

Expanding Chemical and Structural Space in mRNA Display



Minglong Liu
刘明龙

Expanding Chemical and Structural Space in mRNA Display

**Uitbreiding van chemische en structurele ruimte in mRNA-
weergave**

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 20 september 2021 des ochtends te 10.15 uur

door

MINGLONG LIU

geboren op 18 mei 1992
te CHONGQING, China

Promotor:

Prof. dr. G.J.P.H. Boons

Copromotor:

Dr. S.A.K. Jongkees

The reasearch described in thesis was financially supported by a scholarship (NO. 201606010324) to Minglong Liu from the China Scholarship Council (CSC).

The proteins and scaffolds in chapter 4 were provided by Pepscan BV.

Expanding Chemical and Structural Space in mRNA Display

Minglong Liu
刘明龙

2021

Promotor: Prof. dr. G.J.P.H. Boons
Co-promotor: Dr. S.A.K. Jongkees

Front cover and back cover: designed with source from pixabay and biorender. The idea is from *The Little Prince*.

Table of Contents

List of Abbreviations	8
Chapter 1	
Introduction and thesis outline	10
Chapter 2	
Suppression of formylation provides an alternative approach to vacant codon creation in bacterial <i>in vitro</i> translation	40
Chapter 3	
Discovery of Potent Cyclic Peptide E-Selectin inhibitors through mRNA display	68
Chapter 4	
CLIPS/CuAAc utilization in mRNA display for novel peptide scaffolding.....	115
Chapter 5	
Summary and conclusions.....	153
Appendices	158

List of Abbreviations

AMB	4-(Aminomethyl)benzoic
CCL	Chemokines
CCR	Chemokines receptor
CFPS	Cell free protein synthesis
CMP	Cytidine monophosphate
CuAAc	Copper-catalyzed azide-alkyne cycloaddition
DBCO	Dibenzocyclooctyne
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
DHFR	Dihydrofolate reductase
DIBO	Dibenzocyclooctyne
DIC	N,N'-Diisopropylcarbodiimide
DIPEA	N,N-Diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EIC	Extracted ion chromatogram
ESI	Electrospray ionization
FAM	Carboxyfluorescein
FDA	US Food and Drug Administration
FIT	Flexible in vitro translation
FZD5	Frizzled-5
GFP	Green fluorescent protein
GPCR	G protein-coupled receptor
GTP	Guanosine triphosphate
HBS	HEPES buffered saline
HBTU	Hexafluorophosphate Benzotriazole Tetramethyl Uronium
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOBt	1-hydroxybenzotriazole
HPLC	High-performance liquid chromatography

HTS	High throughput screening
IF	Initiation factors
ITC	Isothermal titration calorimetry
LC	Liquid chromatography
MTF	Methionyl-tRNA formyltransferase
nAA	Non-canonical amino acid
NHS	N-Hydroxysuccinimide
NMR	Nuclear Magnetic Resonance
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PTM	Post-translational modification
RF	Release factors
RNA	Ribonucleic Acid
RT	Room temperature
SDS	Sodium dodecyl sulfate
SPAAC	Strain-promoted azide–alkyne cycloaddition
SPPS	Solid Phase Peptide Synthesis
SPR	Surface Plasmon Resonance
TBE	Tris/Borate/EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TEA	Triethylamine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
THPTA	Tris (benzyltriazolylmethyl) amine
TLC	Thin-layer chromatography
UPLC	Ultra high performance liquid chromatography
UV	Ultraviolet

Chapter 1

Introduction and thesis outline

In vitro translation provides a useful toolbox for the chemical biologist because of its multiple applications, such as protein/peptide synthesis¹, protein folding studies² and rapid identification of genes^{3,4}. By reprogramming the genetic code within this system we can incorporate nonstandard amino acids to further expand the applications, such as site-specific labelling of proteins⁵, mRNA display with an expanded chemical space⁶ and to provide handles for biorthogonal reactions on proteins/peptides.^{7,8} This chapter provides a theoretical basis for the experimental chapters that follow, outlining the steps involved in bacterial translation, how this can be used to pan libraries of peptides at ultra-high numerical diversity, how genetic code expansion can be used to extend the chemical diversity within these, and some of the opportunities for even further modification provided by bio-orthogonal reactions. The chapter concludes with an outline of the remainder of the thesis.

1.1 Prokaryotic translation

In general, the translation process can be divided into 3 parts: initiation, elongation and termination.^{9,10} The multiple check points for formation of the initiation complex provide precise control over the start of translation. The backbone amide bond is formed in the ribosome with help from elongation factors. When the ribosome reaches a stop codon, hydrolysis of the peptidyl-tRNA ester bond will happen with assistance from release factors. In the translation process, the ribosome thus has a primary role but receives support from multiple proteins.

Initiation of translation is the rate-limiting step of prokaryotic translation,¹¹ and is a factor in regulation of gene expression.¹² Investigating how initiation begins and how cells control translation requires a large investment of energy and time, analyzing it from structural, mechanistic and dynamical aspects.¹³ As shown in **Fig 1**, translation initiation begins with the 30S subunit. This unit can bind to IF1, the IF2-GTP complex and IF3 in a fast step to form an intermediate complex. If there are fMet-tRNA_{ini} and mRNA existing in the environment, they will slowly bind to this intermediate and change its structure to the 30S pre-initiation complex (30S pre IC). This step is the rate-limiting point for translation initiation. After this, the 50S unit will dock with the 30S pre IC and change the conformation to form the mature 70S unit. The fMet-tRNA_{ini} bound to the P-site of this 70S complex will start to yield peptide with GTP hydrolysis providing energy.

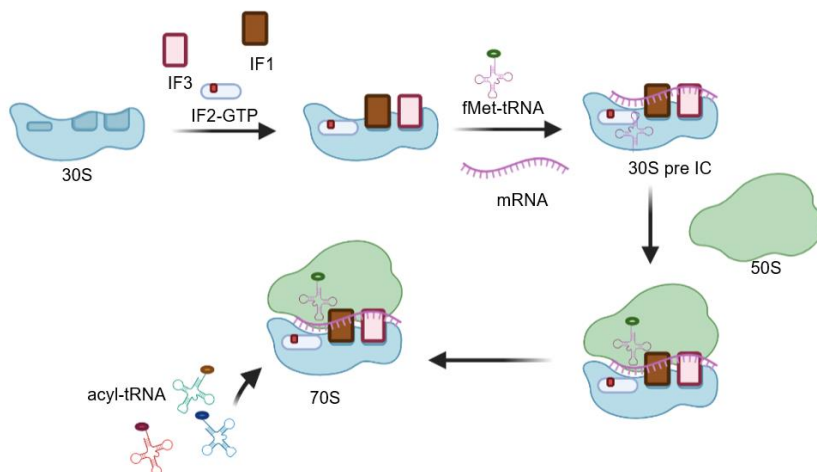


Figure 1. Sequence of binding events for prokaryotic translation initiation. Initiation factors 1-3 bind to the 30S subunit that can then bind mRNA and acylated-tRNA in the rate limiting step to form the 30S pre IC. This then completes assembly with the 50S subunit that then proceeds to elongation.

The quality control of initiation is a complex process and there are several checkpoints. In particular, how the bacteria accurately transfer fMet-tRNA_{ini} instead of Met-tRNA_{Met} onto the initiation site challenged researchers for several decades. In fact, Methionine aminoacyl-tRNA synthetase (MetRS) does not distinguish between tRNA_{ini} and tRNA_{Met} and carries out acylation on both tRNA.¹⁴ There are three factors for this quality control: 1. structural differences in the tRNA; 2. Formyl transferase activity; 3. EF-Tu recognition. Comparing the structure of tRNA_{ini} and tRNA_{Met}, we can see the unique anticodon stem loop of tRNA_{ini}. (**Fig 2**) that allows for selective recognition in initiation, and this difference is found in almost all kingdoms of life.^{15, 16} While the formyl transferase is not essential, since bacteria can still survive without its existence,¹⁷ its activity can selectively block the α -NH₂ of tRNA_{ini} due to the special structure of this acceptor stem and thus provide another checkpoint by preventing this methionine from being used in chain elongation. This selectivity was confirmed by an elegant experiment in which the tRNA_{Met} was mutated with the acceptor stem of tRNA_{ini} and it was found that the methionine in this mutated tRNA can then be formylated.¹⁸ Moreover, formylated fMet-tRNA_{ini} cannot be recognized by elongation factor thermo-unstable (EF-Tu, discussed below in elongation), and Met-tRNA_{Met} cannot be recognized by the 30S pre IC complex. Finally in prokaryotic initiation, recognition of mRNA derives from the Shine-Dalgarno (SD) sequence, a sequence feature on mRNA that only exists in prokaryotes.¹⁹ The SD sequence helps the

partially assembled ribosome attached to mRNA for protein translation at the correct codon (AUG). By this process, bacteria will always incorporate fMet as the first amino acid of proteins (which is subsequently removed after translation), and by combining these factors together bacteria can accurately transfer fMet onto the correct initiation site.

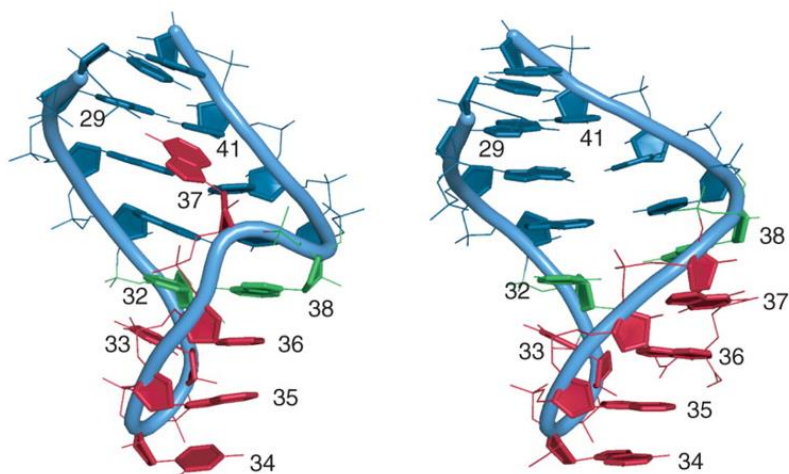


Figure 2. Conformation of the anticodon arm of *Escherichia coli* tRNA_{ini}. Left panel is *E. coli* tRNA_{ini} (PDB ID: 3CW6); Right panel is tRNA_{Phe} (PDB ID: 1EHZ). Bases are colored blue for the stem, red for the anticodon loop, and green to show the unusual base pair extending the anticodon loop (colors based on the left panel and retained per residue in the right). We can see the clear difference between the Cm32•A38 wobble-like base pair and the A37•(G29-C41) base pair of tRNA_{ini} compared to tRNA_{Phe}. Figure adapted from Barraud *et al.*¹⁶

Elongation starts with more aminoacyl-tRNA binding to the ribosome with help from EF-Tu. As shown in **Fig 3**, acyl-tRNAs are transferred to the A site of the ribosome. The peptide grows by amide bond formation between the C-terminus of the previous amino acid in the P site and the N-terminus of the amino acid in the A site. The process is catalyzed by the peptidyl transfer center of the ribosome.²⁰ After amide bond formation, the vacant tRNA will move to the E site and the newly produced peptidyl-tRNA will move to the P site. This process is promoted by elongation factor G (EF-G) with hydrolysis of GTP.²¹ The vacant tRNA in the E site can then be released from the ribosome. This peptide growth will continue until the ribosome reaches a stop codon on the mRNA. In bacteria translation and transcription can happen simultaneously in a coupled process, which is not possible in eukaryotes where these processes are physically separated between the nucleus and cytoplasm.²²

When the ribosome reaches a stop codon, termination of translation is carried out by recognition of three proteins: release factor 1 (reading UAA and UAG codons), 2 (reading UAA and UGA codons) and 3 (auxiliary function).²³ By hydrolysis of the ester bond in the ultimate peptidyl-tRNA, the peptide can be released from the ribosome. Unlike RF1 and RF2, RF3 does not recognize a stop codon but instead is responsible for removing RF1 or RF2 from the ribosome. A ribosome recycling factor (RRF) then triggers disassembly of the final mRNA-tRNA-ribosome complex together with EF-G and allows the entire process to repeat.²⁴

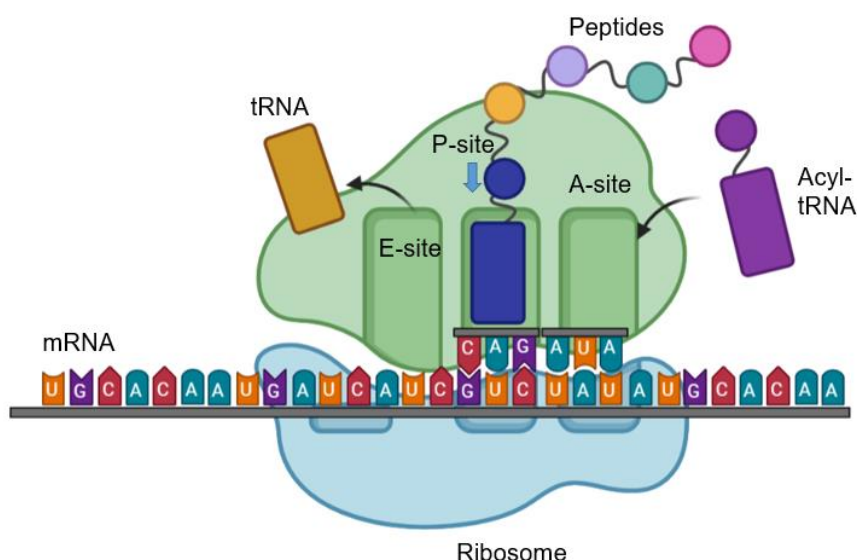


Figure 3. Elongation of translation. Acyl-tRNA are guided to the A-site by EF-Tu, followed by reaction in the peptidyl transfer center between this A-site and the P-site. Deacylated tRNA then moves to the E-site by action of EF-G together with GTP hydrolysis and translocation on the mRNA, where it can then be released.

1.2 Peptide display technologies

1.2.1 High-throughput Screening In Drug Discovery

High throughput screening (HTS) is a technology to assess large numbers of candidates, especially in drug discovery.^{25, 26} The concept of HTS is young, with this term first used in 1991, and before the 2000s there were relatively few articles employing this concept. Thanks to the rapid development of detection methods, such as surface plasmon resonance (SPR)²⁷, mass spectrometry²⁸ and RT-qPCR²⁹, more and more papers have since come out with this concept and have resulted in

discovery of a lot of binders in medical chemistry³⁰. According to FDA statistics, about 29% of approved drugs are from high throughput screening from 2016 to 2017.³¹ Depending on the methods used to construct the library, the size of library can vary from 10^2 to 10^{13} members. The most commonly used high throughput screening technologies in peptide discovery today are phage display, yeast display and mRNA display as illustrated conceptually in **Fig 4**.



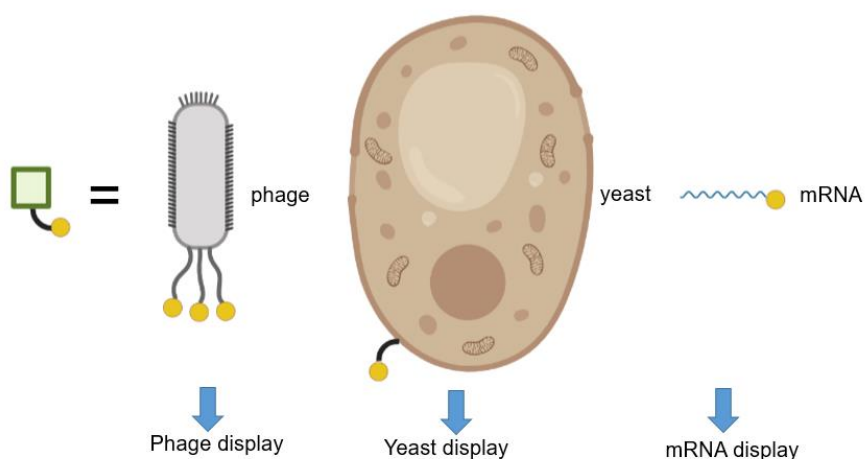


Figure 4. The steps in a general peptide display cycle. In this approach, a DNA-encoded library is constructed and translated to peptides in one of three different systems: bacteria, yeast or *in vitro* translation to give phage display, yeast display and mRNA display (or ribosome display). For all, a target protein is immobilized and the library is panned for affinity to this. The enrichment of binders can then be analyzed and these hits amplified for further rounds. After a few rounds, the sample can be sequenced and hits identified.

1.2.2 Phage Display

In the 1980s, phage display was first described by G.P. Smith. In this method the display of peptides on phage surface was achieved by injecting the corresponding plasmid.³² Phage display was further developed, and improved for application to proteins such as antibodies, by Winter. For these contributions, Smith and Winter were awarded the Nobel Prize in Chemistry in 2018 for their contribution to phage display.³³ The principal of the most common approach to phage display is to ligate a section of DNA encoding a peptide library into the pIII or the pVIII gene of M13 phage, which each encode a coat protein. The hybrid gene and remaining phage genes are inserted into an appropriate strain of *E. coli* bacteria, either directly or by means of a helper phage, to allow production of new phage particles that will then display the library on their coat. By immobilizing a target protein to the surface of a solid support, any phage that display recombinant peptides or proteins that can bind to the target will be recovered while the rest can be washed off. These retained phage can be eluted and amplified by again infecting *E. coli*. Repeated cycles can enrich the binding phage and the last cycle's phage can be sent to be sequenced to identify the ligands.^{34, 35} Some drugs that originated from phage display are already in clinical use,

such as Ecanllantide, Romiplostim and Tanzeum.³⁶ As an “improvement” of phage display, bacteria display can serve a similar function to derive peptide binders as well.³⁷ However, there is no clear and big advantage over phage display, and so bacteria display is not used widely.

1.2.3 Yeast Display

Due to limitations of bacterial protein synthesis, phage display can only present libraries of sequences amenable to prokaryotic translation and so are typically lacking in post-translational modifications such as glycosylation, phosphorylation and sulfation. To overcome this drawback, yeast display was developed by carrying out a similar selection approach on simple eukaryotic cells.³⁸ By incorporating the peptide's or protein's DNA sequences to the Aga2p protein gene, which expresses the Aga2p protein and mediates yeast cell mating, the peptides or proteins encoded by the library can be expressed and carry post-translational modifications while displayed on the yeast cell surface.³⁹ The use of magnetic separation and flow cytometry can help to isolate high affinity binder from this technology. Despite some advantages over phage display in terms of library diversity, yeast display retains some disadvantages, such as a smaller library size and differences in modifications from mammalian cells.³⁹

1.2.4 mRNA Display

Compared to phage display and yeast display, where surface protein fusion is employed, the linkage between expressed peptides and mRNA in mRNA display is by covalent linkage through puromycin.^{40, 41} As such, it is a much smaller assembly. Puromycin is an antibiotic that non-specifically stops bacterial translation by mimicking an acylated tRNA. When attached at the end of an mRNA it gives a covalent connection between the last amino acid of a translated sequence and its encoding sequence. There are 3 common ways to construct a puromycin-linked mRNA library: enzymatic ligation⁴², photo-crosslinking⁴³ and annealing⁴⁴. Of these, enzymatic ligation has favorable characteristics in terms of the stability of the mRNA compared to annealing and ease of use compared to photo-crosslinking. Because of the ease of construction of libraries in an *in vitro* translation system, high throughput screening time is significantly decreased from a few days to a few hours for each round. What's more, the library size (up to 10^{13}) is much bigger since mRNA display isn't limited by transfection. Early mRNA display employed only the standard amino acids, and because of this the cyclization approaches were limited to disulfide bonds or bio-orthogonal cross-linking reactions such as NHS chemistry (see below).⁴⁵ Even though the resulting hits can have strong target-binding affinity, in particular disulfides

can suffer from low peptide stability *in vivo* and NHS crosslinking can be difficult to control and limits library design.

1.2.5 RaPID

Random non-standard peptide integrated discovery (RaPID) is the combination of mRNA display and Flexible *In vitro* Translation (FIT, see section 1.3 below for more detail). By omitting specific amino acids from this translation, vacant codon boxes can be created that can then be decoded with non-standard acyl-tRNAs^{46, 47} to broaden the scope of library.^{48, 49} Using this approach, a range of natural product structural features can be incorporated into the displayed peptides, such as *N*-methylation, macrocyclization and D-stereochemistry.⁵⁰ This grants increased library diversity and allows screening against more challenging targets.^{51, 52} While chemical synthesis of peptide libraries allows access to many more building blocks and has been confirmed useful in screening for inhibitors of protein targets, the library size and binding affinity are quite limited and extensive medicinal chemistry is required to optimize the hits.⁵³ In contrast, RaPID offers a balance of high chemical diversity through reprogramming of some codons with specified nonstandard amino acids for targeted non-standard elements while still retaining enormous library diversity and so often much better optimised initial hits. In particular, macrocyclic peptide libraries are frequently constructed in RaPID because of their remarkable stability, bioavailability and cell permeability. For example, macrocyclic peptide Pakti-L1 showed an *in vitro* IC₅₀ of ~100 nM for inhibition of the cancer related Kinase Akt2⁵², S2iL5 showed a 3-4 nM K_D and IC₅₀ value against SIRT2⁵⁴, and *N*-methylated inhibitors against cancer-related ubiquitin ligase E6AP were found from RaPID and showed activity in a cell culture assay.⁵¹ The utility of genetic code reprogramming via FIT on mRNA display has led to diverse chemical biology applications. However, there are also remain some limitations to RaPID as currently employed. For instance, glycosylated amino acids cannot be reliably translated directly and so cannot be used in the RaPID system and vacant codon creation requires removal of canonical amino acids from the translation.

1.3 Genetic Code Expansion & Reprogramming

In natural translation, only 20 canonical amino acids (AAs) can be utilized for proteins/peptides. Even though these 20 amino acids can create billions of sequences and give many different functions for proteins, there are still limits to protein applications. To attempt to overcome some of these limits, much effort has been invested into expanding the scope of translatable amino acids. Over the last decades, with continual progress in chemistry and biology, various methodologies have been

established to expand the genetic code and thereby enabling more precise manipulation and increased functional evolution of proteins.⁵⁵ Despite this progress, there remains much interest in developing new methods for recoding the translation machinery and incorporating novel side chains or even backbones into proteins.⁵⁶

1.3.1 Incorporation of unnatural amino acids *in vitro*

Compared to *in vivo* translation, there are more options for reprogramming *in vitro* translation due to its less complex nature, easier access to the system, and the lack of need to avoid interfering with cellular survival. Generally, noncanonical amino acids can be charged onto tRNA for incorporating into peptides or proteins during *in vitro* translation in three ways: by aaRS-mediated reprogramming, by flexizymes in the flexible *in vitro* translation (FIT) or by chemo-enzymatic synthesis from truncated tRNA. By using aaRS enzymes, some close analogues of the natural substrates can easily be charged onto tRNA and yield a reprogrammed product if the canonical amino acid is omitted. This substrate scope can further be expanded by directed evolution. In contrast, the FIT system and chemo-enzymatic ligation of acylated nucleotides onto tRNA both can charge a much wider range of nonstandard amino acids, although each of these methods still has its own limits.

aaRS-mediated reprogramming can be a very convenient choice for genetic code reprogramming *in vitro*. For example, MetRS can recognize methionine mimics and transfer them onto the corresponding tRNA and this has been widely used for click handle incorporation, with azidohomoalanine and homopropargylglycine both being well accepted.⁵⁷ Up to now, over 100 noncanonical amino acids have been shown to be efficiently acylated onto tRNA by native aaRS enzymes.^{46, 58} Importantly, many of these acyl-tRNA are also efficiently used in translation and give a high yield of modified proteins/peptides. Genetic code reprogramming by this method is less labor-intensive, at least once an enzyme with appropriate activity has been identified, because most of the amino acids are commercially available and no unstable activated esters need to be isolated. The disadvantage of this approach is also obvious: only close analogues can be reprogrammed.

An alternative to aaRS-mediated reprogramming is flexible *in vitro* translation (FIT). It is a simple but powerful tool to incorporate more diverse building blocks, including many which are poorly or not at all tolerated by aminoacyl-tRNA synthetases (aaRS), into peptides and proteins by *in vitro* translation. Notable examples include *N*-methyl amino acids, D-amino acids and even some β -amino acids.⁵⁹ (**Fig 5A**) Flexizymes are

a type of ribozyme (catalytic RNA) that can prepare a wide range of acyl-tRNAs. The 1st generation of flexizyme was generated by *in vitro* selection from a random sequence of RNA, but could only afford specific amino acylation onto a small range of tRNA.⁶⁰ It only recognizes aromatic amino acids but could not be used as a versatile catalyst because of limited activity and specific recognition of tRNA. To get rid of this limitation, further evolution from the 1st generation flexizyme made it possible to aminoacylate diverse tRNAs, and also expanded the amino acid substrate scope to remove the requirement for an aromatic amino acid.⁶¹ Using this approach, acyl-tRNAs can easily be transferred into peptides on vacant codon boxes to yield the desired nonstandard peptides in just one day.⁵⁹ The FIT system in particular has advantages over *in vivo* reprogramming because of how easy and fast implementation of a new amino acid is. In addition, *in vitro* translation allows precise control over the composition of the reaction, not just for small molecules but also protein factors. Taken together, these advantages make *in vitro* reprogramming in the FIT system attractive for studying and engineering translation. Nevertheless, vacant codon boxes are most often created in this approach by sacrificing canonical amino acids to reassign sense codons to noncanonical amino acids, and so this makes its utility in protein translation quite limited.^{62, 63} In addition, for more extensive reprogramming the ideal concentration of each single component may need to be optimized for efficient translation.

Another chemoenzymatic approach is also possible for genetic code reprogramming, by synthesizing a dinucleotide coupled to an unnatural amino acid and then ligating this to a truncated tRNA. (**Fig 5B**) This more chemical method of charging permits the introduction of nonstandard amino acids that are even more radically different from the canonical set. The core of this technique is the synthesis of an aminoacylated dinucleotide pCpA bearing the desired nonstandard amino acid. Using a mildly activated ester of the amino acid together with a dinucleotide where the only remaining nucleophiles are the 2' and 3' hydroxyls leads to selective charging. Important here is the use of an amine protecting group that can subsequently be removed without damaging the relatively fragile nucleotide ester that results, which is achieved with an *N*-pentenoyl group that is deprotected by aqueous iodine. This aminoacylated dinucleotide building block is then conjugated by a ligation reaction to a suppressor tRNA lacking the 3' terminal CA dinucleotide in its acceptor stem, and the resulting orthogonal tRNAs can be employed directly for *in vitro* translation reactions. To avoid competition with canonical amino acids, this method usually utilizes stop codon suppression. However, for either of these two chemoenzymatic

methods the efficiency of reprogramming is compromised by competition from release factors or natively charged elongation tRNA unless these can be omitted.⁵⁸ While this method in principle can be applied to a wide range of amino acids, its time consuming synthetic process and relatively complex operation hinder its application.

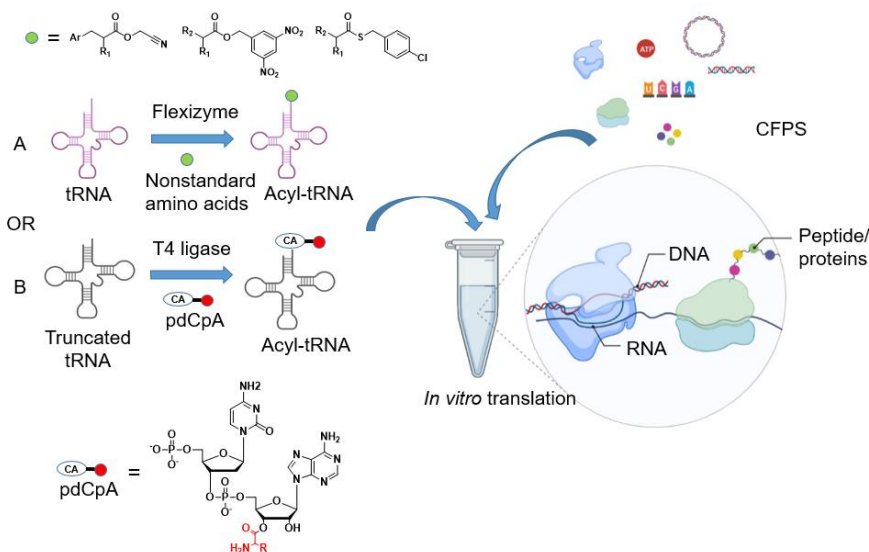


Figure 5. Genetic code reprogramming *in vitro* by generating modified acyl-tRNA. A. Nonstandard amino acids are charged onto tRNA by flexizyme and utilized by *in vitro* translation; B. Nonstandard amino acid-pdCpA is ligated onto truncated tRNA to form the mature acyl-tRNA. These acyl-tRNA can be utilized in cell-free protein synthesis to yield modified peptides or proteins.

1.3.2 Incorporation of unnatural amino acids *in vivo*

Genetic code reprogramming in cells is challenging because it requires partial breaking of the checkpoints in the complex translation system. Compared to *in vitro* translation, reprogrammed *in vivo* translation requires stricter checks as living cells may suffer toxic effects from poor translational control, removing factors to vacate codons is generally not possible, and transferring all required components into cells can also be difficult. These features make it an ambitious task to carry out reprogramming in cells. Despite this, multiple approaches have been developed for unnatural amino acid incorporation onto tRNA.

The tRNA^{Pyl}/PylRS pair is found originally in a small number of archaea and bacteria for decoding a stop codon with pyrrolysine.⁶⁴ The most commonly used approach for

reprogramming in cells is engineering such orthogonal tRNA and aminoacyl tRNA synthetase pairs.⁶⁵ (**Fig 6**) This particular tRNA/aaRS pair has been shown to act orthogonally in bacteria and in eukaryotes,^{66, 67} and using this approach it is possible to charge nonstandard amino acids onto tRNA without interfering with charging of any other amino acids, subsequently directing it to an amber stop codon in translation (UAG). Using directed evolution, *in vivo* reprogramming with such orthogonal pairs has been expanded to ever more divergent pyrrolysine mimics, such as alkynes or azides for CuAAC reaction, probes for photocrosslinking *etc.*⁶⁶ Nevertheless, the interaction between EF-Tu and aminoacyl-tRNA^{Pyl} is relatively poor and stop codon translation is in competition with release factor binding, which both influence the yield. Further optimization has focused on the interaction between EF-Tu and another orthogonal tRNA, that of *Mj*-tRNA^{Tyr} (tRNA from *Methanocaldococcus jannaschii*).⁶⁸ The *Mj*-tRNA/TyrRS pair is orthogonal in *E.coli* and significantly improves the translation yield. Various strategies have been developed for improved fidelity and yield, such as further engineering the tRNA⁶⁹ and aaRS⁶⁶ for not only increased substrate scope but also for improved translation efficiency.

An even more demanding task than genetic code reprogramming (replacing an amino acid or stop codon with another) is expanding the genetic code (increasing the total number of codons available). In one example this was achieved by utilizing four-base codons instead of three. Theoretically, using this approach the codons can be expanded from 4³ (64) to 4⁴ (256) and nAAs can be assigned to the vacant codon boxes.^{70, 71} While this appears to offer a dramatic expansion, a four-base codon overlaps with a three-base codon and so this approach for now still has a fatal drawback in the low fidelity of decoding and replication.

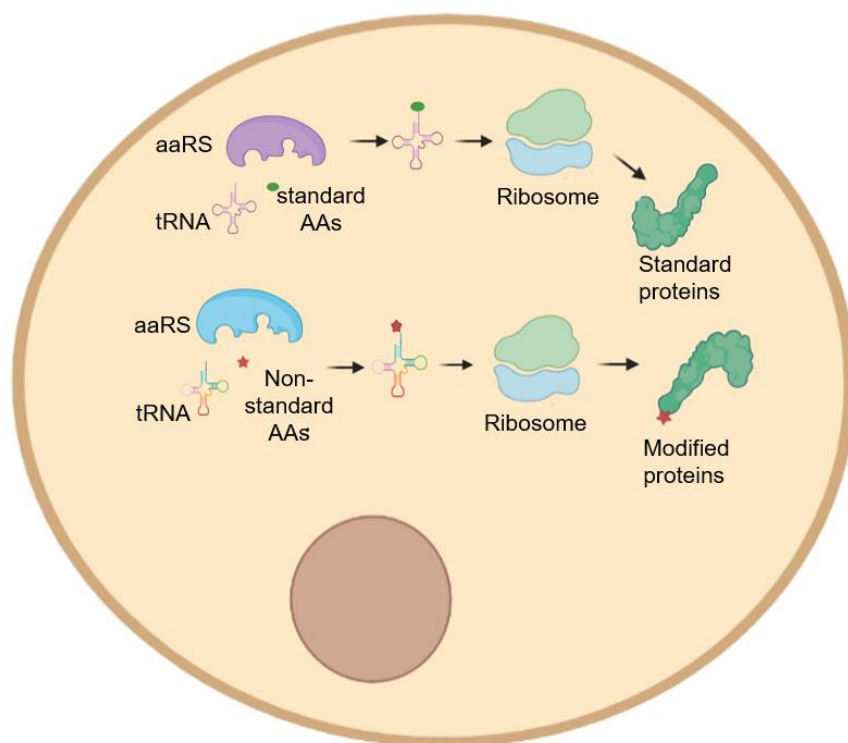


Figure 6. Genetic code reprogramming *in vivo*, showing the orthogonality of nonstandard-tRNA/aaRS pair. Engineering aaRS (or tRNA) can incorporate nonstandard amino acids onto specific tRNA. The acyl-tRNA can be utilized by eukaryotic (or prokaryotic) ribosomes to yield precisely modified proteins.

Substantial advances have been made in the endeavor to engineer the translation apparatus *in vitro* and *in vivo* to allow access to an ever expanding repertoire of non-canonical amino acids. A perhaps more demanding task is to utilize these methodologies in creative ways and to create new opportunities for chemical biology. One good example is genetic code reprogramming applied to drug discovery. The combination of display technologies with genetic code reprogramming holds a lot of promise for drug discovery since reprogrammed translation not only offers increased library diversity directly but also provides access to a range of nonstandard amino acids which can be used for biorthogonal reactions. Such reactions will be further discussed in the next section.

1.4 Chemoselective and biorthogonal reactions

In the past decades, a number of aqueous-compatible reactions have been developed

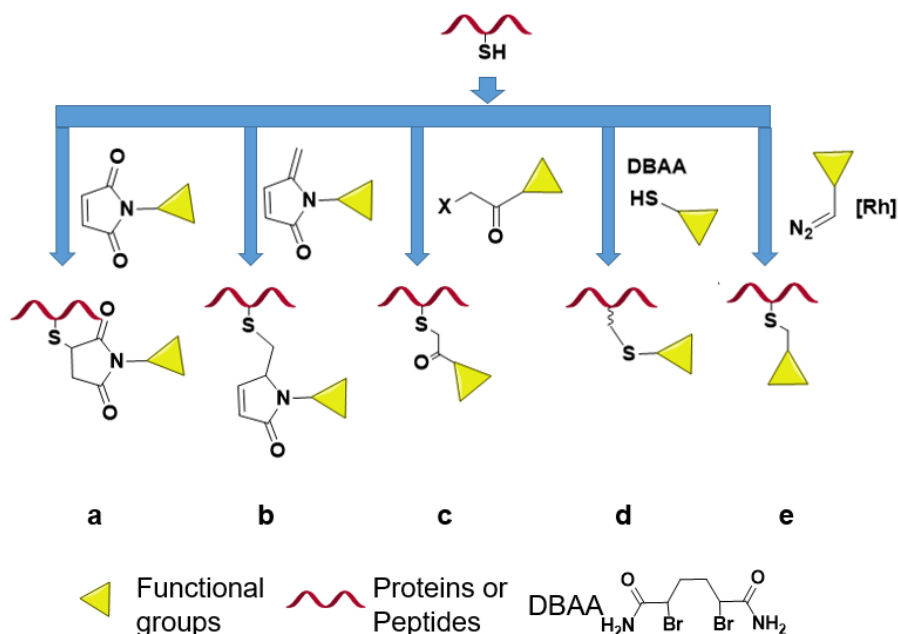
for modification of biomolecules. Chemists continue to try to apply these reactions in ever more complex biological systems to give a new meaning to chemical synthesis. These reactions should be mild and selective, since they must tolerate the many functional groups in living system such as diverse nucleophiles, reductants and oxidants. They also should not influence the biomolecules' structures and functions too much after modification. With application of these reactions, chemists can begin to do reactions on cancer cells or mice instead of in flasks. For example, the strain-promoted azide-alkyne cycloaddition (SPAAC) reaction can be used for imaging in zebrafish or mice, and conjugation of a drug to an antibody can be used to carry it selectively to cancer cells or for precision radiotherapy.⁷² These biorthogonal reactions have motivated chemists to dig into classic chemistry reactions to find new uses in biological systems and develop new applications in chemical biology. In this section, we focus mainly on biorthogonal reactions that can be exploited in mRNA display. Comparing this case to other applications of biorthogonal reactions, modification in mRNA display may ask for a higher reaction rate and conversion because of the low concentration of displayed peptide and nucleic acids have different reactivity profiles from proteins, but it is in other ways more forgiving in reaction selectivity as bulk modification of proteins is not detrimental once enzymatic reactions in library building are complete.

1.4.1 Reactions with canonical amino acids

Cysteine is a good reaction site for selective modification because of its robust nucleophilicity. Modification on cysteine can be achieved selectively, as compared to other nucleophilic side chains such as lysine, tyrosine and histidine, by adjusting the pH. What's more, the low abundance of cysteine in proteins gives better access to single modifications on proteins.⁷³ Maleimide is a classic reagent for exploiting the nucleophilicity of cysteine thiols (**Scheme 1a**) and has good selectivity. There exist a large number of maleimide reagents that are commercially available, including biotin probes and fluorophore labels. However, the maleimide pathway can produce some byproducts because of elimination from conjugated maleimide, hydrolysis of the product or reaction with other side chains such as lysine. To counteract issues with the stability and selectivity of maleimide, other functional groups have also been investigated for reaction with cysteine. Methylene pyrrolene is another tool that exploits a conjugate-addition type of reactivity and exhibits better selectivity and stability over maleimide.⁷⁴ (**Scheme 1b**) Bromo- and iodo-acetamide are also useful and commonly used for thiol modification, in particular for capping off thiols in proteomics.⁷⁵ This reaction was found almost one century ago,⁷⁶ with strong

electrophiles such as α -halocarbonyls able to react with thiols under mild condition (**Scheme 1c**). It has since been shown that genetic code reprogramming of amino acids displaying α -halocarbonyls gives a functionality that can react with thiols spontaneously and selectively during translation. This methodology was applied to synthesize macrocyclic peptides in high-throughput screening such as mRNA display.⁵⁰ What's more, a similar approach using crosslinking between thiols and active halides such as bromoxylene to construct cyclic peptide called CLIPSTM aids in exploring different cyclic peptides in phage display.⁷⁷

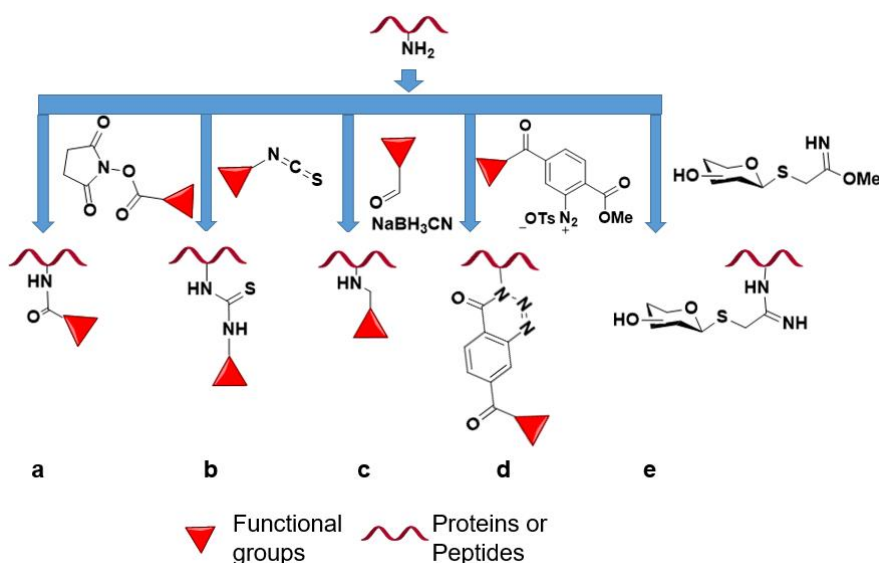
Another conjugation methodology derived from cysteines and based on the idea of polarity inversion (umpolung) is to synthesize dehydroalanine, for example by use of phenylthiosulfonates or dibromoadipic-bisamide (DBAA), then employing the resulting electrophile in a conjugate addition with an exogenous thiol. This approach can be exploited to carry out largely 'traceless' thioglycosylation and biotinylation in proteins as well as in peptide libraries^{78, 79} (**Scheme 1d**), but with loss of amino acid stereochemistry. Finally, it was reported that a rhodium catalyst can mediate the conjugation of diazo compounds with cysteine.⁸⁰ (**Scheme 1e**) However, its reaction rate and stability may limit the development of this approach for *in vitro* and *in vivo* system.⁸¹ Other methods⁸², such as radicals, nitrenes and photo-mediated cross linking are higher risk in mRNA display due to high reactivity and the likelihood of damage to nucleic acids. Selenocysteine, which can be incorporated into proteins, can also react in a similar way to the thiol from cysteine, but because of the increased nucleophilicity of selenols will react faster still. This also gives access to tiered nucleophilicity, from amines to thiols to selenols, for controlled multiple sequential modifications.



Scheme 1. Bioconjugation reactions at thiols.

Aside from thiols, amines are a popular reaction site for efficient bioconjugation. Lysine residues are commonly used because of the high availability of primary amines in natural proteins and peptides. NHS esters (*N*-hydroxysuccinimide) can carry out modifications on lysine under mild conditions.⁸³ (**Scheme 2a**) Even though the conjugation may be compromised somewhat because of some side reactions and low regioselectivity, the problems can be minimized to a certain degree by modifying pH for site-specific modification (for example targeting the *N*-terminus by use of mild acidic conditions to exploit the roughly 2 pK_a units difference in basicity).⁸⁴ Isothiocyanates can also react with lysine residues and can be exploited for labelling *in vivo*.⁸⁵ (**Scheme 2b**) This reaction also has been in use for a long time already, being first reported in 1937.⁸⁶ It has several shortcomings, such as reversible reaction with thiols and phenols and limited buffer choice, but despite this its moderate reactivity and water-stability make it a popular functionality for bioconjugation.⁸³ Reductive amination reactions such as with aldehydes and sodium cyanoborohydride shows amine-selectivity and so also reacts on Lys and the *N*-terminus.⁸⁷ (**Scheme 2c**) The most important drawback of this method is the need to use the water-sensitive reagent sodium cyanoborohydride, which could also have potential side-reactions on proteins such as reducing disulfide bonds, and the associated release of toxic cyanide. Amine-selective modification was also achieved by using diazonium terephthalates,

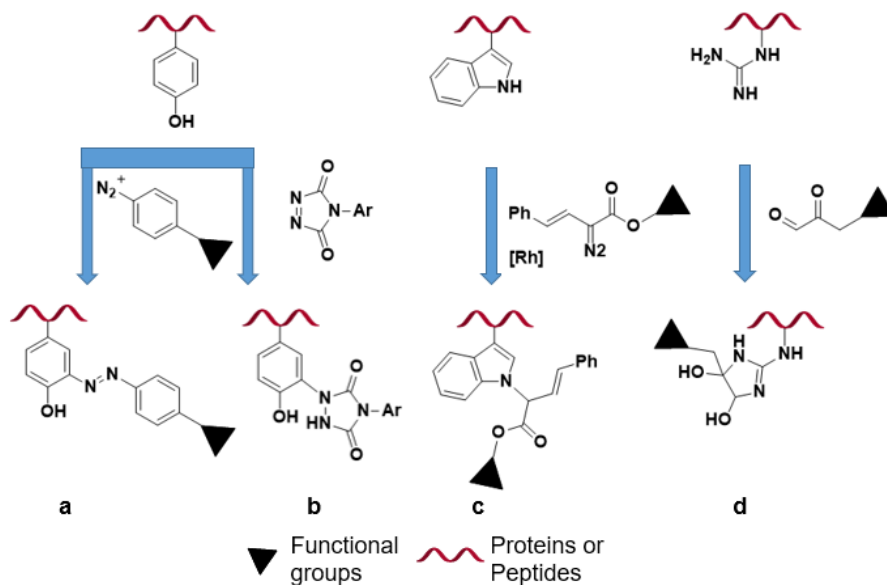
which can form a stable triazine ring. (**Scheme 2d**) This method suffers from the instability of diazonium salts, which limits its widespread use. Another method used for lysine modification is through use of 2-imino-2-methoxyethyl reagents to form an amidine bond. (**Scheme 2e**). RNase A and α -L-rhamnopyranosidase were successfully used to demonstrate site-specific incorporation of a glycan by this method.^{88, 89} The reaction can be finished in a short time, but this methodology has poor regioselectivity and can form byproducts with Tyr and even carboxylate groups.⁹⁰



Scheme 2. Bioconjugation reactions at lysine and the *N*-terminus.

While less common, other canonical amino acids besides lysine and cysteine can also serve as bioconjugation sites, including tyrosine, tryptophan and arginine. Tyrosine has an electron rich aromatic ring in its side chain which can be a chemoselective reaction site. It was found that several diazonium salts can react with tyrosine at the ortho position of its phenol group.^{91, 92} (**Scheme 3a**) As a further development of this approach, 2-naphthol analogues were found to have a faster reaction rate and better selectivity for tyrosine.⁹³ Another method for tyrosine modification is by using azomaleimides.⁹⁴ (**Scheme 3b**) A drawback of this method is the low stability of cyclic diazodicarboxamides in water, which comes from side reactions with lysine. Tryptophan is typically of low abundance in protein sequences (~1%), much like cysteines, but most proteins do contain at least 1 tryptophan residue.⁹⁵ Thus tryptophan provides a useful site for unique modifications, which can

be achieved selectively by diazo compounds with Rh catalyst.⁹⁶ (**Scheme 3c**) However, the excess amount of diazo compound and acidic conditions required (pH 1.5-3.5) limit the applicability, although further optimization may be able to expand the pH range for the reaction.⁹⁷ Arginine, whose side chain has a pK_a value around 12, is mainly protonated in nature. Because of this, most arginine labelling approaches are based on the chemistry of α -dicarbonyl compounds. Site-specific modification on arginine can be achieved by using an α -oxo-aldehyde compound.⁹⁷ (**Scheme 3d**) The conditions for this approach are mild and the reaction is selective since the required geometry prevents reaction with other side chains such as lysine. Overall, these biorthogonal reactions are typically less often employed because of the disadvantages mentioned above, namely low reaction rates, harsh conditions and side reactions.



Scheme 3. Examples of bioconjugation reactions at other amino acids such as tyrosine, tryptophan and arginine.

1.4.2 Reactions with non-canonical amino acids

Even though a wide range of reactions can be used for biorthogonal reactions with canonical amino acids, the increasing demand for new reactivities with improved selectivity, stability and mildness means that it is not sufficient to focus only on this limited set of amino acids. Additionally, the introduction of nonstandard amino acids onto proteins/peptides makes precise site-specific modification a key tool for evolving

and adjusting biological systems *in vitro* or *in vivo*. A general overview of some common biorthogonal reactions not involving the canonical amino acids is shown in **Fig 7**.

