

Aqueous Synthetic Methods and their Applications in DNA-Encoded Chemical Libraries

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Cedric Joseph Stress

aus Nenzlingen, Basel-Landschaft

Basel, 2019

**Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von**

Prof. Dr. Dennis Gillingham

Prof. Dr. Christof Sparr

Basel, den 21. Mai 2019

Prof. Dr. Martin Spiess
Dekan der Philosophisch-
Naturwissenschaftlichen
Fakultät

Abstract

Water is the basis for all living organisms on Earth and most biochemical processes proceed in aqueous environments. To study biological systems, chemists have developed numerous procedures to perform chemical transformations in the presence of water. However, there is still a limited scope of reactions that proceed efficiently and reliably under physiological conditions. The search for new techniques to enable selective and reliable modifications of biomolecules and small molecules alike, has attracted the attention of many researchers in academia and industry. Such "water-friendly" reactions are highly desired for different areas of biochemical research, such as bioconjugation techniques and drug discovery procedures.

Schiff base formation is a workhorse in bioconjugation science, although the reactions proceed slowly under neutral conditions without catalysts. We investigated oxime and hydrazone formations of *ortho*-boronate carbonyl compounds with hydroxylamines and hydrazines. The boronic acid was found to strongly increase the Schiff base formation rate, which enabled the fluorescent labeling of antibodies. Hydrazones with an adjacent boronic acid group undergo a secondary cyclization reaction to form a stable, aromatic boron-heterocycle (BIQ, 4,3-borazaroisquinoline). Upon modulation of the electronic properties of this boron-heterocycle with different substituents, we developed a blue fluorophore that formed upon cyclization to the BIQ product (**Figure 1a**).

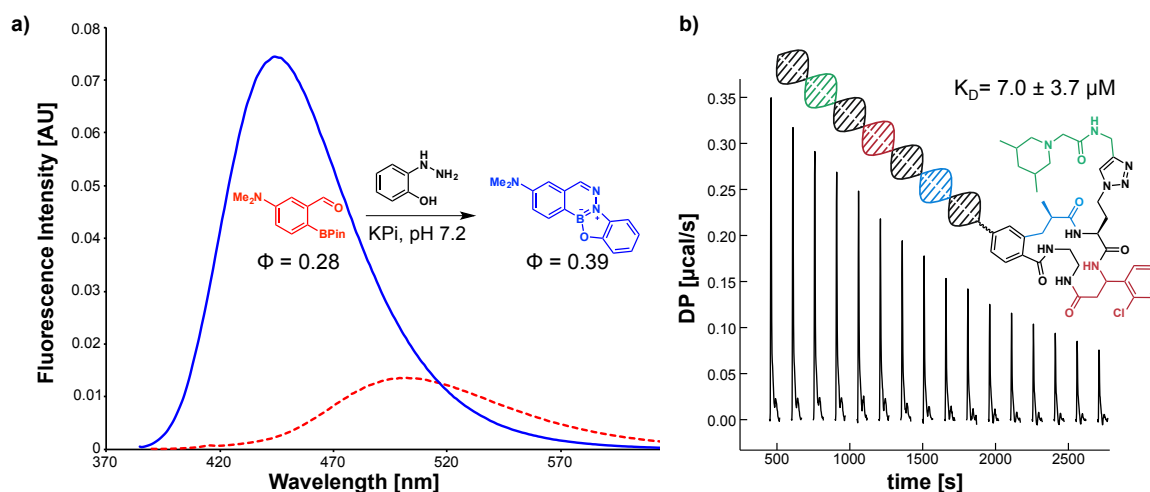


Figure 1. a) Rate-enhanced Schiff base formation of a hydrazone with an *ortho*-boronic acid aldehyde to obtain a fluorogenic boron-heterocycle. b) Generation of a DNA-encoded macrocycle library for the identification of protein binders. The shown macrocycle was determined to bind AGP with low micromolar affinity.

Based on the results from the bioconjugation techniques, we developed a library of macrocycles, in which every compound was attached to a DNA strand containing the information about the macrocycle structure. Despite initial difficulties, we obtained a 1.4 million member DNA-encoded library of natural product-like macrocycles with high scaffold diversity. After thorough analysis of the library properties, we screened the encoded macrocycle collection against three human proteins. Several hits were found and resynthesized without DNA tag. Binding affinities to the target proteins were evaluated by biophysical techniques. Differential scanning fluorimetry enabled the parallel screening of the hit compounds, giving a qualitative measurement for the protein affinities. Isothermal titration calorimetry yielded quantitative dissociation rate constants for the most promising compounds. We discovered a novel macrocyclic ligand for α -1-acid glycoprotein (AGP) with a low micromolar affinity (**Figure 1b**), which holds promise for the development of new drug candidates. The developed encoded library methodology was shown to be well suited for early stage drug discovery.

Table of Contents

List of Abbreviations	VII
<i>Part I: Boron-Assisted Oxime, Hydrazone and BIQ Formation</i>	1
1. Introduction to Bioconjugations with α-Effect Nucleophiles	1
1.1 Bioconjugations	1
1.2 α -Effect Nucleophile Condensations	10
1.3 Fluorescent Molecules in Bioconjugation	13
1.4 Project Description	14
2. Boron-Assisted Oxime Condensations	15
2.1 Characteristics of Boron-Assisted Oxime Condensations	15
2.2 Bioorthogonal Labeling of Immunoglobulin G	19
3. Boron-Assisted Hydrazone and BIQ Formation	21
3.1 Characteristics of Boron-Assisted Hydrazone and BIQ Formation	21
3.2 BIQ Syntheses and Fluorescence Property Analysis	25
4. Conclusions and Perspective	27
5. Experimental Part Oxime, Hydrazone, BIQ	29
5.1 General Information	29
5.2 Compound Synthesis	31
5.3 Stability and Reversibility of Boron-Assisted Oxime Formations	44
5.4 HPLC Assays on Oxime Formation Mechanism	50
5.5 Fluorescent Labeling of Immunoglobulin G	54
5.6 NMR Assay for BIQ Formation	55
5.7 NMR Assay for BIQ Reversibility	57
5.8 pH Effects on BIQ Formation	57
5.9 Effects of Glutathione and Human Serum on BIQ Formation	58
5.10 Hydrazone Reversibility Assay	58
5.11 Fluorescence Measurements	60
<i>Part II: DNA-Encoded Chemical Libraries</i>	61
6. Introduction to DNA-Encoded Chemical Libraries	61
6.1 Combinatorial Chemistry and High-Throughput Screening	61
6.2 Phage Display Libraries	63
6.3 DNA-Encoded Chemical Libraries	64
6.4 Chemical Requirements and Hit Validation	68
6.5 Description of our Natural Product-like DNA-Encoded Macrocyclic Library	71
7. Assembly of a DNA-Encoded Macrocyclic Library	77
7.1 Chemical Library Designs	77

7.2 Library Validation off-DNA.....	82
7.3 Library Validation on-DNA.....	83
7.4 Assembly of the Macrocyclic Library.....	84
8. Properties Evaluation and DNA Damage Analysis of the DEML	99
8.1 Introduction to Physicochemical Properties of Macrocycles	99
8.2 Evaluation of the Physicochemical Properties of the DEML	102
8.3 Quantitative PCR (qPCR).....	104
9. Protein Affinity Selections, Sequencing and Data Analysis	107
9.1 Protein Affinity Selections.....	107
9.2 Next Generation Sequencing (NGS).....	108
9.3 NGS Results Analysis	112
10. Macrocyclic Resynthesis and Protein Binding Affinity Measurements.....	117
10.1 Resynthesis of the Macrocyclics.....	117
10.2 Differential Scanning Fluorimetry (DSF).....	118
10.3 Isothermal Titration Calorimetry (ITC).....	120
11. Conclusions and Perspective	123
12. DNA-Encoded Small Molecule Library	127
12.1 Introduction.....	127
12.2 Design and Assembly of the Small Molecule Library	127
12.3 Perspective.....	133
13. Experimental Part DNA-Encoded Libraries	135
13.1 General Information.....	135
13.2 LC-MS Analysis of DNA-tagged Small Molecules.....	138
13.3 Off-DNA Macrocyclic Synthesis	138
13.4 Test Macrocyclic Synthesis on DNA.....	145
13.5 DNA-Encoded Library Assembly, Synthetic Procedures.....	150
13.6. Quantitative PCR (qPCR) of Chemical Reactions on DNA	160
13.7 Protein Biotinylation of HSA and AGP.....	162
13.8 DEML Protein Target Selections	163
13.9 PCR Amplification of Eluted DNA.....	165
13.10 Next Generation Sequencing (NGS)	168
13.11 Macrocyclic Resynthesis Procedures	170
13.12 Differential Scanning Fluorimetry (DSF).....	192
13.13 Isothermal Titration Calorimetry (ITC).....	192
13.14 Building Block Synthesis	196
13.15 Building Block Encoding and Virtual Library Assembly.....	224
13.16 Synthesis of the Small Molecule Library Scaffold	225
13.17 Validation of the Small Molecule Library Chemistry	226

13.18 Small Molecule Library Assembly and Building Block Validations	230
14. References	241
15. Acknowledgement	247
<i>Part III: Appendix</i>	249
16. Building Blocks	249
16.1 Diversity Element 1	249
16.2 Diversity Element 2	249
16.3 Diversity Element 3	252

List of Abbreviations

2-FPBA	2-Formylphenylboronic acid	DIFO3	Difluorocyclooctyne-3
2-FPTFB	2-Formylphenyltrifluoroborate	DIPEA	<i>N,N,N</i> -Diisopropylethylamine
AA	Amino acid	DKP	Diketopiperazine
ACN	Acetonitrile	DMAP	4-Dimethylaminopyridine
ADME	Absorption, distribution, metabolism, excretion	DMF	Dimethylformamide
AGP	α -1-Acid glycoprotein	DMP	Dess-Martin periodinane
AlogP	Atom-based logarithm of the partition coefficient	DMSO	Dimethylsulfoxide
APS	Ammonium persulfate	DMTMM	4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium
ATP	Adenosine triphosphate	DNA	Deoxyribonucleic acid
BAW	Butanol-acetic acid-water	DNP	2,4-Dinitrophenylhydrazine
BB	Building block	dNTP	Deoxynucleoside triphosphate
BCN	Bicyclononyne	DP	Differential potential
BIQ	4,3-Borazaroisoquinoline	DSF	Differential scanning fluorimetry
BME	2-Mercaptoethanol	DTT	Dithiothreitol
Boc	<i>tert</i> -Butyloxycarbonyl	EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
BODIPY	Boron-dipyrrromethene	EDCCL	Encoded dynamic combinatorial chemical library
bp	base pair	EDTA	<i>N,N,N',N'</i> -Ethylenediamine tetraacetic acid
BPin	Boronic acid pinacol ester	EEDQ	<i>N</i> -Ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline
CA9	Carbonic anhydrase 9	ESAC	Encoded self-assembling chemical library
CHA	Cyclohexylammonium	ESI-MS	Electrospray ionization mass spectrometry
CTA	Cellulose triacetate	FITC	Fluorescein isothiocyanate
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition	Fmoc	Fluorenylmethyloxycarbonyl
DBCO	Dibenzocyclooctyne-amine	GFP	Green fluorescent protein
DBU	1,8-Diazabicyclo(5.4.0)undec-7-ene	GSH	Glutathione
DCM	Dichloromethane	HATU	Hexafluorophosphate azabenzotriazole tetramethyluronium
DE-1	Diversity element 1	HBA	Hydrogen bond acceptor
DE-2	Diversity element 2	HBD	Hydrogen bond donor
DE-3	Diversity element 3		
DE-4	Diversity element 4		
DECL	DNA-encoded chemical library		
DEML	DNA-encoded macrocycle library		

HBTU	Hexafluorophosphate benzotriazole tetramethyl uronium	MPAA	4-Mercaptophenylacetic acid
HMBC	Heteronuclear multiple-bond correlation	mRNA	Messenger ribonucleic acid
HMPA	Hexamethylphosphoric triamide	MtBE	Methyl <i>tert</i> -butyl ether
HMQC	Heteronuclear multiple-quantum correlation	MW	Molecular weight
HOAt	1-Hydroxy-7-azabenzotriazole	MWCO	Molecular weight cut-off
HOSu	<i>N</i> -Hydroxysuccinimide	N/A	Not available
HPLC	High performance liquid chromatography	NADH	Nicotinamide adenine dinucleotide reduced form
HRMS	High-resolution mass spectrometry	NaOAsc	Sodium ascorbate
HS	Human serum	NaTG	Sodium thioglycolate
HSA	Human serum albumin	NGS	Next generation DNA sequencing
HTS	High-throughput screening	NHS	<i>N</i> -Hydroxysuccinimide
IBX	2-Iodoxybenzoic acid	NiAAC	Nickel-catalyzed azide-alkyne cycloaddition
IEDDA	Inverse electron-demand Diels-Alder reaction	NK3	Neurokinin 3
IgG	Human immunoglobulin G	NMM	<i>N</i> -Methylmorpholine
IPE	Diisopropyl ether	NMR	Nuclear magnetic resonance spectroscopy
ISS	<i>In situ</i> sequencing	NOE	Nuclear Overhauser effect
ITC	Isothermal titration calorimetry	Nosyl	2-Nitrophenylsulfonamide
KAT	Potassium acyltrifluoroborate	OBOC	One-bead-one-compound
K_D	Dissociation rate constant	PAGE	Polyacrylamide gel electrophoresis
KHMDS	Potassium hexamethyldisilylamide	PBS	Phosphate buffered saline
KPi	Potassium phosphate buffer	PCR	Polymerase chain reaction
LC	Liquid chromatography	PEG	Polyethyleneglycol
LC-MS	Liquid chromatography coupled to mass spectrometry	PNA	Peptide nucleic acid
LED	Light-emitting diode	PNK	Polynucleotide kinase
logP	Logarithm of the partition coefficient	POI	Protein of interest
MC	Macrocycle	PP	Polypropylene
MIDA	<i>N</i> -Methyliminodiacetic acid	pTsOH	<i>para</i> -Toluenesulfonic acid
MOPS	3-(<i>N</i> -Morpholino)propanesulfonic acid	QED	Quantitative estimation of drug-likeness
		qPCR	Quantitative polymerase chain reaction
		R_f	Retardation factor
		RNA	Ribonucleic acid

Ro5	Rule-of-five	<i>t</i>BuOH	<i>tert</i> -Butanol
RotB	Rotatable bond	TCEP	Tris(2-carboxyethyl)phosphine
RP-HPLC	Reversed phase high performance liquid chromatography	TE	TRIS-EDTA
RP-Silica	Reversed phase C ₁₈ silica	TEAA	Triethylammonium acetate
RT	Room temperature	TEMED	<i>N,N,N',N'</i> -Tetramethylethylene diamine
SD	Standard deviation	TFA	Trifluoroacetic acid
SDS	Sodium dodecylsulfate	TFL	Trifunctional linker
sEH	Soluble epoxide hydrolase	THF	Tetrahydrofuran
SM	Small molecule	TIC	Total ion chromatogram
SM	Starting material	TLC	Thin layer chromatography
smDEML	Small DNA-encoded macrocycle library	T_m	Melting temperature
SMILES	Simplified molecular input line entry specification	TMS	Trimethylsilyl
S_N2	Nucleophilic substitution	TMSP-<i>d</i>₄	3-(Trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid
SOLiD	Sequencing by oligonucleotide ligation and detection	TNKS1	Tankyrase-1
SPAAC	Strain-promoted azide-alkyne cycloaddition	TPSA	Topological polar surface area
T3P	Propanephosphonic acid anhydride	TRIS	Tris(hydroxymethyl)aminomethane
TBE	TRIS-Borate-EDTA buffer	TRITC	Tetramethylrhodamine isothiocyanate
TBTA	Tris((1-benzyl-4-triazolyl)methyl)amine	UPLC-MS	Ultraperformance liquid chromatography coupled to mass spectrometry
		UV	Ultraviolet

*"There is much pleasure to be
gained from useless knowledge."*

Bertrand Russel (1872 - 1970)

Part I: Boron-Assisted Oxime, Hydrazone and BIQ Formation

1. Introduction to Bioconjugations with α -Effect Nucleophiles

In this part, the influences of boron substituents on α -effect nucleophile condensations are presented and an insight into the reaction mechanism is given. The conducted assays led to successful labeling of an antibody and the development of a stable boron-containing turn-on fluorophore. Before the discussion of the experimental results, the most important concepts of bioconjugation are described, followed by an introduction to Schiff base condensations and fluorescent labeling of biomolecules.

1.1 Bioconjugations

Biomolecule Classes

The chemical cross-linking of biological molecules (e.g. proteins, peptides, carbohydrates, nucleic acids and lipids) with another moiety is termed bioconjugation.^[1] The coupling partners of these biomolecules include other biological molecules, synthetic polymers, dyes, drugs, small molecules and many other possibilities. The connection between the two moieties is established with a wide range of different chemical transformations. However, certain reactions or reaction types have proven to be more applicable and useful than others. These reactions must proceed under mild aqueous conditions and in the presence of the biomolecule's (native) functional groups. Bioconjugation reactions usually target nucleophilic and electrophilic moieties on the native biomolecule or synthetically incorporated reactive handles.^[1]

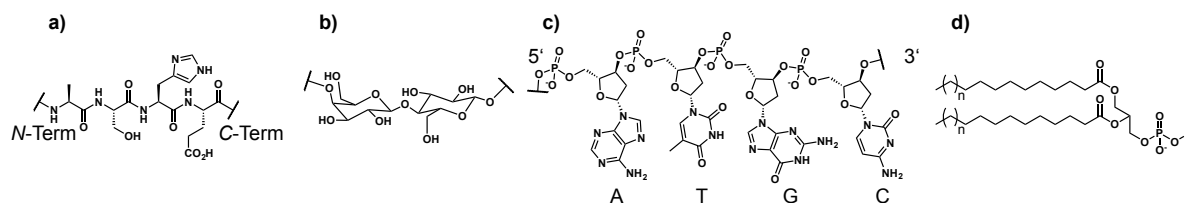


Figure 2. Overview of the most important biomolecule classes. **a)** Proteins and peptides. Shown is the primary structure of an Ala-Ser-His-Glu stretch. **b)** Carbohydrates, with the example of a galactose- β -1,4-glucose stretch. **c)** Oligonucleotides. Displayed is a DNA stretch with the four nucleobases adenine (A), thymine (T), guanine (G) and cytosine (C). RNA contains an additional 2'-hydroxyl group at the ribose unit and thymine is replaced by uracil (U). **d)** Lipids. The shown example is a phospholipid. Typical lengths for the fatty acids are $n = 6-14$ with varying degrees of unsaturation. R represents the common headgroups like ethanolamine, glycerol, choline, serine and inositol.

The most commonly used modification sites in protein/peptide bioconjugations are the lysine sidechain primary amine, the cysteine thiol and the *N*-terminus of the backbone (**Figure 2a**). Other amino acids displaying functionalities in their sidechains, such as carboxylic acid groups (Asp, Glu, C-terminus), alcohols (Ser, Thr, Tyr), other amine groups (Arg, Trp, His) or the methionine thioether, may be considered, but are not as commonly targeted.^[2] Modification sites of proteins may also be chemically introduced during peptide synthesis or by genome alteration to incorporate unnatural amino acid residues during protein expression.^[3-5]

The second class of biomolecules, carbohydrates (**Figure 2b**), natively displays many hydroxyl groups that are susceptible to chemical modification, although carbohydrates show poor nucleophilicity in aqueous solvent. Amino-sugars and sugar-acids, however, display highly reactive moieties amenable to chemical modifications. Such amino- and acid-saccharides are found in glycosaminoglycans like chondroitin and heparin.^[1] The repetitive manner of the sugar building blocks and the limited diversity of functional groups make selective modifications

challenging, which is the major drawback in carbohydrate bioconjugation chemistry. For carbohydrates with a reducing end, the aldehyde or ketone moiety (from the equilibrium between the linear carbonyl form and the cyclic hemiacetal form) allows modifications with amines, hydrazines or hydroxylamines by condensation reactions. Polysaccharides without a reducing end require chemical modification to introduce electrophilic moieties that undergo the bioconjugation reaction. Periodate oxidation transforms vicinal diols of the sugar moieties to aldehydes and cyanogen bromide is used for the generation of cyanate esters. Such electrophilic substituents are well suited to establish the desired chemical linkage with a nucleophilic coupling partner.^[1]

Oligonucleotides, such as DNA and RNA, are very important targets in biochemical research (**Figure 2c**). To study the different mechanisms and functions of these macromolecules, selective chemical modifications are highly desirable. However, nucleic acids are not as readily modified as proteins are. The chemical structure of oligonucleotides comprise a ribose-phosphate polymer backbone with attached nucleobases. DNA consists of deoxyribose sugars with two pyrimidine and two purine nucleobases (cytosine, thymine, adenine and guanine). RNA displays an additional 2'-hydroxyl group at the sugar backbone (ribose) with the same nucleobases as for DNA, except thymine, which is replaced by uracil (uracil lacks the exocyclic methyl group). DNA and RNA undergo Watson-Crick base-pairing and form double-stranded helical structures.^[6] Oligonucleotides usually display a phosphate group at the 5'-terminus and in the case of RNA a vicinal diol at the 3'-terminus. Termini modifications are achieved by carbodiimide chemistry with the phosphate group or periodate oxidation of the RNA's vicinal diol to generate aldehydes. Nucleobase modifications are possible as well, but often require single-stranded oligonucleotides. Cytosine is transformed to 6-sulfo-cytosine with sodium bisulfite for subsequent transamination reactions. The purine bases (adenine and guanine) can be brominated and used for proximate substitution reactions. Due to the repetitive manner of the nucleotides, site-specific modifications are extremely difficult to achieve. Enzymatic methods to introduce chemical modifications have been used for small-scale labeling with radioactive isotopes, biotin, fluorescent dyes or specific functional groups.^[1] The chemical modification of synthetic oligonucleotides is achieved more readily compared to native DNA and RNA because the alterations are introduced by standard phosphoramidite chemistry during solid-phase synthesis. Aminoalkyl and thioalkyl linkers as well as azido and alkyne groups are commonly employed as bioorthogonal modifications. Non-specific bioconjugation has been achieved in a few instances, for example with psoralen (a furanocoumarin) that intercalates with the double-stranded DNA and undergoes a [2+2] cycloaddition with pyrimidines (especially thymine) upon UV irradiation.^[1]

The last group of biomolecules discussed here are lipids (**Figure 2d**). The most commonly occurring lipids in Nature are phospholipids that consist of a glycerol backbone with two fatty acids and a phosphate group attached. The phosphate headgroup is often attached to other functional groups such as choline, glycerol, ethanolamine, serine or inositol. The lipids' modification site is usually the functionalized headgroup, because the fatty acids are mostly unreactive and not exposed to the aqueous media.^[1]

Bioconjugate Reactions with Amines

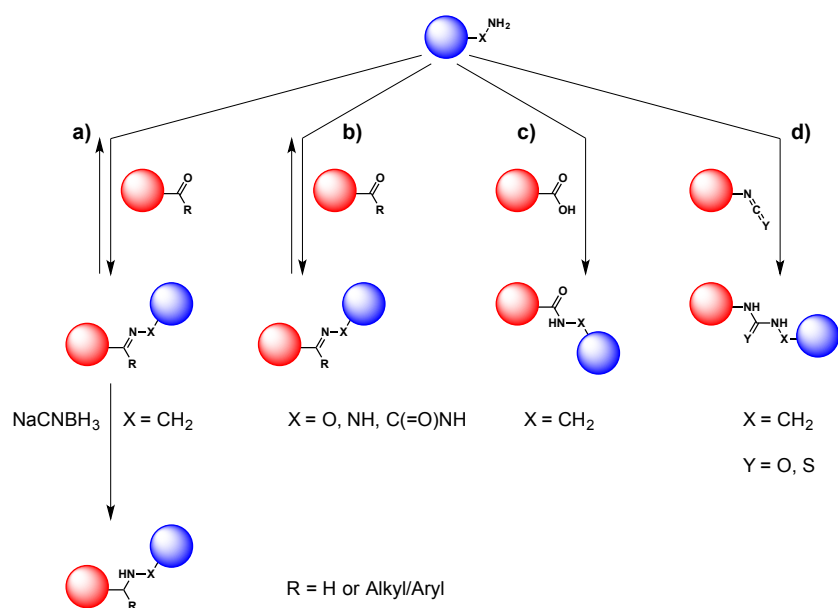


Figure 3. Different modification reactions of amines and α -effect amines. **a)** Condensation with an aldehyde ($\text{R} = \text{H}$) or ketone ($\text{R} = \text{alkyl/aryl}$) to an imine, which can be reduced to a secondary amine (reductive amination). **b)** Condensation reaction of a hydroxylamine ($\text{X} = \text{O}$) to an oxime or a hydrazine ($\text{X} = \text{NH}$)/hydrazide ($\text{X} = \text{C(=O)NH}$) to a hydrazone. **c)** Amide coupling of an amine with a carboxylic acid. For this transformation additional coupling reagents to activate the carboxylic acid are required. **d)** Condensation reaction of an amine with an isocyanate ($\text{X} = \text{O}$) or an isothiocyanate ($\text{X} = \text{S}$) to the urea or thiourea product.

Amines and their derivatives are the most versatile functional group occurring naturally in biological molecules. Especially proteins and peptides often display several sidechain amines from lysine residues that can be chemically modified.^[2,7] In the presence of an aldehyde or a ketone, amines will undergo a condensation reaction to the imine, also known as a Schiff base. However, imine formation is reversible and the hydrolytic stability of this linkage is low. For that reason, reductive amination procedures with sodium cyanoborohydride are employed to form stable, irreversible secondary amines (**Figure 3a**).

α -Effect nucleophiles, such as hydroxylamines, hydrazines and hydrazides, also undergo condensation reactions with carbonyl moieties to form the more stable oxime and hydrazone linkages (**Figure 3b**). These transformations are also reversible, but show good hydrolytic stability. This type of reactions will be discussed in more detail in Chapter 1.2.

A versatile and widespread modification is the acylation reaction of amines (**Figure 3c**). Reaction of a carboxylic acid with an amine generates an irreversible and hydrolytically stable amide. Since carboxylic acids are unreactive, an activating agent (coupling reagent) is required. Under physiological conditions, carbodiimides, such as EDC, are often employed in combination with an additive to generate a more hydrolytically stable activated ester. There is a large collection of coupling reagents and additives with various reaction properties. However, *N*-hydroxysuccinimide and its water-soluble, sulfonated variant are the most commonly used additives in combination with a carbodiimide for activated ester formation. Since succinimidyl esters are very slow in modifying alcohols, phenols and aromatic amines, they are the reagent of choice for selective lysine modifications. Many succinimidyl esters are commercially available or can be synthesized in a simple procedure, prior to bioconjugation.

Iso(thio)cyanates are another group of amine modifying reagents (**Figure 3d**). The formed urea or thiourea products show good hydrolytic stability and the reactions are usually irreversible. However, this substance class is less commonly used, because fewer iso(thio)cyanate

compounds are commercially available as well as their synthesis can be tricky. Fluorophores are often available as isothiocyanates such as fluorescein isothiocyanate (FITC) or tetramethylrhodamine isothiocyanate (TRITC).

