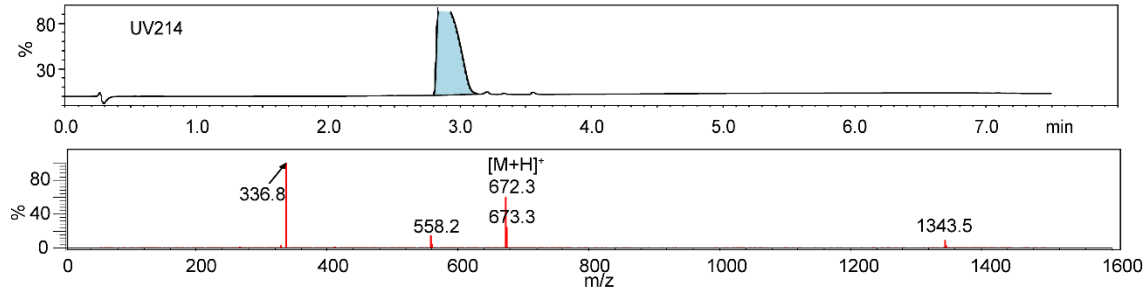
Chemical Formula: $C_{34}H_{41}N_9O_6$
Molecular Weight: 671.76



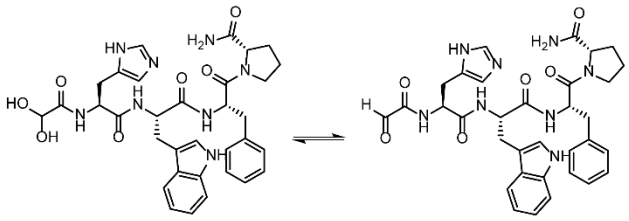
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

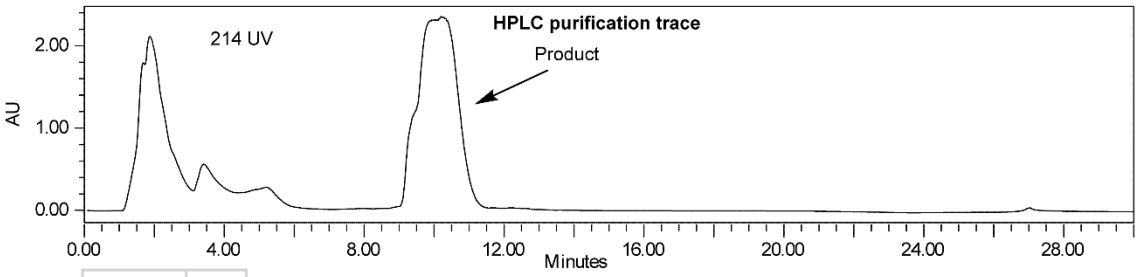


o-HWFP



Chemical Formula: $C_{33}H_{38}N_8O_7$
Exact Mass: 658.29

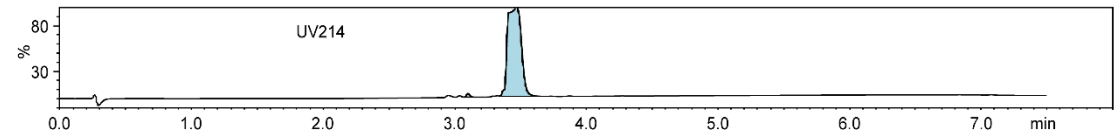
Chemical Formula: $C_{33}H_{36}N_8O_6$
Exact Mass: 640.28

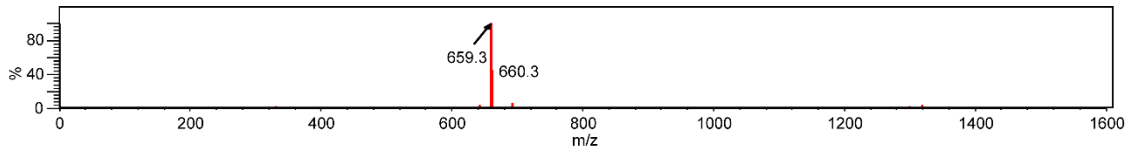


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

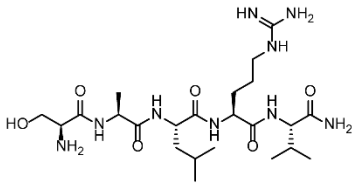
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

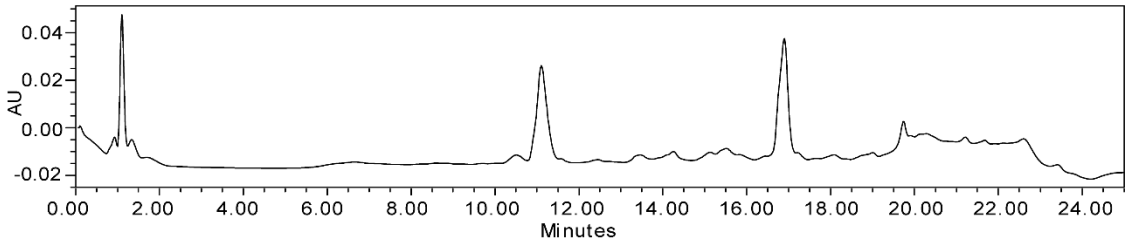




SALRV



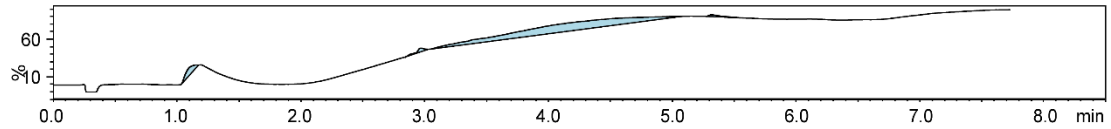
Chemical Formula: $C_{23}H_{45}N_9O_6$
Exact Mass: 543.35

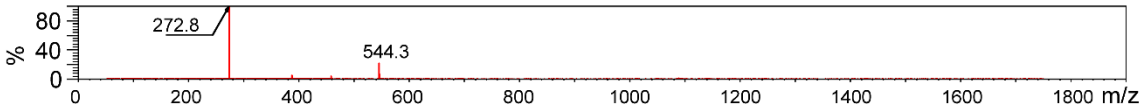


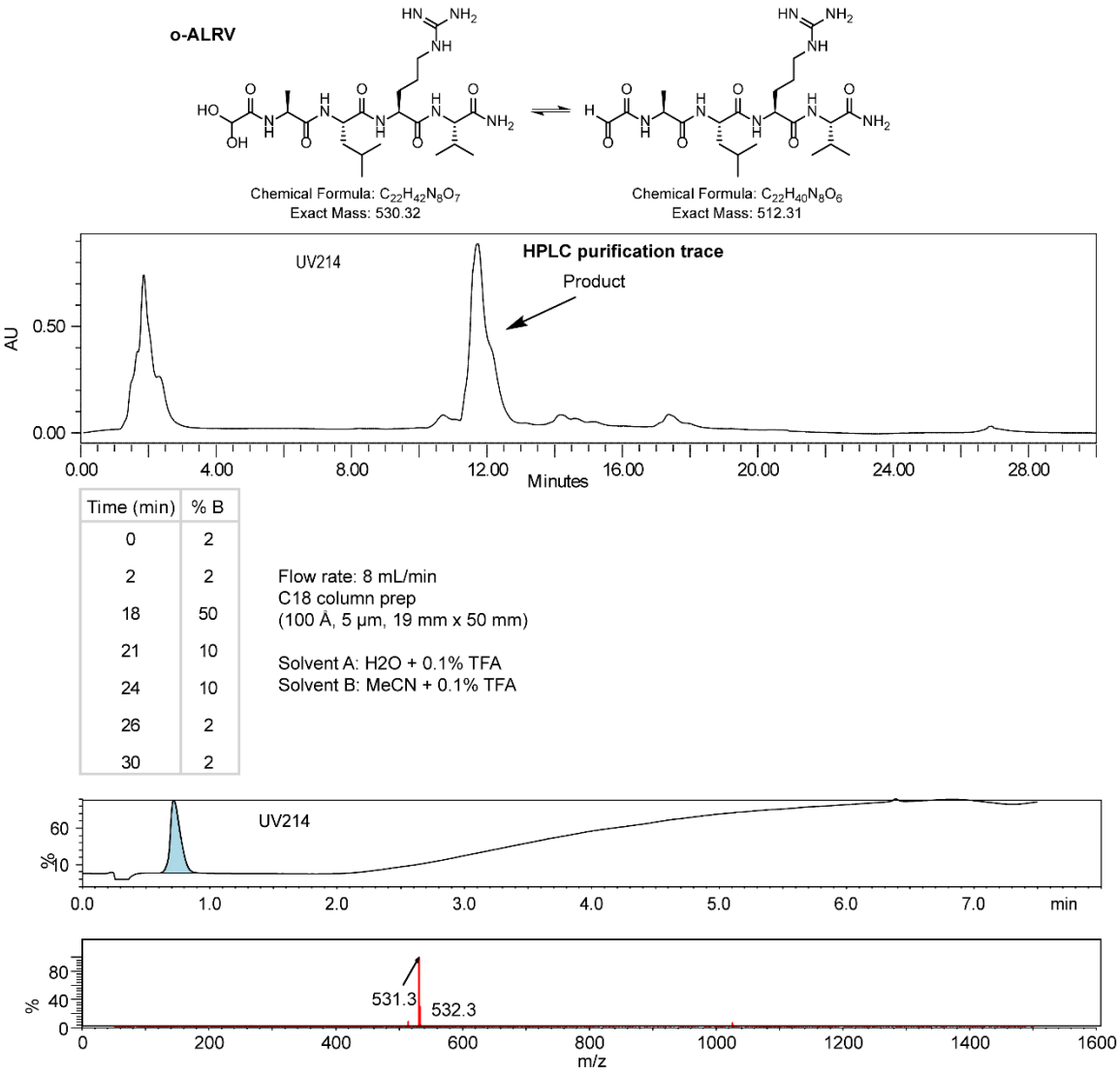
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

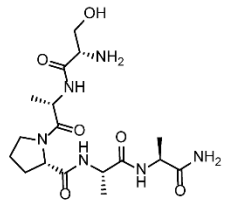
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA





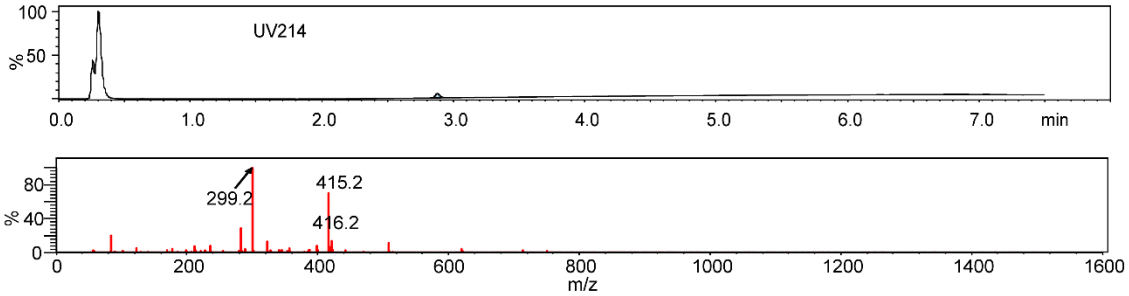

o-ALRV
Chemical Formula: $C_{22}H_{42}N_8O_7$
Exact Mass: 530.32
Chemical Formula: $C_{22}H_{40}N_8O_6$
Exact Mass: 512.31
UV214
HPLC purification trace
Product
AU
0.50
0.00
0.00
4.00
8.00
12.00
16.00
20.00
24.00
28.00
Minutes
Time (min) % B
0 2
2 2
18 50
21 10
24 10
26 2
30 2
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA
UV214
%
60
10
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
min
%
80
40
0
531.3
532.3
0
200
400
600
800
1000
1200
1400
1600
m/z

SAPAA

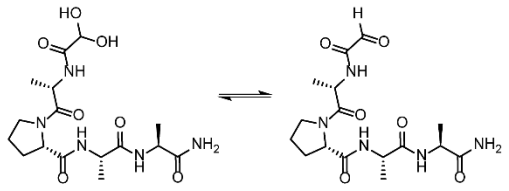


Chemical Formula: $C_{17}H_{30}N_6O_6$
Exact Mass: 414.22

Crude LCMS trace. Not purified before oxidation

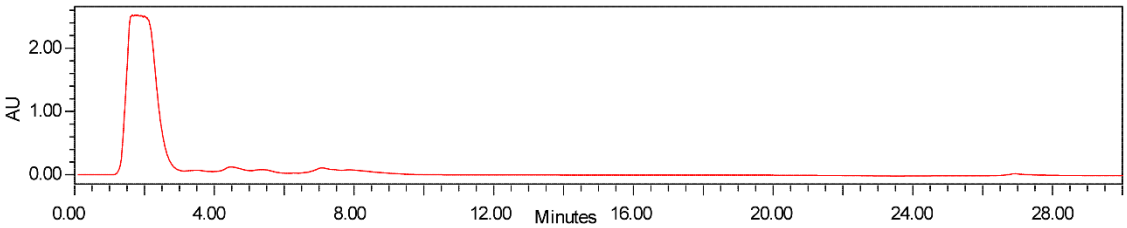


o-APAA



Chemical Formula: $C_{16}H_{27}N_5O_7$
Exact Mass: 401.19

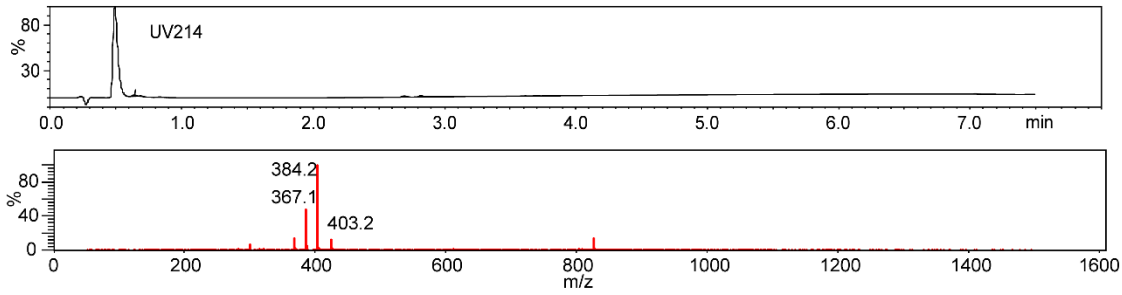
Chemical Formula: $C_{16}H_{25}N_5O_6$
Exact Mass: 383.18



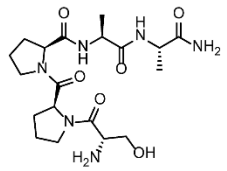
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

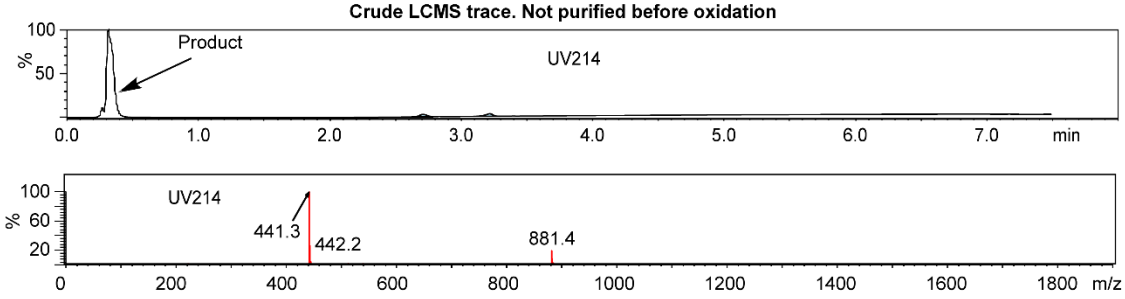


SPPAA

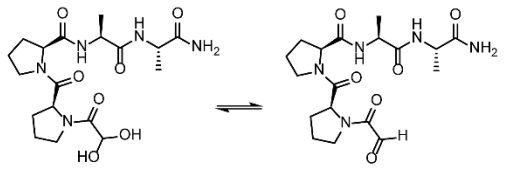


Chemical Formula: $C_{19}H_{32}N_6O_6$
Exact Mass: 440.24

Crude LCMS trace. Not purified before oxidation

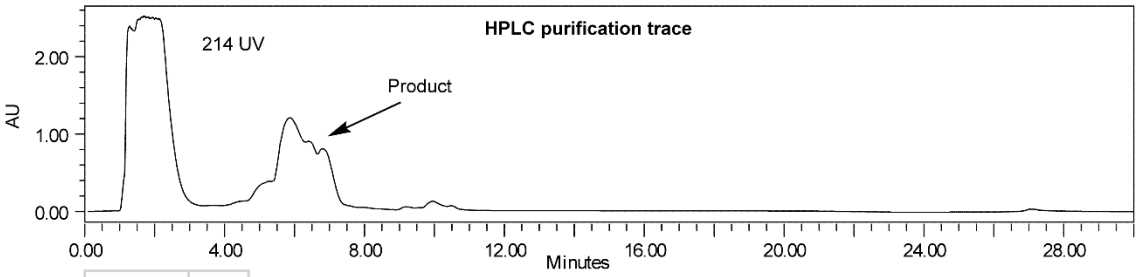


o-PPAA



Chemical Formula: $C_{18}H_{29}N_5O_7$
Exact Mass: 427.21

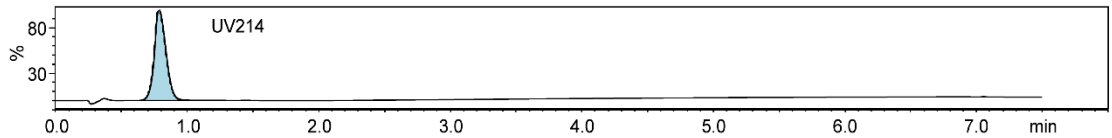
Chemical Formula: $C_{18}H_{27}N_5O_6$
Exact Mass: 409.20

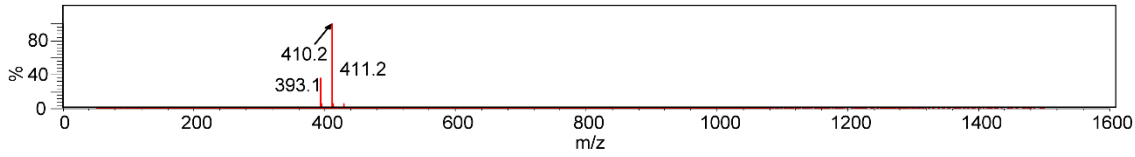


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

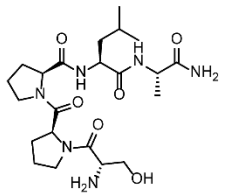
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