The Staudinger reduction is a reaction of azides with phosphine to form an amine, which was discovered about 100 years ago.⁹⁸ In chemical biology, this was developed into a Staudinger ligation by installing an ester on the phosphine to allow capture of an intermediate in the formation of an amide bond.⁹⁹ However, the low stability of the ester and the phosphine (hydrolysis and oxidation, respectively) and a low kinetic rate constant for the reaction ($\sim 10^{-3} \text{ M}^{-1}\text{s}^{-1}$) limits its application in living systems. Other slow biorthogonal reactions ($k \sim 10^{-4} \text{ M}^{-1}\text{s}^{-1}$), such as ligation of a ketone/aldehyde with a hydrazide, have also been developed for bioconjugation, but a source of aldehyde is hard to achieve *in vivo*.¹⁰⁰

The most quoted biorthogonal reaction is copper catalyzed azide-alkyne cycloaddition (CuAAC).¹⁰¹ The CuAAC reaction is easy to handle, because of its high yield without any byproducts, wide tolerance for different conditions and high selectivity.¹⁰² Handles for this 'click' reaction, an azide or alkyne, can be easily incorporated into proteins by misacylation using the MetRS enzyme during *in vitro* translation (See section 1.3). However, copper is toxic to cells above low concentrations ($\sim 1 \text{ mM}$), which limits the application of CuAAC reaction to *in vivo* systems, although ligands can aid in this respect. To avoid the problem of copper toxicity, a copper-free click reaction has been developed. This is achieved by using ring strain to promote azide alkyne cycloaddition (SPAAC).¹⁰³ This reaction typically employs a cyclooctyne activated by fluorination (DIFO) or ring fusion with a three membered ring (BCN) or aromatic rings with one of several linkages (DIBO, DBCO, BARAC), all reacting without a metal catalyst.¹⁰⁴ Even though the SPAAC reaction rate ($1\sim 10 \text{ M}^{-1}\text{s}^{-1}$) is slower than the CuAAC reaction, the more broad applicability of SPAAC for *in vivo* labelling means it has seen extensive use and demonstrates its effectiveness as a tool in chemical biology.

Another breakthrough in biorthogonal reactivity is the inverse electron demand Diels-Alder cycloaddition between tetrazines and strained alkenes or alkynes (such as cyclopropenes and trans-cyclooctenes).¹⁰⁵⁻¹⁰⁷ The reaction happens very rapidly and specifically to form stable adducts. This reaction occurs under mild conditions and in water, with almost no side products. The limiting factors for this reaction are the difficult syntheses of the two reacting components and the stability of the required

alkynes, tetrazines or alkenes, with the half life time of these substrates ranging from days to months.

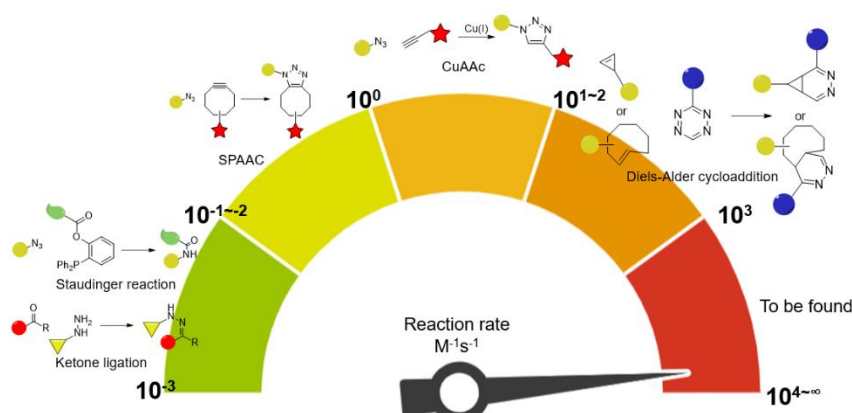


Figure 7. Some Biorthogonal reaction for bioconjugation, arranged by rate constant.

The various handles required for these bioorthogonal reactions can in many cases be incorporated into peptides and proteins using reprogrammed *in vitro* translation, and together provide a powerful tool for both bioorganic synthesis and biochemical studies, such as in site-specific modifications on proteins and in drug discovery. Even though a number of techniques for bioconjugation have already been developed and exploited, there remain limits for application in modifying peptides and proteins because of the complex environments of biological systems. Further fast, convenient, and stable bioconjugation reactions to overcome the limitations mentioned above undoubtedly remain to be found.

1.5 Aim and outline

As an introduction chapter, this part aims to provide the reader with a focused review of the exciting and flourishing field of chemical biology and to provide context for the experimental chapters that follow. We aim to show that *in vitro* translation is a useful tool for protein engineering, and in particular that it can be utilized in mRNA display as well as small scale protein synthesis with precise control over incorporation of different functional groups. We also discuss the further synergy of this approach with biorthogonal reactions and genetic code reprogramming. In the three experimental chapters that follow we present work in which we use an understanding of bacterial translation to further develop *in vitro* genetic code reprogramming, and use mRNA display of peptides with an expanded chemical scope from genetic code reprogramming and bioorthogonal chemistry to discover new ligands for several

proteins relevant to human health.

In Chapter 2, we develop a new strategy for creating a vacant reprogrammable initiation codon. We found that either removal of the formyl donor or inhibition of the formyl transferase is sufficient to prevent formylation of Met-tRNA_{ini} and so liberates the initiation codon. Our method allows easy incorporation of non-canonical functional groups at the initiation site without disrupting the translation of methionine or its analogues at other AUG codons. To further demonstrate our methodology, we apply this protocol to macrocyclization of peptides by CuAAc and dual labelling of proteins with SPAAC.

In Chapter 3, we describe the discovery of tight-binding peptide ligands for E-selectin, chosen as a target because of its crucial role in the inflammatory cascade and in cancer metastasis. We describe literature-guided design of several lectin-focused binding factors: a sulfonate, a vicinal diol and an inverted electron density aromatic ring. We synthesized amino acids incorporating each of these factors and successfully demonstrated their incorporation into peptides by reprogrammed *in vitro* translation. Messenger RNA display was employed to select hits that bind E-selectin from a random library incorporating these factors. Sequencing showed that only one of the binding factors was present (a sulfonate), but synthesis and characterization of the most enriched hits by SPR demonstrated that these had good binding affinity to selectins, including one with a particularly slow off rate at the limit for detection with this method.

In Chapter 4, we investigated the use of tandem CLIPS/CuAAc chemistry to form unique tricyclic peptides in mRNA display. Following initial optimization on synthetic and translated peptides to derive a 'one-pot approach', we validated its use in mRNA display using 2 different scaffolds and 4 different target proteins: FZD5, mAb197, D-CXCL8 and ST6Gal1. Libraries with 3 randomized fragments of 4 amino acids each embedded in these tricycles showed strong and clear enrichment for the first two targets and some enrichment for the remaining two. Sequencing indicated conserved motifs for each target and these were dependent on the scaffold used, while for mAb197 many of the hits found matched a known interacting motif. Together our results indicate that binders have successfully been found. Our methodology and results thus indicate that CLIPS/CuAAc is a valuable addition to the toolbox for macrocyclic drug discovery.

References

1. Khambhati, K.; Bhattacharjee, G.; Gohil, N.; Braddick, D.; Kulkarni, V.; Singh, V., Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. *Front. Bioeng. Biotechnol.* **2019**, *7*, 248-263.
2. Wruck, F.; Katranidis, A.; Nierhaus, K. H.; Büldt, G.; Hegner, M., Translation and folding of single proteins in real time. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 4399-4407.
3. Miyamoto, S.; Qin, J.; Safer, B., Detection of early gene expression changes during activation of human primary lymphocytes by in vitro synthesis of proteins from polysome-associated mRNAs. *Protein Sci.* **2001**, *10*, 423-433.
4. Chong, S., Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. *Curr. Protoc. Mol. Biol.* **2014**, *108*, 16.30.1-16.30.11.
5. van Vught, R.; Pieters, R. J.; Breukink, E., SITE-SPECIFIC FUNCTIONALIZATION OF PROTEINS AND THEIR APPLICATIONS TO THERAPEUTIC ANTIBODIES. *Comput. Struct. Biotechnol. J.* **2014**, *9*, e201402001.
6. Lipovsek, D.; Plückthun, A., In-vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **2004**, *290*, 51-67.
7. Devaraj, N. K., The Future of Bioorthogonal Chemistry. *ACS Cent. Sci.* **2018**, *4*, 952-959.
8. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **2009**, *48*, 6974-6998.
9. Robinson, A.; van Oijen, A. M., Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat. Rev. Microbiol.* **2013**, *11*, 303-315.
10. Rodnina, M. V., Translation in Prokaryotes. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*.
11. Winkler, W. C.; Breaker, R. R., REGULATION OF BACTERIAL GENE EXPRESSION BY RIBOSWITCHES. *Annu. Rev. Microbiol.* **2005**, *59*, 487-517.
12. McCarthy, J. E. G.; Gualerzi, C., Translational control of prokaryotic gene expression. *Trends Genet.* **1990**, *6*, 78-85.
13. Gualerzi, C. O.; Pon, C. L., Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell. Mol. Life Sci.* **2015**, *72*, 4341-67.
14. Nakanishi, K.; Ogiso, Y.; Nakama, T.; Fukai, S.; Nureki, O., Structural basis for anticodon recognition by methionyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* **2005**, *12*, 931-932.
15. Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A.

- C.; Weir, J. R.; Ramakrishnan, V., Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **2006**, *313*, 1935-1942.
16. Barraud, P.; Schmitt, E.; Mechulam, Y.; Dardel, F.; Tisné, C., A unique conformation of the anticodon stem-loop is associated with the capacity of tRNA^{fMet} to initiate protein synthesis. *Nucleic Acids Res.* **2008**, *36*, 4894-4901.
 17. Guillon, J. M.; Mechulam, Y.; Schmitter, J. M.; Blanquet, S.; Fayat, G., Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of *Escherichia coli*. *J. Bacteriol.* **1992**, *174*, 4294-4301.
 18. Lee, C. P.; Dyson, M. R.; Mandal, N.; Varshney, U.; Bahramian, B.; RajBhandary, U. L., Striking effects of coupling mutations in the acceptor stem on recognition of tRNAs by *Escherichia coli* Met-tRNA synthetase and Met-tRNA transformylase. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9262-9266.
 19. Shine, J.; Dalgarno, L., Determinant of cistron specificity in bacterial ribosomes. *Nature* **1975**, *254*, 34-38.
 20. Trobro, S.; Åqvist, J., Mechanism of peptide bond synthesis on the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12395-12400.
 21. Wintermeyer, W., Translocation on the Ribosome. In *Encyclopedia of Biophysics*, Roberts, G. C. K., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 2650-2655.
 22. Jackson, R. J.; Hellen, C. U. T.; Pestova, T. V., The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 113-127.
 23. Zhou, J.; Korostelev, A.; Lancaster, L.; Noller, H. F., Crystal structures of 70S ribosomes bound to release factors RF1, RF2 and RF3. *Curr. Opin. Struct. Biol.* **2012**, *22*, 733-742.
 24. Hirokawa, G.; Demeshkina, N.; Iwakura, N.; Kaji, H.; Kaji, A., The ribosome-recycling step: consensus or controversy? *Trends Biochem. Sci* **2006**, *31*, 143-149.
 25. Inglese, J.; Auld, D. S., High Throughput Screening (HTS) Techniques: Applications in Chemical Biology. *Wiley Encyclopedia of Chemical Biology* **2008**, 1-15.
 26. Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S., Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **2011**, *10*, 188-195.
 27. Maynard, J. A.; Lindquist, N. C.; Sutherland, J. N.; Lesuffleur, A.; Warrington, A. E.; Rodriguez, M.; Oh, S.-H., Surface plasmon resonance for high-throughput ligand screening of membrane-bound proteins. *Biotechnol. J.* **2009**, *4*, 1542-1558.

28. Rohman, M.; Wingfield, J., High-Throughput Screening Using Mass Spectrometry within Drug Discovery. In *High Throughput Screening: Methods and Protocols*, Janzen, W. P., Ed. Springer New York: New York, NY, 2016; pp 47-63.
29. Zhang, H.; Zhu, G., Quantitative RT-PCR assay for high-throughput screening (HTS) of drugs against the growth of *Cryptosporidium parvum* in vitro. *Front Microbiol* **2015**, *6*, 991-991.
30. Sahdeo, S.; Tomilov, A.; Komachi, K.; Iwahashi, C.; Datta, S.; Hughes, O.; Hagerman, P.; Cortopassi, G., High-throughput screening of FDA-approved drugs using oxygen biosensor plates reveals secondary mitofunctional effects. *Mitochondrion* **2014**, *17*, 116-125.
31. Brown, D. G.; Boström, J., Where Do Recent Small Molecule Clinical Development Candidates Come From? *J. Med. Chem.* **2018**, *61*, 9442-9468.
32. Smith, G. P., Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985**, *228*, 1315.
33. The Nobel Prize in Chemistry 2018. NobelPrize.org. .
34. Smith, G. P.; Petrenko, V. A., Phage Display. *Chem. Rev.* **1997**, *97*, 391-410.
35. Kehoe, J. W.; Kay, B. K., Filamentous Phage Display in the New Millennium. *Chem. Rev.* **2005**, *105*, 4056-4072.
36. Mimmi, S.; Maisano, D.; Quinto, I.; Iaccino, E., Phage Display: An Overview in Context to Drug Discovery. *Trends Pharmacol. Sci.* **2019**, *40*, 87-91.
37. Kenrick, S. A.; Daugherty, P. S., Bacterial display enables efficient and quantitative peptide affinity maturation. *Protein engineering, design & selection : PEDS* **2010**, *23*, 9-17.
38. Boder, E. T.; Wittrup, K. D., Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **1997**, *15*, 553-557.
39. Gai, S. A.; Wittrup, K. D., Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **2007**, *17*, 467-473.
40. Amstutz, P.; Forrer, P.; Zahnd, C.; Plückthun, A., In vitro display technologies: novel developments and applications. *Curr. Opin. Biotechnol.* **2001**, *12*, 400-405.
41. Roberts, R. W.; Szostak, J. W., RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12297-12302.
42. Liu, R.; Barrick, J. E.; Szostak, J. W.; Roberts, R. W., [19] Optimized synthesis of RNA-protein fusions for in vitro protein selection. In *Methods Enzymol.*, Academic Press 2000; Vol. 318, pp 268-293.
43. Kurz, M.; Gu, K.; Lohse, P. A., Psoralen photo-crosslinked mRNA–puromycin conjugates: a novel template for the rapid and facile preparation of mRNA–protein fusions. *Nucleic Acids Res.* **2000**, *28*, e83-e83.

44. Ishizawa, T.; Kawakami, T.; Reid, P. C.; Murakami, H., TRAP Display: A High-Speed Selection Method for the Generation of Functional Polypeptides. *J. Am. Chem. Soc.* **2013**, *135*, 5433-5440.
45. Millward, S. W.; Takahashi, T. T.; Roberts, R. W., A General Route for Post-Translational Cyclization of mRNA Display Libraries. *J. Am. Chem. Soc.* **2005**, *127*, 14142-14143.
46. Josephson, K.; Hartman, M. C. T.; Szostak, J. W., Ribosomal Synthesis of Unnatural Peptides. *J. Am. Chem. Soc.* **2005**, *127*, 11727-11735.
47. Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T., Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **2001**, *19*, 751-755.
48. Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H., A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods* **2006**, *3*, 357-359.
49. Roberts, R. W., Totally in vitro protein selection using mRNA-protein fusions and ribosome display. *Curr. Opin. Chem. Biol.* **1999**, *3*, 268-273.
50. Passioura, T.; Suga, H., A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun.* **2017**, *53*, 1931-1940.
51. Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H., Natural Product-Like Macrocyclic N-Methyl-Peptide Inhibitors against a Ubiquitin Ligase Uncovered from a Ribosome-Expressed De Novo Library. *Chem. Biol.* **2011**, *18*, 1562-1570.
52. Hayashi, Y.; Morimoto, J.; Suga, H., In Vitro Selection of Anti-Akt2 Thioether-Macrocyclic Peptides Leading to Isoform-Selective Inhibitors. *ACS Chem. Biol.* **2012**, *7*, 607-613.
53. Liu, R.; Li, X.; Xiao, W.; Lam, K. S., Tumor-targeting peptides from combinatorial libraries. *Adv. Drug Del. Rev.* **2017**, *110-111*, 13-37.
54. Morimoto, J.; Hayashi, Y.; Suga, H., Discovery of Macrocyclic Peptides Armed with a Mechanism-Based Warhead: Isoform-Selective Inhibition of Human Deacetylase SIRT2. *Angew. Chem. Int. Ed.* **2012**, *51*, 3423-3427.
55. Chin, J. W., Expanding and reprogramming the genetic code. *Nature* **2017**, *550*, 53-60.
56. Hammerling, M. J.; Krüger, A.; Jewett, M. C., Strategies for in vitro engineering of the translation machinery. *Nucleic Acids Res.* **2019**, *48*, 1068-1083.
57. Horisawa, K., Specific and quantitative labeling of biomolecules using click chemistry. *Front. Physiol.* **2014**, *5*, 457.
58. Passioura, T.; Suga, H., Reprogramming the genetic code in vitro. *Trends Biochem. Sci* **2014**, *39*, 400-408.

59. Goto, Y.; Katoh, T.; Suga, H., Flexizymes for genetic code reprogramming. *Nat. Protoc.* **2011**, *6*, 779-790.
60. Saito, H.; Kourouklis, D.; Suga, H., An in vitro evolved precursor tRNA with aminoacylation activity. *The EMBO Journal* **2001**, *20*, 1797-1806.
61. Murakami, H.; Saito, H.; Suga, H., A Versatile tRNA Aminoacylation Catalyst Based on RNA. *Chem. Biol.* **2003**, *10*, 655-662.
62. Forster, A. C.; Tan, Z.; Nalam, M. N. L.; Lin, H.; Qu, H.; Cornish, V. W.; Blacklow, S. C., Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6353-6357.
63. Katoh, T.; Passioura, T.; Suga, H., Advances in in vitro genetic code reprogramming in 2014–2017. *Synth.Biol.* **2018**, *3*, ysy008.
64. Srinivasan, G.; James, C. M.; Krzycki, J. A., Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science* **2002**, *296*, 1459-1462.
65. Schmidt, M. J.; Summerer, D., Directed Evolution of Orthogonal Pyrrolysyl-tRNA Synthetases in Escherichia coli for the Genetic Encoding of Noncanonical Amino Acids. In *Noncanonical Amino Acids: Methods and Protocols*, Lemke, E. A., Ed. Springer New York: New York, NY, 2018; pp 97-111.
66. Wan, W.; Tharp, J. M.; Liu, W. R., Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **2014**, *1844*, 1059-1070.
67. Tharp, J. M.; Ehnborn, A.; Liu, W. R., tRNAPyl: Structure, function, and applications. *RNA Biol.* **2018**, *15*, 441-452.
68. Guo, J.; Melançon Iii, C. E.; Lee, H. S.; Groff, D.; Schultz, P. G., Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angew. Chem. Int. Ed.* **2009**, *48*, 9148-9151.
69. Fan, C.; Xiong, H.; Reynolds, N. M.; Söll, D., Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.* **2015**, *43*, e156-e156.
70. Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M., Incorporation of Nonnatural Amino Acids into Streptavidin through In Vitro Frame-Shift Suppression. *J. Am. Chem. Soc.* **1996**, *118*, 9778-9779.
71. Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W., Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464*, 441-444.
72. Porte, K.; Riberaud, M.; Châtre, R.; Audisio, D.; Papot, S.; Taran, F., Bioorthogonal Reactions in Animals. *ChemBioChem* **2021**, *22*, 100-113.

73. Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G., Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem-Asian J.* **2009**, *4*, 630-640.
74. Zhang, Y.; Zhou, X.; Xie, Y.; Greenberg, M. M.; Xi, Z.; Zhou, C., Thiol Specific and Tracelessly Removable Bioconjugation via Michael Addition to 5-Methylene Pyrrolones. *J. Am. Chem. Soc.* **2017**, *139*, 6146-6151.
75. Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R., Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132*, 1960-1965.
76. Goddard, D. R.; Michaelis, L., DERIVATIVES OF KERATIN. *J. Biol. Chem.* **1935**, *112*, 361-371.
77. Chen, S.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C., Peptide Ligands Stabilized by Small Molecules. *Angew. Chem. Int. Ed.* **2014**, *53*, 1602-1606.
78. Jongkees, S. A. K.; Umemoto, S.; Suga, H., Linker-free incorporation of carbohydrates into in vitro displayed macrocyclic peptides. *Chem. Sci.* **2017**, *8*, 1474-1481.
79. van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G., Site-selective glycosylation of proteins: creating synthetic glycoproteins. *Nat. Protoc.* **2007**, *2*, 3185-3194.
80. Kundu, R.; Ball, Z. T., Rhodium-catalyzed cysteine modification with diazo reagents. *Chem. Commun.* **2013**, *49*, 4166-4168.
81. Sevier, C. S.; Kaiser, C. A., Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 836-847.
82. Baslé, E.; Joubert, N.; Pucheault, M., Protein Chemical Modification on Endogenous Amino Acids. *Chem. Biol.* **2010**, *17*, 213-227.
83. Koniev, O.; Wagner, A., Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc. Rev.* **2015**, *44*, 5495-5551.
84. Gaudriault, G.; Vincent, J.-P., Selective labeling of α - or ϵ -amino groups in peptides by the Bolton-Hunter reagent. *Peptides* **1992**, *13*, 1187-1192.
85. Nakamura, T.; Kawai, Y.; Kitamoto, N.; Osawa, T.; Kato, Y., Covalent Modification of Lysine Residues by Allyl Isothiocyanate in Physiological Conditions: Plausible Transformation of Isothiocyanate from Thiol to Amine. *Chem. Res. Toxicol.* **2009**, *22*, 536-542.
86. Todrick, A.; Walker, E., A note on the combination of cysteine with allyl isothiocyanate. *Biochem. J* **1937**, *31*, 297-298.
87. Boutureira, O.; Bernardes, G. J. L., Advances in Chemical Protein Modification.

Chem. Rev. **2015**, *115*, 2174-2195.

88. Robinson, M. A.; Charlton, S. T.; Garnier, P.; Wang, X.-t.; Davis, S. S.; Perkins, A. C.; Frier, M.; Duncan, R.; Savage, T. J.; Wyatt, D. A.; Watson, S. A.; Davis, B. G., LEAPT: Lectin-directed enzyme-activated prodrug therapy. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14527-14532.
89. Bavaro, T.; Filice, M.; Temporini, C.; Tengattini, S.; Serra, I.; Morelli, C. F.; Massolini, G.; Terreni, M., Chemoenzymatic synthesis of neoglycoproteins driven by the assessment of protein surface reactivity. *RSC Advances* **2014**, *4*, 56455-56465.
90. Diethelm, S.; Schafröth, M. A.; Carreira, E. M., Amine-Selective Bioconjugation Using Arene Diazonium Salts. *Org. Lett.* **2014**, *16*, 3908-3911.
91. Schlick, T. L.; Ding, Z.; Kovacs, E. W.; Francis, M. B., Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **2005**, *127*, 3718-3723.
92. Gavriluk, J.; Ban, H.; Nagano, M.; Hakamata, W.; Barbas, C. F., Formylbenzene Diazonium Hexafluorophosphate Reagent for Tyrosine-Selective Modification of Proteins and the Introduction of a Bioorthogonal Aldehyde. *Bioconjugate Chem.* **2012**, *23*, 2321-2328.
93. Chen, S.; Tsao, M.-L., Genetic Incorporation of a 2-Naphthol Group into Proteins for Site-Specific Azo Coupling. *Bioconjugate Chem.* **2013**, *24*, 1645-1649.
94. Ban, H.; Gavriluk, J.; Barbas, C. F., Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-Like Reaction for Tyrosine. *J. Am. Chem. Soc.* **2010**, *132*, 1523-1525.
95. Gilis, D.; Massar, S.; Cerf, N. J.; Rooman, M., Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol.* **2001**, *2*, research0049.1.
96. Antos, J. M.; Francis, M. B., Selective Tryptophan Modification with Rhodium Carbenoids in Aqueous Solution. *J. Am. Chem. Soc.* **2004**, *126*, 10256-10257.
97. Antos, J. M.; McFarland, J. M.; Iavarone, A. T.; Francis, M. B., Chemoselective Tryptophan Labeling with Rhodium Carbenoids at Mild pH. *J. Am. Chem. Soc.* **2009**, *131*, 6301-6308.
98. Staudinger, H.; Meyer, J., Über neue organische Phosphorverbindungen III. Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta* **1919**, *2*, 635-646.
99. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 19-24.
100. Spears, R. J.; Fascione, M. A., Site-selective incorporation and ligation of protein aldehydes. *Org. Biomol. Chem.* **2016**, *14*, 7622-7638.
101. Meldal, M.; Diness, F., Recent Fascinating Aspects of the CuAAC Click Reaction.

102. Hein, J. E.; Fokin, V. V., Copper-catalyzed azide–alkyne cycloaddition (CuAAC) and beyond: new reactivity of copper(i) acetylides. *Chem. Soc. Rev.* **2010**, *39*, 1302-1315.
103. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047.
104. Dommerholt, J.; Rutjes, F. P. J. T.; van Delft, F. L., Strain-Promoted 1,3-Dipolar Cycloaddition of Cycloalkynes and Organic Azides. *Top. Curr. Chem.* **2016**, *374*, 16.
105. Blackman, M. L.; Royzen, M.; Fox, J. M., Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* **2008**, *130*, 13518-13519.
106. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W., Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat. Chem.* **2012**, *4*, 298-304.
107. Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A., Functionalized Cyclopropenes As Bioorthogonal Chemical Reporters. *J. Am. Chem. Soc.* **2012**, *134*, 18638-18643.

Chapter 2

Suppression of formylation provides an alternative approach to vacant codon creation in bacterial *in vitro* translation

This chapter has been published as:

Liu, M.; Thijssen, V., Jongkees, S.A.K. (2020) Angew Chem Int Ed, 59: 21870

Abstract

A particularly powerful approach to controlled protein modification is through the use of genetic code reprogramming. The incorporation of unique functional groups into proteins has a great number of application in biological engineering and mechanistic studies. One remaining challenge, however, is the convenient generation of vacant codons. In this work we target the initiation machinery of *Escherichia coli*, showing that simple restriction of formyl donor or inhibition of the formyl transferase is sufficient to prevent formylation of the acylated initiating tRNA and thereby create a vacant initiation codon that can be reprogrammed by exogenously charged tRNA. Our approach can conveniently generate peptides and proteins tagged *N*-terminally with non-canonical functional groups at up to >99% reprogramming efficiency, in combination with decoding the AUG elongation codons either by native methionine or with further reprogramming by azide- and alkyne-containing near-cognates. We further show that we can carry out both macrocyclization and intermolecular modifications with these click handles and can generate proteins with multiple modifications, emphasizing the applicability of our method to current challenges in peptide and protein chemistry.

2.1 Introduction

Generating proteins with uniform post-translational modifications is essential to the study of the functions of these modifications. Solid phase synthesis significantly increases the efficiency and yield of peptide synthesis,¹ allowing synthesis of peptides with diverse functional groups, including post-translational modifications and fluorescent probes.^{1, 2} Native chemical ligation (NCL) is an important extension of peptide synthesis to proteins synthesis.³ With this reaction, the length of chemically accessible peptides and proteins are considerably extended.⁴ Together these are a powerful set of synthesis tools for modified proteins, but the size of chemically synthesized proteins remains limited. A more convenient method to produce modified proteins is offered by genetic code reprogramming and doing this using *in vitro* translation offers unparalleled flexibility, but at the cost of yield. Thus, for exploratory reactions or highly diverse protein synthesis this is an appropriate tool, while for large-scale production of a single product other methods are likely more efficient.

Reprogramming of the genetic code requires a tRNA charged with an unnatural amino acid and a vacant codon (see section 1.3 for more details). Multiple approaches have been developed for the first requirement and these are generally robust. For tRNA aminoacylation in cells or *in vivo* the most common approach is by engineered tRNA and amino-acyl tRNA synthetase pairs⁵, while *in vitro* more chemoenzymatic approaches are possible such as by ligation of pdCpA-amino acid with truncated tRNA⁶ or catalytic RNA (Flexizyme)⁷. Mis-recognition of near cognates by the endogenous amino-acyl tRNA synthetases also provides a convenient approach, but with severely limited scope.⁸ The first incorporation of unnatural amino acids was published in 1950s by using selenomethionine in place of methionine.⁹ To date, there are several methionine analogs including homopropargylglycine (Hpg) and azidohomoalanine (Aha) that are regularly used in protein biosynthesis with high efficiency by using the MetRS synthetase.^{10, 11} The incorporation of these unnatural amino acids into peptides or proteins enables unique bioorthogonal chemistry for conjugation. In the past 20 years, due to the development of Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC), a diverse set of azide- and alkyne-functionalized unnatural amino acid have been developed to attach probes (e.g. affinity tags¹², fluorescent dyes¹³ and other small molecules⁸), while the azide can be used in the phosphine-mediated Staudinger reaction as well.¹⁴

Despite these successes, the range of amino acids that can be used by wild-type Met-tRNA synthetase is limited even when expanded by mutant Met-tRNA synthetases.

With flexizyme-mediated acylation a broad range of moieties have proven amenable to ribosomal translation, with the *N*-terminus proving particularly permissive in translation and allowing boron clusters¹⁵, exotic peptide fragments¹⁶, and even foldamers¹⁷. In addition, the *N*-terminus of proteins is a common target for chemical modification.¹⁸ Approaches for generating a vacant initiating codon are thus particularly valuable.

One crucial challenge that remains only partially solved is an efficient way to generate a vacant codon to reprogram. The most common current approach is reprogramming of a stop codon¹⁹, and this is also compatible with initiation²⁰. However, this is in competition with release factors, which in many cases results in low efficiency and truncated products. Sense codon reprogramming during *in vitro* translation can efficiently provide a vacant codon, but requires the use of a reconstituted recombinant translation system to allow for omission of canonical amino acids and prevents generation of sequences with all 20 canonical amino acids^{7, 21}.

As an alternative approach towards this same goal, in the current work we explore the possibility of preventing the formylation process in bacterial initiation to create vacant start codons without removal of any canonical amino acids. (**Fig. 1**)

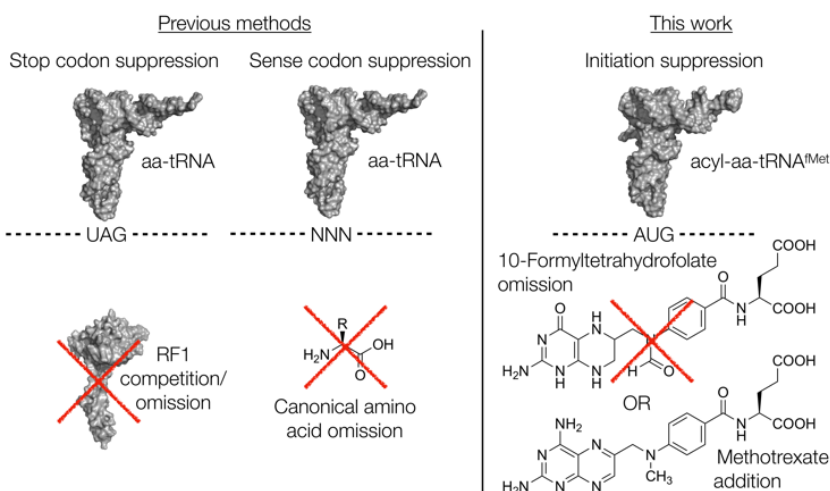


Figure 1: Previous and current approaches to vacant codon creation. aa = amino acid, RF1 = release factor 1

2.2 Results and discussion

2.2.1 Omission of formyl donor

During efforts to generate multiply reprogrammed peptides using an *in vitro* translation system²² we observed that methionine near-cognates such as homopropargylglycine (Hpg) and azidohomoalanine (Aha) competed with flexizyme-charged initiator aminoacyl-tRNA (Ac-Phe-tRNA^{fMet}), giving a mixture of reprogrammed and formyl-Hpg-initiated products (**Fig.2A**). While not unexpected, as initiation with these near-cognates has been reported before²³, it lead us to investigate ways of improving the ratio of desired (reprogrammed) product **P1** to formyl-Hpg product **P1a**. Given that omitting the methionine cognate is not an option in this case as it is required for elongation AUG codon translation, we turned our attention to the bacterial formylation apparatus. In bacterial initiation, the enzyme methionyl-tRNA formyltransferase (MTF) is responsible for transferring a formyl group from 10-formyltetrahydrofolate (10-CHO-THF) to methionyl-tRNA^{fMet} generated by methionyl-tRNA synthetase (MetRS). This formylated methionyl-tRNA (fMet-tRNA^{fMet}) is then recognized by initiation factors (IF1, IF2-GTP, and IF3) and the ribosome to form the pre-30S initiation complex.^{24, 25} The formyl moiety is an important recognition factor for this process, meaning that if we could prevent generation of fMet-tRNA^{fMet} or its analogues it should bias initiation in favor of an exogenously charged aminoacyl-tRNA^{fMet}. The simplest approach to prevent MTF activity is to omit its substrate 10-CHO-THF. Combining a previously reported custom energy solution (SolAFD-) that omits 10-CHO-THF with the commercial translation system PURExpress (NEB),⁷ and adding 25 μ M of initiating tRNA charged by flexizyme with *N*-acetyl phenylalanine, changed the nature of the side product (formylated/non-formylated) and also changed the ratio of desired product to side product from 69:31 **P1/P1a** in the initial test to 92:8 **P1/P1b**. Further increasing the concentration of Ac-Phe-tRNA^{fMet} to 50 μ M was found to give only the desired product **P1** (**Fig.2B**). Simple removal of 10-CHO-THF from the translation is thus sufficient for reprogramming of the initiation codon even in the presence of a substrate for MetRS, provided that sufficient amounts of exogenous aminoacyl-tRNA^{fMet} are added.

2.2.2 Suppression of formylation with MTF inhibitor

As an alternative to 10-CHO-THF omission, we investigated whether MTF activity could be suppressed by a competitive inhibitor to achieve the same result, removing the need for a custom energy solution. One candidate was found in methotrexate, a chemotherapy drug that targets dihydrofolate reductase and a close analogue of

10-CHO-THF, which has been shown to be a modest inhibitor of algal MTF. Adding 25 μM Ac-Phe-tRNA^{fMet} and 100 μM methotrexate to a translation reaction using commercial SolA, which contains an undisclosed amount of 10-CHO-THF, gave an improved fraction of reprogrammed initiation product (83%), with the main side product still being initiation with formylated homopropargylglycine (17%). Increasing the methotrexate and Ac-Phe-tRNA^{fMet} concentrations gave a further improvement (93%; **Fig. 2C**). This indicates that initiation reprogramming by MTF suppression is compatible with the presence of all natural translation components, and may be applicable to efficient initiation reprogramming in cells with a recently reported engineered orthogonal initiation tRNA and aminoacyl-tRNA synthetase pair.²⁶

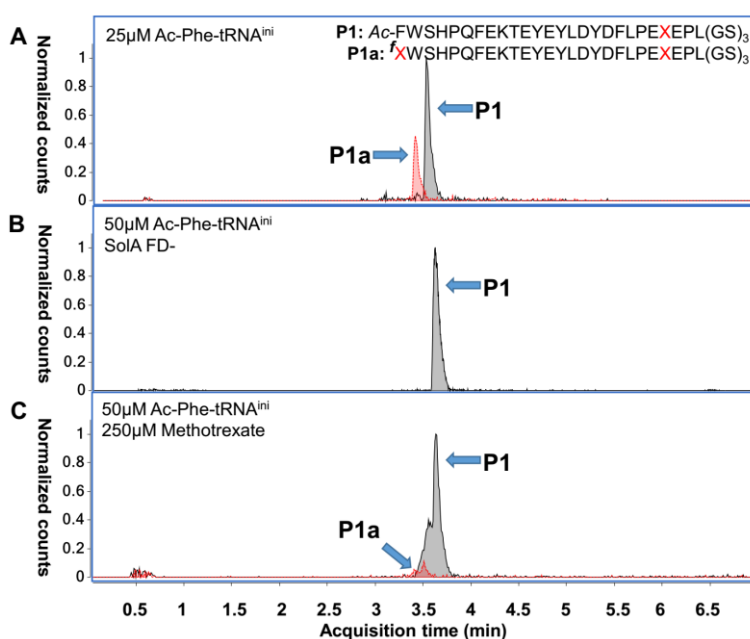


Figure 2. UPLC-MS extracted ion chromatograms (EIC) from analyses of translation tests with reprogrammed initiation for the sequence shown. A) Competition between 25 μM Ac-Phe-tRNA^{fMet} and endogenous initiators (69:31 **P1/P1a**). B) As for A but using SolAFD- and with 50 μM Ac-Phe-tRNA^{fMet} (100% **P1**, byproduct undetectable). C) As for A, but with 250 μM methotrexate and 50 μM Ac-Phe-tRNA^{fMet} (93:7 **P1/P1a**). X = homopropargylglycine, ^f = N-formylation.

2.2.3 Application to peptides and proteins

With these encouraging results, we next investigated the scope of the 10-CHO-THF omission approach for initiation with amino acids bearing different functional groups

as *N*-acyl modifications (**Tables 1** and **S1**).

Table 1. Product ratios for reprogrammed translation with various initiating amino acid and methionine analogue combinations. Aha: azidohomoalanine, Hpg: homopropargylglycine, AMB: 4-aminomethylbenzoic acid, FAM: carboxyfluorescein, Pyn: 4-pentynoic acid.

Entry	Initiator	AUG Elongator	Product ratio ^[a]
1	Ac-Phe	Aha	100:nd ^[b]
2	ClAc-Tyr	Hpg	100:nd
3	ClAc-Tyr	Aha	100:nd
4	AMB-Phe	Hpg	77:4:14:5
5	AMB-Phe	Aha	65:5:14:16
6	(5/6)FAM-Phe	Hpg	78:4:9:9
7	(5/6)FAM-Phe	Aha	72:7:11:10
8	Biotin-Phe	Hpg	100:nd
9	Biotin-Phe	Aha	100:nd
10	Ac-Phe	Met	100:nd
11	Pyn-Phe	Aha	100:nd

[a] Product ratios determined by relative peak areas in UPLC-MS EIC traces (reported as ratio of desired product : *N*-formyl methionine (analogue) side product : α -amino methionine (analogue) side product : initiator-truncated side product). [b] nd; none detected.

In addition to simple capping by an acetyl group, we investigated a chloroacetyl group for thioether head-to-sidechain cyclisation with cysteine, an aminomethylbenzyl group for fluorogenic oxidative cyclization/derivatization reactions with hydroxytryptophan, a fluorescein derivative, and biotin (Table.1) For decoding of the elongation AUG codon we included two 'clickable' near cognates of methionine, homopropargylglycine and azidohomoalanine. An additional trace side product that could be detected in some cases was a peptide starting from the second amino acid ('initiator-truncated'), which is observed for initiation reprogramming by methionine omission in less efficient cases. A control reaction using SolAFD- but with no flexizyme-mediated reprogramming (still containing the Met analogue Hpg) afforded peptide product with the initiating amino acid present in the formylated form (**P2a**; 25%), to our surprise,

as well as the free amine (**P2b**; 63%) and the initiator-truncated peptide (**P2c**; 12%; **Fig. 3**). To investigate whether this initiator-truncated product can be exploited as a further alternative to initiation with formyl-methionine, we tested whether addition of excess un-acylated *in vitro*-transcribed tRNA^{fMet} to the translation mix would improve the yield and purity of initiator-truncated product. An additional 50 μ M tRNA^{fMet} indeed gave clean production of an abundant amount of this initiator-truncated product. This is consistent with reported upregulation of expression of tRNA^{fMet} as an adaptation to deletion or inhibition of MTF in *E. coli*, allowing initiation of protein synthesis to resume.²⁷

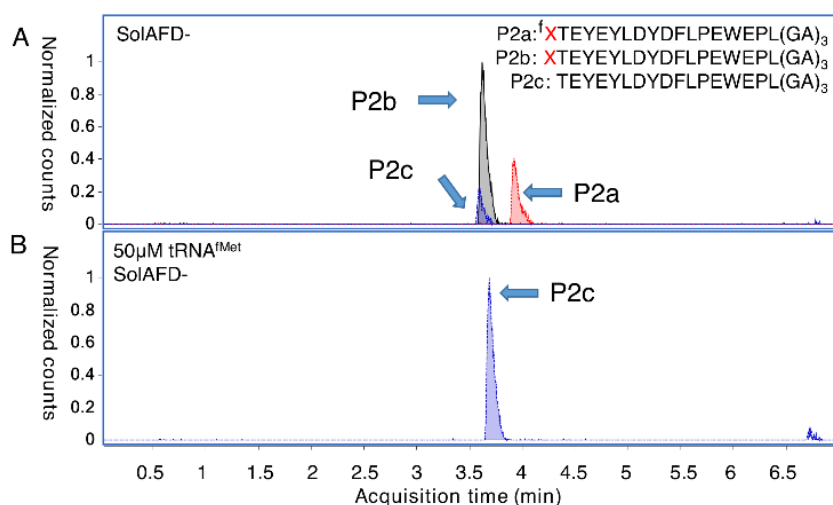


Figure 3. EIC from UPLC-MS analyses of translation without reprogramming initiation. A) Control translation in the absence of both formyl donor and reprogrammed initiator tRNA (25:63:12 **P2a/P2b/P2c**). B) As for A but with additional uncharged tRNA^{fMet}. X = homopropargylglycine, ^f = N-formylation

Given that we were able to show translation of clickable near-cognates of methionine along with reprogramming of the initiating position, we investigated the application of copper-catalyzed azide-alkyne cycloaddition reactions to macrocyclization of the products.²⁸ A peptide **P3** initiated with *N*-(4-pentynoyl) phenylalanine ('Pyn-Phe') and containing a downstream azidohomoalanine was treated with Cu^I directly in the translation mixture. Clean conversion to the macrocyclized product **P4** was shown by retention shift of the isobaric product (**Fig. 4**) and further confirmed by treatment with tris(2-carboxyethyl)phosphine (TCEP), which gave no detectable azide reduction. This click macrocyclization reaction is particularly convenient as it was carried out without any intermediate purification or removal of oxygen, and is anticipated to be

compatible with mRNA display as well as further modifications such as glycosylation at cysteine residues.²⁹

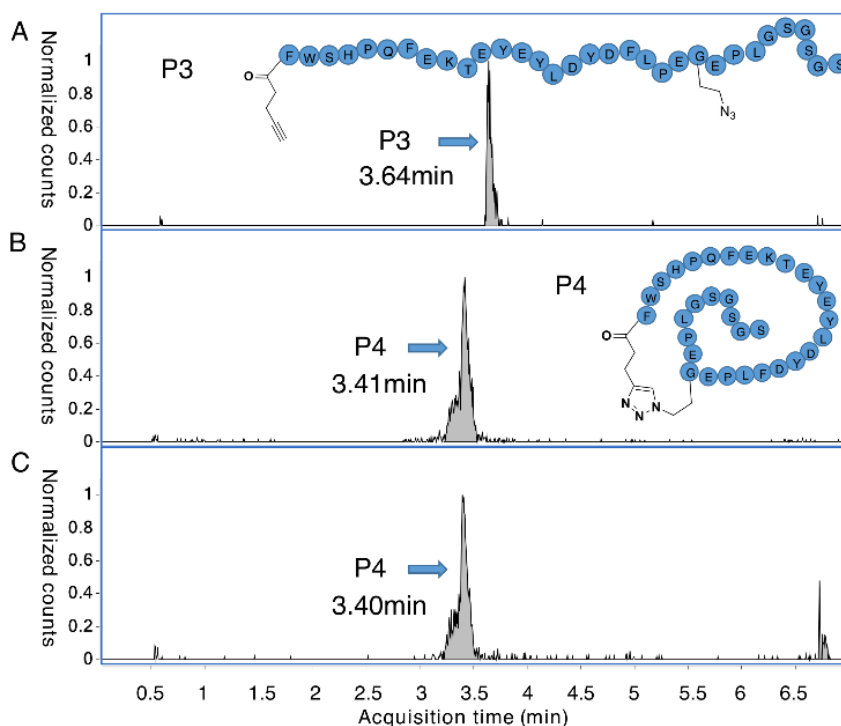


Figure 4. EIC from UPLC-MS analyses of peptide macrocyclization following reprogrammed translation. A) Peptide translation with an alkynyl initiator and downstream azidohomoalanine (100% **P3**, no detectable side products). B) The same peptide after treatment with Cu^I. C) Macrocyclized product from B treated with 35 mM TCEP for one hour (no reduction detectable).

Next we tested translation of a small panel of proteins with (+)-biotinyl phenylalanine as the initiating amino acid and azidohomoalanine reprogramming the elongation AUG codons, with formylation suppressed by methotrexate inhibition of MTF. Chosen as model proteins were a nanobody, a protein used in folding studies (*Escherichia coli* dihydrofolate reductase), and a carrier protein used in conjugate vaccine development (tetanospasmin toxoid). Protein products were reacted with DBCO-Sulfo-Cy3 in a copper-free click reaction under mild denaturing conditions, then pulled down by magnetic streptavidin beads and analyzed by SDS-PAGE gel detected by both fluorescence and silver staining (**Fig. 5**). Pull-down by biotin confirms the initiation reprogramming while attachment of fluorophore confirms

orthogonal reprogramming of the elongator AUG codons, together demonstrating facile generation of doubly reprogrammed proteins amenable to further covalent modification. Quantification of yield was estimated by densitometry of the band from silver staining of a non-clicked product referenced to a standard curve of bovine serum albumin to give 40 ng/μL (**Fig S10**), consistent with the manufacturer's estimated yield for non-reprogrammed translation. We envisage our method to be suited for applications such as generating immobilized proteins for protein interaction experiments, conjugate vaccines carrying multiple epitopes, and generating proteins with multiple fluorophores for conformational studies.

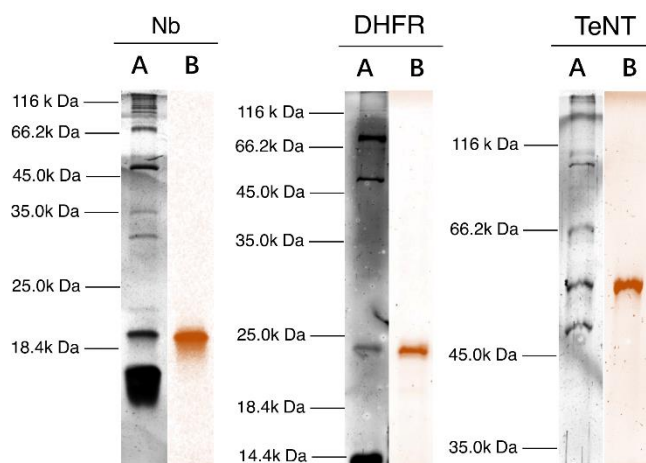


Figure 5. SDS-PAGE showing translation of proteins under a multiply reprogrammed genetic code (Biotin-Phe initiation and Aha decoding of AUG elongator codons). A: silver staining, B: Cy3 fluorescence (color applied in processing), Nb: nanobody, DHFR: dihydrofolate reductase, TeNT: tetanospasmin toxoid.

2.3 Conclusion

We have shown here that suppression of formylation is sufficient to liberate the initiation codon in *E. coli*-derived *in vitro* translation, achievable by removing the formyl donor or by inhibiting the enzyme MTF. We have demonstrated reprogramming of the genetic code on this vacated codon using amino acids acylated with non-canonical functional groups such as reactive handles, a fluorophore, and biotin, and we have shown this to be applicable to protein translation as well as peptides. This initiation reprogramming is compatible with translation of downstream AUG codons by methionine, or exploiting mis-acylation of near cognates by MetRS to incorporate clickable handles to provide convenient access to doubly modified proteins. The

method we report here provides a novel and practical approach to creating a vacant initiation codon, and allows rapid access to N-terminal reprogramming with a broad range of chemical functionality using simple molecular biology techniques and easily accessible reagents.