There are many more amine-modifying reactions with other electrophilic groups, such as sulfonyl chlorides, epoxides or cyanate esters. However, these reactions often show low chemoselectivity for amines, hydrolytic stability and slow reaction rates.^[1]

Bioconjugate Reactions with Thiols

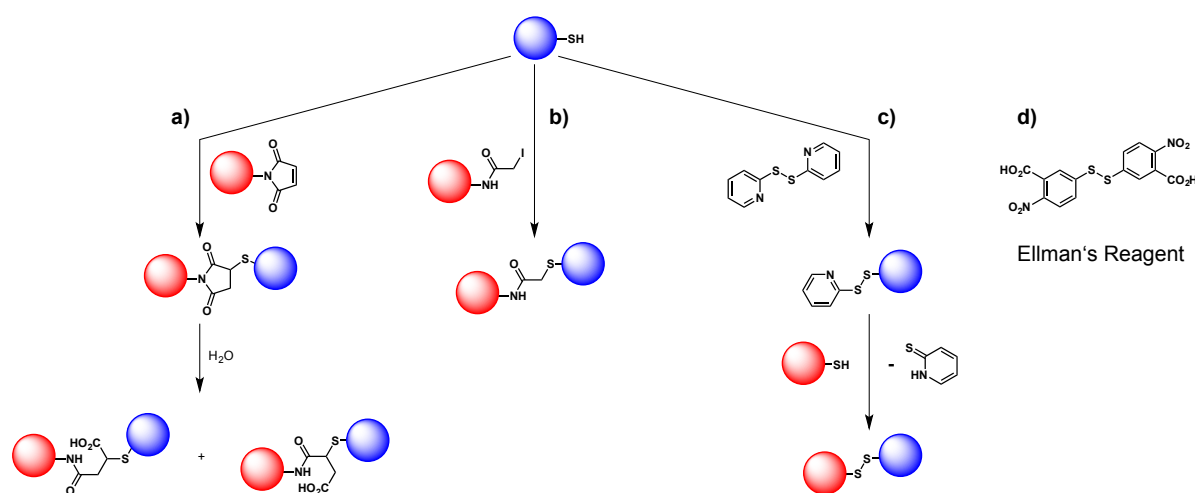


Figure 4. Modification reactions of thiols. **a)** Reaction of a thiol with the maleimide Michael acceptor. The thiosuccinimide moiety subsequently hydrolyses to a mixture of succinimic acid derivatives. **b)** Reaction of a thiol with iodoacetamide in a nucleophilic substitution (S_N2). **c)** Disulfide exchange with the activated disulfide 2,2'-dipyridyldisulfide. **d)** Chemical structure of 5,5'-dithiobis-(2-nitrobenzoic acid), the Ellman's reagent, another disulfide activating compound.

Thiols are a functional group frequently targeted for modifications.^[2,7] They are good nucleophiles in a broader pH range and are less abundant in biomolecules than amines. Thiols often arise from cysteine residues and they are prone to forming disulfide bridges with other thiol groups. For this reason, reducing agents are added prior to modification to generate the reactive thiol group *in situ*. The scarcity of thiols, along with their reactivity, enable high selectivity for bioconjugation reactions.^[1]

A common modification of thiols is achieved by Michael addition with maleimides (**Figure 4a**). Although amines are able to perform an aza-Michael addition with maleimides (pH > 8.5), thiols preferentially react with them due to their higher nucleophilicity at lower pH values (pH 6.5-7.5). Maleimides are slowly hydrolyzed in an aqueous environment even after thioether formation,^[8] however the generated succinimic acid derivative keeps a stable connection between the linked moieties.^[1]

The good nucleophilic properties of the thiol group facilitate modifications by substitution. Alkyl halides and haloacetamides readily undergo substitution reactions with thiols (**Figure 4b**). Iodo compounds (and especially iodoacetamides) are preferred due to their higher reaction rates compared to other halides. The drawback of these iodocompounds is their sensitivity to light, which generates iodine (I₂) that interacts with biomolecules.^[1]

Another approach for thiol modifications is the formation of activated disulfides (**Figure 4c**). This procedure is selective for thiols and does not show hydrolytic difficulties. For the activation of a thiol group, pyridine disulfides or the Ellman's reagent (**Figure 4d**) are commonly employed. The mixed, activated disulfide undergoes a second reaction with a sulfur nucleophile to form the desired disulfide connection. Pyridine-2-thione and 2-nitro-5-thiobenzoic acid are formed as

byproducts, which are both detectable in the UV, allowing measurement and quantification of the modification progress. However, this conjugation technique is sensitive to reduction, which renders the connection impractical (or in some applications desirable) for *in vivo* experiments due to high intracellular reducing agent concentrations (e.g. glutathione).^[1]

Bioconjugate Reactions with Hydroxyls

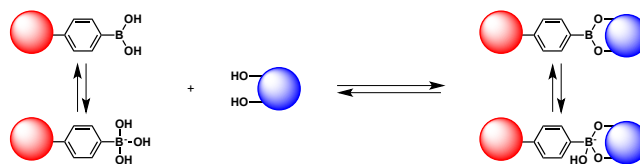


Figure 5. Bioconjugation of vicinal diols with boronic acids. The boronic acids are in equilibrium in aqueous medium with the corresponding boronates.

Aromatic and aliphatic hydroxy groups are much less reactive in aqueous medium than amine or thiol moieties. In some of the presented reactions (e. g. ester formation with succinimidyl active esters), hydroxy groups may generate side products, although, only in minor quantities. Selective hydroxy modifications are more difficult compared to amines and thiols. Boronic acids provide a possibility to reversibly modify diols (**Figure 5**). This technique has been employed in carbohydrate chemistry for modification of aliphatic diols^[9–11] or sensing catechols such as dopamine.^[12,13] Phenylboronic acid derivatives may be incorporated into proteins, other biomolecules, gels or a solid support to enable a connection with carbohydrates or other diol-containing compounds.^[1]

Bioconjugate Reactions with Click Chemistry

In 2001, Barry Sharpless and co-workers defined the concept of click chemistry for certain reaction types.^[14] They defined the prerequisite of such reactions to be modular, wide in scope, stereospecific, with simple reaction conditions, to give high yields and inoffensive byproducts. Furthermore, the starting materials and reagents must be readily available, the solvent must be benign (such as water) and the products must be easily isolated by simple, non-chromatographic methods. These characteristics are typically only achieved if the reactions display a high thermodynamic driving force, usually greater than 20 kcal mol⁻¹. The reaction classes that they defined to fulfill these conditions include cycloadditions (1,3-dipolar cycloadditions and Diels-Alder reactions), nucleophilic substitution chemistry (ring-opening of strained heterocycles such as epoxides, aziridines and episulfonium ions), non-aldol type carbonyl chemistry (urea formation, oxime ethers, hydrazones and amides) and additions to carbon-carbon multiple bonds (epoxidation, dihydroxylation, aziridination, sulfonyl halide addition and Michael additions of nucleophiles). From this definition, several of the previously discussed reactions (oxime condensation, hydrazone formation, (thio)urea formation with iso(thio)cyanates, Michael addition of sulfur nucleophiles to maleimides) fulfill the criteria to be "click" reactions.

Due to its widespread use, the copper(I)-catalyzed azide-alkyne cycloaddition reaction (CuAAC) has become known simply as "the click reaction" over the last couple of years.^[15] For this reason, other reactions that fulfill the criteria put forth by Sharpless are depicted as click-type reactions and the term click reaction is now reserved for the CuAAC. In the following sections, we discuss three frequently used click-type reactions, the CuAAC, the strain-promoted [3+2] azide-alkyne cycloaddition (SPAAC) and the inverse electron-demand Diels-Alder reaction (IEDDA).

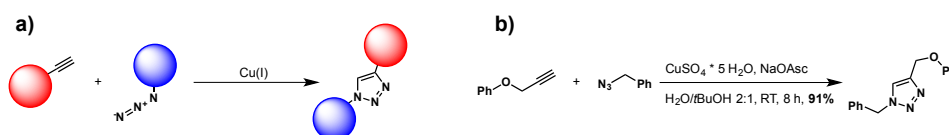


Figure 6. a) General scheme of the copper-catalyzed azide-alkyne click reaction. b) Representative example of a click reaction with *in-situ* generated copper(I) catalyst.^[16]

1,3-Dipolar cycloaddition reactions, which describe the formation of cyclic products from a 1,3-dipole and a multiple bond system, the dipolarophile, have been known for decades and have seen extensive use in organic synthesis. In 1963, Huisgen published reports about the formation and mechanism of 1,3-dipolar cycloadditions^[17] along with the kinetics of these reactions.^[18] The most frequently used dipolarophiles are alkenes and alkynes, however, carbonyl compounds can also undergo 1,3-dipolar cycloaddition reactions. There are many different dipole molecules such as azomethine ylides, nitrones, diazoalkanes and azides. The most popular 1,3-dipolar cycloaddition is by far the reaction between an azide and an alkyne. This transformation requires high temperatures and yields a mixture of 1,4- and 1,5-disubstituted 1,2,3-triazoles. In 2002, two groups (Sharpless and Meldal) reported the selective formation of 1,4-disubstituted 1,2,3-triazoles at ambient temperature with the use of a copper(I) catalyst (**Figure 6**).^[16,19] The click reaction has since then been extensively applied to the development of new small molecules^[15,20], but also to bioconjugation reactions with proteins^[21], carbohydrates^[22], oligonucleotides^[23] and even living cells.^[24] Azides and alkynes are functional moieties that are easily accessible, versatile and show high stability towards non-reducing environments with very high selectivity. However, the employed copper(I) catalysts are sensitive to oxidation and they can also generate reactive oxygen species from the copper-catalyzed reduction of oxygen (O₂) by ascorbate.^[24,25] These reactive oxygen species can potentially damage biomolecules, leading to loss of biological activity. Furthermore, the oxidized ascorbic acid (dehydroascorbate) potentially induces damage by electrophilic interactions with lysine, arginine and cysteine sidechains.^[25] This renders the copper catalysts toxic to most living cells. Special copper ligands were developed to increase the reaction rates of the click chemistry and to minimize the oxidative damage of the biomolecules.^[24,26,27]

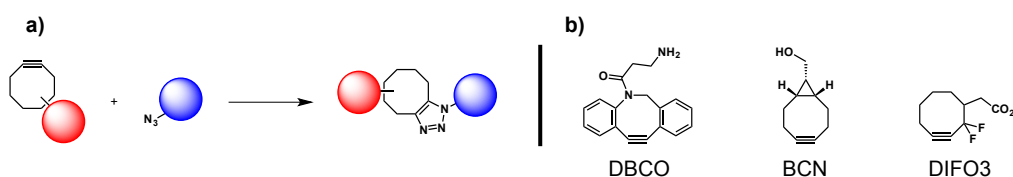


Figure 7. Strain-promoted azide-alkyne cycloaddition reaction. a) General reaction scheme of a strained cyclooctyne and an azide. b) Several examples of strained cyclooctynes. DBCO = Dibenzocyclooctyne-amine, BCN = Bicyclononyne, DIFO3 = Difluorocyclooctyne-3.

An alternative click-type reaction was rediscovered in 2004 by the Bertozzi group.^[28] The reaction between strained alkynes (usually cyclooctynes or larger rings) and azides had been reported in the 1960's, but were forgotten for a long time (**Figure 7a**). Bertozzi *et al.* reported the use of modified cyclooctynes for strain-promoted [3+2] alkyne-azide cycloadditions (SPAAC) for the selective modification of biomolecules and living cells.^[28] This reaction has become popular due to the lack of potentially harmful catalysts.^[29] Series of novel cyclooctynes were synthesized (**Figure 7b** shows a few examples of commonly employed cyclooctynes) to increase the reaction rates (up to 0.9 M⁻¹s⁻¹) so that the process became useful for low concentration modifications of biomolecules under physiological conditions.^[29] The driving force of this transformation was found to be the highly favorable enthalpic release of ring-strain (nearly 18 kcal mol⁻¹).^[28] This technique has been applied to many different applications, such as tumor cell labeling^[30] and nucleotide

modifications within single cells.^[31] Nevertheless, the SPAAC also possesses several disadvantages, such as lack of regioselectivity for the reaction product, which yields triazole mixtures. Aqueous solubility of the strained alkynes may also be a concern, which may demand the use of organic solvents to ensure the conjugation reaction can proceed. The reduced stability of certain alkyne reactants may cause problems, however, several stable cyclooctynes have been developed. The size of the employed strained alkynes for the bioorthogonal connections should also be considered in respect to the physicochemical properties of the alkyne reactants.^[32]

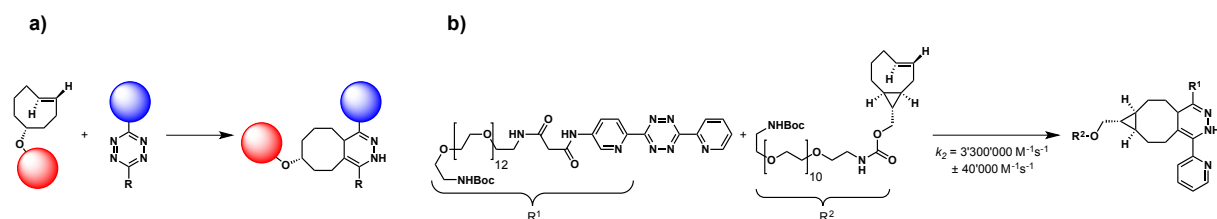


Figure 8. a) General scheme for the inverse electron-demand Diels-Alder reaction. b) The fastest inverse electron-demand Diels-Alder reaction known for bioconjugation applications.^[33]

Another type of cycloaddition is the Diels-Alder reaction of tetrazines with different dienophiles in an inverse electron-demand assembly. This means that, in contrast to classic Diels-Alder reactions, the diene (tetrazine) is electron-poor and the dienophile is the electron-rich species.^[34] The reaction of tetrazines with different dienophiles, such as cyclopropenes, cyclooctynes and terminal alkenes has been described, however, the reaction with *trans*-cyclooctene dienophiles attracts the most interest for bioconjugation reactions.^[33] The mechanism for this reaction proceeds via a [4+2] cycloaddition step to form a strained, bicyclic intermediate, which undergoes a Retro-Diels-Alder reaction to generate one molecule of nitrogen (N₂) and the 4,5-dihydropyridazine intermediate, which usually isomerises to the 1,4-dihydropyridazine product (**Figure 8a**).^[35] In certain cases, the 1,4-dihydropyridazine is further oxidized to the corresponding pyridazine.^[34] The inverse electron-demand Diels-Alder reaction (IEDDA) is the fastest bioconjugation method to date. Optimization of the tetrazine and *trans*-cyclooctene structures led to the development of a conjugation reaction with a second order rate constant of 3'300'000 M⁻¹s⁻¹ (**Figure 8b**).^[33] This reaction is more than 10'000 times faster than the best CuAAC reactions.^[34] Along with the high selectivity and the mild conditions of the IEDDA, it represents the almost perfect reaction type for demanding biological applications. However, the stability of the *trans*-cyclooctene^[36] and the tetrazine^[37] under certain conditions is a limitation that needs to be addressed. Furthermore, the complicated synthesis of the two coupling partners is a major limitation of this procedure.^[34] The IEDDA has proven valuable for many applications in biomedical research including cell labeling, *in vivo* imaging or antibody modifications.^[38]









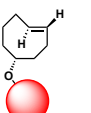
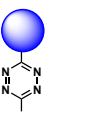


Next to the presented bioconjugation methods, there are many other chemical processes that are potentially useful for bioconjugation applications.^[7,39] Suzuki-Miyaura cross coupling was used for the modification of proteins and oligonucleotides,^[40–42] selective peptide modifications by amide couplings with potassium acyltrifluoroborates (KAT) were performed^[43] and even multicomponent reactions, such as the Passerini or Ugi reactions, were employed.^[44]

Reaction Rates

There are several criteria that must be fulfilled for a chemical reaction to be useful in bioconjugation. First, the chemoselectivity of the reaction is important to avoid undesired side reactions. The stability and availability of the reagents must be considered, as well as the size and solubility of the newly introduced connection. The reaction conditions must also be compatible with the biomolecules. Toxic reactants or catalysts may be used for *in vitro* coupling with certain (synthetic) biomolecules, but modification techniques with living cells and organisms

are not possible with toxic compounds. A 1:1 stoichiometry is often required for the labeling of large biomolecules at high dilution to achieve the required chemoselectivity and to avoid undesired side reactions. For that purpose, the employed reaction needs to proceed efficiently at very low concentrations, which requires a high reaction rate constant.^[45]

Table 1. Reaction rates of different bioconjugation methods. The values were adapted from literature reports.^[45,46]

Reaction	Coupling Partner 1	Coupling Partner 2	Reaction Rate [M ⁻¹ s ⁻¹]	Comments
Oxime Condensation ^[a]			1.3 x 10 ⁻³	- aniline catalysis - neighbouring group effects - acid catalysis
Maleimide-Thiol addition			734	- poor stability of <i>N</i> -aryl maleimides and the ligation products.
CuAAC			10-100	- toxic copper catalyst - reactive oxygen species may be generated
SPAAC			0.96	- no catalyst necessary - potential thiol attack
IEDDA			3'300'000	- stability of reagents - synthetically complex building blocks
KAT			22	- acidic additive necessary for faster reaction rates

[a] uncatalyzed reaction.

A list of different bioconjugation reactions with the highest rate constants known to date and some remarks concerning the reagents and reaction conditions are depicted in **Table 1**. The majority of the presented reactions show second order rate constants between 1 and 1000 M⁻¹s⁻¹. Exceptions are the uncatalyzed oxime condensation that is three magnitudes slower and the IEDDA, which is up to three magnitudes faster than the other listed transformations. Depending on the nature of the bioconjugation reaction and the coupling partners, one should choose the best fitting method in terms of reagents, toxicity, selectivity and reaction speed. For example, live imaging of cell processes requires extremely high reaction rate constants that are only provided by the IEDDA, whereas the unspecific *in vitro* labeling of a simple protein with a fluorophore is perfectly achieved by thiol-maleimide Michael addition or an isothiocyanate modification of the lysine sidechains.

Cross-Linking Strategies

As we discussed in the previous sections, there are many different functional groups that have been employed for bioconjugation assays. However, biomolecules usually display just a few reactive moieties such as amines, thiols, carboxylic acids and alcohol groups. For example, cycloadditions require special functional groups such as azides, alkynes, dienes or tetrazines. Such moieties need to be chemically introduced into the biomolecule. In the simplest implementation, a biomolecule label (e.g. a fluorophore) is directly introduced by the reaction with naturally occurring functional groups. However, it is often required or beneficial to introduce

artificial functions, change certain groups to another functionality or install a chemical spacer. Chemical functional linkers are very powerful tools to achieve this task. The cross-linking of two biomolecules is often just possible with the incorporation of a functionalized spacing unit. Synthetic spacers may also be required to maintain enzyme activity after modification with certain functionalities or for biomolecule immobilization techniques.

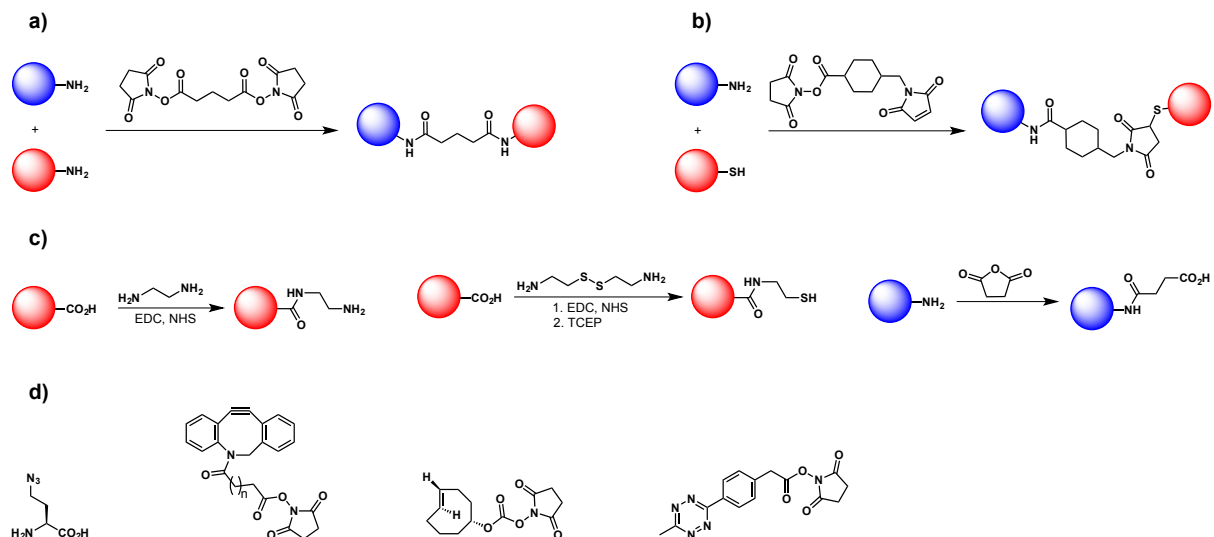


Figure 9. **a)** Covalent linkage of two amines with disuccinimidyl glutarate. **b)** Covalent linkage of an amine with a thiol compound using an *N*-hydroxysuccinimide/maleimide linker. **c)** Functional groups interconversion examples. From left to right: carboxylic acid to amine, carboxylic acid to thiol, amine to carboxylic acid. Further functional group interconversions are possible with other reagents. **d)** A few reagent examples of the introduction of cycloaddition functional groups into biomolecules.

Figure 9a and **Figure 9b** show examples of the crosslinking of molecules with either the same or different functional groups. Many different linkers have been used for the construction of such crosslinks with the naturally occurring reactive groups in biomolecules.^[1] **Figure 9c** shows a few examples of the interconversion of functional groups with the shortest possible handle. Interconversions of carboxylic acids to amines or thiols as well as amines to carboxylic acids or thiols are achieved within a few simple reaction steps.^[1] For the more elaborate cycloaddition reactions, the functional groups always need to be incorporated into the biomolecule either by functional group modifications (usually amines and thiols) or during the synthetic/enzymatic production of the biomolecule (**Figure 9d**).^[3,32,34]

Another frequently used linking strategy is to take advantage of the biotin and avidin (or streptavidin) interaction. Biotin and streptavidin undergo the strongest non-covalent biological connection known ($K_D = 4 \times 10^{-14}$ M).^[47] Immobilization of streptavidin on a solid support is commonly employed to immobilize other biomolecules on a solid support or to selectively "fish out" biotin-labeled compounds from mixtures (pull-down assay).^[1]

1.2 α -Effect Nucleophile Condensations

The reactions of carbonyl groups with different nucleophiles have been widely used in organic chemistry. In 1964, Jencks reported on this class of transformations and showed that the reaction with amine nucleophiles proceeds over a tetrahedral hemiaminal intermediate, which further undergoes a dehydration step to yield the final condensed product (Schiff bases).^[48] The dehydration step was found to be rate-limiting and the whole process is reversible. Amine nucleophile condensations are generally slow under neutral conditions and require acid catalysis.

Nucleophiles with an adjacent heteroatom with unshared electron pairs are called α -nucleophiles. Hydroxylamines, hydrazines, the hypochlorite ion and peroxy anions are examples of this type of influenced nucleophiles. Certain members of these nucleophiles show higher reactivity compared to their non- α -modified versions (e.g. hydroxylamines are more reactive than amines) and are therefore termed α -effect nucleophiles.^[49] The special properties of these nucleophiles have been further investigated in earlier reports.^[49,50] We limit the following discussion about α -effect nucleophile condensations to the reactions of hydroxylamines, hydrazines and hydrazides with carbonyl compounds (aldehydes and ketones).

Despite the higher nucleophilicity of α -effect nucleophiles, condensation reactions with carbonyl compounds are still very slow at neutral conditions. Most attempts on increasing the reaction rate have focused on lowering the barrier of the dehydration step. The use of low pH values^[51] and the addition of aniline derived catalysts to speed up the reactions have been reported.^[52-54] Recent reports showed that aniline derivatives with *ortho* proton donors even further increased the reaction rates of oxime/hydrazone condensations.^[55-57] Interestingly, the attachment of a phosphate group in *ortho* position to an aldehyde also increased the reaction rate with hydrazines/hydrazides by an order of magnitude.^[58] In a very recent assay, it was also shown that the oxime condensation might be accelerated by the increase of the salt (NaCl) concentration in the reaction mixture. This is an important finding for modifications of biological systems, since increased salt concentrations are easily applicable, whereas complicated catalytic procedures or laboriously introduced accelerating groups can show incompatibilities with the biological system or even toxicity to the biomolecules/cells.^[59]

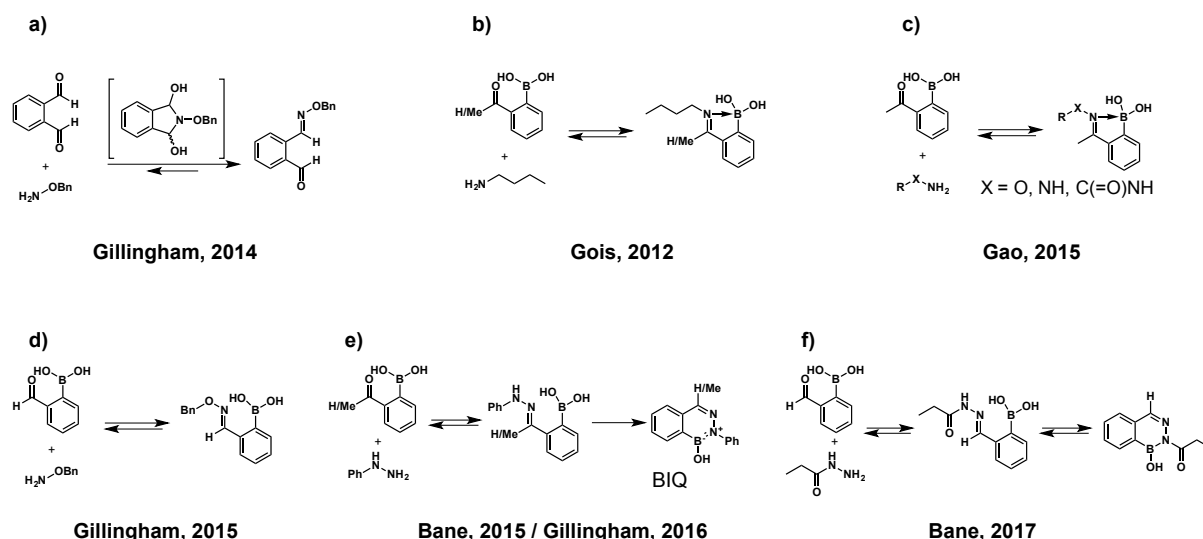


Figure 10. Overview of the most recent reports about (α -effect) amine nucleophile condensations under neutral conditions that do not make use of catalysts or *ortho* proton-donating groups. **a)** *Ortho*-dialdehydes showed to increase oxime formation rates due to the formation of a *bis*-hemiaminal intermediate.^[60] **b)** Gois *et al.* reported an increased stability of imines with *ortho*-boronate carbonyl compounds.^[61] **c)** Ketone boronic acids displayed highly increased reaction rates for oxime and hydrazone formations in a reversible manner.^[62] **d)** Boronic acid aldehydes were found to facilitate oxime condensations under neutral conditions with high rate constants.^[63] **e)** Boronic acid carbonyl compounds formed a stable, irreversible, aromatic product with hydrazines in a two steps process.^[64,65] **f)** Hydrazides underwent condensation with aldehydes to rapidly form a reversible, heterocyclic boron product.^[66] BIQ = 4,3-borazaroisquinoline.