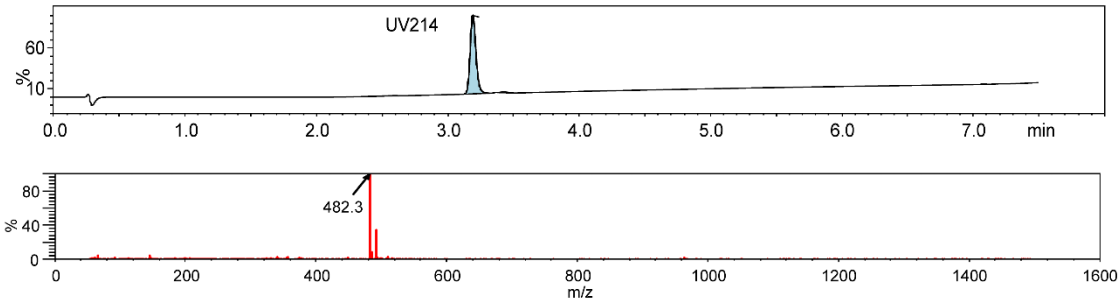


SPPLA

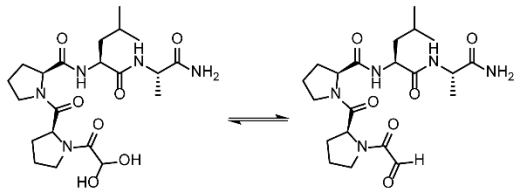


Chemical Formula: $C_{22}H_{38}N_6O_6$
Exact Mass: 482.29

Crude LCMS trace. Not purified before oxidation

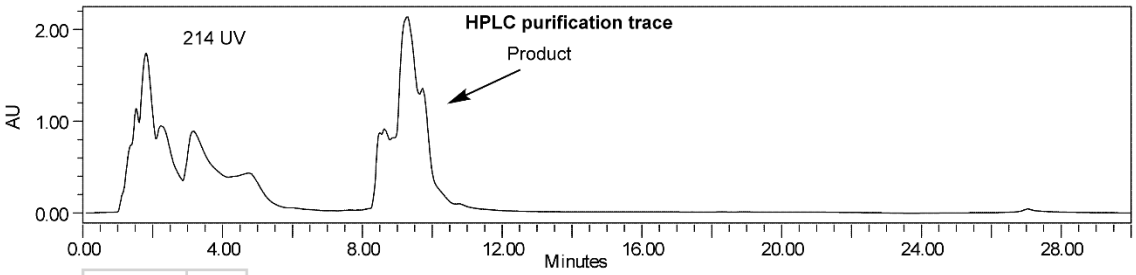


o-PPLA



Chemical Formula: $C_{21}H_{35}N_5O_7$
Exact Mass: 469.25

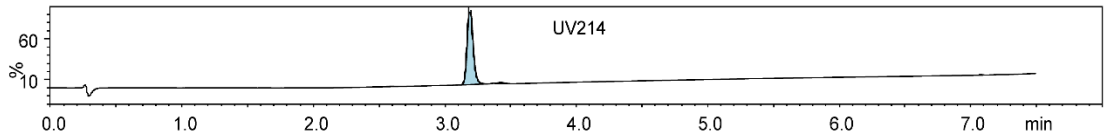
Chemical Formula: $C_{21}H_{33}N_5O_6$
Exact Mass: 451.24

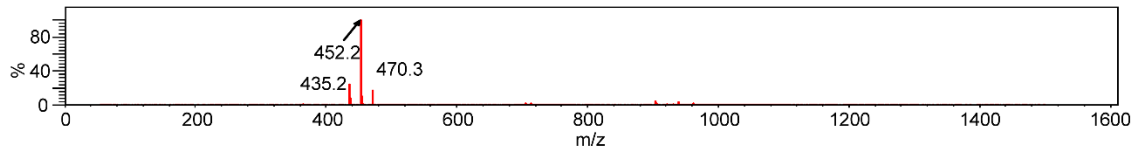


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

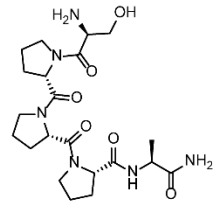
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



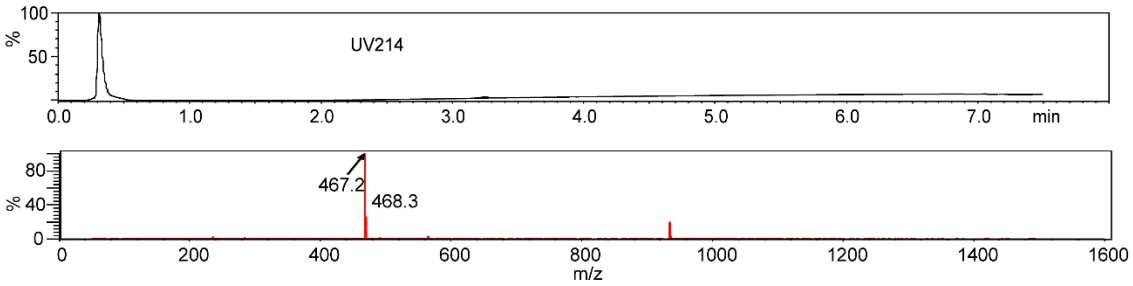


SPPPA

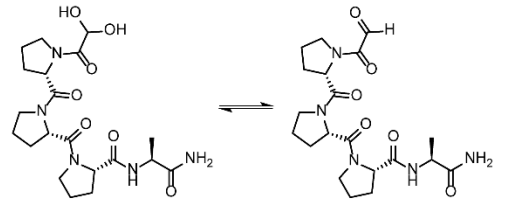


Chemical Formula: $C_{21}H_{34}N_6O_6$
Exact Mass: 466.25

Crude LCMS trace. Not purified before oxidation

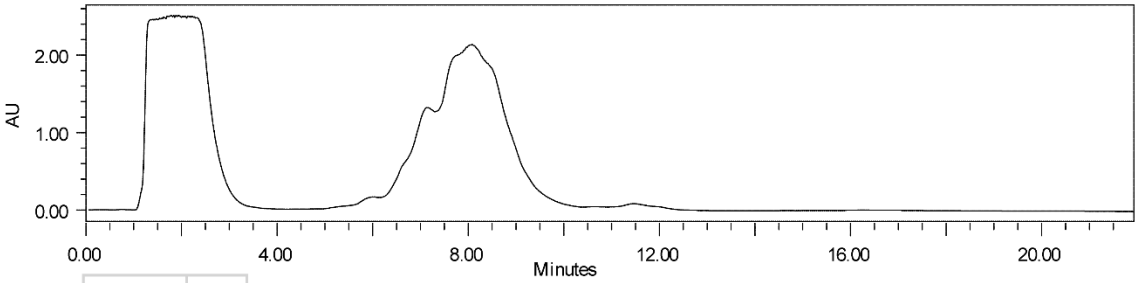


o-PPPA



Chemical Formula: $C_{20}H_{31}N_5O_7$
Exact Mass: 453.22

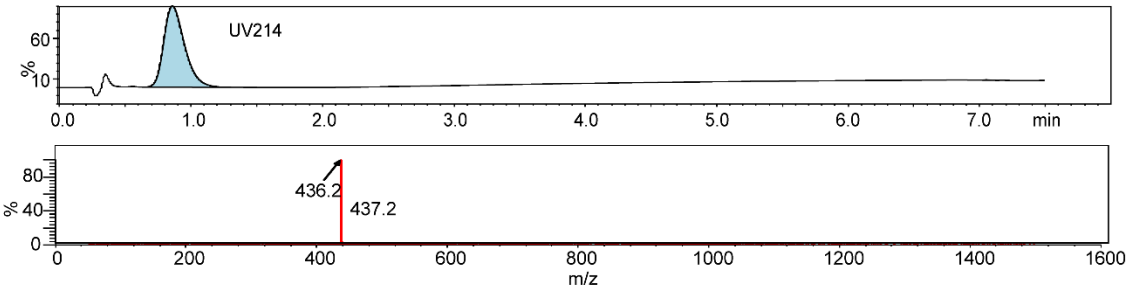
Chemical Formula: $C_{20}H_{29}N_5O_6$
Exact Mass: 435.21



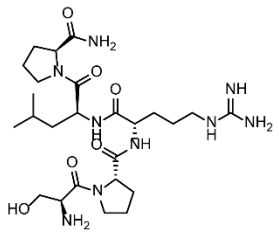
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

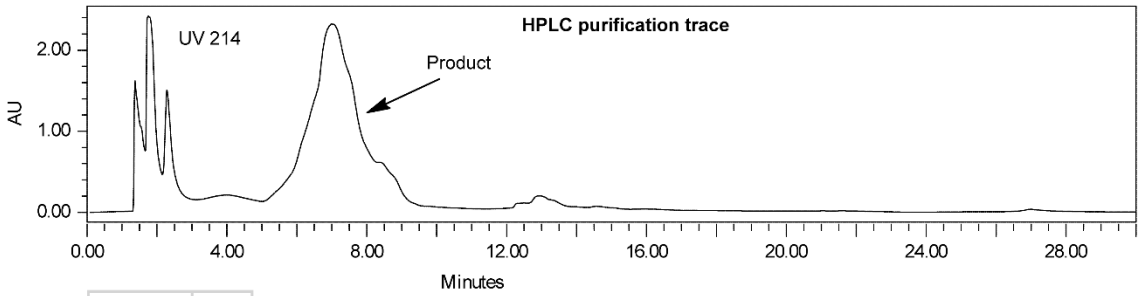
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



SPRLP



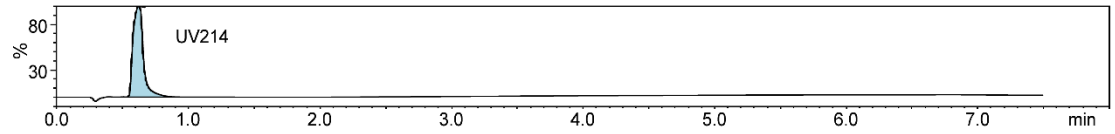
Chemical Formula: $C_{25}H_{45}N_9O_6$
Exact Mass: 567.35

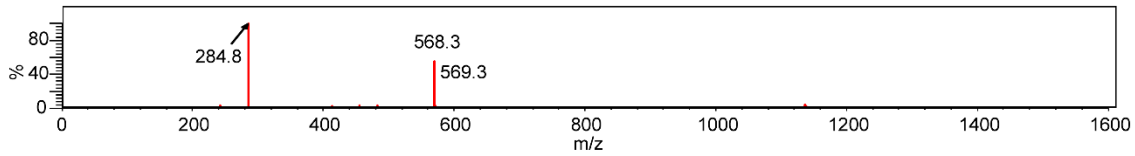


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

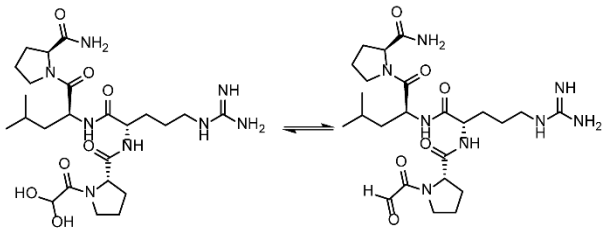
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



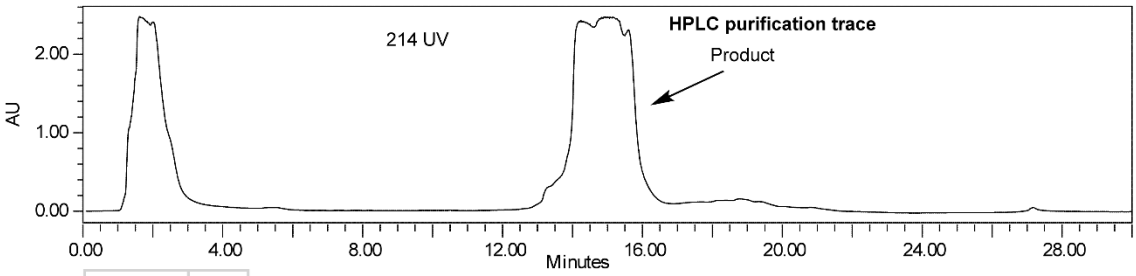


o-PRLP



Chemical Formula: $C_{24}H_{42}N_8O_7$
Exact Mass: 554.32

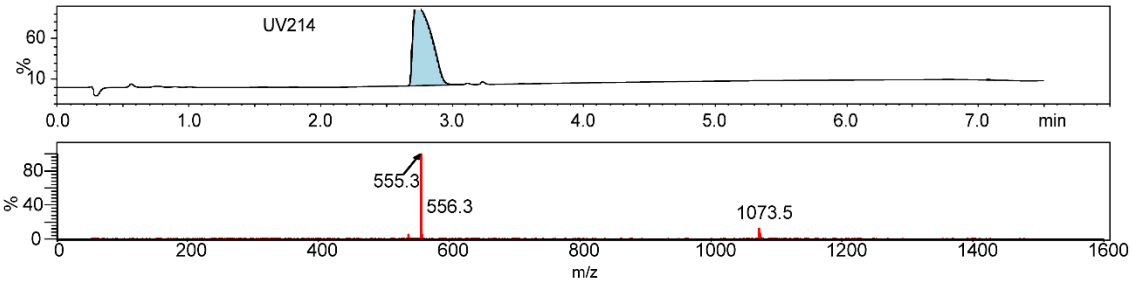
Chemical Formula: $C_{24}H_{40}N_8O_6$
Exact Mass: 536.31



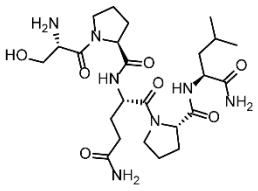
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

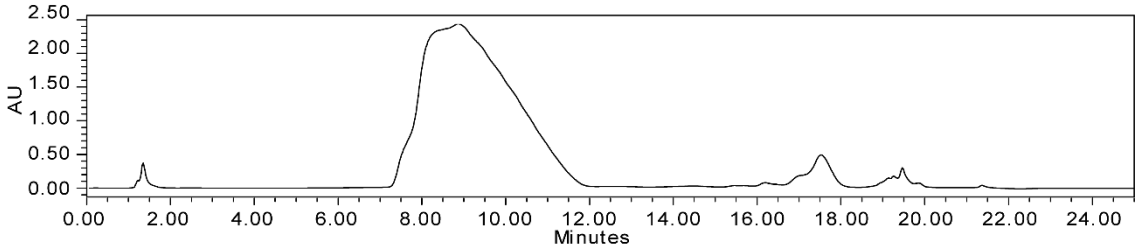
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



SPQPL



Chemical Formula: $C_{24}H_{41}N_7O_7$
Exact Mass: 539.31

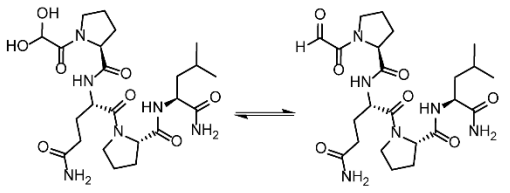


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

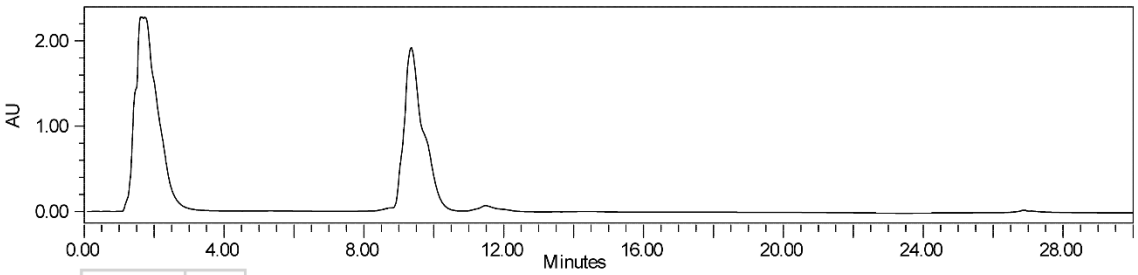
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

o-PQPL



Chemical Formula: $C_{23}H_{38}N_6O_8$
Exact Mass: 526.28

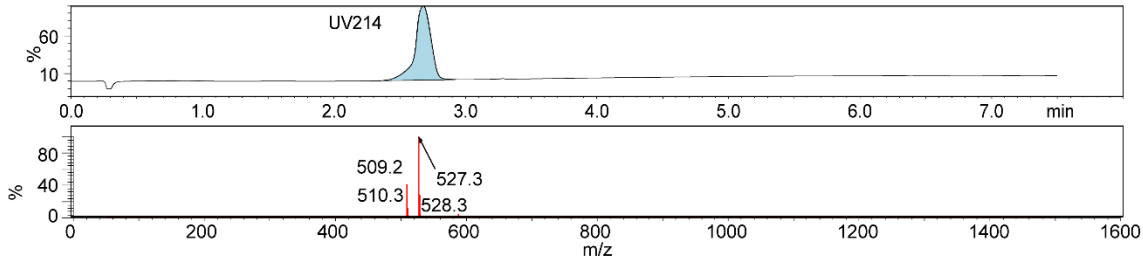
Chemical Formula: $C_{23}H_{36}N_6O_7$
Exact Mass: 508.26



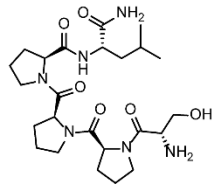
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

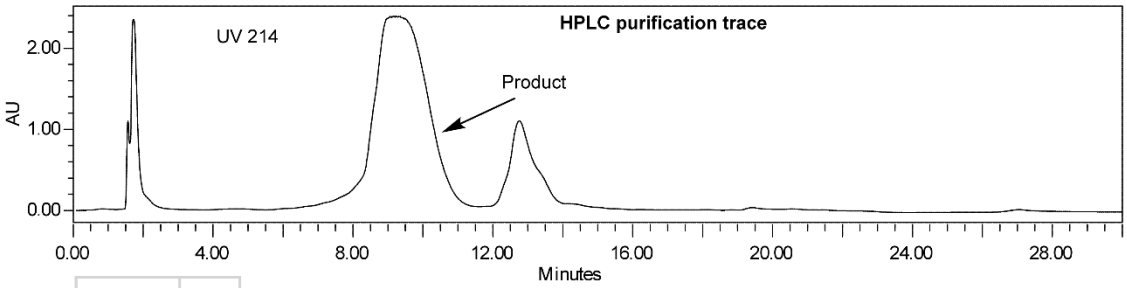
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



SPPPL



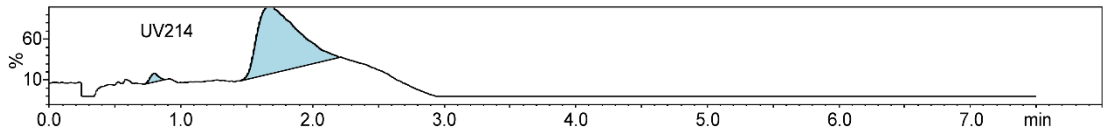
Chemical Formula: $C_{24}H_{40}N_6O_6$
Exact Mass: 508.30

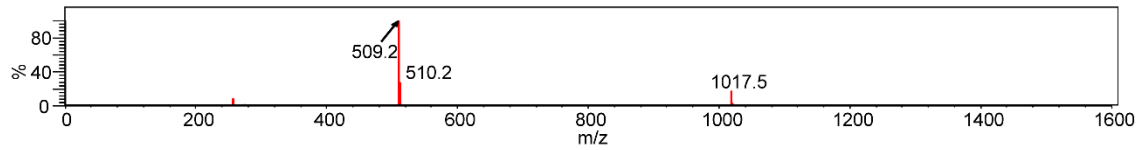


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

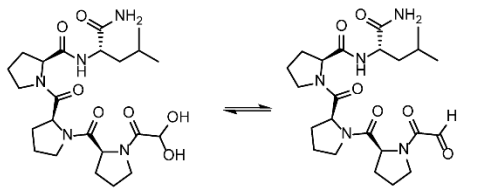
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



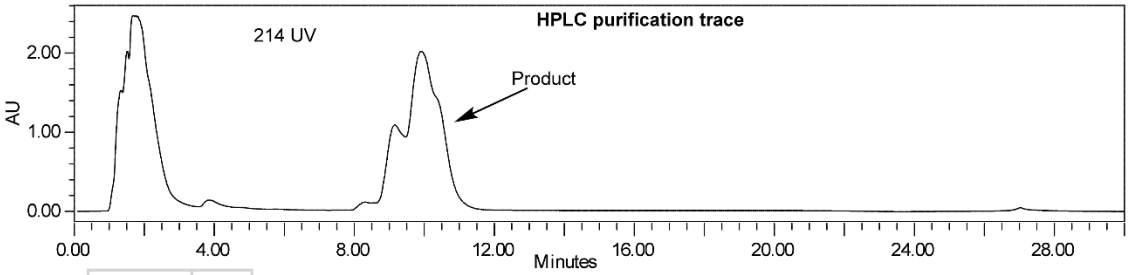


o-PPPL



Chemical Formula: $C_{23}H_{37}N_5O_7$
Exact Mass: 495.27

Chemical Formula: $C_{23}H_{35}N_5O_6$
Exact Mass: 477.26

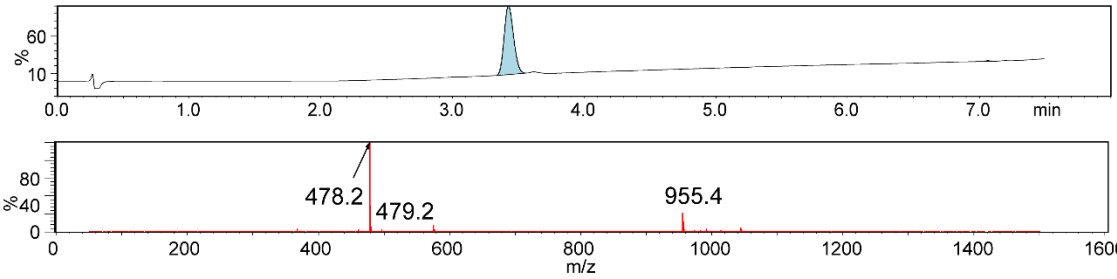


Minutes

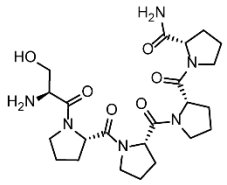
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H_2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