2.4 Experimental

General Materials

Chemicals were purchased from Sigma Aldrich or Alfa Aesar and used without further purification. PURExpress (E6840S) was purchased from New England Biolabs. UPLC-MS experiments were performed on an Agilent 6560 Ion Mobility Q-TOF LC/MS with an Agilent 1290 Infinity LC system. Experiments were performed using an Agilent Eclipse Plus C18 RRHD 1.8 μm , 2.1 X 50 mm column. A seven-minute time program was used with a 10-70% acetonitrile gradient. All the HPLC-grade solvents were purchased from Biosolve BV Netherlands. Gels were imaged on an Amersham Imager 600 or Bio-Rad ChemiDocTM Touch Gel Imager. Ac-Phe-CME³⁰, *Biotin*-Phe-CME³¹, *AMB*-Phe-CME³², and ClAc-Tyr-CME²¹ were prepared as previously reported.

Model template

Peptide templates were prepared by multi-step assembly PCR using synthetic DNA primers (IDT Europe, Belgium) and isolated by precipitation with 70% ethanol in 0.3 M NaCl, with final DNA and resulting peptide sequences detailed below. The linear DNA template for nanobody translation was prepared by two-step tailed PCR using the primers listed below and plasmid DNA as template. DNA products were quantified by A_{260} and stored in solution at -20 °C.

Template 1.

Peptide sequence: MWSHPQFEKTEYEYLDYDFLPEMEPLGSGSGS

DNA sequence:

TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGTGGAGCCA
TCCGCAGTTTGAGAAGACCGAATACGAATACCTGGATTACGATTTTCTGCCGGA
AATGGAACCGCTGGGCAGCGGCAGCGGCAGCTAGGACGGGGGGCGGAAA

Template 2.

Peptide sequence: MCGSGCSGAMSRYEVDWRGRGSAMG

DNA sequence:

TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGTGCGGCAG
CGGCTGCAGCGGCGCGATGAGCCGCTATGAAGTGGATTGGCGCGGCCGCGG
CAGCGCGATGGGCTAGGACGGGGGGCGGAAA

Template 3

Peptide sequence: MTEY EYLDYDFLPEWEPLGAGAGA

DNA sequence:

TAATACGACTCACTATAGGGTAACTTTAAGAAGGAGATATACATATGACCGAATA
CGAATACTTAGATTACGATTTTTTACCTGAAUGGGAACCTTTAGGTGCAGGCGCA
GGCGCATAGGACGGGGGGCGGAAA

Primers for nanobody DNA preparation

Forward primer 1 (T7 promotor insertion)

TAATACGACTCACTATAGGGTAACTTTAAGAAGGAGATATACATATG

Forward primer 2

TTTAAGAAGGAGATATACATATGGAAGTTCAGCTGGTTGAATCTG

Reverse primer 1 (T7 terminator insertion)

TTTCCGCCCCCGTCTTAAGCTTCCGGTTCGTAATC

Aminoacylation testing of non-canonical amino acids

Aminoacylation reactions (5 μ L) were carried out with the following standard conditions: A mixture of 3 μ L containing 125 pmol of μ helix and 125 pmol eFx in 0.25 M HEPES-KOH buffer (pH 7.5) was heated at 95 $^{\circ}$ C for 2 minutes and subsequently allowed to cool to room temperature for 5 minutes. 1 μ L 3 M $MgCl_2$ solution was then added, followed by an incubation at room temperature for another 5 minutes. The mixture was then cooled on ice, followed by the addition of 1 μ L 25 mM amino acid solution in DMSO. The reaction was left to incubate on ice for a varied amount of time (as detailed in the figures below). The reaction was quenched with 20 μ L 0.3 M NaOAc buffer (pH 5.2). This was followed by the addition of 50 μ L Ethanol to precipitate the product. This mixture was then centrifuged at 13000 rpm for 15 minutes to form a pellet of the precipitated product. The supernatant was removed and the pellet was washed with 30 μ L 70% EtOH, after which it was again centrifuged for 5 minutes at 13000 rpm. The supernatant was removed and the pellet was allowed to air dry for 5 minutes, after which the pellet was dissolved in 16 μ L RNA loading buffer. All aminoacylation samples were loaded onto a 20% PAGE gel and run in acidic buffer (sodium acetate pH 5.2) for 150 minutes at 120 V. Gels were subsequently incubated with TBE and then stained with SYBR Green II (Thermo-Fisher, USA). Product ratios were determined by densitometry in ImageJ.

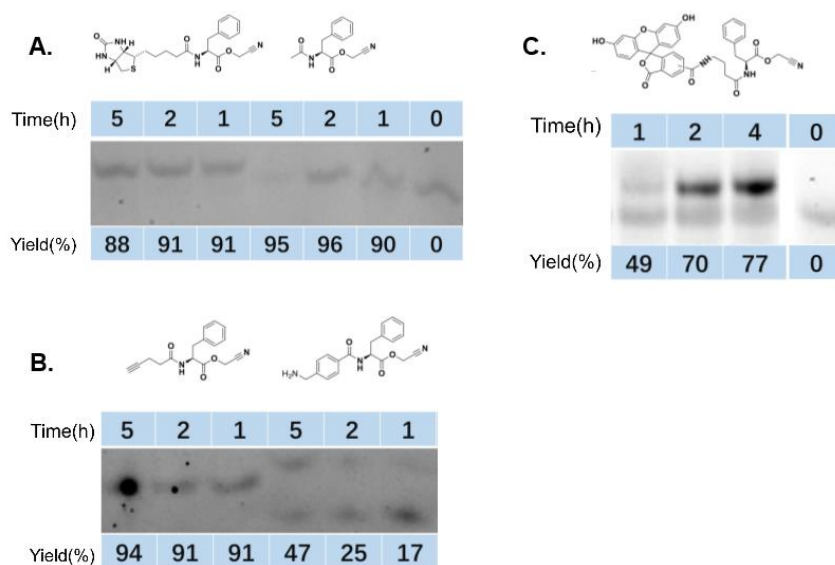


Figure 6. Aminoacylation tests for different amino acids. **(A)** The aminoacylation test for *Biotin*-L-Phe-CME. **(B)** The aminoacylation test for *Pyn*-L-Phe-CME and *AMB*-L-Phe-CME. **(C)** The aminoacylation test for *FAM*-Phe-CME. Note that *FAM*-Phe-CME has inherent fluorescence that likely interferes with this assay, and so quantification of its efficiency is not accurate but product can be confirmed to be present.

Aminoacylation of tRNA^{fMet}

Aminoacylation reactions (5 μ L) were carried out with the following standard conditions: A mixture of 3 μ L of 125 pmol of tRNA^{fMet} and 125 pmol eFx in 0.25 M HEPES-KOH buffer (pH 7.5) was heated at 95 °C for 2 min followed by an incubation at room temperature for 5 minutes. Then, 1 μ L MgCl₂ solution (3 M) was added and the mixture was left to incubate another 5 min at room temperature. The mixture was then cooled on ice, followed by the addition of 1 μ L amino acid solution in DMSO (25 mM). The reaction was left to incubate on ice for 2 hours, then quenched with 20 μ L NaOAc buffer (0.3 M, pH 5.2). This was followed by the addition of 50 μ L ethanol to precipitate the product. This mixture was then centrifuged at 13000 rpm for 15 min to form a pellet of the precipitated product. The supernatant was removed and the pellet was vortexed with 40 μ L NaOAc buffer (0.1 M, pH 5.2) in 70% EtOH then centrifuged for 10 minutes at 13000 rpm, and this process was repeated for two washes. A final wash was then performed (without vortexing) using 30 μ L 70% ethanol in water and followed by a 3 min centrifugation at 13000 rpm. Following removal of the supernatant, the pellet was left to air dry for 5 minutes and stored at -20 °C until use.

***In Vitro* Translation and UPLC-MS Methods**

For translation using PURExpress solution A Δ (aa, tRNA), *in vitro* translation

reactions were prepared on ice with the following composition:

20% (vol%) PURExpress solution A Δ (aa, tRNA)

30% (vol%) PURExpress solution B

1 $\mu\text{g} \cdot \mu\text{L}^{-1}$ *E. coli* tRNA

25 μM or 50 μM acylated tRNAⁱⁿⁱ

10 $\mu\text{g} \cdot \mu\text{L}^{-1}$ DNA template or plasmid

0.5 mM 20 amino acid mix (-Methionine, + Hpg (or Aha))

0, 100 or 250 μM Methotrexate

Volume was adjusted to a final volume of 5 μL with milliQ water and the reaction mixture incubated at 37 °C for 30 minutes (or 1 hour in the case of protein translation).

For peptides: After translation, 3 volumes of acetonitrile were added and the sample was centrifuged at 13000 rpm for 5 minutes. Samples were analyzed using UPLC-MS.

For proteins: After translation, 5 μL glycine buffer (0.5 M, pH=10) was added and the solution was incubated for another hour. After the incubation a Vivaspin 500 centrifugal concentrator (10 kDa) was used to remove excess amino acids, then DBCO-Sulfo-Cy3 (Jena Bioscience) was added to a final concentration of 100 μM and the reaction mixture incubated for 3 hours at 37 °C in denaturing conditions (4 M urea, 150 mM NaCl). The mixture was then incubated with streptavidin beads (Invitrogen Dynabeads M-280 Streptavidin) for 20 minutes, followed by washing two times with PBS-T. Proteins were eluted off the beads by incubating at 95 °C for 5 minutes with SDS loading buffer. Samples were then analyzed by running them on a SDS-PAGE gel.

Translation using custom energy solution (SolAFD-)

In vitro translation reactions were prepared on ice with the following composition:

14% (vol%) SolAFD-

30% (vol%) PURExpress solution B

25 μM or 50 μM acylated tRNA^{fMet}

50 ng DNA template

0.5 mM 20 amino acid mix (-Methionine, + Hpg (or Aha))
0, 100 or 250 μ M Methotrexate

Volume was adjusted to a final volume of 5 μ L with milliQ water and the reaction mixture incubated at 37 °C for 30 min. After translation, 3 volumes of acetonitrile were added and the sample was centrifuged at 13000 rpm for 5 minutes. Samples were analyzed using UPLC-MS.

Click reaction after translation (Based on 5 μ L translation volume)

In a small tube, 0.5 μ L 20 mM CuSO₄ solution and 1 μ L 50 mM THPTA solution were premixed, changing color to light blue immediately. This mixture was then added to the translation mixture. This was followed by addition of 1 μ L 100 mM aminoguanidine hydrochloride solution and 1 μ L 100 mM sodium ascorbate solution. The tube was then sealed and the reaction was allowed to incubate overnight at 37 °C. After the incubation, 3 volumes of acetonitrile were added and the samples were centrifuged at 13000 rpm for 5 minutes. Samples were analyzed using UPLC-MS.

TCEP reduction

After the click reaction, 1 μ L 350 mM TCEP solution was added and incubated at 37 °C for 1 hour. After the incubation 3 volumes of acetonitrile were added and the samples were centrifuged at 13000 rpm for 5 minutes. Samples were analyzed using UPLC-MS.

Mass extraction

Extracted ion chromatograms were generated by adding all product mass peaks identified for a given calculated mass from 2+ through 5+ ionization states (not all observed in all cases, depending on sequence) and including cation adducts where relevant (Na, K, NH₄) with a 20 ppm error range around the observed mass, summed across all isotopes with significant signal above baseline noise. Calculated and observed masses shown are for the most abundant isotope peaks rather than the monoisotopic peaks in all cases, to avoid issues with low signal to noise for minor side products.

Table and Figures for UPLC-MS

Table S1. Product ratios and reaction conditions for reprogrammed translation with various initiating amino acid and methionine analogue combinations. Aha: azidohomoalanine, Hpg: homopropargylglycine, AMB: 4-aminomethylbenzoic acid,

FAM: carboxyfluorescein, Pyn: 4-pentynoic acid.

Entry	Initiator	Template	Conditions	AUG Elongator	Product ratio ^[a]
1	Ac-Phe	1	SolAA(aa, tRNA) 25µM aa-tRNA ^{fMet}	Hpg	69:31:0:0
2	Ac-Phe	1	SolAFD- 25µM aa-tRNA ^{fMet}	Hpg	92:8:0:0
3	Ac-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Hpg	100:nd ^[b]
4	Ac-Phe	2	SolAFD- 50µM aa-tRNA ^{fMet}	Aha	100:nd
5	ClAc-Tyr	1	SolAFD- 50µM aa-tRNA ^{fMet}	Hpg	100:nd
6	ClAc-Tyr	1	SolAFD- 50µM aa-tRNA ^{fMet}	Aha	100:nd
7	AMB-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Hpg	77:4:14:5
8	AMB-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Aha	65:5:14:16
9	(5/6)FAM-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Hpg	78:4:9:9
10	(5/6)FAM-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Aha	72:7:11:10
11	Biotin-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Hpg	100:nd
12	Biotin-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Aha	100:nd
13	-	3	SolAFD-	Hpg	0:25:63:12
14	-	3	SolAFD- 50µM tRNA ^{fMet}	Hpg	0:0:0:100
15	Ac-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Met	100:nd
16	Pyn-Phe	1	SolAFD- 50µM aa-tRNA ^{fMet}	Aha	100:nd
17	Ac-Phe	1	SolAA(aa, tRNA) 100µM Methotrexate 25µM aa-tRNA ^{fMet}	Hpg	83:17:0:0
18	Ac-Phe	1	SolAA(aa, tRNA) 250µM Methotrexate 50µM aa-tRNA ^{fMet}	Hpg	93:7:0:0

[a] Product ratios determined by relative peak areas in UPLC-MS EIC traces (showing ratio of desired product: *N*-formyl methionine (analogue) side product: α -amino methionine (analogue) side product: initiator-truncated side product). [b] nd; none detected.

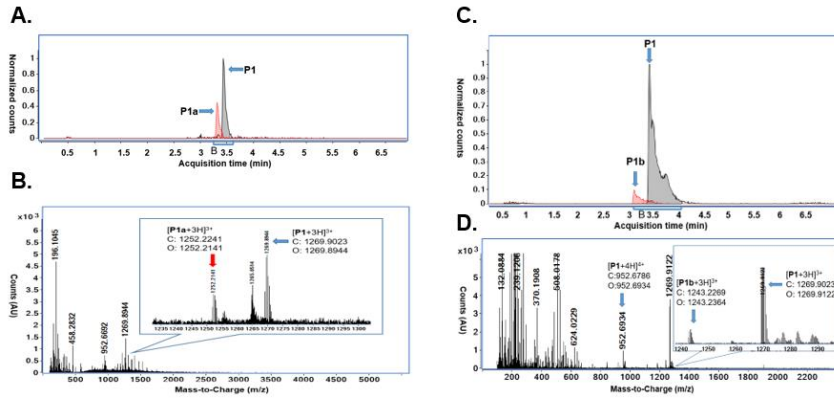


Figure S1. The EIC and mass spectrum for Table S1, entry 2: translation of peptide sequence 'MWSHPQFEKTEYEYLDYDFLPEMEPLGSGSGS' using SolAFD-supplemented with 25 μ M Ac-Phe-tRNA^{fMet} (initiation) and homopropargylglycine (elongation). A. The EIC for **P1** and **P1b**. B. The mass spectrum (2.87-3.82 min, 101 scans) of **P1** and **P1b**. The EIC and mass spectrum for Table S1, entry 3: translation of peptide sequence 'MWSHPQFEKTEYEYLDYDFLPEMEPLGSGSGS' using SolAFD-supplemented with 50 μ M Ac-Phe-tRNA^{fMet} (initiation) and homopropargylglycine (elongation). C. The EIC for **P1**. D. The mass spectrum (3.50-3.81 min, 34 scans) of **P1**.

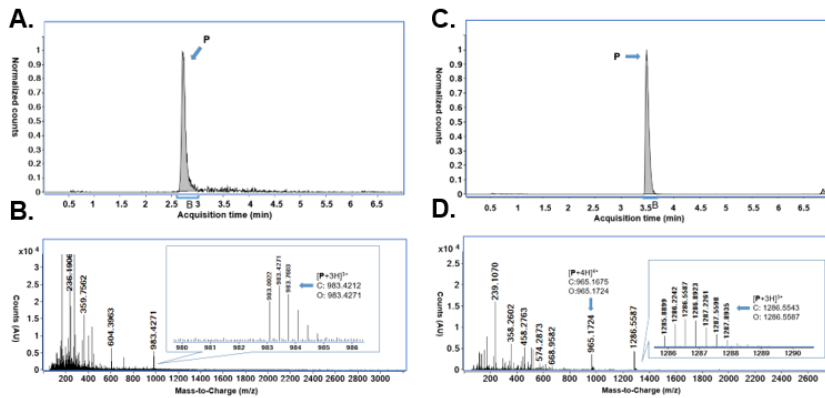


Figure S2. The EIC and mass spectrum for Table S1, entry 4: translation of peptide sequence 'MFCGSGCAMSRYEVDWRGRGSAMGSGSGS' using SolAFD-supplemented with 50 μ M Ac-Phe-tRNA^{fMet} (initiation) and azidohomoalanine (elongation). A. The EIC for Peptide. B. The mass spectrum (2.62-2.92 min, 55 scans) of peptide. The EIC and mass spectrum for Table S1, entry 5: translation of peptide sequence 'MWSHPQFEKTEYEYLDYDFLPEMEPLGSGSGS' using SolAFD-

supplemented with 50 μM ClAc-Tyr-tRNA^{fMet} (initiation) and homopropargylglycine (elongation). C. The EIC for Peptide and possible byproduct. D. The mass spectrum (3.44-3.63 min, 62 scans) of peptide.

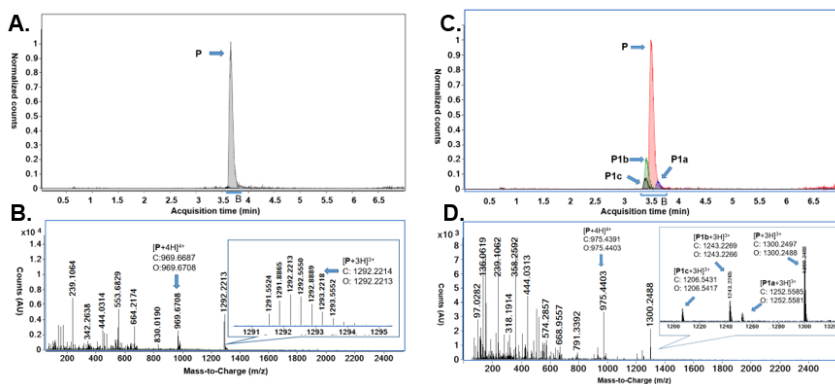


Figure S3. The EIC and mass spectrum for Table S1, entry 6: translation of peptide sequence 'MWSHPQFEKTEYEYLDYDFLPMEPLGSGSGS' using SolAFD-supplemented with 50 μM ClAc-Tyr-tRNA^{fMet} (initiation) and azidohomoalanine (elongation). A. The EIC for Peptide and possible byproduct. B. The mass spectrum (3.59-3.81 min, 70 scans) of peptide. The EIC and mass spectrum for Table S1, entry 7: translation of peptide sequence 'MWSHPQFEKTEYEYLDYDFLPMEPLGSGSGS' using SolAFD- supplemented with 50 μM AMB-Phe-tRNA^{fMet} (initiation) and homopropargylglycine (elongation). C. The EIC for Peptide and byproduct. D. The mass spectrum (3.35-3.76 min, 127 scans) of peptide.

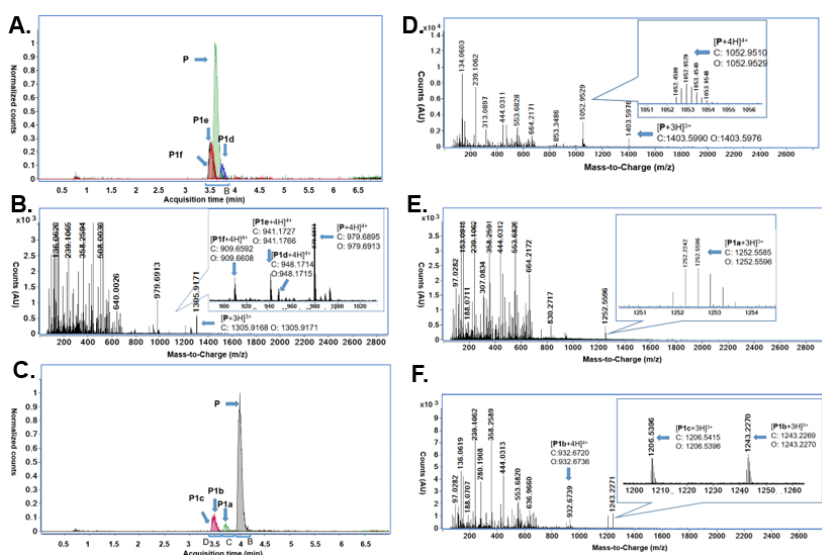


Figure S4. The EIC and mass spectrum for Table S1, entry 8: translation of peptide sequence 'MWSHPQFEKTEYEYLDYDFLPEMEPLGSGSGS' using SolAFD-supplemented with 50 μ M AMB-Phe-tRNA^{fMet} (initiation) and azidohomoalanine (elongation). A. The EIC for Peptide and byproduct. Red, **P1f**; Grey, **P1e**. B. The mass spectrum (3.48-3.88 min, 122 scans) of peptide. The EIC and mass spectrum for Table S1, entry 9: translation of peptide sequence 'MWSHPQFEKTEYEYLDYDFLPEMEPLGSGSGS' using SolAFD- supplemented with 50 μ M (5/6) FAM-Phe-tRNA^{fMet} (initiation) and homopropargylglycine (elongation). C. The EIC for Peptide and byproduct. Red, **P1b**; Blue, **P1c**. D. The mass spectrum (3.93-4.17 min, 77 scans) of peptide. E. The mass spectrum (3.65-3.83 min, 57 scans) for **P1a**. F. The mass spectrum (3.44-3.61 min, 53 scans) for **P1b** and **P1c**.

(elongation). A. The EIC for Peptide and byproduct. B. The mass spectrum (3.75-4.03 min, 87 scans) of peptide. The EIC and mass spectrum for Table S1, entry 13: translation of peptide sequence 'MTEYELDYLDFLEWEPLGAGAGA' using SolAFD- supplemented with homopropargylglycine (initiation and elongation). C. The EIC for Peptide and byproduct. D. The mass spectrum (3.56-4.07 min, 157 scans) of peptide.

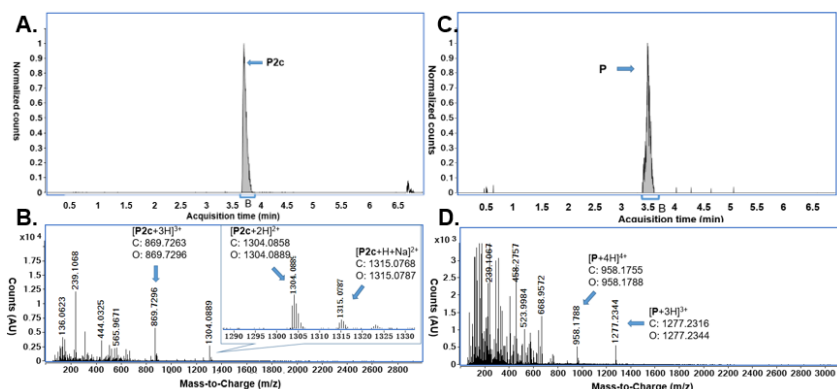
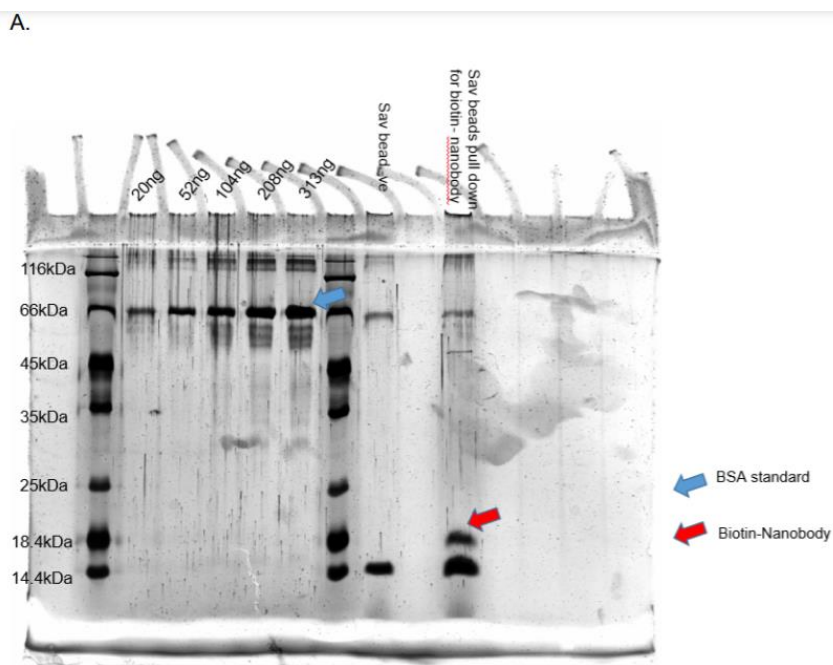


Figure S7. The EIC and mass spectrum for Table S1, entry 14: translation of peptide sequence 'MTEYELDYLDFLEWEPLGAGAGA' using SolAFD- supplemented with 50 μ M tRNA^{fMet} and homopropargylglycine (elongation). A. The EIC for Peptide and byproduct. B. The mass spectrum (3.65-3.84 min, 62 scans) of peptide. The EIC and mass spectrum for Table S1, entry 15: translation of peptide sequence 'MWSHPQFEKTEYELDYLDFLEMEPLGSGSGS' using SolAFD- supplemented with 50 μ M Ac-Phe-tRNA^{fMet} (initiation) and methionine (elongation). C. The EIC for Peptide. D. The mass spectrum (3.43-3.65 min, 70 scans) of peptide.

tRNA) supplemented with 25 μ M Ac-Phe-tRNA^{fMet} (initiation), 100 μ M Methotrexate and homopropargylglycine (elongation). A. The EIC for Peptide **P1** and **P1a**. B. The mass spectrum (3.34-3.68 min, 57 scans) of peptide. The EIC and mass spectrum for Table S1, entry 18: translation of peptide sequence 'MWSHPQFEKTEYEYLDYDFLPEMEPLGSGSGS' using NEB SolA Δ (aa, tRNA) supplemented with 50 μ M Ac-Phe-tRNA^{fMet} (initiation), 250 μ M Methotrexate and homopropargylglycine (elongation). C. The EIC for Peptide **P1** and **P1a**. D. The mass spectrum (3.43-3.85 min, 45 scans) of peptide. II, [P1+2H+Na]³⁺. C:1277.5640. O: 1277.5681. III, [P1+2H+K]³⁺. C:1282.5543. O: 1282.5749. IV, [P1+H+NH₄+K]³⁺. C:1287.8953 O: 1287.8904.



B.

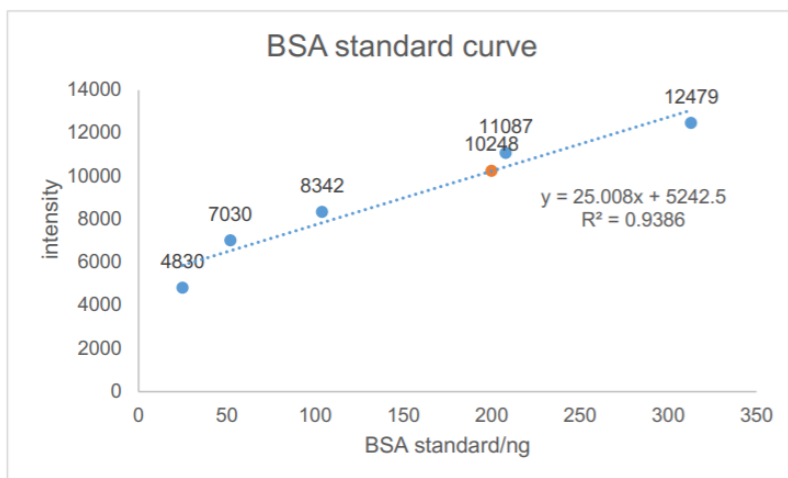
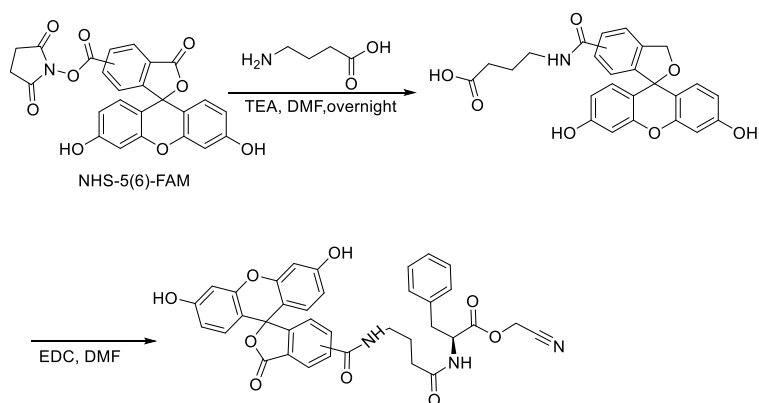


Figure S10. Initiation reprogramming in protein expression (translated following the general method with 250 μ M methotrexate and 50 μ M acyl-tRNA). (A) Protein yield quantification for nanobody reprogrammed translation (no DBCO-Sulfo-Cy3 modification) by BSA dilution series. (B) Resulting BSA standard curve (blue), with biotin-nanobody band intensity (orange) giving an estimated yield of 200 ng nanobody from 5 μ L translation (40 μ g/mL, cf. manufacturer's estimated expected yield of 10-200 μ g/mL)

Synthesis of molecules



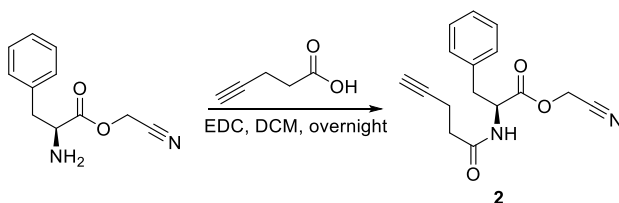
1

N-(*N*-(5/6-Carboxyfluoresceinamide)-4-aminobutyric)-phenylalanine cyanomethyl ester

5/6-FAM (378 mg, 1 mmol) was dissolved in 5 mL DMF, followed by addition of EDC (286 mg, 1.5 mmol) and NHS (126 mg, 1.2 mmol). The reaction mixture was then

stirred at room temperature for 2 hours. This was then followed by the addition of 4-aminobutyric acid (103 mg, 1 mmol) and 3 equiv of TEA (420 μ L, 303 mg, 3 mmol). The reaction mixture was stirred at room temperature overnight, followed by the addition of Phe-CME (204 mg, 1 mmol), *N,N'*-diisopropylcarbodiimide (626 μ L, 4 mmol,) and ethyl isonitrosocyanoacetate (566 mg, 4 mmol). The reaction mixture was then stirred at room temperature overnight again. The solution was then poured into 100 mL brine and extracted with 100 mL ethyl acetate. The organic layer was dried over anhydrous sodium sulfate and concentrated under vacuum. The crude product was purified by silica column chromatography (DCM/MeOH/CH₃COOH = 10:1:0.01, *R_f* = 0.5) to obtain 100 mg (16%) compound **1** as a yellow foam. ¹H NMR (400 MHz, CD₃OD) δ : 8.42 (s, 1H, -NH-Ph-H), 8.21 – 8.04 (m, 2H, -NH-Ph-H), 7.61 (m, 0.5H, -NH-Ph-H), 7.37 – 7.04 (m, 5H, Ph-H), 6.78 – 6.44 (m, 6H, HO-Ph-H), 4.85 (s, 2H, -CH₂-CN), 4.51 – 4.27 (m, 1H, -NH-CH-CO-), 3.69 – 3.37 (m, 2H, -NH-CH₂-), 3.23 – 2.95 (m, 2H, Ph-CH₂-), 2.54 – 2.14 (m, 2H, -NH-CH₂-CH₂-), 2.00 – 1.67 (m, 2H, -CO-CH₂-CH₂-) ppm. ¹³C NMR (101 MHz, CD₃OD) δ 213.11, 180.78, 170.83, 169.44, 154.06, 153.15, 135.41, 134.55, 133.08, 130.21, 129.62, 128.98, 128.74, 128.09, 128.05, 125.73, 124.83, 123.34, 122.39, 118.73, 113.75, 110.94, 109.12, 103.61, 55.05, 50.07, 42.69, 36.95, 31.65, 28.87, 26.48, 26.38, 26.27. HRMS (ESI⁺): [M+H]⁺ Calculated for [C₃₆H₃₁N₃O₉]⁺: required 648.1977 m/z, found: 648.1970 m/z.

Note: this product is a mixture, attaching to positions 5 and 6 of the fluorescein.



N-(4-pentynoic)-Phenylalanine cyanomethyl ester

Phenylalanine cyanomethyl ester (134 mg, 0.65 mmol, 1 equiv) was dissolved in anhydrous DCM, followed by addition of 4-pentynoic acid (95 mg, 0.98 mmol, 1.5 equiv) and EDC (186 mg, 0.98 mmol, 1.5 equiv), then the reaction mixture stirred at room temperature overnight. The solvent was subsequently removed under vacuum and the residue was purified by silica column chromatography (PE/EA = 1:1, *R_f* = 0.3) to obtain 180 mg (65%) compound **2** as a colorless oil. ¹H NMR (400 MHz, CDCl₃) δ : 7.37 – 7.27 (m, 3H, Ph-H), 7.18 – 7.13 (m, 2H, Ph-H), 6.08 (d, *J* = 7.6 Hz, 1H, -NH-CO-), 4.94 (dt, *J* = 7.7, 6.2 Hz, 1H, -NH-CH-), 4.79 (d, *J* = 15.7 Hz, 1H, -CH₂-CN), 4.69 (d, *J* = 15.7 Hz, 1H, -CH₂-CN), 3.23 – 3.09 (m, 2H, Ph-CH₂-), 2.53 – 2.44 (m, 2H, CH \equiv C-CH₂-), 2.44 – 2.38 (m, 2H, -CO-CH₂-CH₂-) 1.96 (t, *J* = 2.6 Hz, 1H, CH \equiv C-) ppm.

^{13}C NMR (101 MHz, CDCl_3) δ 181.91, 170.22, 134.88, 129.16, 128.91, 127.55, 113.67, 82.52, 69.61, 52.96, 48.87, 37.58, 34.93, 25.54, 14.58, 14.16 ppm. HRMS (ESI $^+$): $[\text{M}+\text{H}]^+$ Calculated for $[\text{C}_{16}\text{H}_{17}\text{N}_2\text{O}_3]^+$: required 285.1234 m/z, found: 285.1227 m/z.

2.5 References

1. Palomo, J. M., Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides. *RSC Advances* **2014**, *4*, 32658-32672.
2. Meienhofer, J., 3 - Peptide Synthesis: A Review of the Solid-Phase Method. In *Hormonal Proteins and Peptides*, Li, C. H., Ed. Academic Press 1973; pp 45-267.
3. Nilsson, B. L.; Soellner, M. B.; Raines, R. T., Chemical synthesis of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91-118.
4. Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O., Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**, *119*, 7328-7443.
5. Schmidt, M. J.; Summerer, D., Directed Evolution of Orthogonal Pyrrolysyl-tRNA Synthetases in *Escherichia coli* for the Genetic Encoding of Noncanonical Amino Acids. In *Noncanonical Amino Acids: Methods and Protocols*, Lemke, E. A., Ed. Springer New York: New York, NY, 2018; pp 97-111.
6. Duca, M.; Chen, S.; Hecht, S. M., Aminoacylation of transfer RNAs with one and two amino acids. *Methods* **2008**, *44*, 87-99.
7. Goto, Y.; Katoh, T.; Suga, H., Flexizymes for genetic code reprogramming. *Nat. Protoc.* **2011**, *6*, 779-790.
8. Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M., Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proceedings of the National Academy of Sciences* **2006**, *103*, 9482.
9. Cowie, D. B.; Cohen, G. N., Biosynthesis by *Escherichia coli* of active altered proteins containing selenium instead of sulfur. *Biochim. Biophys. Acta* **1957**, *26*, 252-261.
10. Saleh, A. M.; Wilding, K. M.; Calve, S.; Bundy, B. C.; Kinzer-Ursem, T. L., Non-canonical amino acid labeling in proteomics and biotechnology. *J. Biol. Eng.* **2019**, *13*, 43.
11. Lang, K.; Chin, J. W., Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114*, 4764-4806.
12. Link, A. J.; Tirrell, D. A., Cell Surface Labeling of *Escherichia coli* via Copper(I)-Catalyzed [3+2] Cycloaddition. *J. Am. Chem. Soc.* **2003**, *125*, 11164-11165.
13. Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A., Selective Dye-Labeling of Newly

- Synthesized Proteins in Bacterial Cells. *J. Am. Chem. Soc.* **2005**, *127*, 14150-14151.
14. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proceedings of the National Academy of Sciences* **2002**, *99*, 19.
15. Yin, Y.; Ochi, N.; Craven, T. W.; Baker, D.; Takigawa, N.; Suga, H., De Novo Carborane-Containing Macrocyclic Peptides Targeting Human Epidermal Growth Factor Receptor. *J. Am. Chem. Soc.* **2019**, *141*, 19193-19197.
16. Goto, Y.; Suga, H., Translation Initiation with Initiator tRNA Charged with Exotic Peptides. *J. Am. Chem. Soc.* **2009**, *131*, 5040-5041.
17. Rogers, J. M.; Kwon, S.; Dawson, S. J.; Mandal, P. K.; Suga, H.; Huc, I., Ribosomal synthesis and folding of peptide-helical aromatic foldamer hybrids. *Nat. Chem.* **2018**, *10*, 405-412.
18. Rosen, C. B.; Francis, M. B., Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **2017**, *13*, 697-705.
19. Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G., A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **1989**, *244*, 182.
20. Mamaev, S.; Olejnik, J.; Olejnik, E. K.; Rothschild, K. J., Cell-free N-terminal protein labeling using initiator suppressor tRNA. *Anal. Biochem.* **2004**, *326*, 25-32.
21. Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H., Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. *ACS Chem. Biol.* **2008**, *3*, 120-129.
22. Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T., Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **2001**, *19*, 751-755.
23. Horiya, S.; Bailey, J. K.; Temme, J. S.; Guillen Schlippe, Y. V.; Krauss, I. J., Directed Evolution of Multivalent Glycopeptides Tightly Recognized by HIV Antibody 2G12. *J. Am. Chem. Soc.* **2014**, *136*, 5407-5415.
24. Zorzet, A.; Pavlov, M. Y.; Nilsson, A. I.; Ehrenberg, M.; Andersson, D. I., Error-prone initiation factor 2 mutations reduce the fitness cost of antibiotic resistance. *Mol. Microbiol.* **2010**, *75*, 1299-1313.
25. Antoun, A.; Pavlov, M. Y.; Lovmar, M.; Ehrenberg, M., How initiation factors tune the rate of initiation of protein synthesis in bacteria. *The EMBO Journal* **2006**, *25*, 2539-2550.
26. Tharp, J. M.; Ad, O.; Amikura, K.; Ward, F. R.; Garcia, E. M.; Cate, J. H. D.; Schepartz, A.; Söll, D., Initiation of Protein Synthesis with Non-Canonical Amino Acids In Vivo. *Angew. Chem. Int. Ed.* **2020**, *59*, 3122-3126.

27. Nilsson, A. I.; Zorzet, A.; Kanth, A.; Dahlström, S.; Berg, O. G.; Andersson, D. I., Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *Proceedings of the National Academy of Sciences* **2006**, *103*, 6976.
28. Sako, Y.; Morimoto, J.; Murakami, H.; Suga, H., Ribosomal Synthesis of Bicyclic Peptides via Two Orthogonal Inter-Side-Chain Reactions. *J. Am. Chem. Soc.* **2008**, *130*, 7232-7234.
29. Jongkees, S. A. K.; Umemoto, S.; Suga, H., Linker-free incorporation of carbohydrates into in vitro displayed macrocyclic peptides. *Chemical Science* **2017**, *8*, 1474-1481.
30. Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H., A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods* **2006**, *3*, 357-359.
31. Saito, H.; Kourouklis, D.; Suga, H., An in vitro evolved precursor tRNA with aminoacylation activity. *The EMBO Journal* **2001**, *20*, 1797-1806.
32. Yamagishi, Y.; Ashigai, H.; Goto, Y.; Murakami, H.; Suga, H., Ribosomal Synthesis of Cyclic Peptides with a Fluorogenic Oxidative Coupling Reaction. *ChemBioChem* **2009**, *10*, 1469-1472.

Chapter 3

Discovery of Potent Cyclic Peptide E-Selectin inhibitors through mRNA display

Part of this chapter has been published as:

Johansen-Leete, J.; Passioura, T.; Foster, S.; Bhusal, R. P.; Ford, D.; **Liu, M.**; Jongkees, S. A. K.; Suga, H.; Stone, M. J.; Payne, R. J. (2020) J Am Chem Soc, 142:9141

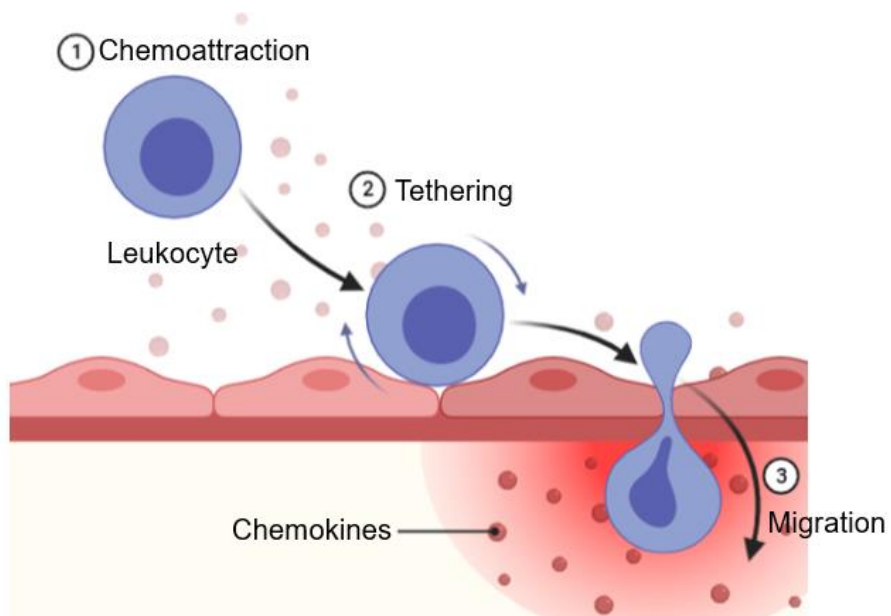
Abstract

Selectins (E-, P- and L-selectins) are a family of calcium dependent-lectins which mediate the essential tethering and rolling step in leukocyte adhesion. In this role they are involved in the inflammatory cascade, and they have also been implicated as an adhesion molecule in cancer metastasis. E-selectin in particular is expressed on endothelial cells of bone marrow and skin and its ligands are commonly found throughout the human body, including PSGL-1, ESL-1 and CD44. The interactions of carbohydrates with lectins is typically not strong, and so rather than a glycomimetic route we sought to find *de novo* peptide ligands for E-selectin. To drive their binding to the sugar-binding site we designed a set of three amino acids with side chains that can exploit typical sugar-protein interactions. These binding factors were synthesized and their incorporation into peptides by the FIT system (Flexible *in vitro* translation) was validated before building them into a random peptide library in mRNA display format. From panning of this library we identified several possible binders for E-selectin, including several hits containing a sulfonate from one of the novel interaction factors. The peptides were synthesized by solid phase peptide synthesis (SPPS) and binding verified by SPR to be of high affinity, with several peptides having ~1-20nM K_D . These binders show potential as lead compounds for application in inhibition of E-selectin in its role as an immune cell recognition protein, as an inhibitor of tumor metastasis, or as a targeting molecule for selective drug delivery.

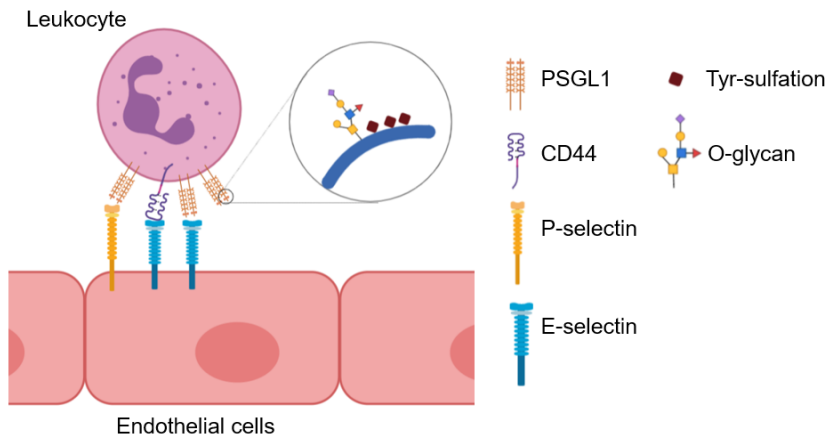
3.1 Introduction

The selectins are transmembrane lectins that mediate leukocyte rolling on the vascular surface. There are 3 types of selectins: P-selectin, E-selectin and L-selectin, named for the cells on which they are predominantly expressed (platelet cells, endothelial cells, and leukocytes).¹ All the selectins have an *N*-terminal Ca^{2+} -dependent lectin domain, an epidermal growth factor (EGF)-like module, a series of repeated complement-binding protein-like domains, a transmembrane domain and a cytoplasmic tail.² All of them are single chain transmembrane glycoproteins,^{3,4} and bind to glycoprotein ligands. They are essential for the immune system by helping leukocytes recruitment and then mediating the inflammatory cascade. In the past decades, selectins have been shown to contribute to multiple diseases, such as sickle cell anemia⁵, rheumatoid arthritis⁶ and psoriasis⁷. P- and E- selectins are upregulated in these patients. Furthermore, it was shown that selectins are involved in cancer migration,³ similarly allowing tumor adhesion to endothelial cells under flow conditions and extravasation into tissue.

A.



B.



C.

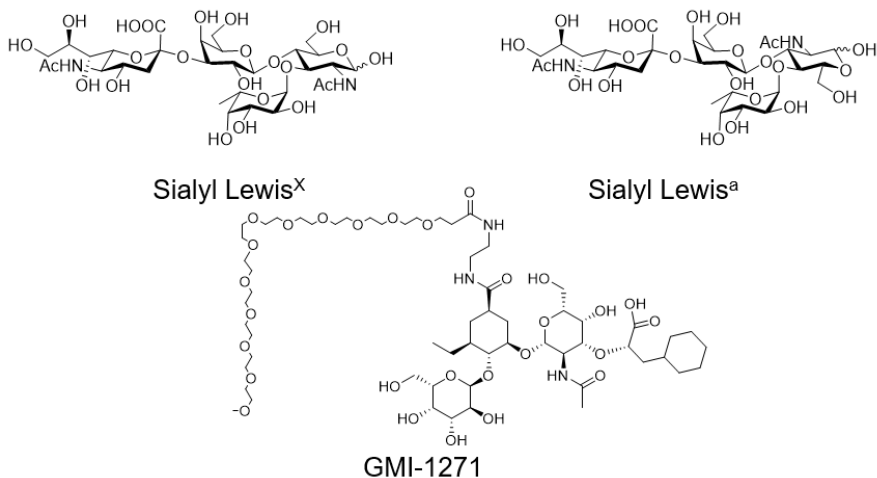


Figure 1. The role of selectins in the inflammatory cascade and their ligands. A.1. A leukocyte in the circulatory system is attracted by chemokines. 2. selectin ligands expressed on the leukocyte can recognize and bind to selectins. This interaction will slow down the leukocyte by causing it to roll slowly along the surface of the endothelial cells before complete adhesion. 3. leukocytes extravate to the site of injury and mediate inflammation; B. Selectins are crucial factors for the inflammatory cascade. The ligands expressed on the leukocyte can recognize the selectins on endothelial cells. P-selectin and E-selectin share consensus repeat units. PSGL-1 is one such ligand and can recognize both P-selectin and E-selectin. At the *N*-terminus of PSGL-1, is unique motif containing three sulfates on tyrosine (tyr-46, 48, 51) and one O-glycosylation site (thr-57). The O-glycan contains sialyl Lewis X, which is commonly found in selectin-binding ligands. Another ligand is a special variant of CD44. Its

sialylfucosylated structure is involved in binding with E-selectin; C. ligands for E-selectin. Natural ligands Sialyl Lewis X and Sialyl Lewis A both have a GlcNAc core, a fucose and a sialic acid. GMI-1271 is a glycomimetic ligand based on these and with pan-selectin binding, including E-selectin, and conserves a glycan core.