Several methodologies to improve the formation of imine-type condensations without the need for catalysts or proton donating groups have been shown. Our group reported the formation of stable oximes with the use of *ortho*-dialdehydes under neutral conditions (**Figure 10a**). The intermediate *bis*-hemiaminal structure formed very rapidly with rate constants of approximately $500 \text{ M}^{-1}\text{s}^{-1}$. While the addition step was very quick, the subsequent dehydration step to form the actual oxime was found to be slow ($k_2 = 1.2 \cdot 10^{-5} \text{ s}^{-1}$), about two to three orders of magnitude slower than typical oxime condensations under equal conditions. Along with the very large overall equilibrium constant ($>10^7 \text{ M}^{-1}$), this implied an energetically very stable *bis*-hemiaminal intermediate. For that reason, the nucleophile addition step was found to be practically irreversible, which also illustrates a successful bioconjugation. DNA labeling experiments and peptide hydroxylamine test reactions with common biological additives rendered this reaction suitable for more advanced bioorthogonal transformations.^[60]

In 2012, the Gois group published a procedure for the formation of hydrolytically stable imine formations with the potential for selective reversibility. They discovered that 2-carboxylphenylboronic acids underwent rapid imine formation with primary amines under neutral conditions, while the generated iminoboronates showed exceptionally high stability towards hydrolysis (**Figure 10b**). They concluded that an interaction between the imine nitrogen and the boron atom is responsible for this high stability. Upon the addition of dopamine, fructose or glutathione, compounds that are known to interfere with boronic acids, the hydrolysis rate of the iminoboronates highly increased,^[61] which supported this assumption.^[61]

Based on these results, our group published almost simultaneously with the Gao group reports about oxime and hydrazone formations with 2-carboxylphenylboronic acids. The Gao group investigated oxime and hydrazone formations with ketone boronic acids, whereas our group focused on the formation of oximes with boronic acid aldehydes (**Figure 10c and d**).^[62,63] While both carbonyl functional groups underwent rapid imine-type conjugations with rate constants $>10^3 \text{ M}^{-1}\text{s}^{-1}$, the ketone Schiff base compounds were highly susceptible to hydrolysis. Aldehyde Schiff bases, however, showed a much better hydrolytic stability and were only reversible upon

the addition of a different hydroxylamine or boronic acid aldehyde in excess. Our group also successfully showed that common biological additives like glutathione, sucrose, lysozyme and human serum did not interfere with Schiff base formation. Only the oxidation of the phenylboronic acid to a phenol in human serum was observed as a side reaction, however, the oxime formation was determined to be faster than the oxidation process, which ensured a successful conjugation. Lysozyme, which was modified with boronic acid aldehydes in previous reports,^[61] remained unaffected in this assay, showing that hydroxylamines outcompete any nucleophilic amino acid moieties.^[63]

In a further report, we showed that the mechanism for boron-assisted oxime formation most likely proceeds over the generation of a five-membered hemiaminal boronate. This intermediate strongly facilitates the rate limiting dehydration step, due to the elimination of a boronate instead of a water molecule. We supported this mechanistic assumption by substrate variation; changing the position of the boronic acid from the aldehyde moiety to the hydroxylamine coupling partner did not show any acceleration of the condensation compared to uncatalyzed oxime condensations under neutral conditions. Additionally, the use of coordinatively saturated boronates (trifluoroboronates and MIDA esters) showed a diminished reaction rate enhancement in comparison to free boronic acids. We also showed that ketoximes were more susceptible to exchange reactions than aldoximes. An X-ray structure of an aldoxime showed no iminoboronate formation, which explained the higher hydrolytic stability compared to ketoximes. The latter are known to form iminoboronate structures, which act as an activation for hydrolysis.^[64]

The Bane lab and our group published reports about boron-assisted hydrazone formation that leads to the generation of an irreversible 4,3-borazaroisoquinoline (BIQ) product (**Figure 10e**). The initial hydrazone formation between hydrazines and *ortho*-boronate carbonyl compounds was found to be very fast ($>10^3 \text{ M}^{-1}\text{s}^{-1}$). However, the hydrazone underwent a secondary reaction, forming the aromatic, heterocyclic BIQ. This second process proceeded more slowly ($10^{-3} - 10^{-2} \text{ s}^{-1}$), but formed a stable product. We showed that the fused aromatic ring system could be used to generate fluorescent molecules by the introduction of electron-donating substituents.^[64] The Bane lab showed the usefulness of this methodology in orthogonal protein labeling.^[65] BIQs and their preparation are known since the 1960s,^[67,68] but were forgotten until the 1990's, when Groziak *et al.* confirmed the structure of the BIQs and extended the scope of this compound class.^[69,70] BIQs have been considered as unique pharmacophores, and indeed were shown to have growth inhibitory effects against different types of bacteria.^[71-74]

In 2017, the Bane and Gao groups reported the formation of heterocyclic structures upon the synthesis of hydrazones with acyl hydrazides (**Figure 10f**). Unlike hydrazines, these acyl hydrazides did not form a stable BIQ, because the B-N bond formation in the heterocyclic structure was reversible. Semicarbazides, on the contrary, formed stable compounds and upon the introduction of an amine substituent in α -position to a hydrazide, a tricyclic structure was generated that was also stable enough for protein modifications.^[66,75]

In summary, iminoboronate, acyl hydrazone and ketoxime formations are rapidly reversible, which enables dynamic, covalent modification applications. Aldoxime formation is slowly reversible and therefore perfectly suited for targeted slow release and exchange assays. A stable, irreversible linkage is obtained upon the treatment of carbonyl boronic acids with hydrazines that form stable, heteroaromatic BIQs in a two steps process.^[57,64,76]

1.3 Fluorescent Molecules in Bioconjugation

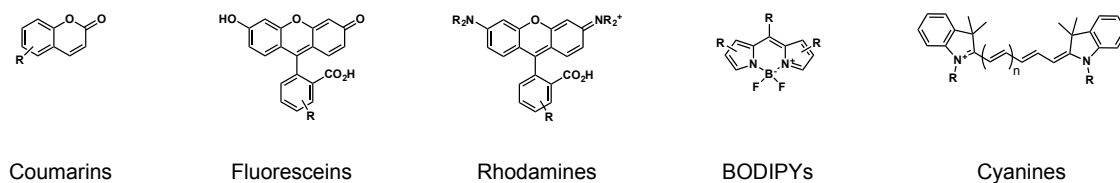


Figure 11. Overview of the most important fluorophore classes that are regularly used in chemical labeling of biomolecules. BODIPY = Boron-dipyrromethene.

Fluorophores are one of the most important classes of small molecules for bio-imaging techniques. Due to the high sensitivity and selectivity, fast and simple data acquisition, multiparametric nature and the spatial and temporal resolution, fluorescent tagging is the method of choice for biomolecule labeling and observation of dynamic cellular processes. In contrast with radioactivity-based techniques, fluorescence is non-invasive, which makes it safe to use with living cells. In principle, fluorescence detection can be achieved down to one single fluorescent molecule, owing to its high sensitivity.^[77] The biggest problem in biological systems is the intrinsic fluorescence of naturally occurring molecules, such as tryptophan, NADH or flavins. For that purpose, many different fluorophores have been developed that display different absorption and emission spectra. **Figure 11** shows the structures of the most important fluorophore classes that have been employed for bioorthogonal fluorescence labeling.^[78]

Another advantage of fluorescence imaging over other labeling methods is the potential modulation of the signal. A non-fluorescent or only weakly fluorescent molecule can be designed in order to increase its fluorescence upon structural alterations or by environmental changes (e.g. metal complexation or pH changes).^[79] Such structures are called turn-on fluorophores.^[80] In an ideal case, the turn-on fluorophore displays no fluorescence until a specific modification occurs. However, a certain amount of fluorescence is observed with the parent molecule. Nevertheless, if the reaction gives a significant signal increase and a shift in emission wavelength (for biological applications a red shift is desired due to the biomolecules' intrinsic fluorescence in the blue-green), the background signal is not problematic. Boron-containing fluorophores are very interesting for the development of such turn-on fluorophores. BODIPY is by far the most widely applied boron-based fluorophore and it has been shown that turn-on fluorophores may be generated using the BODIPY core structure.^[81,82] Other boron-containing heterocycles were found to act as fluorophores as well.^[83] We and the Gao group published reports about the generation of turn-on fluorophores upon BIQ formations with hydrazines^[64] and semicarbazides.^[84]

1.4 Project Description

After reporting the fast formation of oximes with adjacent boronic acid substituents, we emphasized the evaluation of the hydrolytic stability of aldoximes, since ketoximes showed a rather low resistance towards hydrolysis.^[62] Furthermore, several reversibility assays gave an idea about the kinetics of the back reaction and a direct comparison about the preference between aldoximes and ketoximes. We also investigated the mechanism of the boron-assisted oxime formations by repositioning of the boronic acid in the oxime condensation system and masking of the boronic acid by different coordinative saturation methods. This methodology was applied to the fluorescent labeling of proteins.

Another focus was the evaluation of boron-assisted condensation reactions between hydrazines and carbonyl boronic acids. The formation of the hydrazone and the subsequent intramolecular cyclization reaction were of particular interest along with potential influences on the reaction. For the first time, we showed the time-resolved observation of the BIQ formation with NMR spectroscopy. We also synthesized a series of BIQs to develop a novel turn-on fluorophore that might be used for bioorthogonal fluorescent labeling.

2. Boron-Assisted Oxime Condensations

2.1 Characteristics of Boron-Assisted Oxime Condensations

Stability and Reversibility

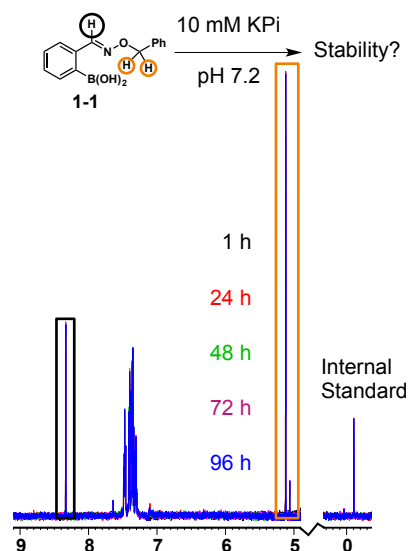


Figure 12. NMR stability measurements of **1-1** at 100 μM concentration over four days. The oxime (black) and benzylic (orange) protons are highlighted. The used internal standard was 3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid (TMSP- d_4).

Measurements of the hydrolytic stability of aldoxime **1-1** at a 100 μM concentration in pH 7.2 phosphate buffer revealed a high resistance against hydrolysis (**Figure 12**). Over the course of three days, we found integration differences of less than 5%, which lies within the margin of the integration accuracy. After four days, we found an integration reduction of 15% compared to the initial measurement, which indicated a slow hydrolysis of the product (see **Figure 23** for detailed NMR spectra with integrals).

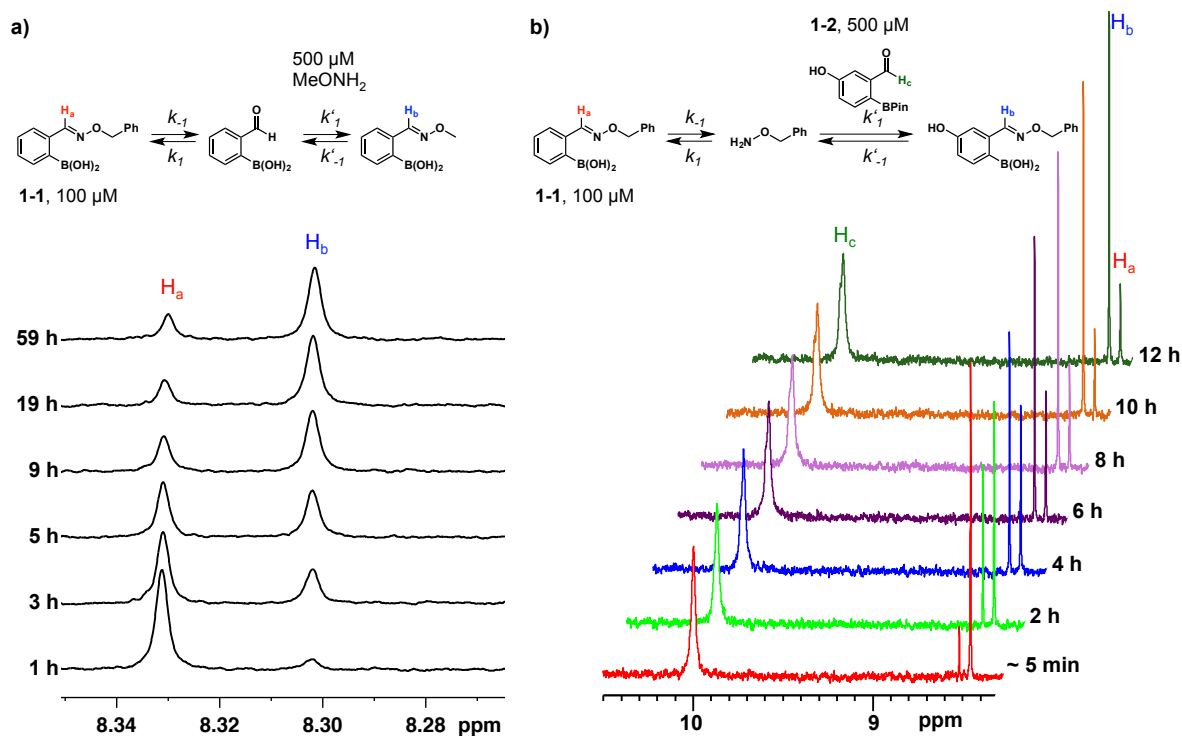


Figure 13. a) NMR reversibility assay between aldoxime **1-1** and a five-fold excess of methylhydroxylamine (MeONH₂). The conversion to the new oxime proton H_b is displayed. b) Time-course NMR assay for the reversibility of aldoxime **1-1** with a five-fold excess of boronic acid aldehyde **1-2**. The oxime and aldehyde proton changes are displayed. The assays were conducted with 10 mM KPi, pH 7.2 in D₂O:CD₃CN 4:1.

To find evidence for the reversibility of oxime formation and to determine the aldoxime hydrolysis rate, we conducted two time-course NMR assays, in which we directly observed the generation of new aldoxime products. In the first assay (**Figure 13a**), we treated **1-1** with a five-fold excess of methylhydroxylamine (MeONH₂). The formation of the new oxime species was clearly observed by NMR and we found that equilibrium was established after 10-19 hours. For a final proof of the oxime reversibility, we replaced methylhydroxylamine by boronic acid aldehyde **1-2** and again observed the formation of a new oxime species with equilibrium established after 10-12 hours. This second experiment was necessary to be able to distinguish between hydrolysis and direct nucleophilic attack of methylhydroxylamine at the aldoxime **1-1**. From these results, we found that the oxime formation rate constant must be much larger than the rate constant for aldoxime hydrolysis ($k_1 \gg k_{-1}$). For that reason, we could take the formation of the new product as a direct measure of the aldoxime hydrolysis rate. Intermediate hydrolysis products could not be observed in any of these assays. The rate constant for this particular aldoxime hydrolysis was calculated to be $4.2 \pm 0.4 \times 10^{-5} \text{ s}^{-1}$. From the rate constant of oxime formation ($11.1 \pm 0.3 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$), we could calculate an equilibrium constant (k_1/k_{-1}) of $2.6 \pm 0.3 \times 10^8 \text{ M}^{-1}$.^[63]

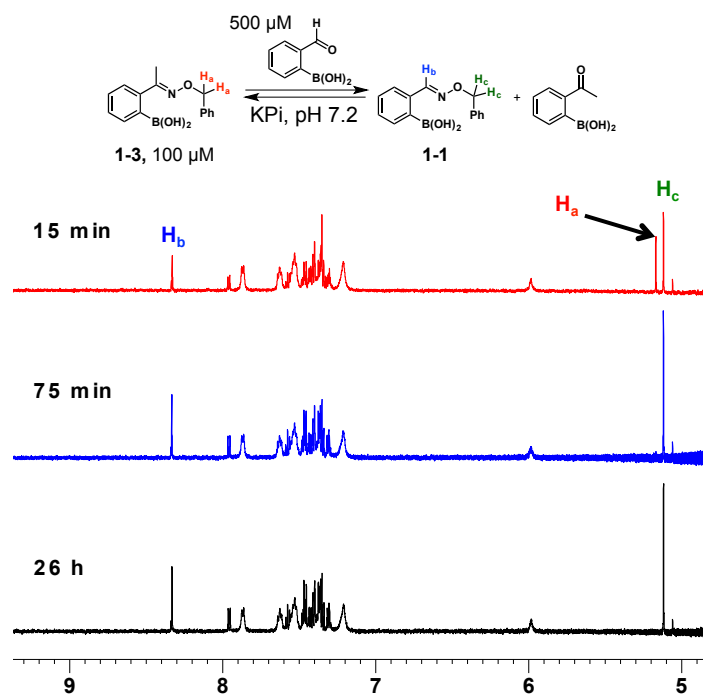


Figure 14. Time-course NMR assay for the determination of the ketoxime reversibility in the presence of a boronic acid aldehyde (2-formylphenylboronic acid, 2-FPBA in this case).

In earlier reports, it was shown that ketone hydrazones showed fast reversibility in aqueous medium.^[62] To explore the reversibility of ketoximes in comparison to aldoximes, we prepared ketoxime **1-3** and treated it with a five-fold excess of 2-formylphenylboronic acid (2-FPBA). Time-course NMR analysis revealed that ketoxime hydrolysis and subsequent aldoxime formation rapidly occurred (**Figure 14**). After 15 minutes, about 60% of the ketoxime had been transformed to aldoxime **1-1** and after 75 minutes, compound **1-1** was the only detectable oxime species. We concluded that the ketoxime nitrogen is more prone to binding to the boron and formation of an iminoboronate species. Unlike non-boron containing imines, iminoboronates were reported to be activated for hydrolysis.^[61] An X-ray structure of our model compound **1-1** revealed no iminoboronate interaction in the ground state.^[64] We supposed that the energetic reason for this difference is related to the hypothesis that the nitrogen basicity of the Schiff base determines the stability of oximes and hydrazones.^[85] The additional σ -donation from the ketoxime methyl substituent should render its nitrogen more basic and therefore more susceptible to protonation or interaction with a Lewis acid.

Mechanistic Insights

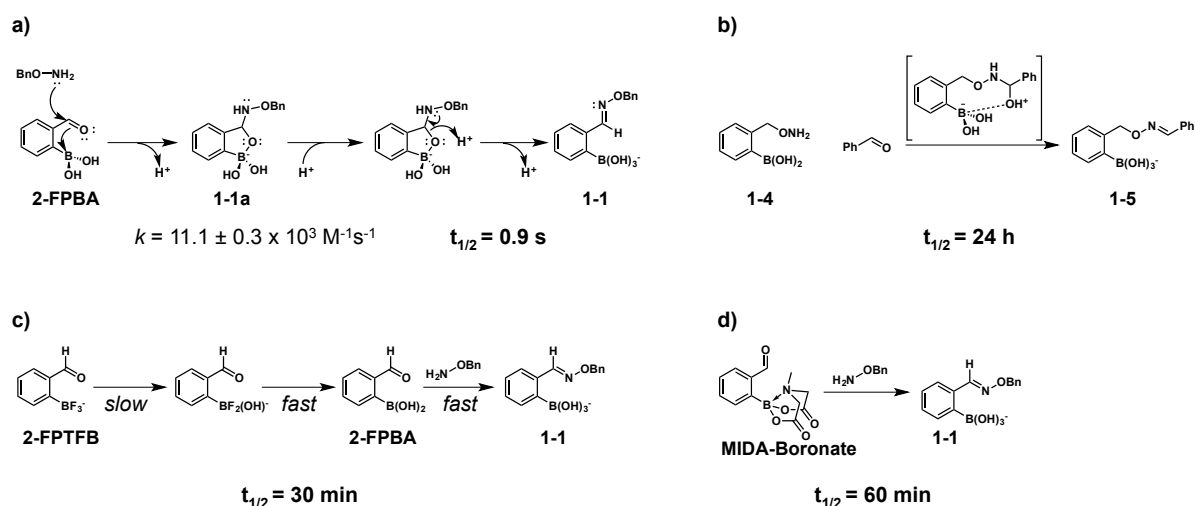


Figure 15. Experiments to give insight into the mechanism of oxime formation. All reactions were conducted in pH 7.2 phosphate buffer at 100 μM (a, c and d) and 1 mM (b) concentrations. **a)** Putative mechanism of the boron-assisted oxime formation. Reaction rates were adapted from our first publication of that topic.^[63] **b)** Prove of the importance of the five-membered ring intermediate. An 8-membered intermediate did not promote an efficient coupling. **c)** Prove of the importance for a free boron coordination site. Coordination saturation with trifluoroborates reduces the oxime condensation rate. **d)** MIDA-boronates also slow down the oxime condensation. The reaction supposedly proceeds over the same mechanism as for trifluoroborates. 2-FPBA = 2-formylphenylboronic acid, 2-FPTFB = 2-formylphenyltrifluoroborate.

From the initial reports of our group about boron-accelerated oxime condensations, we know that the positioning of the boronic acid plays a very important role, thus only *ortho*-boronic acid aldehydes/ketones show increased reaction rates compared to uncatalyzed oxime formations under neutral conditions.^[63] We hypothesized that the transformation of the aldehyde to the oxime proceeds over a five-membered, cyclic boron-hemiaminal intermediate (**1-1a** in **Figure 15a**). The rate limiting dehydration step further occurs over the elimination of a boronate rather than a water molecule. This process should be far more facile since the protonation of the cyclic boron-hemiaminal intermediate should take place at around neutral pH.^[86–88] The shown mechanism was also supported by the report about iminoboronate formation of carbonyl boronic acids with primary amine sidechains in proteins.^[61] A potential interaction of the hemiaminal (**1-1a**) nitrogen with the boron atom could be excluded because such an interaction would lead to an activation of the hemiaminal towards hydrolysis (or the backreaction) due to iminoboronate formation.

The importance of the cyclic intermediate was further proven by the reaction of aminoxybenzylboronic acid **1-4** with benzaldehyde (**Figure 15b**). In that case, the formation of the product would proceed over an 8-membered cyclic intermediate. In fact, we found a reaction rate that was comparable to oxime condensations at neutral pH without catalysis or substituent-induced reaction rate enhancement.

The role of the boronic acid was further elucidated by the use of boron moieties without a free coordination site (masking). 2-Formylphenyltrifluoroborate (**2-FPTFB**) and its corresponding MIDA ester (**MIDA-boronate**) were employed for that purpose (**Figure 15c** and **d**). The reaction rates were much faster than conventional oxime condensations, however, boronates with a free coordination site showed a 2000-4000 times higher reaction rate. We hypothesized that with coordinatively saturated boronates the rate-limiting step of the reaction was the hydrolysis of the trifluoroborate (or the MIDA ester) prior to the actual oxime condensation. Two reports showed that the hydrolysis of the first fluoride (or the B-N interaction in the **MIDA-boronate**) was rate-limiting in other reactions with boronic acids, whereas the further hydrolysis proceeded much

faster. The reported reaction rates for these processes fit our observations in the oxime forming reactions and supported our proposed mechanism.^[89,90]

2.2 Bioorthogonal Labeling of Immunoglobulin G

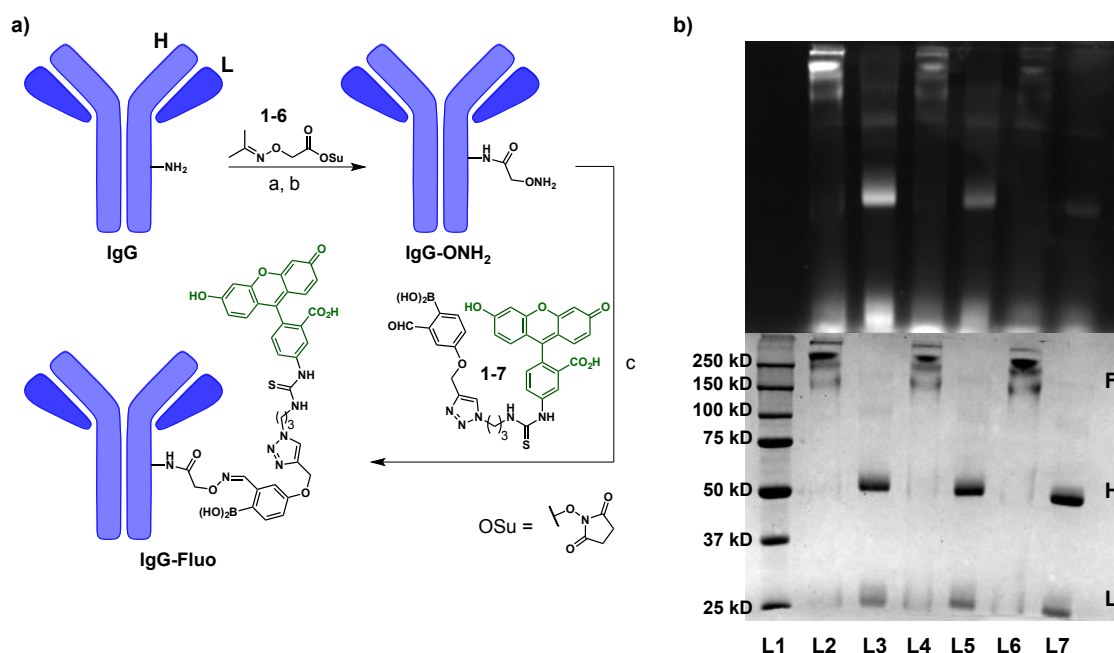


Figure 16. a) Schematic representation of the IgG modification with a hydroxylamine handle and subsequent fluorescein labeling with boronic acid aldehyde **1-7**. IgG is represented in cartoon and only one primary amine group is shown for simplicity. H = Heavy chain; L = Light chain; F = Full protein. Reaction conditions: a) 50 mM phosphate buffer, 1 mM EDTA, pH 7.5, DMSO, 35°C, 30 min. b) NH₂OH hydrochloride, 50 mM phosphate buffer, 25 mM EDTA, pH 7.5, 36°C, 2 h. c) 50 mM phosphate buffer, 1 mM EDTA, pH 7.5, DMSO, 37°C, 80 min. **b)** 10% SDS-PAGE analysis (200 V, 60 min) of the IgG fluorescent labeling. Top: Fluorescein excitation; Bottom: Coomassie staining. **L1:** Protein standards, **L2:** IgG-Fluo, **L3:** IgG-Fluo reduced, **L4:** Control experiment 1, IgG-ONH₂ + 2-FPBA + **1-7**, **L5:** Control experiment 1, IgG-ONH₂ + 2-FPBA + **1-7** reduced, **L6:** Control experiment 2, IgG + **1-7**, **L7:** Control experiment 2, IgG + **1-7** reduced. 2-FPBA = 2-formylphenylboronic acid. Protein reduction was achieved by treatment with 2-mercaptoethanol (BME).

Protein modifications are one of the most important tasks in biochemical research. Especially the fluorescent labeling of proteins opens up the possibility to directly observe them in cell assays or in complex mixtures. The introduction of unique handles creates the desired selectivity.

IgG is the most abundant class of antibodies in human serum. There are four different subclasses of IgG (IgG1-IgG4), of which IgG1 and IgG2 are the most commonly found variants.^[91] A schematic representation of IgG is depicted in **Figure 16a**. It has a molecular weight of 146 kD (except IgG3 with 170 kD) and consists of two heavy chains (H, ~ 55 kD) and two light chains (L, ~25 kD) that are assembled in the typical Y-shape of antibodies.^[92] Electrostatic interactions and disulfide bridges hold the four chains together. IgG interacts with different antigen classes (soluble protein antigens, membrane proteins, bacterial polysaccharides, allergens) and triggers the immune response.^[91]

To demonstrate the value of our boron-assisted oxime formation technique, we labeled immunoglobulin G (IgG) with the fluorescein-boronic acid aldehyde **1-7**. IgG contains about 90 lysine residues, of which up to 30 are accessible for chemical modification.^[93] We used the activated succinimidyl ester **1-6** to randomly generate the corresponding amides with the accessible primary amines (**Figure 16a**). The acetone-oxime protecting group was selectively removed by the addition of a large excess of hydroxylamine hydrochloride. Treatment of the hydroxylamine antibody **IgG-ONH₂** with the fluorescent aldehyde boronic acid **1-7** resulted in the

desired labeling (**Figure 16b**, lanes 2 and 3). To exclude random binding of the dye to the protein, we ran two control experiments. In a first assay, we treated the hydroxylamine antibody (**IgG-ONH₂**) with a large excess of 2-FPBA to selectively block all hydroxylamines (and potential remaining amines). Despite treatment with an excess of the fluorescent dye **1-7** for 1 h, we only found weakly fluorescent protein bands (**Figure 16b**, lanes 4 and 5). The remaining fluorescence most likely arose from the slow reversibility of the oxime linkage (compare **Figure 13b**). The second control experiment used unmodified IgG, which was treated with dye **1-7**. Polyacrylamide gel analysis showed very weakly fluorescent bands (**Figure 16b**, lanes 6 and 7), which most likely arose from iminoboronate formation with the accessible lysine residues. As described in Chapter 1.2, iminoboronates are relatively unstable modifications. Most of the fluorescent label was removed during gel analysis.

In summary, we showed the successful modification of the IgG antibody with a hydroxylamine handle and subsequent fluorescent labeling. The fluorescein compound was selectively attached via oxime formation. Due to the reversible nature of the oxime linkage, dynamic covalent modifications are possible, while undesired background modifications remain low.