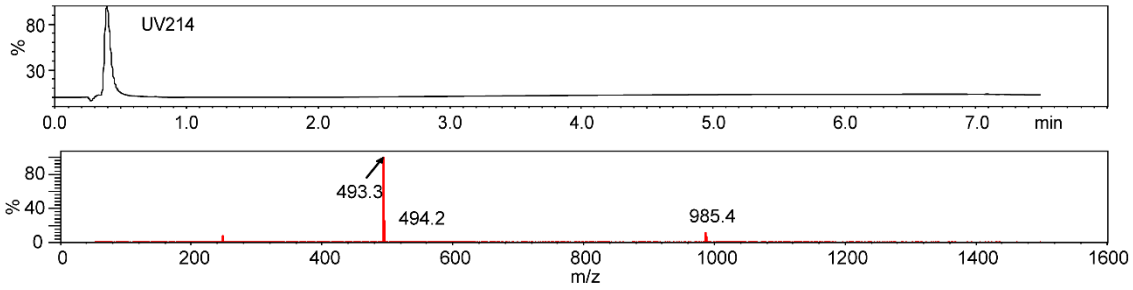


SPPPP

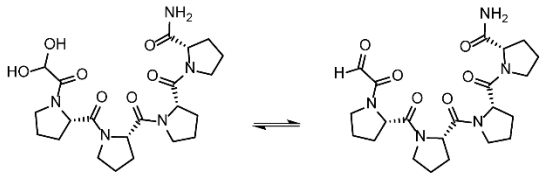


Chemical Formula: $C_{23}H_{36}N_6O_6$
Exact Mass: 492.27

Crude LCMS trace. Not purified before oxidation

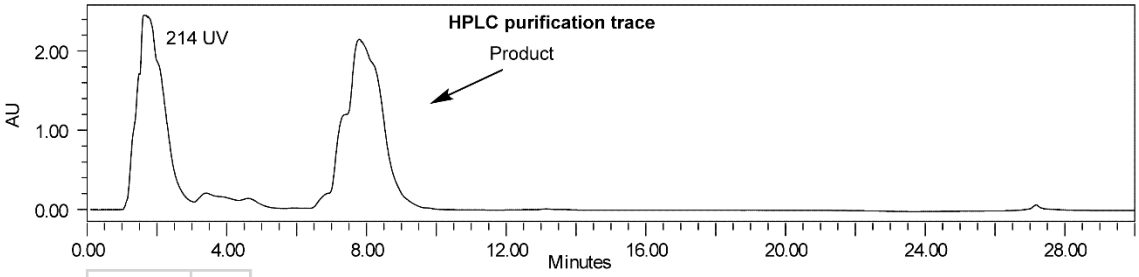


o-PPPP



Chemical Formula: $C_{22}H_{33}N_5O_7$
Exact Mass: 479.24

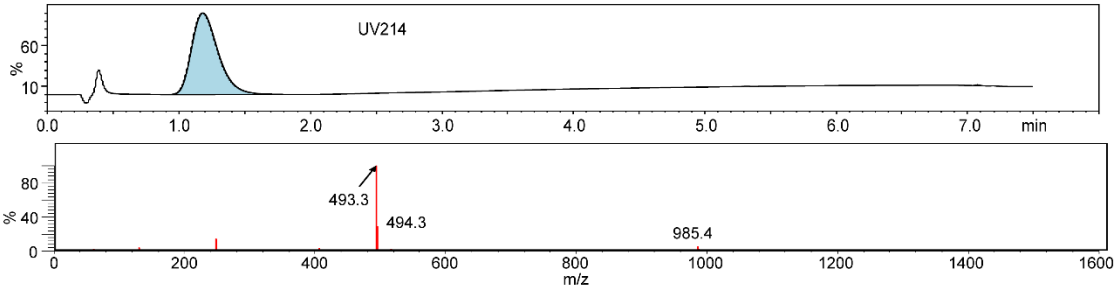
Chemical Formula: $C_{22}H_{31}N_5O_6$
Exact Mass: 461.23



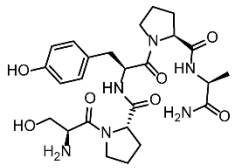
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

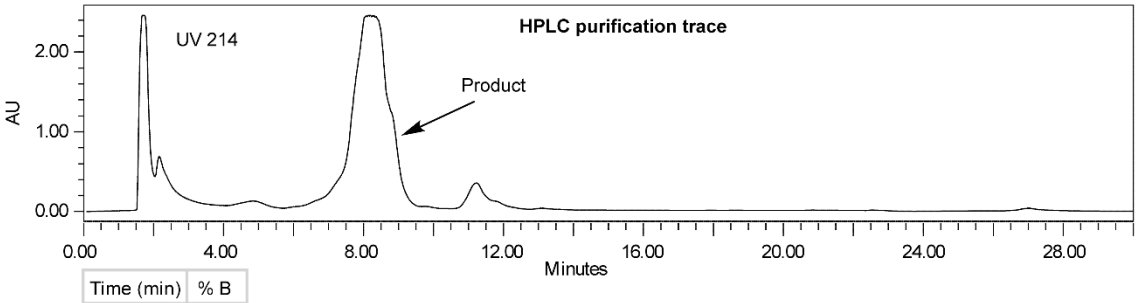
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



SPYPA



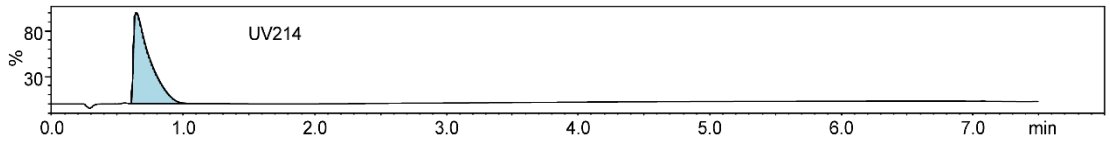
Chemical Formula: $C_{25}H_{36}N_8O_7$
Exact Mass: 532.26

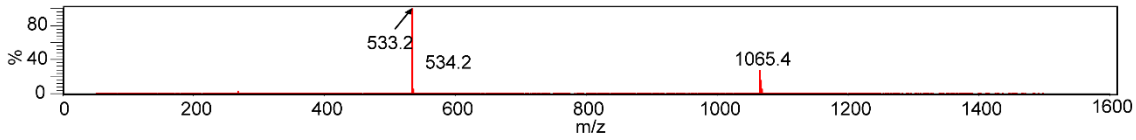


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

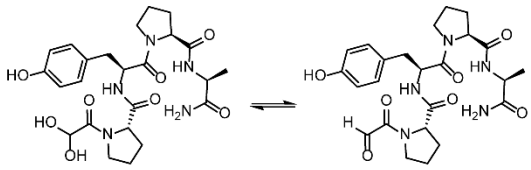
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



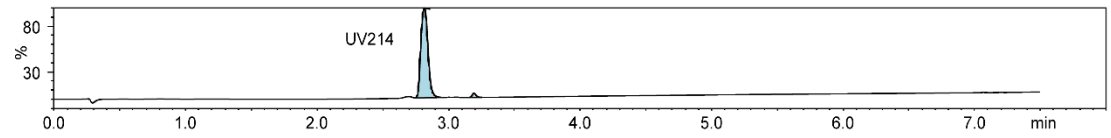


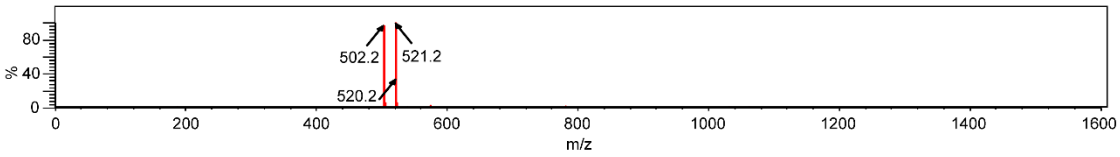
o-PYPA



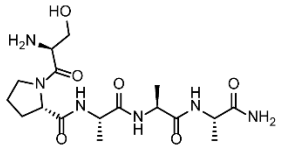
Chemical Formula: $C_{24}H_{33}N_5O_8$
Exact Mass: 519.23

Chemical Formula: $C_{24}H_{31}N_5O_7$
Exact Mass: 501.22



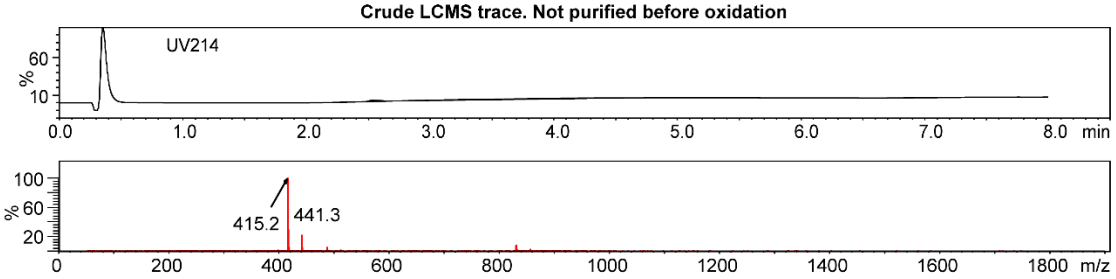


SPAAA

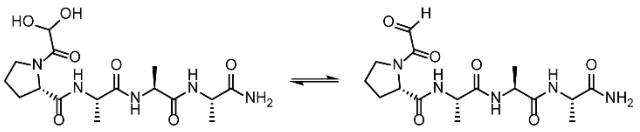


Chemical Formula: $C_{17}H_{30}N_6O_6$
Exact Mass: 414.22

Crude LCMS trace. Not purified before oxidation

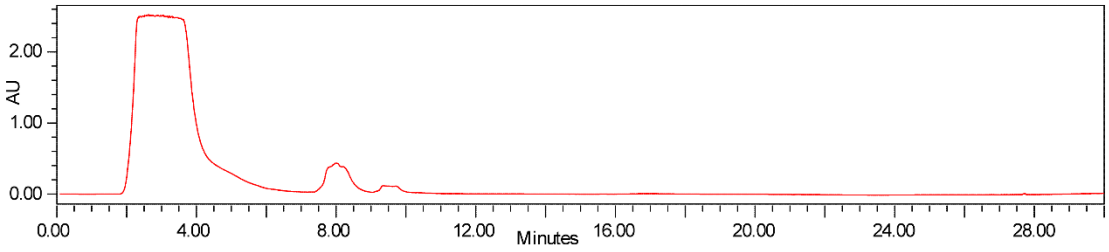


o-PAAA



Chemical Formula: $C_{16}H_{27}N_5O_7$
Exact Mass: 401.19

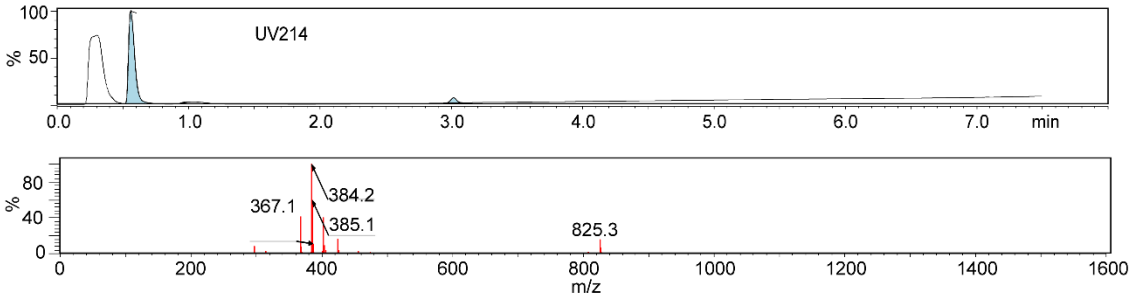
Chemical Formula: $C_{16}H_{25}N_5O_6$
Exact Mass: 383.18



Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



Appendix B-21: MATLAB script for DE Analysis

```
clear all; close all;
clc
Dir='';
File = '';
%SAVEto = [File(1:end-4) ''];  % keep blank if don't want to save
SET{1} = [19 20 21];        %
SET{2} = [22 23 24];        %
SET{3} = [25 26 27];        %
SET{4} = [28 29 30];
SET{5} = [31 32 33];
SET{6} = [34 35 36];
TEST_SET = 4;
CONTROL_SETS = [1];
HITS2DISPLAY = 20; % maximum numer of hits to display
SHOWaminoACIDS = [1 2 3 4 5];
CLUSTERbyH = 1;  % if you want your hits to be clustered by Hamming dist.
PLOT_VOLCANO = 1; % set to 1 if you want to see the actual volcano plot
%%%%%%% volcano plot parameters here
p_cutoff = 0.05;          % p-value cutoff  (use 0.9 if dont care abt p)
R_cutoff = 5;             % ratio cutoff
MaxX=14;                  % maximum on the X-scale (if plotting volcano)
vert_cutoff = 0.00001;    % maximum on the Y-scale (if plotting volcano)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%% do not change things beyond this point
%%%%%%%%%%%%%%%%%%%%%%%% unless you know what you are doing
REREAD=1;
if REREAD
[Nuc, AA, Fr] = readMulticolumn('Dir', Dir, 'File', File, ...
                    'column', 1:max(cell2mat(SET)),...
```

```
                    'skip', 2, 'output', 'normalized+1');

end
% select only the aminoacids you want to see
cAA = char(AA);
AA=cellstr(cAA(:,SHOWaminoACIDS));
SQUARE=zeros(size(Fr,1),1);
IX=zeros(size(Fr,1),numel(CONTROL_SETS));
i=0;
for j=CONTROL_SETS
   i=i+1;
   ratio(:,i) = mean(Fr(:,SET{TEST_SET}), 2) ./ mean(Fr(:,SET{j}), 2);
   [~,confi(:,i)] = ttest2(Fr(:,SET{TEST_SET})',Fr(:,SET{j})',....
                    p_cutoff,'both','unequal');
   IX(:,i) = (confi(:,i) <= p_cutoff) & (ratio(:,i) >= R_cutoff);
   SQUARE = SQUARE + ratio(:,i).^2;
   if PLOT_VOLCANO
      subplot(1,numel(CONTROL_SETS),i);
      plot(log2 (ratio(:,i)),...
         -log10(confi(:,i)),'d',...
         'MarkerSize',4,...
         'MarkerFaceColor',0.5*[1 1 1],...
         'MarkerEdgeColor',0.5*[1 1 1]); hold on;
      plot( log2 (ratio(find(IX(:,i)),i)),...
          -log10(confi(find(IX(:,i)),i)),'d',...
              'MarkerSize',4,...
              'MarkerFaceColor','r',...
              'MarkerEdgeColor','r'); hold on;
      line([log2(R_cutoff) MaxX],[-log10(p_cutoff) -log10(p_cutoff)]);
      line([ log2(R_cutoff)   log2(R_cutoff)],...
         [-log10(p_cutoff) -log10(vert_cutoff)]);
```

```
        xlim([-MaxX MaxX]);
    end
end
    R2 = sqrt(SQUARE);
   IXall = find((sum(IX,2)==size(IX,2))); %hits that satisfy all criteria
    %IXall = find( (sum(IX,2)>=5) ); %hits that satisfy 5 criteria
    hits    = char(AA(IXall,:));
    Rhits   = ratio(IXall,:);
    R2hits  = R2(IXall);
%%%%%%%%%%%%    this    is    part    where    hits    are    clustered    by    H-dist
%%%%%%%%%%%%%%
% figure(2)
% if size(hits,1)>3
%     Y = pdist(hits,'hamming');
%     Z = linkage(Y,'complete');
%     [H,T,perm] = dendrogram(Z,0,'colorthreshold',20);
%     set(H,'LineWidth',2)
%     for i =1:size(hits,1)
%         label{i} = i;
%     end
%     set(gca,'XTick', 1:1:size(hits,1), 'XTickLabel',label);
%     hits    = hits(perm,:);
%     Rhits   = Rhits(perm,:);
%     R2hits  = R2(perm);
%     IXall   = IXall(perm);
% end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% display all results as heat map
figure(3)
if size(IXall,1)>=HITS2DISPLAY
```

```
   N=HITS2DISPLAY;  % display only the first or defined number of hits
else
   N=size(IXall,1); %display all
end
FrPPM = round(10^6*Fr);   % convert normalized fraction frequency to PPM
imagesc( log10([FrPPM(IXall(1:N),:) ratio(IXall(1:N),:) ]+1) );
set(gca,'YTick', 1:1:N, 'YTickLabel',cellstr(hits(1:N,:)),'TickDir','out',...
   'FontName','Courier New','FontSize',14);
set(gca,'XTick', 1:1:size(Fr,2)+4, 'TickDir','out');
jet1=jet;
jet1(1,:)=[0.4 0.4 0.4];
colormap(jet1);
colorbar;
% generate a plain text table for saving or copy from command line
S = char(32*ones(size(hits,1),2));
L = [ S(:,1) char(124*ones(size(hits,1),1)) S(:,1)];
F  = FrPPM(IXall,:);   % display frequency in ppm
toSave = [hits     S  ];
for i=1:numel(SET)
   for j=1:numel(SET{i})
      toSave = [toSave num2str(F(:,SET{i}(j))) S];
   end
   toSave = [toSave L];
end
toSave = [toSave S  num2str(round(Rhits)) L];
disp(toSave);
if ~isempty(toSAVE)
   fs = fopen(fullfile(Dir,toSAVE),'w');
   RET = char(10*ones(size(toSave,1),1));
   fprintf( fs, '%s\r\n', [toSave RET]');
   fclose all;
```

```
end
% or you can just copy paste the results from the command line
```

Appendix B-22: MATLAB Script for 20x20 Plots Generation

```
%close all;
addpath (genpath(pwd));
D = '';
F = 'VT_unfiltered_Feb.txt';
SAVEfile = '';  % keep blank if don't want to save
SET{1} = 19:21;        % B-SX4, captured: size 6e7 PFU
SET{2} = 22:24;        % SX74, captured: size 2e4 PFU
SET{3} = 25:27;       % B-SX4, captured on biotin-blocked beads: 3e3 PFU
SET{4} = 28:30;        % input of B-SX4,
SET{5} = 31:34;        % input of SX74,
SET{6} = 35:36;       % input of B-SX4, captured on biotin-blocked beads
PFU(1) = 6e7;
PFU(2) = 2e4;
PFU(3) = 3e3;
PFU(4) = 1e10;
PFU(5) = 1e10;
PFU(6) = 1e10;
INPUTS = [4 5 6];
TEST_SET = 1;
SHOWaminoACIDS = [19 20 21 22];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
SAVEDir = '';
saveIMG = 'YEBreactive1';
SAVEto = 'YEBreactive1.txt';
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%% find the largest element in the SET
MAX = 0;
for i=1:numel(SET)
    if max(SET{i}) > MAX
        MAX = max(SET{i});
    end
end
% define the numeric values for the input columns
input_columns = [];
for i = 1:numel(INPUTS)
    input_columns = [input_columns SET{INPUTS(i)}];
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%
REREAD = 1; % change to zero if you're rerun the script
if REREAD
    [Nu0, AA0, Fr0] = readMulticolumn('Dir', D, 'File', F, ...
                                'column', 1:MAX,...
                                'skip', 2, 'output', 'raw');
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%
% find the most common SDBs
cNuc = char(Nu0);
SDB = cellstr( cNuc(:,7:27) );
[uSDB,frSDB]=uniqueCOMB( SDB, ones (size(SDB,1), 1 ) );
% most common SDB is the SX4 SDB #1:
j=1;
SDB{j} = ['~~~~~~' uSDB{j} '~~~~~~~~~~~~~~~~~~~~~~~~~'];
```