3.1.1 Selectins and inflammation

In the inflammatory cascade, leukocytes tether to and roll on the vascular cell surface. They are then arrested and pass between endothelial cells into the site of injury. (Fig 1A) This cascade is initialized with stimulation by secretagogues, such as thrombin or histamine, which are released from tissue-resident sentinel leukocytes when they come into contact with pathogens.⁸⁻¹⁰ P-selectin and E-selectin are subsequently presented on the endothelial surface and recognized by their ligands leading to the initial adhesion of leukocytes. Binding of selectins to their ligands on leukocytes can also activate different signaling pathways,¹¹ movement from the blood into tissues and finally recruitment to the site of inflammation.¹²

3.1.2 Selectin and cancer migration

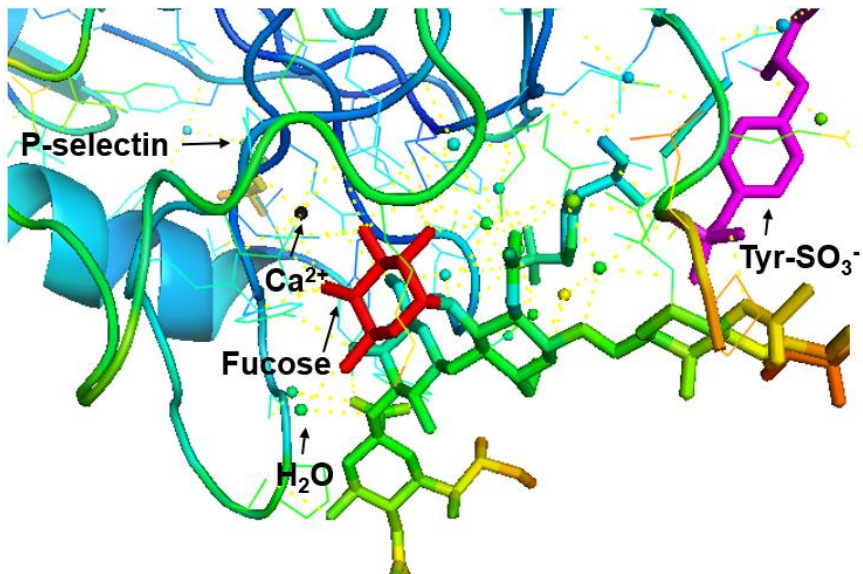
Selectins mediate not only interaction among leukocytes, platelets and endothelial cells, but also cancer metastasis through the recruitment of myeloid-derived cells.¹³⁻¹⁵ It is well established that the cancer cell surface glycosylation patterns change.^{16, 17} As results, adhesion, mobilization and migration are also changed in these cells. Structurally, selectin ligands consist of different glycan moieties. These ligands, expressed on the tumor cell surface, commonly encounter selectins, which are presented on leukocyte, platelets and endothelial cells and finally promote cancer metastasis.^{18, 19} P-selectin and E-selectin facilitate platelet-tumor cell interactions to promote the adhesion of cancer cells.¹⁹⁻²¹ Numerous *in vitro* and *in vivo* experiments have showed the correlation between selectins and tumor migration.^{21, 22} Furthermore, some clinical trials have shown that inhibition of selectins can indeed block breast cancer metastases.^{14, 19, 22-25}

3.1.3 Selectin ligands

Monocytes, dendritic cells and macrophages can express E-selectin ligands on their surface and thereby initiate the immune response.²⁶ There are several glycan-containing molecules that are common ligands for selectins, including P-selectin glycoprotein ligand 1 (PSGL-1),²⁷ CD44,²⁸ E-selectin ligand 1 (ESL-1)²⁹ and glycolipids³⁰. (Fig 1B) Among these ligands, PSGL-1 is the most characterized molecule since it can bind to all three selectins (P-, E- and L-selectin) with a relatively

high affinity.³¹ The binding motif is at the *N*-terminus of PSGL-1. The sulfated tyrosine residues (Y-46,48,51) and *O*-glycosylation on threonine (T-57) found in this region are crucial for this binding.³² (**Fig 2A**) A further study showed that PSGL-1 recognition by selectins induces several signaling pathways, such as mitogen-activated protein kinases, the Ras pathway, Src family kinase activation and extracellular signal-regulated kinase-1(ERK-1) activation.³³ While slightly different from P-selectin, the binding requirement of E-selectin is still well conserved: E-selectin can bind a large range of glycan epitopes, which typically contain a lactosamine backbone (Glc β 1-4GlcNAc; GlcNAc, *N*-acetylglucosamine).³⁴ The most prototypical E-selectin binding determinant is sialyl Lewis X (sLeX, Sia α 2-3Gal β 1-4(Fuca1-3)GlcNAc). From co-crystal structures of ligands bound to selectins we can see calcium ions mediate and stabilize the interaction between glycan and selectin, which is a prototypical characteristic of the C-type lectin. (**Fig 2B**) In addition, another isomeric epitope sialyl Lewis A (sLeA, Sia α 2-3Gal β 1-3(Fuca1-3) GlcNAc) preserves E-selectin binding activity.⁷ Some sulfated derivatives of Le^x and Le^a can also show binding affinity.³⁵

A.



B.

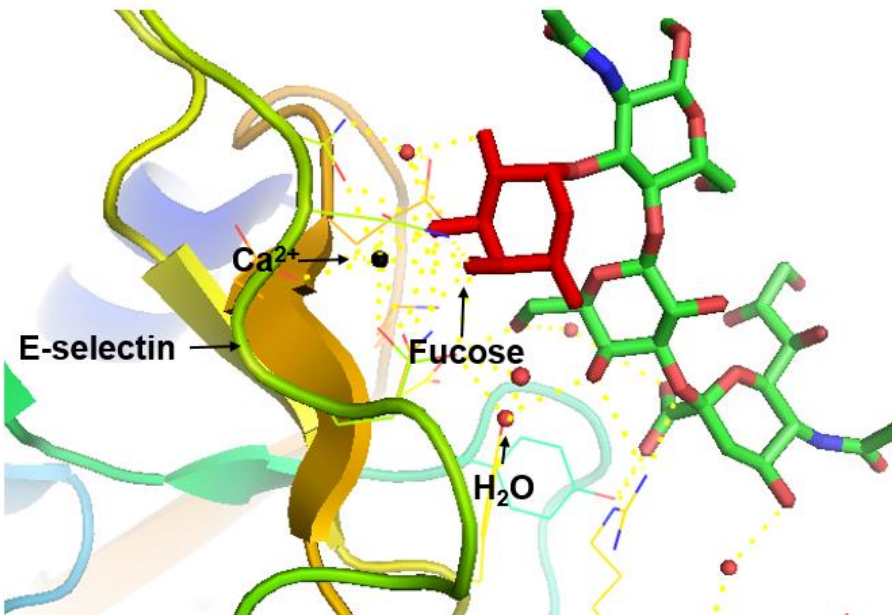


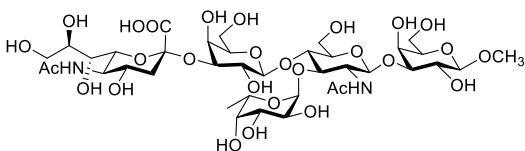
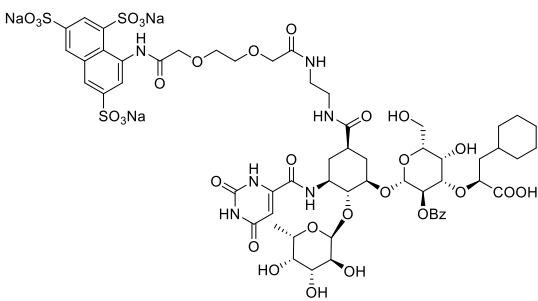
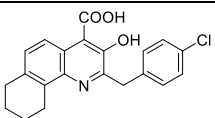
Figure 2. Ligands of selectins. A) The co-crystal structure of a PSGL-1 peptide fragment and P-selectin (PDB ID: 1g1s). Glycopeptide ligand residues are shown as rainbow sticks while selectin protein is shown as cartoon, with pink for tyrosine sulfate and red for fucose. Black dots represent calcium ions. B) The crystal structure of sialyl Lewis X and E-selectin (PDB ID: 1g1t), color-coded as above with additionally red

dots standing for water.

3.1.4 Targeting selectins in drug development

Due to the critical role of selectins in the pathogenesis of both the inflammatory cascade and cancer metastasis, several candidates have been developed to inhibit the adhesion between selectins and their ligands. One approach was to base these on mimics of sialyl Lewis X (sLeX) or tyrosine sulfate. Among these small molecules, GMI-1271(**Fig 1C**), developed by GlycoMimetics, is in clinical trials for acute myeloid leukemia and multiple myeloma (NCT02306291, NCT02744833, NCT02811822, NCT03616470, NCT03701308), although several of these have recently returned negative results because of side effects and safety problems. Despite the uncertain prospect for small molecule inhibitors, antibodies that bind to selectins have been developed to prevent inflammatory disorders and showed quite promising outcomes.^{24, 36} In 2019, crizanlizumab was approved for use by the FDA in treatment of sickle cell disease, in which selectin-mediated adhesion is involved in vaso-occlusive crises. It is the first selectin-targeting drug to be approved,^{37, 38} but likely not the last.

Table 1. Selectin antagonists and clinical trials

Name and structure	Disease	Status
 <p>Cylexin (CY-1503)</p>	Cardio-vascular injury	stopped ³⁹
 <p>Rivipansel (GMI-1070)</p>	Sickle cell crisis	Stopped ⁴⁰
Crizanlizumab (monoclonal antibody)	Sickle Cell Disease	approved ³⁸ , 41-47
	venous thrombotic diseases	Stopped ⁴⁸

PSI-697		
Inclacumab (RO4905417) (monoclonal antibody)	Peripheral Arterial Disease	Phase II ⁴⁹

Even though quite a lot of effort has been devoted to selectin-targeted molecules, there is still no real small-molecule medicine that has reached the market. Early research indeed showed that selectin inhibitors may be powerful weapons for vascular disease and cancers and some of these lead compounds were quite promising, such as GMI-1070 which unfortunately stopped at phase III. Despite this, the success of an antibody suggests that a more 'biological-like' approach may be more fruitful. We hope to capture the interaction of an antibody in a form more amenable to mass-production – a macrocyclic peptide.

3.1.5 Outline

Given that selectins are a promising target for a range of disorders, we turned our attention to E-selectin as a target for generating inhibitors from mRNA display. The RaPID (Random nonstandard Peptide Integrated Discovery) system is a powerful technology that fuses FIT (Flexible *in Vitro* translation, a methodology utilizing flexizymes integrated with a custom-made *in vitro* translation apparatus for genetic code reprogramming) and mRNA display to achieve high throughput selection of around 10^{13} peptides.⁵⁰ (see section 1.2&1.3 for more details) (**Fig 3**) In this, a randomized mRNA library is linked to puromycin at its 3'-end and introduced into *in vitro* translation to produce a random peptide library. When the ribosome reaches the puromycin, an amide bond is formed between the mRNA and the peptide. By using the FIT system, peptides can be synthesized and displayed with nonstandard elements, such as unusual side chains and macrocyclising functional groups (see section 4.1 for a discussion of the advantages of macrocyclic peptides). Following translation, reverse transcription produces a more stable peptide-mRNA-cDNA library and the resulting fusion can be panned across immobilized protein. The recovered tight-binding members of the library can be amplified with PCR for further rounds and once sufficient enrichment of binding sequences is achieved, deep sequencing of the DNA can identify candidate hits.

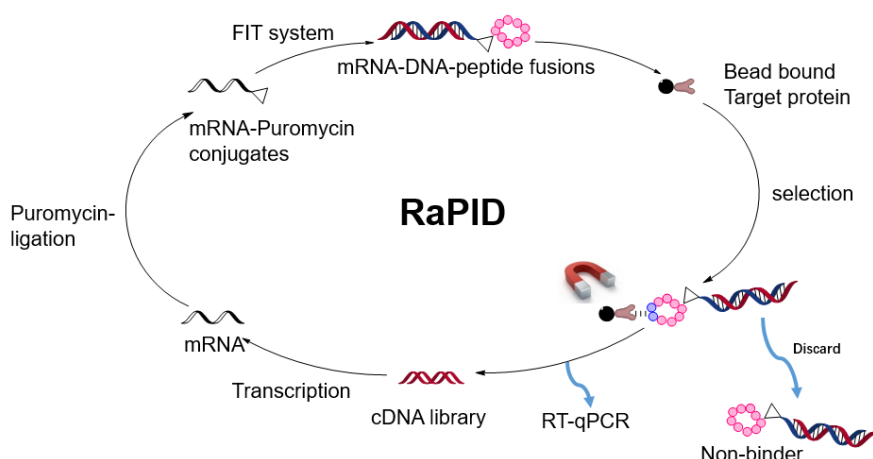


Figure 3. The cycle of the RaPID system. DNA libraries are assembled from commercial oligos. After transcription and puromycin ligation to the 3' end of the resulting mRNA, the mRNA library is formed. Translation of the mRNA library in the FIT system results in a macrocyclic peptide library which is conjugated to the cognate mRNA. Reverse transcription is carried out to generate the cDNA of each mRNA molecule. This cDNA-mRNA-peptide library is panned against the immobilized target to derive high affinity hits. The cDNA is recovered and amplified by PCR and used for further rounds. After multiple rounds of selection, a sample is sent for deep sequencing after which hits can be identified and further characterized.

In this chapter, we describe a selection against E-selectin from a library of random cyclic peptides with total length of 17 amino acids by means of the RaPID system. To improve the chances of tight-binding and inhibitory hits, we describe the synthesis and reprogrammed translation of several binding factors to incorporate into this library, aiming to mimic elements of the glycan-protein interaction. From this selection, multiple hits were identified and several of these showed low nanomolar binding ($K_D \sim 1\text{-}20\text{ nM}$) affinity for E-selectin and so are potential leads for medicinal chemistry work.

3.2 Results and discussion

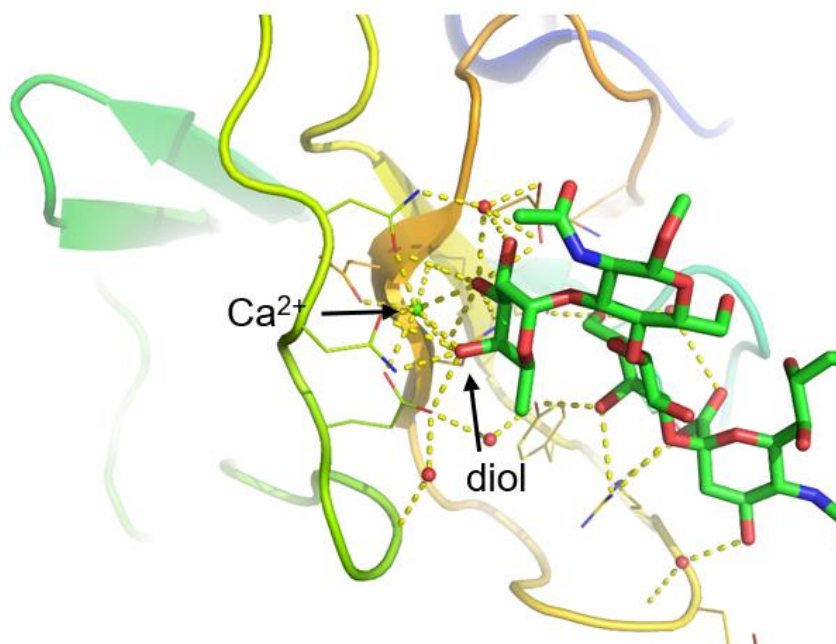
3.2.1 Binding factors for E-selectin

The glycan binding motif for E-selectin is highly conserved, with a large part of the interaction strength coming from coordination of a calcium ion by vicinal diols in the sugar and the side chains of E-selectin. (**Fig .4A**). This chelation binding mode is not

accessible to peptides containing the standard set of 20 amino acids. Work on other lectins has suggested that the limiting case of this type of interaction is a terminal acyclic 1,2-diol. The co-crystal structure of lectin and glycan indicated that such a terminal hydroxyl group can coordinate to a protein bound calcium ion only if it has no further substitution because of a highly restricted binding pocket.⁵¹ To allow the panel of interacting factors designed here to be as broadly applicable as possible, we considered that a terminal 1,2-diol should be one binding factor to incorporate in our E-selectin targeting library and that glycerol-modified cysteine (Cys-diol) provided a convenient means to incorporate this.

Analysis of an ensemble of crystal structures for carbohydrate binding pockets showed that aliphatic hydrophobic residues are disfavored while aromatic side chains are favored.⁵² Notably tryptophan, tyrosine and histidine have a higher frequency than other amino acids. The data presented in that work showed that an electronic effect plays a vital role for this binding: relatively electron-poor carbohydrate C-H bonds in a CH- π interaction with the electron rich aromatic residues. These observations have been validated by computational studies⁵³ and in solution-phase NMR.⁵⁴ A natural binding partner for these ubiquitous electron-rich aromatic groups in carbohydrate binding sites is a counterpart inverse electron density aromatic ring, such as in pentafluorophenylalanine (5F-Phe). As shown in **Fig 4B**, the electrostatic potential maps of Trp and pentafluorophenylalanine are complementary and so allow a strong direct stacking interaction.

A.



B.

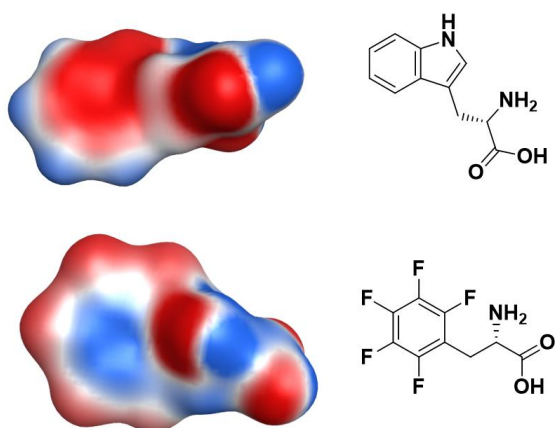


Figure 4. Binding factors for E-selectin. **A.** The binding between E-selectin and sLeX, emphasising the importance of calcium chelation. The green dot in the middle is a calcium ion, sialyl Lewis X is shown as sticks colored by atom type (green = carbon, red = oxygen, blue = nitrogen), and tightly bound water molecules are shown as red spheres. PDB ID: 1G1T. **B.** The complementary electrostatic potential map of Trp and pentafluorophenylalanine. Red indicates electron density, blue indicates electron deficiency.

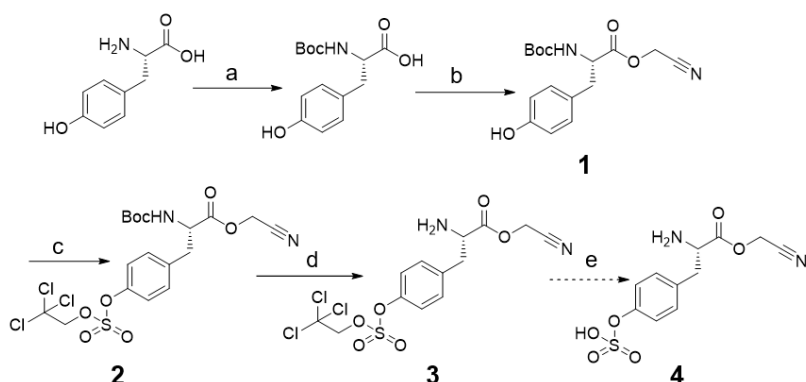
Looking further into binding studies of PSGL-1 and its mutations with E-selectins, it is apparent that E-selectin requires O-glycosylation but that sulfates do not appear to increase the binding significantly.⁵⁵ Despite the data that suggests that deletion of tyrosine sulfate on PSGL-1 will not influence the binding to E-selectin,⁵⁵ we nonetheless considered sulfate as a relevant binding factor when designing a library for selectins in general. What's more, literature shows that some sulfated glycans can bind to E-selectin.³⁵ GMI-1070, the glycomimetic based on PSGL-1, contains a sulfonate as one binding motif, and it shows efficient selectin binding and inhibition (NCT02187003). Taking this into account, we thought that a sulfate or sulfonate is certainly a reasonable binding factor for a lectin-focused library in general, and there would be no harm in also including this for E-selectin. Given that P-selectin in particular has an affinity for tyrosine sulfation (Tyr(SO₃)) we considered this (or a stable mimic) as the most logical form for including this in our library.

We thus identified 3 potential binding factors to include in a lectin-targeted library to supplement the chemical diversity of the standard 20 amino acids: a diol, an inverse electron-density aromatic ring and a sulfate (or sulfonate).

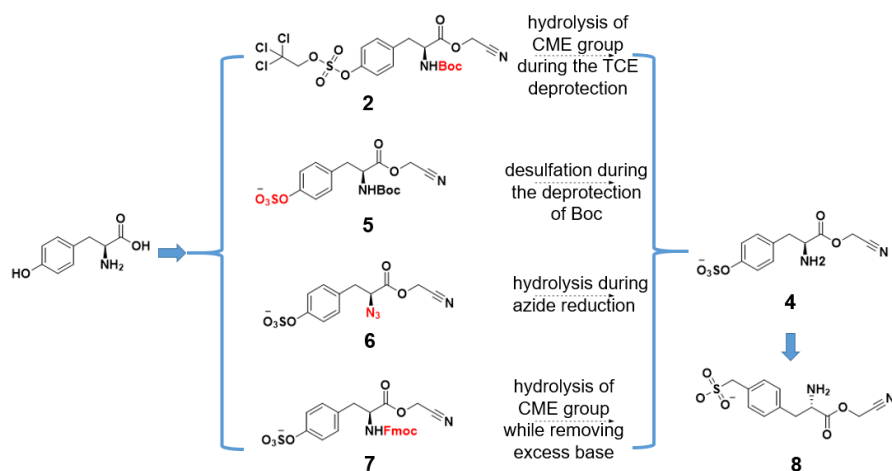
3.2.2 Synthesis of binding factor for E-selectin

To incorporate the identified binding elements into a peptide library using the FIT system we need these as an active ester on the carboxylic acid in an amino acid, and so we decided to synthesize Cys-diol-DBE (dinitrobenzyl ester), 5F-Phe-CME (cyanomethyl ester) and Tyr(SO₃)-CME. We started synthesis with the most challenging target, Tyr(SO₃)-CME. The synthetic route for this building block is shown in **Scheme 1A**. We installed a Boc protecting group onto tyrosine, then synthesized the Boc-tyrosine cyanomethyl ester by using chloroacetonitrile to afford compound **1**. After that, we incorporated chlorosulfuric acid 2,2,2-trichloroethyl ester under basic condition to achieve compound **2**. This was converted to compound **3** by treating with TFA. Until this point, all steps were accomplished without problems. However, we could not obtain compound **4** when we tried to use zinc reduction to do the final deprotection. Unfortunately, we could only find the hydrolysis product of **4**. In this reaction it proved very difficult to find a balance between removal of the final protecting group and stability of the cyanomethyl ester.

A.



B.

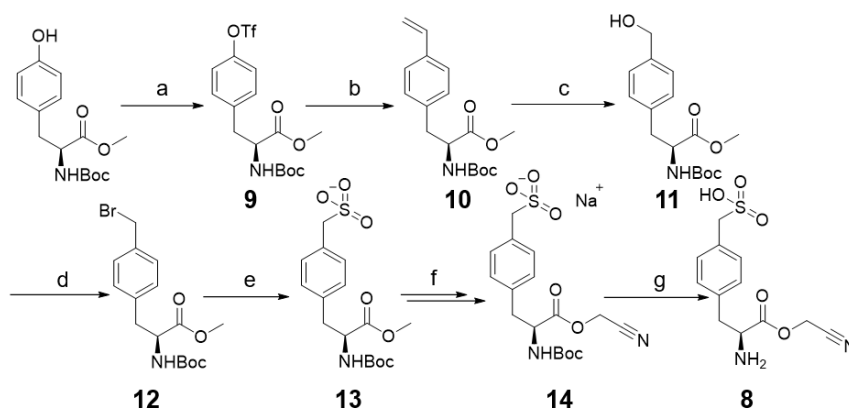


Scheme 1. The synthesis plan for an activated ester of tyrosine sulfate. A. The attempted synthesis route for CME-activated tyrosine sulfate. a. Boc₂O, sat. NaHCO₃; b. Chloroacetonitrile, Triethylamine; c. 2,2,2-trichloroethyl sulfurochloride, triethylamine; d. Trifluoroacetic acid, triisopropylsilane; e. Zinc dust, ammonium formate. B. Tests of alternate routes for synthesis of CME-activated tyrosine sulfate, and critical issues encountered in each route.

Following several attempts at this deprotection, we also explored several alternate synthesis strategies. (**Scheme 1B**) Unprotected sulfate was successfully installed on tyrosine with Boc deprotection to give compound 5. However the sulfate predictably proved too unstable in all deprotection conditions, such as different concentrations of formic acid or even heating in water. Next, we turned to using azide as a protecting group for the amine. We transferred the sulfate onto the *N*-azido-tyrosine and got compound 6 without problem. After that, we attempted to use trimethylphosphine or

hydrogen to reduce the azide. However, the cyanomethyl ester again proved to be too unstable under these conditions. Finally, we chose Fmoc as an amine protecting group (compound **7**) and used DBU or dimethylamine for deprotection. This again proved to be too harsh for the CME stability, affording only hydrolyzed product. Nevertheless, later research showed that a tyrosine sulfate ester can be achieved with an Fmoc protection strategy by the unusual choice of a DBE ester activating group instead of the CME typically used for aromatic amino acids.⁵⁶

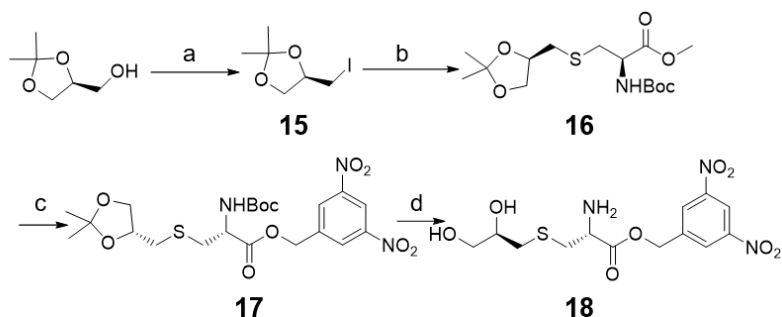
Due to the difficulties encountered in synthesis of compound **4**, we tuned our attention to a C-sulfonate (compound **8**) which is relatively stable in basic and acidic conditions. While this was more tolerant to deprotection conditions, it required a much more elaborate synthesis route to introduce the sulfonate, which proceeded as follows. (**Scheme 2**) We started from Boc-Tyr-OMe and installed a triflate on the phenol to get compound **9**. Using this we carried out Stille coupling with tributyl(vinyl)tin to afford compound **10**. Ozone oxidation and sodium borohydride reduction were carried out to convert the alkene to the hydroxymethane (compound **11**). Bromination of this was carried with NBS (*N*-bromosuccinimide) and triphenylphosphine to afford compound **12**. From this we synthesized the sulfonate by boiling **12** with sodium sulfite solution to displace the bromide. Finally, we hydrolyzed the methyl ester with lithium hydroxide, installed the cyanomethyl ester by chloride displacement from chloroacetonitrile, and deprotected the Boc with formic acid. The final deprotection proceeded smoothly in this case, and we obtained the final product **8** in one pot for the final three steps. The synthesis of **8** was accomplished over 11 steps from tyrosine with a 9% overall yield.



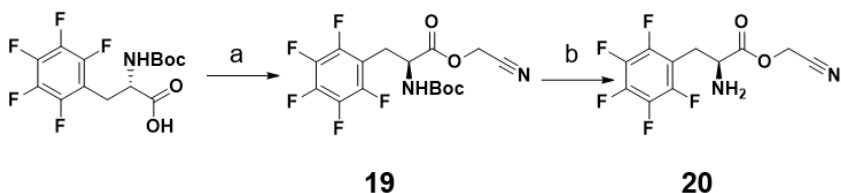
Scheme 2. The synthesis of Phenylalanine methylsulfonate. a, Triflic anhydride, DCM/Pyridine, 88%; b, tributyl vinyl tin, Lithium chloride, Pd(PPh₃)₂Cl₂, DMF, 90 °C, 85%; c, 1). ozone, -78 °C. 2) NaBH₄, 60%; d, Triphenylphosphine,

N-bromosuccinimide, 83%; e, Na₂SO₃, H₂O, 90 °C, 95%; f, Lithium hydroxide, THF/H₂O then chloroacetonitrile, DMAP, TEA, DMF, 33% over two steps; g, HCOOH, 80%.

To generate an amino acid with a sidechain vicinal diol, we considered multiple candidates for conjugation to an amine (Lys), a hydroxyl (Ser) and a thiol (Cys). Ultimately cysteine was chosen for the simplicity of thiol conjugation. synthesis of Cys-diol-DBE was achieved with the route below. (**Scheme 3**) Iodomethyl-2,2-dimethyldioxolane (**15**) was acquired with triphenylphosphine and iodine from commercial available isopropylidenglycerol in enantiomerically pure form. Compound **15** was reacted with Boc-Cys-OMe under basic conditions to form **16**. Hydrolysis of the methyl ester was achieved by lithium hydroxide, and 3,5-dinitrobenzyl chloride was added to form the activated ester. Finally, deprotection was carried out in pure formic acid to get the target product **18** with a 1.7% overall yield over 5 steps. A large part of this poor yield was because of product loss during the final purification and so this can likely be optimized, but sufficient material was acquired to proceed.



Scheme 3. The synthesis of Cys-diol-DBE. a, Triphenylphosphine, iodine, 90°C, 84%; b, Boc-Cys-OMe, DIPEA, 71%; c, Lithium hydroxide, then 3,5-dinitrobenzyl chloride; d, HCOOH, 3% over two steps.



Scheme 4. The synthesis of 5F-Phe-CME. a, Chloroacetonitrile, TEA, 97%; b, HCOOH, 78%;

Finally, the synthesis of pentafluorophenylalanine cyanomethyl ester followed the

reported literature.⁵⁷ The *N*-Boc-pentafluoro phenylalanine (Boc-5F-Phe) starting material was commercial available, and on this we installed cyanomethyl ester and then deprotected the Boc group with formic acid to get target compound **20** with a 76% overall yield. (**Scheme 4**)

3.2.3 Verification in *in vitro* translation

With the 3 candidate binding factors in hand, we continued with testing of their incorporation into peptides by reprogrammed translation. First, we determined the acylation yield with the appropriate flexizyme by using a truncated tRNA acceptor stem mimic, microhelix (22 nt).⁵⁸ Analysis of the reaction progress was by denaturing polyacrylamide gel electrophoresis (PAGE). For all three, maximum product formation was observed in 2 hours with acceptable yield (**Fig 5**), although notably less for Cys-diol-DBE (commonly observed for dFx-catalyzed acylation). Henceforth, we used these optimized condition to charge these amino acids onto tRNA.

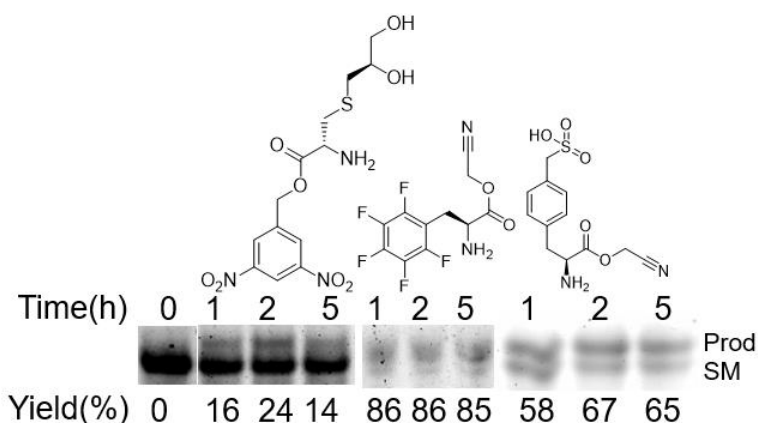


Figure 5. Aminoacylation testing for the three glycan binding site targeting factors. 'Prod'= acylated microhelix; 'SM'= unreacted microhelix. The relative yield was calculated by densitometry in ImageJ.

With aminoacylation confirmed, we next tested the compatibility with reprogrammed *in vitro* translation. For this we tested two different tRNA bodies. After aminoacyl-tRNA synthetases (aaRS) charge amino acids onto their compatible cognate tRNAs⁵⁹, the carrier protein EF-Tu is tasked with transferring the aa-tRNA into the ribosome.⁶⁰ However, EF-Tu has different binding affinities to different tRNAs and amino acids.⁶¹ Too strong or too weak binding will influence the translation efficiency, and the binding affinity for the tRNA in each case compensates the varied binding to amino acids. Considering the possible influence of EF-Tu binding on translation of an amino acid

with unknown binding affinity, we tested 5F-Phe, Cys-diol and Phe(*p*-CH₂SO₃H) translation with two different tRNA bodies: EnAsn and EnGlu (engineered forms of the Asn and Glu tRNA), which have relatively strong binding and weak binding to EF-Tu (respectively). For this test, we prepared a test DNA template encoding a peptide which has a single Trp in the sequence, giving a convenient codon for reprogramming. The three new amino acid building blocks were each charged onto tRNA^{EnGlu}_{CCA} and tRNA^{EnAsn}_{CCA} by flexizyme. After purification, the tRNA pellets were utilized directly for *in vitro* translation. Following precipitation of the translation system proteins by addition of acetonitrile, the resulting solutions were analyzed by UPLC-MS and with the results shown in **Fig 6-8**. For the sulfonate reprogramming, we found a peak in the mass spectrum corresponding to the product using the EnAsn tRNA body, while the EnGlu body tRNA didn't give a detectable product peak (**Fig 6A**). For Cys-diol reprogramming both EnGlu and EnAsn tRNA gave a strong product peak, showing this unnatural amino acid appears to have no preference for tRNA body (**Fig.7**). Finally, for 5F-Phe reprogramming the EnGlu tRNA body showed a higher yield than the EnAsn tRNA body (**Fig.8**). All three amino acids were thus able to be charged onto tRNA and translated into peptides.

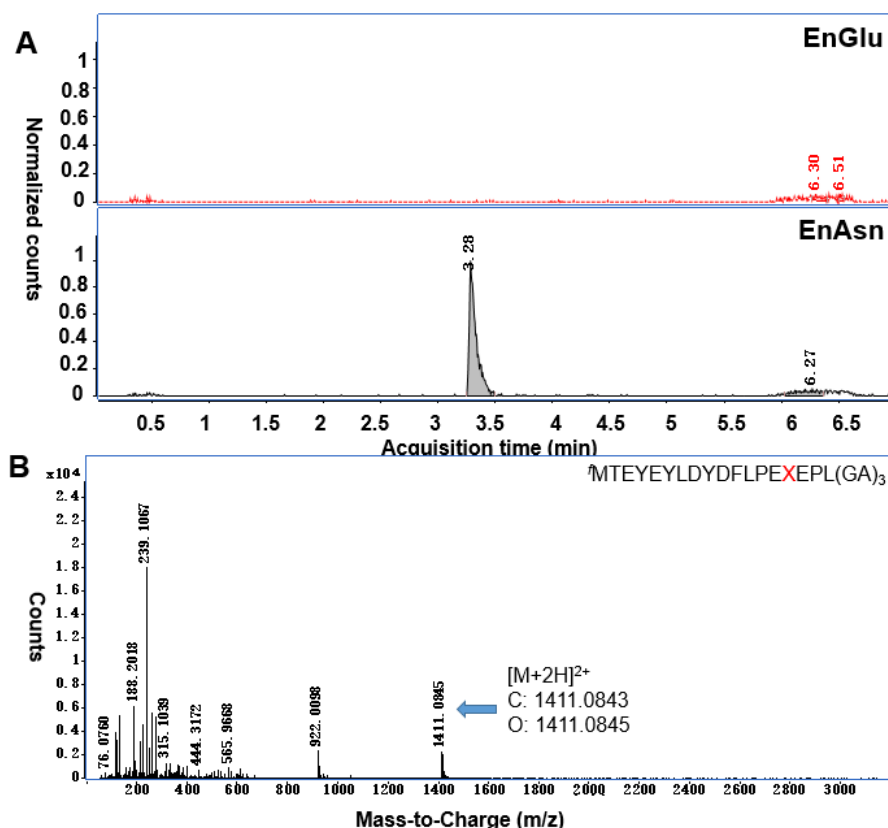


Figure 6. UPLC-MS analysis of Phe(*p*-CH₂SO₃H) reprogramming. A. EIC for the desired peptide mass with the respective tRNA body as shown; B. mass spectrum of the reprogrammed peptide, with sequence as shown. X= Phe(*p*-CH₂SO₃H).

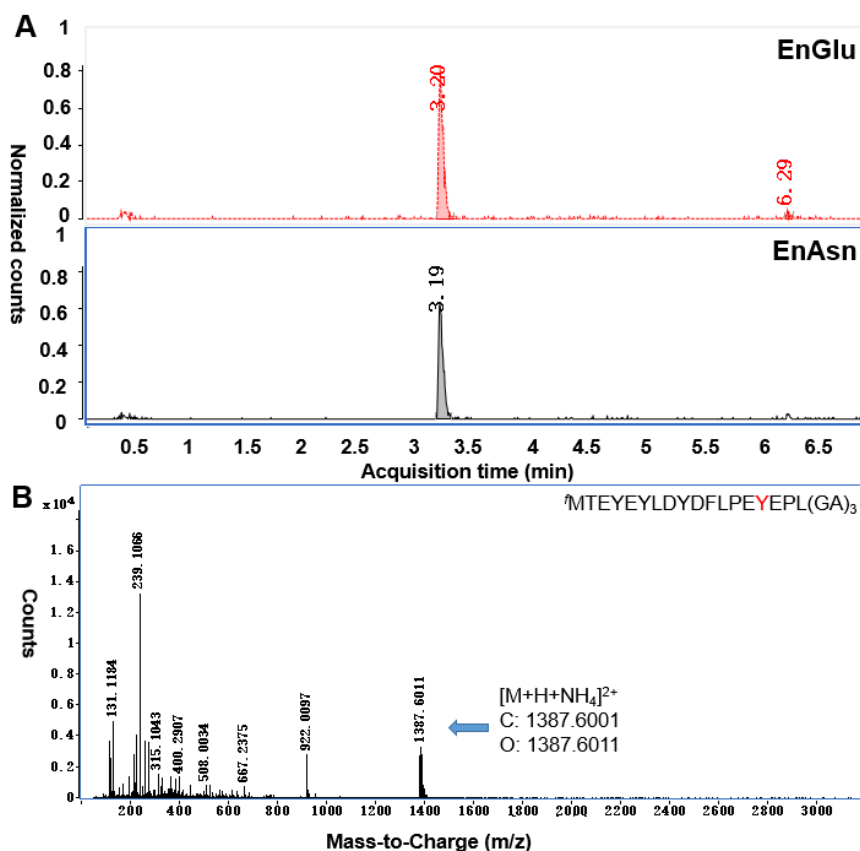


Figure 7. UPLC-MS analysis of Cys-diol reprogramming. A. EIC for the desired peptide mass with the respective tRNA body as shown; B. mass spectrum of the reprogrammed peptide, with sequence as shown. Y=Cys-diol.

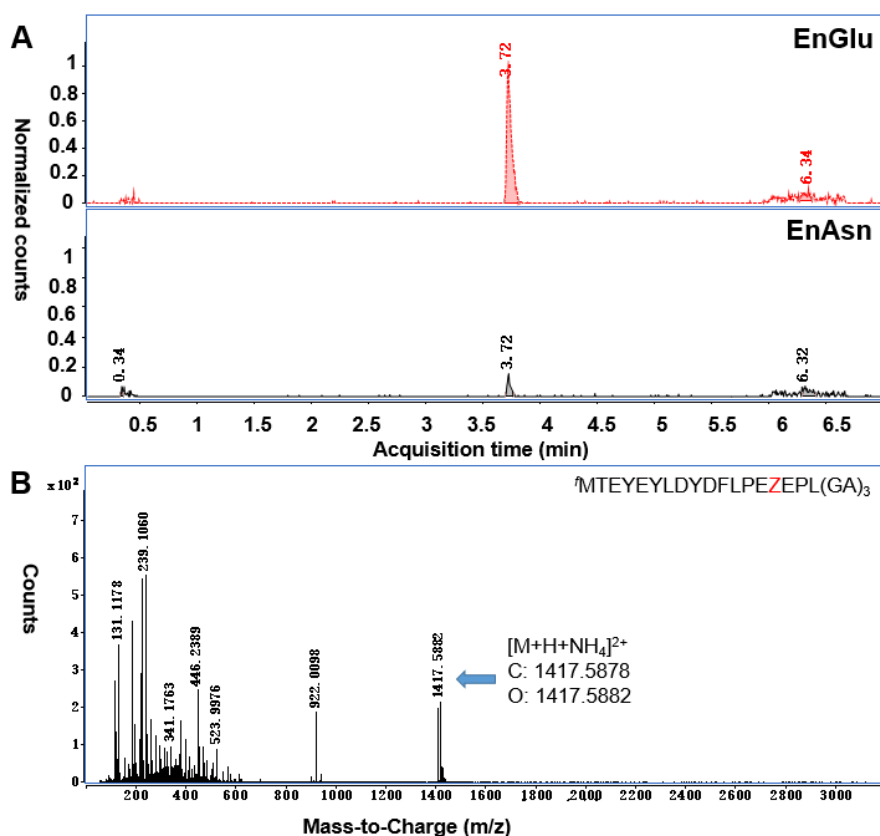


Figure 8. UPLC-MS analysis of 5F-Phe reprogramming. A. EIC for the desired peptide mass with the respective tRNA body as shown; B. mass spectrum of the reprogrammed peptide, with sequence as shown. Z= pentafluorophenylalanine.

3.2.4 Library design and selection

In order to find novel sequences that bind to E-selectin we designed a macrocyclic peptide library incorporating the above three binding factors for use in the RaPID system. At the RNA level, the library contained a ribosome binding site, an AUG start codon, 15 NNK codons (N= A, G, C or U, K=A or U), a UGC codon (Cys) and a sequence encoding a GSGSGS peptide spacer between mRNA and peptide. By omitting Met and Ala from the translation reactions with this template we can generate three vacant codons. For this, Ala was chosen as an amino acid that is unlikely to contribute significantly to binding and also to facilitate any future ‘alanine scanning’ studies of interacting sequences. The AUG start codon was used to initiate translation with ClAc-L-Tyr and ClAc-D-Tyr in two separate translation reactions to generate two libraries, with the change in stereochemistry potentially giving access to different

peptide conformations. The chloroacetyl group can react with spontaneously with the C-terminal cysteine following translation to cyclize the peptide (or with another cysteine if this arises by chance in the random region of the library). Finally, the AUG elongation codon was used to incorporate 5F-Phe, while the GCG and GCU alanine codons were used to install Phe(*p*-CH₂SO₃H) and Cys-diol, respectively.

A.

1st	2nd				3rd
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
					A
C	Leu	Ser	Stop	Trp	G
	Leu	Pro	His	Arg	U
					C
A					A
	Leu	Pro	Gln	Arg	G
	Ile	Thr	Asn	Ser	U
G					C
					A
	5F-Phe*	Thr	Lys	Arg	G
	Val	Phe (<i>p</i> -CH ₂ SO ₃ H)	Asp	Gly	U
					C
	Val	Cys-diol	Glu	Gly	A
					G

B.

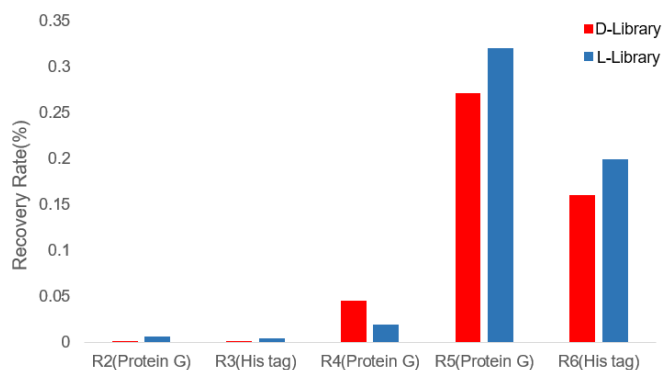


Figure 9. Selection of macrocyclic peptides against E-selectin. **A.** codon table for the reprogramming used to create the targeted library. (*ClAc-D/L-Tyr in initiation) **B.** Enrichment of peptides binding E-selectin across 6 rounds using the RaPID system.

Red bars represent the recovery per round from the D-Tyr initiated library and blue bars represent the recovery per round from the L-Tyr initiated library. Round number is indicated on the X-axis, along with the target immobilization used for that round.

An E-selectin construct with N-terminal Fc- and C-terminal His-tags was immobilized on Protein G magnetic beads and his-tag magnetic beads (respectively) for panning with these libraries. These proteins were used in alternating rounds to decrease false-positive hits from binding to the functionalized magnetic beads directly. Both libraries showed signs of enrichment from the 4th round as measured by qPCR of recovered cDNA (**Fig 9**), and after 6 rounds of selection were judged to be complete (not enriching further) so samples from all rounds were sent for deep sequencing. The dominant peptide sequences found are shown in **Fig 10**.

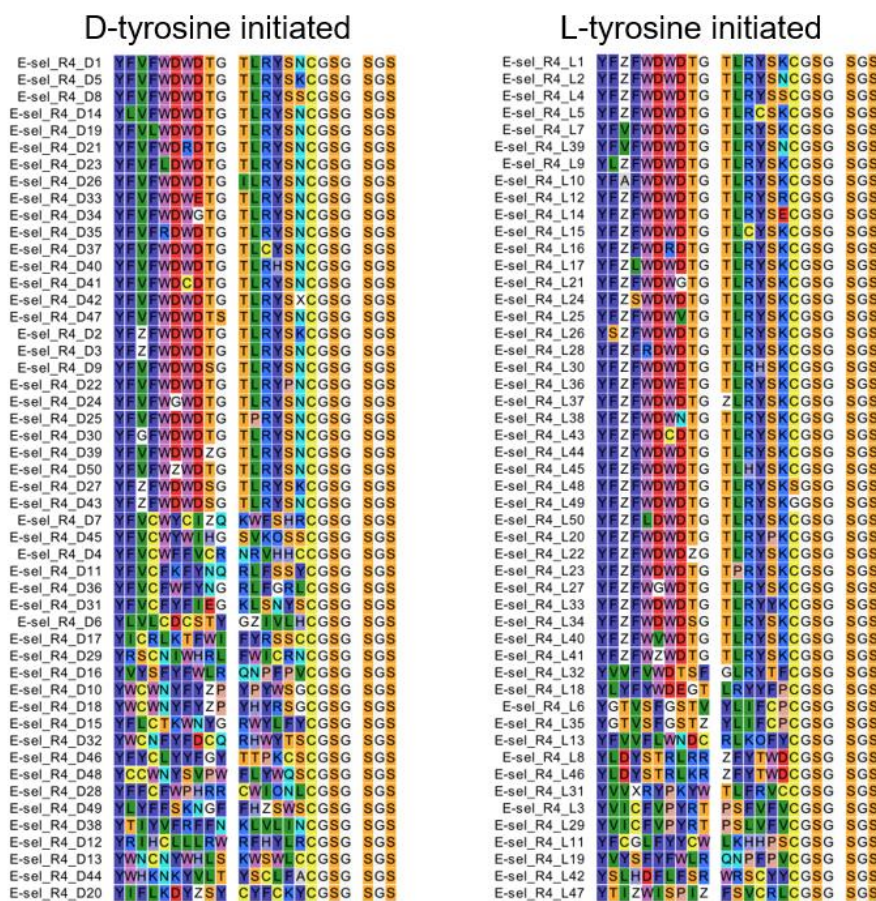


Figure 10. Sequencing results for the selections against E-selectin. The top 50 most abundant unique peptides from the 4th round are shown ordered by sequence conservation as one-letter codes with additionally Z=Phe(p-CH₂SO₃H), displayed in

Rasmol colors. Sequence naming reflects the initiating amino acid stereochemistry (D/L) and the sequencing abundance rank (number, *e.g.* L1 = most abundant unique sequence from the L-Tyr initiated library).