3. Boron-Assisted Hydrazone and BIQ Formation

3.1 Characteristics of Boron-Assisted Hydrazone and BIQ Formation

After the successful elucidation of the boron-assisted oxime formation characteristics and the application in antibody labeling, we turned our attention to boron-assisted hydrazone formations. Despite hydrazones show a higher susceptibility to hydrolysis than oximes,^[85] we considered the boron-assisted hydrazone formation a valuable approach for the development of an irreversible bioconjugation. Several reports from the 1960's showed that hydrazones with a boronic acid substituent in *ortho* position undergo an intramolecular cyclization reaction to 4,3-borazaro-isoquinolines (BIQ).^[67,68] The synthesis and the exact BIQ structure had been shown in several publications,^[69,70,73,74] however, applications in bioconjugation had not been reported up to that point. Shortly before our own publication, the Bane group reported their analysis of BIQ formation with reaction kinetics and applicability in protein labeling.^[65] We were pleased to find that our own results on the kinetic parameters fit well with the findings of the Bane lab. We included direct observation of BIQ formation by ¹H NMR analysis, showing that the cyclization to BIQ was irreversible. The influence of pH and additives on the reaction rate was also investigated.

BIQ Formation and Reversibility

The first step in the synthesis of a BIQ is the formation of a hydrazone with an adjacent boronic acid. We think that the mechanism for this condensation proceeds through the same pathway as for boron-mediated oxime formations (**Figure 15a**). Unlike oximes, hydrazones undergo a second, intramolecular cyclization reaction to form the heteroaromatic BIQ. **Figure 17** shows a reaction example between aldehyde **1-2** and phenylhydrazine to form the BIQ product **1-8**. We found that hydrazone formation proceeded rapidly at low micromolar concentrations with subsequent intramolecular cyclization to the BIQ product.

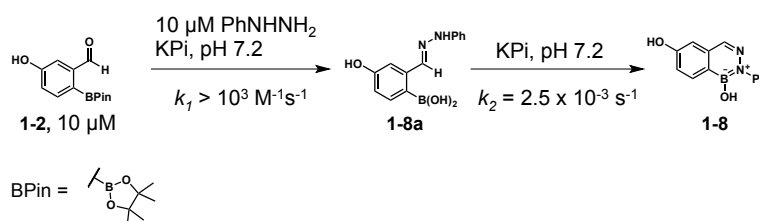


Figure 17. Two steps formation of BIQs. The hydrazone condensation proceeds rapidly, whereas the cyclization to the aromatic BIQ is a slow, intramolecular process. The rate constants were adapted from the kinetic UV measurements (**Table 2**).

Although BIQ formation is observable by UV spectroscopy and LC-MS analysis, we decided to use a time-course NMR assay to show the formation of the BIQ. In **Figure 18a**, the generation of a new proton signal (H_b) right next to the hydrazone proton H_a was observed. Due to the high reaction rate, the formation of the hydrazone intermediate **1-8a** was not traceable by NMR.

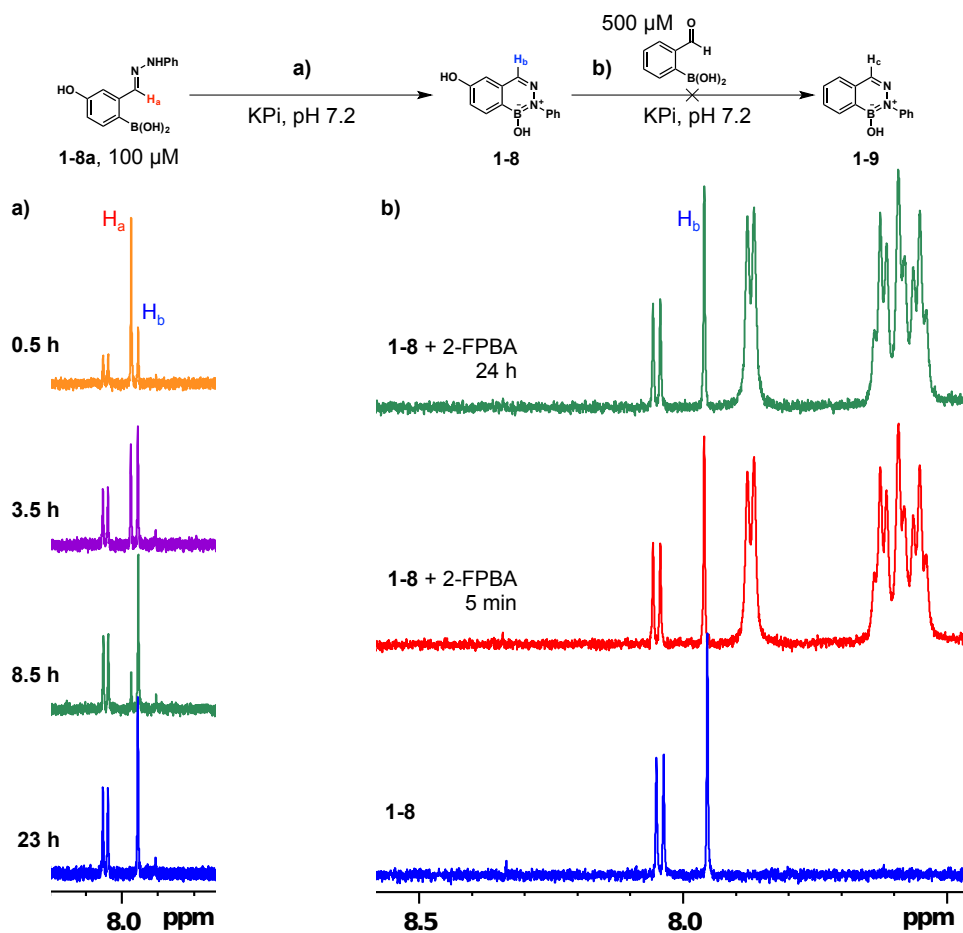


Figure 18. a) BIQ formation observed by time-course NMR analysis. The detection of the aldehyde and the hydrazine starting materials was not possible due to the high reaction rate in hydrazone formation. A complete conversion from the hydrazone proton H_a to the BIQ proton H_b was detected. b) NMR reversibility assay of BIQ **1-8**. A five-fold excess of 2-FPBA was added to BIQ **1-8** to potentially form the new BIQ **1-9**.

The high stability of BIQs towards hydrolysis had been shown in earlier reports,^[65,67,68] however, to the best of our knowledge, a final piece of evidence of the irreversibility of BIQ formation had not been shown. For that purpose, we treated BIQ **1-8** with a five-fold excess of 2-FPBA and recorded several ¹H NMR spectra after various time points (**Figure 18b**). We observed no changes within 24 hours, which led us to the conclusion that the formation of BIQ is irreversible under the applied aqueous conditions. An explanation for the high stability is found in the aromatic structure of the BIQ, which stabilizes the system and renders the former hydrazone carbon atom unsusceptible to the nucleophilic attack of a water molecule.

Reaction Kinetics, Influence of pH and Biological Additives

The formation and the structural confirmation of BIQs was shown in several reports (see previous section), however, the analysis of the reaction rates for BIQ formation were still unknown until the report published by the Bane group.^[65] We were pleased to find that their reaction rate determinations showed similar results to our own. pH and biological additives were found to influence the rate of BIQ formation (**Table 2**). The time resolution of UV analysis for the formation of hydrazone **1-8a** was not high enough to enable an accurate determination of the rate constant. However, a lower bound of $k_1 > 10^3 \text{ M}^{-1}\text{s}^{-1}$ could be established from the first few seconds of the hydrazone UV measurement. The cyclization to the BIQ proceeded much slower and could be measured by UV absorption at 340 nm at 10 μM in both starting materials (**Figure 19**). We chose 340 nm as the observation wavelength due to the high absorption of the hydrazone intermediate

and the low absorption of the BIQ. In fact, we used the decrease in hydrazone absorption as an indirect measurement of the BIQ formation. For accurate rate constant determination, we first measured the extinction coefficient ϵ of the hydrazone intermediate **1-8a**, which is also pH dependent. We ran these measurements with a 50% excess of hydrazine at a known aldehyde concentration. The determined maximum absorption was considered as the point of complete conversion to the hydrazone **1-8a**. Although BIQ formation immediately starts after mixing the two starting materials, the fraction of BIQ at the highest absorption point was negligible due to the much higher hydrazone formation rate.

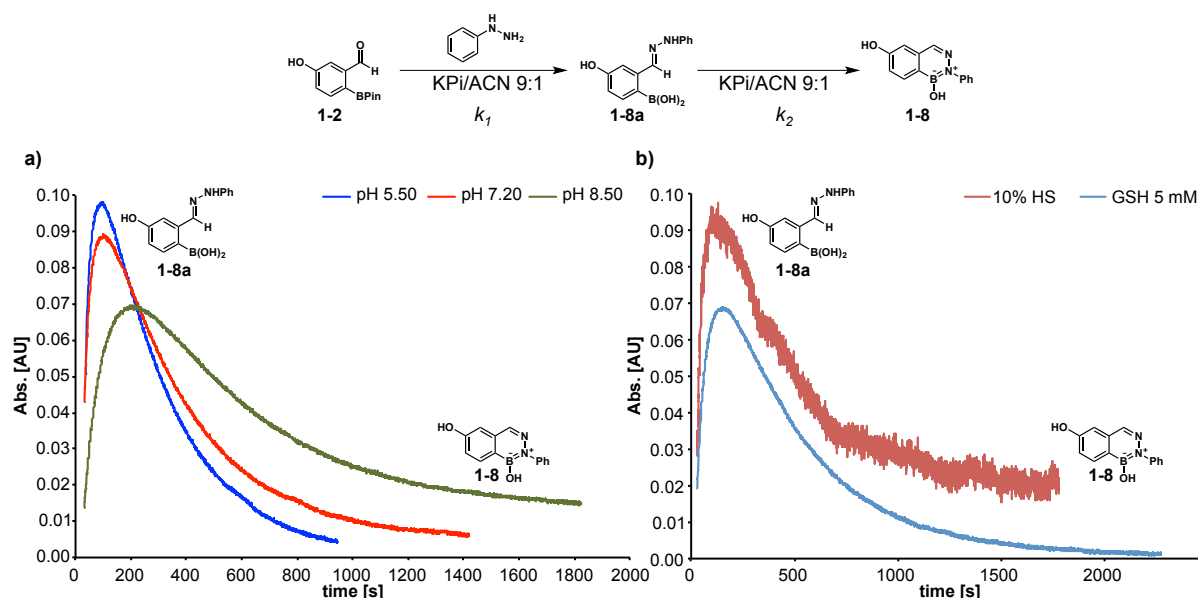


Figure 19. Time-course UV measurements (340 nm) of the BIQ formation. All assays were performed at a 10 μM concentration with 1:1 stoichiometry. **a)** Measurements of the BIQ formation under different pH values. Lowered pH increases the BIQ formation rate, increased pH reduces the BIQ formation rate. **b)** Influence of human serum (HS) and glutathione (GSH) on the BIQ reaction rate under neutral conditions. Both additives decreased the BIQ formation kinetics.

As shown in **Figure 19** and **Table 2**, the pH of the buffer influenced the formation rate of the BIQ. While the rate constant under neutral (pH 7.20) conditions was found to be $2.6 \times 10^{-3} \text{ s}^{-1}$, a decrease to pH 5.50 or an increase to pH 8.50 resulted in a 50% higher or 50% lower reaction rate, respectively. Biological additives, such as human serum (HS) and glutathione (GSH) also influenced the reaction kinetics. Both rate constants were determined to be in between the pH 7.20 and pH 8.50 rates. In the case of human serum, a maximum amount of 10% could be added, because higher HS concentrations deteriorated the signal-to-noise ratio of the UV signal.

Table 2. Summary of the measured BIQ reaction rates (k_2) dependent on pH and additives. The reaction rates of the hydrazone formation (k_1) could not be measured with high accuracy, but a lower bound of $>10^3 \text{ M}^{-1}\text{s}^{-1}$ could be established from the measurement results.

Condition	Extinction coefficient [$\text{M}^{-1}\text{cm}^{-1}$]	Rate constant [s^{-1}]	Standard deviation [s^{-1}]
pH 5.50	$\epsilon = 11390$	$k_2 = 3.95 \times 10^{-3}$	1.32×10^{-4}
pH 7.20	$\epsilon = 10006$	$k_2 = 2.57 \times 10^{-3}$	1.19×10^{-4}
pH 8.50	$\epsilon = 7810$	$k_2 = 1.32 \times 10^{-3}$	1.36×10^{-4}
HS (10%)	$\epsilon = 10006$	$k_2 = 1.68 \times 10^{-3}$	1.04×10^{-4}
GSH (5 mM)	$\epsilon = 10006$	$k_2 = 2.22 \times 10^{-3}$	8.64×10^{-5}

Reversibility of Hydrazones

In analogy to boron-assisted oxime condensations, we suspected that hydrazone formation was reversible as well. Gao *et al.* reported the reversibility of hydrazones from the reaction of 2-acetylphenylboronic acid with acetylhydrazine.^[62] The Bane lab also reported the formation of BIQ-like heterocycles in the reaction of hydrazides and *ortho*-carbonyl phenylboronic acids. These BIQ-type products formed dimers and showed reversibility dependent on the pH of the buffer.^[66] We could confirm the reversibility of hydrazides by our own assays with benzhydrazide and different boronic acid aldehydes. However, the reversibility of hydrazones from the reaction of *ortho*-carbonyl substituted phenylboronic acids with a hydrazine had not been confirmed yet.

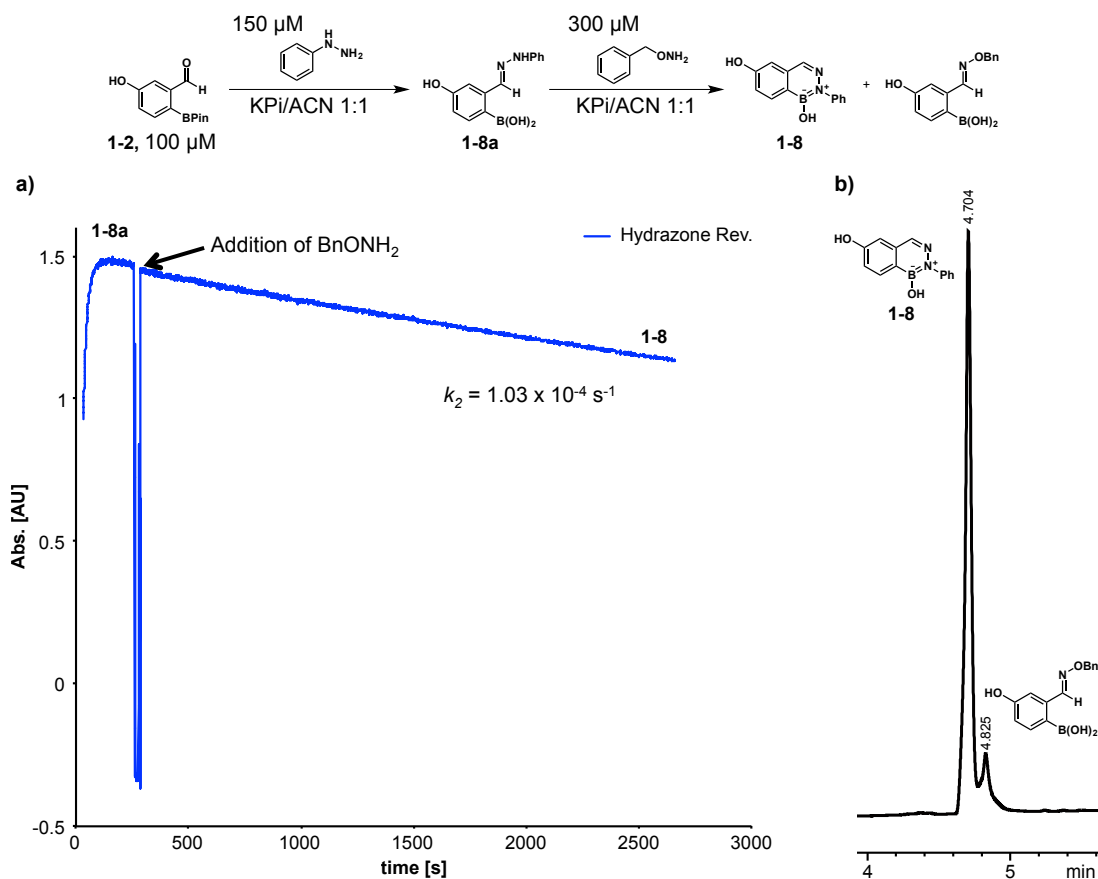


Figure 20. a) UV measurement of BIQ **1-8** formation with the addition of an excess of benzylhydroxylamine after 3.5 min. The reaction was carried out in 1:1 phosphate buffer/acetonitrile to slow down the BIQ formation. b) UPLC-MS analysis of the reaction products after 13 h. 85% of the BIQ product was found, along with 15% of the corresponding oxime product.

We designed an experiment, in which we could show that the intermediate hydrazone is reversible, however, BIQ formation was still found to be faster. Our model aldehyde **1-2** was treated with phenylhydrazine at 100 μM and after a few minutes when all of the aldehyde had been converted to the hydrazone, a three times excess of benzylhydroxylamine was added to the mixture (**Figure 20a**). Subsequent analysis of the reaction products revealed a BIQ:oxime ratio of 17:3 (**Figure 20b**). This result showed that the BIQ formation was faster than the hydrazone hydrolysis with subsequent oxime formation and only a minor amount of the hydrazone was hydrolyzed during the reaction. In addition, the organic solvent ratio in this assay had to be increased to 50% acetonitrile in order to slow down BIQ formation by a factor of 25. Under conditions with a lower organic solvent ratio (10%), no oxime product was found at all. With this assay, we could demonstrate the reversibility of the intermediate hydrazone boronic acid, along with the dependence of the BIQ formation on the amount of organic solvent.

3.2 BIQ Syntheses and Fluorescence Property Analysis

BIQ Synthesis

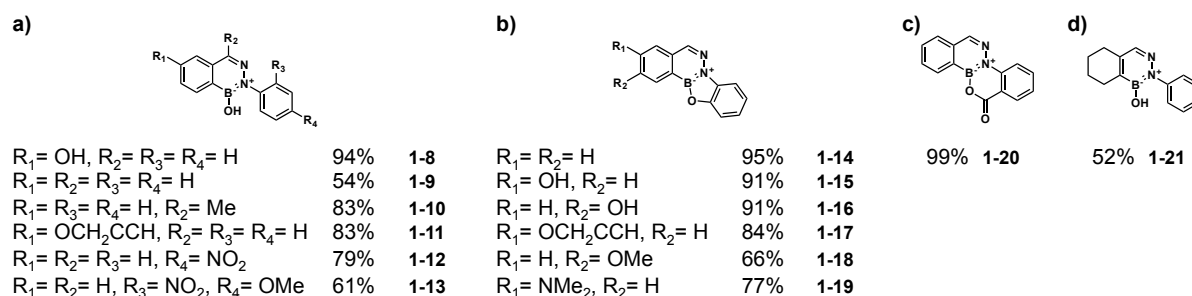


Figure 21. a) BIQ series with a B-OH group. b) BIQ series with fused ring systems consisting of three six-membered and one five-membered ring. c) A BIQ with a fused ring system consisting of four six-membered rings. d) BIQ that was generated from an alkene boronic acid, containing an aliphatic ring.

After the evaluation of the BIQ formation characteristics, we enlarged the scope of the reaction by introducing different substituents (**Figure 21a**). We synthesized a completely unsubstituted BIQ (**1-9**), as well as compounds with methyl (**1-10**), alkoxy (**1-11** and **1-13**) and nitro substituents (**1-12** and **1-13**). Performing the reaction with 2-hydroxyphenylhydrazine yielded a series of BIQs with fused ring systems and hydroxy, alkoxy and alkylamino substituents (**1-14 - 1-19**, **Figure 21b**). The aromatic alcohol of the hydrazine coupling partner underwent a dehydration step with the boron-hydroxy group to form the five-membered heterocycle. The same process occurred with 2-carboxyphenylhydrazine, giving a BIQ with a fused ring system, composed of four six-membered rings (**1-20**, **Figure 21c**). We also showed that the BIQ formation process was not restricted to arylboronic acid carbonyls, but an alkeneboronic acid with an aldehyde in the *ortho*-position was also suitable for BIQ formation (**1-21**, **Figure 21d**). We also considered the synthesis of an indole-based BIQ, which we could obtain from the reaction of an indoleboronic acid aldehyde and 2-hydroxyphenylhydrazine. However, the compound turned out to be unstable in solution. Complete decomposition was observed after a few hours. Moreover, the indole BIQ formed dimers over the boron and oxygen atoms instead of the intramolecular cyclization with the aromatic hydroxyl group.

The BIQ synthesis did not require any precautions in terms of inertness. The reaction proceeded best in a water:DMSO 4:1 mixture, even though some starting materials and intermediates/products were poorly soluble under these conditions. The reaction proceeded to completeness within two hours, whereby no side products were observed. Only in the case of BIQs with electron-withdrawing nitro-substituents (**1-12** and **1-13**), prolonged reaction times and acidification were necessary to drive the reaction to completion. **Figure 21** shows that the BIQs were in general obtained in high yields (>75%) with only few exceptions. Loss of product mainly occurred during the work-up of the reaction mixtures.

Fluorescence Property Analysis

The fused ring structure and modularity of BIQs opens up the potential for the creation of new fluorogenic molecules. Especially the *de novo* construction of the boron-nitrogen aromatic ring might enable turn-on fluorescence. The fluorescence potential of boron-heterocycles has been shown in previous reports^[83,84] and the BODIPY fluorophores, a boron-containing compound family, are frequently used for fluorescence assays.^[81]

We initially measured the UV and fluorescence properties of the BIQs without fused ring systems (**Figure 21a**). Although all of these compounds showed fluorescence at excitation wavelengths

<300 nm, the efficiencies of most of these compounds was not useful for practice. The fluorescence emission at higher excitation wavelengths was weak. We expected the B-OH group to potentially act as a fluorescence quencher, which is why we synthesized the second series of BIQs (**Figure 21b**) that contained the additional annulation. We also hoped to achieve a substantial red shift by this additional annulation and the decoration of the BIQs with different substituents. In general, the hydroxyl and alkoxy substituents did not improve the fluorescence properties of these molecules (**1-15** - **1-18**). Incorporation of the electron-rich dimethylamino group in the boronic acid component (**1-19c**), however, created a good blue fluorophore (**Figure 22**).

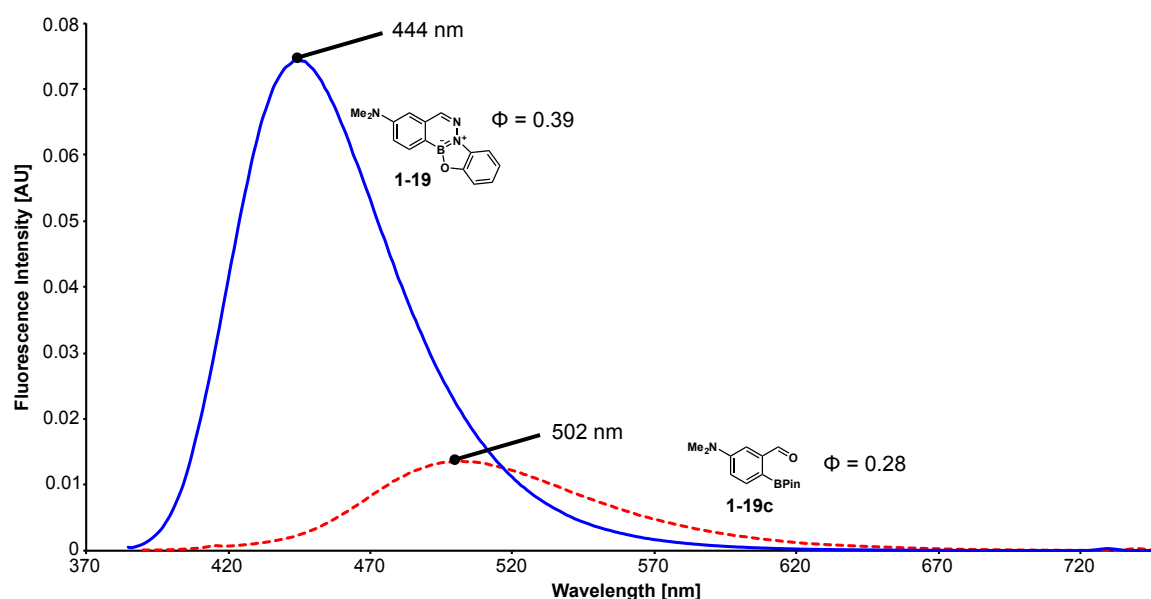


Figure 22. Fluorescence emission spectra of BIQ **1-19** and its precursor aldehyde **1-19c**. The spectra were recorded at 10 μ M in DMSO. Excitation wavelength **1-19**: 365 nm, excitation wavelength **1-19c**: 370 nm.

The aldehyde precursor **1-19c** showed fluorescence in DMSO with a maximum at 502 nm and a fluorescence quantum yield (Φ , ratio of fluorescently emitted photons per absorbed photon, value between 0 and 1) of 28%. BIQ formation with **1-19c** yielded a five-fold increase in fluorescence and an improvement in fluorescence quantum yield ($\Phi = 39\%$) with maximum emission at 444 nm. A remarkable finding for BIQ **1-19** and its precursor **1-19c** were the large Stokes shifts of 83 nm and 126 nm, respectively (see **Figure 30**). The Stokes shift is defined as the difference between excitation and emission maximum. A large Stokes shift (>80 nm) is beneficial, due to the reduced risk of interferences between the excitation source and the fluorescent emission. This is, in particular, important for fluorescent assays with high signal-to-noise ratios.^[94] Many common fluorophores exhibit rather small Stokes shifts (e.g. fluorescein with 24 nm or BODIPY 505 with 10 nm Stokes shifts). Protonation of the dimethylamino substituent did not change any fluorescence properties. The major drawback of our turn-on fluorophore was the blue shift in the emission spectrum upon BIQ formation. For biological applications, a red shift is desirable, because in the region of blue fluorophores, many other biological molecules (e.g. tryptophan, NADH, flavins) also show fluorescence emission. However, fine-tuning of the substitution pattern of the BIQ might reveal an improved turn-on fluorophore that emits in the green or even yellow to red spectrum.

4. Conclusions and Perspective

In bioconjugation techniques, α -effect-amine condensations are frequently used for the modification of biomolecules. We showed that the boron-assisted oxime condensation proceeded at very high reaction rates, was stable in aqueous solution for more than three days and a slow reversibility was found upon the addition of a second aldehyde boronate or hydroxylamine. Ketoximes were found to be more susceptible to hydrolysis (and therefore for reversibility) than aldoximes. The proposed mechanism, proceeding over an intermediate, cyclic hemiaminal boronate, was supported by different experiments. Placing the boronic acid adjacent to the hydroxylamine showed little to no enhancement of the reaction rate compared to oxime formations under neutral conditions without boronic acid substituent. The reason for this behavior was the eight-membered cyclic intermediate, which was much less favorable than the five-membered ring intermediate. Masking of the boronic acid by covalent saturation with fluorine atoms (trifluoroborate) or the MIDA ester also resulted in a diminished reaction rate for boron-assisted oxime condensations. These results showed the importance of the free coordination site at the boron atom for the enhancement of the reaction rate.

We used the accelerated oxime formation technique to label immunoglobulin G with a fluorescein tag. The experimental results revealed a high degree of labeling and a high stability towards hydrolysis. The oxime link was found to be stable under the conditions of SDS-PAGE analysis with β -mercaptoethanol added as a reducing agent.

Boron-assisted hydrazone condensations were further explored to determine the characteristics of this conjugation type. Hydrazones with adjacent boronic acid substituents undergo a cyclization reaction to stable, heteroaromatic compounds, the borazaroisoquinolines (BIQs). We showed that BIQs formed much slower than the hydrazone precursors, however, the BIQ formation was found to be stable to hydrolysis and irreversible. Influences on BIQ formation by pH and additives could be determined as well. An acidic environment enhances the rate of BIQ formation, whereas increases in pH value slowed the reaction down. Additionally, high amounts of organic solvents decreased the BIQ reaction rate. Biological additives, such as human serum and glutathione, were found to have minor influences on BIQ generation. In a competition experiment, we found that hydrazones were reversible as well with the addition of a hydroxylamine. However, BIQ formation proceeded much faster than the hydrolysis of the hydrazone with subsequent oxime formation. Analysis of the reaction products revealed only a small amount of the oxime product, despite the excess of added hydroxylamine.