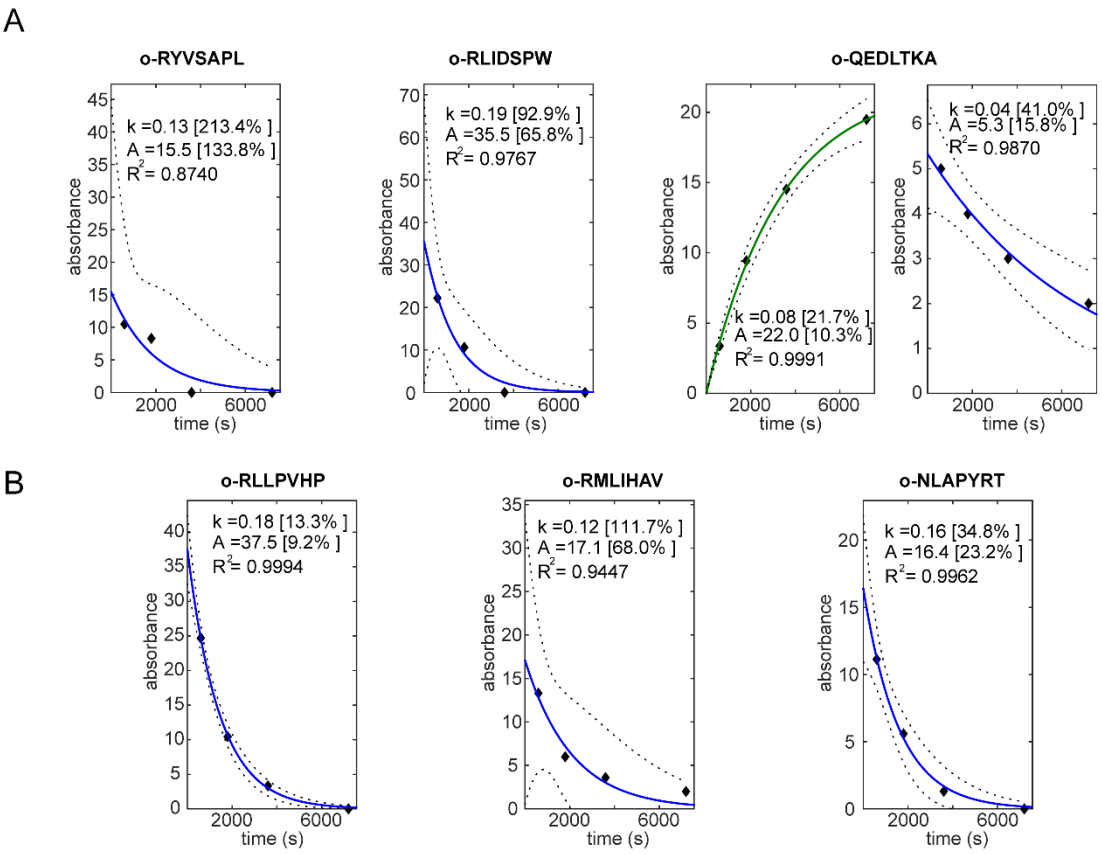
```
SX4     = 'AG~~~~~~~~~~~~~~GG~GG~GG~';
% found all the library members that have this SDB
[S1, ~] = unfiltered2filtered(Nu0, SDB{j}, 1); fprintf('.');
% trim the library to that SDB
trNuc = cNuc(S1, numel(SDB{j})+1 : end);
% filter out the SX4 library
[S2, ~] = unfiltered2filtered(trNuc,SX4, 1);
Nu = Nu0 ( S1(S2), :);
AA = AA0 ( S1(S2), :);
Fr = Fr0 ( S1(S2), :);
%%
% select only the aminoacids you want to see, SELECT ONLY SX4 motifs!!
cAA = char(AA);
AA=cellstr(cAA(:,SHOWaminoACIDS));
%%%%%% combine the frequencies
[AA,Fr]=uniqueCOMB(AA,Fr);
%%%% select only the amino acids reliably present in the input
inFr = sum( Fr(: ,input_columns), 2);
reliable = find( inFr > 2);
%% plot the input

VAR =       {'Nuc',AA,...
             'LookAt',[1 2 3 4],...
             'disp400x400',1,'disp20x400',0,...
             'maxFreq', 10000, 'scale', 'log2',...
             'grid',1,...
             'hydrophobicity','janin',...
             'sample',0};
visual400x400gen('figure', figure(1), 'Fr', inFr, VAR{:});
visual400x400gen('figure', figure(2), 'Fr', sum( Fr(:, SET{1}), 2), VAR{:});
visual400x400gen('figure', figure(3), 'Fr', sum( Fr(:, SET{2}), 2), VAR{:});
```

```
visual400x400gen('figure', figure(4), 'Fr', sum( Fr(:, SET{3}), 2), VAR{:});
Nu = Nu ( reliable, :);
AA = AA ( reliable, :);
Fr1 = Fr ( reliable, :);
inFr = inFr(reliable, :);
%%  This is where I calculate the Si, form which I then calculate the C
for i =1 : numel(SET)
   Sa(i) = PFU(i) / sum(sum(Fr(:,SET{i})));
   Si{i} = Sa(i)*sum(Fr1(:,SET{i}),2);
end
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
VAR =       {'Nuc',AA,...
            'LookAt',[1 2 3 4],...
            'disp400x400',1,'disp20x400',0,...
            'maxFreq', 3000, 'scale', 'log2',...
            'grid',1,...
            'hydrophobicity','janin',...
            'sample',0};
visual400x400gen('figure', figure(5), 'Fr', inFr, VAR{:});
visual400x400gen('figure', figure(6), 'Fr', sum( Fr1(:, SET{1}), 2), VAR{:});
visual400x400gen('figure', figure(7), 'Fr', sum( Fr1(:, SET{2}), 2), VAR{:});
visual400x400gen('figure', figure(8), 'Fr', sum( Fr1(:, SET{3}), 2), VAR{:});
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This is the MOST IMPORTANT LINE, this is where I calculate the conversion
% it is calculated from Si (PFU-like values), which is

C = (Si{1} - Si{2} - Si{3} ) ./ ( Si{4} + Si{5} + Si{6} );
```

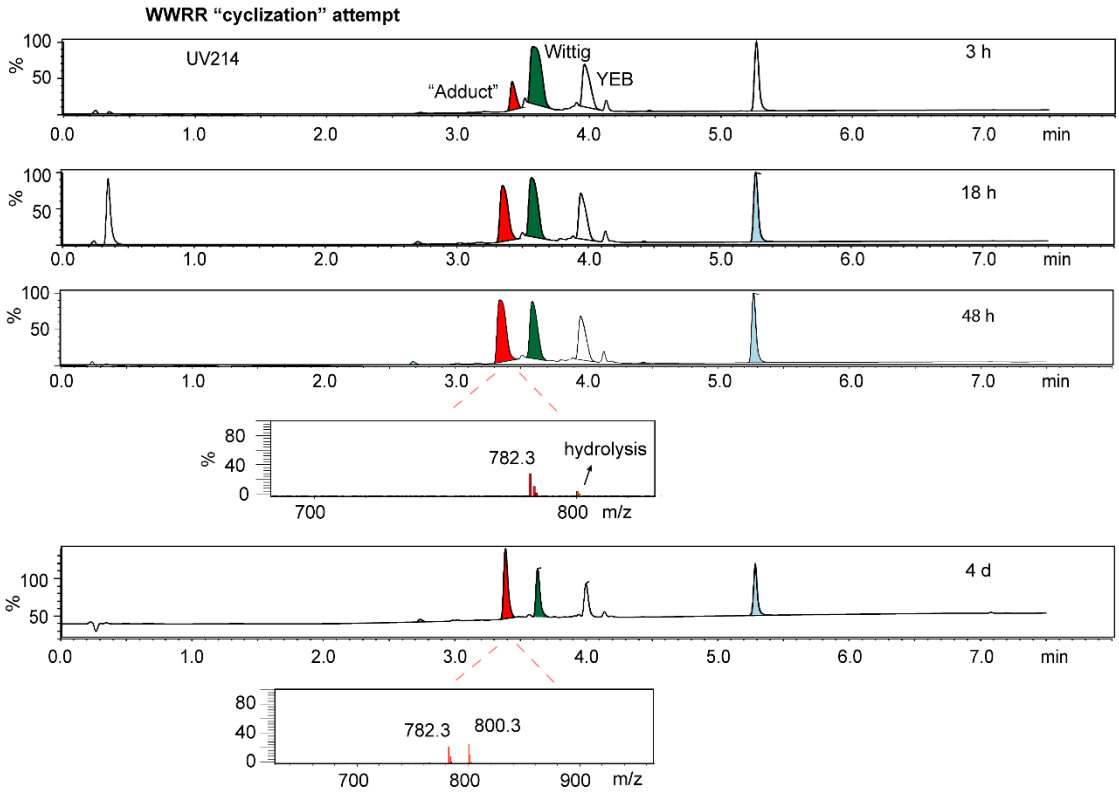
```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%find inities in the ratio and replaced them by a finite large number.
INF = find(isinf(C));
MAX = max(C(C<inf));
C(INF,:)=MAX*2;
%zero doesn't plot well on log-scale; replace zeros by some small numbers
NEG = find(C(:,1)<0);
ZER = find(C(:,1)==0);
% define small number as the smallest non-zero number in the set
%make them the same as in SX4 set
%MIN = min(C(C>0));
MIN=2^-13;
% replace zeros & negative numbers by a 1/2 and 1/3 of the smallest number
C(ZER,:)=MIN/3;
C(NEG,:)=MIN/4;
% find zero/zero = NaN and replace them by 1/5 of the smallest number
NANS = find(isnan(C(:,1)));
C(NANS,:)=MIN/5;
% note that in the "reliable dataset, there will be no INF and no NANS
% data cannot be sampled because the inputs are floating point ratios not integer counts
[dist400]=visual400x400gen('Nuc',AA, 'Fr',log2(C)-log2(MIN/4),...
                'LookAt',[1 2 3 4],...
                'disp400x400',1,'disp20x400',0,...
                'maxFreq', 10,... log2(MAX*2)-log2(MIN/4), ...
                'scale', 'lin',...
                'grid',1,...
                'figure', figure(10),...
                'hydrophobicity','janin',...
                'sample',0);
        saveas(gcf,saveIMG,'epsc')
```

```
%%%%% save the ratio data and sequence data
S = char( 32*ones(size(AA,1),1) );
toSave = [char(AA)     S num2str(C,'%10.5e')  ];
if ~isempty(SAVEfile)
    fs = fopen(fullfile(SAVEDir,SAVEfile),'w');
    RET = char(10*ones(size(toSave,1),1));
    fprintf( fs, '%s\r\n', [toSave RET]');
    fclose all;
end
%%%%%% plot histogram of ratio
figure(30);
[b,x]=hist( log2(C), floor(log2(MIN/4)) :0.5: log2(MAX*2) );
bar(x,b);
  set(gca,'yscale','log',...
      'xtick', floor(log2(MIN/4)) : 1 : ceil(log2(MAX*2)) ,...
      'TickDir','out');
%xlim([log2(MIN/5); ceil(log2(MAX*2)) ]);
xlim([log2(MIN/5); -4 ]);
saveas(gcf,[saveIMG 'Hist'],'epsc')
figure(100);
plot(Si{1},Si{2},'.k')
hold on
plot(Si{1},Si{3},'.r')
plot(Si{1},Si{4},'.c')
LIM = [0.001 1e8];
plot(LIM, LIM,'-b');
set(gca,'xscale','log','yscale','log');
xlim(LIM);
ylim(LIM);
```

Appendix C: Supporting information for chapter 4



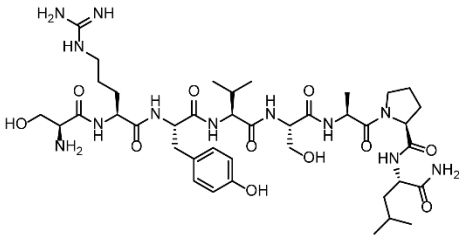
Appendix C-1. Wittig rate for validated sequences from (A) 1% biotinylation selection and (B) 0.1% biotinylation selection



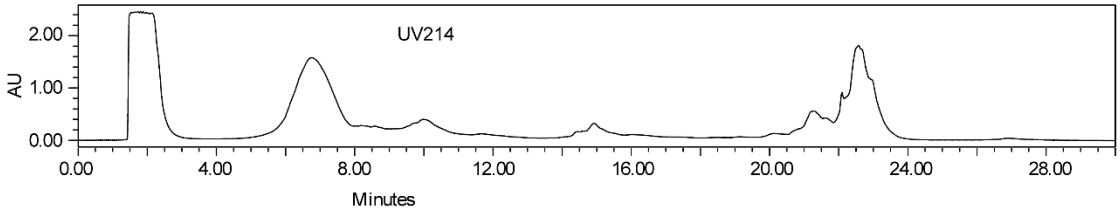
Appendix C-2. Hydrolysis of "cyclization adduct" for WWRR sequence. Reaction conditions were [Ald-WWRR]=5.5 mM, [YEb]=11mM in MOPS 200 mM at pH 6.5.

Appendix C-3. HPLC purity and LCMS traces of synthesized peptides

SRYVSAPL



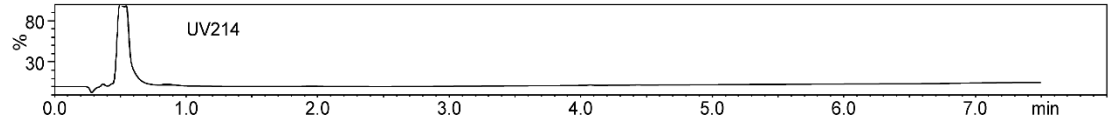
Chemical Formula: $C_{40}H_{66}N_{12}O_{11}$
Exact Mass: 890.50

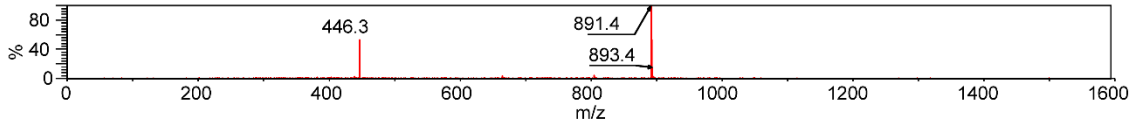


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

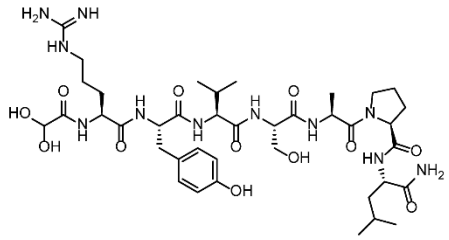
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

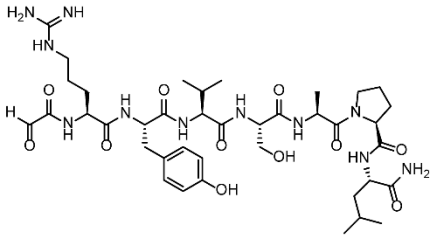




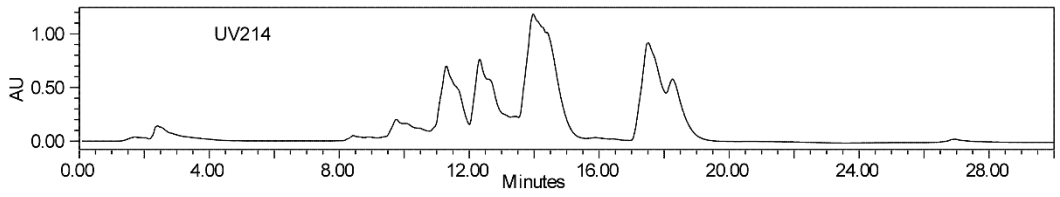
o-RYVSAPL



Chemical Formula: $C_{39}H_{63}N_{11}O_{12}$
Exact Mass: 877.47



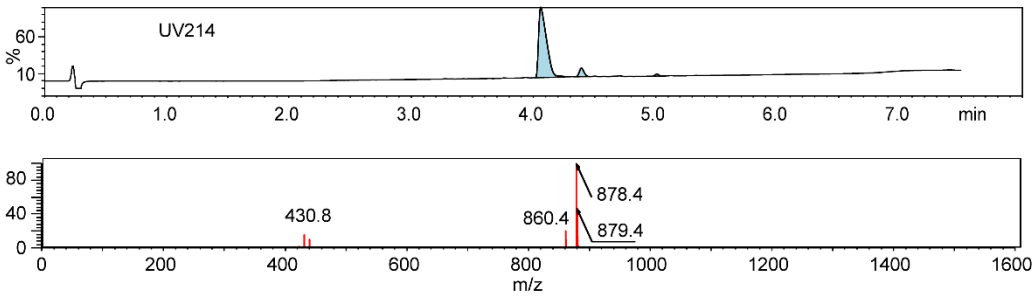
Chemical Formula: $C_{39}H_{61}N_{11}O_{11}$
Exact Mass: 859.46



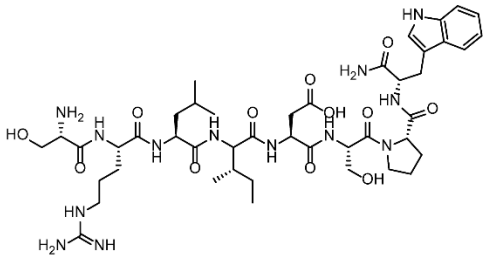
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

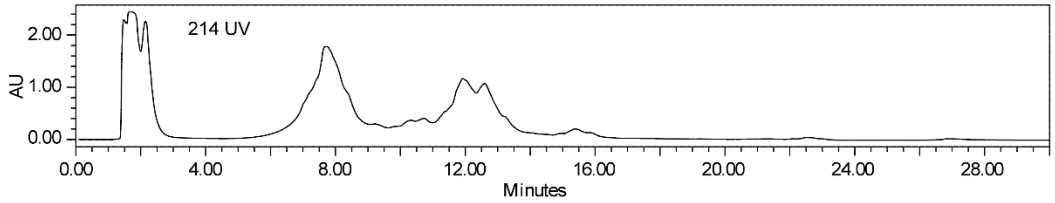
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



SRLIDSPW



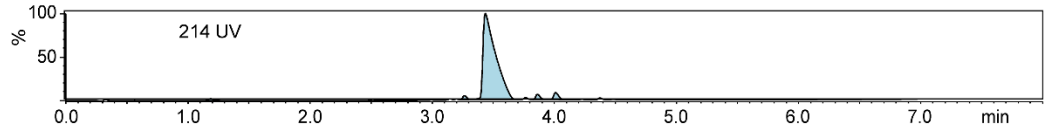
Chemical Formula: $C_{44}H_{69}N_{13}O_{12}$
Exact Mass: 971.52

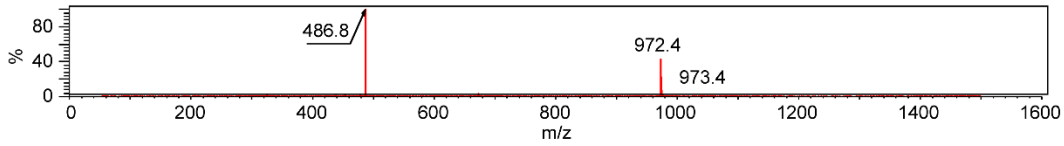


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

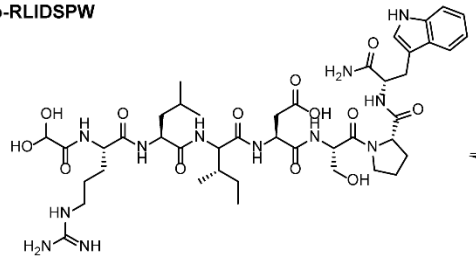
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