To our surprise, the sequencing results showed that the only amino acid binding factor that appeared to be enriched was Phe(*p*-CH₂SO₃H). Even though literature indicates that removing the sulfate in PSGL-1 does not influence the binding between PSGL-1 and E-selectin, we still found a substantial fraction of sulfonates in the most abundant hits. However, Cys-diol and 5F-Phe, intended as more general glycan interaction mimics, were not observed. Furthermore, we also observed a relatively large fraction of sequences with Cys appearing in the random region of the library (*e.g.* D7). Such sequences form a lariat, and the conservation of parts of the 'tail' suggests it also contains residues important for binding.

3.2.5 Assessing binding of the peptide hits

From the sequencing results we identified several conserved motifs, which suggests a successful selection. We next set out to test the binding affinity and location of these peptides. To qualitatively identify the candidates most likely to have high affinity for E-selectin, we chose 10 sequences out of these enriched libraries to validate in a pull-down experiment which follows the same mRNA display protocol as the initial selection but with a single defined sequence instead of a library. We assembled the DNA encoding each peptide by PCR and transcribed these to the corresponding mRNA. Following pull-down we calculated the ratio of recovered cDNA to input cDNA, which gives a recovery rate that is indicative of binding affinity. (**Fig. 11**) From this recovery ratio and the positive to negative ratio, we identified the most likely strong binder as L1, which also matches its position as the 'winner' of that selection. Other peptides such as D1, L4, L5, L6 also appear to have good binding affinity to E-selectin while others such as D2, D3, L2, L3 and L7 did not show good recovery. However, it should be noted that within the set that recovers well there is often a poor correlation between recovery ratio and binding affinity as other factors such as translation yield can influence this.

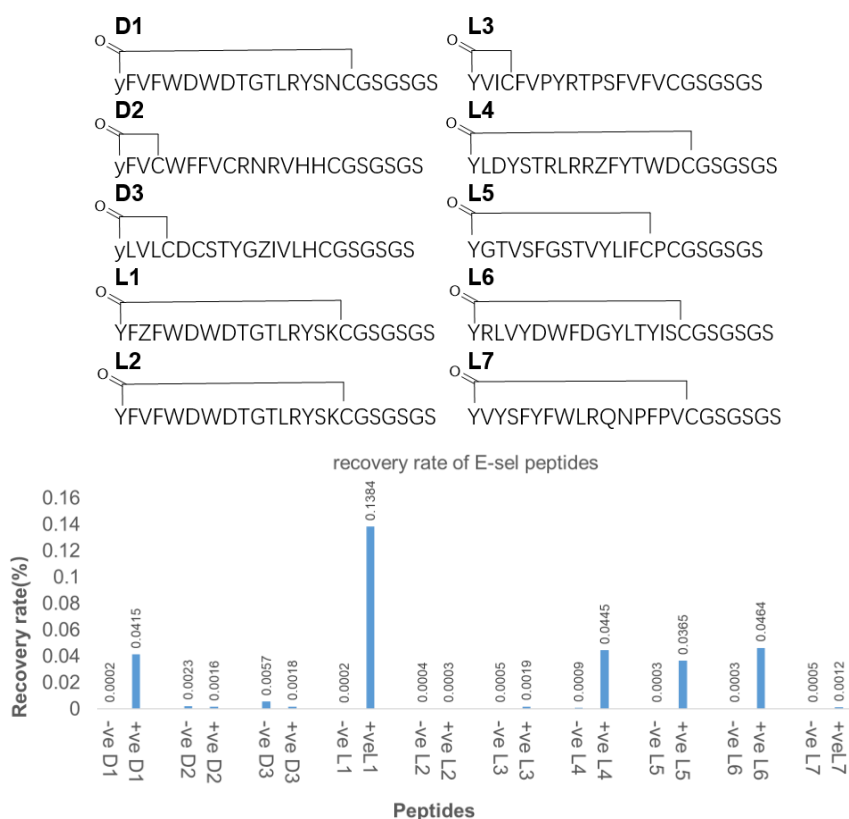


Figure 11. Validation of binding by pull-down for 10 peptides. 'y'= D-Tyr, 'Z'= Phe(*p*-CH₂SO₃H), with all the peptides cyclized by thioether bridges formed between chloroacetyl tyrosine and cysteine as indicated in the semi-structural representation. "+ve", pull-down of cDNA from E-selectin immobilized magnetic beads; "-ve", pull-down from magnetic beads only.

With the pull-down results indicating several potential candidates, we next quantified the strength of binding. There are several methods we can use to measure K_D (dissociation constant),⁶² such as ITC (isothermal titration calorimetry)⁶³, FP (fluorescence polarization)⁶⁴ and SPR (surface plasmon resonance)⁶⁵. We initially chose FP to measure the K_D as this is a solution method that does not need a lot of the relatively expensive commercial proteins used in the selection, and so we synthesized these 10 peptides by solid phase synthesis with a GGK(5/6-Fam)G (with 5/6-Fam=5/6-Carboxyfluorescein) linker at the C-terminus for the required fluorescence. Unfortunately, after multiple attempts varying protein concentration and detergents, we could not see any clear and consistent polarization change. Rather than persist with this approach and risk wasting our limited protein stocks, we instead

turned to SPR.

SPR is a robust method for binding analysis that can provide separate kinetic constants for binding and dissociation as well as the equilibrium binding constant. By immobilizing ligand on the chip and flowing the surface with analyte, a response is generated that reflects binding status from minute reflection angle changes. Changes in this signal are translated into kinetic values. To measure the binding of our peptides we thus immobilized E-selectin on the chip to then flow over the surface with varying concentrations of peptides. Of note is that the selectins are fairly large proteins (~100 kDa), and so the signal change from binding of ~2 kDa peptides is relatively low. To ensure the immobilized selectins on our chip were stable and in an active conformation we used an established peptide ligand for E-selectin as a control, the linear sequence DITWDQLWDLMK.⁶⁶ Subsequently we carried out measurements with the fluorescent peptides from the above selection using a single cycle kinetics model. The association and disassociation rate constants of these peptides are shown in **Table 2**.

Notable in these results is that some of the SPR data did not match with the pull-down experiment data. The ‘winners’ of the two selections, D1 and L1, did not show strong affinity to E-selectin on SPR (but binding was still detected). For L1 we observed a much slower on rate and this contributes heavily to its low K_D . In the selection the relatively long binding equilibration step may have compensated for this. Of all the peptides tested, D3 showed the highest binding affinity to E-selectin with a fast k_{on} rate and a slow k_{off} rate. This was surprising, as this peptide showed no recovery in the pull-down assay above (and reproducibly so). In fact, the k_{off} of D3 is at the measurement limit of the biacore T100 used, so this K_D has a larger error than the other peptides and can be seen as an upper limit for binding affinity. Moreover, D3 has a different structure from the other peptides: it has two cycles, one being cyclized by the chloroacetyl group and a cysteine and the other being formed by a disulfide bond. This may allow it to access a conformation not available to the other hits, which may contribute to its higher affinity for E-selectin. Aside from D3, L6 also showed a high affinity for E-selectin. Unlike D3 and L1, L6 only contains canonical amino acids in its sequence, potentially making it easier to scale up. Finally, peptides D1 and L2 have almost the same sequence except that they have different stereochemistry for the initiating amino acid (and a N/K substitution at the C-terminus). These differences appear to have only a modest effect on binding affinity, suggesting that in this case the stereochemistry of the first amino acid is not crucial for the peptide’s binding

conformation.

Table 2. binding affinity of selected peptides to E-selectin.

Name	Sequence	K _D (nM)	K _{on} (M ⁻¹ .s ⁻¹)	K _{off} (s ⁻¹)	χ ²	U-value
D1	yFVFWDWDTGTLRYSNCGGkG	241	1.54*10 ³	3.72*10 ⁻³	0.668	3
D3	yLVLCD CSTYGZIVLHCGGkG	~1*	1.15*10 ⁵	3.56*10 ⁻⁴	3.830	43
L1	YFZFWDWDTGTLRYSKCGGkG	>1000	375	4.30*10 ⁻³	0.528	15
L2	YFVFWDWDTGTLRYSKCGGkG	110	3.02*10 ⁴	3.32*10 ⁻³	0.817	4
L4	YLDYSTRLRRZFYTWDCGGkG	67	3.96*10 ⁴	2.63*10 ⁻³	0.469	5
L6	YRLVYDWF DG YLT YISCGGkG	18	8.37*10 ⁴	1.48*10 ⁻³	0.262	4
+ve	DITWDQLWDLMK	20	5.28*10 ⁴	1.07*10 ⁻³	1.05	7

Peptides were capped with a chloroacetyl group and cyclized with the first cysteine. k = Lys(5,6-Fam), Z= Phe(*p*-CH₂SO₃H), y=D-Tyr. **k_{off}* is at the lower measurement limit of biacore measurement and so cannot be accurately determined.

Encouraged by this first set of SPR data, we next wanted to investigate if the sulfonate is crucial for strong binding to E-selectin and if the bicyclic structure of D3 improves its binding affinity. Based on these questions, we synthesized some variants of D3, D1 and L4: D3 with phenylalanine instead of its methylsulfonate derivative (D3F), D3 with the second and third cysteine replaced with alanine and so only the *N*-terminal macrocycle (D3AA), D1 with the phenylalanine methylsulfonate in its conserved position in other related sequences (D1Z), and L4 with phenylalanine instead of its methylsulfonate derivative (L4F). We also wanted to ensure that the fluorophore was not contributing heavily to binding, so also synthesized D3 without the fluorescein-containing C-terminal linker (D3no Fam). The binding data for these peptides are shown in **Table 3**.

Table 3. binding affinity of variant peptides to E-selectin.

Name	Sequence	K_D (nM)	$K_{on}(M^{-1}.s^{-1})$	$K_{off}(s^{-1})$	χ^2	U-value
D1	yFVFWDDWTGTLRYSNCGGkG	241	$1.54 \cdot 10^3$	$3.72 \cdot 10^{-3}$	0.668	3
D1Z	yFZFWDWDTGTLRYSNCG	48	$8.51 \cdot 10^4$	$4.09 \cdot 10^{-3}$	0.949	5
D3	yLVLCDCSTYGZIVLHCGGkG	$\sim 1^*$	$1.15 \cdot 10^5$	$3.56 \cdot 10^{-4}$	3.830	43
D3no Fam	yLVLCDCSTYGZIVLHCG	4.6	$3.30 \cdot 10^4$	$1.53 \cdot 10^{-4}$	0.411	12
D3AA	yLVLCDASTYGZIVLHAG	19	$1.75 \cdot 10^5$	$3.34 \cdot 10^{-3}$	0.308	3
D3F	yLVLCDCSTYGFIVLHCG	51	$2.20 \cdot 10^5$	$1.11 \cdot 10^{-2}$	0.252	5
L4	YLDYSTRLRRZFYTWDGCGkG	67	$3.96 \cdot 10^4$	$2.63 \cdot 10^{-3}$	0.469	5
L4F	YLDYSTRLRRFFYTWDGCG	106	$3.98 \cdot 10^4$	$4.23 \cdot 10^{-3}$	1.84	1

Peptides were capped with a chloroacetyl group and cyclized with the first cysteine. k = Lys(5,6-Fam), Z= Phe(*p*-CH₂SO₃H), y=D-Tyr. The corresponding SPR result from Table 2 is shown in orange and substitutions are highlighted in red.

Comparing both L4 and D3 with their non-sulfonated variants L4F and D3F clearly shows that the sulfonate indeed contributes substantially to the binding affinity of these peptides to E-selectin, with the dominant effect being faster dissociation without sulfonate. Similarly, for D1 the introduction of a sulfonated residue in D1Z also increases its binding affinity, in this case by increasing the association rate. As mentioned earlier, previous studies of PSGL-1 binding to E-selectin showed that elimination of the sulfate on PSGL-1 did not substantially influence the interaction.⁵⁵ It is thus unclear what these sulfonates are interacting with. Comparing D3 and D3AA, we found that the bicyclic structure of D3 indeed increases its binding strength. Even though the disulfide bond is not a stable structure in cells, considering that selectins are expressed on the surface of endothelial cell surface we anticipate that this disulfide bond can survive *in vivo* experiments and so is worth retaining. Finally, D3no FAM shows that the fluorescein is making only a modest contribution to both the on and off rates.

After characterising the binding of these peptides with E-selectin, and especially seeing the influence of the sulfonate on affinity, we were curious about their binding site within the protein. E-selectin is a glycan-binding protein, with its most well-known ligand being sLeX with a K_D value around 800 μ M.³² The positive control peptide used in SPR, which came from phage display, does not have clear competition with sLeX but does disrupt E-selectin interactions in a more biological setting (further discussed below). By carry out competition experiments, again in pull-down format, we sought an initial indication of the binding site(s) of these peptides. Pull-down was carried out

in the mRNA display system as above for the initial binding tests, while to saturate the binding site of E-selectin we used 8 mM sLex (10 times K_D) and 500 nM control peptide (25 times K_D). The results of this are shown in **Table 4**.

Table 4. Competition binding test by pull-down and RT-qPCR

Name	No competition recovery (%)	sLex competition recovery (%)	Control peptide competition recovery (%)	Bead-only recovery (%)
D1	0.0435	0.0176(0.40)	0.0019(0.04)	5.76×10^{-4}
L2	0.0124	0.0128(1.03)	0.0052(0.42)	2.54×10^{-3}
L4	0.0878	0.0554(0.63)	0.0052(0.06)	1.45×10^{-4}
L6	0.1551	0.0373(0.24)	0.0127(0.08)	3.09×10^{-3}

Parenthetical values indicate the ratio with competition over that without (values substantially below 1 indicate competition).

From these competition results it is apparent that D1, L4 and L6 bind to the same site as the positive control peptide. L2 does show some decrease in binding when in competition with this control peptide, but this decrease is an order of magnitude less and so may indicate only partial overlap of binding sites. L6 showed the largest decrease in recovery when in competition with sLex at 76%, which could indicate that it is binding in or near the glycan-binding pocket. D1 and L2 also show a small decrease, but even less so than L6. These small changes in recovery in the competition setting may be a result of the difficulty in saturating such a weak binder. It has previously been shown that the binding site of the control peptide and sLex do not overlap. This was confirmed by an experiment with E-selectin immobilized on plates that was subsequently bound by radio-labelled control peptide and cold sLex, which were found to not influence each other's binding. Despite this, in cell flow assays with E-selectin pre-incubated with this control peptide the attachment of neutrophils was still blocked.⁶⁶ This was interpreted as suggesting that the peptide binds adjacent to, but not in, the sugar binding site, and so in the context of a larger glycopeptide rather than glycan alone the interaction is still inhibited. Given that we

see tight binding for our peptides and that it is competitive with this control peptide, we predict that our peptides may also be able to inhibit the biologically relevant interaction despite the lack of clear competition by SLeX.

3.3 Conclusion

In this work we designed several lectin-targeted binding factors for incorporation into peptide libraries in an mRNA display setting: Phe(*p*-CH₂SO₃H), Cys-diol, 5F-Phe. As a first implementation these were used in a RaPID selection against E-selectin. Following enrichment of binding peptides and their identification by high-throughput sequencing, the most promising candidates were assessed in pull-down format before synthesis on solid phase. Several of these were found to bind E-selectin with low nanomolar affinity binders, one of which had an extremely slow off-rate. Only the sulfonate binding factor was found to be enriched, in contrast to expectations from literature, but further characterization showed these to be important contributors to binding. Competition experiments suggested a binding site for these peptides adjacent to or partially overlapping with the sugar binding site. The peptides identified in this work will in future be explored as E-selectin inhibitor leads for use in treating inflammatory disorders and cancer metastasis, as well as targeting agents for precision drug delivery.

3.4 Experimental

General methods.

All reagents, unless otherwise stated, were purchased from Sigma-Aldrich or Alfa Aesar and used without further purification. PURExpress (E6840S and E6850S) was purchased from New England Biolabs. ¹H and ¹³C NMR spectra were recorded on Varian 400 MHz or Bruker 600 MHz. Column chromatography was carried out on silica gel G-60. Analytic HPLC-MS was performed on an automated HPLC system (Shimadzu LC-20) equipped with UV/VIS detector at 214/280 nm on a C18 column (particle size: 5 μm, 150x4.6 mm) at a flow rate of 1 mL.min⁻¹ using a linear gradient of buffer A (0.1% Formic acid in CH₃CN/H₂O 5:95 v/v) and buffer B (0.1% Formic acid in CH₃CN/H₂O 95:5 v/v) from 10-70% B over 30 min, connected to a Bruker daltonics microTOF. Preparative RP-HPLC was performed on an automated preparative HPLC system equipped with a UV/VIS detector at 214/280nm using a C18 column (particle size: 10 μm, 250x22 mm) at a flow rate of 12.5 mL.min⁻¹ using a linear gradient of buffer A (0.1% TFA in CH₃CN/H₂O 5:95 v/v) and buffer B (0.1% TFA in CH₃CN/H₂O 95:5 v/v) from 10-70% B over 60 min. High resolution mass and UPLC-MS

measurements were acquired on an Agilent 6560 Ion Mobility Q-TOF MS with an Agilent 1290 Infinity LC system. Gel were imaged on a Bio-Rad ChemiDoc™ Touch Gel Imager. ClAc-Tyr-CME (both enantiomers) were prepared as previously described.⁶⁷ The NNK library and test template were prepared by multi-step assembly PCR using synthetic DNA primers (IDT Europe, Belgium) and isolated by precipitation with 70% ethanol in 0.3 M NaCl, with final DNA and resulting peptide sequences detailed below.

Model template

DNA sequence:

TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGACCGAATA
CGAATACTTAGATTACGATTTTTTACCTGAAUGGGAACCTTTAGGTGCAGGCGCA
GGCGCATAGGACGGGGGGCGGAAA

Resulting peptide: MTEYEYLDYDFLPEWEPLGAGAGA

Amino-acylation of non-canonical amino acids

Testing acylation yield

Aminoacylation reactions (5 µL) were carried out with the following standard conditions: A mixture of 3 µL containing 125 pmol of µhelix and 125 pmol eFx in 0.25 M HEPES-KOH buffer (pH 7.5) was heated at 95 °C for 2 minutes and subsequently allowed to cool to room temperature for 5 minutes. Then, 1 µL 3 M MgCl₂ solution was added, followed by an incubation at room temperature for another 5 minutes. The mixture was then cooled on ice, followed by the addition of 1 µL 25 mM amino acid solution in DMSO. The reaction was left to incubate on ice for a varied amount of time (as detailed in **Fig. 5**). The reaction was quenched with 20 µL 0.3 M NaOAc buffer (pH 5.2). This was followed by the addition of 50 µL Ethanol to precipitate the product. This mixture was then centrifuged at 13000 rpm for 15 minutes to form a pellet of the precipitated product. The supernatant was removed and the pellet was washed with 30 µL 70% EtOH, after which it was again centrifuged for 5 minutes at 13000 rpm. The supernatant was removed and the pellet was allowed to air dry for 5 minutes, after which the pellet was dissolved in 16 µL RNA loading buffer. All aminoacylation samples were loaded onto a 20% PAGE gel and run in acidic buffer (sodium acetate pH 5.2) for 150 minutes at 120 V. Gels were subsequently incubated with TBE and then stained with SYBR Green II (Thermo-Fisher, USA) by incubation with 1X dye in TBE for 5 min. Product ratios were estimated by

densitometry in ImageJ.

Aminoacylation of tRNA

Aminoacylation reactions (5 μL) were carried out with the following standard conditions: A mixture of 125 pmol of tRNA and 125 pmol eFx (for CME activation) or dFx (for DBE activation), in 3 μL of 0.25 M HEPES-KOH buffer (pH 7.5) was heated at 95 $^{\circ}\text{C}$ for 2 min followed by an incubation at room temperature for 5 minutes. Then, 1 μL MgCl_2 solution (3 M) was added and the mixture was left to incubate another 5 min at room temperature. The mixture was then cooled on ice, followed by the addition of 1 μL amino acid solution in DMSO (5 mM final concentration). The reaction was left to incubate on ice for 2 hours. The reaction was then quenched with 20 μL NaOAc buffer (0.3 M, pH 5.2). This was followed by the addition of 50 μL ethanol to precipitate the product. This mixture was then centrifuged at 13000 rpm for 15 min to form a pellet of the precipitated product. The supernatant was removed and the pellet was vortexed with 40 μL NaOAc buffer (0.1 M, pH 5.2) in 70% EtOH then centrifuged for 10 minutes at 13000 rpm, and this process was repeated for two further washes. A final wash was then performed (without vortexing) using 30 μL 70% ethanol in water and followed by a 3 min centrifugation at 13000 rpm. Following removal of the supernatant, the pellet was left to air dry for 5 minutes and stored at -20 $^{\circ}\text{C}$.

In vitro translation and UPLC-MS test

Translation reactions were prepared on ice with the following composition:

20% (v/v) PURExpress solution Δ (aa, tRNA)

30% (v/v) PURExpress solution B

1 $\mu\text{g} \cdot \mu\text{L}^{-1}$ *E. coli* tRNA

25 μM acylated tRNA_{ini}

0.5 mM each 19 amino acid mix (-Trp)

10 $\text{ng} \cdot \mu\text{L}^{-1}$ DNA template

The mixture was adjusted to a final volume of 5 μL with milliQ water and incubated at 37 $^{\circ}\text{C}$ for half an hour. After translation, 15 μL of acetonitrile were added and the sample was centrifuged at 13000 rpm for 5 minutes before samples were analyzed by UPLC-MS. Extracted ion chromatograms were generated for product mass peaks in Agilent Masshunter Qualitative Navigator.

Synthesis of library and selection

Library construction and screening were performed largely as previously described.⁵⁶ A DNA template containing 15 NNK codons was synthesized from commercially

available primers by extension and PCR. The corresponding mRNA library was transcribed by T7 polymerase and covalently linked to Puromycin by T4 RNA ligase. The resulting library was translated in the PURExpress *in vitro* translation system, with Met and Ala withheld from the amino acid pool, following the manufacturer's protocol. Each reaction additionally contained 1.0 μ M puro-mRNA library, 62 μ M tRNA^{EnGlu}_{CGC}-Cys-diol, 31 μ M tRNA^{EnGlu}_{CAU}-5F-Phe, 31 μ M tRNA^{EnAsn}_{GGC}-Phe(*p*-CH₂SO₃H) and 31 μ M tRNA_{ini}-N-CIAC-*L*- (or *D*-) Tyr in two separate translation reactions. Each translation was performed at 5 μ L scale except for the 1st round of selection, which was carried out at 10 μ L scale. Following translation and release of mRNA-peptide from ribosomes with EDTA, cDNA was produced by reverse transcription using protoscript II reverse transcriptase. The cDNA-mRNA-peptide libraries were then panned against 1 pmol of Fc-E-selectin-6His (Biolegend) immobilized on Dynabeads™ Protein G or Dynabeads™ His tag magnetic beads (ThermoFisher) at 4 °C for 30 min. The beads were washed 3 times with PBS-T (10mM phosphate pH 7.4, 140 mM NaCl, 0.02% v/v Tween-20). The output was quantified by qPCR and cDNA was recovered by PCR using forward primer TAATACGACTCACTATAGGGTAACTTTAAGAAGGAGATATACATATG and reverse primer TTTCCGCCCCCGCTCCTAGCTGCCGCTGCCGCTGCCGCA. Following 5 rounds of selection, the final enriched DNA library was sequenced by MiSeq and results extracted by Python script and analyzed in CLC sequencer viewer.

Peptide synthesis.

Peptides were synthesized by automated Fmoc solid phase synthesis on a Biotage Syro II or CEM Liberty Blue Microwave synthesizer with tentagel S-RAM resin from Rapp polymere. The noncanonical amino acid Fmoc-Phe(*p*-CH₂SO₃H)-OH was synthesized as previously described⁶⁸ and was coupled without sidechain protection using HBTU activation. Peptide couplings were performed using Fmoc protected amino acids (4 equiv), HOBt (4 equiv), HBTU(4 equiv) and DIPEA (8 equiv) without heating in the Syro II, or Fmoc protected amino acids (5 equiv), DIC (5 equiv) and Oxyma (5 equiv) with microwave heating in the Liberty Blue, all in DMF. Upon completion of SPPS, each peptide was further coupled with chloroacetic acid under the same conditions but with no final deprotection, treated with a cleavage cocktail (TFA/H₂O/TIPS/TES 90:5:2.5:2.5), and cyclized as previously described.⁶⁹

Table 5. ESI-MS data for synthetic peptides (all as C-terminal amides)

Name	Sequence	Calculate mass [M+2H] ²⁺	Observed mass [M+2H] ²⁺
D1	yFVFWDWDGTGLRYSNCGGkG	1436.1	1436.2
D3	yLVLCDCSTYGZIVLHCGGkG	1371.0	1371.0
L1	YFZFWDWDGTGLRYSKCGGkG	1514.1	1514.0
L2	YFVFWDWDGTGLRYSKCGGkG	1443.1	1443.2
L4	YLDYSTRLRRZFYTWDCGGkG	1549.2	1549.1
L6	YRLVYDWF DG YLT YISCGGkG	1438.1	1438.2
D3AA	yLVLCDASTYGZIVLHAG	1038.5	1038.5
D3no Fam	yLVLCDCSTYGZIVLHCG	1070.5	1070.5
D3F	yLVLCDCSTYG FIVLHCG	1023.0	1023.0
D1Z	yFZFWDWDGTGLRYSNCG	1206.3	1206.5
L4F	YLDYSTRLRRFFYTWDCG	1202.1	1202.1
Linear +ve	DITWDQLWDLMK	781.9	781.9

Z=Phe(*p*-CH₂SO₃H), y=D-Tyr, k=Lys(5/6-Fam). All the peptides except linear +ve were cyclized by a thioether bridge from the N-terminal chloroacetic acid and cysteine.

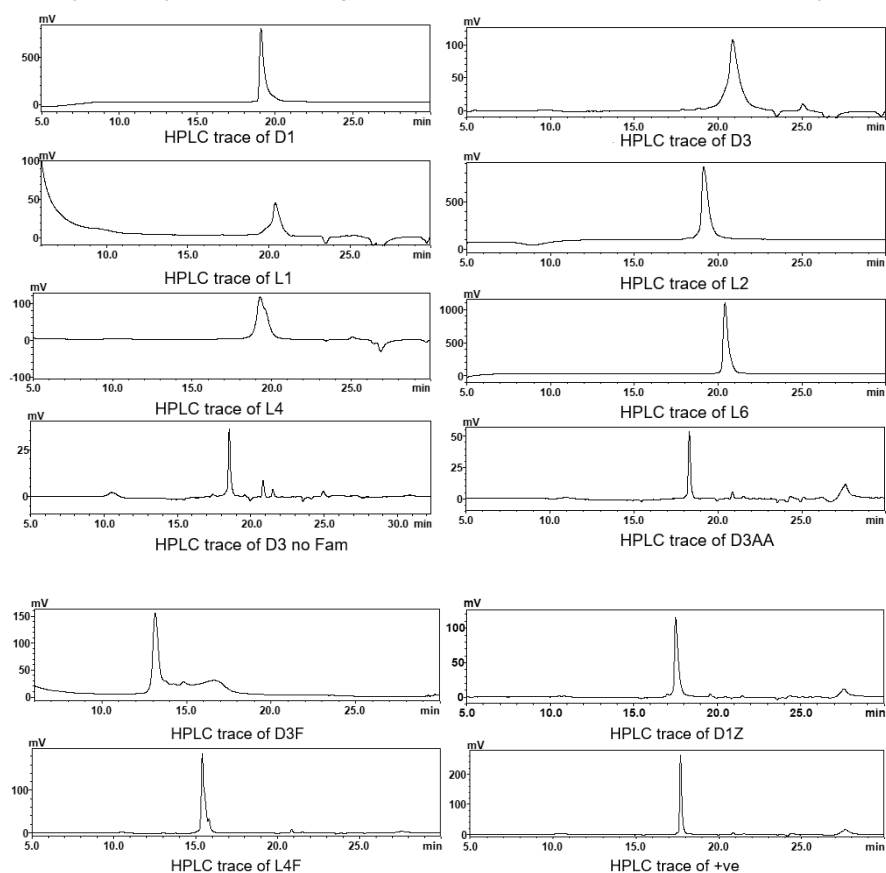
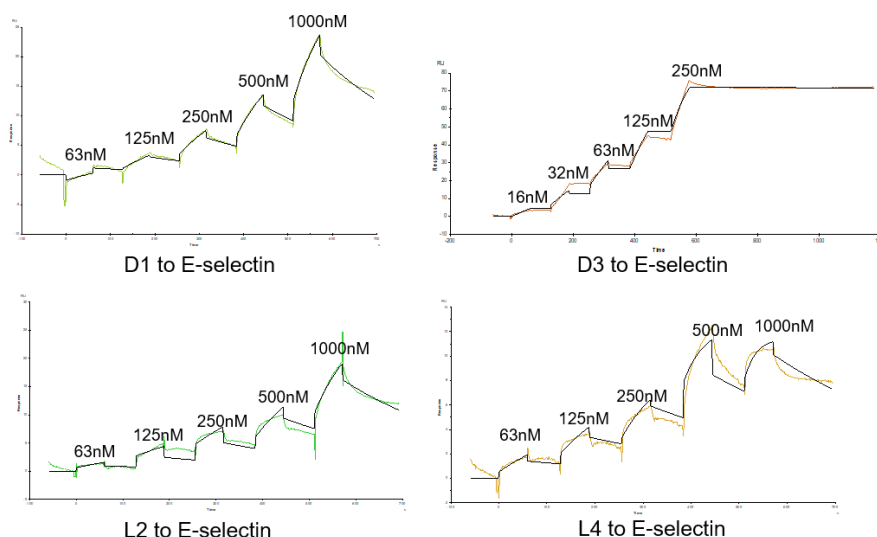
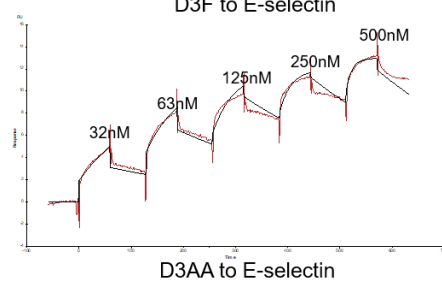
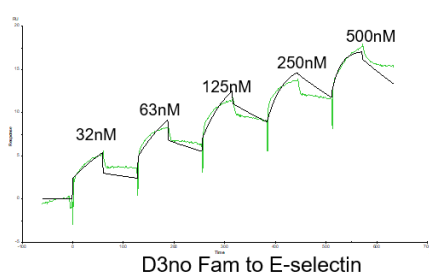
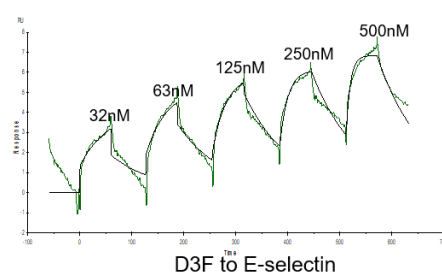
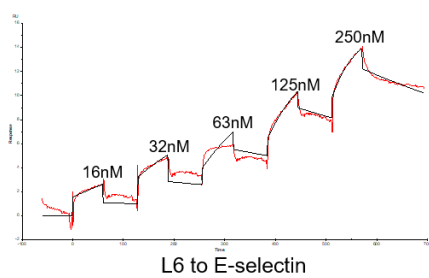


Figure 12. Analytical HPLC traces of purified synthetic peptides.

SPR characterization of peptide binding

A Biacore T100 (GE healthcare) was used for measurement of the binding kinetics of peptides with E selectin at 25 °C. The E-selectin was immobilized on a CM5 S sensor chip (Cytiva, GE healthcare Life Sciences) using the amine coupling kit (Cytiva, GE healthcare Life Sciences) as previously described.⁷⁰ Briefly, the CM5 chip was preactivated by 70 μ L of a mixture of EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) and NHS (N-hydroxysuccinimide). 70 μ L of protein solution (50 mM NaOAc pH 5.0, 50 ng. μ L⁻¹) was injected into channel 2 over 7 minutes. Channel 1 was used as a reference channel, with BSA being immobilized in a similar manner. After protein immobilization, the remaining active ester was blocked by injection of 70 μ L of ethanolamine (1 M, pH 8.5). The protein-immobilized chip was rinsed with HBS-T buffer (10 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 0.005 % Tween-20) until a stable baseline was acquired. To perform the binding assays, analytes were injected at a flow rate of 50 μ L.min⁻¹ for 1 min. The sensor chip surface was regenerated with glycine-HCl (10 mM, pH 2.0) for further experiments. All binding kinetic parameters were obtained from single-cycle kinetics with 5 different concentrations by a 1:1 fitting model from BIAevaluation software. For D3 the final dissociation step was extended (600 vs. 60 s) to attempt to better define the off-rate.





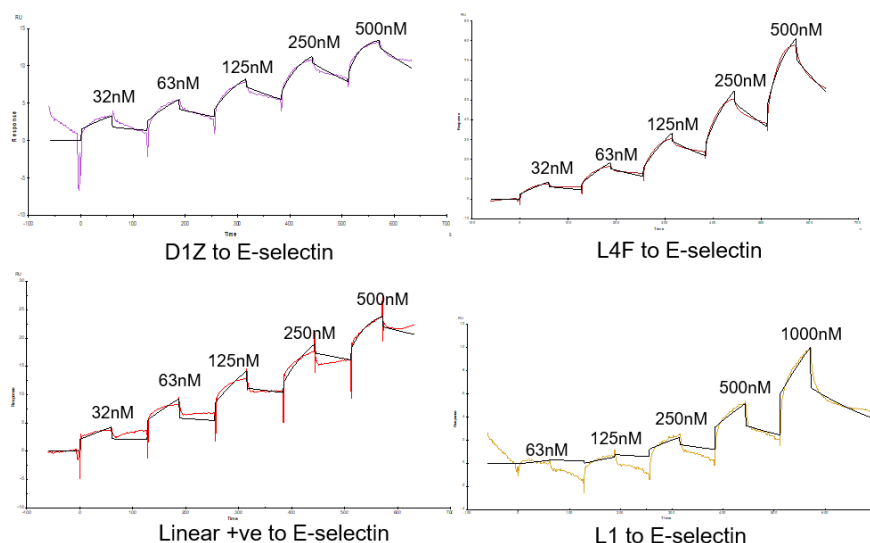
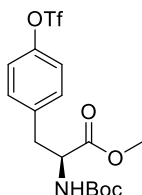


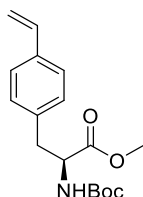
Figure 13. The sensorgrams for binding of peptides to E-selectin, with peptide concentrations as noted above the trace. In each case the colored line is actual data and the black line shows the fitting.

Synthetic procedures and compound characterization



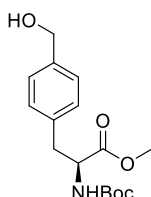
9: *N*-tert-butyloxycarbonyl-*O*-trifluoromethylsulfonyl-*L*-tyrosine methyl Ester

To an ice-cold solution of Boc-protected tyrosine methyl ester (2.52 g, 8.52 mmol,) in 8 mL dry DCM/Pyridine (4:1) was added Tf₂O (1.43 mL, 1 equiv) in dropwise fashion. The reaction mixture was stirred in an ice bath for 2.5 hr before the reaction was quenched with 50 mL water and an equal volume of DCM was added. The organic layer was separated and then washed with 50 mL 0.5 N NaOH, 50 mL water, 50 mL 1 N HCl (2 times) and again with water. The organic layer was dried over anhydrous Na₂SO₄, filtered, concentrated and partially purified by flash column chromatography (PE/EA 5:1) to obtain 3.19 g product as a white solid. This slightly impure product was used for next step without further purification.



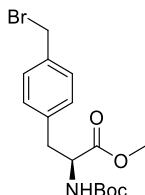
10: N-tert-butyloxycarbonyl-*p*-vinyl-*L*-phenylalanine methyl ester

To a dry and nitrogen-purged flask containing a solution of protected tyrosine compound **9** (1.28 g, 3 mmol) in 9 mL DMF was added LiCl (254 mg, 2 equiv, 6 mmol), Pd(PPh₃)Cl₂ (105 mg, 0.05 equiv, 0.15 mmol) and tributyl vinyl tin (1.05 g, 1.1 equiv, 3.3 mmol). The reaction mixture was stirred at 90 °C for 16 hrs, then diluted with 50 mL water and extracted 3 times with 50 mL EA. The combined organic layer was washed with 150 mL brine, dried with anhydrous Na₂SO₄, filtered, and purified by flash column chromatography (PE/EA 4:1) to get a colorless oil (0.79 g, yield 85% over two steps). NMR data match previous reports.⁶⁸



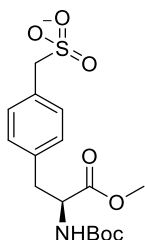
11: N-tert-butyloxycarbonyl-(*p*-hydroxymethyl)-*L*-phenylalanine methyl Ester

Through a solution of compound **10** (200 mg, 0.65 mmol) in 20 mL MeOH was bubbled ozone at -78 °C. The mixture was stirred until the solution changed to light blue. After that, nitrogen was bubbled through the solution for 10 min and then 5 mL Me₂S was added. The mixture was allowed to react another 1 hr at -78 °C, then NaBH₄ (37.8 mg, 5 equiv, 3.22 mmol) was added and allowed to react at -78 °C for 10 min. The reaction mixture was then allowed to warm up to room temperature slowly and stirred for another 15 min. Saturated NH₄Cl was added to quench the reaction, then 50 mL EA was added to extract the product, which was then dried over anhydrous Na₂SO₄, filtered, and concentrated under vacuum. The product was purified by flash column chromatography (PE/EA 3:1) to afford a light yellow solid (130 mg, yield 60%). NMR data match previous reports.⁶⁸



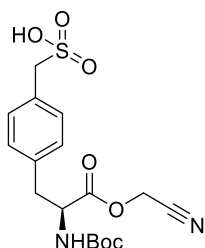
12: (p-bromomethyl)-L-phenylalanine methyl ester

Compound **11** (110 mg, 0.35 mmol) was dissolved in 3 mL dry THF at 0 °C. Triphenylphosphine (279 mg, 3 equiv, 1.05 mmol) was then added and after 5 min this was followed by NBS (190 mg, 3 equiv, 1.05 mmol). The resulting solution was then stirred at room temperature for 16 hrs. After dilution of the mixture with 10 mL Et₂O and washing with 10 mL saturated Na₂S₂O₃ and water, the organic layer was dried with anhydrous Na₂SO₄, filtered and concentrated, and the product purified by flash column chromatography (PE/EA 5:1) to get a white solid (112 mg, yield 83%). NMR data match previous reports.⁶⁸



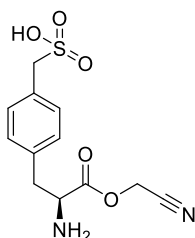
13: N-tert-butyloxycarbonyl-(p-sulfomethyl)-L-phenylalanine methyl ester

To a solution of compound **12** (12 mg, 0.03 mmol) in 5 mL dioxane was added 2 mL of a 1 N solution of sodium sulfite in water. The mixture was allowed to stir under reflux for 1hr. The solvent was then removed under vacuum and the product was purified by flash column chromatography (DCM/MeOH 10:1) to obtain a white foam (14 mg, yield=95%). NMR data match previous reports.⁶⁸



14: *N*-tert-butyloxycarbonyl-(*p*-sulfomethyl)-*L*-phenylalanine acetonitrile ester

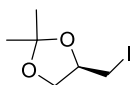
To a solution of compound **13** (14 mg, 0.038 mmol) in 5 mL THF/H₂O (4:1), was added LiOH (5 mg, 3 equiv, 0.11 mmol). The reaction mixture was stirred at room temperature for 1 hr before it was neutralized with Amberlite IR-120H acidic resin, filtered, and the solvent removed under vacuum. The crude product was dissolved in 2.5 mL DMF, and then DMAP (0.5 mg, 4 μmol, 0.1 equiv), TEA (15 μL, 0.1 mmol, 2.5 equiv), and chloroacetonitrile (2.5 mL) were added. The reaction mixture was stirred at room temperature for 5 hrs and the solvent subsequently removed under vacuum. The resulting crude product was purified by flash column chromatography (DCM/MeOH 5:1) to obtain a light yellow foam (5 mg, yield 33%). ¹H NMR (400 MHz, D₂O) δ 7.40 (d, *J* = 7.8 Hz, 2H), 7.30 (d, *J* = 7.7 Hz, 2H), 4.97 (s, 2H), 4.57 (dd, *J* = 9.2, 5.5 Hz, 1H), 4.17 (s, 2H), 3.14 (m, 2H), 1.37 (s, 3H) ppm. ¹³C NMR (101 MHz, D₂O) δ 171.99, 130.34, 130.27, 129.14, 114.98, 56.24, 54.56, 49.42, 36.03, 27.22 ppm. HRMS (ESI⁺): [M+Na]⁺ Calculated for [C₁₇H₂₂N₂O₇NaS]⁺: required 421.1040 m/z, found: 421.1141 m/z.



8: (*p*-sulfomethyl)-*L*-phenylalanine acetonitrile ester

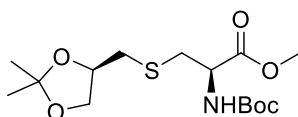
Compound **14** (10 mg, 0.025 mmol) was dissolved in 1 mL formic acid and the reaction mixture was stirred at room temperature for 1 hr. The formic acid was subsequently removed under vacuum and the product purified by flash column chromatography to afford a white foam (6 mg, yield 80%). ¹H NMR (400 MHz, D₂O) δ 8.42 (s, 2H), 7.48 (d, *J* = 8.0 Hz, 2H), 7.36 (d, *J* = 8.0 Hz, 2H), 5.12 (d, *J* = 1.2 Hz, 2H), 4.63 (dd, *J* = 7.9, 5.8 Hz, 1H), 4.22 (s, 2H), 3.56 – 3.45 (m, 1H), 3.33 (m, 1H) ppm. ¹³C NMR (101 MHz, D₂O) δ 168.88, 132.99, 131.67, 131.11, 130.89, 129.52,

129.44, 56.43, 53.67, 50.40, 34.99 ppm. HRMS (ESI⁺): [M+H]⁺ Calculated for [C₁₂H₁₅N₂O₅S]⁺: required 299.0696 m/z, found: 299.0817 m/z.



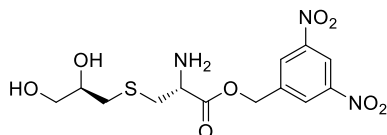
15: (R)-4-(iodomethyl)-2,2-dimethyl-1,3-dioxolane.

To a solution of (R)-(-)-2,2-Dimethyl-1,3-dioxolane-4-methanol (3.2 g, 24.2 mmol) in toluene (40 mL) was added PPh₃ (7.6 g, 29.1 mmol, 1.2 equiv), imidazole (4.9 g, 72.6 mmol, 3 equiv) and iodine (8.0 g, 31.5 mmol, 1.3 equiv). The reaction mixture was stirred at 90 °C for 3 hours and then the solvent was removed under vacuum. The resulting residue was dissolved in DCM, washed with saturated Na₂S₂O₃ then brine and dried over Na₂SO₄ and filtered. The solvent was again removed under vacuum and the oil was purified by flash column chromatography to afford compound **15** as a colorless liquid (4.9 g, yield 84%). NMR data match previous reports.⁷¹



16: Methyl N-(tert-butoxycarbonyl)-S-(((R)-2,2-dimethyl-1,3-dioxolan-4-yl) methyl)-L-cysteinate

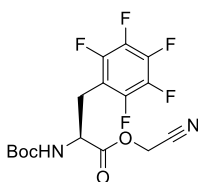
To a solution of compound **15** (402 mg, 1.7 mmol) in anhydrous DMF was added DIPEA (643 mg, 5.0 mmol, 2.9 equiv), followed by compound **3** (300 mg, 1.28 mmol, 0.75 equiv). The reaction mixture was stirred at 85 °C for 4 hours, then the solvent was removed under vacuum and the resulting residue purified by flash column chromatography (PE/EA 3:1) to give a colorless oil (318 mg, yield 71%). NMR data match previous reports.⁷¹



18: 3,5-dinitrobenzyl S-((R)-2,3-dihydroxypropyl)-L-cysteinate

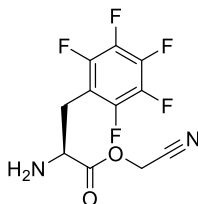
To a solution of compound **16** (200 mg, 0.57 mmol) in 5 mL THF/H₂O (4:1) solvent was added LiOH (21 mg, 3.5 mmol, 6 equiv) and the reaction mixture was stirred at room temperature. After 3 hours the THF was removed under vacuum and the remaining water was diluted with a further 10 mL before the reaction was carefully

quenched to pH 2-3 with saturated citric acid solution. This aqueous fraction was extracted 3 times with 15 mL DCM, dried over Na₂SO₄, filtered and concentrated under vacuum. The crude product was subsequently redissolved in 10 mL anhydrous DCM and to this was added 3,5-dinitrobenzyl chloride (185 mg, 0.85 mmol, 1.5 equiv), DIPEA (146 mg, 1.14 mmol, 2 equiv) and DMAP (3 mg, 0.057 mmol, 0.1 equiv). The reaction was allowed to proceed with stirring at room temperature overnight. The solvent was then removed under vacuum and the resulting solid was redissolved in formic acid. After 3 hours the deprotected product was purified by flash column chromatography (DCM/MeOH 10:1) to afford a colorless oil (6 mg, yield 3%). ¹H NMR (600 MHz, CDCl₃) δ 9.03 (t, *J* = 2.1 Hz, 1H), 8.60 (s, 2H), 5.48 – 5.36 (m, 2H), 5.03 (tt, *J* = 8.0, 4.7 Hz, 1H), 4.25 (dd, *J* = 11.5, 4.0 Hz, 1H), 4.18 (dd, *J* = 11.4, 6.4 Hz, 1H), 3.21-3.05 (m, 2H), 2.79 (dd, *J* = 14.1, 4.4 Hz, 1H), 2.72 (dd, *J* = 14.1, 7.3 Hz, 1H) ppm. ¹³C NMR (151 MHz, CDCl₃) δ 169.85, 148.74, 139.37, 128.07, 118.90, 68.94, 66.13, 65.12, 51.06, 36.50, 35.03 ppm. HRMS (ESI⁺): [M+H]⁺ Calculated for [C₁₃H₁₈N₃O₈S]⁺: required 376.0809 m/z, found: 376.0799 m/z.