The remarkable structure of BIQs prompted us to synthesize a series of substituted BIQs with fused ring systems for the determination of the fluorescence properties. While most of the compounds showed only weak fluorescence at low emission wavelengths, one BIQ, with an electron-rich diethylamino substituent, showed good fluorescence properties. Although the aldehyde precursor revealed fluorescence in the blue region, upon BIQ formation a five-fold enhancement of the fluorescence intensity was found, along with an increase in fluorescence quantum yield. This behavior was considered as a turn-on fluorescence event. The fact that the turn-on fluorescence, upon BIQ formation, causes a blue-shift of the emission maximum was the main drawback due to the intrinsic fluorescence of several biomolecules in this region of the electromagnetic spectrum. The fluorescent BIQ and its precursor showed both a large Stokes shift (>80 nm), which is beneficial for applications in biochemistry.

Our findings on boron-assisted oxime and hydrazone/BIQ formations, along with the results from other research groups, enabled the construction of a guidebook for α -effect-amine condensations. Ketoxime and acylhydrazones were determined to be rapidly reversible and might be applied in dynamic covalent chemistry. Aldoximes are slowly hydrolyzed and may find

applications for targeted slow release or exchange. Hydrazones lead to the formation of BIQs, which are irreversible and might be used for permanent linking applications.

Boron-mediated Schiff base formations with hydroxylamines and hydrazines may find various applications for bioorthogonal labeling. However, the technique might be restricted to *in vitro* applications, since boronic acids interfere with carbohydrates that are found on the surface of cells. For that reason, the transport of the boronic acid carbonyl compound into the cell may be a big challenge. Masking of the boronic acid by coordination saturation might be considered, however, the hydrolytic stability of such boronates is usually not high enough to prevent interactions with carbohydrates in a cellular environment and meanwhile slows down the rate of the Schiff base formation.

Boronic acids are prone to oxidation, a property that may be used for targeted release of a cargo in close proximity to cancer cells. Boron-mediated oximes might be used as the linker between the cargo (e.g. an anti-cancer drug) and a transporter molecule. The prerequisite for such a system is that the linker is protected from other nucleophiles by the transporter moiety or by encapsulation of the assembly within a vesicle. The transporter molecule ensures the selective binding to the target cell proteins and the increased concentration of oxidizing agents near tumor cells destroys the boronic acid substituent. This oxidation leads to a destabilization of the oxime linkage, which enables the release of the cargo molecule to kill the tumor cell.

Due to the functionality of boronic acids, a secondary reaction, such as Suzuki-Miyaura cross-coupling, may also be considered after completed Schiff base formation. This duality of boronic acid carbonyl compounds opens up the potential for the creation of new, interesting molecules, applicable in early drug-discovery. Due to the high availability of halogenated compounds, large combinatorial libraries may be accessible by this technique.

The BIQ system may be expanded by substituent variation in order to create new, red-fluorescent molecules or turn-on fluorophores. The ease of synthesis of this compound class will enable the creation of large BIQ collections. However, BIQs are not restricted to the area of fluorescent molecules, they also represent interesting scaffolds for potential pharmacophores or functional materials.

5. Experimental Part Oxime, Hydrazone, BIQ

5.1 General Information

Reagents and Solvents

Reagents and solvents were purchased from Sigma-Aldrich, Alfa Aesar, Apollo Scientific Ltd., Fluka, Fluorochem, Fisher Scientific, Enamine, TCI, Bachem, Biosolve or Acros and were used as received. 1,4-dioxane was freshly distilled from sodium benzophenone ketyl under dry nitrogen prior to use. Buffers and HPLC eluents were prepared with nanopure water (resistivity 18.2 M Ω). KPi refers to (sodium or potassium) phosphate buffer at the stated concentration and pH.

Chromatographic Purification and Isolation

Flash chromatography was performed on SilicaFlash® gel P60 40-63 μ m (230-400 mesh, SiliCycle, Quebec) according to Still^[95] or on a Biotage Isolera four with SilicaFlash® gel packed cartridges. Reversed phase flash chromatographies were run on a Biotage Isolera four with self-packed columns (LiChroprep RP-18, 40-63 μ m silica from Merck). Preparative RP-HPLC was carried out on a Shimadzu Prominence UFLC Preparative Liquid Chromatograph.

Method A: Gemini NX-C₁₈, 5 μ m, 110 Å, 21.2 x 250 mm from Phenomenex with a flow rate of 20 mL/min, gradient: 1% (3 min)-99% (25 min)-99% (3 min) (B), monitoring and collecting the products at 254 nm. Buffer (A): 0.1% TFA (v/v) in H₂O, Buffer (B): MeCN.

The crude compound mixtures were injected as MeCN solutions. Concentration under reduced pressure was performed by rotatory evaporation at 40°C water bath temperature. Aqueous product fractions were frozen in liquid N₂ and lyophilized on a Christ Alpha 2-4 LDplus flask lyophilizer at 0.3 mbar or below.

Chromatographic Analysis

Analytical TLC was performed on Silica gel 60 F254, 0.25 mm pre-coated glass plates (Merck) and visualized by fluorescence quenching under UV light at 254 nm and subsequent KMnO₄ staining.

ESI-MS and LC-MS

ESI-MS spectra were recorded on a Bruker Esquire3000 spectrometer by direct injection in positive or negative polarity of the ion trap detector. Compounds were injected as MeOH, MeCN or H₂O solutions. High-resolution mass spectra (HRMS) were recorded by the mass spectrometric service of the University of Basel on a Bruker maXis 4G QTOF ESI mass spectrometer.

UPLC-MS was carried out on an Agilent 1290 Infinity system equipped with an Agilent 6130 Quadrupole ESI-MS using a C₁₈ column (ZORBAX Eclipse Plus RRHD, 1.8 μ m, 2.1 x 50 mm) from Agilent with a flow rate of 0.45 mL/min at 40°C. Buffer (A): 0.1% (v/v) formic acid in H₂O/1% MeCN (v/v), Buffer (B): 0.1% (v/v) formic acid in MeCN/1% H₂O (v/v) using the following gradient: 5-90% (3.5 min)-90% (1 min) (B), ESI-MS in positive ion mode of the ion trap.

Gel Electrophoresis

Gel electrophoresis was performed with a Bio-Rad PowerPac HV high-voltage power supply. Gels were prepared with an area of 83 x 83 mm and a thickness of 1.0 mm with 10 wells.

Denaturing protein gels were prepared with a 5% stacking gel (approx. 2 cm high) and a resolving gel of different percentages (approx. 6 cm). The stacking gel was prepared from 40% acrylamide/bis-acrylamide 37.5:1 (Fisher Scientific) with 0.1% (m/v) SDS, 0.2% (v/v) TEMED and 0.4% (v/v) APS (10% APS solution in H₂O) in 125 mM TRIS, pH 6.8. The resolving gel was prepared from 40% acrylamide/bis-acrylamide 37.5:1 (Fisher Scientific) with 0.1% (m/v) SDS, 0.1% (v/v) TEMED and 0.3% (v/v) APS (10% APS solution in H₂O) in 375 mM TRIS, pH 8.8. Protein samples were treated with 2X loading dye (66 mM TRIS pH 6.8, 2% (m/v) SDS, 0.01% (m/v) bromophenol blue, 30% (v/v) glycerol) and denatured at 95°C for 5 min prior to gel loading. Visualization was achieved with Coomassie Blue staining. Gel imaging and analysis was performed on a Bio-Rad ChemiDoc MP system.

Running buffer: 193 mM glycine, 25 mM TRIS, 0.1% (m/v) SDS.

Nuclear Magnetic Resonance (NMR)

¹H, ¹³C and 2D-NMR spectra were acquired on BrukerAvance (250, 400, 500 or 600 MHz proton frequency) spectrometers at 298.15 K. Chemical shifts (δ values) are referenced to the solvent's residual peak and reported in ppm. Multiplicities are reported as follows: s = singlet, sbr = broad singlet, d = doublet, t = triplet, q = quartet, p = quintet, m = multiplet or unresolved and combinations of these multiplicities (e.g. dd, dt, td etc.). Coupling constants *J* are given in Hz. Due to the quadrupole effect, carbon atoms attached to a boron atom are invisible in ¹³C spectra.

Special NMR Assays

NMR experiments were performed at 298 K either on a Bruker Avance III NMR spectrometer operating at 600.13 MHz and equipped with a direct observe 5 mm BBFO smart probe or a Bruker Avance III HD NMR spectrometer operating at 600.13 MHz equipped with a cryogenic 5 mm four-channel QCI probe (H/C/N/F). All ¹H experiments were recorded using the same parameters (256 scans, relaxation delay of 4.0 s and an acquisition time of 2.00 s) and were processed using TopSpin software from Bruker. Line broadening of 0.5 Hz and automated baseline correction was applied after manual phase correction. Chemical shifts were referenced to the residual solvent peak of acetonitrile (1.94 ppm) and TMSP-*d*₄ (3-(trimethylsilyl)-2,2',3,3'-tetra-deuteriopropionic acid) was used as an internal standard. The pH (no correction for deuterium) was adjusted by addition of concentrated DCl from ABCR (DCl, 38% wt% in D₂O, 99.5% atom%D) measured with a 827 pH Lab - Metrohm equipped with a glass minitrode, which was pre-rinsed with D₂O. Standard NMR tubes (throw away quality) were rinsed with D₂O and dried in the oven prior to use. Stock solutions were mixed in the NMR tube, vortexed and sonicated before measurement.

UV Assays for BIQ Experiments

All measurements were conducted on a Shimadzu UV-1800 UV-Vis spectrophotometer (190-1100 nm, bandwidth 1 nm) using 10 x 10 mm quartz cuvettes from Hellma Analytics (3500 μ L). The Human Serum (HS) was of type AB (male). Cuvettes were rinsed with deionized H₂O followed by technical EtOH and dried under a N₂ stream prior to use. For all UV measurements the boronic acid was dissolved in the solvent and measured for 10 s. The hydrazine stock solution (freshly prepared, could not be used for longer than 1 h) was directly added into the cuvette, gently shaken three times and the measurement was continued. Formation of the

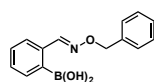
hydrazone and BIQ were recorded in triplicate at 340 nm. The absorbance was measured every 0.5 s over a period of about 1500 s. The recorded data was analyzed with MS Excel. Hydrazone formation was too fast to determine exact rate constants by the described method. The intramolecular BIQ formation showed first order behavior and was measured by the signal decrease at 340 nm (only hydrazone visible). Extinction coefficients for the hydrazone intermediate at 340 nm were determined by addition of 1.5 equivalents of the corresponding hydrazine to a known concentration of aldehyde (typically 10 μM), assuming that the highest absorbance is equal to complete conversion to the hydrazone (Lambert-Beer law).

Fluorescence Measurements

Fluorescence emission measurements were performed on a Shimadzu RF-5301PC spectrofluoro-photometer with 3 nm spectral bandwidth, using 10 x 10 mm quartz cuvettes from Hellma Analytics (3500 μL). Quantum yield measurements were done on a Hamamatsu Quantaaurus-QY Absolute PL quantum yields measurement system C11347-11 with an integrating sphere. All samples were prepared as 10 μM samples in DMSO. UV and fluorescence files were processed with MS Excel.

5.2 Compound Synthesis

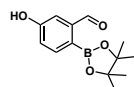
(E)-2-(((Benzyloxy)imino)methyl)phenyl)boronic acid **1-1**



2-Formylphenylboronic acid (50.0 mg, 333.0 μmol , 1.0 eq.) and *O*-benzylhydroxylamine hydrochloride (58.5 mg, 367.0 μmol , 1.1 eq.) were dissolved in MeOH (2 mL) and the mixture was stirred at RT for 21 h, after which UPLC-MS analysis showed complete conversion to the desired product. The solvent was removed by rotary evaporation and the residue was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 25 g, H₂O/MeCN, UV) to yield the desired product as a white solid (73 mg, **86%**). Analytical data was in agreement with reported data.^[63]

¹H NMR (400 MHz, CD₃CN) δ /ppm: 8.50 (s, 1H), 7.76 – 7.70 (m, 1H), 7.62 (dd, *J* = 7.4, 1.7 Hz, 1H), 7.46 – 7.36 (m, 5H), 7.34 – 7.26 (m, 2H), 6.48 (s, 2H), 5.17 (s, 2H).

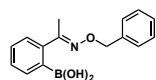
2-Formyl-4-hydroxyphenylboronic acid pinacol ester **1-2**



Under an inert atmosphere, 2-bromo-5-hydroxybenzaldehyde (500.0 mg, 2.5 mmol, 1.0 eq.), *bis*(pinacolato)diboron (821.0 mg, 3.23 mmol, 1.3 eq.), Pd(dppf)Cl₂ (182.0 mg, 249.0 μmol , 10 mol%) and potassium acetate (732.0 mg, 7.5 mmol, 3.0 eq.) were suspended in freshly distilled dioxane (10 mL). The mixture was heated to reflux (110°C) for 3 h. After cooling, the solvent was removed by rotary evaporation. The residue was diluted with DCM (40 mL) and washed with H₂O (3 x 40 mL). The combined aqueous layers were extracted with DCM (3 x 15 mL) and the combined organic layers were washed with brine (40 mL). The organic phase was dried over Na₂SO₄, concentrated by rotary evaporation and the crude was purified by flash column chromatography (Silica, 110 g, cyclohexane \rightarrow 10:1 \rightarrow 7.5:1 \rightarrow 5:1 cyclohexane:EtOAc, *R_f* = 0.3, UV/KMnO₄). Product containing fractions were combined and concentrated *in vacuo* to yield the desired product as a white solid (300 mg, **49%**). Analytical data was in agreement with reported data.^[96]

^1H NMR (400 MHz, DMSO- d_6) δ /ppm: 10.45 (s, 1H), 10.26 (s, 1H), 7.68 (d, J = 8.2 Hz, 1H), 7.25 (d, J = 2.5 Hz, 1H), 7.06 (dd, J = 8.1, 2.5 Hz, 1H), 1.31 (s, 12H).

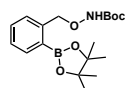
(*E*)-2-(1-((Benzyloxy)imino)ethyl)phenylboronic acid **1-3**



2-Acetylphenylboronic acid (50.0 mg, 305.0 μmol , 1.0 eq.) and *O*-benzylhydroxylamine hydrochloride (53.5 mg, 335.0 μmol , 1.1 eq.) were dissolved in MeOH (2 mL) and the mixture was stirred at RT for 15 h, after which UPLC-MS analysis showed full conversion to the desired product. The solvent was removed by rotary evaporation and the residue was purified by preparative RP-HPLC (Method A). Product containing fractions were combined and lyophilized to yield the desired product as a white solid (56 mg, **68%**). Analytical data was in agreement with reported data.^[63]

^1H NMR (400 MHz, CD_3CN) δ /ppm: 7.47 – 7.40 (m, 4H), 7.40 – 7.31 (m, 4H), 7.31 – 7.24 (m, 1H), 5.19 (s, 2H), 2.23 (s, 3H).

2-(*N*-Boc-aminooxymethyl)phenylboronic acid pinacol ester **1-4a**



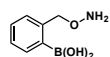
Under a N_2 atmosphere in a dry Schlenk tube, *N*-Boc-hydroxylamine (33.6 mg, 250.0 μmol , 1.5 eq.) and 2-bromomethylphenylboronic acid pinacol ester (50.0 mg, 170.0 μmol , 1.0 eq.) were dissolved in dry MeCN (0.5 mL) and cooled to 0°C . NaH (9.0 mg, 340.0 μmol , 2.0 eq.) was slowly added and the mixture was stirred for 30 min. The cooling bath was removed and the mixture was stirred for another 4 h whereupon UPLC-MS analysis showed complete conversion. The mixture was diluted with H_2O (3 mL) and EtOAc (3 mL) and the layers were separated. The aqueous layer was extracted with EtOAc (2 x 5 mL) and the combined organic layers were washed with H_2O (5 mL), brine (5 mL), dried over Na_2SO_4 and the solvent was removed by rotary evaporation. The residue was purified by preparative RP-HPLC (Method A, no TFA) and product containing fractions were combined and lyophilized to yield the desired product as a white solid (20 mg, **34%**).

^1H NMR (400 MHz, CDCl_3) δ /ppm: 7.81 (dt, J = 7.3, 1.0 Hz, 1H), 7.45 – 7.42 (m, 2H), 7.35 – 7.30 (m, 1H), 7.18 (s, 1H), 5.14 (s, 2H), 1.46 (s, 9H), 1.35 (s, 12H).

^{13}C NMR (101 MHz, CDCl_3) δ /ppm: 156.71, 141.82, 135.98, 131.03, 129.63, 127.72, 83.97, 81.52, 77.87, 28.37, 25.02.

HRMS (ESI): $\text{C}_{18}\text{H}_{29}\text{BNO}_5^+$ *calcd.*: 350.2133, *found*: 350.2137.

2-Aminooxymethylphenylboronic acid **1-4**



1-4a (35.0 mg, 100.0 μmol , 1.0 eq.) was dissolved in CH_2Cl_2 (0.5 mL) while cooling to 0°C . Trifluoroacetic acid (300.0 μL , 3.9 mmol, 39.1 eq.) was added dropwise over 2 min. After 10 min, the cooling bath was removed and the mixture was stirred at RT for 20 min. H_2O (0.25 mL) and MeCN (0.1 mL) were added and the biphasic mixture was stirred at room temperature for 3 h. The volatiles were removed by rotary evaporation and the mixture was purified by preparative RP-HPLC (Method A, no TFA). Product containing fractions were combined and lyophilized to

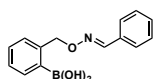
yield the desired product as a white solid (5 mg, **30%**). NMR showed a mixture of the free boronic acid and pinacol ester in a ratio of 3:2.

^1H NMR (400 MHz, DMSO- d_6) δ /ppm: 7.68 – 7.64 (m, 1H), 7.52 – 7.49 (m, 1H, pinacol ester), 7.22 – 7.18 (m, 2H), 7.17 – 7.13 (m, 2H, pinacol ester), 7.10 – 7.07 (m, 1H), 7.04 – 7.01 (m, 1H, pinacol ester), 5.04 (s, 2H), 4.78 (s, 2H, pinacol ester), 1.14 (s, 12H, pinacol ester).

^{13}C NMR (500 MHz, 2D NMR, DMSO- d_6) δ /ppm: 142.64, 138.25, 132.12, 126.29, 126.14, 124.64, 72.83 (boronic acid); 140.02, 131.5, 126.13, 126.00, 123.52, 79.22, 71.65, 25.88 (pinacol ester).

HRMS (ESI): $\text{C}_{13}\text{H}_{21}\text{BNO}_3^+$ *calcd.*: 250.1609, *found*: 250.1611 (pinacol ester).

(*E*)-(2-(((Benzylideneamino)oxy)methyl)phenyl)boronic acid **1-5**



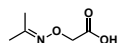
1-4 (3.0 mg, 20.0 μmol , 1.0 eq.) and benzaldehyde (2.8 μL , 30.0 μmol , 1.5 eq.) were dissolved in MeOH (0.2 mL) and the solution was stirred at RT overnight. The mixture was directly purified by preparative RP-HPLC (Method A). Product containing fractions were combined and lyophilized to yield the desired product as a white solid (1 mg, **22%**).

^1H NMR (500 MHz, CD_3CN) δ /ppm: 8.17 (s, 1H), 7.67 – 7.64 (m, 1H), 7.57 – 7.54 (m, 2H), 7.44 – 7.39 (m, 5H), 7.35 – 7.31 (m, 1H), 6.88 (s, 2H), 5.29 (s, 2H).

^{13}C NMR (500 MHz, 2D NMR, DMSO- d_6) δ /ppm: 150.97, 142.00, 135.17, 132.66, 131.15, 130.80, 130.65, 129.72, 128.46, 127.92, 76.94.

HRMS (ESI): $\text{C}_{14}\text{H}_{14}\text{BNNaO}_3^+$ *calcd.*: 278.0959, *found*: 278.0961.

2-(((Propan-2-ylideneamino)oxy)acetic acid **1-6a**



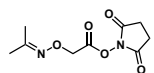
Carboxymethoxylamine hemihydrochloride (500.0 mg, 2.3 mmol, 1.0 eq.) was dissolved in acetone (6.6 mL, 89.3 mmol, 39.1 eq.) and H_2O (0.5 mL). The solution was stirred at RT for 2 h, after which the volatiles were removed by rotary evaporation and the residue was lyophilized to yield the desired product as a white solid (543 mg, **91%**).

^1H NMR (400 MHz, DMSO- d_6) δ /ppm: 12.57 (sbr, 1H), 4.44 (s, 2H), 1.82 (s, 3H), 1.79 (s, 3H).

^{13}C NMR (101 MHz, DMSO- d_6) δ /ppm: 171.34, 155.35, 69.54, 21.17, 15.50.

HRMS (ESI): $\text{C}_5\text{H}_9\text{NNaO}_3^+$ *calcd.*: 154.0475, *found*: 154.0476.

Succinimidyl 2-(((propan-2-ylideneamino)oxy)acetate **1-6**



1-6a (50.0 mg, 381.0 μmol , 1.0 eq.), *N*-hydroxysuccinimide (65.8 mg, 572.0 μmol , 1.5 eq.) and DMAP (4.7 mg, 38.1 μmol , 10 mol%) were dissolved in DCM (3 mL) and the mixture was cooled in an ice bath. Diisopropylcarbodiimide (89.6 μL , 572.0 μmol , 1.5 eq.) was added dropwise over 2 min. The mixture was stirred in the ice bath for 1 h, warmed to RT and continued stirring overnight. The precipitate was filtered off and the filtrate was washed with aqueous HCl (50 mM, 4 x 10 mL), brine (10 mL) and was dried over Na_2SO_4 . The solvent was removed by rotary evaporation, the residue was suspended in EtOAc (3 mL) and placed in the freezer at -20°C

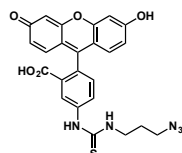
overnight. The precipitate was filtered off, washed with EtOAc (3 x 2 ml, -20°C) and the combined filtrates were concentrated *in vacuo* to yield the desired product as slightly yellow, thick oil that slowly crystallized (90 mg, **99%**). The product was used without further purification.

^1H NMR (400 MHz, CDCl_3) δ /ppm: 4.90 (s, 2H), 2.84 (s, 4H), 1.92 (s, 3H), 1.88 (s, 3H).

^{13}C NMR (101 MHz, CDCl_3) δ /ppm: 168.86, 165.91, 158.17, 68.07, 25.72, 21.77, 15.93.

HRMS (ESI): $\text{C}_9\text{H}_{12}\text{N}_2\text{NaO}_5^+$ *calcd.*: 251.0638, *found*: 251.0640.

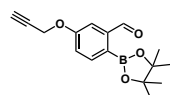
Azidopropyl-(fluoresceiny)thiourea **1-7a**



A solution of fluorescein isothiocyanate (50.0 mg, 128.0 μmol , 1.0 eq.) in DMF (1 mL) was added dropwise to a solution of 3-azidopropylamine **2-1** (14.0 mg, 140.0 μmol , 1.1 eq.) in DMF (2 mL) under a nitrogen atmosphere. The solution was stirred for 20 h at RT, after which the mixture was directly purified by preparative RP-HPLC (Method A). Product containing fractions were combined and lyophilized to yield the desired product as an orange solid (30 mg, **48%**). Analytical data was in agreement with reported data.^[97]

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 10.13 (brs, 2H), 9.97 (brs, 1H), 8.22 – 8.15 (m, 2H), 7.73 (dd, $J = 8.2, 4.4$ Hz, 1H), 7.18 (d, $J = 8.3$ Hz, 1H), 6.67 (d, $J = 2.2$ Hz, 2H), 6.62 (d, $J = 1.8$ Hz, 1H), 6.59 (s, 1H), 6.57 (d, $J = 2.2$ Hz, 2H), 3.59 (t, $J = 5.6$ Hz, 2H), 3.44 (t, $J = 6.7$ Hz, 2H), 1.85 (p, $J = 6.8$ Hz, 2H).

2-Formyl-4-(propargyloxy)phenylboronic acid pinacol ester **1-7b**



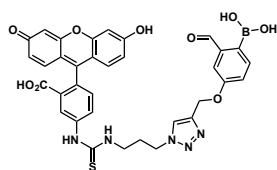
Under a N_2 atmosphere **1-2** (50.0 mg, 200.0 μmol , 1.0 eq.) and propargyl bromide (26.9 μL , 240.0 μmol , 1.3 eq., 80% solution in toluene) were dissolved in DMF (0.5 mL) and K_2CO_3 (41.8 mg, 300.0 μmol , 1.5 eq.) was added. The mixture was stirred at room temperature for 21 h, after which the mixture was diluted with H_2O (15 mL) and EtOAc (15 mL). The layers were separated and the aqueous layer was washed with EtOAc (2 x 15 mL). The combined organic layers were washed with brine (10 mL), dried over Na_2SO_4 and the solvent was removed by rotary evaporation. The crude product was purified by flash chromatography (Silica, 5 g, 10:1 cyclohexane/EtOAc, UV/ KMnO_4 , $R_f = 0.24$) to yield the desired product as a white solid (37 mg, **64%**).

^1H NMR (400 MHz, CDCl_3) δ /ppm: 10.67 (s, 1H), 7.89 (d, $J = 8.3$ Hz, 1H), 7.57 (d, $J = 2.7$ Hz, 1H), 7.19 (dd, $J = 8.3, 2.7$ Hz, 1H), 4.77 (d, $J = 2.4$ Hz, 2H), 2.53 (t, $J = 2.4$ Hz, 1H), 1.37 (s, 12H).

^{13}C NMR (101 MHz, CDCl_3) δ /ppm: 194.72, 159.92, 143.64, 138.11, 120.66, 111.63, 84.38, 77.89, 76.23, 55.94, 25.02.

LRMS (ESI): $\text{C}_{16}\text{H}_{19}\text{BNaO}_4^+$ *calcd.*: 309.13, *found*: 309.24.

Fluorescein boronic acid aldehyde **1-7**



In a 1.5 mL Eppendorf tube **1-7a** (25.0 mg, 51.1 μmol , 1.0 eq.), **1-7b** (16.1 mg, 56.2 μmol , 1.1 eq.) and sodium ascorbate (6.1 mg, 30.6 μmol , 0.6 eq.) were dissolved in *t*-BuOH:H₂O (1:1, 310 μL each). CuSO₄ pentahydrate (3.8 mg, 15.3 μmol , 0.3 eq.) was added and the mixture was agitated at RT for 21 h, after which UPLC-MS analysis showed complete conversion of the starting material. The mixture was diluted with MeCN (1 mL) and purified by preparative RP-HPLC (Method A). Product containing fractions were combined and lyophilized to yield the product as an orange solid (11 mg, **31%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ /ppm: 10.20 (s, 1H), 10.12 (brs, 2H), 9.97 (brs, 1H), 8.31 (s, 1H), 8.24 – 8.13 (m, 3H), 7.76 – 7.71 (m, 1H), 7.61 (d, *J* = 8.2 Hz, 1H), 7.50 (d, *J* = 2.6 Hz, 1H), 7.30 (dd, *J* = 8.2, 2.7 Hz, 1H), 7.19 (d, *J* = 8.3 Hz, 1H), 7.09 – 7.01 (m, 1H), 6.67 (d, *J* = 2.3 Hz, 2H), 6.59 (s, 1H), 6.57 (d, *J* = 2.3 Hz, 1H), 5.24 (s, 2H), 4.46 (t, *J* = 7.0 Hz, 2H), 3.58 – 3.52 (m, 2H), 2.16 (p, *J* = 7.1 Hz, 3H).

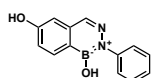
¹³C NMR (500 MHz, 2D NMR, DMSO-*d*₆) δ /ppm: 193.83, 159.18, 158.33, 151.53, 142.04, 141.02, 140.87, 135.19, 131.37, 129.36, 128.72, 126.20, 124.36, 123.77, 119.43, 116.36, 113.33, 112.25, 109.39, 101.93, 82.72, 60.93, 47.02, 40.80, 28.88.

HRMS (ESI): C₃₅H₃₁BN₅O₉S⁺ *calcd.*: 708.1930, *found*: 708.1932 (mono-methyl boronate).