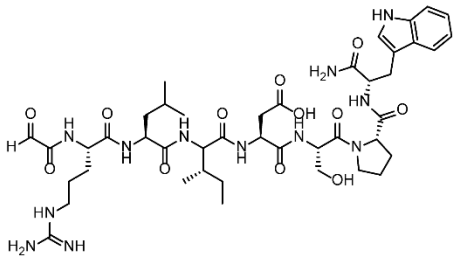




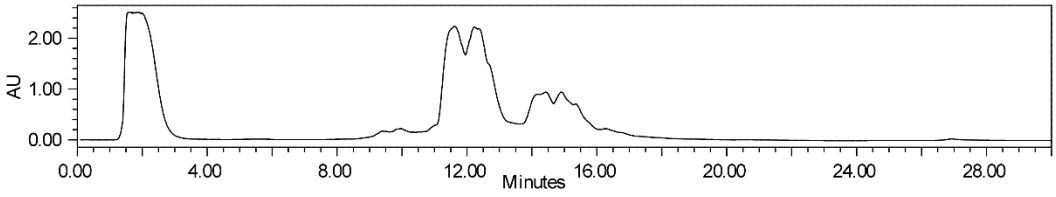
o-RLIDSPW



Chemical Formula: $C_{43}H_{66}N_{12}O_{13}$
Exact Mass: 958.49



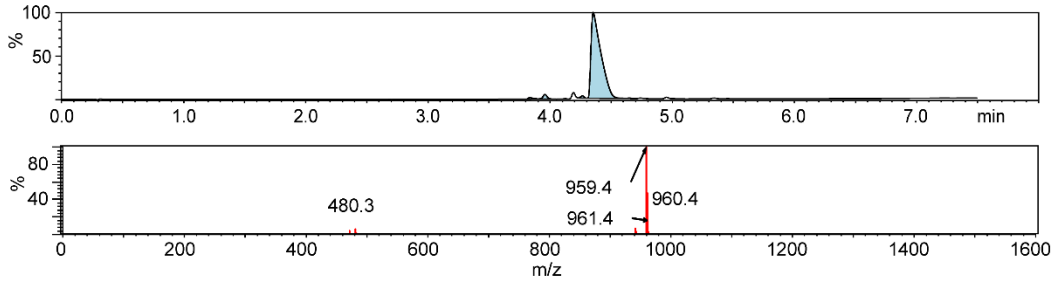
Chemical Formula: $C_{43}H_{64}N_{12}O_{12}$
Exact Mass: 940.48



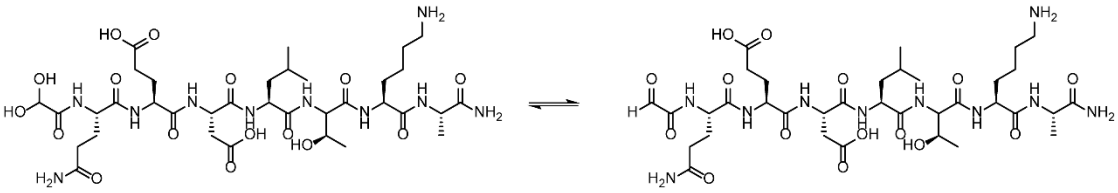
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

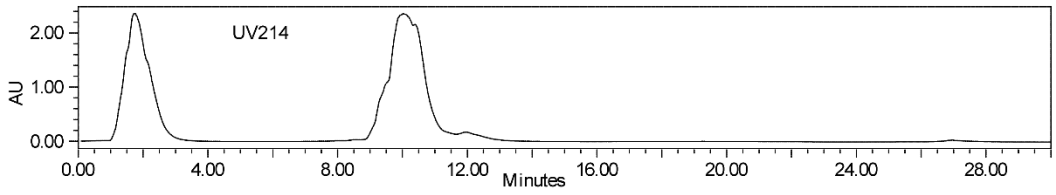


o-QEDLTKA



Chemical Formula: $C_{35}H_{60}N_{10}O_{16}$
Exact Mass: 876.42

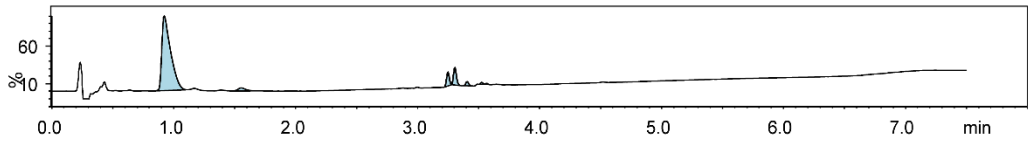
Chemical Formula: $C_{35}H_{58}N_{10}O_{15}$
Exact Mass: 858.41

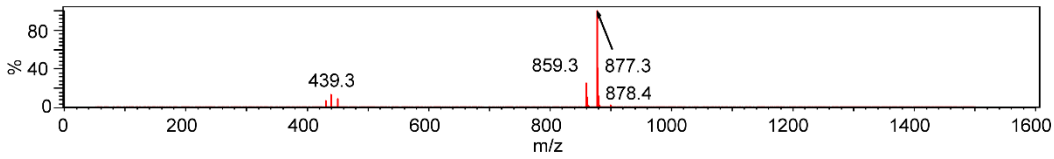


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

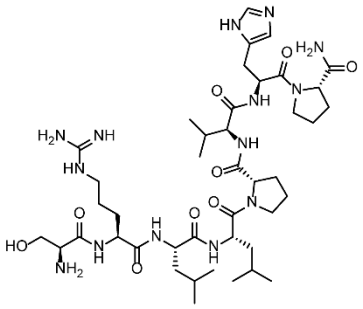
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

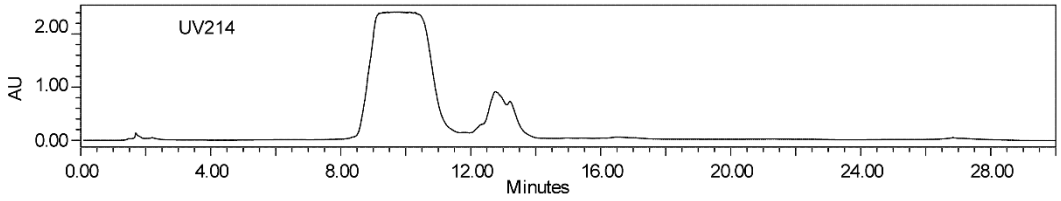




SRLLPVHP



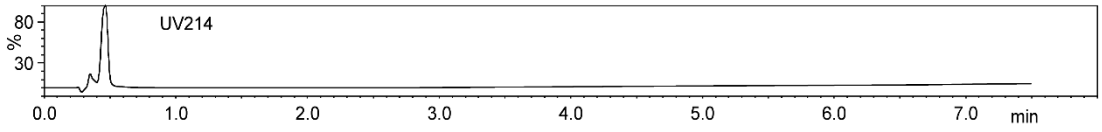
Chemical Formula: $C_{42}H_{72}N_{14}O_9$
Exact Mass: 916.56

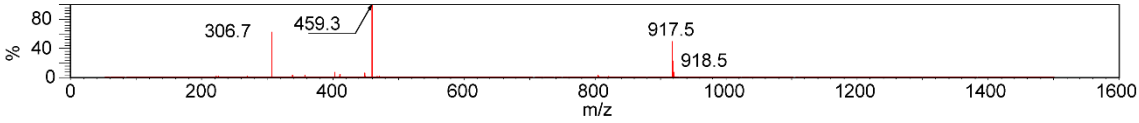


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

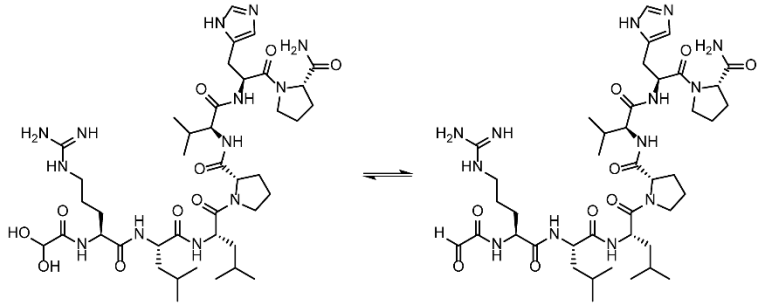
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



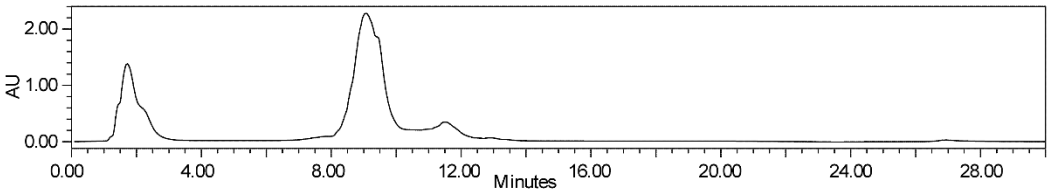


o-RLLPVHP



Chemical Formula: $C_{41}H_{69}N_{13}O_{10}$
Exact Mass: 903.53

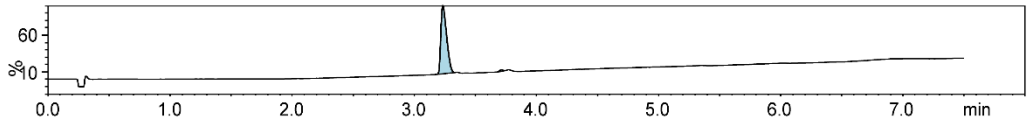
Chemical Formula: $C_{41}H_{67}N_{13}O_9$
Exact Mass: 885.52

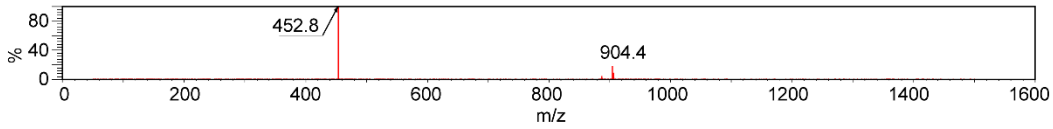


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

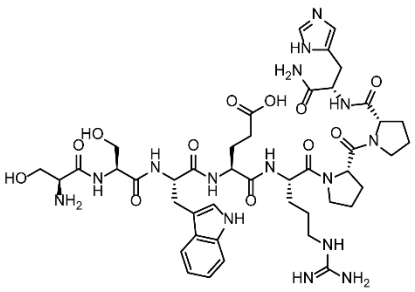
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

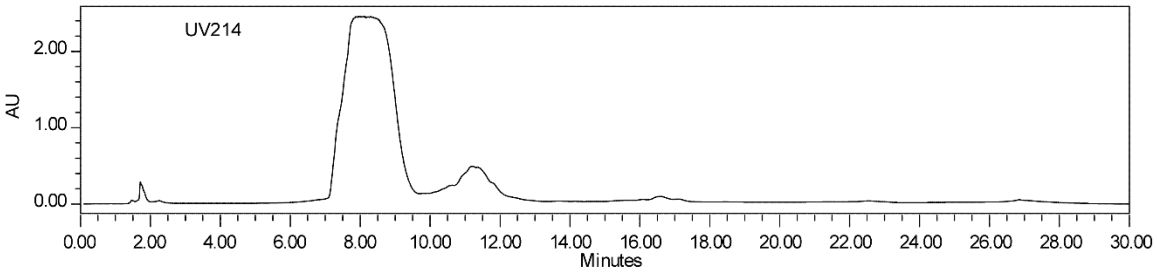




SSWERPPH



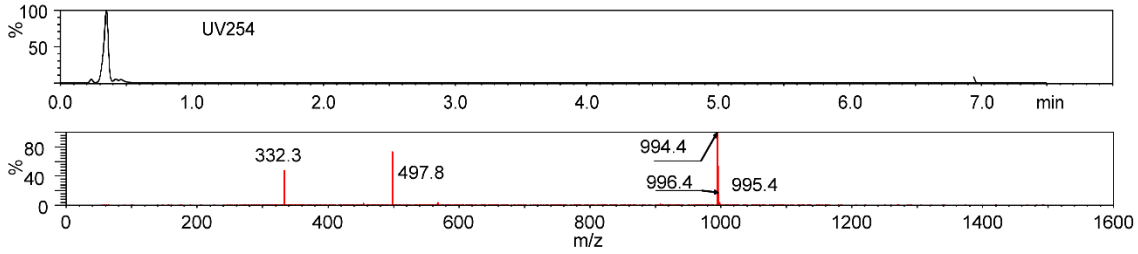
Chemical Formula: $C_{44}H_{63}N_{15}O_{12}$
Exact Mass: 993.48



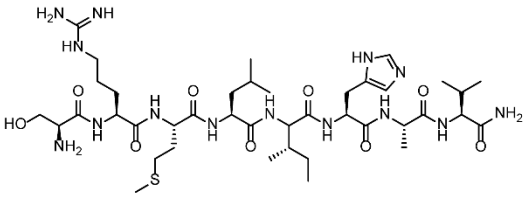
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

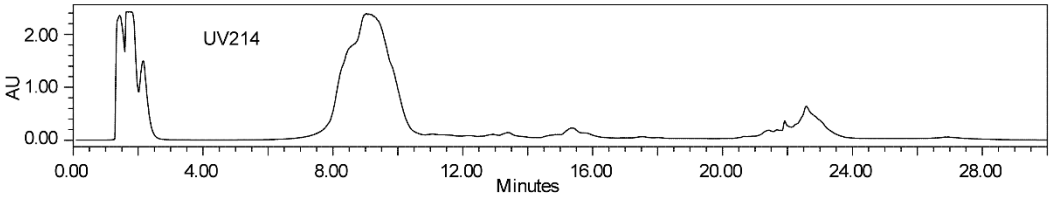
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



SRMLIHAV



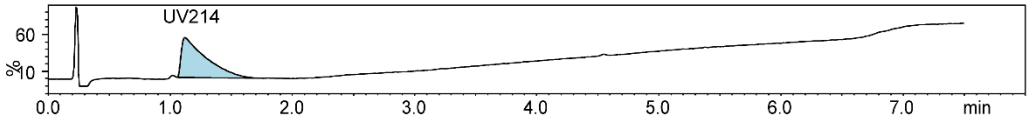
Chemical Formula: $C_{40}H_{72}N_{14}O_9S$
Exact Mass: 924.53

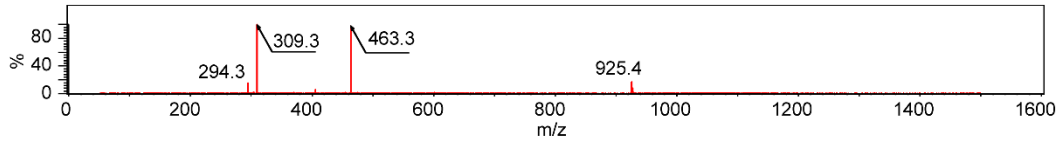


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

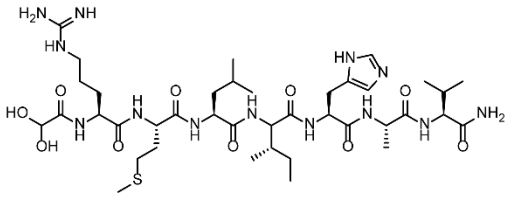
Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

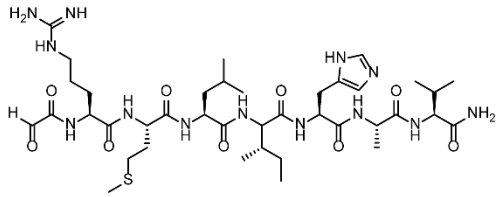




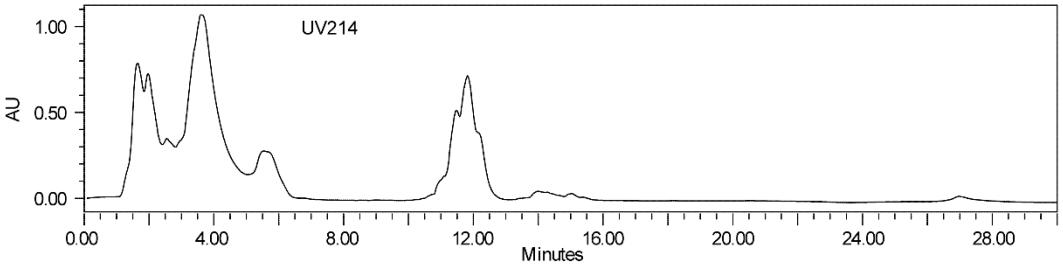
o-RMLIHAV



Chemical Formula: $C_{39}H_{69}N_{13}O_{10}S$
Exact Mass: 911.50



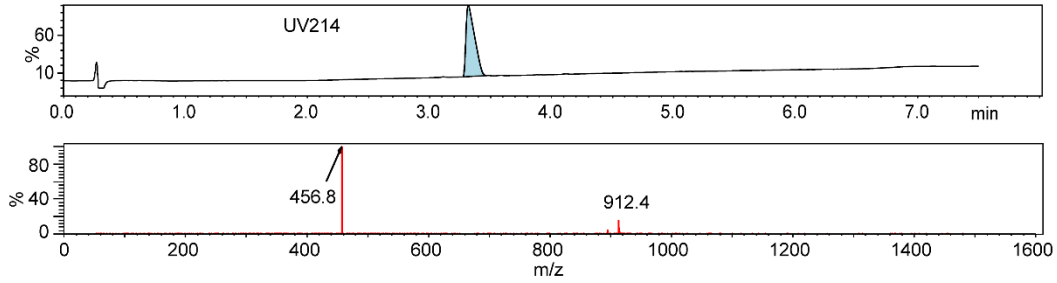
Chemical Formula: $C_{39}H_{67}N_{13}O_9S$
Exact Mass: 893.49



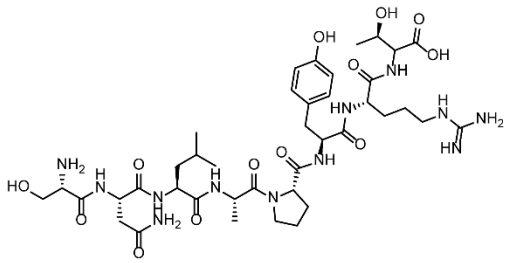
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

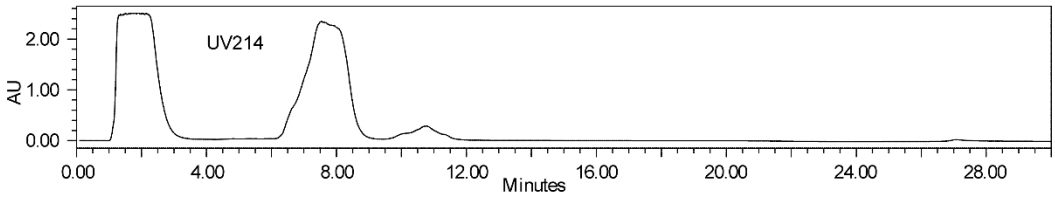
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



SNLAPYRT



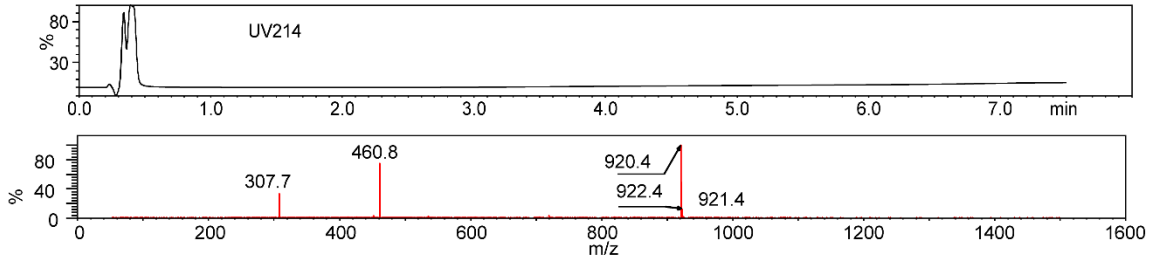
Chemical Formula: $C_{40}H_{64}N_{12}O_{13}$
Exact Mass: 920.47