19: N-Boc-pentafluoro-L-phenylalanine-cyanomethyl ester

N-Boc-pentafluoro-L-phenylalanine (167 mg, 0.47 mmol) was dissolved in 2 mL chloroacetonitrile and to this was added TEA (0.2 mL, 1.4 mmol, 3 equiv). The reaction mixture was stirred at room temperature overnight and the solvent was subsequently removed under vacuum before the resulting residue was purified by flash column chromatography (PE/EA 1:1) to yield a colorless oil (180 mg, yield 97%) NMR data match previous reports.⁵⁷



20: Pentafluoro-L-phenylalanine-cyanomethyl ester

N-Boc-pentafluoro-L-phenylalanine-cyanomethyl ester (67 mg, 0.17 mmol) was dissolved in 1 mL pure formic acid and the solution stirred at room temperature for

3 hrs. The solvent was subsequently removed under vacuum and the residue purified by flash column chromatography (PE/EA 1:1) to afford a colorless oil (38 mg, yield 78%). NMR data match previous reports.⁵⁷

3.5 References

1. McEver, R. P., Selectins: initiators of leucocyte adhesion and signalling at the vascular wall. *Cardiovasc. Res.* **2015**, *107*, 331-339.
2. McEver, R. P.; Zhu, C., Rolling Cell Adhesion. *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 363-396.
3. Gout, S.; Tremblay, P.-L.; Huot, J., Selectins and selectin ligands in extravasation of cancer cells and organ selectivity of metastasis. *Clin. Exp. Metastasis* **2008**, *25*, 335-344.
4. Ley, K., The role of selectins in inflammation and disease. *Trends Mol. Med.* **2003**, *9*, 263-268.
5. Rees, D. C.; Williams, T. N.; Gladwin, M. T., Sickle-cell disease. *The Lancet* **2010**, *376*, 2018-2031.
6. Kuuliala, A.; Eberhardt, K.; Takala, A.; Kautiainen, H.; Repo, H.; Leirisalo-Repo, M., Circulating soluble E-selectin in early rheumatoid arthritis: A prospective five year study. *Ann. Rheum. Dis.* **2002**, *61*, 242-246.
7. Silva, M.; Videira, P. A.; Sackstein, R., E-Selectin Ligands in the Human Mononuclear Phagocyte System: Implications for Infection, Inflammation, and Immunotherapy. *Front. Immunol.* **2018**, *8*, 1878.
8. Phillipson, M.; Kubes, P., The neutrophil in vascular inflammation. *Nat. Med.* **2011**, *17*, 1381-1390.
9. Ley, K.; Laudanna, C.; Cybulsky, M. I.; Nourshargh, S., Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nat. Rev. Immunol.* **2007**, *7*, 678-689.
10. Sadik, C. D.; Kim, N. D.; Luster, A. D., Neutrophils cascading their way to inflammation. *Trends Immunol.* **2011**, *32*, 452-460.
11. Luo, B. H.; Carman, C. V.; Springer, T. A., Structural basis of integrin regulation and signaling. *Annu. Rev. Immunol.* **2007**, *25*, 619-647.
12. Phillipson, M.; Heit, B.; Colarusso, P.; Liu, L.; Ballantyne, C. M.; Kubes, P., Intraluminal crawling of neutrophils to emigration sites: a molecularly distinct process from adhesion in the recruitment cascade. *The Journal of Experimental Medicine* **2006**, *203*, 2569.
13. Borsig, L.; Wong, R.; Hynes, R. O.; Varki, N. M.; Varki, A., Synergistic effects of L- and P-selectin in facilitating tumor metastasis can involve non-mucin ligands and

implicate leukocytes as enhancers of metastasis. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2193-2198

14. Läubli, H.; Stevenson, J. L.; Varki, A.; Varki, N. M.; Borsig, L., L-Selectin Facilitation of Metastasis Involves Temporal Induction of α -Fut7-Dependent Ligands at Sites of Tumor Cell Arrest. *Cancer Res.* **2006**, *66*, 1536.

15. Häuselmann, I.; Roblek, M.; Protsyuk, D.; Huck, V.; Knopfova, L.; Grässle, S.; Bauer, A. T.; Schneider, S. W.; Borsig, L., Monocyte Induction of E-Selectin-Mediated Endothelial Activation Releases VE-Cadherin Junctions to Promote Tumor Cell Extravasation in the Metastasis Cascade. *Cancer Res.* **2016**, *76*, 5302.

16. Carvalho, S.; Catarino, T. A.; Dias, A. M.; Kato, M.; Almeida, A.; Hessling, B.; Figueiredo, J.; Gärtner, F.; Sanches, J. M.; Ruppert, T.; Miyoshi, E.; Pierce, M.; Carneiro, F.; Kolarich, D.; Seruca, R.; Yamaguchi, Y.; Taniguchi, N.; Reis, C. A.; Pinho, S. S., Preventing E-cadherin aberrant N-glycosylation at Asn-554 improves its critical function in gastric cancer. *Oncogene* **2016**, *35*, 1619-1631.

17. Stowell, S. R.; Ju, T.; Cummings, R. D., Protein glycosylation in cancer. *Annu. Rev. Pathol.* **2015**, *10*, 473-510.

18. Witz, I. P., The selectin-selectin ligand axis in tumor progression. *Cancer Metastasis Rev.* **2008**, *27*, 19-30.

19. Läubli, H.; Borsig, L., Selectins promote tumor metastasis. *Semin. Cancer Biol.* **2010**, *20*, 169-177.

20. Erpenbeck, L.; Schon, M. P., Deadly allies: the fatal interplay between platelets and metastasizing cancer cells. *Blood* **2010**, *115*, 3427-3436.

21. Zhong, L.; Simoneau, B.; Tremblay, P.-L.; Gout, S.; Simard, M. J.; Huot, J., E-Selectin-Mediated Adhesion and Extravasation in Cancer. In *Encyclopedia of Cancer*, Schwab, M., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2017; pp 1618-1624.

22. Kang, S.-A.; Blache, C. A.; Bajana, S.; Hasan, N.; Kamal, M.; Morita, Y.; Gupta, V.; Tsolmon, B.; Suh, K. S.; Gorenstein, D. G.; Razaq, W.; Rui, H.; Tanaka, T., The effect of soluble E-selectin on tumor progression and metastasis. *BMC Cancer* **2016**, *16*, 331.

23. Price, T. T.; Burness, M. L.; Sivan, A.; Warner, M. J.; Cheng, R.; Lee, C. H.; Olivere, L.; Comatas, K.; Magnani, J.; Kim Lyerly, H.; Cheng, Q.; McCall, C. M.; Sipkins, D. A., Dormant breast cancer micrometastases reside in specific bone marrow niches that regulate their transit to and from bone. *Sci. Transl. Med.* **2016**, *8*, 340ra73.

24. Barthel, S. R.; Gavino, J. D.; Descheny, L.; Dimitroff, C. J., Targeting selectins and selectin ligands in inflammation and cancer. *Expert Opin. Ther. Targets* **2007**, *11*, 1473-1491.

25. Natoni, A.; Macauley, M. S.; O'Dwyer, M. E., Targeting Selectins and Their Ligands in Cancer. *Front. Oncol.* **2016**, *6*, 93.
26. Zarbock, A.; Ley, K.; McEver, R. P.; Hidalgo, A., Leukocyte ligands for endothelial selectins: specialized glycoconjugates that mediate rolling and signaling under flow. *Blood* **2011**, *118*, 6743-6751.
27. Fuhlbrigge, R. C.; Kieffer, J. D.; Armerding, D.; Kupper, T. S., Cutaneous lymphocyte antigen is a specialized form of PSGL-1 expressed on skin-homing T cells. *Nature* **1997**, *389*, 978-981.
28. Dimitroff, C. J.; Lee, J. Y.; Rafii, S.; Fuhlbrigge, R. C.; Sackstein, R., Cd44 Is a Major E-Selectin Ligand on Human Hematopoietic Progenitor Cells. *J. Cell Biol.* **2001**, *153*, 1277-1286.
29. Steegmaler, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H. P.; Kleuser, B.; Vestweber, D., The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth factor. *Nature* **1995**, *373*, 615-620.
30. Alon, R.; Feizi, T.; Yuen, C. T.; Fuhlbrigge, R. C.; Springer, T. A., Glycolipid ligands for selectins support leukocyte tethering and rolling under physiologic flow conditions. *J. Immunol.* **1995**, *154*, 5356-5366.
31. Mehta, P.; Cummings, R. D.; McEver, R. P., Affinity and Kinetic Analysis of P-selectin Binding to P-selectin Glycoprotein Ligand-1. *J. Biol. Chem.* **1998**, *273*, 32506-32513.
32. Somers, W. S.; Tang, J.; Shaw, G. D.; Camphausen, R. T., Insights into the Molecular Basis of Leukocyte Tethering and Rolling Revealed by Structures of P- and E-Selectin Bound to SLeX and PSGL-1. *Cell* **2000**, *103*, 467-479.
33. Carlow, D. A.; Gossens, K.; Naus, S.; Veerman, K. M.; Seo, W.; Ziltener, H. J., PSGL-1 function in immunity and steady state homeostasis. *Immunol. Rev.* **2009**, *230*, 75-96.
34. Sackstein, R., Glycoengineering of HCELL, the Human Bone Marrow Homing Receptor: Sweetly Programming Cell Migration. *Ann. Biomed. Eng.* **2012**, *40*, 766-776.
35. Yuen, C. T.; Lawson, A. M.; Chai, W.; Larkin, M.; Stoll, M. S.; Stuart, A. C.; Sullivan, F. X.; Ahern, T. J.; Feizi, T., Novel sulfated ligands for the cell adhesion molecule E-selectin revealed by the neoglycolipid technology among O-linked oligosaccharides on an ovarian cystadenoma glycoprotein. *Biochemistry* **1992**, *31*, 9126-9131.
36. Fuhlbrigge, R. C.; Weishaupt, C., Adhesion molecules in cutaneous immunity. *Semin. Immunopathol.* **2007**, *29*, 45-57.
37. FDA approves first targeted therapy to treat patients with painful complication of

sickle cell disease.

38. Drug Approval Package: ADAKVEO (crizanlizumab-tmca).
39. Kerr, K.; Auger, W. R.; Marsh, J. J.; Comito, R. M.; Fedullo, R. L.; Smits, G. J.; Kapelanski, D. P.; Fedullo, P.; Channick, R. N.; Jamieson, S.; Moser, K. M., The Use of Cylexin (CY-1503) in Prevention of Reperfusion Lung Injury in Patients Undergoing Pulmonary Thromboendarterectomy. *Am. J. Respir. Crit. Care Med.* **2000**, *162*, 14-20.
40. Safety Of Rivipansel (GMI-1070) In The Treatment Of One or More Vaso-occlusive Crises In Hospitalized Subjects With Sickle Cell Disease. <https://ClinicalTrials.gov/show/NCT02433158>.
41. Study of Dose Confirmation and Safety of Crizanlizumab in Pediatric Sickle Cell Disease Patients. <https://ClinicalTrials.gov/show/NCT03474965>.
42. Pharmacokinetics and Pharmacodynamics Study of SEG101 (Crizanlizumab) in Sickle Cell Disease (SCD) Patients With Vaso- Occlusive Crisis (VOC). <https://ClinicalTrials.gov/show/NCT03264989>.
43. Study of Two Doses of Crizanlizumab Versus Placebo in Adolescent and Adult Sickle Cell Disease Patients. <https://ClinicalTrials.gov/show/NCT03814746>.
44. A Study to Evaluate the Safety and Efficacy of Crizanlizumab in Sickle Cell Disease Related Priapism. <https://ClinicalTrials.gov/show/NCT03938454>.
45. Study Exploring the Effect of Crizanlizumab on Kidney Function in Patients With Chronic Kidney Disease Caused by Sickle Cell Disease. <https://ClinicalTrials.gov/show/NCT04053764>.
46. MAP to Provide Access to Crizanlizumab, for Sickle Cell Disease Patients. <https://ClinicalTrials.gov/show/NCT03720626>.
47. Platform Study of Novel Ruxolitinib Combinations in Myelofibrosis Patients. <https://ClinicalTrials.gov/show/NCT04097821>.
48. Study Evaluating PSI-697 in Patients With Scleritis. <https://ClinicalTrials.gov/show/NCT00367692>.
49. A Study of Single Dose Inclacumab in Japanese Healthy Volunteers Compared to Caucasian Healthy Volunteers. <https://ClinicalTrials.gov/show/NCT01815827>.
50. Passioura, T.; Suga, H., A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun.* **2017**, *53*, 1931-1940.
51. Wesener, D. A.; Wangkanont, K.; McBride, R.; Song, X.; Kraft, M. B.; Hodges, H. L.; Zarling, L. C.; Splain, R. A.; Smith, D. F.; Cummings, R. D.; Paulson, J. C.; Forest, K. T.; Kiessling, L. L., Recognition of microbial glycans by human intelectin-1. *Nat. Struct. Mol. Biol.* **2015**, *22*, 603-610.
52. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N., Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem.*

Soc. **2015**, *137*, 15152-15160.

53. Fernández-Alonso, M. d. C.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G., Molecular Recognition of Saccharides by Proteins. Insights on the Origin of the Carbohydrate–Aromatic Interactions. *J. Am. Chem. Soc.* **2005**, *127*, 7379-7386.

54. Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J., Carbohydrate–Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46*, 946-954.

55. Martinez, M.; Joffraud, M.; Giraud, S.; Baisse, B.; Bernimoulin, M. P.; Schapira, M.; Spertini, O., Regulation of PSGL-1 Interactions with L-selectin, P-selectin, and E-selectin: ROLE OF HUMAN FUCOSYLTRANSFERASE-IV AND -VII. *J. Biol. Chem.* **2005**, *280*, 5378-5390.

56. Johansen-Leete, J.; Passioura, T.; Foster, S. R.; Bhusal, R. P.; Ford, D. J.; Liu, M.; Jongkees, S. A. K.; Suga, H.; Stone, M. J.; Payne, R. J., Discovery of Potent Cyclic Sulfopeptide Chemokine Inhibitors via Reprogrammed Genetic Code mRNA Display. *J. Am. Chem. Soc.* **2020**, *142*, 9141-9146.

57. Passioura, T.; Liu, W.; Dunkelmann, D.; Higuchi, T.; Suga, H., Display Selection of Exotic Macrocyclic Peptides Expressed under a Radically Reprogrammed 23 Amino Acid Genetic Code. *J. Am. Chem. Soc.* **2018**, *140*, 11551-11555.

58. Goto, Y.; Katoh, T.; Suga, H., Flexizymes for genetic code reprogramming. *Nat. Protoc.* **2011**, *6*, 779-790.

59. McClain, W. H., Rules that Govern tRNA Identity in Protein Synthesis. *J. Mol. Biol.* **1993**, *234*, 257-280.

60. Weijland, A.; Harmark, K.; Cool, R. H.; Anborgh, P. H.; Parmeggiani, A., Elongation factor Tu: a molecular switch in protein biosynthesis. *Mol. Microbiol.* **1992**, *6*, 683-688.

61. Dale, T.; Uhlenbeck, O. C., Amino acid specificity in translation. *Trends Biochem. Sci* **2005**, *30*, 659-665.

62. Pathare, B.; Tambe, V.; Patil, V., A review on various analytical methods used in determination of dissociation constant. *Int. J. Pharm. Pharm. Sci.* **2014**, *6*, 26-34.

63. Freyer, M. W.; Lewis, E. A., Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. In *Methods Cell Biol.*, Academic Press 2008; Vol. 84, pp 79-113.

64. Jameson, D. M.; Ross, J. A., Fluorescence Polarization/Anisotropy in Diagnostics and Imaging. *Chem. Rev.* **2010**, *110*, 2685-2708.

65. Nguyen, H. H.; Park, J.; Kang, S.; Kim, M., Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications. *Sensors* **2015**, *15*.

66. Martens, C. L.; Cwirla, S. E.; Lee, R. Y. W.; Whitehorn, E.; Chen, E. Y. F.; Bakker, A.; Martin, E. L.; Wagstrom, C.; Gopalan, P.; Smith, C. W.; Tate, E.; Koller, K. J.;

- Schatz, P. J.; Dower, W. J.; Barrett, R. W., Peptides Which Bind to E-selectin and Block Neutrophil Adhesion. *J. Biol. Chem.* **1995**, *270*, 21129-21136.
67. Kawakami, T.; Murakami, H.; Suga, H., Ribosomal Synthesis of Polypeptoids and Peptoid–Peptide Hybrids. *J. Am. Chem. Soc.* **2008**, *130*, 16861-16863.
68. Roosenburg, S.; Laverman, P.; Joosten, L.; Eek, A.; Oyen, W. J. G.; de Jong, M.; Rutjes, F. P. J. T.; van Delft, F. L.; Boerman, O. C., Stabilized ¹¹¹In-Labeled sCCK8 Analogues for Targeting CCK2-Receptor Positive Tumors: Synthesis and Evaluation. *Bioconjugate Chem.* **2010**, *21*, 663-670.
69. Goldbach, L.; Vermeulen, B. J. A.; Caner, S.; Liu, M.; Tysoe, C.; van Gijzel, L.; Yoshisada, R.; Trellet, M.; van Ingen, H.; Brayer, G. D.; Bonvin, A. M. J. J.; Jongkees, S. A. K., Folding Then Binding vs Folding Through Binding in Macrocyclic Peptide Inhibitors of Human Pancreatic α -Amylase. *ACS Chem. Biol.* **2019**, *14*, 1751-1759.
70. Myung, J. H.; Gajjar, K. A.; Pearson, R. M.; Launiere, C. A.; Eddington, D. T.; Hong, S., Direct Measurements on CD24-Mediated Rolling of Human Breast Cancer MCF-7 Cells on E-Selectin. *Anal. Chem.* **2011**, *83*, 1078-1083.
71. Wu, W.; Li, R.; Malladi, S. S.; Warshakoon, H. J.; Kimbrell, M. R.; Amolins, M. W.; Ukani, R.; Datta, A.; David, S. A., Structure–Activity Relationships in Toll-like Receptor-2 Agonistic Diacylthioglycerol Lipopeptides. *J. Med. Chem.* **2010**, *53*, 3198-3213.

Chapter 4

CLIPS/CuAAc utilization in mRNA display for novel peptide scaffolding

Abstract

Peptides, being able to form large contact surfaces, are promising ligands for difficult-to-target proteins. However, their cell permeability and stability are long-lasting problems for linear peptides applied *in vivo*. Cyclization of peptides is one solution for this problem, and diverse methods have been developed to achieve this based on chemoselective reactions. Among these approaches, a combined CLIPS/CuAAC cyclization scaffold can produce a tricyclic geometry which promises unique backbone conformational effects that could allow it to provide improved hits. In this work we used LC-MS to validate the application of this scaffolding reaction on reprogrammed peptides produced by *in vitro* translation. Further verification in an mRNA display setting was achieved by use of biotinylated biorthogonal probes with complementary reactivity. To conclusively demonstrate that this combination of CLIPS/CuAAC and mRNA display provides a robust and useful strategy for the discovery of tricyclic peptides, we chose 4 different protein targets and two different scaffold cores with which to test our methodology. All selections showed signs of enrichment, with two being particularly strong. Sequencing showed enrichment of a known binding motif for one target and unique hits per scaffold/protein combination, together strongly indicating a successful enrichment of unique hits that either engage the target directly with the scaffold or achieve a unique conformation determined by the scaffold.

4.1 Introduction

Compared to small molecule inhibitors of protein surfaces or protein-protein interaction interfaces, peptides have a lot of advantages such as large interaction surfaces, plasticity, and selectivity.¹ Moreover, peptides can be produced by both chemical synthesis or biological expression, offering increased flexibility, and they are cheaper to produce and easier to store than antibodies. Display technologies, such as phage display², yeast display³ and mRNA display⁴, have made it increasingly straightforward and fast to find new bioactive peptides. However, linear peptides still have a crucial limitation because of low stability *in vivo*. There are several ways to remove this obstacle, such as *N*-terminal or *C*-terminal modifications⁵, amide bond variation⁶, unnatural amino acid addition⁷ and cyclization⁸⁻¹⁰, among which the most accessible methodology directly in a display setting is macrocyclization. Compared to linear peptides, cyclic peptides have better metabolic stability, selectivity and more constrained conformations (**Fig 1**). Their unique structures make cyclic peptide initial leads more promising. Even though most cyclic peptides remain not “drug-like” according to the rules suggested by Veber,¹¹ to date more than 60 peptides have been approved for clinical use. It was estimated that the yearly market for peptides is around \$25.4 billion, and this number likely will continue to rise.¹²

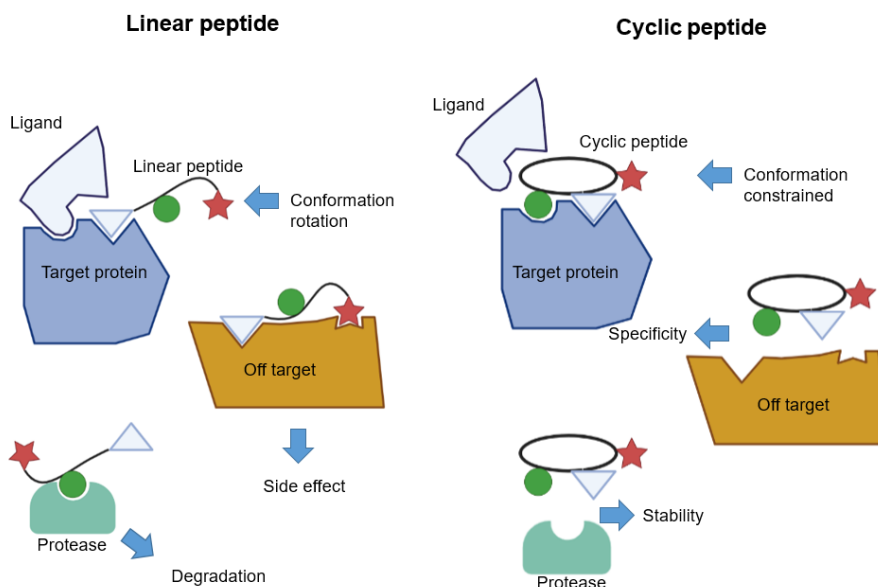


Figure 1. Advantages of cyclic peptides (right panel) compared to linear peptides (left panel). Due to their conformational flexibility, linear peptides often reach equilibrium with natural ligands and do not fully displace them. What's more, linear peptides can have side effects by binding to alternate targets in these other conformations. Finally,

linear peptides can be decomposed more easily by proteases because of their backbone accessibility. As a comparison, cyclic peptides have constrained conformations, offering better target engagement, specificity and resistance to proteases.

Over the last decades, many cyclization methods have been developed. In this section, we mainly focus on the approaches that have proven suitable for use in more biologically complex settings (**Fig 2**). These cyclization methods are able to be combined with mRNA display or phage display to construct unique cyclic peptide libraries of extremely high diversity. Of these, the simplest and most convenient method is by oxidizing cysteine residues to form disulfide bonds.¹³ However, because of stability issues, this method has somewhat fallen out of favor. As an alternative, thioether cyclization was developed and used in mRNA display by Suga et al.¹⁴ In this, an α -chloroacetylated amino acid is incorporated *N*-terminally to form a thioether bond with a downstream cysteine in the library. This reaction is highly effective and typically gives high conversion spontaneously following translation. Besides the chloroacetyl group, dibromoxylens are also used to cyclize two cysteines to produce thioether-based cross-linking, although this does not proceed spontaneously.¹⁵ Similar scaffolds that cross-link cysteine based on conjugate addition have also been developed, with the scaffold choice influencing peptide conformation.¹⁶⁻¹⁸

NHS chemistry can also be used for conjugation of lysine side-chains and/or the *N*-terminus to construct a macrocycle.¹⁹ This reaction is also compatible with mRNA display, but can be difficult to apply cleanly because of the abundance of amines in biomolecules and the associated limitations on lysine placement in library design, as well as the instability of the reagent in aqueous medium. The CuAAC reaction has been utilized to construct cyclic peptides and carry out selections, with the intramolecular CuAAC reaction being favored over intermolecular reaction.²⁰ Cyclization options are further extended by the $K_3Fe(CN)_6$ catalyzed reaction between benzylamine and 5-hydroxyindole to form a fluorescent connection with an unusual aromatic core that may offer unusual effects on conformation.²¹ While this reaction is reported to be fast, being finished within five minutes, this has yet to be applied in a published selection.

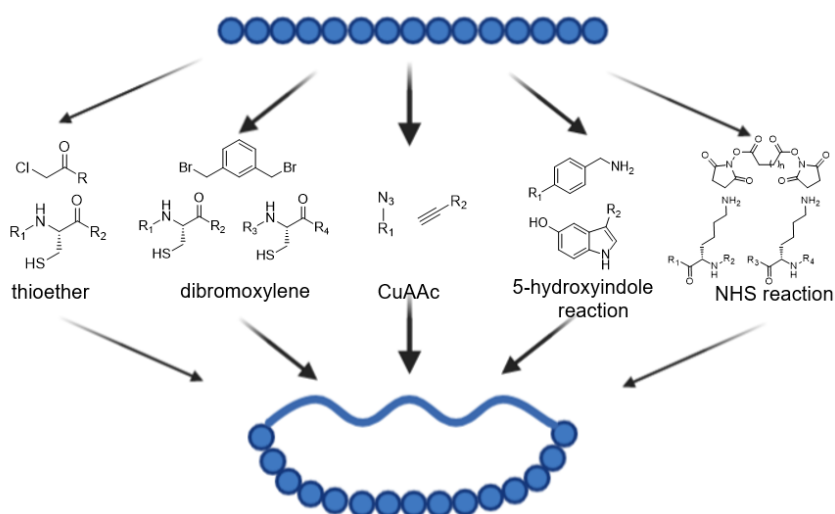


Figure 2. Common cyclization methods applicable to mRNA display

Some of the cyclization methods above are achievable in SPPS (solid phase peptide synthesis), but are difficult to apply in a more biological setting because of the requirement for non-canonical amino acids. Fortunately, genetic code reprogramming technologies can incorporate non-canonical amino acids into translated peptides and proteins. Of the various approaches for this, the most convenient way is to use the aaRS enzymes where these have a suitable substrate scope. Using these, diverse functional groups such as azides, alkynes and alkene have been incorporated into proteins, and so can also be included in displayed peptide libraries.²² Notably, azidohomoalanine (Aha) is widely used in genetic code reprogramming for azide incorporation through mis-acylation by Met-RS and can be applied in reactions such as Staudinger ligation²³, CuAAC²⁴ coupling and SPAAC coupling²⁵. (See section 1.3 & 1.4.2 for more details)

Recently, Timmerman *et al.* reported a synthetic methodology that combines cysteine cyclization and azide-alkyne cycloaddition to together synthesize tricyclic peptides, referred to as CLIPS/CuAAC, which provides access to unique peptide geometry (**Fig 3**). Different scaffolds give different levels of conversion: flexible scaffolds, such as Th1 and Th2, promote the formation of tricycles while a rigid scaffold, Th3, does not yield a clean product. This methodology is thus heavily dependent on a flexible scaffold to reliably form a stable tricyclic peptide.²⁶ Having azides present at the termini and cysteine in the center also gives better results for tricycle formation, and free rotation about a bond in the middle of the scaffold allows for reliable generation

of a single product. This approach has been shown to give convenient access to unusual cyclic peptide topology in a modular manner, but what is currently missing is a means of finding new sequences to exploit this topology. Given that the complex environment of an *in vitro* translation system has been shown to be compatible with both click reactions and bromoxylene cross-linking of cysteines, we considered it likely that the CLIPS/CuAAc approach should be applicable in mRNA display.

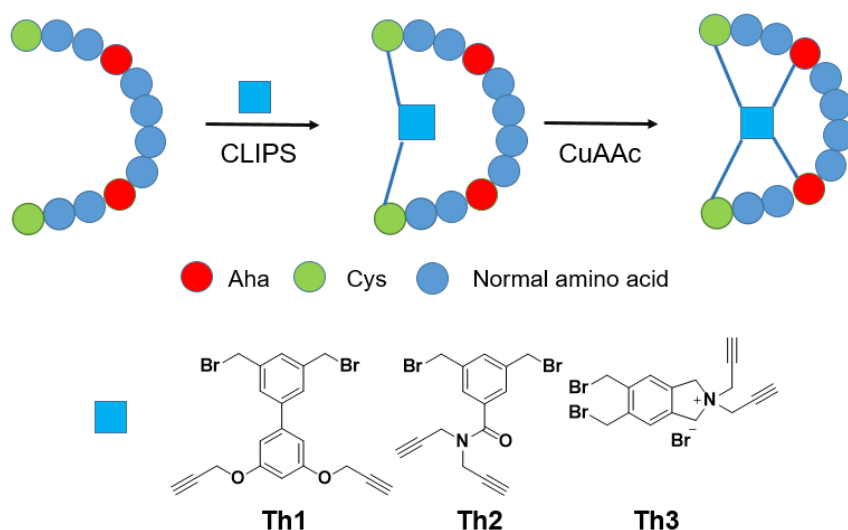


Figure 3. CLIPS/CuAAc reaction. A scaffold reacts first with cysteine to form a macrocyclic peptide, and subsequently closes by intramolecular CuAAc to afford a unique tricyclic peptide structure.

In this chapter, we demonstrate the applicability of CLIPS/CuAAc chemistry as a post-translational cyclisation strategy for mRNA display of tricyclic peptides. Following initial testing by LC-MS with synthetic then reprogrammed translated peptides in the complex translation environment, further validation was carried out based on pull-down in mRNA display with thiol- and azide-reactive biotinylated probes. Using our optimized “one-pot” CLIPS/CuAAc approach we carried out proof-of-principle selections with two scaffolds against four target proteins: FZD5, mAb197, CXCL8 and ST6Gal1. Of these, strong enrichment was found for two of the proteins, and on sequencing these showed unique and converged sequence pools in all cases with the two different scaffolds, suggesting binding is determined by conformation induced by the scaffold.

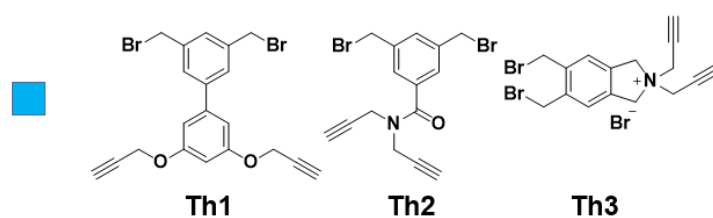
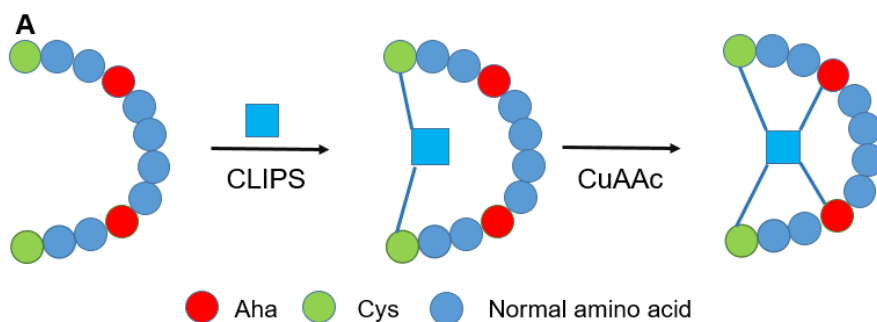
4.2 Results and discussion

4.2.1 Verification by UPLC-MS from synthetic and translated peptides

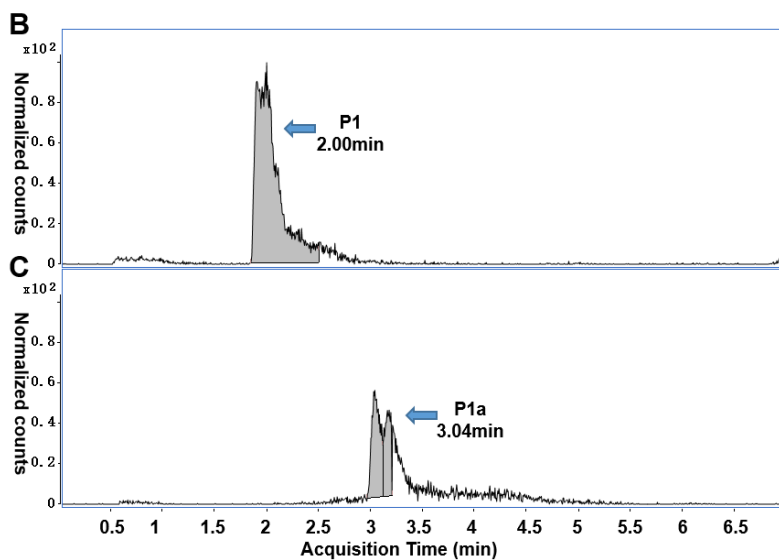
There have previously been reports of CLIPS-based approaches to construct macrocyclic peptide libraries in phage display^{27, 28} and mRNA display²⁹. While copper cytotoxicity limits CuAAc application in phage display, mRNA display (which is based on *in vitro* translation) has a good tolerance for copper and so makes it an appealing setting to exploit this novel multicyclization approach. While the CLIPS/CuAAc methodology is robust on synthetic peptides, the same reaction in the presence of the PURExpress translation system (a commercial *in vitro* translation kit) or similar can potentially see interference from several factors. For example, bacterial translation proteins or the thiols required for reducing conditions could interfere with the CLIPS reaction and decrease the yield, while purification to remove these components could cause a loss of library diversity. Copper chelation and redox chemistry is also a possible concern. Due to the limited and sometimes not reproducible yield of translation, we decided to begin testing with a synthetic peptide diluted into the translation system at a low but known concentration, as a mimic of the real *in vitro* translation product.

The peptide H-CGSGCA[Aha]SRYEVDWRGRGSA[Aha]G-NH₂ ([Aha]= azidohomoalanine) was produced synthetically and then diluted into a mock translation reaction (no template DNA) at 10 μ M concentration, which is similar to the yield expected from *in vitro* translation. Given that the CuAAc reaction is intramolecular and so expected to be the faster of the two, our primary target was to confirm the efficiency of the CLIPS reaction. We adjusted the pH to 8.5 by adding phosphate buffer and then added scaffold Th1 dissolved in DMF/H₂O (1:1) to 1 mM final concentration. The reaction was incubated at 37 °C for 1 hour and analyzed by UPLC-MS. EIC traces confirmed that the CLIPS reaction went well, showing clear but noisy traces for both the unmodified (P1) and modified (P1a) peptides in the translation mixture (**Fig 4**). No remaining starting material could be found, and products from double modification, unclosed rings, or side reactions with beta-mercaptoethanol also could not be identified. While the product EIC trace, and to a lesser extent the starting material, seems to show two peaks, this was the case for all peaks in the run and so attributed to a pressure jump during injection, although it could also reflect slowly interconverting conformers. Regardless, this experiment confirmed that the correct product was forming. The exact masses of both the starting and product peptides match well with the calculated values, although many other peaks are visible. These are derived from the translation system and in many cases have a

higher intensity than the desired peak.



Sequence: H-CGSGCA[Aha]SRYEVDWRGRGSA[Aha]G-NH₂



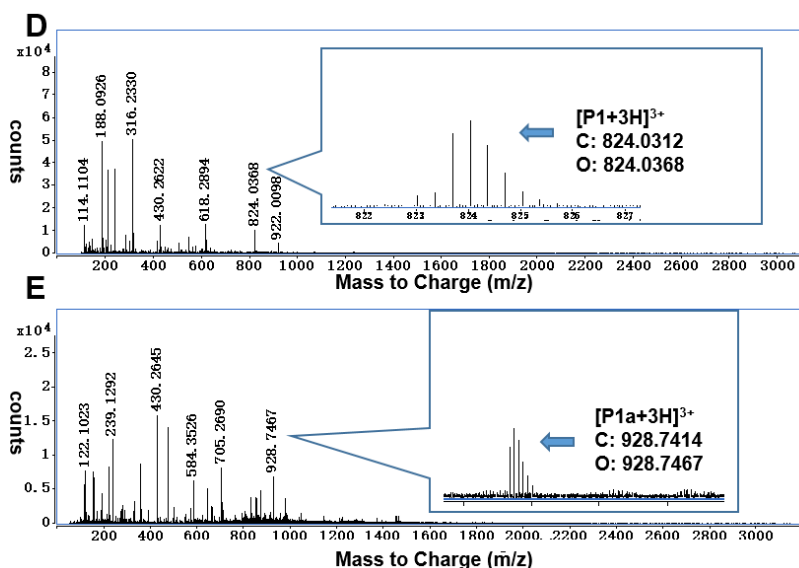
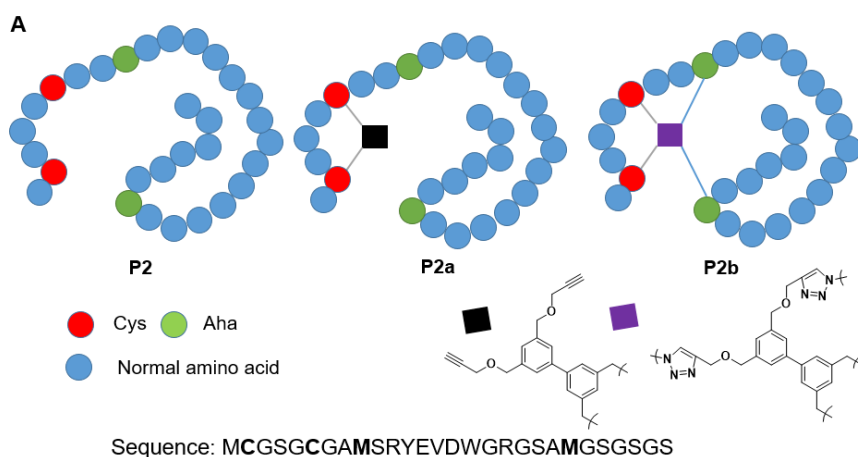


Figure 4. CLIPS reaction of synthetic peptide in the presence of an *in vitro* translation system. A. Cartoon depiction of the reaction sequence in CLIPS/CuAAc scaffolding. B. EIC trace for peptide P1 diluted in the translation mixture; C. EIC trace following CLIPS reaction to form P1a in the translation mixture (EIC for starting material mass showed no peak).; D, The mass spectrum for P1 (1.85-2.30 min); E. The mass spectrum for P1a (2.98-3.24 min).

Encouraged by these results, we next turned our attention to the behavior of the azide in the system and the use of a translated peptide. We designed primers and synthesized DNA encoding the same test peptide as above but with a “GSGSGS” tail. With the addition of azidohomoalanine instead of methionine into the translation reaction, which is tolerated by the MetRS enzyme,³⁰ we can easily incorporate the required azide at AUG codons. To keep this test as close to the synthetic peptide as possible, we opted to put an amino acid with an unreactive group at the initiation position, and so used an acylating ribozyme (flexizyme) to charge initiator tRNA with *N*-Ac-Phe.³¹ This product was confirmed by UPLC-MS, showing the desired mass peak for peptide P2. (**Fig 5.**) With this reprogrammed test peptide translating well we then tested both steps of the CLIPS/CuAAc reaction, analyzing samples after the peptide thiols reacted with the scaffold, treated as for the synthetic peptide above (sample A, expected product is P2a), and after subsequent CuAAc intramolecular reaction, treated with CuSO₄ and sodium ascorbate in a sealed tube with THPTA ligand and aminoguanidine as scavenger to prevent side-reactions (sample B, expected product is P2b). Because the CuAAc reaction does not cause a change in

mass, sample B was also subsequently treated with TCEP to convert any remaining azides to amines and so give a clear mass change if the CuAAc reaction had not occurred. Reaction progress was again assessed by retention time and mass changes on UPLC-MS. These experiments showed that the CLIPS reaction again went well under these conditions, with no P2 starting peptide left in sample A. The CuAAc reaction also appeared to progress smoothly, with P2b exhibiting a clearly different retention time from P2a (2.75 min vs 3.61 min) and little to no peak remaining at the retention time corresponding to P2b in sample B, as well as no sign of a reduced peptide peak derived from unreacted azide. All of these reflect good conversion in the CuAAc reaction, and this reaction was found to not require degassing or an inert atmosphere as previously reported for an intermolecular click reaction in mRNA display.³² However, the background of sample B was found to be particularly noisy, which may arise from ion suppression following the CuAAc reaction and TCEP reduction or alternate counterions such as Cu^{2+} . Together these results show that a peptide can be created with the correct reactive handles for CLIPS/CuAAc reaction by reprogrammed *in vitro* translation, and that reaction proceeds well on this peptide without purifying away from the translation reaction mixture.



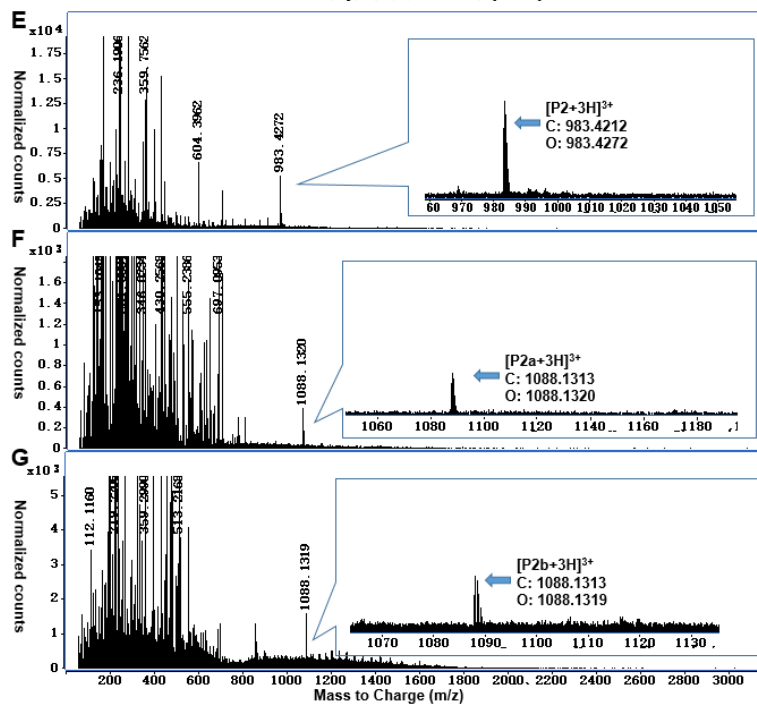
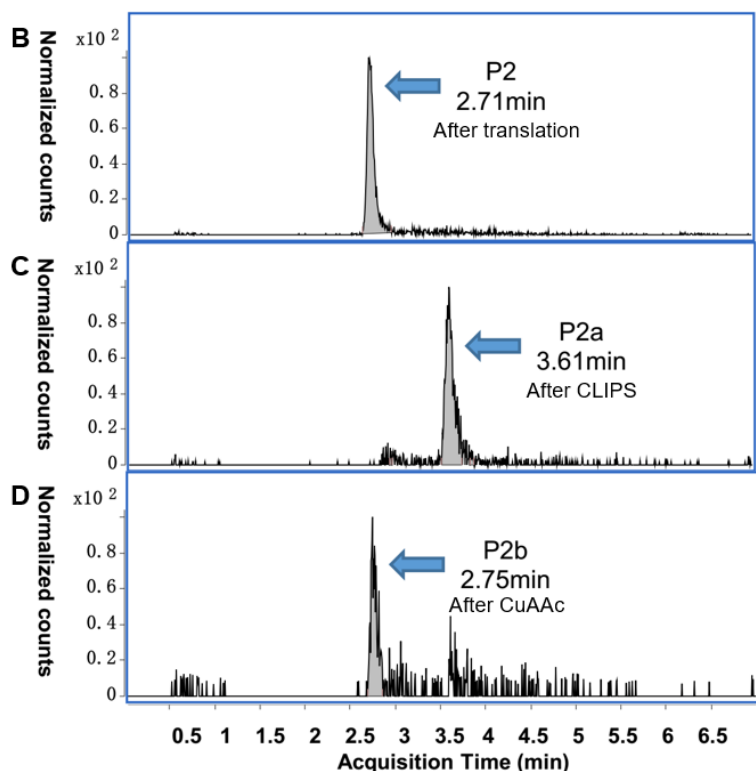


Figure 5. CLIPS/CuAAc tricyclization of a peptide from reprogrammed *in vitro* translation. A. Cartoon representation of the reaction and the sequence of the peptide (Met codons translated as Aha) B. EIC trace for the translated reprogrammed peptide P2; C. EIC trace for peptide following CLIPS reaction to form P2a (EIC for starting material mass showed no peak); D. EIC for peptide following CLIPS/CuAAc and subsequent TCEP reduction to form P2b (EIC for reduced peptide showed no peak); E. The mass-spectrum for P2 (2.61-2.94 min); F. The mass spectrum for P2a (3.52-3.79 min); G. The mass spectrum for P2b (2.69 min-2.88 min).

4.2.2 Verification by pull-down in mRNA display

With these encouraging results from UPLC-MS on synthetic and translated peptides, we continued on to testing the methodology in mRNA display. Since the modification does not give a change that is easily detectable in this setting, to confirm the product of CLIPS and CuAAc reaction in mRNA display we instead used biorthogonal handles to extract any free thiol or free azide that was not fully converted. This indirect pull-down approach has been used before to confirm a library modification reaction in phage display.²⁷ Here we used a similar design but different reactive handles. For free thiol extraction we used an old but robust handle, maleimide. The synthesis of Biotin-Lys-maleimide (Maleimide-Biotin) following the reported literature.³³ To capture free azides we used SPAAC, thereby aiming to avoid interference from an additional CuAAc reaction. The required DBCO-PEG4-Biotin (DBCO-Biotin) is commercially available. All the pull-down experiments were carried out with streptavidin magnetic beads to capture modified mRNA-displayed library. With these maleimide and DBCO probes in hand, we began to test the reaction by using the methodology described above following the work flow illustrated in **Fig 6**. Despite being conceptually relatively simple, this approach proved difficult to carry out in a reproducible manner and took extensive optimization.

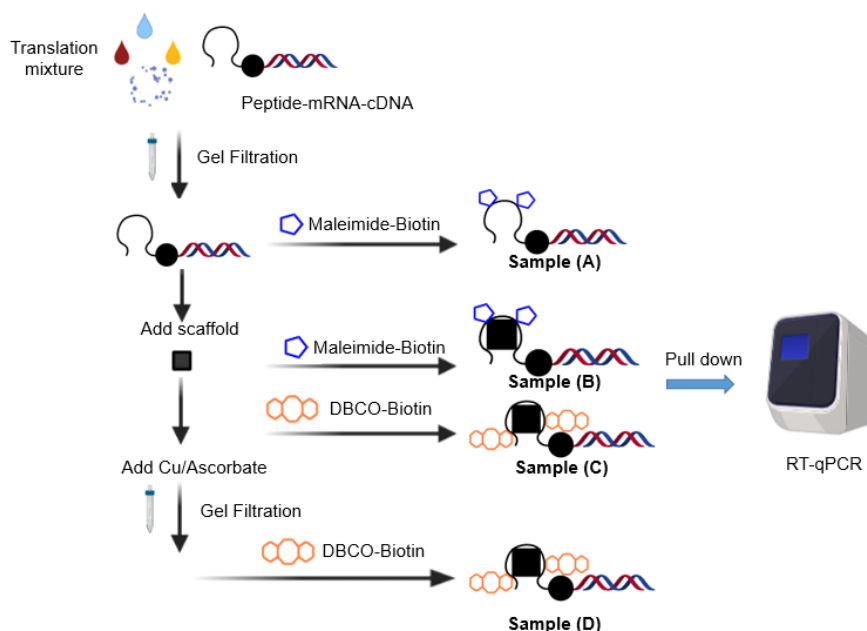


Figure 6. The workflow for qPCR-based quantification in a sequential CLIPS/CuAAC modification of an mRNA-displayed peptide. After *in vitro* translation and reverse transcription, the peptide-mRNA-cDNA is purified by gel filtration. The free thiols on the peptide can be reacted with maleimide-biotin and pulled down with streptavidin to give the thiol input sample (A). After the scaffold is added by CLIPS reaction, any remaining thiol can be extracted by again reacting with maleimide-biotin (B) and the difference in pull-down recovery used as an indirect readout of the conversion in the CLIPS reaction. Similarly, the free azides can be reacted with DBCO-biotin and pulled down with streptavidin to give an azide input level (C). The remaining reaction mixture can then be subjected to the CuAAC reaction conditions. Following purification after the CuAAC reaction any remaining azides can again be coupled to biotin and pulled down in the same manner (D). All samples can then be quantified based on the attached cDNA tag in qPCR and by comparing the recovery ratio (accounting for any losses in purification) the conversion in the CuAAC reaction can be indirectly assessed.