General Procedure for the Synthesis of 4,3-Borazaroisoquinolines (BIQ)

Boronic acid aldehyde (1.0 eq.) and phenylhydrazine (1.4 eq.) were suspended in DMSO/H₂O (1:4) and the suspension was stirred at room temperature for 2 h. The mixture was diluted with MeCN until a clear solution was obtained. The product was purified by preparative RP-HPLC (Method A) and the product-containing fractions were lyophilized.

2-Phenylbenzo[d][1,2,3]diazaborinine-1,6(2H)-diol **1-8**



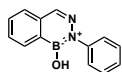
1-8 was synthesized according to the general procedure. **1-2** (20.0 mg, 80.6 μmol , 1.0 eq.) and phenylhydrazine (11.1 μL , 113.0 μmol , 1.4 eq.) were mixed in the solvent (1 mL). The product was isolated as a white solid (18 mg, **94%**).

¹H NMR (400 MHz, CD₃CN) δ /ppm: 8.06 – 8.02 (m, 1H), 8.01 (s, 1H), 7.54 – 7.50 (m, 2H), 7.45 – 7.40 (m, 2H), 7.29 – 7.24 (m, 1H), 7.17 – 7.13 (m, 1H), 7.12 – 7.10 (m, 1H).

¹³C NMR (500 MHz, 2D NMR, CD₃CN) δ /ppm: 160.63, 147.20, 140.05, 138.50, 134.04, 129.37, 126.33, 125.90, 119.23, 112.15.

HRMS (ESI): C₁₃H₁₂BN₂O₂⁺ *calcd.*: 239.0986, *found*: 239.0989.

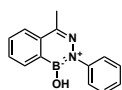
2-Phenylbenzo[d][1,2,3]diazaborinin-1(2H)-ol **1-9**



1-9 was synthesized according to the general procedure. 2-Formylphenylboronic acid (2-FPBA) (5.0 mg, 33.3 μmol , 1.0 eq.) and phenylhydrazine (4.6 μL , 46.7 μmol , 1.4 eq.) were mixed in the solvent (1 mL). The product was isolated as a white solid (4 mg, **54%**). Analytical data was in agreement with reported data.^[68]

^1H NMR (400 MHz, CD_3CN) δ /ppm: 8.19 (d, $J = 7.6$ Hz, 1H), 8.14 (s, 1H), 7.82 – 7.74 (m, 2H), 7.68 (ddd, $J = 8.2, 4.9, 3.5$ Hz, 1H), 7.57 – 7.51 (m, 2H), 7.49 – 7.41 (m, 2H), 7.32 – 7.26 (m, 1H), 6.32 (s, 1H).

4-Methyl-2-phenylbenzo[d][1,2,3]diazaborinin-1(2H)-ol **1-10**



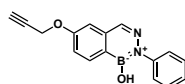
1-10 was synthesized according to the general procedure. 2-Acetylphenylboronic acid (5.0 mg, 30.5 μmol , 1.0 eq.) and phenylhydrazine (4.2 μL , 42.7 μmol , 1.4 eq.) were mixed in the solvent (250 μL). The product was isolated as a yellow solid (6 mg, **83%**).

^1H NMR (400 MHz, CD_3CN) δ /ppm: 8.21 – 8.17 (m, 1H), 7.95 – 7.91 (m, 1H), 7.82 – 7.77 (m, 1H), 7.70 – 7.65 (m, 1H), 7.56 – 7.51 (m, 2H), 7.47 – 7.41 (m, 2H), 7.30 – 7.24 (m, 1H), 2.58 (s, 3H).

^{13}C NMR (500 MHz, HMQC, HMBC, CD_3CN) δ /ppm: 147.01, 144.40, 136.10, 132.29, 132.00, 129.78, 129.38, 126.44, 126.33, 126.04, 20.27.

HRMS (ESI): $\text{C}_{14}\text{H}_{14}\text{BN}_2\text{O}^+$ *calcd.*: 237.1194, *found*: 237.1198.

2-Phenyl-6-(propargyloxy)benzo[d][1,2,3]diazaborinin-1(2H)-ol **1-11**



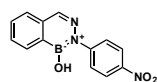
19 was synthesized according to the general procedure. **1-7b** (5.0 mg, 17.5 μmol , 1.0 eq.) and phenylhydrazine (2.4 μL , 24.5 μmol , 1.4 eq.) were mixed in the solvent (250 μL). The product was isolated as a white solid (4 mg, **83%**).

^1H NMR (400 MHz, CD_3CN) δ /ppm: 8.14 – 8.11 (m, 1H), 8.08 (s, 1H), 7.55 – 7.51 (m, 2H), 7.46 – 7.41 (m, 2H), 7.32 – 7.29 (m, 2H), 7.29 – 7.27 (m, 1H), 6.25 (s, 1H), 4.89 (d, $J = 2.4$ Hz, 2H), 2.86 (t, $J = 2.3$ Hz, 1H).

^{13}C NMR (500 MHz, HMQC, HMBC, CD_3CN) δ /ppm: 160.83, 147.03, 139.86, 138.27, 133.82, 129.29, 126.56, 125.91, 119.21, 110.93, 79.23, 77.11, 56.59.

HRMS (ESI): $\text{C}_{16}\text{H}_{14}\text{BN}_2\text{O}_2^+$ *calcd.*: 277.1143, *found*: 277.1146.

2-(4-Nitrophenyl)benzo[d][1,2,3]diazaborinin-1(2H)-ol **1-12**



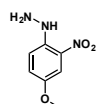
2-Formylphenylboronic acid (10.0 mg, 66.7 μmol , 1.0 eq.) and 4-nitrophenylhydrazine (22.0 mg, 93.4 μmol , 1.4 eq., 65% in H_2O) were dissolved in the solvent (1.0 mL). The suspension was acidified with aqueous HCl (100 μL , 1N) and stirred for 32 h at RT. Purification was performed according to the general procedure. The product was isolated as a yellow solid (14 mg, **79%**).

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ /ppm: 9.64 (s, 1H), 8.44 (d, $J = 7.4$ Hz, 1H), 8.33 – 8.27 (m, 3H), 8.03 – 7.97 (m, 2H), 7.88 – 7.78 (m, 2H), 7.71 (td, $J = 7.1, 1.5$ Hz, 1H).

^{13}C NMR (101 MHz, $\text{DMSO-}d_6$) δ /ppm: 152.28, 143.52, 140.86, 134.70, 132.00, 131.97, 129.63, 127.40, 123.96, 123.81.

HRMS (ESI): $\text{C}_{13}\text{H}_{11}\text{BN}_3\text{O}_3^+$ *calcd.*: 268.0888, *found*: 268.0893.

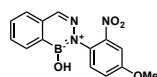
2-Nitro-4-methoxyphenylhydrazine **1-13a**



4-Methoxy-2-nitroaniline (5.0 g, 29.7 mmol, 1.0 eq.) was suspended in HCl (35%, 12.5 mL) and H_2O (12.5 mL). The mixture was cooled in an ice/salt bath to -10°C . A solution of NaNO_2 (2.3 g, 32.7 mmol, 1.1 eq.) in H_2O (5 mL) was added dropwise over 10 min, keeping the temperature between -10°C and -5°C (no gas evolution). The dark solution was stirred for 10 min, after which it was slowly transferred to a solution of SnCl_2 (16.9 g, 89.2 mmol, 3.0 eq.) in HCl (35%, 20 mL) while cooling in an ice/salt bath. The suspension was stirred for 1 h in the cooling bath and was then warmed to RT in the course of 45 minutes. The precipitate was filtered off and dissolved in H_2O (180 mL). NaOAc (15 g) was added, whereby a dark red precipitate formed. The solid was filtered off, washed with H_2O (15 mL) and was dried *in vacuo* overnight. The crude product was recrystallized in cyclohexane:EtOH (1:1, 80 mL) with hot filtration over G4. The desired product was isolated as a dark red solid (853 mg, **16%**). Analytical data was in agreement with the reported data.^[98]

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ /ppm: 8.99 (s, 1H), 7.61 (d, $J = 9.5$ Hz, 1H), 7.40 (d, $J = 2.9$ Hz, 1H), 7.29 (ddd, $J = 9.5, 3.0, 0.7$ Hz, 1H), 4.66 (s, 2H), 3.74 (s, 3H).

2-(4-Methoxy-2-nitrophenyl)benzo[d][1,2,3]diazaborinin-1(2H)-ol **1-13**



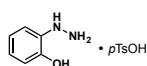
2-Formylphenylboronic acid (10.0 mg, 66.7 μmol , 1.0 eq.) and **1-13a** (17.1 mg, 93.3 μmol , 1.4 eq.) were dissolved in the solvent (1.0 mL). The suspension was acidified with aqueous HCl (100 μL , 1N) and was stirred for 67 h at 70°C . Purification was performed according to the general procedure. The product was isolated as a brown solid (12 mg, **61%**).

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ /ppm: 9.12 (sbr, 1H), 8.36 – 8.31 (m, 1H), 8.15 (s, 1H), 7.84 – 7.76 (m, 2H), 7.70 – 7.65 (m, 1H), 7.57 (d, $J = 9.0$ Hz, 1H), 7.51 (d, $J = 3.1$ Hz, 1H), 7.35 (dd, $J = 8.9, 3.0$ Hz, 1H), 3.89 (s, 3H).

^{13}C NMR (101 MHz, $\text{DMSO-}d_6$) δ /ppm: 157.19, 145.62, 139.74, 135.16, 132.07, 131.67, 131.65, 130.25, 129.28, 127.26, 119.17, 109.03, 56.12.

HRMS (ESI): $\text{C}_{14}\text{H}_{13}\text{BN}_3\text{O}_4^+$ *calcd.*: 298.0994, *found*: 298.0996.

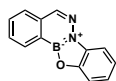
2-Hydroxyphenylhydrazine pTsOH salt **1-14a**



2-Hydroxyaniline (3.0 g, 27.5 mmol, 1.0 eq.) was suspended in EtOH (22 mL) and concentrated HCl (35%, 2.4 mL, 27.5 mmol, 1.0 eq.) was added. The mixture was cooled in an ice/salt bath. Isopentyl nitrite (4.1 mL, 30.2 mmol, 1.1 eq.) in EtOH (13 mL) was slowly added, keeping the temperature between -5°C and 0°C . The solution was stirred for 45 min, after which it was slowly transferred to a cooled solution of SnCl_2 (10.4 g, 55.0 mmol, 2.0 eq.) and pTsOH monohydrate (5.2 g, 27.5 mmol, 1.0 eq.) in EtOH (35 mL), keeping the temperature between 0°C and 10°C . The suspension was stirred in the ice bath for 30 min and was then warmed to RT over the course of 30 min. Et_2O (70 mL) was added and the suspension was stirred for another 10 min. The solid was filtered off and washed with Et_2O (4 x 25 mL). The obtained material was suspended in cyclohexane (20 mL) and heated to reflux. EtOH (50 mL) was added portionwise to obtain a clear solution. The solution was slowly cooled down to RT and was then left in the fridge (5°C) overnight. The precipitate was filtered off, washed with cyclohexane (3 x 15 mL) and dried *in vacuo* to yield the desired product as white plates (4.4 g, **54%**). Analytical data was in agreement with the reported data.^[99]

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 9.97 (s, 1H), 9.66 (s, 3H), 7.49 – 7.45 (m, 2H), 7.11 (d, $J = 7.9$ Hz, 2H), 6.92 (dd, $J = 7.6, 1.5$ Hz, 1H), 6.88 – 6.82 (m, 2H), 6.79 (ddd, $J = 7.6, 6.5, 2.3$ Hz, 1H), 2.29 (s, 3H).

Benzo[d]benzo[4,5][1,3,2]oxazaborolo[3,2-b][1,2,3]diazaborinine **1-14**



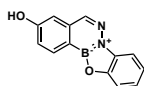
1-14 was synthesized according to the general procedure. 2-Formylphenylboronic acid (10.0 mg, 66.7 μmol , 1.0 eq.) and **1-14a** (27.7 mg, 93.4 μmol , 1.4 eq.) were mixed in the solvent (1.0 mL). The product was isolated as a white solid (14 mg, **95%**).

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 8.73 (s, 1H), 8.34 (d, $J = 7.4$ Hz, 1H), 8.18 (d, $J = 7.9$ Hz, 1H), 7.96 (td, $J = 7.3, 1.4$ Hz, 1H), 7.88 (td, $J = 7.4, 1.1$ Hz, 1H), 7.82 (dd, $J = 7.5, 1.3$ Hz, 1H), 7.62 (dd, $J = 8.0, 0.9$ Hz, 1H), 7.35 (td, $J = 7.5, 1.3$ Hz, 1H), 7.29 (td, $J = 7.7, 1.5$ Hz, 1H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ /ppm: 148.12, 144.64, 135.75, 134.35, 131.99, 130.62, 130.37, 129.49, 123.17, 122.89, 113.69, 110.35.

HRMS (ESI): $\text{C}_{13}\text{H}_{10}\text{BN}_2\text{O}^+$ *calcd.*: 221.0881, *found*: 221.0883.

Benzo[d]benzo[4,5][1,3,2]oxazaborolo[3,2-b][1,2,3]diazaborinin-3-ol **1-15**



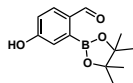
1-15 was synthesized according to the general procedure. **1-2** (15.0 mg, 60.5 μmol , 1.0 eq.) and **1-14a** (25.1 mg, 84.6 μmol , 1.4 eq.) were mixed in the solvent (1.0 mL). The product was isolated as a white solid (13 mg, **91%**).

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 10.56 (sbr, 1H), 8.55 (s, 1H), 8.16 (d, $J = 8.3$ Hz, 1H), 7.75 (dd, $J = 7.5, 1.1$ Hz, 1H), 7.55 (dd, $J = 7.8, 1.1$ Hz, 1H), 7.43 (d, $J = 2.1$ Hz, 1H), 7.35 – 7.27 (m, 2H), 7.24 (td, $J = 7.7, 1.4$ Hz, 1H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ /ppm: 160.75, 148.19, 144.37, 138.17, 134.51, 132.44, 122.92, 122.81, 120.36, 113.74, 113.59, 110.12.

HRMS (ESI): C₁₃H₁₀BN₂O₂⁺ *calcd.*: 237.0830, *found*: 237.0832.

2-Formyl-5-hydroxyphenylboronic acid pinacol ester **1-16a**



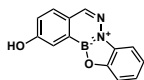
Under an Ar atmosphere, 2-bromo-4-hydroxybenzaldehyde (250.0 mg, 1.2 mmol, 1.0 eq.), bis(pinacolato)diboron (410.6 mg, 1.6 mmol, 1.3 eq.), Pd(dppf)Cl₂ (91.0 mg, 120 μmol, 10 mol%) and KOAc (366.2 mg, 3.7 mmol, 3.0 eq.) were suspended in freshly distilled dioxane (5 mL). The mixture was heated to reflux (100°C) for 16 h. After cooling, the solvent was removed by rotary evaporation. The residue was diluted with CH₂Cl₂ (20 mL) and was washed with H₂O (3 x 20 mL) and brine (20 mL). The organic layer was dried over Na₂SO₄, concentrated by rotary evaporation and purified by flash column chromatography (10:1→ 5:1→ 3:1 cyclohexane:EtOAc, R_f= 0.33, UV/KMnO₄) to yield the desired product as a slightly orange solid (141 mg, **46%**).

¹H NMR (400 MHz, CDCl₃) δ/ppm: 10.33 (s, 1H), 7.89 (d, *J* = 8.5 Hz, 1H), 7.23 (d, *J* = 2.6 Hz, 1H), 6.96 (dd, *J* = 8.5, 2.6 Hz, 1H), 1.39 (s, 12H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 193.13, 160.16, 134.78, 132.52, 131.34, 121.89, 117.64, 84.71, 25.00.

HRMS (ESI): C₁₃H₁₇BNaO₄⁺ *calcd.*: 271.1112, *found*: 271.1114.

Benzo[d]benzo[4,5][1,3,2]oxazaborolo[3,2-b][1,2,3]diazaborinin-3-ol **1-16**



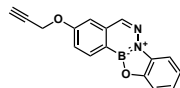
1-16 was synthesized according to the general procedure. **1-16a** (15.0 mg, 60.5 μmol, 1.0 eq.) and **1-14a** (25.1 mg, 84.6 μmol, 1.4 eq.) were mixed in the solvent (1.0 mL). The product was isolated as a white solid (13 mg, **91%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 10.63 (sbr, 1H), 8.54 (s, 1H), 8.02 (d, *J* = 8.3 Hz, 1H), 7.77 (dd, *J* = 7.5, 1.2 Hz, 1H), 7.58 (dd, *J* = 7.8, 1.0 Hz, 1H), 7.55 (d, *J* = 2.6 Hz, 1H), 7.38 – 7.29 (m, 2H), 7.26 (td, *J* = 7.8, 1.5 Hz, 1H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 159.48, 148.06, 144.41, 134.58, 132.09, 128.91, 122.97, 122.90, 121.27, 114.07, 113.69, 110.35.

HRMS (ESI): C₁₃H₁₀BN₂O₂⁺ *calcd.*: 237.0830, *found*: 237.0832.

3-(Propargyloxy)benzo[d]benzo[4,5][1,3,2]oxazaborolo[3,2-b][1,2,3]diazaborinine **1-17**



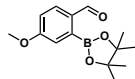
1-17 was synthesized according to the general procedure. **1-7b** (15.0 mg, 52.4 μmol, 1.0 eq.) and **1-14a** (21.7 mg, 73.4 μmol, 1.4 eq.) were mixed in the solvent (1.0 mL). The product was isolated as a white solid (12 mg, **84%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 8.65 (s, 1H), 8.28 (d, *J* = 8.5 Hz, 1H), 7.79 (dd, *J* = 7.5, 1.3 Hz, 1H), 7.75 (d, *J* = 2.4 Hz, 1H), 7.58 (d, *J* = 8.0 Hz, 1H), 7.51 (dd, *J* = 8.5, 2.5 Hz, 1H), 7.33 (td, *J* = 7.5, 1.2 Hz, 1H), 7.27 (td, *J* = 7.7, 1.4 Hz, 1H), 5.02 (d, *J* = 2.4 Hz, 2H), 3.66 (t, *J* = 2.4 Hz, 1H).

^{13}C NMR (101 MHz, DMSO- d_6) δ /ppm: 159.93, 148.20, 144.24, 137.78, 134.39, 132.33, 123.15, 122.90, 120.13, 113.69, 112.76, 110.28, 78.92, 78.77, 55.79.

HRMS (ESI): $\text{C}_{16}\text{H}_{12}\text{BN}_2\text{O}_2^+$ *calcd.*: 275.0986, *found*: 275.0989.

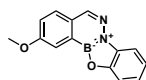
2-Formyl-5-methoxyphenylboronic acid pinacol ester **1-18a**



Under a N_2 atmosphere **1-16a** (70.0 mg, 282.0 μmol , 1.0 eq.) was dissolved in DMF (1 mL) and K_2CO_3 (58.5 mg, 423.0 μmol , 1.5 eq.) was added. MeI (21.1 μL , 339 μmol , 1.2 eq.) was added dropwise and the mixture was stirred at room temperature overnight. The mixture was diluted with H_2O (15 mL) and extracted with EtOAc (3 x 15 mL). The combined organic layers were washed with H_2O (2 x 15 mL) and brine (15 mL). The organic layers were concentrated by rotary evaporation and the residue was purified by flash column chromatography (cyclohexane \rightarrow 20:1 \rightarrow 10:1 \rightarrow 5:1 cyclohexane:EtOAc, $R_f = 0.26$, UV/ KMnO_4) to yield the desired product as slightly yellow oil (18 mg, **24%**). Analytical data was in agreement with reported data.^[100]

^1H NMR (400 MHz, CDCl_3) δ /ppm: 10.37 (s, 1H), 7.94 (d, $J = 8.6$ Hz, 1H), 7.29 (d, $J = 2.6$ Hz, 1H), 7.03 (ddd, $J = 8.6, 2.7, 0.6$ Hz, 1H), 3.90 (s, 3H), 1.40 (s, 12H).

2-Methoxybenzo[d]benzo[4,5][1,3,2]oxazaborolo[3,2-b][1,2,3]diazaborinine **1-18**



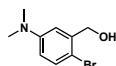
1-18 was synthesized according to the general procedure. **1-18a** (16.0 mg, 61.0 μmol , 1.0 eq.) and **1-14a** (25.3 mg, 85.5 μmol , 1.4 eq.) were mixed in the solvent (1.0 mL). The product was isolated as a white solid (10 mg, **66%**).

^1H NMR (400 MHz, DMSO- d_6) δ /ppm: 8.62 (s, 1H), 8.11 (d, $J = 8.8$ Hz, 1H), 7.80 (dd, $J = 7.7, 1.3$ Hz, 1H), 7.74 (d, $J = 2.7$ Hz, 1H), 7.60 (dd, $J = 7.8, 1.0$ Hz, 1H), 7.52 (dd, $J = 8.7, 2.7$, 1H), 7.34 (td, $J = 7.6, 1.2$ Hz, 1H) 7.28 (td, $J = 7.7, 1.6$ Hz, 1H), 3.98 (s, 3H).

^{13}C NMR (101 MHz, DMSO- d_6) δ /ppm: 160.57, 148.02, 144.12, 134.44, 131.65, 129.96, 123.03, 122.86, 120.96, 113.60, 111.48, 110.37, 55.68.

HRMS (ESI): $\text{C}_{14}\text{H}_{12}\text{BN}_2\text{O}_2^+$ *calcd.*: 251.0986, *found*: 251.0989.

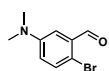
2-Bromo-5-(dimethylamino)phenylmethanol **1-19a**



In a Schlenk tube under an inert atmosphere, 3-(dimethylamino)benzyl alcohol (142.0 μL , 992.0 μmol , 1.05 eq.) was dissolved in MeCN (4 mL). NBS (168.0 mg, 942.0 μmol , 1.0 eq.) was added portionwise over 10 min, keeping the temperature at RT with a water bath. After 17 h, NMR analysis showed no more conversion to the product. The solvent was removed by rotary evaporation and the residue was purified by flash column chromatography (Silica, 40 g, 5:1 \rightarrow 3:1 \rightarrow 2:1 cyclohexane:EtOAc, $R_f = 0.43$, UV/ KMnO_4). Product-containing fractions were combined and concentrated *in vacuo* to yield the desired product as a white solid (162 mg, **75%**). Analytical data was in agreement with the reported data.^[101]

^1H NMR (400 MHz, CDCl_3) δ /ppm: 7.35 (d, $J = 8.8$ Hz, 1H), 6.87 (s, 1H), 6.66 – 6.49 (m, 1H), 4.69 (s, 2H), 2.96 (s, 6H), 2.00 (s, 1H).

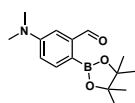
2-Bromo-5-dimethylaminobenzaldehyde **1-19b**



Under an argon atmosphere, oxalyl chloride (45.8 μL , 535.0 μmol , 1.2 eq.) was dissolved in DCM (2.2 mL) and cooled to -70°C in dry ice/acetone bath. A solution of DMSO (80.3 μL , 1.1 mmol, 2.6 eq.) in DCM (1.1 mL) was added dropwise over 5 min, while keeping the temperature below -63°C . The mixture was stirred for 15 min. A solution of **1-19a** (100.0 mg, 435.0 μmol , 1.0 eq.) in DCM (1.1 mL) was added over 5 min and the mixture was stirred for 15 min. Triethylamine (242.0 μL , 1.7 mmol, 4.0 eq.) was added dropwise and the yellow solution was stirred for 15 min. The cooling bath was removed and the solution was stirred at RT for 30 min, after which TLC showed complete conversion. The solution was washed with H_2O (5 x 10 mL) and the combined aqueous layers were extracted with DCM (3 x 5 mL). The combined organic layers were washed with brine (15 mL), dried over Na_2SO_4 and were concentrated by rotary evaporation. The residue was purified by flash column chromatography (Silica, 13 g, cyclohexane \rightarrow 20:1 cyclohexane:EtOAc, $R_f = 0.3$, UV/ KMnO_4). Product-containing fractions were combined and concentrated *in vacuo* to yield the desired product as a yellow solid (76 mg, **77%**). Analytical data was in agreement with reported data.^[101]

^1H NMR (400 MHz, CDCl_3) δ /ppm: 10.32 (s, 1H), 7.50 (d, $J = 8.8$ Hz, 1H), 7.32 (d, $J = 3.3$ Hz, 1H), 7.03 (d, $J = 8.4$ Hz, 1H), 3.02 (s, 6H).

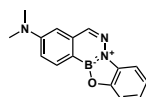
2-Formyl-4-(dimethylamino)phenylboronic acid pinacol ester **1-19c**



Under an argon atmosphere, **1-19b** (50.0 mg, 219.0 μmol , 1.0 eq.), bis(pinacolato)diboron (72.4 mg, 285.0 μmol , 1.3 eq.), Pd(dppf) Cl_2 (16.0 mg, 21.9 μmol , 10 mol%) and KOAc (64.5 mg, 658.0 μmol , 3.0 eq.) were suspended in freshly distilled dioxane (1.5 mL). The mixture was heated to reflux for 3 h. After cooling, the solvent was removed by rotary evaporation and the residue was diluted with CH_2Cl_2 (20 mL). The solution was washed with H_2O (3 x 20 mL), followed by brine (20 mL), was dried over Na_2SO_4 and concentrated by rotary evaporation. The residue was purified by flash column chromatography (Silica, 13 g, cyclohexane \rightarrow 20:1 \rightarrow 15:1 \rightarrow 10:1 cyclohexane:EtOAc, $R_f = 0.27$, UV/ KMnO_4). Product-containing fractions were combined and concentrated *in vacuo* to yield the desired product as a yellow solid (43 mg, **71%**). Analytical data was in agreement with reported data.^[102]

^1H NMR (400 MHz, CDCl_3) δ /ppm: 10.70 (s, 1H), 7.88 (d, $J = 8.1$ Hz, 1H), 7.49 (s, 1H), 7.21 (s, 1H), 3.08 (s, 6H), 1.36 (s, 12H).

N,N-dimethylbenzo[*d*]benzo[4,5][1,3,2]oxazaborolo[3,2-*b*][1,2,3]diazaborinin-3-amine **1-19**



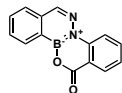
1-19 was synthesized according to the general procedure. **1-19c** (15.0 mg, 54.5 μmol , 1.0 eq.) and **1-14a** (22.6 mg, 76.3 μmol , 1.4 eq.) were mixed in the solvent (1.0 mL). The product was isolated as a white solid (11 mg, **77%**).

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 8.52 (s, 1H), 8.11 – 8.06 (m, 1H), 7.72 (dd, $J = 7.7, 1.2$ Hz, 1H), 7.53 (dd, $J = 8.0, 1.0$ Hz, 1H), 7.31 – 7.25 (m, 3H), 7.21 (td, $J = 7.6, 1.5$, 1H), 3.09 (s, 6H).

^{13}C NMR (101 MHz, DMSO- d_6) δ /ppm: 152.54, 148.19, 145.00, 137.87, 134.57, 131.32, 122.57, 122.49, 116.18, 113.35, 109.77, 109.70, 39.50.

HRMS (ESI): $\text{C}_{15}\text{H}_{15}\text{BN}_3\text{O}^+$ *calcd.*: 264.1303, *found*: 264.1305.

14H-benzo[d]benzo[4,5][1,2,3]diazaborinino[3,2-b][1,3,2]oxazaborinin-14-one **1-20**



1-20 was synthesized according to the general procedure. 2-FPBA (11.0 mg, 73.4 μmol , 1.0 eq.) and 2-hydrazinobenzoic acid (19.4 mg, 103.0 μmol , 1.4 eq.) were mixed in the solvent (1.0 mL). The product was isolated as a white solid (18 mg, **99%**).

^1H NMR (400 MHz, DMSO- d_6) δ /ppm: 8.42 (s, 1H), 8.20 – 8.15 (m, 1H), 8.13 – 8.05 (m, 2H), 7.90 – 7.85 (m, 1H), 7.83 – 7.76 (m, 2H), 7.74 (td, $J = 7.4, 1.1$, 1H), 7.30 – 7.24 (m, 1H).

^{13}C NMR (101 MHz, DMSO- d_6) δ /ppm: 160.75, 145.82, 142.16, 135.38, 133.87, 131.50, 130.31, 130.25, 130.15, 127.62, 122.77, 115.14, 114.42.