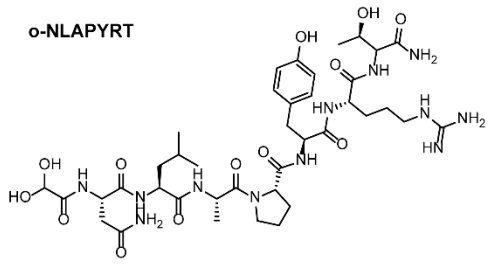
Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

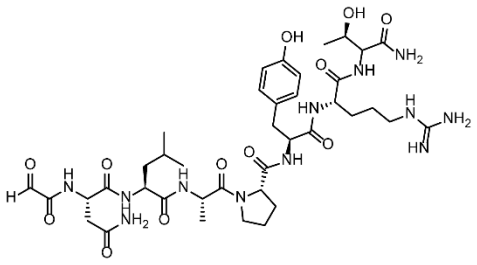
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA



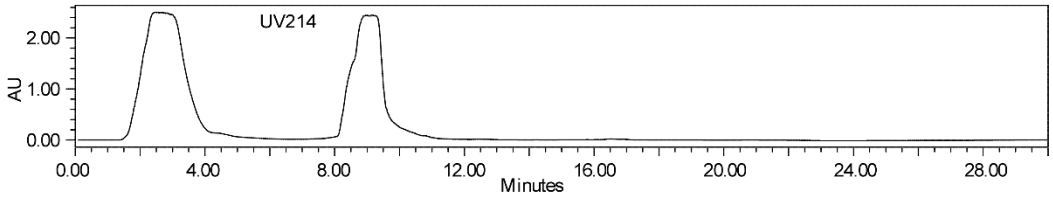
o-NLAPYRT



Chemical Formula: $C_{39}H_{62}N_{12}O_{13}$
Exact Mass: 906.46



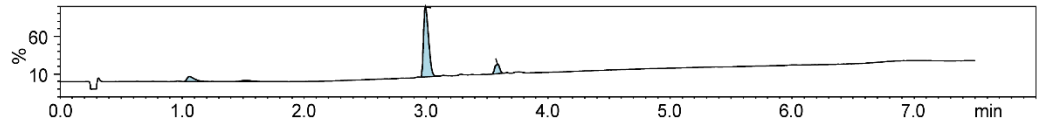
Chemical Formula: $C_{39}H_{60}N_{12}O_{12}$
Exact Mass: 888.45

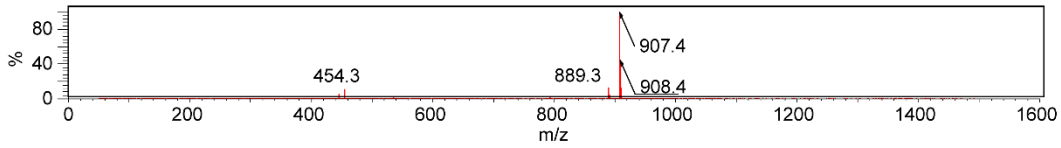


Time (min)	% B
0	2
2	2
18	50
21	10
24	10
26	2
30	2

Flow rate: 8 mL/min
C18 column prep
(100 Å, 5 µm, 19 mm x 50 mm)

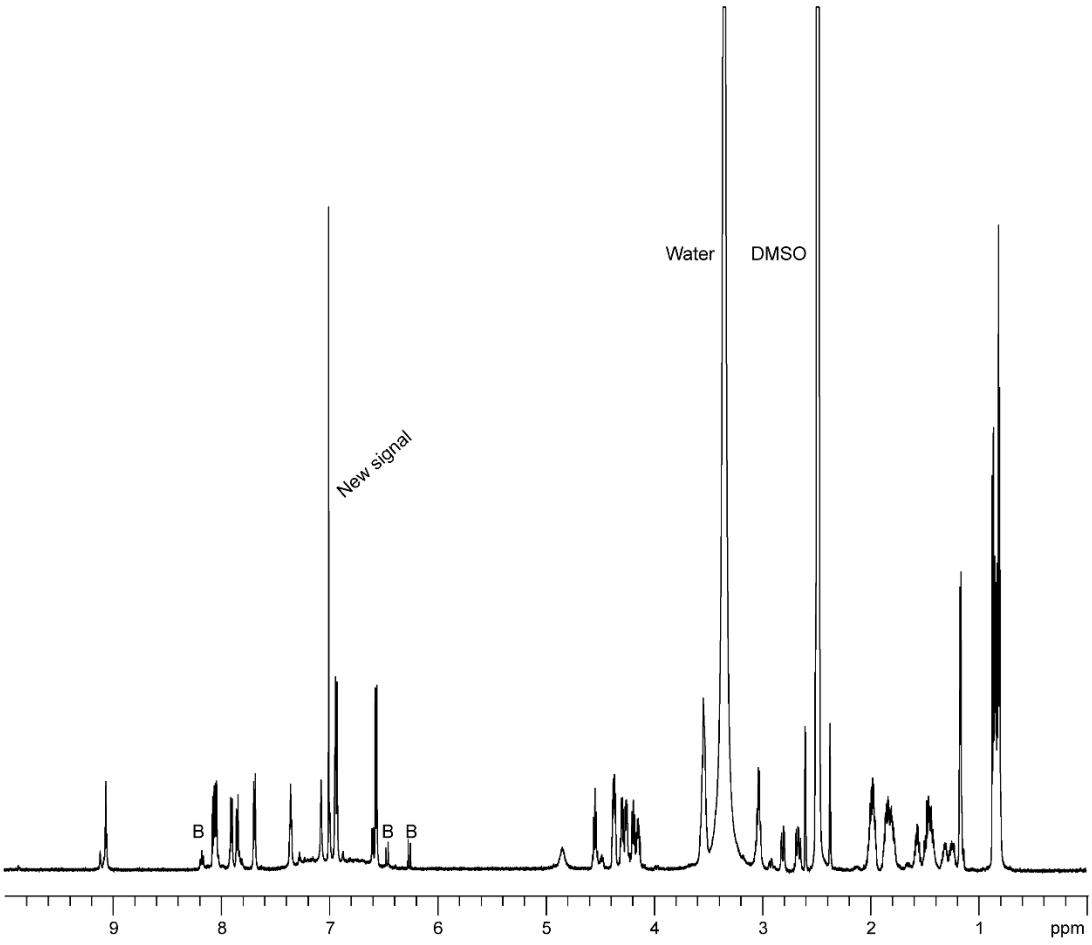
Solvent A: H2O + 0.1% TFA
Solvent B: MeCN + 0.1% TFA

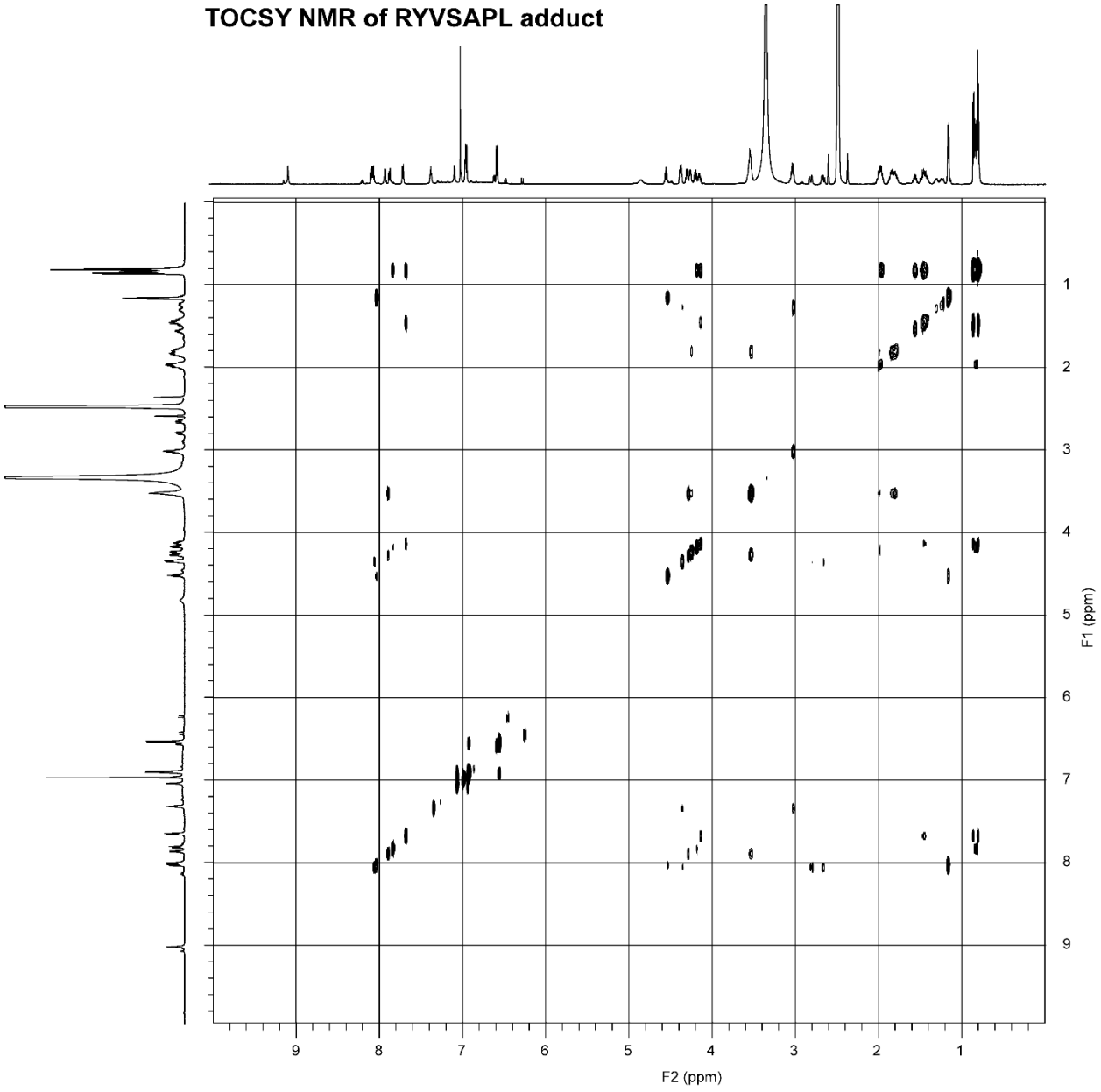


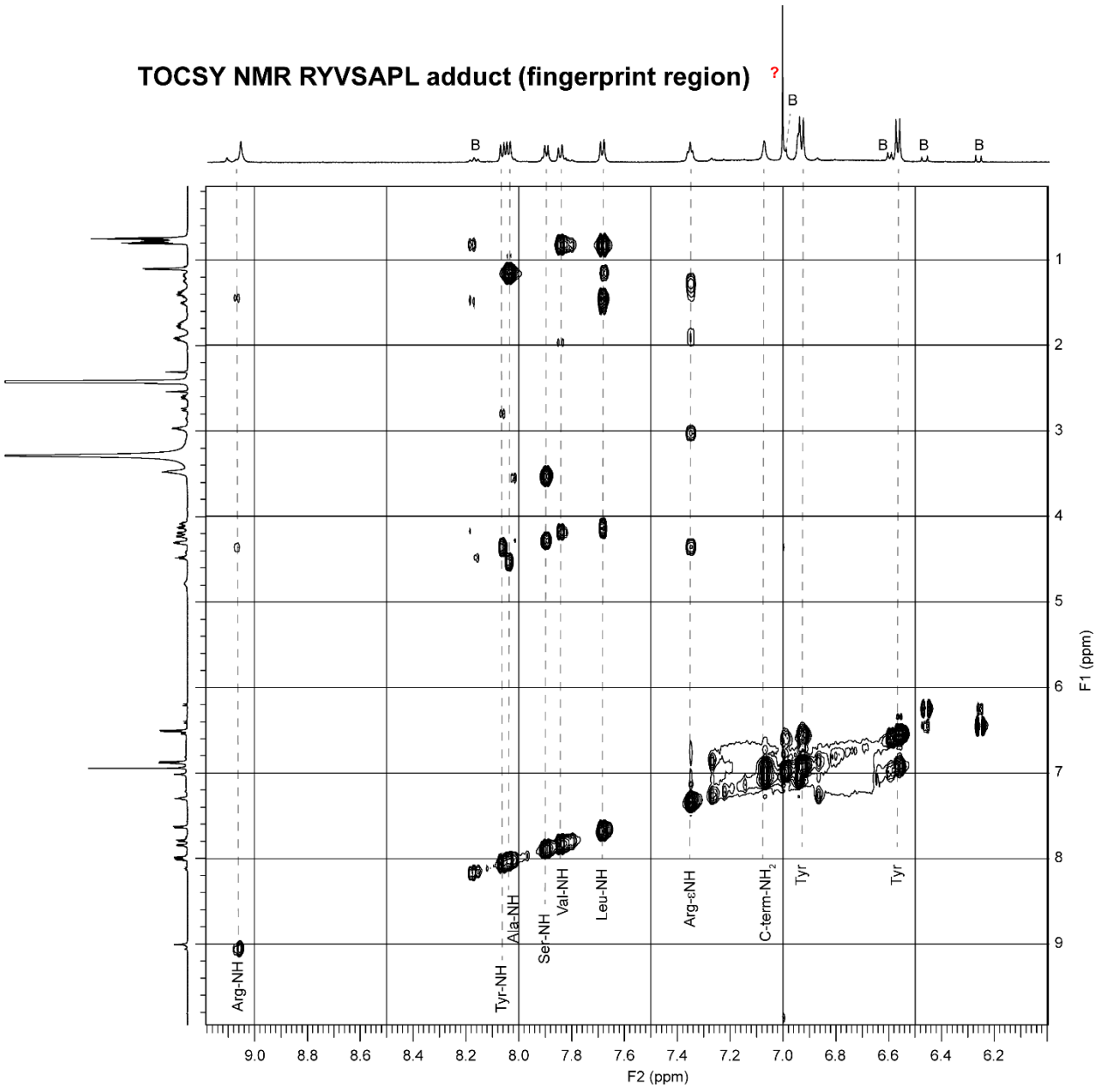


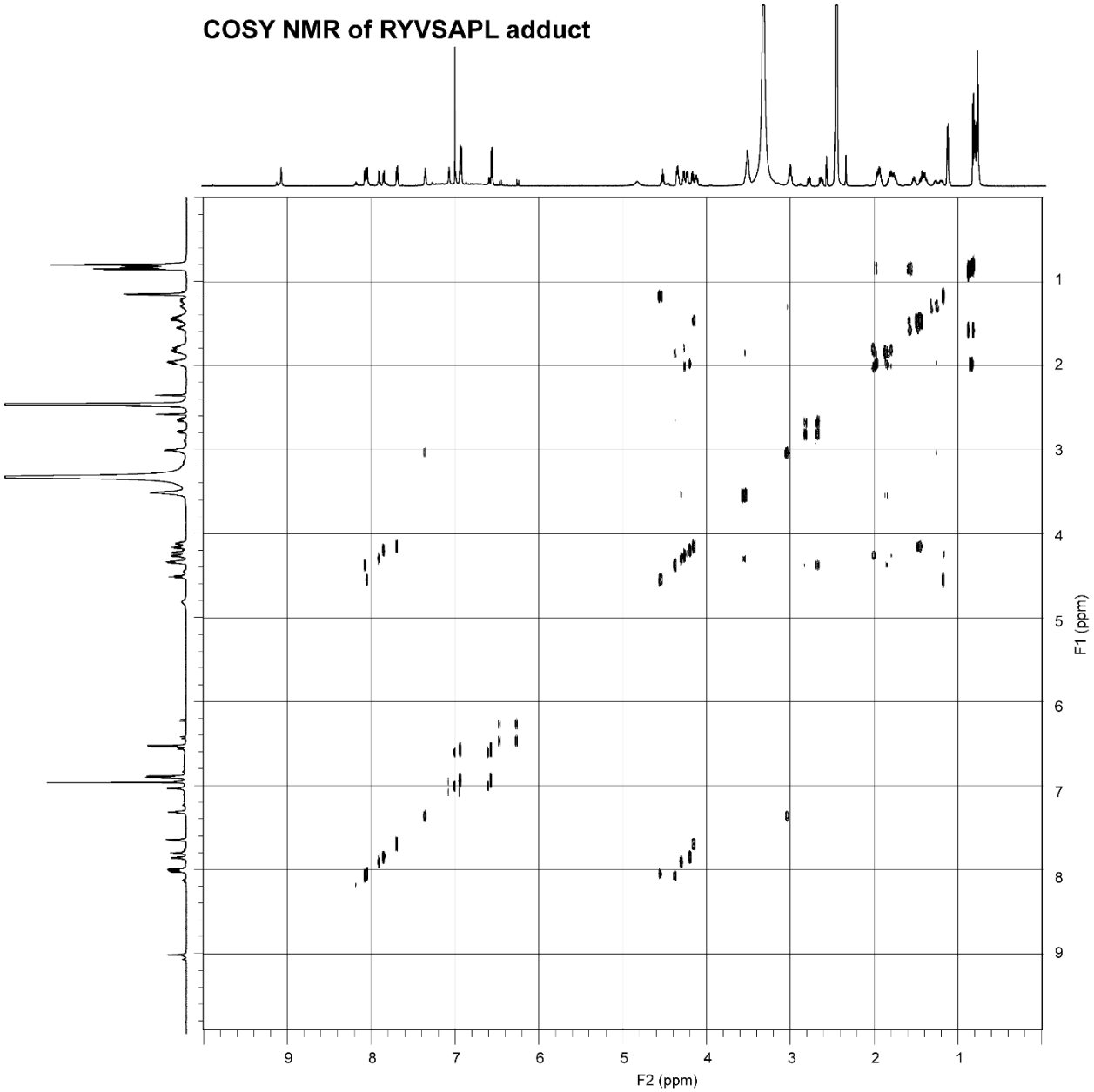
Appendix C-4: NMR spectra for SRYVSAPL and SRLIDSPW "cyclization" adducts