For this approach, we assembled the DNA template encoding a test peptide with the sequence “M**C**GSG**C**SGA**M**SRYEVDWRGRGSA**M**GSGSGS”. In this, the blue “M” represents azide incorporation in the form of methionine replacement with azidohomoalanine (initiation being still by Ac-Phe) and the red “C” shows the cysteine locations that will react with the scaffold. Notably, this template also includes a 3’ untranslated section to allow ligation to the puromycin-containing linker required for

mRNA display. Following *in vitro* transcription by T7 RNA polymerase and said ligation with T4 RNA ligase, modified mRNA was then used for *in vitro* translation and reverse transcription. Despite the positive tests above, we opted to remove the majority of the small molecule contaminants from the solution by use of spin filter or gel filtration in a spin column format (in two separate tests) because of the added complexity of the pull-down tests. The resulting partially purified mixture was split into 4 samples: A, B, C, D. Sample A was reacted with a maleimide-biotin probe as a reading of input total thiols. Samples B, C, D were all treated with scaffold Th1 (2 mM final, added from a DMF/H₂O mix stock solution) at pH 8.0 and allowed to react at 42 °C for two hours. After that, sample B was treated with DBCO-PEG4-Biotin to react with all free azides to give a readout of total azides. Sample B and sample A should match each other since they are both input peptide with 1:1 matched azides and thiols. However, considering the differences in efficiency of the linker chemistry used, it was reasonable to expect that there might be small differences in practice. For this reason, these were not directly compared, rather only using each to compare with the later output using the same chemistry. Sample C was treated with maleimide to capture any unreacted thiols, while sample D continued on to the CuAAc reaction by using CuSO₄/sodium ascorbate as for the UPLC-MS testing above (carried out at 42 °C for two hours to compensate the low peptide concentration). The resultant solution was finally also treated with DBCO-PEG4-Biotin to pull down any remaining unreacted azides. All these samples were then pulled down with streptavidin beads and the captured cDNA tag quantified by qPCR. The conversion rate for each step was then calculated from the ratios (A-B)/A and (C-D)/C. During this indirect verification of CLIPS/CuAAc reaction on mRNA display the biggest problem was poor repeatability. Some of the data showed good conversion while later replicates did not. Even though we are confident of our methodology, poorly reproducible results would give problems for application in selections. This may reflect a problem in the CLIPS/CuAAc methodology, but given the UPLC-MS results we deemed this unlikely and considered that it may be a detection issue – the inverse pull-down approach involves several more reaction and purification steps than the modification reaction itself.

Looking into the workflow in more detail, several possible reasons for these problems were identified: 1. the selectivity and stability of maleimide is too low.³⁴ Indeed, maleimide is an efficient probe but its limitations are well known to chemists and biologists. The maleimide half-life is short in aqueous solutions. What's worse, when the pH is above 7 the selectivity of maleimide will be low and it can give some by-products from reaction with free amines or even hydroxyls; 2. Copper may bind to

amide bonds from proteins, or copper might be inactivated because of free thiols, meaning its catalytic ability will be limited. To make it even worse, the low concentration of peptide in an mRNA-display setting can decrease the CuAAc reaction rate; 3. The SPAAC reaction for free azide capture may be slow and is difficult to drive to full conversion for azides. In some cases, SPAAC can also give by-products that would also influence our result.³⁵

Based on these factors, we considered some approaches for optimization. To improve the selectivity and stability of maleimide, we changed the pH for the maleimide reaction from 7.4 to 6.5. As an alternative to the SPAAC, we investigated the use of TCEP/NHS chemistry to pull down any remaining azides after conversion to amines, with there being no lysines in the test sequence. However, testing of the NHS approach showed too high background modification to be useful. We thus continued with the SPAAC reaction, but to decrease the background the concentration of DBCO-PEG4-Biotin was decreased. To remove potential influence from translation reaction components, such as proteins and free peptides, we adjusted our purification approach as well. Using P/C/I (phenol-chloroform-isoamyl alcohol) extraction would allow removal of proteins and ethanol precipitation would further purify and also give a higher concentration of template peptide for the reactions. Finally, based on the lack of purification in our successful LC-MS test we thought it should be reasonable to combine the CLIPS/CuAAc reactions together into a 'one-pot' reaction with no further purification between the steps. We also considered that the design of our test template may not be optimal. According to the literature,²⁶ the CLIPS/CuAAc reaction will give a higher conversion when the two azides are terminal and the two cysteines central. We thus designed a new test template DNA encoding the peptide "MHEWACKYPICSRNYMGAGAGA" (amino acids color-coded as above), which would incorporate Aha onto the initiation site and a downstream methionine elongation codon, still by MetRS (and thus also removes the need for orthogonal reprogramming for the initiation codon). In designing this, we also tried to cover as many different amino acids as possible to ensure that our tests were a fair assessment of the selectivity of the CLIPS/CuAAc reactions. Finally, taking the eventual mRNA display library design into account, we added a "GAGAGA" spacer where the cDNA-mRNA-puromycin will attach and so minimize its interference in target binding.

Following these adjustments, the new workflow was as shown below (**Fig. 7**). After reverse transcription then phenol/chloroform extraction and ethanol precipitation, a sample was taken from a single translation reaction and the remainder split into four

aliquots. Two of these were used to assess the initial pull-down rate without CLIPS/CuAAc reaction and the remaining two subjected first to CLIPS/CuAAc and then to the same pull-down. The recovery rate for each of the four samples was then calculated from the ratio of pull-down cDNA compared to the initial sample cDNA, and the conversion was calculated from the difference in recovery with and without the CLIPS/CuAAc reaction for each of the two functional groups (**Table 1**). The 'input' recovery for each reaction type reflects the amount of total translated peptide. Normally this value will vary from 0.05-1% for a successful pull-down, depending on various factors including different reprogramming conditions, target binding affinity and library dilution. From the data, it is clear that the DBCO and maleimide input recoveries differ somewhat. This might be caused by differences in labeling efficiency as well as operator variability across the many steps in this challenging protocol. The 'output' recovery indicates the unreacted azide or thiol levels.

Based on the UPLC-MS testing we expected close to 100% conversion for both, and the data from the pull-down testing do agree with this, but still with high variability. Replicate one shows a lower conversion for the CLIPS reaction, but it is unlikely that the subsequent CuAAc reaction could progress to 100% without efficient CLIPS reaction. In replicate two the CLIPS reaction shows good conversion, but the CuAAc shows higher recovery after reaction, which arises from a considerably lower input value (for reasons unknown) rather than a high output value and so is still consistent with acceptable CuAAc conversion. In the third and final replicate both percentage conversions are favorable, but the inputs are both unusually high (the DBCO more so) and so this may be over-estimating conversion. Across three independent replicates we see a high median conversion for both steps (the median being taken to account for the strong effect of outliers on the mean in this case, but with such a small dataset we acknowledge it may also give a skewed perspective), which overall is taken as being at least indicative that the CLIPS/CuAAc reaction is proceeding well enough. We contend that the variability is likely caused by the additional steps required for detection of any remaining functional groups rather than a problem with the CLIPS/CuAAc reaction itself. These data do not show that the mRNA display setting is incompatible with the CLIPS/CuAAc reaction, although they may indicate challenges in reproducibility. These results were thus taken as sufficient indication of plausibility that we could proceed to a real test – a set of proof-of-principle selections that would provide the ultimate evidence that this approach is successful.

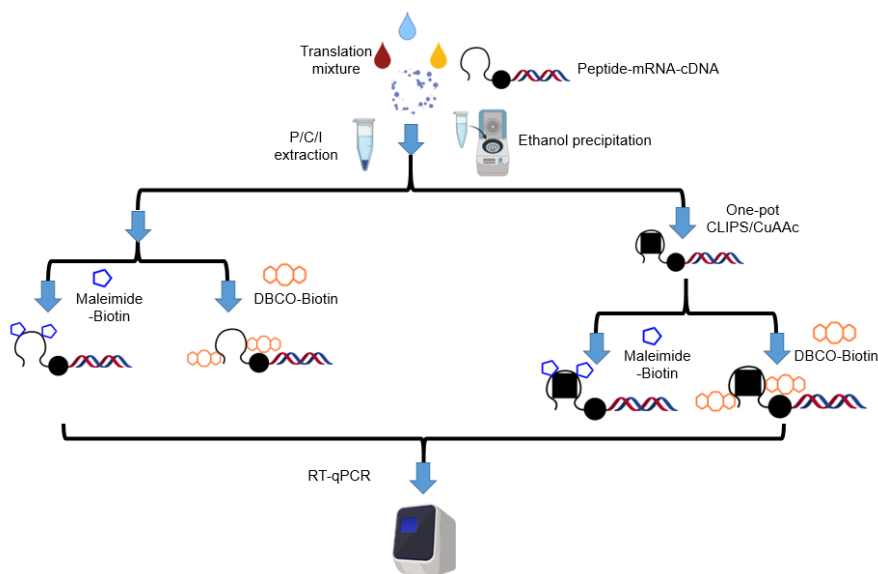


Figure 7. The workflow for qPCR-based quantification of a “one-pot” approach to CLIPS/CuAAC modification of an mRNA-displayed peptide. Following translation, reverse transcription, and purification, samples are reacted with maleimide-biotin and DBCO-biotin either with or without first carrying out the CLIPS/CuAAC modification. All samples are then quantified based on the attached cDNA tag in qPCR and by comparing the recovery ratio (accounting for any losses in purification) the conversion in the CLIPS and CuAAC reactions can be indirectly assessed.

Table 1: Conversion in each step of the inverse-pulldown quantification of CLIPS/CuAAC modification of an mRNA-displayed peptide.

Processed qPCR outputs (%)	1st repeat	2nd repeat	3rd repeat	Median
Maleimide input recovery	0.122	0.123	0.212	
DBCO input recovery	0.119	0.037	0.862	
Maleimide output recovery	0.085	0.001	0.016	
DBCO output recovery	0.006	0.048	0.051	
Calculated conversion				
Maleimide conversion	30%	99%	93%	93%
Click conversion	95%	-29%	95%	95%

4.2.3 mRNA display with CLIPS/CuAAc tricyclization

With a clearly positive test from UPLC-MS for the CLIPS/CuAAc reaction in the presence of the *in vitro* translation components and a positive indication for further compatibility with mRNA display, we moved on to a set of test selections. For this, we used the test template as a model to design a largely randomized library with sequence “MXXXXCXXXXCXXXXMGAGAGA” where ‘M’ again stands for an AUG codon which will incorporate azidohomoalanine, ‘C’, ‘G’, ‘A’ stand for Cysteine, Glycine and Alanine, and ‘X’ stands for random amino acids (NNK codons). After *in vitro* translation of this library and reverse transcription of the attached mRNA tag, the dominant physicochemical characteristics of the much larger mRNA/cDNA duplex allow its purification by phenol/chloroform extraction and ethanol precipitation before the one-pot CLIPS/CuAAc reaction is carried with our optimized conditions. We further increased cyclization and conformational diversity by using two different scaffolds, Th1 and Th2 (Fig 8).

With this approach we will achieve a library of an estimated 10^{13} tricyclic peptides to pan against target proteins immobilized on magnetic beads. In order to remove sequences that may give non-specific binding to either the magnetic beads or the protein’s affinity tag, we used counter-selection against the beads alone (no proteins immobilized) and in later rounds a counter-selection protein (only the tag or the same tag on an unrelated protein). To evaluate the enrichment across rounds of selection we calculated a ‘recovery rate’ from the ratio of recovered cDNA after enrichment and a sample of the total input cDNA in each round. An increasing cDNA recovery rate across several rounds of selections indicates the likely enrichment of target-binding peptides. Deep sequencing of these enriched samples will then give candidate hits for the target protein that can be synthesized and evaluated. For this proof-of-principle set we chose 4 different proteins as targets, based on varied difficulty in selection and also aiming for some biological relevance to any hits found (primarily in cancer biology), which will be outlined below.

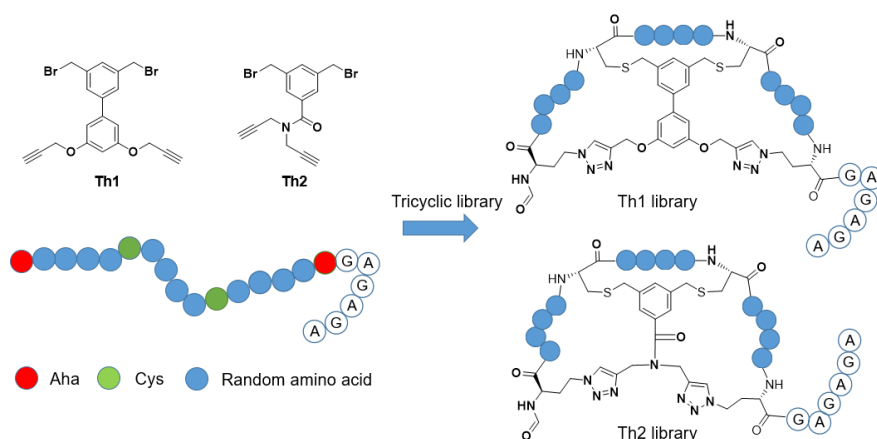
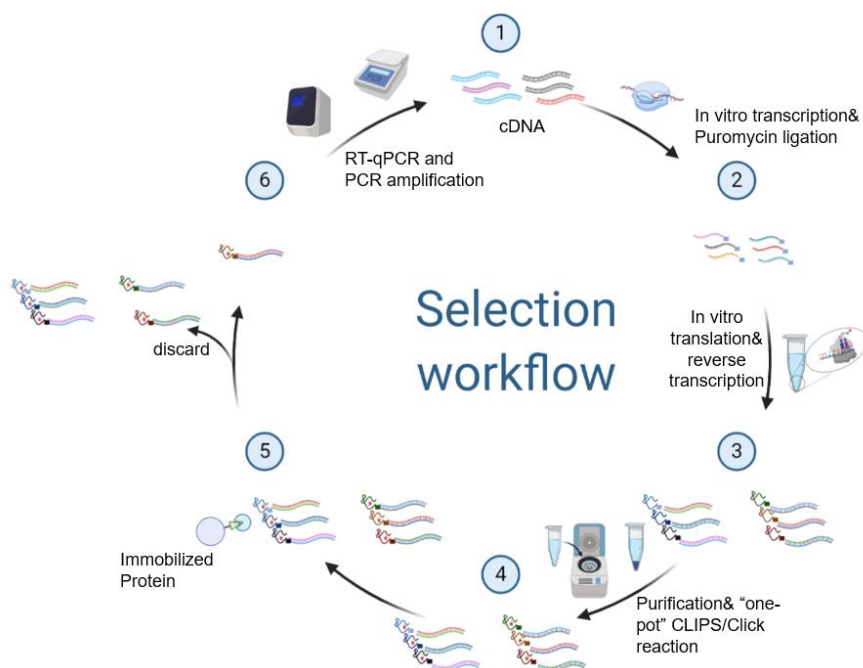
A**B**

Figure 8. The library design and workflow for CLIPS/CuAAC selection. A. Cartoon representation of library design, the Th1 and Th2 scaffolds, and a semi-structural representation of the tricyclic architecture. B. The selection workflow, based on modification of a standard mRNA display approach. Unusual steps are the phenol/chloroform extraction, ethanol precipitation and one-pot CLIPS/CuAAC

reaction carried out after reverse transcription. 1) A DNA library is transcribed by T7 RNA polymerase and puromycin is attached to the mRNA by T4 RNA ligase. 2) The puro-mRNA library is translated in vitro, with azidohomoalanine replacing methionine, and the mRNA tag reverse transcribed to form an mRNA/cDNA hybrid duplex. 3) The peptides are purified by phenol/chloroform extraction and ethanol precipitation, then cyclised around a CLIPS/CuAAc scaffold in a one-pot approach. 4) The tricyclic library is panned over an immobilized protein on magnetic beads. 5) Non-binding peptides are removed by stringent washing. 6) Binding sequences are quantified by qPCR and amplified by PCR, allowing the cycle to repeat (or sequenced).

Frizzled-5 is a 7-transmembrane domain protein involved in governing cell polarity, cell proliferation and many other processes.³⁶ The frizzled family of proteins are receptors for Wnt proteins, as well as other ligands. They are widely expressed, consistent with their role in regulating essential cellular functions. Wnt signaling pathways have received a lot of attention in cancer research, with a FZD-targeting antibody now on the market (Vantictumab).³⁷ FZD5 specifically is overexpressed in several types of cancer, such as breast cancer, colorectal cancer and prostate cancer.^{38, 39} By selecting peptides against Fc-tagged FZD5 immobilized on protein G beads we may be able to block the Wnt signaling pathway and find inhibitors as leads for cancer treatment.

CCR7 is a more typical G-protein, and so also a transmembrane protein, which is expressed in dendritic cells, memory T cells and NK cells.⁴⁰ What's more, CCR7 is also found on various cancer cells, such as breast cancer⁴¹, nonsmall lung cancer⁴² and esophageal cancer⁴³, where it has a role in tumor growth and lymph node metastasis.^{44, 45} CCR7 contributes to organizing thymic architecture and function, being involved in the lymph node homing inflammatory cascade together with its ligands CCL19 and CCL21.⁴⁶ The antibody mAb197 recognizes CCR7 and blocks the recognition site for the chemokine CCL19. This antibody was chosen as a target protein (immobilized on protein G magnetic beads) primarily as a validation system since a binding motif is known, but new peptide ligands raised against mAb197 may also be able to function as decoy receptors for CCL19 and CCL21 and so may be able to tune the immune system.⁴⁷

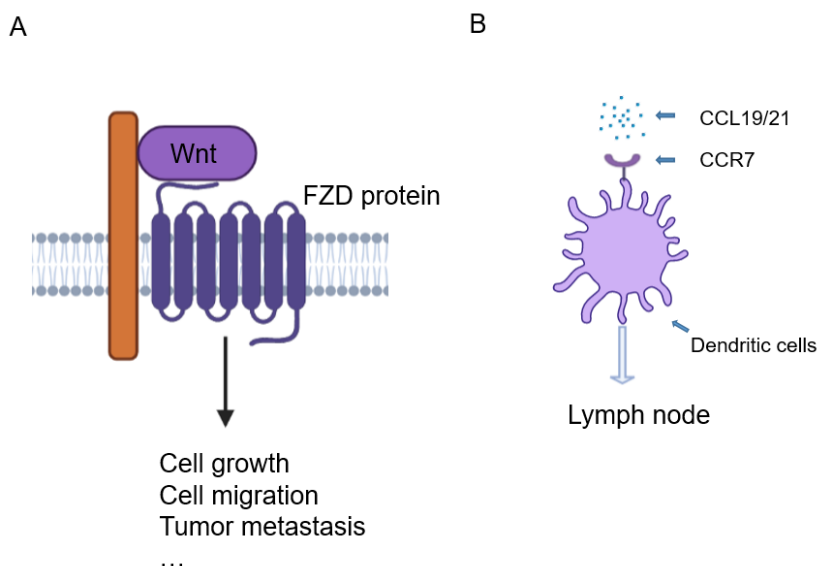


Figure 9. The function of Frizzled proteins and the role of CCR7. A. The Wnt signaling pathway and FZD protein. B. CCL19/CCL21-CCR7 axis directing dendritic cells to the lymph node.

CXCL8 is a chemokine produced by macrophages and endothelial cells.⁴⁸ The most common receptors for CXCL8 are CXCR1 and CXCR2. It is a key mediator of the inflammatory cascade.⁴⁹ CXCL8 can be secreted by tumor cells and its interaction between CXCR1/2 is critical for cancer progression and migration.⁵⁰ By screening for binders to the chemokine CXCL8 we may be able find leads for the inflammatory cascade and cancer progression. As outlined in the introduction, proteolytic stability *in vivo* is a lasting concern in peptide drug discovery. All D-amino acid containing peptides have been found to have better biostability over the all L version.⁵¹ Because of its small size we were able to use biotinylated synthetic D-CXCL8 immobilized on streptavidin beads to select for L-peptides, which after enrichment would yield all-D peptides to bind to the natural all-L protein in a mirror image display strategy. For this target we envisaged the tricyclic library giving the possibility of partially enveloping the small protein similar to a host-guest complex.

ST6Gal1 is an intracellular protein that transfers sialic acid onto galactose-containing substrates to form an alpha-2,6 linkage. As a critical post-translational modification, sialylation mediates a large number of functions including binding of toxins and pathogens, 'self'-recognition by the immune system, and fertilization.⁵² Moreover,

research shows that ST6Gal1 is upregulated in several tumor cells, such as breast cancer⁵³, glioma cancer⁵⁴ and gastric cancer.⁵⁵ Aberrant glycosylation of cells changes their features and is a vital feature and landmark of cancer, found to contribute substantially to tumor migration, proliferation and invasion.⁵⁶ Given the widespread over-sialylation of cells in multiple cancers, selection for tricyclic peptides binding to and hopefully inhibiting ST6Gal1 (immobilized on streptavidin following modification with biotin-PEG4-NHS) again may give leads for anticancer agents that prevent tumors from evading immune detection. Also notable for this target is that several selection campaigns using a simple macrocyclic library againsts this target have not lead to inhibitors, and so this and the fully synthetic D-CXCL8 were in part chosen because they represent particularly challenging targets.

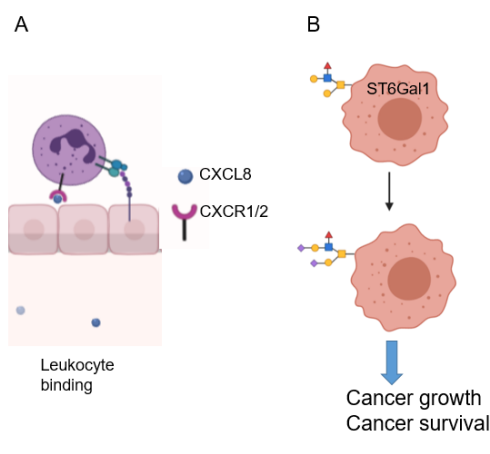
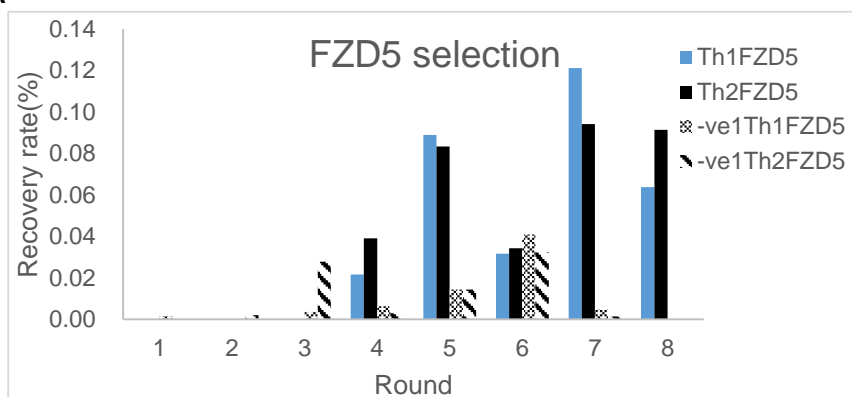


Figure 10. The function of CXCL8 and ST6Gal1. A. CXCL8 and its receptor mediating inflammatory cascade. B. ST6Gal1 mediates sialylation on the cancer cell surface and promotes cancer growth and survival.

The selection against FZD5 proceeded well, as shown by the recovery rate increasing across rounds (**Fig. 11**). The cDNA recovery rate increased dramatically in the 4th and 5th rounds, but there was also an increase in the negative recovery (indicating some proportion of the library binding to the beads alone). To obtain more target-specific binding peptides, we changed the selection protocol from the 6th round by adding IgG₁ as a counter-selection protein, as a decoy target for any Fc-tag binding peptides, together with more rounds of counter-selection against unloaded beads. Thus for the 6th and 7th rounds the negative selections were with half unloaded protein G magnetic beads and half IgG₁ immobilized on the same. This is likely the reason for the decline in recovery in the 6th round. After this decrease, the recovery rate rebounded even

with counter-selection to give strong recovery with target and negligible recovery without in the 8th and final round. From deep sequencing of positive rounds 5 and 8 as well as negative rounds 6 and 8 (to identify and thus rule out any Fc binders), we found that the sequences were highly conserved. In this selection the CLIPS-cyclized core appeared to be conserved across both scaffolds, which is plausible given both have an *m*-dibromoxylene moiety, but the outer loops were unique to each. In the earlier rounds more unique hits for the two scaffolds were also identified. There thus appear to be many candidate inhibitors that will in future be tested for binding (in ELISA format) as synthetic peptides with and without scaffold, and cross-reacting with the other scaffold, to confirm its role in peptide binding.

A



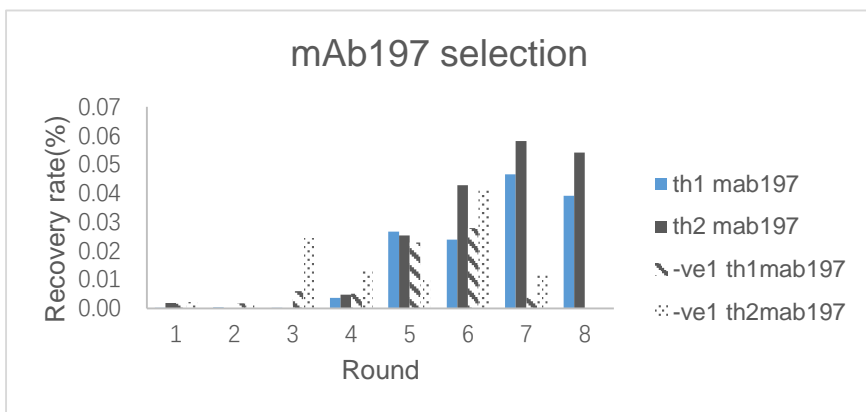
B

Th1								Th2																																			
th1_rd8_1	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_1	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_11	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_15	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_15	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_17	M	Y	R	T	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_16	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	T	G	A	G	A	th2_rd8_19	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_18	M	R	A	V	E	C	Y	R	N	A	C	L	E	R	R	M	G	A	G	A	th2_rd8_20	M	Y	R	S	P	C	Y	R	N	A	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_2	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	T	G	A	G	A	th2_rd8_21	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_22	M	R	A	V	E	C	C	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_22	M	Y	R	S	P	C	Y	R	N	C	L	O	K	M	G	A	G	A	G	A		
th1_rd8_23	M	R	A	V	E	C	H	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_23	M	Y	R	A	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_25	M	R	A	V	E	C	Y	R	N	V	C	L	G	R	R	M	G	A	G	A	th2_rd8_24	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	T	G	A	G	A	G	A	
th1_rd8_32	M	R	A	D	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_27	M	Y	R	S	P	C	H	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A
th1_rd8_33	M	R	A	V	E	C	Y	R	N	V	C	L	K	R	R	M	G	A	G	A	th2_rd8_28	M	Y	R	S	P	C	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A
th1_rd8_34	M	R	A	V	E	C	Y	R	N	M	C	L	E	R	R	M	G	A	G	A	th2_rd8_8	M	Y	R	P	P	C	Y	R	N	C	L	O	K	M	G	A	G	A	G	A		
th1_rd8_40	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_8	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	T	G	A	G	A	G	A	
th1_rd8_46	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	K	G	A	G	A	th2_rd8_10	M	Y	K	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_49	M	R	A	G	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_12	M	Y	G	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_5	M	R	A	V	E	C	Y	R	N	V	C	L	V	R	R	M	G	A	G	A	th2_rd8_14	M	Y	R	S	P	C	Y	R	N	V	R	C	L	O	K	M	G	A	G	A	G	A
th1_rd8_50	M	R	A	V	E	C	Y	R	N	E	C	L	E	R	R	M	G	A	G	A	th2_rd8_16	M	Y	R	S	P	C	Y	R	N	V	C	P	L	O	K	M	G	A	G	A	G	A
th1_rd8_6	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	T	G	A	G	A	th2_rd8_18	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_9	M	R	A	A	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_2	M	Y	R	S	P	C	Y	R	N	V	C	L	O	R	M	G	A	G	A	G	A	
th1_rd8_47	M	R	A	V	E	C	Y	R	N	V	C	L	V	R	R	T	G	A	G	A	th2_rd8_25	M	Y	R	S	O	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_10	M	R	A	V	G	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_26	M	Y	R	S	P	C	Y	R	N	V	C	L	R	R	M	G	A	G	A	G	A	
th1_rd8_12	M	R	A	V	K	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_29	M	Y	R	S	P	C	Y	R	N	V	C	L	E	R	M	G	A	G	A	G	A	
th1_rd8_13	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	O	M	G	A	G	A	th2_rd8_30	M	Y	R	S	P	C	Y	H	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_14	M	R	D	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_31	M	Y	R	S	P	C	Y	R	S	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_17	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_32	M	C	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_19	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_33	M	Y	S	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_20	M	R	A	V	E	C	Y	R	N	V	C	S	E	R	R	M	G	A	G	A	th2_rd8_34	M	Y	R	S	P	C	Y	R	D	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_21	M	R	A	V	E	C	Y	R	N	V	C	L	E	H	R	M	G	A	G	A	th2_rd8_35	M	Y	R	S	P	C	Y	R	N	V	C	L	K	K	M	G	A	G	A	G	A	
th1_rd8_24	M	R	A	V	E	C	H	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_36	M	Y	R	S	P	C	Y	R	N	V	C	L	L	K	M	G	A	G	A	G	A		
th1_rd8_26	M	R	A	V	E	C	Y	S	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_37	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A
th1_rd8_27	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_38	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_28	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_4	M	Y	R	S	P	C	Y	R	N	V	C	L	R	K	M	G	A	G	A	G	A	
th1_rd8_29	M	R	D	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_41	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_3	M	C	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_42	M	Y	R	S	T	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_39	M	C	A	V	E	C	Y	R	N	V	C	L	E	R	R	T	G	A	G	A	th2_rd8_44	M	Y	R	S	P	C	Y	R	N	V	C	M	O	K	M	G	A	G	A	G	A	
th1_rd8_30	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_46	M	Y	R	S	P	C	Y	R	N	V	C	O	L	K	M	G	A	G	A	G	A	
th1_rd8_31	M	R	A	V	E	C	Y	R	D	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_47	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_36	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	I	M	G	A	G	A	th2_rd8_48	M	D	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_37	M	R	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_49	M	Y	R	S	P	C	Y	R	N	V	C	L	L	K	M	G	A	G	A	G	A	
th1_rd8_38	M	H	A	V	E	C	Y	R	N	V	C	L	E	R	R	T	G	A	G	A	th2_rd8_5	M	Y	R	S	P	C	Y	R	N	V	C	L	L	R	M	G	A	G	A	G	A	
th1_rd8_4	M	H	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_50	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_41	M	R	A	V	E	C	Y	C	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_9	M	Y	R	S	P	C	Y	R	N	V	C	L	H	K	M	G	A	G	A	G	A	
th1_rd8_42	M	R	A	V	E	C	Y	R	N	V	C	L	E	C	R	M	G	A	G	A	th2_rd8_11	M	Y	T	T	N	C	Y	R	N	V	C	L	R	T	G	A	G	A	G	A		
th1_rd8_43	M	R	A	V	E	S	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_13	M	Y	T	T	N	C	Y	R	N	V	C	L	R	T	G	A	G	A	G	A		
th1_rd8_44	M	R	A	V	E	C	Y	R	K	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_3	M	Y	T	T	N	C	Y	R	N	V	C	L	L	R	M	G	A	G	A	G	A	
th1_rd8_45	M	S	A	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_40	M	Y	T	T	N	C	Y	R	N	V	C	L	L	R	M	G	A	G	A	G	A	
th1_rd8_48	M	R	T	V	E	C	Y	R	N	V	C	L	E	R	R	T	G	A	G	A	th2_rd8_39	M	Y	T	T	N	C	Y	R	N	V	C	L	L	K	M	G	A	G	A	G	A	
th1_rd8_8	M	R	T	V	E	C	Y	R	N	V	C	L	E	R	R	M	G	A	G	A	th2_rd8_43	M	Y	T	T	N	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	G	A	
th1_rd8_35	M	Y	R	S	P	C	Y	R	N	V	C	L	O	K	M	G	A	G	A	th2_rd8_7	M	D	T	T	N	C	Y	R	N	V	C	L	R	M	G	A	G	A	G	A			
th1_rd8_7	M	D	S	Y	C	N	T	V	C	L	E	R	R	M	G	A	G	A	G	A	th2_rd8_45	M	K	Y	C	D	C	Y	R	K	V	C	A	E	T	M	G	A	G	A	G	A	

Figure 11. Selection results for FZD5. A. The selection enrichment across rounds, with the blue bar for scaffold Th1 and black bar for Th2 positive enrichments, respectively, as well as the first counter-selection in each round ('-ve1', hashed). B. Sequencing results from the final enrichment round for each of the two scaffolds. Sequences are arranged by alignment and named based on abundance rank in high-throughput sequencing. Amino acids are displayed in Rasmol colors.

The selection against mAb197 also proceeded relatively smoothly (**Fig. 12**). The library again showed enrichment from the 4th round, but did not reach the same level of recovery as with FZD5. This sort of variability with target is not abnormal, and does not necessarily reflect the affinity of the final hits. The selection also gave relatively high recovery in the counter-selection rounds, especially in the 4th round. As for mAb197, we again increased the counter-selection stringency from the 6th round by introducing mouse IgG2a to remove affinity tag binders. This is the likely cause of the decrease in enrichment for the 6th round with Th1. After a total of 8 rounds, we again got enrichment to reasonable recovery and a healthy positive/negative ratio and so submitted the positive rounds 7 and 8, as well as negative rounds 6 and 7, for deep sequencing. Analysis of the sequencing data again showed high convergence, in this case on a slightly higher number of motifs. In this case the two scaffolds gave unique results. A high proportion of these also show a known binding motif of “-[aromatic]-A-E-” (personal communication, pepscan BV). Surprisingly, there are some mutations at high abundance in the fixed cyclisation residues in both libraries, with a fixed cysteine mutating to arginine (UGC to CGC) or tyrosine (UGC to UAC) and a compensating second cysteine appearing upstream in the randomized region (e.g. th1_rd8_2, th2_rd8_3). In the latter example an azidohomoalanine also appears in the second loop, with concomitant loss of conservation of the final AUG codon. The unique sequences found in the two enriched datasets as well as the different mutations to conserve the cyclisation again suggest that selection of binding peptides has been successful and that the different scaffolds indeed influence the binding.

A.



B

Th1	Th2
th1_rd8_1 MKYYDCRKY CYAETMGAGA GA	th2_rd8_1 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_13 MKYYDCRKY CYAETMGAGA GA	th2_rd8_15 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_16 MKYYDCGKY CYAETMGAGA GA	th2_rd8_22 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_20 MKYYDCRKY CYAETMGAGA GA	th2_rd8_23 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_25 MKYYDCRKY CYAETMGAGA GA	th2_rd8_24 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_33 MKYYDCRKY CYAETMGAGA GA	th2_rd8_32 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_36 MKYYDCRKY CYAETMGAGA GA	th2_rd8_40 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_38 MKYYDCRKY CYAETMGAGA GA	th2_rd8_7 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_4 MKYHDCRKY CYAETMGAGA GA	th2_rd8_10 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_10 MRYHDCRKY CYAETMGAGA GA	th2_rd8_11 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_19 MKYYDCRKY CYAETMGAGA GA	th2_rd8_12 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_22 MKYYDCRKY CYAETMGAGA GA	th2_rd8_14 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_23 MKYYDCRKY CYAETMGAGA GA	th2_rd8_18 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_24 MEYYDCRKY CYAETMGAGA GA	th2_rd8_21 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_26 MKYYDCRKY CYAETMGAGA GA	th2_rd8_26 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_27 MKYYDCRKY CYAETMGAGA GA	th2_rd8_28 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_28 MYYDCRKY CYAETMGAGA GA	th2_rd8_34 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_29 MKYYDCRKY CYAETMGAGA GA	th2_rd8_35 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_3 MEYYDCRKY CYAETMGAGA GA	th2_rd8_37 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_30 MEYYDCRKY CYAETMGAGA GA	th2_rd8_44 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_31 MKYYDCRKY CYAETMGAGA GA	th2_rd8_47 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_34 MKYYDCRKY CYAETMGAGA GA	th2_rd8_50 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_37 MEYYDCRKY CYAETMGAGA GA	th2_rd8_17 MKERPCTYAE CQSPSTMGAGA GA
th1_rd8_39 MKYYDCRKY CYAETMGAGA GA	th2_rd8_2 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_40 MKYYDCRKY CCAETMGAGA GA	th2_rd8_36 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_41 MKYYDCRKY CHAETMGAGA GA	th2_rd8_46 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_42 MKYYDCRKY CYAETMGAGA GA	th2_rd8_8 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_43 MKYYDCRKY CYAETMGAGA GA	th2_rd8_9 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_46 MKYYDCRKY CYAETMGAGA GA	th2_rd8_13 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_47 MKYYDCRKY CYAETMGAGA GA	th2_rd8_29 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_49 MKYYDCRKY CYAETMGAGA GA	th2_rd8_48 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_5 MKYYDCRKY CYAETMGAGA GA	th2_rd8_5 MKERPCTYAE CQSPSMGAGA GA
th1_rd8_8 MKYYDCRKY CYAETMGAGA GA	th2_rd8_31 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_17 MEYYDCRKY CYAETMGAGA GA	th2_rd8_4 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_2 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_49 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_21 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_30 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_35 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_41 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_9 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_39 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_12 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_19 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_50 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_43 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_18 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_16 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_7 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_20 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_15 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_27 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_48 MCKYFRNDLV CDFPRMGAGA GA	th2_rd8_3 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_44 MPNDVCFNSE CDDHMGAGA GA	th2_rd8_38 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_6 MPNDVCFNSE CDDHMGAGA GA	th2_rd8_25 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_45 MPNDVCFNSE CDDHMGAGA GA	th2_rd8_42 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_11 MSKSDCLYAE CYQMGAGA GA	th2_rd8_45 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_14 MSKSDCLYAE CYQMGAGA GA	th2_rd8_33 MNTYYCAEG CIGDLMGAGA GA
th1_rd8_32 MNTYYCAEG CIGDLMGAGA GA	th2_rd8_6 MNTYYCAEG CIGDLMGAGA GA

Figure 12. Selection results for mab197, displayed as in Fig. 11.

In contrast to the first two targets, selection against D-CXCL8 proved more challenging (**Fig. 13**). Indeed, while an increase in recovery was observed, the final recovery rate of this selection after 8 rounds was much lower than for mAb197 and FZD5. This is not too surprising, as this is a small protein with little surface area for binding. For this selection we did not need to carry out counter-selection with a protein. Despite the low absolute recovery, there was an increase in recovery across the rounds and negligible recovery is seen without target, and so it remains possible that we may be able to find some hits from further analysis of the deep sequencing of the enriched library. In these sequencing results we found numerous sequences also found in the FZD5 and mAb197 libraries, which may indicate some contamination or indexing issues. In addition to this, we also found some frame shift sequences which would not construct a tricyclic peptide. Because of the poor enrichment and the issues with contamination we will first focus subsequent effort on the clearly successful selections above.

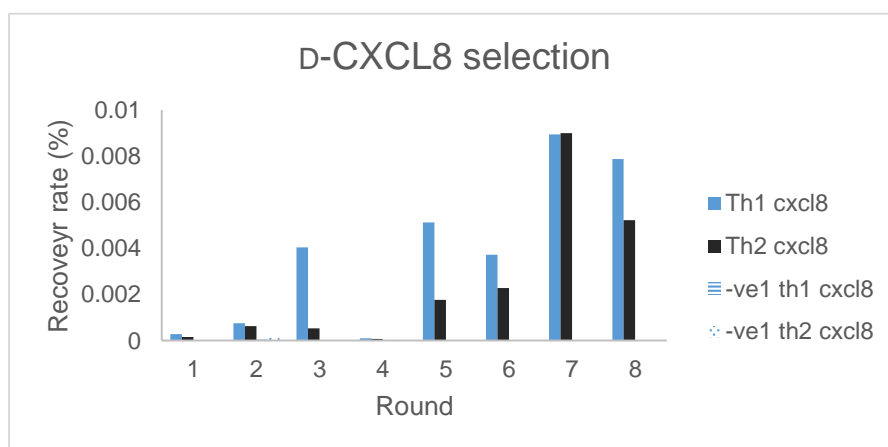


Figure 13. Selection results for D-CXCL8. The selection enrichment across rounds, with the blue bar for scaffold Th1 and black bar for Th2 positive enrichments.

Selection against ST6Gal1 also proved to continue to be stubborn (**Fig. 14**). The recovery rate does seem to increase, but this trend does not continue and the absolute value is not high. However, the ratio of positive/negative recovery was reasonable and so it is unlikely that streptavidin binders have enriched. Across the 8 rounds of selection, we did not get a high recovery rate even though ST6Gal1 is a relatively large and globular protein with an open active site. Why this target is proving difficult is unclear. Despite the uncertain results, the final rounds were submitted for deep sequencing. This showed similar results to the D-CXCL8 selection, with the

sequence pool containing some peptides from FZD5 and mAb197 as well as some frame shift sequences that will disrupt the tricyclic peptide formation. Also for this target we decided to focus our initial attention elsewhere.

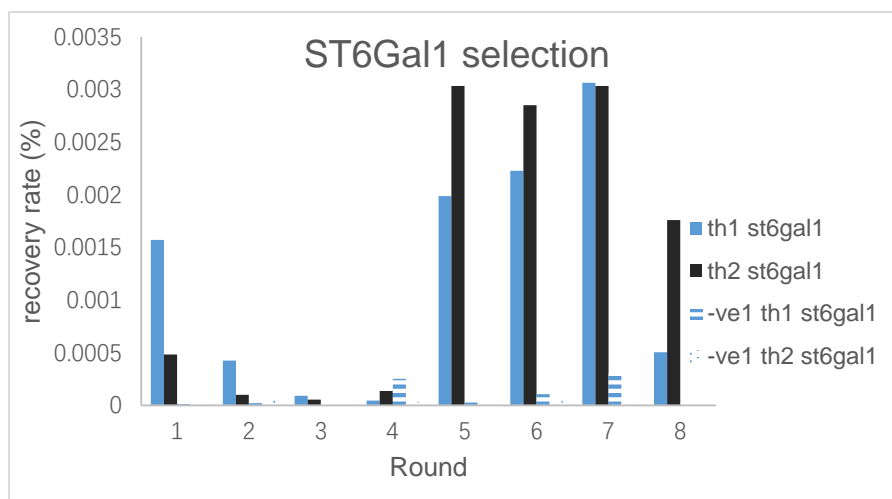


Figure 14. Selection results for ST6Gal1. The selection enrichment across rounds, with the blue bar for scaffold Th1 and black bar for Th2 positive enrichments.

In these proof-of-principle selections with a tricyclic library we got positive results with FZD5 and mAb197 but less conclusive results with D-CXCL8 and ST6Gal1. It is possible that other scaffold and loop combinations will allow more successful selections against these more difficult targets, for example making the loops longer to better envelop the D-CXCL8 or making a single longer loop to enter the active site pocket of ST6Gal1. Regardless of the binding affinity and biological activity of the hits found, our results show that our protocol for one-pot CLIPS/CuAAc modification is compatible with selection by mRNA display. What's more, different scaffolds give different sequences and thus seem to give different conformations or interactions. Also notable is that we do not see a lot of sequences with extra cysteine and azidohomoalanine residues in the random region of the library, a risk with the fully random library design used here. If these were to come up in the random region it would disrupt cyclisation and the resulting peptide conformation, and this seems to be selected against which is consistent with our interpretation that these react with the scaffold to give a tricyclic product in display and that the tricyclization is important for binding. The only cases where a cysteine or azidohomoalanine in the random region were conserved is to compensate for a mutation in one of the conserved residues. We thus conclude that our proof-of-principle CLIPS/CuAAc selection were

successful, with the reaction being compatible with mRNA display and allowing selection of tricyclic peptides.

4.3 Conclusion

Overall, we applied the CLIPS/CuAAC approach to peptides in an *in vitro* translation system and verified the product by UPLC-MS. To further validate this application, we used biotinylating probes based on SPAAC and maleimide to pull down any unreacted azides and thiols, respectively, in tricyclic peptides displayed on mRNA. We used this test system to carry out some optimizations to the modification protocol for application in mRNA display, resulting in a “one-pot” CLIPS/CuAAC protocol which decreases the total reaction time. We designed a tricyclic library with three random loops of four amino acids each and used two different scaffolds to further diversify the libraries before selecting against four different target proteins with relevance to cancer. Two of these targets showed good enrichment with both scaffolds, and gave promising sequences for hits. Selection against mAb197 also resulted in a high percentage of sequences containing a known binding motif in different loops. Different sequences and mutations in selections with the two different scaffold indicate that the scaffold in the library indeed influences the conformation and binding. These hits against mAb197 and FZD5 will in future be tested in ELISA format with synthetic peptide with and without scaffold, and cross-reacting with the other scaffold, as further verification. The results shown in this chapter show that the CLIPS/CuAAC reaction is compatible with mRNA display and could prove to be a powerful tool for peptide drug discovery.

4.4 Experimental

General Methods

All reagents were purchased from Sigma-Aldrich or equivalent chemical suppliers and used without further purification unless otherwise stated. The synthetic peptide for testing, the two scaffolds (Th1 and Th2), and all target proteins were provided by Pepscan BV, except ST6Gal1 which was provided by Prof. Geert-Jan Boons. PURExpress (E6840S and E6850S) was purchased from New England Biolabs. All the other enzymes used in selections were purchased from New England Biolabs. UPLC-MS data were acquired on an Agilent 6560 Ion Mobility Q-TOF LC/MS with an Agilent 1290 Infinity LC system on an Agilent Eclipse Plus C18 RRHD 1.8 μm , 2.1 X 50 mm column. A seven-minute time program was used with a gradient of 10-70% acetonitrile containing 0.1% formic acid. Primers for the tricyclic library and test templates were purchased from IDT Europe (Belgium) and full templates prepared by multi-step assembly PCR amplification before being isolated by precipitation with 70%

ethanol in 0.3 M NaCl. RT-qPCR was carried out on a BIO-RAD™ CFX96.

CLIPS reaction of synthetic peptide in the translation system

Synthetic peptide was dissolved in the translation system to 10 μ M. Phosphate buffer (pH 8.5) was added to 40 mM, followed by scaffold Th1 to 1 mM (final DMF/H₂O 1:3). The reaction mixture was incubated at 37 °C for one hour. Excess protein was precipitated by adding 3 volumes of acetonitrile and the sample was centrifuged for 5 min at 15k rcf then analyzed by UPLC-MS.

CLIPS/CuAAc reaction on translated peptide

A translation reaction was prepared using the PURExpress (Δ aa, tRNA) kit. *In vitro* translation reactions were prepared on ice with the following composition:

- 20% (v/v) PURExpress solution A (Δ aa, tRNA)
- 30% (v/v) PURExpress solution B
- 1 μ g. μ L⁻¹ *E. coli* tRNA
- 50 μ M Ac-Phe acylated tRNAⁱⁿⁱ, previously prepared⁵⁷
- 10 μ g. μ L⁻¹ DNA template
- 0.5 mM each 20 amino acid mix (-Methionine, +Aha)
- Volume was adjusted to 5 μ L with milliQ water

The translation reaction was incubated at 37 °C for 30 minutes. After translation, phosphate buffer (pH 8.5) and Scaffold Th1 (4 mM stock in DMF/H₂O=1/1) were added to final concentrations of 40 mM and 2 mM, respectively. The mixture was allowed to incubate at 37 °C for one hour. For direct analysis, excess protein was precipitated by adding 3 volumes of acetonitrile and the sample was centrifuged for 5 min at 15k rcf then analyzed by UPLC-MS.

For the subsequent CuAAc reaction, the following was added:

- 0.5mM CuSO₄
- 2.5mM THPTA
- 5mM sodium ascorbate
- 5mM aminoguanidine hydrochloride

The tube was then sealed and the reaction was allowed to incubate at 42 °C for two hours. For analysis after the incubation, excess protein was precipitated by adding 3 volumes of acetonitrile and the sample was centrifuged for 5 min at 15k rcf then analyzed by UPLC-MS.

For azide reduction after the click reaction, TCEP.HCl was added to a final

concentration 35 mM and the resulting solution incubated at 37 °C for one hour. After the incubation, excess protein was precipitated by adding 3 volumes of acetonitrile and the sample was centrifuged for 5 min at 15k rcf then analyzed by UPLC-MS..