HRMS (ESI): $\text{C}_{14}\text{H}_{10}\text{BN}_2\text{O}_2^+$ *calcd.*: 249.0830, *found*: 249.0832.

2-Bromocyclohex-1-ene-1-carbaldehyde **1-21a**



In a dry 100 mL flask under an inert atmosphere DMF (3.6 mL, 45.9 mmol, 3.0 eq.) was dissolved in DCM (50 mL) and cooled in an ice bath to 0°C. PBr_3 (3.6 mL, 38.2 mmol, 2.5 eq.) was slowly added and the mixture was stirred while cooling for 1 h. A solution of cyclohexanone (1.6 mL, 15.3 mmol, 1.0 eq.) in DCM (5 mL) was slowly added to the suspension while cooling. The mixture was slowly warmed to RT and stirred for 22 h. The reaction was quenched with H_2O (50 mL) while cooling at 0°C. The mixture was neutralized by addition of solid NaHCO_3 in small portions. The biphasic mixture was separated and the aqueous layer was extracted with DCM (3 x 50 mL). The combined organic layers were washed with brine (50 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The residue was purified by flash column chromatography (Silica, 250 g, 2% \rightarrow 3% \rightarrow 5% EtOAc in cyclohexane, $R_f = 0.21$ in 2% EtOAc/cyclohexane, UV, KMnO_4 or DNP stain) to yield the desired product as yellow liquid (1.4 g, **49%**).

^1H NMR (400 MHz, CDCl_3) δ /ppm: 10.02 (s, 1H), 2.75 (tt, $J = 6.3, 2.3$ Hz, 2H), 2.28 (tt, $J = 5.8, 2.3$ Hz, 2H), 1.80 – 1.73 (m, 2H), 1.72 – 1.65 (m, 2H).

^{13}C NMR (101 MHz, CDCl_3) δ /ppm: 193.90, 143.75, 135.42, 38.98, 25.14, 24.41, 21.24.

HRMS (ESI): $\text{C}_7\text{H}_{10}\text{BrO}^+$ *calcd.*: 188.9910, *found*: 188.9913.

2-(4,4,5,5-Tetramethyl-1,3,2-dioxaborolan-2-yl)cyclohex-1-ene-1-carbaldehyde **1-21b**



In a dry 50 mL flask under an Ar atmosphere, **1-21a** (256.0 mg, 1.4 mmol, 1.0 eq.) was dissolved in dioxane (7 mL), followed by *bis*(pinacolato)diboron (378.0 mg, 1.5 mmol, 1.1 eq.) and potassium acetate (266.0 mg, 2.7 mmol, 2.0 eq.). The mixture was degassed by bubbling argon through the mixture for 10 min. $\text{Pd}(\text{dppf})\text{Cl}_2$ (49.5 mg, 67.7 μmol , 5 mol%) was added and the mixture was heated to 80°C for 2 h, after which UPLC-MS analysis showed full conversion to the desired product. The solvent was removed by rotary evaporation and the crude material was

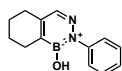
dissolved in EtOAc (40 mL) and was washed with H₂O (2 x 20 mL). The combined aqueous layers were extracted with EtOAc (2 x 20 mL) and the combined organic layers were washed with brine (20 mL), dried over Na₂SO₄ and were concentrated by rotary evaporation. The crude was purified by flash column chromatography (Silica, 33 g, cyclohexane:EtOAc 10:1, R_f = 0.24, UV/KMnO₄/DNP). Product-containing fractions were combined and concentrated *in vacuo* to yield the desired product as yellow oil (258 mg, **81%**).

¹H NMR (400 MHz, CDCl₃) δ/ppm: 9.96 (s, 1H), 2.42 – 2.37 (m, 2H), 2.28 – 2.22 (m, 2H), 1.62 (p, *J* = 3.2 Hz, 4H), 1.31 (s, 12H).

¹³C NMR (126 MHz, CDCl₃) δ/ppm: 194.71, 149.31, 84.27, 29.72, 24.78, 22.60, 21.96, 21.23.

HRMS (ESI): C₁₃H₂₂BO₃⁺ *calcd.*: 237.1657, *found*: 237.1653.

2-Phenyl-5,6,7,8-tetrahydrobenzo[d][1,2,3]diazaborinin-1(2H)-ol **1-21**



1-21b (32.0 mg, 135.5 μmol, 1.0 eq.) was dissolved in DMSO (1.4 mL) and H₂O (1.6 mL). Phenylhydrazine (18.7 μL, 190.0 μmol, 1.4 eq.) was added and the suspension was stirred at RT for 10 min. For better solubility, MeCN (1 mL) was added and the mixture was stirred at RT for another 2 h. UPLC-MS analysis showed 50% BIQ product and 50% Pin protected hydrazone intermediate. TFA (3 drops) was added, followed by MeCN (500 μL). The solution was purified by preparative RP-HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (16 mg, **52%**).

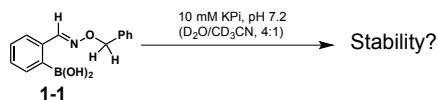
¹H NMR (500 MHz, DMSO-*d*₆) δ/ppm: 7.83 (s, 1H), 7.49 – 7.43 (m, 2H), 7.40 (s, 1H), 7.39 – 7.33 (m, 2H), 7.23 – 7.17 (m, 1H), 2.48 – 2.41 (m, 4H), 1.74 – 1.67 (m, 2H), 1.67 – 1.61 (m, 2H).

¹³C NMR (126 MHz, DMSO-*d*₆) δ/ppm: 146.18, 141.95, 139.52, 128.06, 125.02, 124.37, 26.85, 25.26, 21.70.

HRMS (ESI): C₁₃H₁₆BN₂O⁺ *calcd.*: 227.1350, *found*: 227.1350.

5.3 Stability and Reversibility of Boron-Assisted Oxime Formations

NMR Oxime Stability Assay



Stock solutions:

- Phosphate buffer: 12.8 mM potassium phosphate (K_3PO_4) in D_2O , pH 7.2, adj. with DCl
- Oxime: 500.0 μM **1-1** in CD_3CN
- Internal Standard: 275.0 μM $\text{TMSP-}d_4$ in D_2O

Procedure:

Phosphate Buffer (430 μL) was mixed inside the NMR tube with the oxime **1-1** (110 μL) and the internal standard (10 μL). The final volume was 550 μL with the following final concentrations: 100 μM oxime **1-1**, 10 mM phosphate buffer and 5 μM $\text{TMSP-}d_4$ with a ratio of 4:1 phosphate buffer: CD_3CN . The sample was locked to the CD_3CN residual peak and shimmed before recording the first ^1H NMR spectrum.

The stability of oxime **1-1** was determined by recording ^1H NMR spectra every 7 h for the first day and then one spectrum per day for another three days. The integration of the oxime proton (black) was compared to the internal standard integral after 1, 24, 48, 72 and 96 h (**Figure 23**).

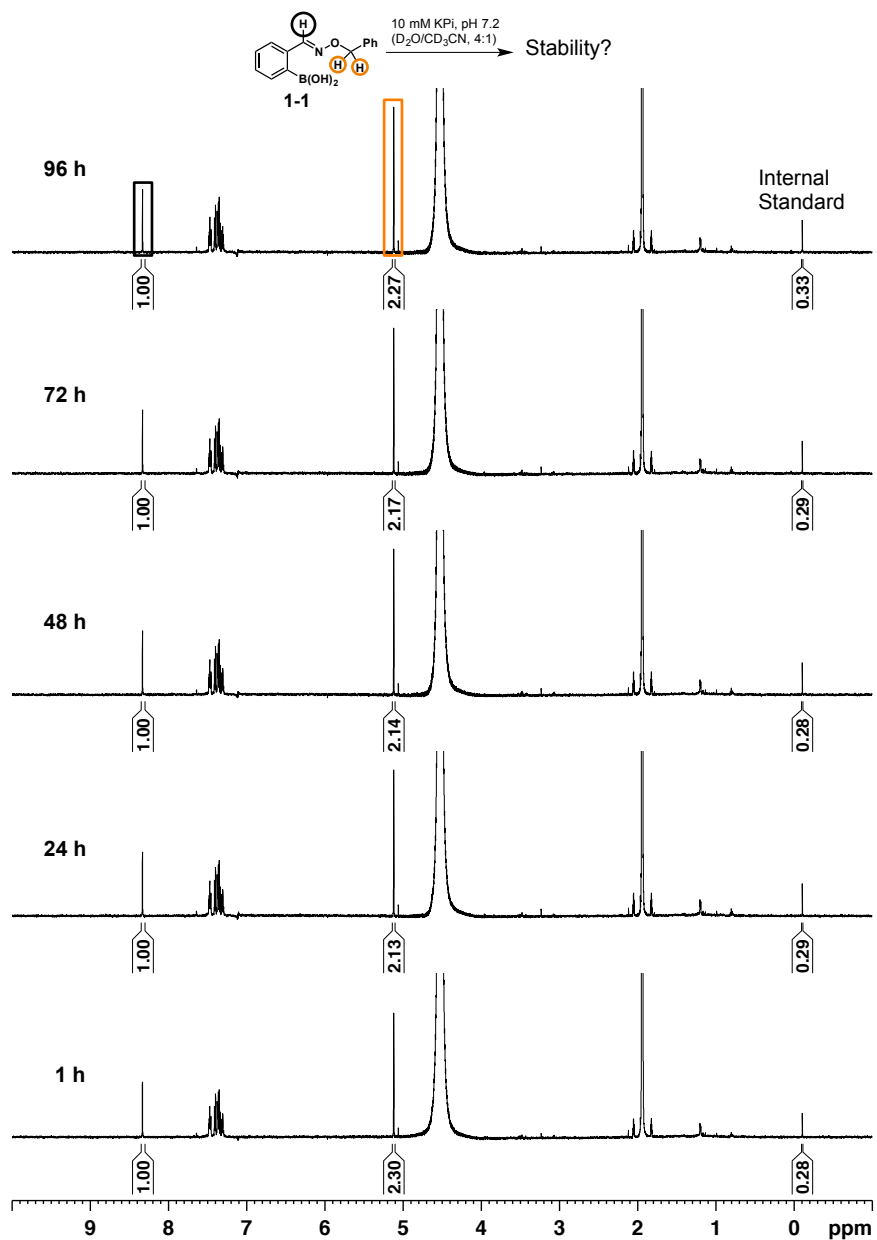
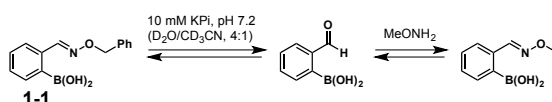


Figure 23. NMR stability assay over four days. TMSP-*d*₄ was used as internal standard. In the first three days a concentration difference of <5% was observed, which was attributed to errors in integration. After 96 h a concentration difference of -15% oxime was found.

First NMR Reversibility Assay Aldoxime



Stock solutions:

- Phosphate buffer: 13.1 mM potassium phosphate (K_3PO_4) in D_2O , pH 7.2, adj. with DCI
- Oxime: 500.0 μM **1-1** in CD_3CN
- Hydroxylamine: 27.5 mM methoxylamine hydrochloride in D_2O
- Internal Standard: 275.0 μM $TMSP-d_4$ in D_2O

Procedure:

Phosphate Buffer (420 μL) was mixed inside the NMR tube with the oxime **1-1** (110 μL), the hydroxylamine (10 μL) and the internal standard (10 μL). The final volume was 550 μL with the following final concentrations: 100 μM oxime **1-1**, 500 μM hydroxylamine, 10 mM phosphate buffer and 5 μM $TMSP-d_4$ with a ratio of 4:1 phosphate buffer: CD_3CN . The sample was locked to the CD_3CN residual peak and shimmed before recording the first 1H NMR spectrum.

The reversibility of oxime **1-1** was determined by recording 1H NMR spectra every 2 h for a total of 60 h. The oxime protons are indicated with an arrow (**Figure 24**). A zoom-in of the NMR spectra is found in **Figure 13a**.

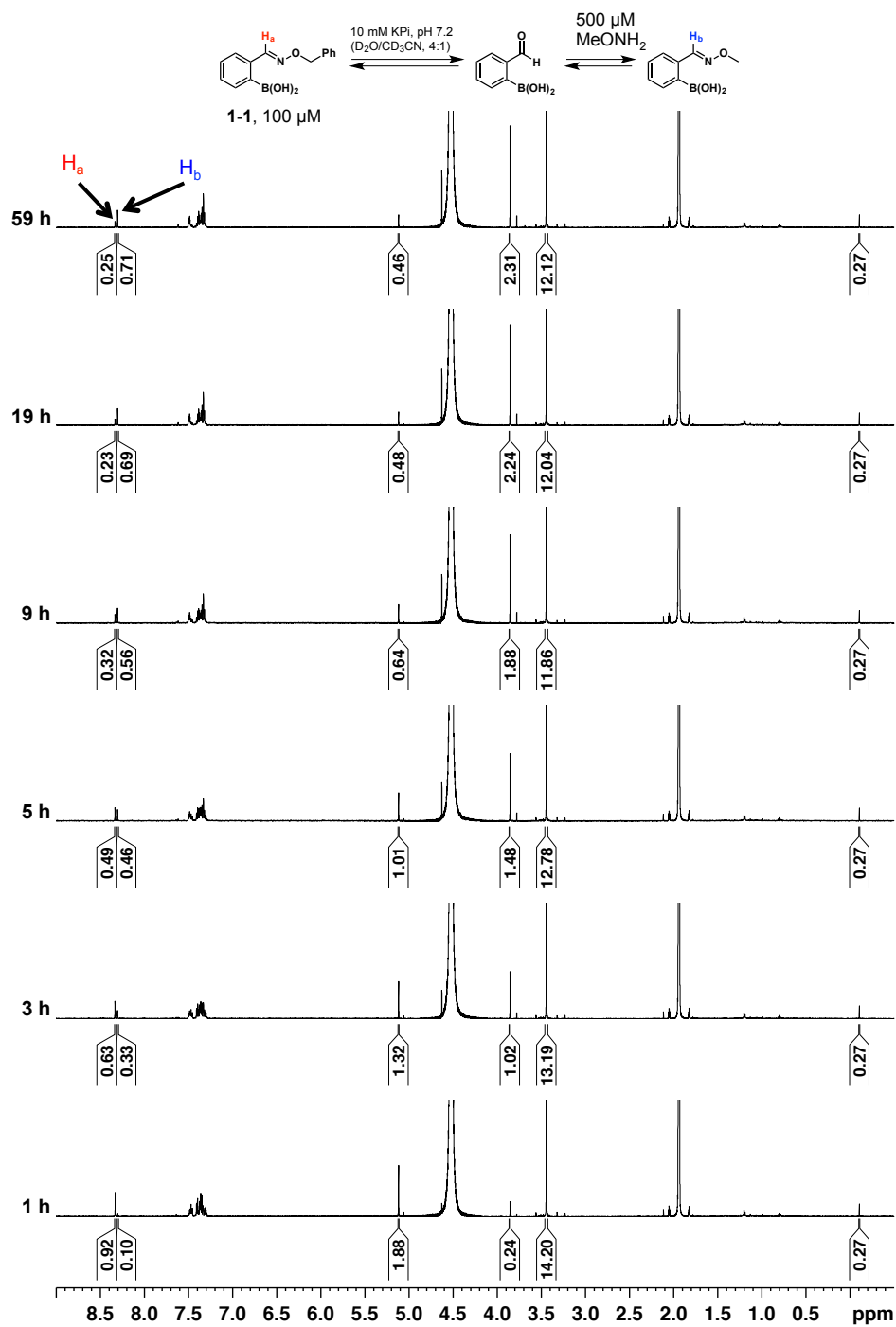
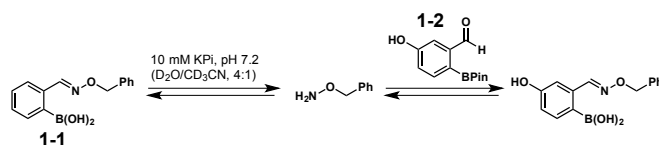


Figure 24. NMR assay for the reversibility of aldoxime **1-1** with an excess of methylhydroxylamine. NMR spectra were recorded over the course of 60 h. Complete formation of a new oxime species was observed within 10-19 h.

Second NMR Reversibility Assay Aldoxime



Stock solutions:

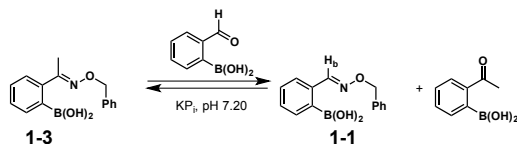
- Phosphate buffer: 13.1 mM potassium phosphate (K_3PO_4) in D_2O , pH 7.2, adj. with DCl
- Oxime: 550.0 μM **1-1** in CD_3CN
- Aldehyde boronic acid: 27.5 mM **1-2** in CD_3CN
- Internal Standard: 275.0 μM TMSP- d_4 in D_2O

Procedure:

Phosphate Buffer (430 μL) was mixed inside the NMR tube with the oxime **1-1** (100 μL), the aldehyde boronic acid (10 μL) and the internal standard (10 μL). The final volume was 550 μL with the following final concentrations: 100 μM oxime **1-1**, 500 μM aldehyde boronic acid, 10 mM phosphate buffer and 5 μM TMSP- d_4 with a ratio of 4:1 phosphate buffer: CD_3CN . The sample was locked to the CD_3CN residual peak and shimmed before recording the first 1H NMR spectrum.

The reversibility of oxime **1-1** was determined by recording 1H NMR spectra every 2 h for a total of 28 h. The stacked NMR spectra are found in **Figure 13b**.

NMR Reversibility Assay of Ketoximes



Stock solutions:

- Phosphate buffer: 13.1 mM potassium phosphate (K_3PO_4) in D_2O , pH 7.3
- Ketoxime: 550.0 μM **1-3** in CD_3CN
- Aldehyde boronic acid: 27.5 mM 2-FPBA in CD_3CN
- Internal Standard: 275.0 μM TMSP- d_4 in D_2O

Procedure:

Phosphate Buffer (430 μL) was mixed inside the NMR tube with the ketoxime **1-3** (100 μL), the aldehyde boronic acid (10 μL) and the internal standard (10 μL). The final volume was 550 μL with the following final concentrations: 100 μM ketoxime **1-3**, 500 μM aldehyde boronic acid, 10 mM phosphate buffer and 5 μM TMSP- d_4 with a ratio of 4:1 phosphate buffer: CD_3CN . The sample was locked to the CD_3CN residual peak and shimmed before recording the first 1H NMR spectrum.

Reversibility measurements were conducted every 30 min for a total of 26 h. The integrals were compared to the signal of TMSP- d_4 . After 75 min the equilibrium was almost reached. Disappearance of the methyl signals was due to H-D exchange (**Figure 25**).

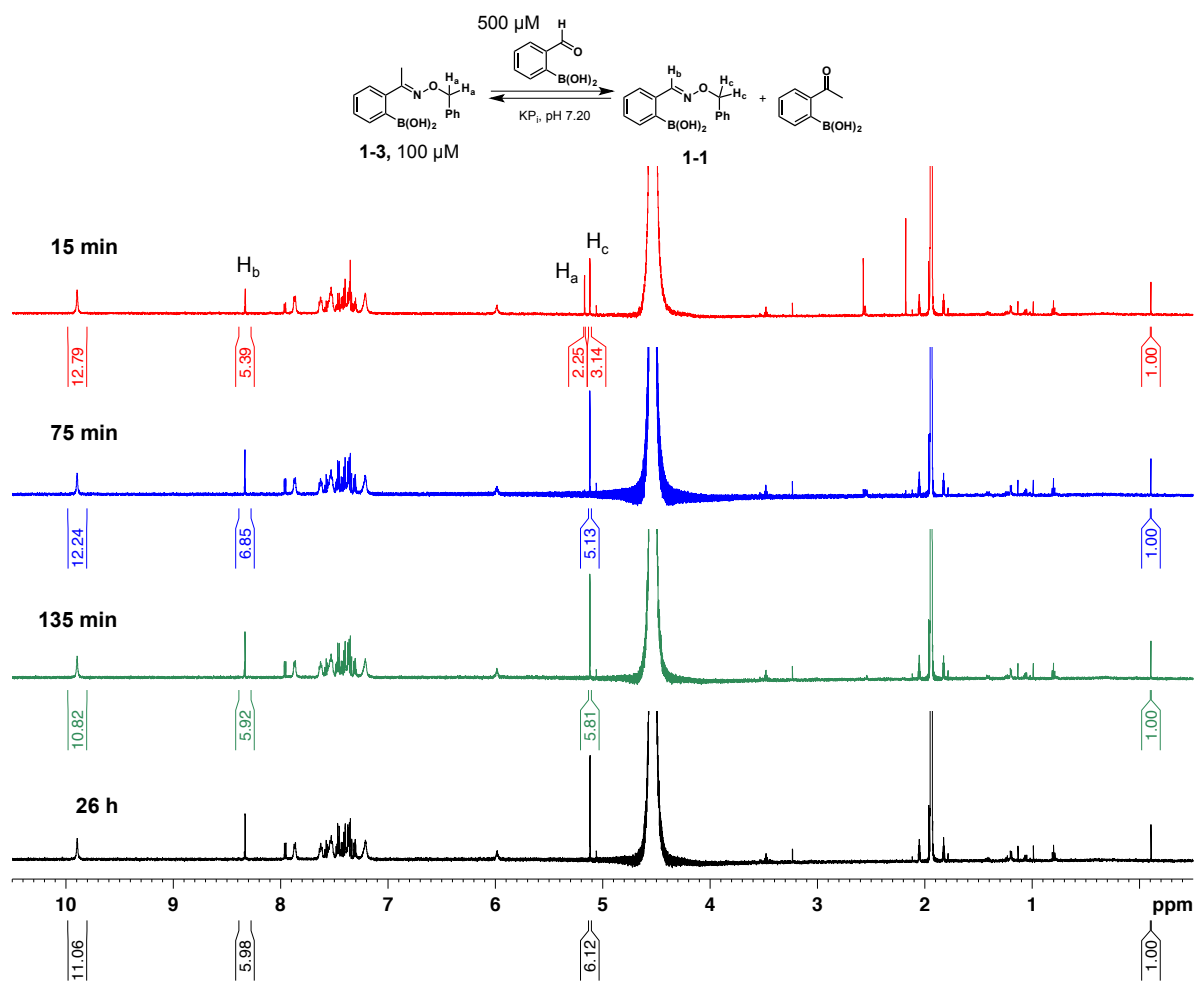
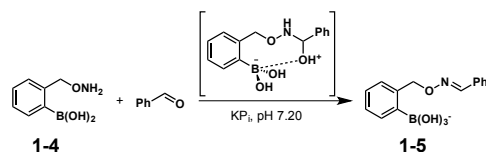


Figure 25. NMR assay for the determination of the ketoimine reversibility with an excess of 2-FPBA. Four traces (out of 52) are shown. The equilibrium was reached after 75 min.

5.4 HPLC Assays on Oxime Formation Mechanism

Oxime Formation with **1-4**



Stock solutions:

Benzaldehyde: 50 mM in HPLC grade MeCN

1-4: 5 mM in HPLC grade MeCN

Phosphate buffer (KPi): 100 mM in *dd* H₂O, pH 7.2

Procedure:

Phosphate buffer (156 μ L), benzaldehyde (4 μ L) and **1-4** (40 μ L) were mixed in a PP HPLC vial to obtain a final concentration of 1 mM.

Sample and references were analyzed by RP-HPLC on an Agilent 1100 LC system equipped with a YMC-Gel ODS-A 10 μ m 4.6 x 150 mm column from Dr. Maisch GmbH at 25°C. Buffer A (A): 0.1% TFA (v/v) in H₂O/1% MeCN (v/v), Buffer B (B): 0.1% TFA (v/v) in MeCN/1% H₂O (v/v). Gradient: 1% (2 min) - 99% (24 min) - 99% (2 min) (B). Samples were measured after 1 min, 30 min, 60 min, 90 min and 15 h with injection volumes of 20 μ L and UV absorption detection at 254 nm. Conversions were calculated according to the peak area change of the benzaldehyde signal. New peaks at 17.522 min and 19.295 min of the 15 h run were collected and analyzed by UPLC-MS (**Table 3** and **Figure 26**).

Table 3. Conversion of **1-4** and benzaldehyde to the oxime product **1-5**. Percentages were calculated by the peak area change of the benzaldehyde signal.

Time	Conversion to 1-5
1 min	< 1%
30 min	<1%
60 min	1%
90 min	2%
15 h	44%

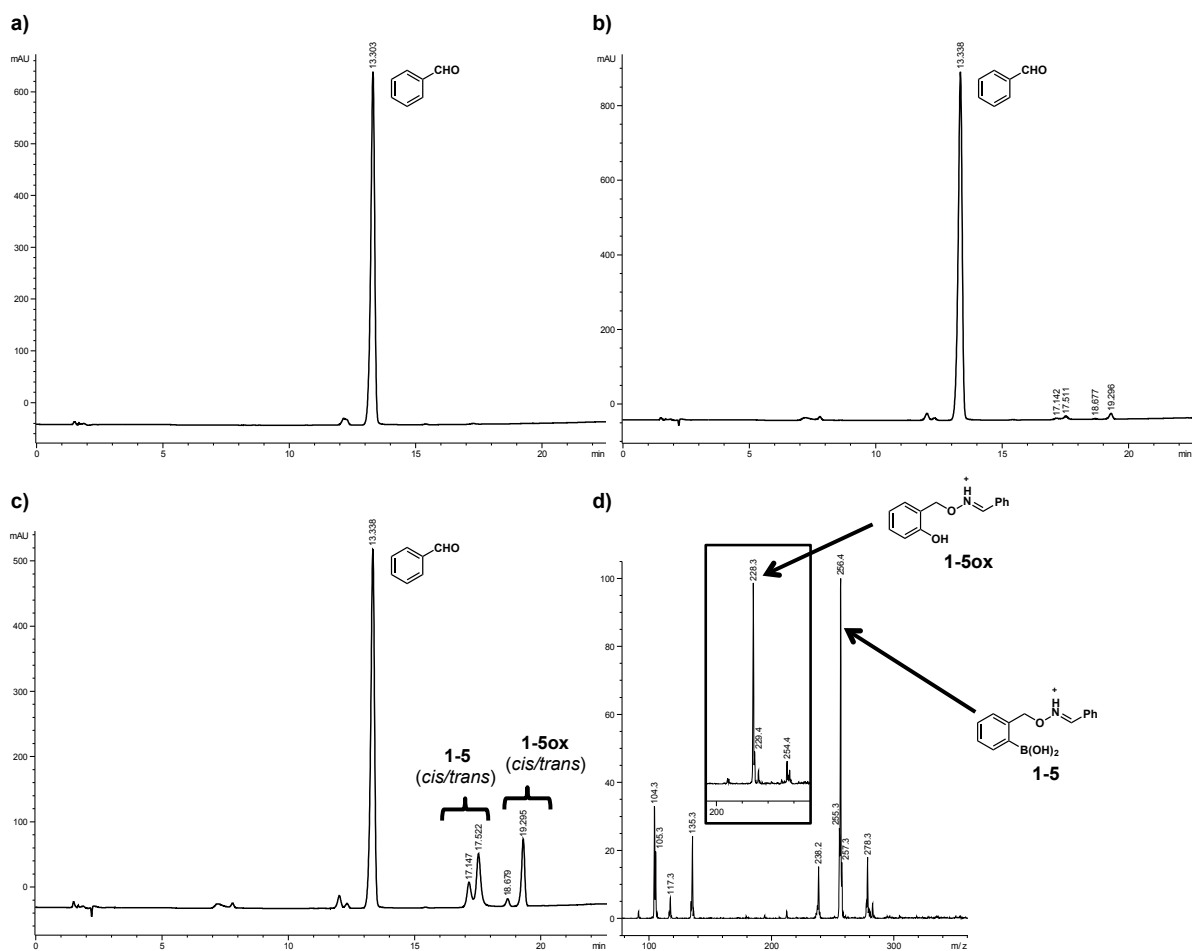
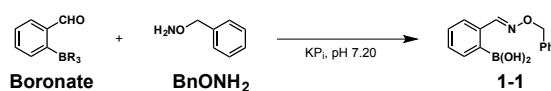


Figure 26. a) HPLC trace of 1 mM benzaldehyde. b) HPLC trace after 60 min reaction time. c) HPLC trace after 15 h reaction time. The peaks at 17.522 and 19.295 min were collected and analyzed by UPLC-MS. d) ESI-MS spectra of the collected peaks. The desired product was found in the 17.522 min fraction, the oxidized product (MS trace inset) was found in the 19.295 min fraction.