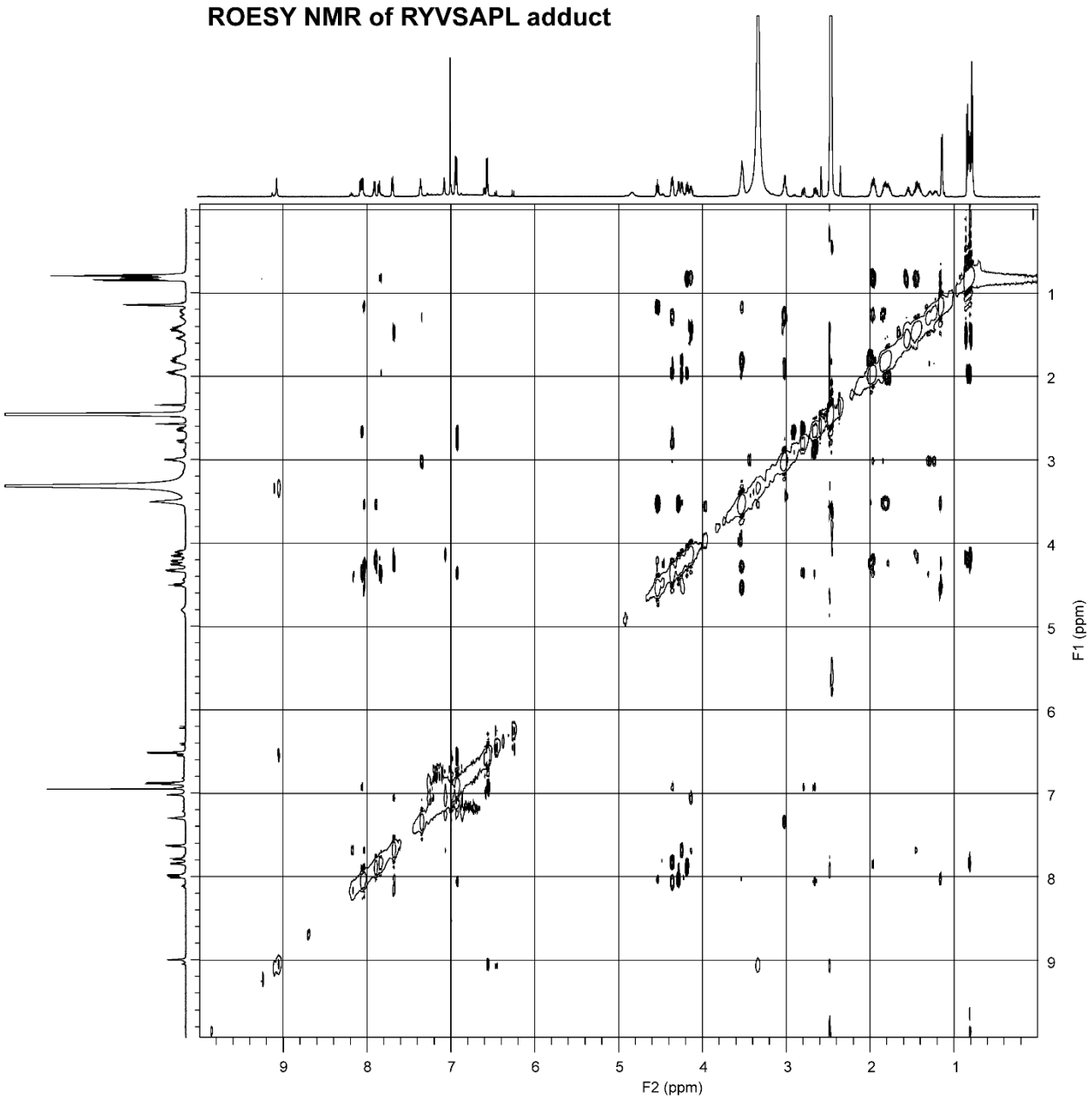
1H NMR of RYVSAPL adduct

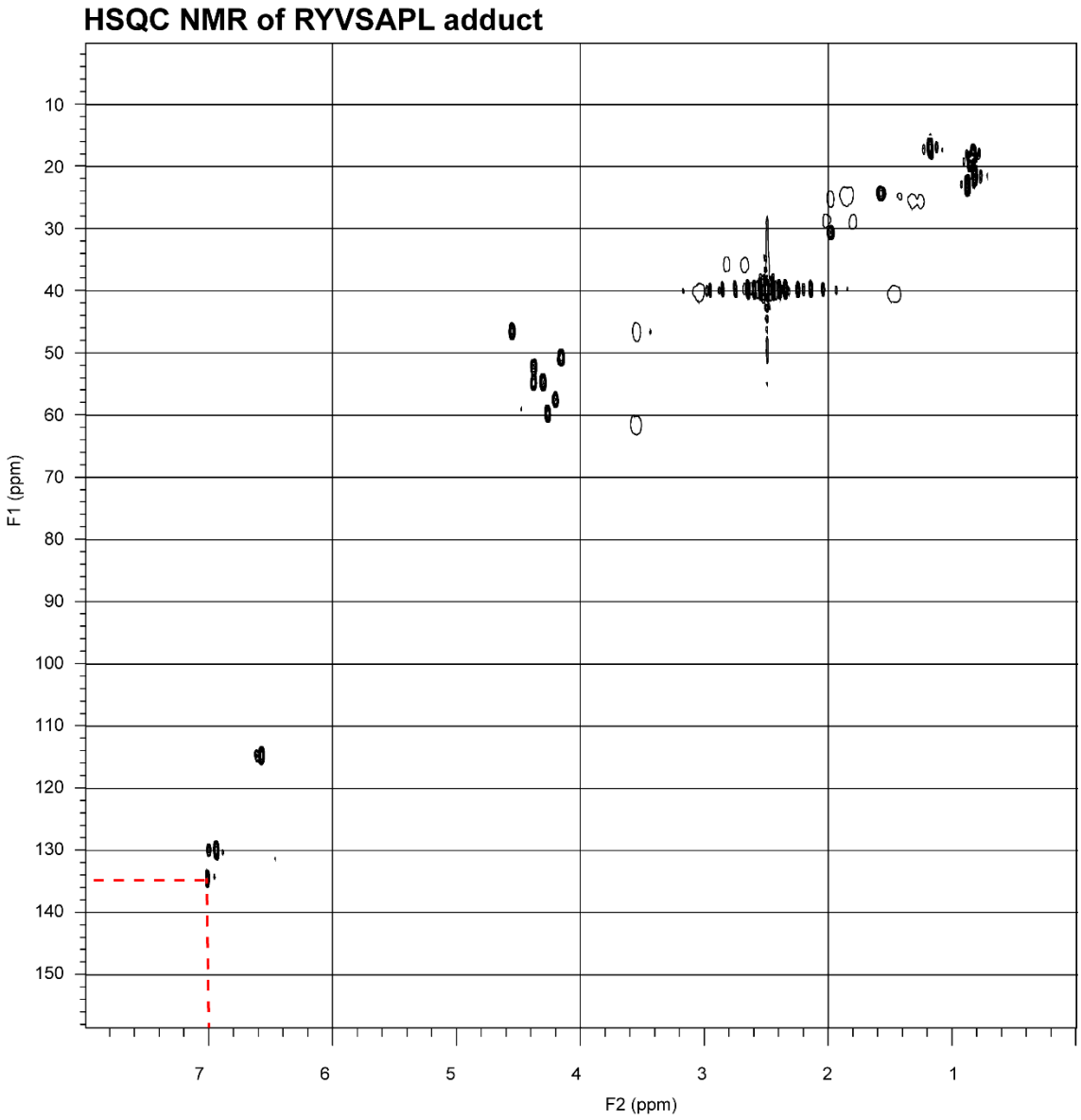


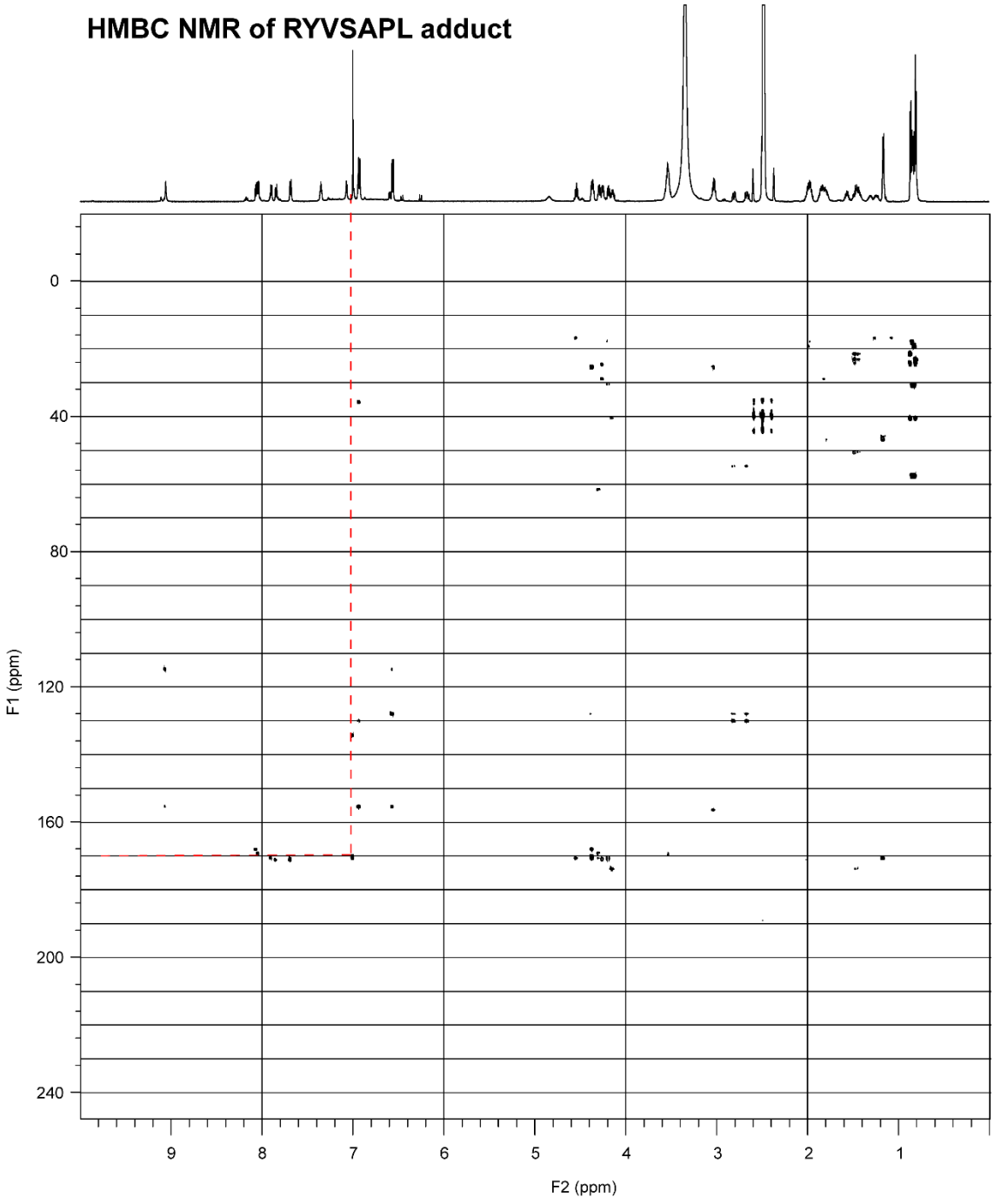

TOCSY NMR of RYVSAPL adduct
F1 (ppm)
F2 (ppm)

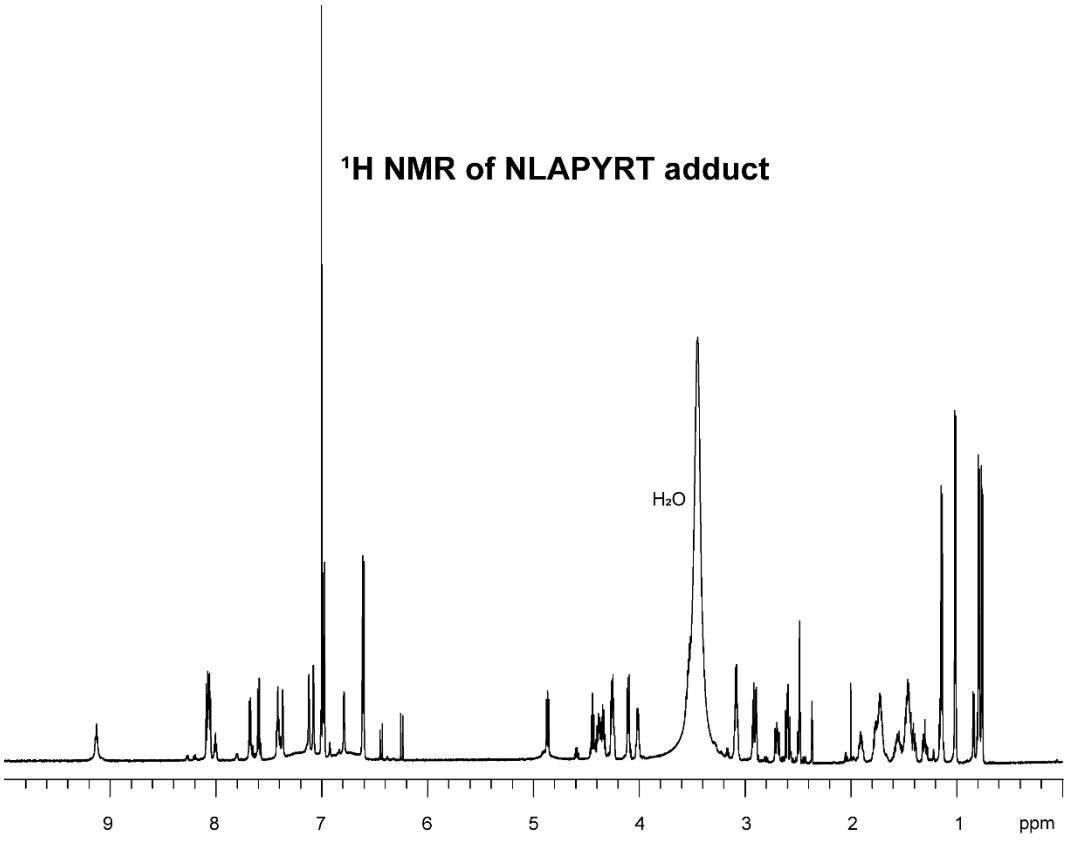

TOCSY NMR RYVSAPL adduct (fingerprint region) ?
B
B
B
B
B
Arg-NH
Tyr-NH
Ala-NH
Ser-NH
Val-NH
Leu-NH
Arg-εNH
C-term-NH2
Tyr
Tyr
F2 (ppm)
F1 (ppm)

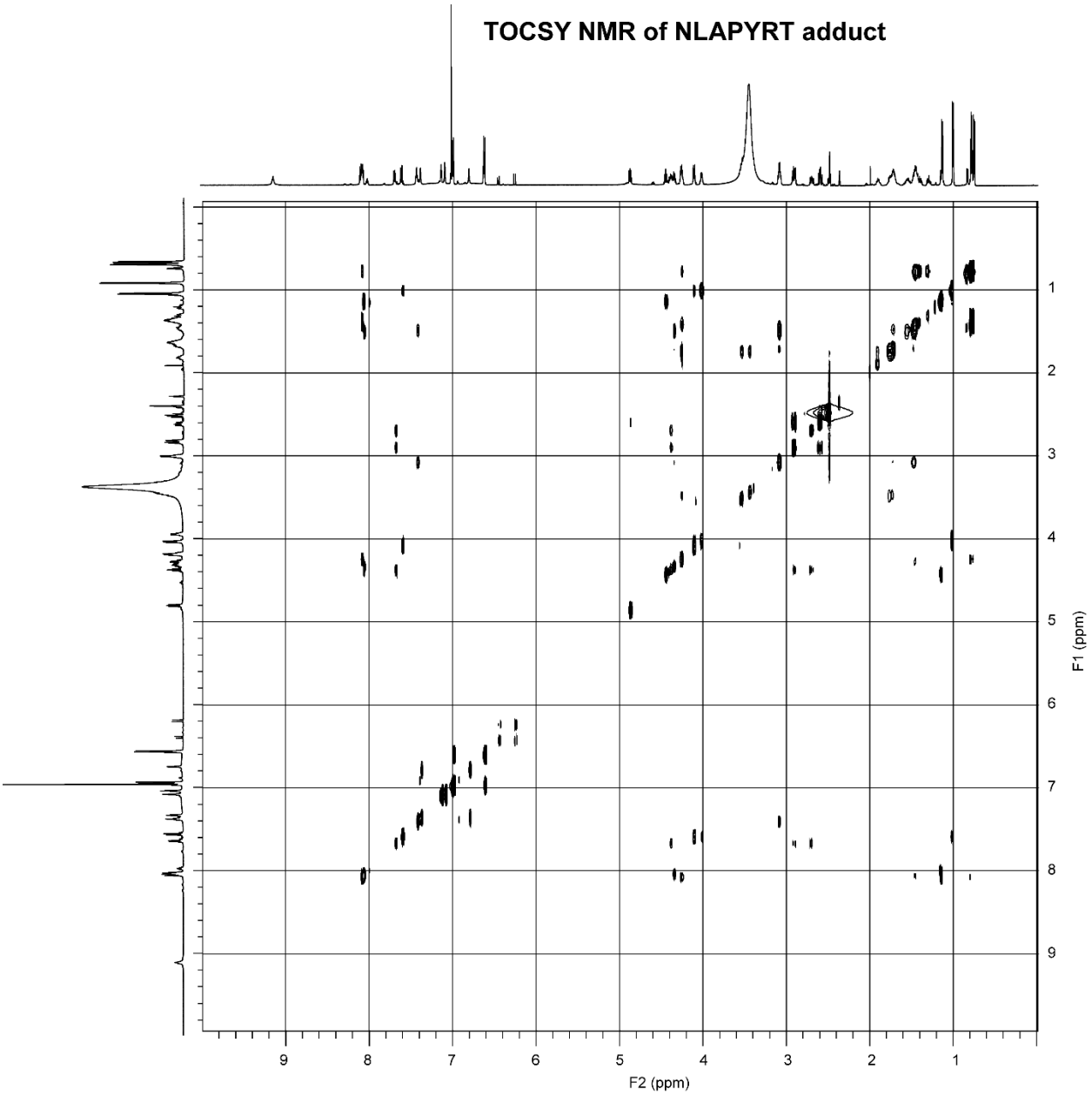

COSY NMR of RYVSAPL adduct
F2 (ppm)
F1 (ppm)

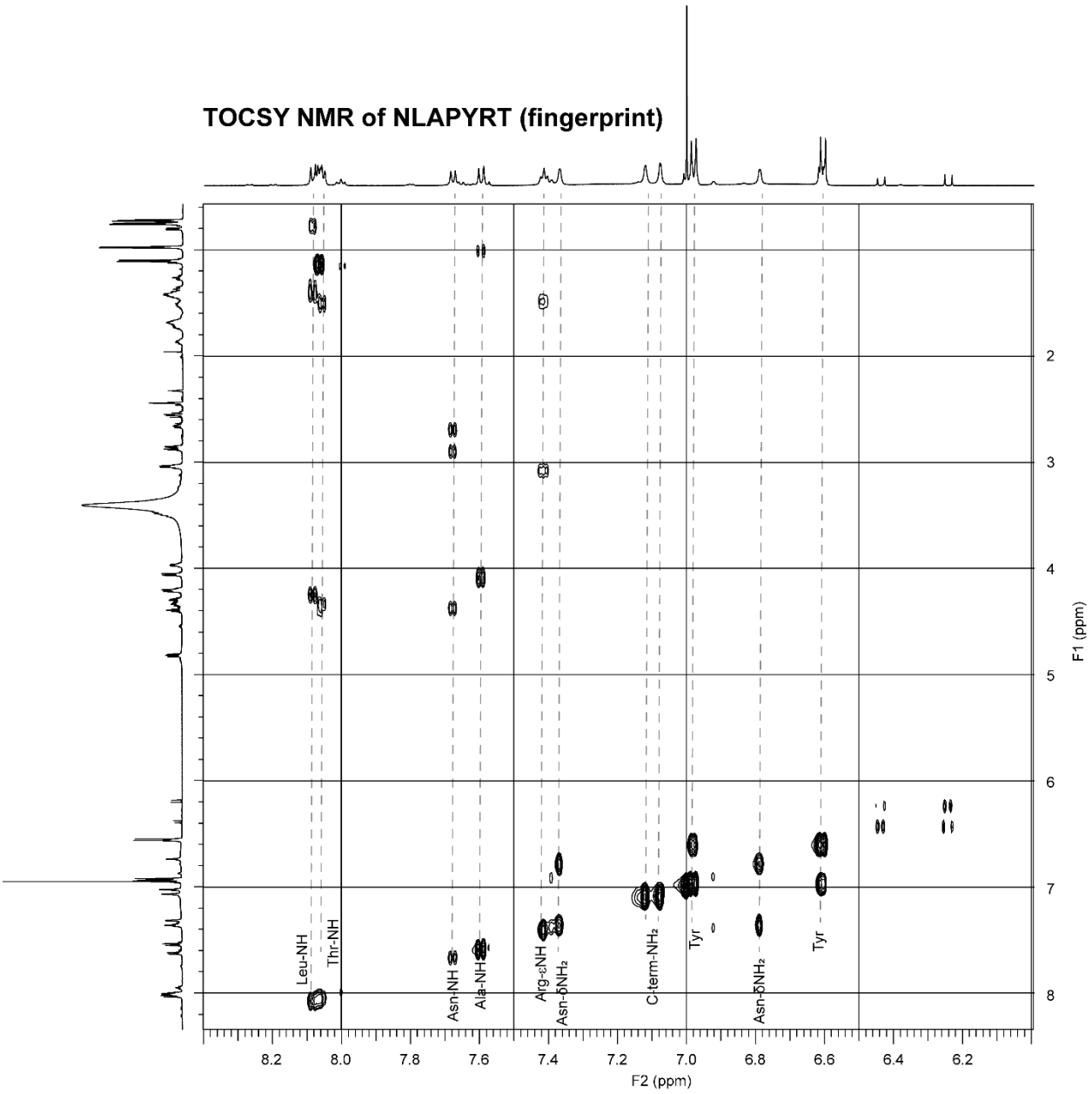

ROESY NMR of RYVSAPL adduct
F2 (ppm)
F1 (ppm)