“One-pot” CLIPS/CuAAc reaction

After *in vitro* translation (at 5 µL scale) and reverse transcription were finished (total volume 10 µL), the reaction mixture was diluted to 30 µL with MilliQ water. The water layer was extracted with 30 µL P/C/I solution and then 30 µL C/I solution (Amresco). The final solution was adjusted to 100 mM NaCl, 10 µg glycogen, and 67% ethanol. The solution was vortexed and then centrifuged at 15k rcf for 15 min, the supernatant discarded and the pellet washed with 50 µL of 70% ethanol. The sample was then centrifuged at 15 rpm for 3 min, the supernatant removed again and the pellets air dried for 3 min. The pellet was redissolved in phosphate buffer (50 mM, pH 8.0), and the CLIPS/CuAAc reaction was carried with the following composition

50 mM phosphate buffer, pH 8.0

2 mM scaffold

500 µM CuSO₄

2.5 mM THPTA

5 mM aminoguanidine

5 mM sodium ascorbate

The solution was incubated at 42 °C for two hours and purified again by ethanol precipitation as above.

Model template

Template 1

Sequence: MCGSGCSGAMSRYEVDWRGRGSAMGSGSGS

DNA sequence:

TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGTGCGGGTC
TGGATGCGGTGCGATGAGCCGCTATGAGGTGGATTGGCGCGGCCGCGGGTCG
GCAATGGGTAGCGGCAGCGGCAGCTAGGACGGGGGGCGGAAA

Template 2

Sequence: MHEWACKYPICSRNYMGAGAGA

DNA sequence:

TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGCATGAATG
GGCGTGCAAATATCCGATTTCAGCCGCAACTATATGGGTGCAGGCGCAGGCG
CATAGGACGGGGGGCGGAAA

Library

Sequence: MXXXXCXXXXCXXXXMGAGAGA

DNA sequence:

TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGNNKNNKNN
KNNKTGCNNKNNKNNKNNKTGCNNKNNKNNKNNKATGGGTGCAGGCGCAGGC
GCATAGGACGGGGGGCGGAAA

Biotinylation of ST6Gal1

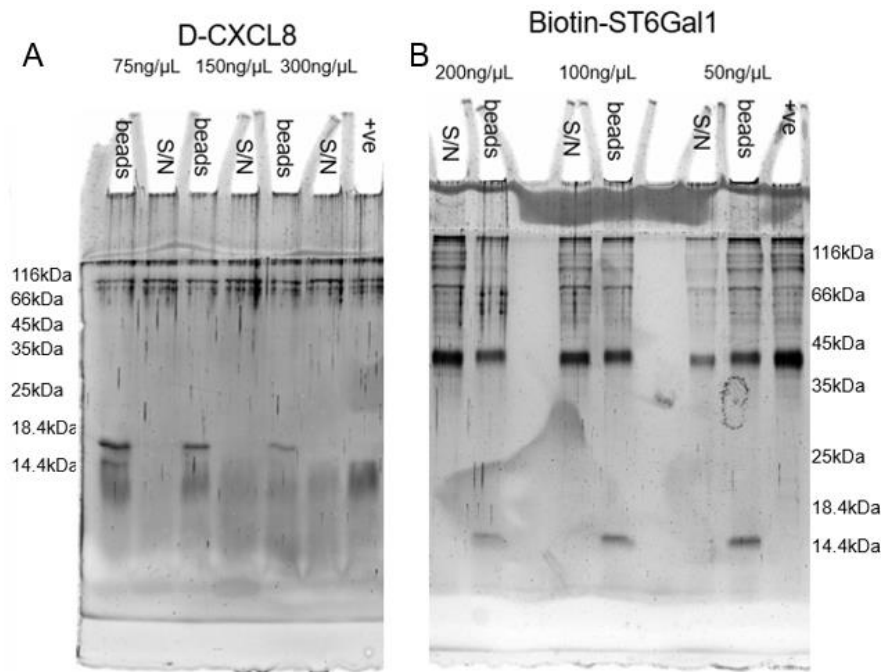
ST6Gal1 (1 mg.ml⁻¹) was incubated with 10 eq. of NHS-PEG4-Biotin at 4 °C overnight in PBS buffer (10 mM Phosphate pH 7.4, 140 mM NaCl). After reaction, the protein was purified with a 10 kDa cut off spin filter and washed 2 times with 500 µL storage buffer. The resulting solution was quantified by absorbance at 280 nm and diluted to 0.5 mg.ml⁻¹ for further use.

Synthesis of library and selection

Library construction and screening were performed largely as previously described.³¹ A DNA template containing 12 NNK codons was synthesized from commercially available primers by extension and PCR. The corresponding mRNA library was transcribed by T7 RNA polymerase and covalently linked to Puromycin linker by T4 RNA ligase. The resulting puro-mRNA library was translated at 1.0 µM under a reprogrammed genetic code in a 5 µL *in vitro* translation reaction, omitting Met and replacing it with Aha (double scale for the first round). Following translation and release of the mRNA-peptide from ribosomes with EDTA, cDNA was produced by reverse transcription. The cDNA-mRNA-peptide library was then submitted to the one-step CLIPS-CuAAc protocol above, and the resulting modified libraries panned against 1 pmol target protein (or for D-CXCL8, 2.5 pmol) immobilized on Dynabeads™ Protein G or M-280 Streptavidin magnetic beads (ThermoFisher) at 4 °C for 30 min with constant gentle inversion. The beads were subsequently washed 3 times with 20 µL PBS-T (10 mM phosphate pH 7.4, 140 mM NaCl, 0.02 v/v % Tween-20). The output was quantified by qPCR and cDNA was recovered by PCR using forward primer TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATG and reverse primer TTTCCGCCCCCGTCCTAGCTGCCGCTGCCGCTGCCGCA. Following 8 rounds of selection, the final enriched DNA library was sequenced by MiSeq and results extracted by Python script and analyzed in CLC sequencer viewer.

Assessing protein binding capacity on magnetic beads

A series of incubations with 200 ng protein each was set up with varied volumes of magnetic bead slurry (Protein G magnetic beads or Streptavidin magnetic beads, for Fc and biotin tags respectively). After incubation at 4 °C for 30 mins with gentle inversion, the supernatant was removed and the beads were washed 3 times with 20 μ L 1X PBS-T. After that, SDS loading buffer was added to the beads and to the supernatant, which were then incubated at 95 °C for 5 min. Samples were loaded on an SDS-PAGE gel (10%, or 14% for D-CXCL8) and run at 200 V for 30 min, then protein visualized by silver stain.⁵⁸ (**Fig. 15**).



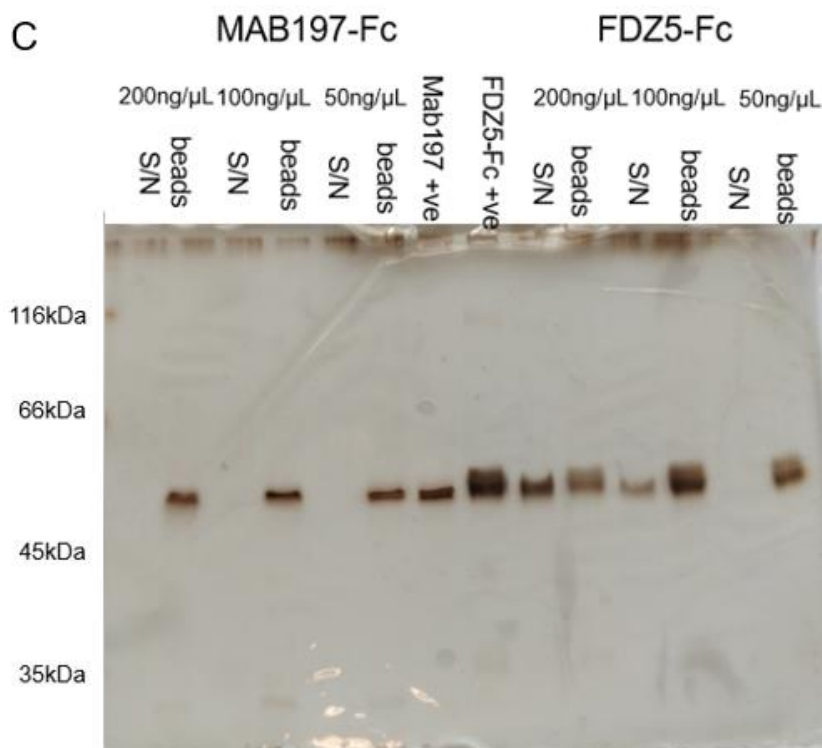


Figure 15. Optimization of protein binding to magnetic beads for A. D-CXCL8; B. Biotin-ST6Gal1; C. mAb197 and FZD5.

Reference

1. Rubin, S.; Qvit, N., Cyclic Peptides for Protein-Protein Interaction Targets: Applications to Human Disease. **2016**, *26*, 199-221.
2. Smith, G. P.; Petrenko, V. A., Phage Display. *Chem. Rev.* **1997**, *97*, 391-410.
3. Eric, V. S.; Lauren, R. P.; Yong Ku, C.; Eric, T. B., A Decade of Yeast Surface Display Technology: Where Are We Now? *Combinatorial Chem. High Throughput Screening* **2008**, *11*, 127-134.
4. Wang, H.; Liu, R., Advantages of mRNA display selections over other selection techniques for investigation of protein-protein interactions. *Expert. Rev. Proteomics.* **2011**, *8*, 335-346.
5. Brinckerhoff, L. H.; Kalashnikov, V. V.; Thompson, L. W.; Yamshchikov, G. V.; Pierce, R. A.; Galavotti, H. S.; Engelhard, V. H.; Slingsluff Jr, C. L., Terminal modifications inhibit proteolytic degradation of an immunogenic mart-127-35 peptide: Implications for peptide vaccines. *Int. J. Cancer* **1999**, *83*, 326-334.

6. Jung-Mo, A.; Nicholas, A. B.; Mary, T. M.; Kim, D. J., Peptidomimetics and Peptide Backbone Modifications. *Mini-Rev. Med. Chem.* **2002**, *2*, 463-473.
7. Yin, H., Constrained Peptides as Miniature Protein Structures. *ISRN Biochemistry* **2012**, *2012*, 692190.
8. Synthesis of Special Peptides and Peptide Conjugates. *Peptides: Chemistry and Biology* **2009**, 365-409.
9. Qvit, N.; Reuveni, H.; Gazal, S.; Zundevich, A.; Blum, G.; Niv, M. Y.; Feldstein, A.; Meushar, S.; Shalev, D. E.; Friedler, A.; Gilon, C., Synthesis of a Novel Macrocyclic Library: Discovery of an IGF-1R Inhibitor. *J. Comb. Chem.* **2008**, *10*, 256-266.
10. Qvit, N.; Hatzubai, A.; Shalev, D. E.; Friedler, A.; Ben-Neria, Y.; Gilon, C., Design and synthesis of backbone cyclic phosphorylated peptides: The IκB model. *Biopolymers* **2009**, *91*, 157-168.
11. Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D., Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615-2623.
12. Fosgerau, K.; Hoffmann, T., Peptide therapeutics: current status and future directions. *Drug Discov. Today* **2015**, *20*, 122-128.
13. Koniev, O.; Wagner, A., Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc. Rev.* **2015**, *44*, 5495-5551.
14. Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H., Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. *ACS Chem. Biol.* **2008**, *3*, 120-129.
15. Guillen Schlippe, Y. V.; Hartman, M. C. T.; Josephson, K.; Szostak, J. W., In Vitro Selection of Highly Modified Cyclic Peptides That Act as Tight Binding Inhibitors. *J. Am. Chem. Soc.* **2012**, *134*, 10469-10477.
16. Deyle, K.; Kong, X.-D.; Heinis, C., Phage Selection of Cyclic Peptides for Application in Research and Drug Development. *Acc. Chem. Res.* **2017**, *50*, 1866-1874.
17. Bellotto, S.; Chen, S.; Rentero Rebollo, I.; Wegner, H. A.; Heinis, C., Phage Selection of Photoswitchable Peptide Ligands. *J. Am. Chem. Soc.* **2014**, *136*, 5880-5883.
18. Angelini, A.; Diderich, P.; Morales-Sanfrutos, J.; Thurnheer, S.; Hacker, D.; Menin, L.; Heinis, C., Chemical Macrocyclization of Peptides Fused to Antibody Fc Fragments. *Bioconjugate Chem.* **2012**, *23*, 1856-1863.
19. Millward, S. W.; Takahashi, T. T.; Roberts, R. W., A General Route for Post-Translational Cyclization of mRNA Display Libraries. *J. Am. Chem. Soc.* **2005**, *127*,

14142-14143.

20. Hacker, D. E.; Hoinka, J.; Iqbal, E. S.; Przytycka, T. M.; Hartman, M. C. T., Highly Constrained Bicyclic Scaffolds for the Discovery of Protease-Stable Peptides via mRNA Display. *ACS Chem. Biol.* **2017**, *12*, 795-804.
21. Yamagishi, Y.; Ashigai, H.; Goto, Y.; Murakami, H.; Suga, H., Ribosomal Synthesis of Cyclic Peptides with a Fluorogenic Oxidative Coupling Reaction. *ChemBioChem* **2009**, *10*, 1469-1472.
22. Saleh, A. M.; Wilding, K. M.; Calve, S.; Bundy, B. C.; Kinzer-Ursem, T. L., Non-canonical amino acid labeling in proteomics and biotechnology. *J. Biol. Eng.* **2019**, *13*, 43.
23. Saxon, E.; Bertozzi, C. R., Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287*, 2007.
24. Link, A. J.; Tirrell, D. A., Cell Surface Labeling of Escherichia coli via Copper(I)-Catalyzed [3+2] Cycloaddition. *J. Am. Chem. Soc.* **2003**, *125*, 11164-11165.
25. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A Strain-Promoted [3 + 2] Azide-Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047.
26. Richelle, G. J. J.; Ori, S.; Hiemstra, H.; van Maarseveen, J. H.; Timmerman, P., General and Facile Route to Isomerically Pure Tricyclic Peptides Based on Templated Tandem CLIPS/CuAAC Cyclizations. *Angew. Chem. Int. Ed.* **2018**, *57*, 501-505.
27. Ng, S.; Derda, R., Phage-displayed macrocyclic glycopeptide libraries. *Org. Biomol. Chem.* **2016**, *14*, 5539-5545.
28. Diderich, P.; Bertoldo, D.; Dessen, P.; Khan, M. M.; Pizzitola, I.; Held, W.; Huelsken, J.; Heinis, C., Phage Selection of Chemically Stabilized α -Helical Peptide Ligands. *ACS Chem. Biol.* **2016**, *11*, 1422-1427.
29. Bashiruddin, N. K.; Suga, H., Construction and screening of vast libraries of natural product-like macrocyclic peptides using in vitro display technologies. *Curr. Opin. Chem. Biol.* **2015**, *24*, 131-138.
30. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 19-24.
31. Johansen-Leete, J.; Passioura, T.; Foster, S. R.; Bhusal, R. P.; Ford, D. J.; Liu, M.; Jongkees, S. A. K.; Suga, H.; Stone, M. J.; Payne, R. J., Discovery of Potent Cyclic Sulfopeptide Chemokine Inhibitors via Reprogrammed Genetic Code mRNA Display. *J. Am. Chem. Soc.* **2020**, *142*, 9141-9146.
32. Horiya, S.; Bailey, J. K.; Temme, J. S.; Guillen Schlippe, Y. V.; Krauss, I. J., Directed Evolution of Multivalent Glycopeptides Tightly Recognized by HIV Antibody

- 2G12. *J. Am. Chem. Soc.* **2014**, *136*, 5407-5415.
33. Abt, D.; Schmidt, B. V. K. J.; Pop-Georgievski, O.; Quick, A. S.; Danilov, D.; Kostina, N. Y.; Bruns, M.; Wenzel, W.; Wegener, M.; Rodriguez-Emmenegger, C.; Barner-Kowollik, C., Designing Molecular Printboards: A Photolithographic Platform for Recodable Surfaces. *Chem. EUR. J.* **2015**, *21*, 13186-13190.
34. Zhang, Y.; Zhou, X.; Xie, Y.; Greenberg, M. M.; Xi, Z.; Zhou, C., Thiol Specific and Tracelessly Removable Bioconjugation via Michael Addition to 5-Methylene Pyrrolones. *J. Am. Chem. Soc.* **2017**, *139*, 6146-6151.
35. van Geel, R.; Pruijn, G. J. M.; van Delft, F. L.; Boelens, W. C., Preventing Thiol-Yne Addition Improves the Specificity of Strain-Promoted Azide-Alkyne Cycloaddition. *Bioconjugate Chem.* **2012**, *23*, 392-398.
36. Huang, H.-C.; Klein, P. S., The Frizzled family: receptors for multiple signal transduction pathways. *Genome Biol.* **2004**, *5*, 234-234.
37. Tai, D.; Wells, K.; Arcaroli, J.; Vanderbilt, C.; Aisner, D. L.; Messersmith, W. A.; Lieu, C. H., Targeting the WNT Signaling Pathway in Cancer Therapeutics. *The Oncologist* **2015**, *20*, 1189-1198.
38. Zeng, C.-M.; Chen, Z.; Fu, L., Frizzled Receptors as Potential Therapeutic Targets in Human Cancers. *International Journal of Molecular Sciences* **2018**, *19*, 1543-1560.
39. Expression of FZD5 in cancers.
40. Sharma, N.; Benechet, A. P.; Lefrançois, L.; Khanna, K. M., CD8 T Cells Enter the Splenic T Cell Zones Independently of CCR7, but the Subsequent Expansion and Trafficking Patterns of Effector T Cells after Infection Are Dysregulated in the Absence of CCR7 Migratory Cues. *J. Immunol.* **2015**, *195*, 5227.
41. Rizeq, B.; Malki, M. I., The Role of CCL21/CCR7 Chemokine Axis in Breast Cancer Progression. *Cancers (Basel)* **2020**, *12*.
42. Takanami, I., Overexpression of CCR7 mRNA in nonsmall cell lung cancer: Correlation with lymph node metastasis. *Int. J. Cancer* **2003**, *105*, 186-189.
43. Ding, Y.; Shimada, Y.; Maeda, M.; Kawabe, A.; Kaganoi, J.; Komoto, I.; Hashimoto, Y.; Miyake, M.; Hashida, H.; Imamura, M., Association of CC Chemokine Receptor 7 with Lymph Node Metastasis of Esophageal Squamous Cell Carcinoma. *Clin. Cancer. Res.* **2003**, *9*, 3406-3412.
44. Mashino, K.; Sadanaga, N.; Yamaguchi, H.; Tanaka, F.; Ohta, M.; Shibuta, K.; Inoue, H.; Mori, M., Expression of Chemokine Receptor CCR7 Is Associated with Lymph Node Metastasis of Gastric Carcinoma. *Cancer Res.* **2002**, *62*, 2937.
45. Ding, Y.; Shimada, Y.; Maeda, M.; Kawabe, A.; Kaganoi, J.; Komoto, I.; Hashimoto, Y.; Miyake, M.; Hashida, H.; Imamura, M., Association of CC Chemokine

Receptor 7 with Lymph Node Metastasis of Esophageal Squamous Cell Carcinoma. *Clin. Cancer. Res.* **2003**, *9*, 3406.

46. Förster, R.; Davalos-Misslitz, A. C.; Rot, A., CCR7 and its ligands: balancing immunity and tolerance. *Nat. Rev. Immunol.* **2008**, *8*, 362-371.

47. Wang, X.; Wang, L.; Miao, L.; Zhao, R.; Wu, Y.; Kong, X., CC-chemokine receptor 7 and its ligand CCL19 promote mitral valve interstitial cell migration and repair. *J Biomed Res* **2015**, *29*, 456-464.

48. Hedges, J. C.; Singer, C. A.; Gerthoffer, W. T., Mitogen-Activated Protein Kinases Regulate Cytokine Gene Expression in Human Airway Myocytes. *Am. J. Respir. Cell Mol. Biol.* **2000**, *23*, 86-94.

49. Harada, A.; Sekido, N.; Akahoshi, T.; Wada, T.; Mukaida, N.; Matsushima, K., Essential involvement of interleukin-8 (IL-8) in acute inflammation. *J. Leukocyte Biol.* **1994**, *56*, 559-564.

50. Ha, H.; Debnath, B.; Neamati, N., Role of the CXCL8-CXCR1/2 Axis in Cancer and Inflammatory Diseases. *Theranostics* **2017**, *7*, 1543-1588.

51. Feng, Z.; Xu, B., Inspiration from the mirror: D-amino acid containing peptides in biomedical approaches. *Biomol Concepts* **2016**, *7*, 179-187.

52. Varki, A., Sialic acids in human health and disease. *Trends Mol. Med.* **2008**, *14*, 351-360.

53. Lu, J.; Isaji, T.; Im, S.; Fukuda, T.; Hashii, N.; Takakura, D.; Kawasaki, N.; Gu, J., β -Galactoside α 2,6-Sialyltransferase 1 Promotes Transforming Growth Factor- β -mediated Epithelial-Mesenchymal Transition*. *J. Biol. Chem.* **2014**, *289*, 34627-34641.

54. Yamamoto, H.; Oviedo, A.; Sweeley, C.; Saito, T.; Moskal, J. R., Alpha2,6-sialylation of cell-surface N-glycans inhibits glioma formation in vivo. *Cancer Res.* **2001**, *61*, 6822-6829.

55. Gretschel, S.; Haensch, W.; Schlag, P. M.; Kemmner, W., Clinical Relevance of Sialyltransferases ST6GAL-I and ST3GAL-III in Gastric Cancer. *Oncology* **2003**, *65*, 139-145.

56. Garnham, R.; Scott, E.; Livermore, K. E.; Munkley, J., ST6GAL1: A key player in cancer (Review). *Oncol. Lett.* **2019**, *18*, 983-989.

57. Goto, Y.; Katoh, T.; Suga, H., Flexizymes for genetic code reprogramming. *Nat. Protoc.* **2011**, *6*, 779-790.

58. Chevallet, M.; Luche, S.; Rabilloud, T., Silver staining of proteins in polyacrylamide gels. *Nat. Protoc.* **2006**, *1*, 1852-1858.

Chapter 5

Summary and conclusions

Summary and perspective

This thesis describes work on *in vitro* translation engineering and application: bacterial initiation reprogramming for peptides and proteins, and mRNA display with genetic code reprogramming and chemical post-translational modification. During testing of incorporation of methionine mimics, we found that suppression of formylation in bacterial translation makes a vacant initiation site for reprogramming without influencing decoding of elongator AUG codons. This observation formed the basis for work presented in chapter 2. By exploiting genetic code reprogramming in mRNA display we can add additional binding factors and handles to carry out post-translational modification to construct macrocyclic or unique tricyclic peptides. On its own this display technology is a robust and fast peptide selection technology to discover potent binders for target proteins, and with the modifications described in chapters 3 and 4 we can access increasingly structurally diverse libraries.

Chapter 1 presents a theoretical background for the subsequent research chapters. In this an overview of bacterial translation is introduced, talking about the precise control of the initiation process from structural and dynamic aspects. How bacteria accurately transfer fMet-tRNA_{ini} onto the initiation site instead of the Met elongation site was outlined, as well as how the elongation and termination of translation work in the ribosome. Following this, peptide selection technologies such as yeast, phage, and mRNA display were introduced, and the promise of incorporating genetic code reprogramming into particularly the last of these was emphasized. Finally, an overview was presented of biorthogonal reactions potentially compatible with these display technologies. This started with some classic reactions on standard amino acids and then went further on the reactions of nonstandard amino acids, emphasizing the synergy of this approach with genetic code reprogramming as a means to access new chemical handles.

Chapter 2 describes a new reprogramming strategy for vacant codon creation in the context of genetic code reprogramming in *Escherichia coli*-derived *in vitro* translation, with suppression of formylation being sufficient to liberate the initiation codon. We showed that this was achievable by removing the formyl donor or by inhibiting the enzyme MTF. Initially, we found that omitting the formyl donor in translation could give a better ratio of reprogrammed initiation product to methionine analogue incorporation. Further optimizing the amount of acylated tRNA, we successfully got clean translation

of initiation-reprogrammed peptides. In addition, we found that adding an inhibitor of the formyl transferase could give the same result. To further demonstrate the scope of reprogramming of the genetic code on this vacated codon, we used amino acids acylated with non-canonical functional groups such as reactive handles, a fluorophore, and biotin. All of these applications showed good yields and product/byproduct ratios. This initiation reprogramming was shown to be compatible with translation of downstream AUG codons by methionine, or exploiting mis-acylation of near cognates by MetRS to incorporate clickable handles to provide convenient access to doubly modified proteins. After analyzing the byproduct peaks, we found evidence of a truncated product which could be caused by unacylated tRNA_{ini} initiating translation. Addition of excess tRNA_{ini} gave this initiator-truncated form as the only product found, providing a convenient route to *N*-terminally processed peptides and proteins without additional enzymatic steps. Finally, we tested translation of a small panel of proteins with biotinyl-phenylalanine as the initiating amino acid and azidohomoalanine reprogramming of the elongation AUG codons using our optimized conditions with a formyl transferase inhibitor. Reaction with DBCO-Sulfo-Cy3 verified the presence of azidohomoalanine, and pull-down by magnetic streptavidin beads verified the orthogonal initiation reprogramming. Together the results show our method to be suited for applications such as generating immobilized proteins for protein interaction experiments, conjugate vaccines carrying multiple epitopes, and generating proteins with multiple fluorophores for conformational studies. Overall, the method we report here provides rapid access to *N*-terminal protein reprogramming with a broad range of chemical functionality using simple molecular biology techniques and easily accessible reagents.

Chapter 3 details a selection against E-selectin using genetic code reprogramming in mRNA display, as well as initial assays of hits arising from this. To increase the targeting and binding affinity of macrocyclic peptide libraries to lectins, we incorporated several sugar-like binding elements into the library. We synthesized several such binding factors as activated amino acids for flexizyme-mediated aminoacylation: phenylalanine methylsulfonate, cysteine-diol and pentafluoro-phenylalanine. Subsequently we verified incorporation of these in amino acyl tRNA and then peptides in reprogrammed *in vitro* translation. After optimization of translation, we used these in a selection campaign against E-selectin starting from a library of random peptides with total length of 17 amino acid and cyclized by head to sidechain thioether formation. After 5 rounds of selection we were able to observe clear enrichment of binding sequences. These were amplified, purified and submitted

for high-throughput sequencing. Of note was that only the methyl sulfonate binding factor was found to have been enriched in these. Several potential binding sequences were picked out based on sequence conservation and enrichment, and pull-down experiments using clonally pure sequences in mRNA display were used to validate these. In these, competition with sugar showed only a small effect, but a known peptide ligand of E-selectin completely ablated binding and so these are likely binding to the same site. We subsequently synthesized some of the more promising peptides by SPPS and measured selectin binding by SPR. Several of these peptides, such as those coded D3 and L6, were found to have low nanomolar binding affinity to E-selectin and a very slow off rate. By synthesis and testing of analogues, also in SPR, we found that sulfonation indeed increased the binding affinity, even though the literature suggests that mutation of the natural ligand PSGL-1 to remove sulfation does not influence the binding to E-selectin. These binders show promise as lead compounds for application in inhibition of E-selectin in its role as an immune cell recognition protein or as targeting factors for selective delivery of chemotherapy.

Chapter 4 demonstrates the utility of chemical post-translational modification in mRNA display to access new structures by employing CLIPS/CuAAc chemistry. CLIPS/CuAAc chemistry has previously been employed for synthetic peptides to form a uniquely well defined tricyclic architecture. What was missing was a method to discover new sequences that would exploit this structure, and so we sought to apply it in mRNA display. This was initially tested using synthetic peptide in inactive *in vitro* translation solution and subsequently with the same sequence translated, with reactions followed by UPLC-MS. We next sought to verify the methodology in the context of mRNA display using two probes, selective for free thiols and azides, to extract any unreacted peptide-mRNA-cDNA sample and calculated the conversion by RT-qPCR. After some optimization, we designed a peptide library consisting of three random regions of four amino acids each and flanked by two cysteines in the core and two azidohomoalanine residues at the termini. This single peptide sequence was combined with two different scaffolds to further increase structural diversity and both libraries were then panned against 4 target proteins for a total of 8 selections. These were chosen to cover a range of more reliable and more challenging targets, including a synthetic all-D stereochemistry protein. After selection, amplification and high-throughput sequencing, strong candidates for potent binders of two of these proteins were found while the remaining two proteins showed some potential candidates but less convincing enrichment. For one of these targets a binding motif is known, and this was indeed found in the sequencing results. Furthermore, unique sequences

were found in all cases with the two different scaffolds, suggesting they are important for peptide binding in each case.

Appendices

Nederlandse Samenvatting

List of Publications

Acknowledgement

Curriculum Vitae

Nederlandse Samenvatting

Dit proefschrift beschrijft het werk aan een *in vitro* translatie techniek en de toepassing daarvan in bacteriële initiatie reprogrammering van peptiden en eiwitten, mRNA display met genetische code reprogrammering en post-translationele modificaties. Tijdens experimenten met als doel het inbouwen van methionine mimics, observeerden wij dat suppressie van formylatie in bacteriële translatie een potentiële manier was voor het creëren van een vacant initiatie site voor reprogrammering zonder daarbij het elongation codon AUG te beïnvloeden. Deze observatie vormde de basis voor het werk dat wordt beschreven in hoofdstuk 2 van dit proefschrift. Door middel van genetische code reprogrammering in mRNA display kunnen wij additionele binding factoren en handvatten introduceren voor het post-translationeel modificeren van peptiden tot macro-cyclische of unieke tricyclische peptiden. Op zichzelf is deze screening techniek al een robuuste en snelle manier voor het vinden van sterke peptide binders van een target eiwit, maar verdere modificaties beschreven in hoofdstuk 3 en 4 maken het verder mogelijk om een library te creëren met nog diversere structuren.

Hoofdstuk 1 geeft een overzicht van de huidige theoretische achtergrond voor de opvolgende hoofdstukken. Hierin wordt een overzicht van bacteriële translatie gegeven, ingaand op de structurele en dynamische aspecten voor de precieze controle over het initiatie proces. Er wordt uitgelegd hoe bacteriën specifiek fMet-tRNA_{ini} overdragen op de initiatie site maar niet op de elongatie site, verder wordt behandeld hoe elongatie en terminatie van translatie werken in het ribosoom. Dit wordt gevolgd door een overzicht van verschillende selectie technieken zoals yeast, phage en mRNA display, waarin met name de laatste wordt behandeld met betrekking tot gebruik van genetische code reprogrammering. Tenslotte worden bioorthogonale reacties geïntroduceerd die potentieel gebruikt kunnen worden in deze display technieken. Hierin wordt gestart met reacties die gebruikt worden op de natuurlijke aminozuren en gevolgd door reacties op onnatuurlijke aminozuren, met name de synergie van genetische code reprogrammering en deze reacties voor het verkrijgen van nieuwe chemische handvatten in eiwitten wordt behandeld.

Hoofdstuk 2 beschrijft een nieuwe reprogrammering strategie voor het creëren van vacant codons in de context van een genetische code reprogrammering van E.coli-afgeleide translatiemix, waarin wordt getoond dat suppressie van formylatie voldoende is voor het vrijmaken van het initiatie codon. We laten zien dat dit mogelijk is door het verwijderen van de formyl-donor of het inhiberen van het enzym MTF. Als

eerste werd bevonden dat het weglaten van de formyl-donor in de translatie, een betere ratio tussen gewenst gereprogrammeerd initiatie product en ongewenst methionine analoog product gaf. Door het verder optimaliseren van de geacetyleerde tRNA concentratie, kon er een zuivere en heldere piek van het gereprogrammeerde initiatie product worden verkregen. Hiernaast vonden we dat door toevoeging van een inhibitor van het formyl transferase enzym, hetzelfde resultaat bereikt kan worden. Voor het verder tonen van de bruikbaarheid van deze techniek, hebben we gebruik gemaakt van geacetyleerde aminozuren met niet-natuurlijke functionele groepen zoals reactieve handvatten, een fluorofoor en biotine. Al deze voorbeelden resulteerden in een goede opbrengst van product als ook een goede ratio tussen product en bijproduct. We laten zien dat deze initiatie reprogrammering compatibel is met translatie van downstream AUG codons door methionine, alsook het gebruik maken van de misincorporatie van aminozuren lijkend op methionine door MetRS waardoor click groepen geïntroduceerd kunnen worden voor het verkrijgen van dubbel-gemodificeerde eiwitten. Na analyse van de bijproducten van deze aanpak, werd de aanwezigheid van initiator-verkort product gevonden welke waarschijnlijk gevormd was door initiatie van translatie met behulp van ongeacetyleerd tRNA_{ini}. Het toevoegen van een overmaat van tRNA_{ini} leidde tot uitsluitend de formatie van dit initiator-verkort product, die gebruikt kan worden voor de directe productie van vrije N-terminale peptiden en eiwitten zonder benodigde additionele enzymatische stappen. Tenslotte, is er een klein panel van eiwitten met initiatie van biotinylnyl fenylalanine en azidohomoalanine op het elongatie AUG codon geproduceerd gebruikmakend van onze geoptimaliseerde condities met formyl transferase inhibitor. Doormiddel van reactie met DBCO-Sulfo-Cy3 werd de aanwezigheid van azidohomoalanine aangetoond, gevolgd door het aantonen van biotinylnyl fenylalanine door gebruik van Pull-down met magnetische streptavidine beads. De resultaten van onze techniek tonen aan dat deze ingezet kan worden voor het produceren van geïmmobiliseerde eiwitten voor eiwit-interactie experimenten, alsook geconjugeerde vaccins geladen met meerdere epitopen en het verkrijgen van eiwit met meerdere fluoroforen voor conformationele studies. In zijn geheel tonen wij hier een techniek die de snelle en makkelijk uitvoerbare productie mogelijk maakt van N-terminaal gereprogrameerde eiwitten uitgerust met een breed bereik aan chemische functionaliteiten, gebruik makend van simpele moleculaire biologische technieken en makkelijk verkrijgbare reagentia.

Hoofdstuk 3 beschrijft een peptide selectie tegen E-Selectin waarbij gebruik wordt gemaakt van genetische code reprogrammering in mRNA display, alsook initiële

proeven met hits voortkomend uit deze selectie. Er is gebruik gemaakt van het inbouwen van een aantal suiker-nabootsende elementen in de library met als doel de kans op het vinden van sterk-bindende en actieve peptiden tegen deze lectinen te vergroten. Hiervoor was het nodig een aantal van deze suiker-nabootsende elementen te synthetiseren als geactiveerde aminozuren om gebruik te kunnen maken van flexizyme-mediated aminoacylering. Fenylalanine methylsulfonate, cysteïne-diol en pentafluorofenylalanine zijn in deze geactiveerde vorm gesynthetiseerd en vervolgens getest in het gereprogrammeerde in vitro translatie systeem voor het verkrijgen van aminoacyl tRNA en peptiden met deze aminozuren. Na het optimaliseren van de condities voor deze aminozuren in de translatie, is er een selectie uitgevoerd tegen E-Selectin met een library van willekeurige peptiden met een lengte van 15 variabele aminozuren en een head to sidechain cyclisatie. Na 5 rondes van selectie, was er sprake van een duidelijke verrijking van bindende peptiden in de library. Deze zijn vervolgens geamplificeerd, gezuiverd en opgestuurd voor high-throughput sequencing. De sequencing resultaten lieten zien dat alleen het methylsulfonate aminozuur aanwezig was in de verrijkte library. Er zijn toen een aantal potentiële bindende sequenties gekozen op basis van verrijking en sequentie conservation, vervolgens is de binding van deze peptiden gevalideerd door gebruik te maken van een pull-down experiment met klonaal zuiver sequenties in mRNA display. Alhoewel competitie met een bekende suiker ligand van E-Selectin maar een klein effect liet zien, was er een duidelijk effect zichtbaar bij competitie met een peptide ligand van het eiwit. Hieruit werd geconcludeerd dat binding van de geselecteerde peptiden hoogstwaarschijnlijk plaatsvindt op dezelfde plek als dit peptide ligand. Vervolgens hebben we de meest belovende peptiden gesynthetiseerd met SPPS en de binding affiniteit bepaald door middel van SPR. Meerdere peptiden, bijvoorbeeld degene met de code D3 en L6, vertoonden een laag nanomolaire binding affiniteit voor E-selectin met een zeer langzame off rate. Door het synthetiseren en testen van analogen in SPR, werd bevonden dat sulfatie de binding affiniteit verhoogt in tegenspraak met de literatuur waarin wordt gesuggereerd dat een sulfatie-verliezende mutatie in het natuurlijke ligand PSGL-1 niet leidt tot effect op binding. Deze geselecteerde peptide binders zijn veelbelovende potentiële lead compounds die toegepast kunnen worden voor inhibitie van E-selectin in zijn rol als immuun cell herkenning eiwit of als targeting factor voor doelgerichte chemotherapie.

Hoofdstuk 4 laat de bruikbaarheid zien van de mogelijkheid tot uitvoeren van chemische post-translationele modificaties in mRNA display, leidend tot nieuwe diverse structuren met behulp van CLIPS/CuAAc chemie. Het is bekend dat door

middel van CLIPS/CuAAC chemie het mogelijk is om unieke gedefinieerde tricyclische peptiden te vormen. Het ontbreekt echter aan een goede methode voor het vinden van nieuwe sequenties die gebruik maken van deze unieke structuur. mRNA display is bij uitstek de methode die dit mogelijk kan maken en dus zijn we begonnen met het testen van deze aanpak. Deze reacties werden eerst getest in een opzet van synthetisch peptide toegevoegd aan een inactief translatie systeem met daarnaast een translatie van het peptide met dezelfde sequentie, gevolgd door analyse met behulp van UPLC-MS. In de volgende stap om de methode in context van mRNA display te verifiëren werd er gebruik gemaakt van twee probes, één selectief reagerend met thiolen en de ander voor azides, waarmee het niet gereageerde deel van beide reacties op het peptide-mRNA-cDNA complex geanalyseerd kon worden doormiddel van Pull-down gevolgd door RT-qPCR analyse. Na het optimaliseren van reactie condities hebben we een peptide library ontworpen, die 3 random regions met elk 4 willekeurige aminozuren bevat. Deze random regions zitten ingesloten tussen twee cysteïnes in de kern van de structuur en twee azidohomoalanines aan de terminus. Om de diversiteit te vergoten zijn twee verschillende scaffolds gebruikt op de ontworpen peptide library, resulterend in twee aparte librarys die gebruikt zijn voor het screenen tegen 4 verschillende target eiwitten. Deze 4 eiwitten zijn gekozen aangezien ze een goede weergave geven van betrouwbare en uitdagende targets. Een uitdagende target in deze selectie eiwitten is een synthetisch eiwit volledige bestaande uit D-aminozuren. Na de selectie, amplificatie en high-throughput sequencing, zijn er potentieel sterke binders gevonden voor twee van deze eiwitten. Voor een van deze eiwit targets is het binding motief bekend, welke ook gevonden is het sequencing resultaat. Verder zijn er unieke sequenties gevonden voor beide scaffolds, die erop wijzen dat ze belangrijk zijn voor specifieke binding met de eiwit target.

List of Publications

1. **Liu, M.**; Thijssen, V., Jongkees, S.A.K. (2020) *Angew Chem Int Ed*, 59: 21870
2. Johansen-Leete, J.; Passioura, T.; Foster, S.; Bhusal, R. P.; Ford, D.; **Liu, M.**; Jongkees, S. A. K.; Suga, H.; Stone, M. J.; Payne, R. J. (2020) *J Am Chem Soc*, 142:9141
3. Goldbach, L.; Vermeulen, B.J.A.; Caner, S.; **Liu, M.**; Tysoe, C.; van Gijzel, L.; Yoshisada, R.; Trellet, M.; van Ingen, H.; Brayer, G.D.; Bonvin, A.M.J.J.; Jongkees, S.A.K. (2019) *ACS Chem Biol*, 14:1751
4. **Liu, M.**; Niu, Y; Wu, Y-F; Ye, X-S. (2016) *Org Lett*, 18:1836

Acknowledgement

Over writing, it finally came to this part. When I looked back into the past 4 years, everything seemed to happen a few moments ago. I felt like it's just yesterday that the airplane landed and I put my footprint on this new earth for first time. Along the journey, so many people have helped me. Their support is behind this thesis. Herein, I would like to spend some time to express my gratitude to them even though words couldn't describe all my thankfulness.

Hi Seino, thank you for everything! The first sight of you depressed me since I have to look up :) If I have cervical spine problems, it must be your fault. As for research, I really thank you for suggestions and discussions about the problems I met during the experiments; for guiding me on how to carry chemical biology experiments; for teaching me some basic knowledge about translation and selection; for showing me how to use SPPS and LC-MS; for sharing your experience in the research; for helping me practice my presentation and skills for speeches; for *et al*... Except for these, I really admire your enthusiasm for research; creativity for ideas and management of lab affiliated stuff; outstanding skills on writing and logic for description. Even though I could not climb that high level, I still improved myself a lot. Finally, I really enjoy our group atmosphere such as group dinner and some games. It made everybody know each other well. However, I still remembered: you said that you can tolerate strong chili while suffering from my spicy chicken. Then I came to know, you may have misunderstanding of Chinese chili. All in all, Seino, thanks!

I also would like to thank Prof. Geert-Jan Boons gave me the chance to finish my PhD research in chemical biology and drug discovery department. Even though I majored in Pharmaceutical science for my bachelor and master program, I have little experience on biochemical studies. Without your guide and supervision, I think I couldn't finish these projects. I am very grateful for the discussion and meeting, at beginning of my project, it's a little tough, but I learned a lot as you shared your precious experience in chemistry and biology.

Vito, thank you! As colleague for about 4 years, you helped me a lot in the academic research. You are always helpful when somebody needs help. And some of our collaboration is showed in thesis, mainly in chapter 2. I greatly appreciate your support on these projects. Except for lab work, you also help me a lot in getting used to Dutch

life style, especially on tax and allowance problems. You said that you are from south of Netherlands, and I am also from south of China. Both of us are not high, maybe this is the discrimination from planet on southern people.

Many thanks to other Seino's group member, Ryoji, Rick, Lisa, Cerissa, Helena, Lin, Miha, David, Lloyd, Najoua, Avand, Wendy. Thank you all for the help at lab and also the happiness off working hours. Every group dinner pictures freeze the happy moments. Thank you guys for the good-bye movie and it really moved me a lot.

I would also like to thank all the CBDD group members. Thanks, John! With your technical support and peptides synthesis experience, my research got easier. Javier, thank you for the guide on using UPLC-MS and LC-MS, these robust machines helped me a lot on my project. Gerlof, thanks for sharing abundant protein experience. Arwin, thanks for your solid support in chemical lab. Margreet, many thanks for your help on protein works and microarray measurements. To other dedicated members of CBDD: Roland, Tom, Dirk, Ed, Linda, Rob, Justyna, Roosmarijn, Nino, Luca, Julia, Victor, Nuria, Diksha, Frederik, Apoorva, Ivan, Mehman, Rosanne, Ingrid, Tim, Yvette, Pieter, Gael, Hanna, Cindy, Ilhan, Enrico, Pouya. Thank you for all the help during every stage of my PhD life.

I would also like to some friends from OCC group, ICC group, pharmaceuticals group. They provide helpful suggestions and lab support on my projects. I really appreciated all the tips and help in my past 4 years.

接下来, 请允许我用中文表达我的谢意。首先, 感谢凤爪在过去四年的陪伴。已经十余年, 从相遇, 相识, 相知到相伴, 过去的种种犹如回马灯一般在脑海浮现。尽管内心的感情无比丰富, 可临落笔, 写下的只言片语却略显单薄。虽然并不想将一份简单的致谢写成一份告白信, 但无论如何, 也希望这短短的一段话能传达出自己的部分感情。在过去四年里, 我能想起的回忆都是快乐, 在巴黎迪士尼看烟花, 在布拉格的广场吃冰淇淋面包, 在冰岛空旷的大地上奔驰, 在海牙的中国大使馆领证……我很庆幸, 我们能记录大多数这些片段。当老了的时候, 看到照片, 就能回忆起那一瞬间的感受, 那就够了。

感谢岛爷, 2016 年 10 月 10 日, 没有你的接机, 我们估计会迷路好一阵子。在实验中, 岛爷也帮助了我许多, 让我快速地熟悉实验室。最近虽然不能多多聚餐, 但是在过往的几年中, 每次聚餐都是非常快乐, 只是, 岛爷你的 5L 啤酒承诺什么时候兑现。我们不接受养乐多, 岛嫂那儿我们可以帮你拦着。一起加入 CBDD 组的几位小伙伴, 宪轲, 度伸, 很高兴和你们在同一个组生活学习, 感谢你们在实验中的帮助, 也很怀念那些一起

聚餐的日子。虽然在荷兰，我们的娱乐项目不多，但是我们的聚餐还非常丰富的。同样的，也感谢刘燕燕，你的存在帮助我们了解了好多八卦（doge）。韦萱，乐檬，同样也感谢你们在组里对我的帮助。还有已经毕业的浩哥，欣爷，感谢你们在我们刚来荷兰生活的时候的亲切指导教诲，并给出好多实用建议，下次回国找你们喝酒，希望这一天不会太久。文静师姐，光允师兄，永志师兄，同样也感谢你们在实验和生活中对我的帮助。OCC 组李婧师姐，健明师兄，现在你们已经顺利学成毕业回国，但依然感谢你们在过去几年借过的试剂，当然还有忘不了的聚餐。当然，还有其他几位 DDW 的小伙伴，相洁，凡石，陈婧师姐，杨逸，你们也在我的实验中帮助良多。

一起打羽毛球的小伙伴呀，金学长，斌哥，超哥，涛哥，玉珑师姐，飞哥，杨敏，鑫姐，栋哥，亮哥，禹哥，孙季，方莉师姐，小莉你们最近还好吗？在过去四年，我曾经有过的最艰难的那段时间，是羽毛球和你们，让我在生活中收获更多动力。我送走了一位又一位的师兄师姐，每一次的留影，默默看着上面的熟悉面孔变少，而现在，我即将成为其中一员。远方的朋友，希望有机会我们还能一起打球。

同样的，我也很想感谢以前在叶组一起奋斗的小伙伴们，老罗，小兰姐，宁宁师兄，硕哥，成程姐，肖老师，孙校长…怀念以前一起出门烧烤的日子。除此之外，我也要感谢叶组的各位老师对我的教导，感谢叶老师，牛老师，吴老师，熊老师和李勤老师。

邹绵成，我想单独把你提出来放在一段。作为 6 年的室友，真的太熟悉了。但是还是得说，前年冰岛之行真的辛苦你啦，今年如果有机会希望能继续帮我们开车。

最后，我要把我的感谢献给在背后默默支持我的家人们，我的爸爸，妈妈，爷爷，奶奶，外公，外婆，二孃，姑爷，舅舅，舅妈，二姨，二姑爷，三爸，三妈，么爸，么妈。有了你们的支持和教育，我才会在这遥远的 8000 公里之外的陌生土地有今日的成果。这篇论文，是我以往四年研究成果的展示，也是对你们的一份感激之情。再次感谢你们的理解和支持！

Curriculum Vitae

Minglong Liu was born on May 18th, 1992 in Tongliang, Chongqing, China. He grew up to 7 years old and then moved to Jiangjin, another town in Chongqing province, besides Yangzi River, which is the longest river in China. He attended primary school, middle school and primary school there. In 2010, after graduated from high school, he was enrolled in Peking University where he majored in Pharmaceutical science for a 6-year program. He obtained his bachelor degree in 2014 under the supervision of Prof. Xin-Shan Ye. After finished his master program under the supervision of Prof. Xin-Shan Ye at 2016, he moved to Netherlands to pursue a PhD program with the support from Chinese Scholarship Council (CSC) at the same year. During his 4-years study at chemical biology and drug discovery department, he mainly focused on the genetic code reprogramming and mRNA display. The results are described in this thesis.