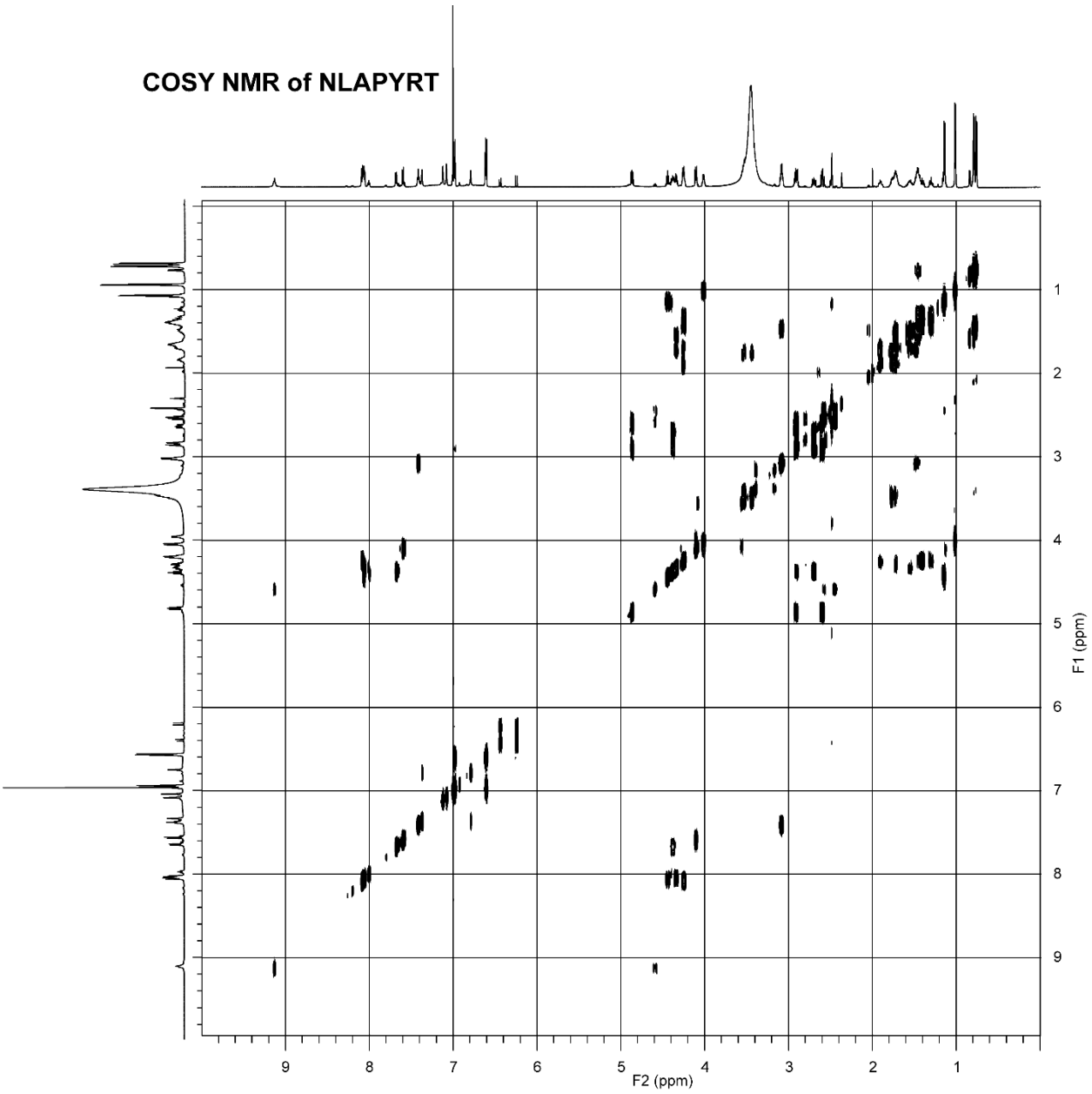

HSQC NMR of RYVSAPL adduct
F1 (ppm)
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
7
6
5
4
3
2
1
F2 (ppm)

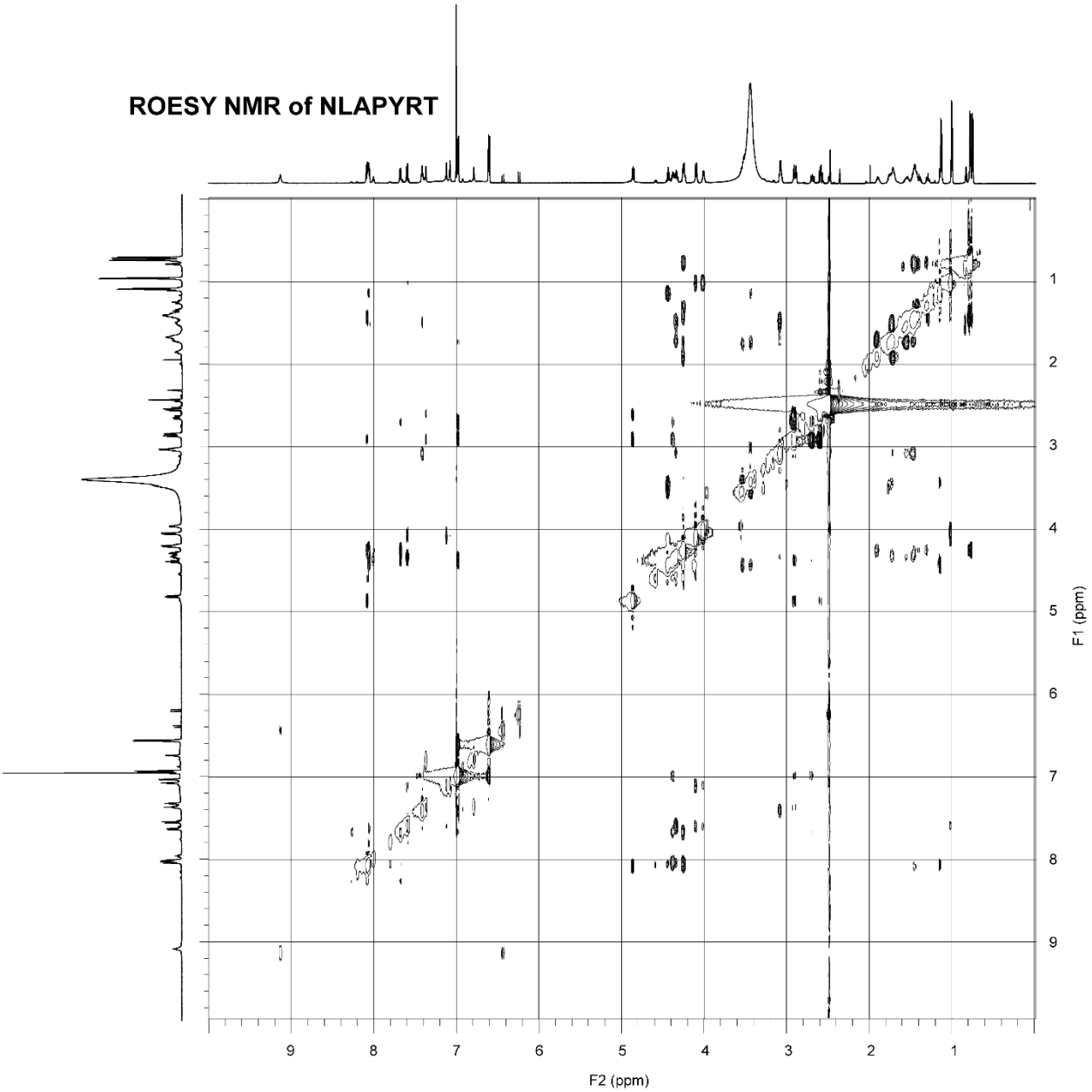

HMBC NMR of RYVSAPL adduct
0
40
80
120
160
200
240
F1 (ppm)
9
8
7
6
5
4
3
2
1
F2 (ppm)

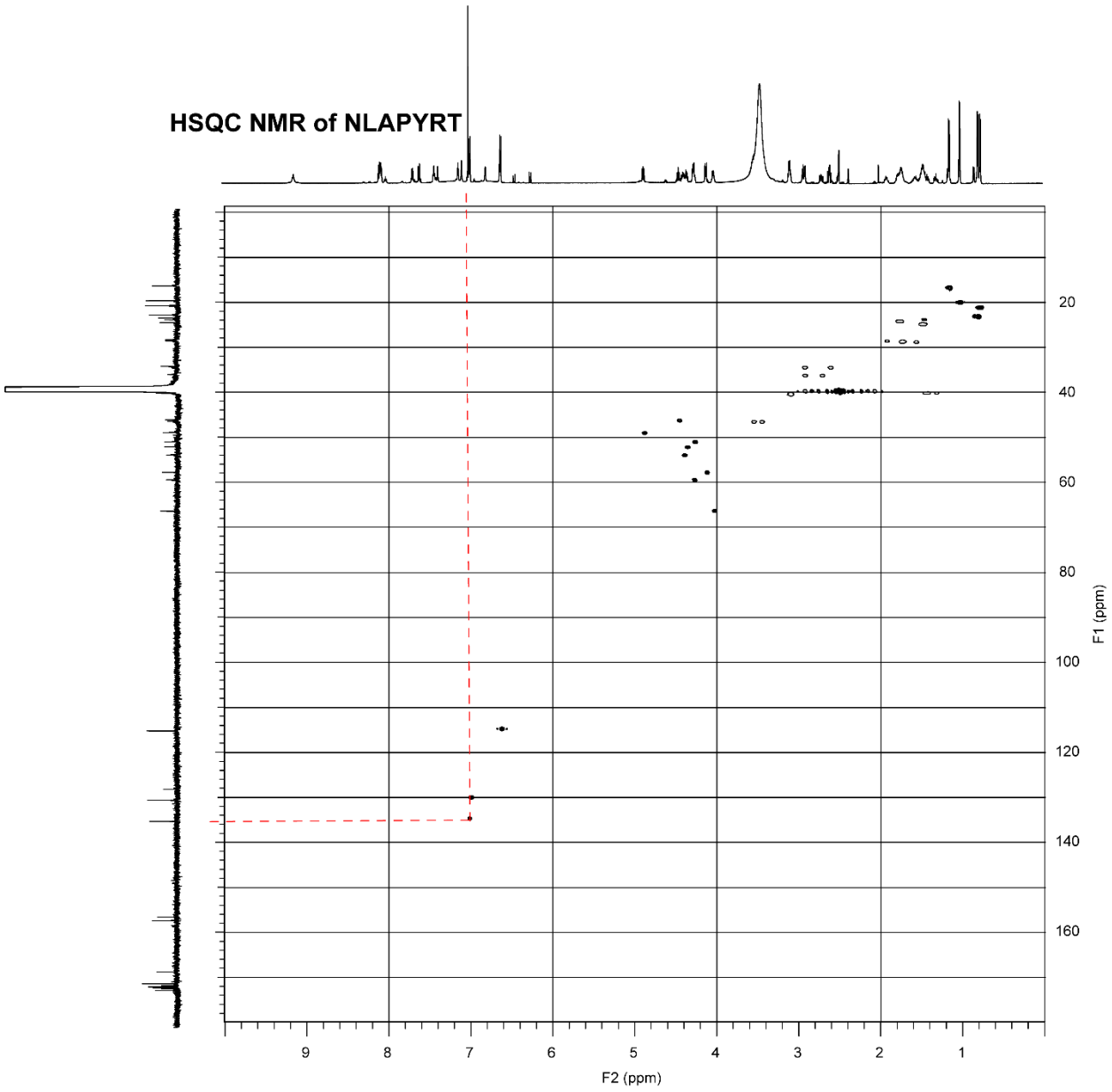

¹H NMR of NLAPYRT adduct
H₂O
9
8
7
6
5
4
3
2
1
ppm

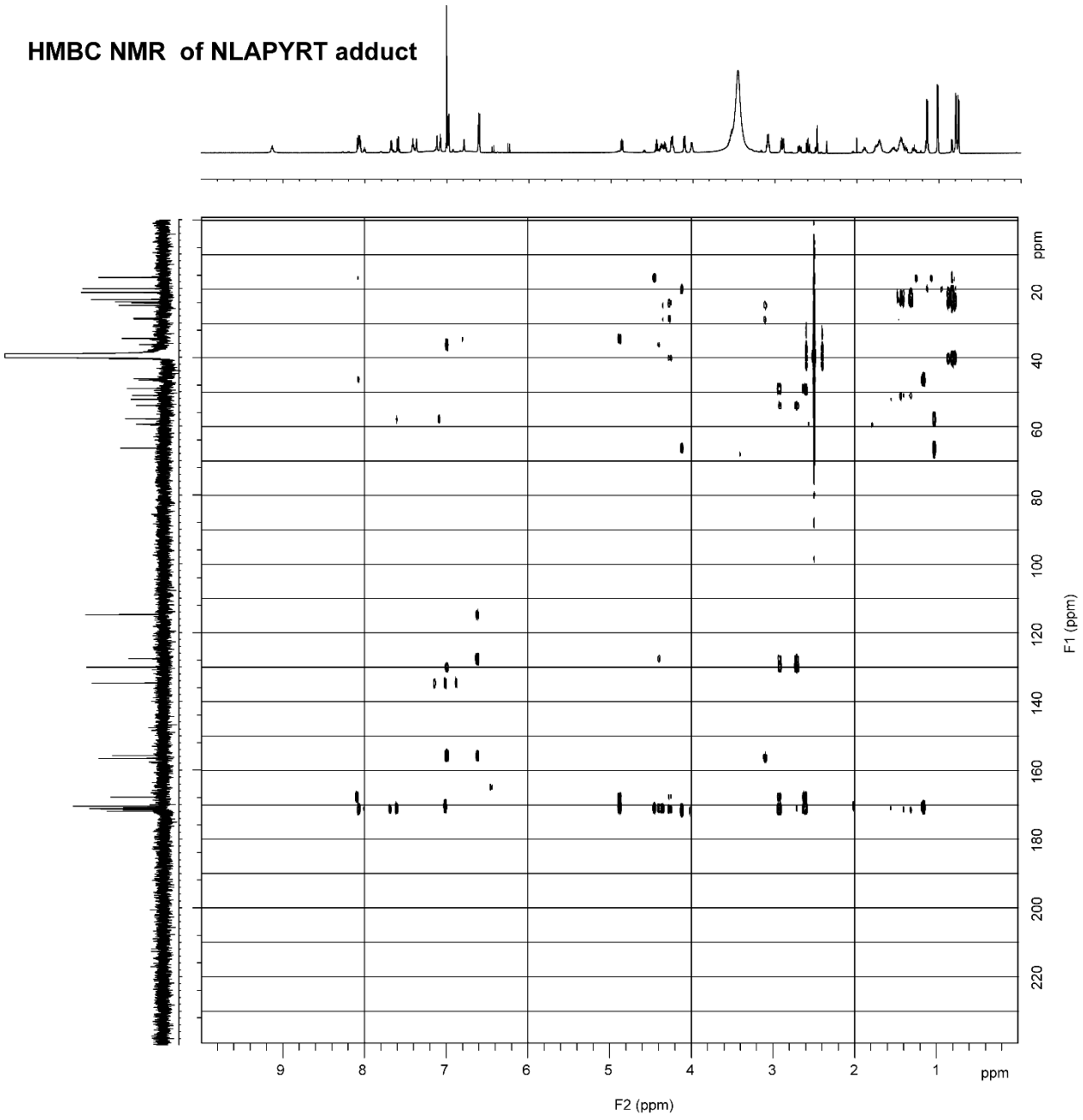

TOCSY NMR of NLAPYRT adduct
F2 (ppm)
F1 (ppm)

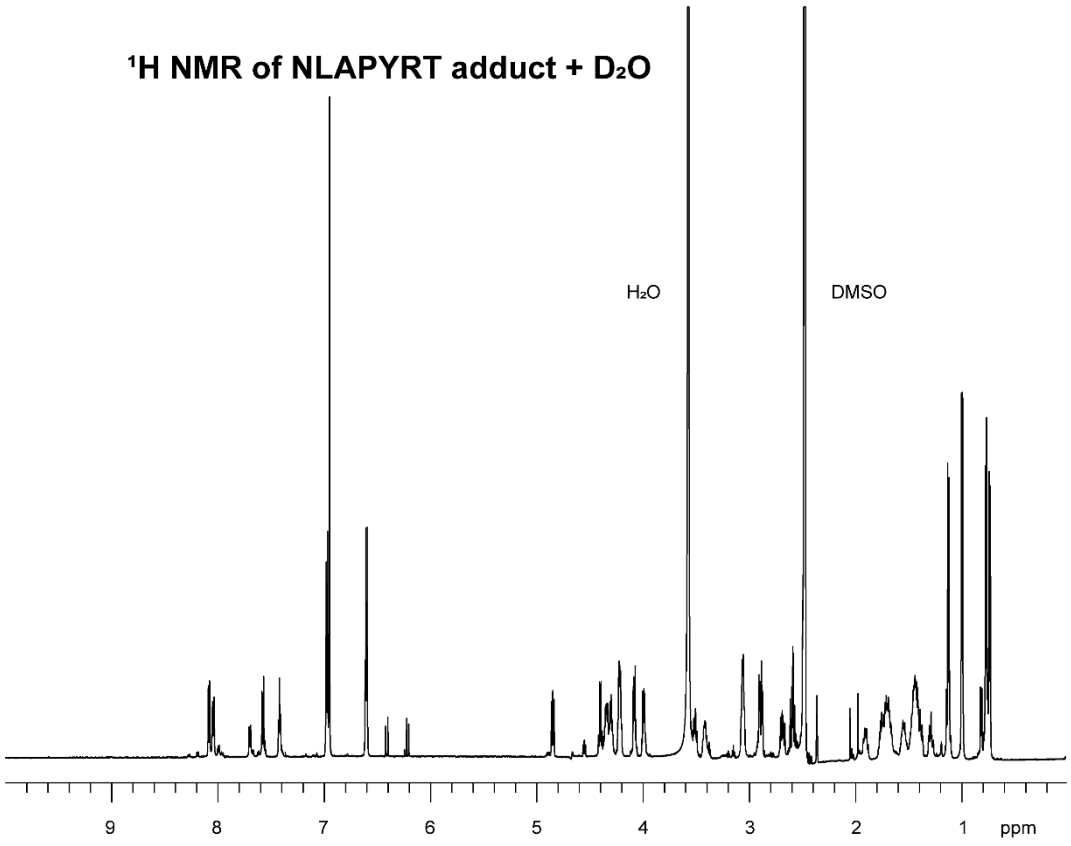

TOCSY NMR of NLAPYRT (fingerprint)
Leu-NH
Thr-NH
Asn-NH
Ala-NH
Arg-εNH
Asn-δNH₂
C-term-NH₂
Tyr
Asn-δNH₂
Tyr
F2 (ppm)
F1 (ppm)


COSY NMR of NLAPYRT
F2 (ppm)
F1 (ppm)


ROESY NMR of NLAPYRT
F1 (ppm)
F2 (ppm)


HSQC NMR of NLAPYRT
20
40
60
80
100
120
140
160
F1 (ppm)
9
8
7
6
5
4
3
2
1
F2 (ppm)


HMBC NMR of NLAPYRT adduct
ppm
20
40
60
80
100
120
140
160
180
200
220
F1 (ppm)
9
8
7
6
5
4
3
2
1
ppm
F2 (ppm)


¹H NMR of NLAPYRT adduct + D₂O
H₂O
DMSO
9
8
7
6
5
4
3
2
1
ppm