We proved the formation of the desired product **1-5** by LC-MS analysis of the newly formed peaks. Oxidation of the boronic acid (**1-5ox**) was also observed in a substantial amount.

Oxime Condensation with Triply Coordinated Boronates



Stock solutions:

Boronate 1: 5 mM potassium 2-formylphenyltrifluoroborate (2-FPTFB) in HPLC grade MeCN

Boronate 2: 5 mM 2-formylphenylboronic acid MIDA ester in HPLC grade MeCN

Benzylhydroxylamine (BnONH₂): 5 mM in HPLC grade MeCN

Phosphate buffer: 100 mM in *dd* H₂O, pH 7.2

Procedure:

Phosphate buffer (192 μ L), boronate 1 or 2 (4 μ L) and BnONH₂ (4 μ L) were mixed in a PP HPLC vial to obtain a final concentration of 100 μ M.

Samples and references were analyzed by RP-HPLC on an Agilent 1100 LC system equipped with a Zorbax Eclipse XDB-C8 5 μ m 4.6 x 150 mm column from Agilent at 25°C. Buffer A (A): H₂O/1% MeCN (v/v), Buffer B (B): MeCN/1% H₂O (v/v). Gradient: 1% (2 min) - 99% (24 min) - 99% (2 min) (B). The samples were measured after 1 min, 30 min, 60 min and 90 min with injection volumes of 20 μ L and UV absorption detection at 254 nm. For 2-FPTFB, conversions were calculated according to the area increase of the oxime **1-1** product signal compared to a 100 μ M product reference sample. For the MIDA boronate the area increase of the oxime **1-1** product signal or the area decrease of the starting material could be used for conversion calculations (**Table 4** and **Figure 27**).

Table 4. Conversion of the triply coordinated boronates to the oxime product **1-1**. Percentages were calculated by the peak area changes of the product or the starting material.

Time	2-FPTFB	MIDA boronate
1 min	5%	12%
30 min	51%	26%
60 min	76%	46%
90 min	93%	60%

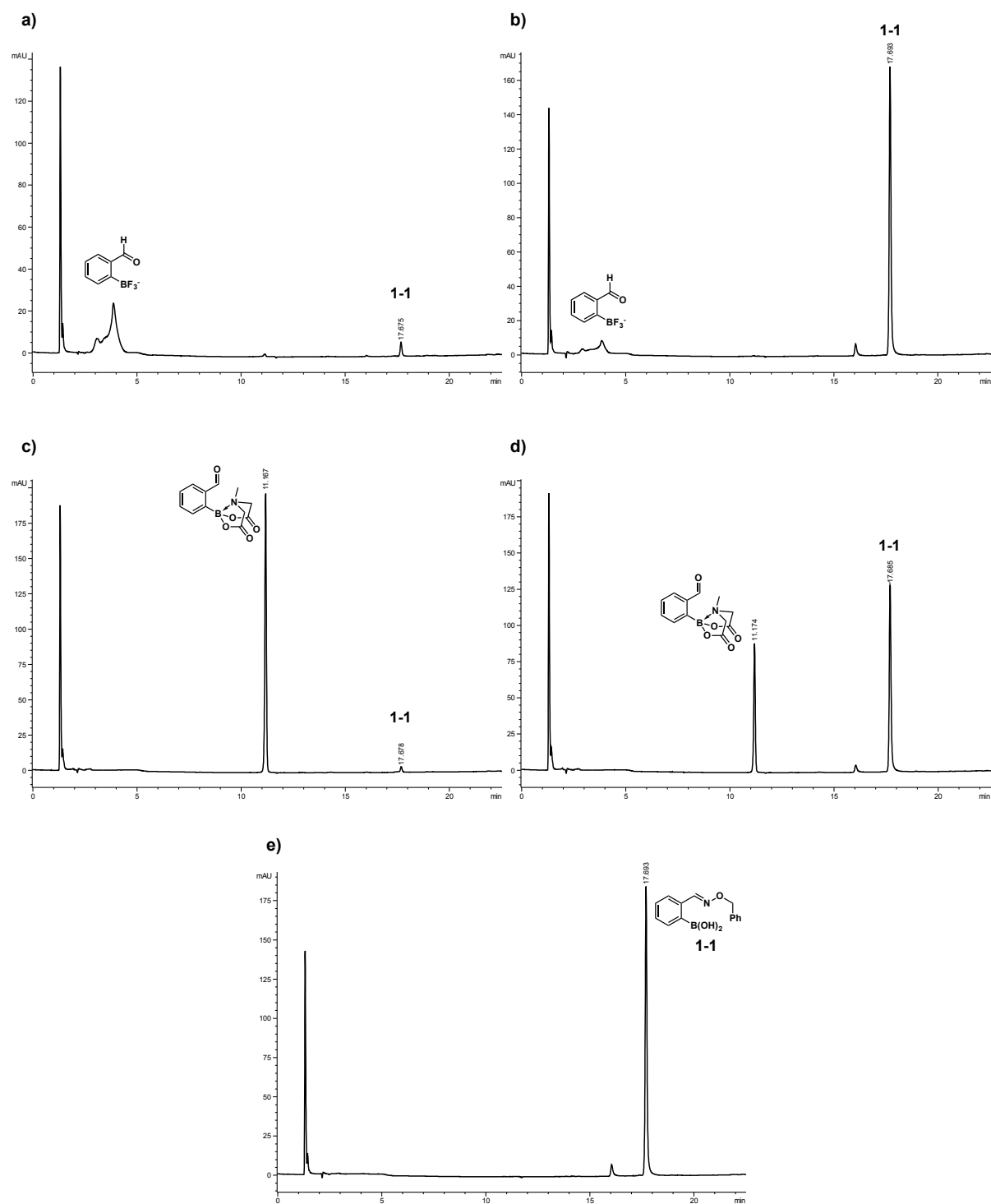
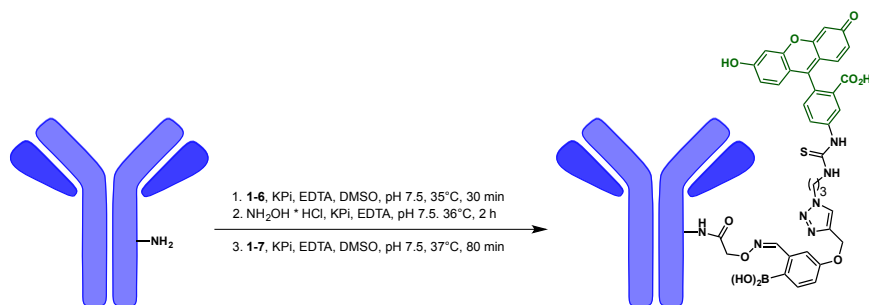


Figure 27. HPLC chromatograms of the oxime conversion assays with triply coordinated boron species. **a)** Oxime condensation with 2-FPTFB after 1 min reaction time. **b)** Oxime condensation with 2-FPTFB after 90 min reaction time. **c)** Oxime condensation with MIDA boronate after 1 min reaction time. **d)** Oxime condensation with MIDA boronate after 90 min reaction time. **e)** Oxime 1-1 product reference chromatogram.

5.5 Fluorescent Labeling of Immunoglobulin G



Stock solutions:

Phosphate Buffer: 50 mM Na_2HPO_4 , 1 mM EDTA in dd H_2O , adjusted to pH 7.5

Hydroxylamine: 500 mM NH_2OH hydrochloride in 50 mM NaH_2PO_4 /25 mM EDTA, pH 7.5 buffer

Antibody IgG: 0.96 mg/ml IgG in phosphate buffer. Concentration measured by the Nanodrop at 280 nm ($\text{MW} = 146'000$ g/mol, $\epsilon = 210'000$ $\text{M}^{-1}\text{cm}^{-1}$)

Active ester **1-6**: 65 mM **1-6** in DMSO

Fluorophore **1-7**: 10 mM **1-7** in DMSO

Hydroxylamine incorporation:

A solution of **1-6** (10.0 μL , 650.0 nmol, 99.4 eq.) was added to the antibody solution (1.0 mL, 6.5 nmol, 1.0 eq.) and the mixture was agitated at 35°C for 30 min. Purification was achieved by dialysis against phosphate buffer (2 x 1 L) at 4°C for 6 h and 18 h, respectively. For the oxime deprotection, hydroxylamine hydrochloride solution (80.0 μL , 40.0 μmol , 6115.2 eq.) was added to the antibody and the mixture was agitated at 36°C for 2 h. The modified antibody was dialyzed against phosphate buffer (2 x 1 L) at 4°C for 6 h and 18 h, respectively, to yield a final concentration of 0.742 mg/ml.

Fluorescent Labeling:

An aliquot of the modified IgG solution (50.0 μL , 254.1 pmol, 1.0 eq.) was treated with the fluorophore **1-7** solution (10 μL , 100.0 nmol, 393.5 eq.). The mixture was agitated at 37°C for 80 min. The success of the labeling was controlled by SDS-PAGE analysis (200 V, 60 min) and fluorescent bands analysis in the gel imager. Final protein staining was performed with Coomassie Blue.

1. Control experiment:

An aliquot of the modified IgG solution (25.0 μL , 127.1 pmol, 1.0 eq.) was treated with 2-FPBA (2.0 μL , 25.0 mM in DMSO, 50.0 nmol, 393.4 eq.) at 37°C for 45 min. Then, fluorophore **1-7** solution (5 μL , 50.0 nmol, 393.4 eq.) was added and the reaction mixture was agitated at 37°C for 1 h. The reaction was analyzed by SDS-PAGE separation (200 V, 60 min) and fluorescent bands analysis in the gel imager. Final protein staining was performed with Coomassie Blue.

2. Control experiment:

An aliquot of the native IgG solution (25.0 μL , 163.5 pmol, 1.0 eq.) was diluted with phosphate buffer (7 μL) and was mixed with the fluorophore **1-7** solution (6.4 μL , 64.0 nmol, 391.4 eq.). The mixture was agitated at 37°C for 1h. The reaction was analyzed by SDS-PAGE separation (200 V, 60 min) and fluorescent bands analysis in the gel imager. Final protein staining was performed with Coomassie Blue.

The results of the IgG fluorescent labeling and the control experiments are shown in **Figure 16**.

5.6 NMR Assay for BIQ Formation

Stock solutions:

207.5 μM **1-2** in CD_3CN

5.5 mM phenylhydrazine in CD_3CN

275 μM TMSP- d_4 in D_2O

20.75 mM potassium phosphate buffer in D_2O , pH 7.25

The stock solutions were mixed in a NMR tube with the following final concentrations: 100 μM **1-2**, 100 μM phenylhydrazine, 5 μM TMSP- d_4 , 10 mM potassium phosphate buffer with a final volume of 550 μL and a ratio of 1:1 KPi in $\text{D}_2\text{O}/\text{CD}_3\text{CN}$. The sample was locked to the residual CD_3CN peak and shimmed prior to the first measurement. BIQ formation measurements were conducted every 30 min over 23 h. The integrals were compared to the signal of TMSP- d_4 . After 8.5 h the reaction was complete (**Figure 28**).

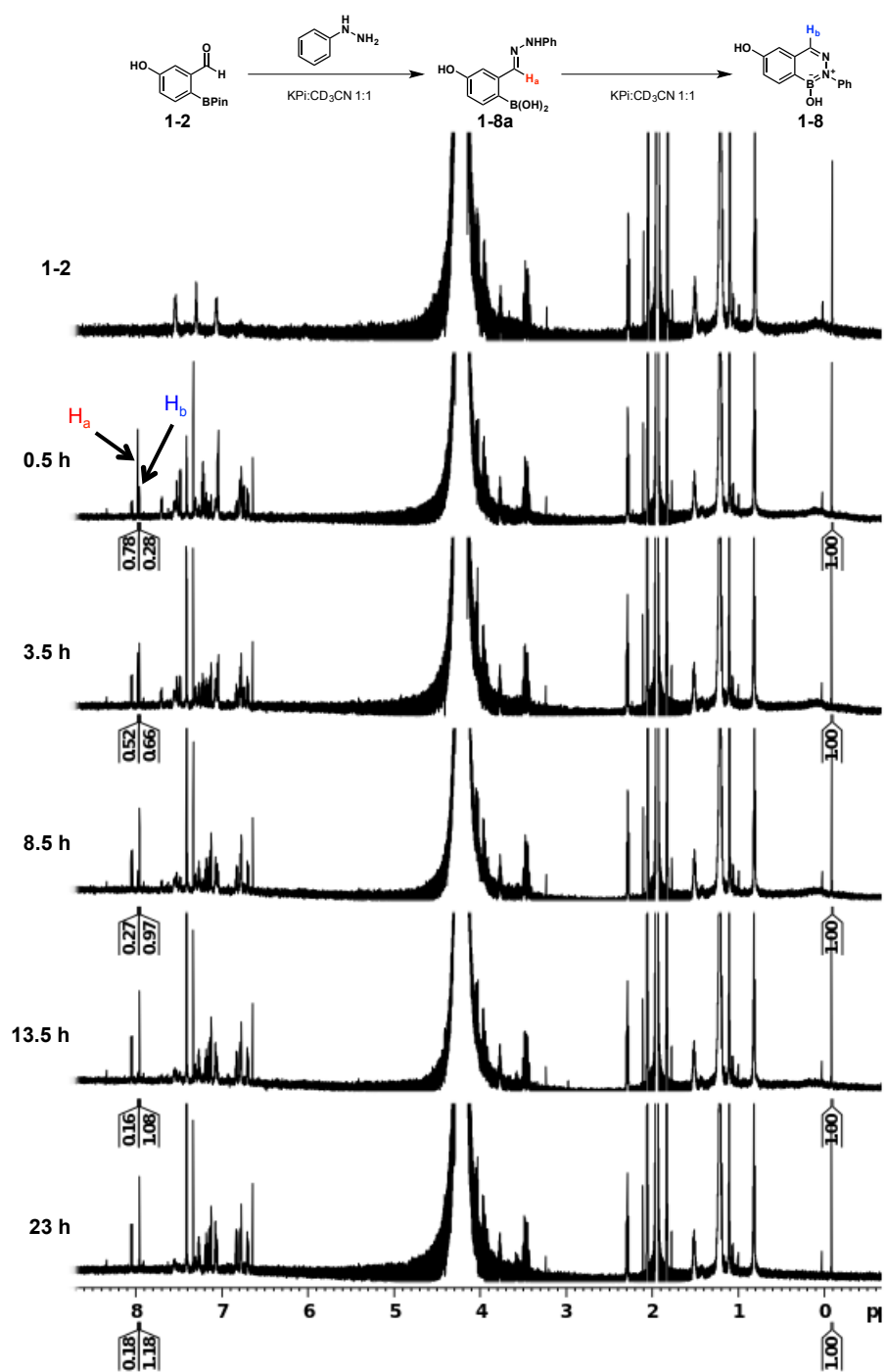
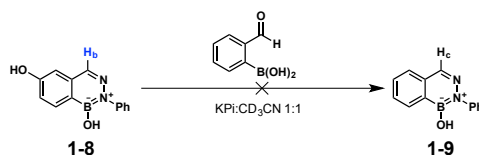


Figure 28. Time-course NMR of the BIQ formation over 23 hours. The reaction was carried out at 100 μ M concentration in 1:1 KPi:CD₃CN. The hydrazone proton (H_a) and the BIQ proton (H_b) are specified. The peak at -0.1 ppm corresponds to the internal standard (TMS-*d*₄). A zoom of the hydrazone/BIQ protons is found in **Figure 18a**.

5.7 NMR Assay for BIQ Reversibility



Stock solutions:

207.5 μ M **1-8** in CD₃CN

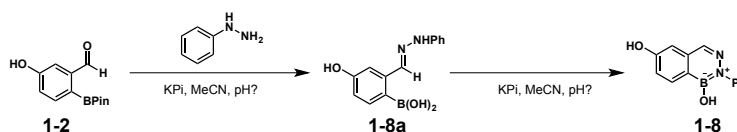
27.5 mM 2-FPBA in CD₃CN

275 μ M TMSP-*d*₄ in D₂O

20.75 mM potassium phosphate buffer in D₂O, pH 7.25

The stock solutions were mixed in a NMR tube with the following final concentrations: 100 μ M **1-8**, 500 μ M 2-FPBA, 5 μ M TMSP-*d*₄, 10 mM potassium phosphate buffer with a final volume of 550 μ L and a ratio of 1:1 KPi in D₂O/CD₃CN. The sample was locked to the residual CD₃CN peak and shimmed prior to the first measurement. BIQ reversibility measurements were conducted every 30 min over 24 h, starting with the first measurement after 10 min. The integrals were compared to the signal of TMSP-*d*₄. After 24 h no changes were observed. The comparison of the NMR spectra is found in **Figure 18b**.

5.8 pH Effects on BIQ Formation



Stock solutions:

10.04 μ M **1-2** in 100 mM potassium phosphate buffer pH 5.50/MeCN 9:1

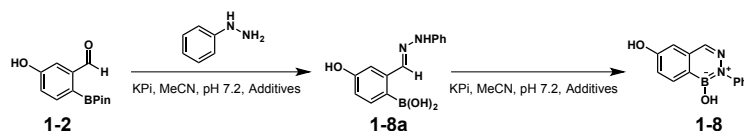
10.04 μ M **1-2** in 100 mM potassium phosphate buffer pH 7.20/MeCN 9:1

10.04 μ M **1-2** in 100 mM potassium phosphate buffer pH 8.50/MeCN 9:1

2.5 mM phenylhydrazine in *dd* H₂O/MeCN 1:1

2490 μ L of the **1-2** stock solution was measured for 10 s, after which 10 μ L of the phenylhydrazine stock solution was added and measured. Result analysis and calculations were performed with MS Excel. The accuracy of the calculated reaction kinetics is given by the calculated standard deviation (SD). The UV traces and the rate constant calculations are shown in **Figure 19a** and **Table 2**.

5.9 Effects of Glutathione and Human Serum on BIQ Formation



Stock solutions:

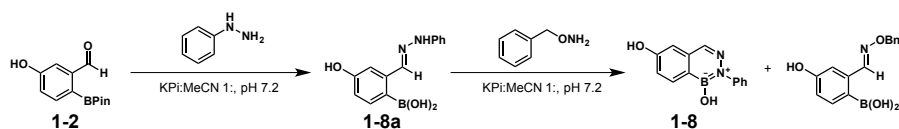
10.04 μM **1-2** in 100 mM potassium phosphate buffer pH 7.20/MeCN/HS 8:1:1

10.04 μM **1-2** in 100 mM potassium phosphate buffer pH 7.20 with 5 mM GSH/MeCN 9:1

2.5 mM phenylhydrazine in MeCN

2490 μL of **1-2** stock solution was measured for 10 s, after which 10 μL of the phenylhydrazine stock solution was added and measured. Result analysis and calculations were performed with MS Excel. The accuracy of the calculated reaction kinetics is given by the calculated standard deviation (SD). The UV traces and the rate constant calculations are shown in **Figure 19b** and **Table 2**.

5.10 Hydrazone Reversibility Assay



Stock solutions:

100.4 μM **1-2** in 100 mM potassium phosphate buffer pH 7.20/MeCN 1:1

25 mM phenylhydrazine in MeCN

25 mM benzylhydroxylamine in MeCN

2490 μL of **1-2** stock solution was measured for 10 s, after which 15 μL (1.5 eq.) of the phenylhydrazine stock solution was added and measured. After 3.5 min 30 μL (3 eq.) of the hydroxylamine stock solution were added. BIQ formation turned out to be much slower compared to previous results even though a ten times higher concentration was used. This showed the decelerating effect of higher organic solvent concentrations. The kinetic results were analyzed and processed with MS Excel. After 13 h, the mixture was analyzed by UPLC-MS. About 15% of the corresponding oxime product was observed, proving that the BIQ formation was faster than the hydrazone hydrolysis. The result of the measurement is shown in **Figure 20** and **Figure 29**.

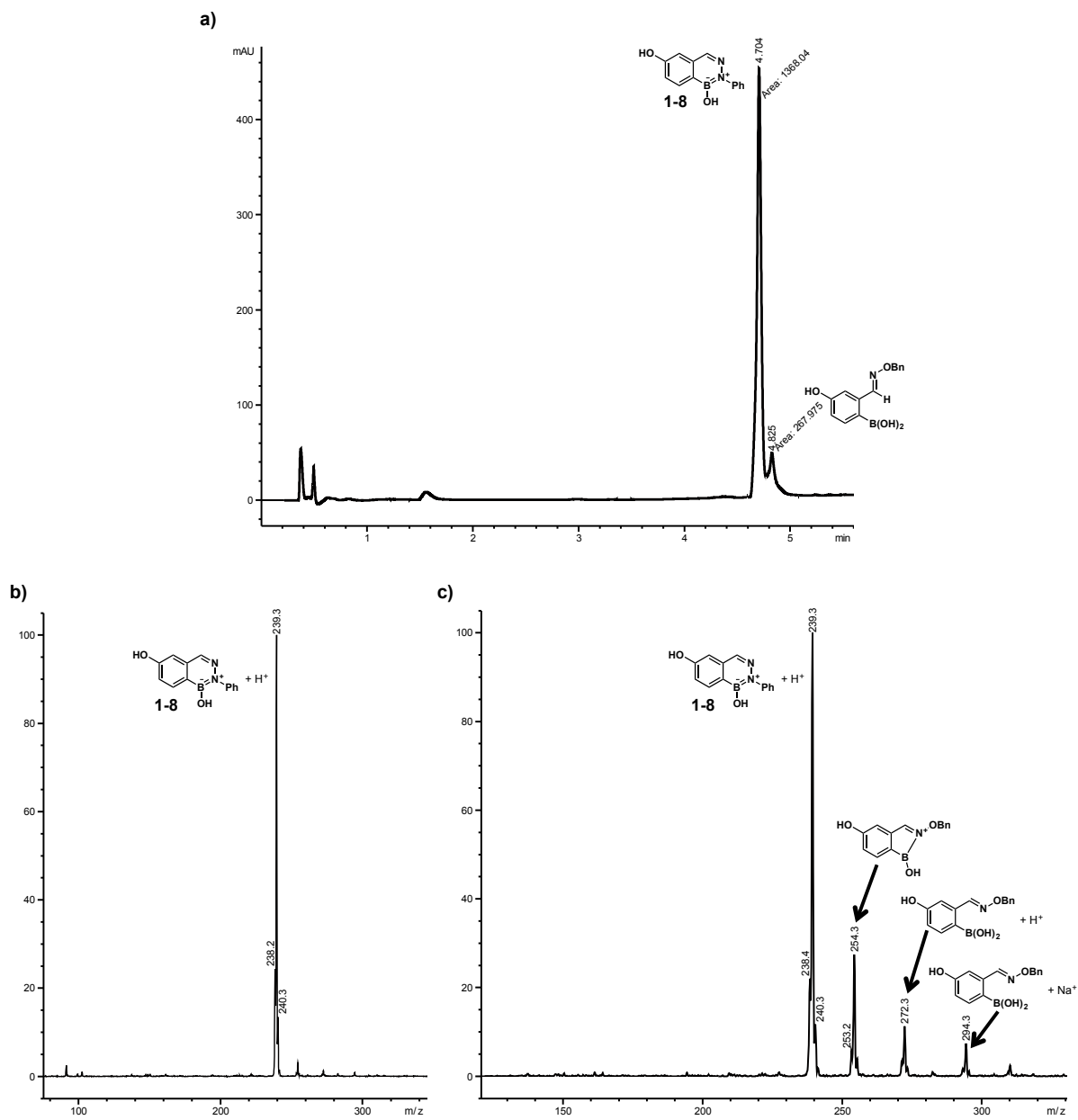


Figure 29. **a)** UPLC chromatogram of the hydrazone reversibility assay after 13 hours reaction time. Approximately 15% oxime was found next to the BIQ 1-8 product. **b)** MS trace of the BIQ product. **c)** MS trace of the oxime product. Due to peak overlap, the BIQ ion is found as the most intense signal. In contrast to the BIQ, the oxime also showed a dehydration ion.

5.11 Fluorescence Measurements

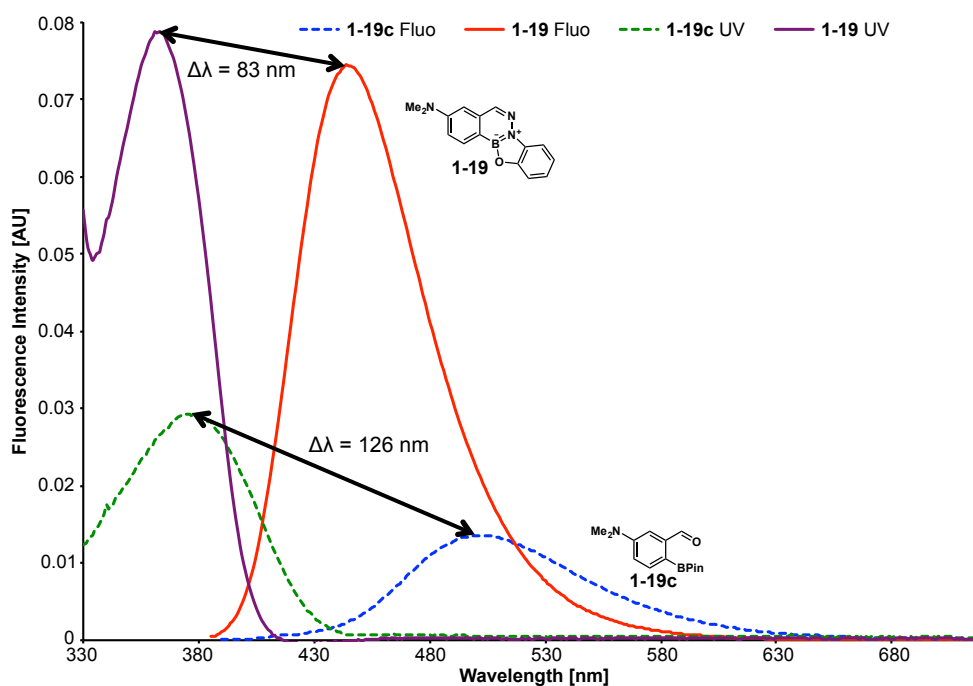


Figure 30. UV and fluorescence spectra of **1-19** and **1-19c** in DMSO at 10 μM concentration. Arrows indicate the Stokes shift. For additional fluorescence properties, see **Figure 22**.

Quantum yields were determined at 10 μM concentration in DMSO with excitation wavelengths 370 nm and 365 nm for **1-19c** and **1-19** respectively.

$$\Phi_{(370 \text{ nm})} = 0.279 \text{ (1-19c)}$$

$$\Phi_{(365 \text{ nm})} = 0.386 \text{ (1-19)}$$

6. Introduction to DNA-Encoded Chemical Libraries

The development of DNA-encoded libraries is the focus of this part with an emphasis on the development of a natural product-like encoded macrocycle library. Library assembly, properties evaluation and protein affinity screening results are presented. Before the discussion of the experimental assays, a detailed introduction to the main concepts of combinatorial chemistry and encoding of chemical libraries is given.

6.1 Combinatorial Chemistry and High-Throughput Screening

The discovery of new lead compounds for drug development is an essential task for modern pharmaceutical research. The generation of large molecule collections is very tedious and costly due to the synthesis, purification and characterization of each substance. Combinatorial chemistry offers the potential to simultaneously generate large compound collections within a relatively short time.^[103] These libraries usually range from several thousands to up to billions of different members. Library assembly is often performed by the chemical linkage of available building blocks in a repetitive manner.^[104] Screening of these compound collections against biological targets (e.g. proteins) may lead to the discovery of new and interesting binders that are then further optimized with medicinal chemistry tools, often with the help of combinatorial methodologies. The concepts of combinatorial chemistry were developed in the 1980's when, for the first time, hundreds of peptides were synthesized in parallel on a solid support.^[105,106] In the following decade, a large number of new library technologies were evolved, such as one-bead-one-compound (OBOC) libraries,^[107] phage display technology^[108] and DNA-encoded chemical libraries.^[109]

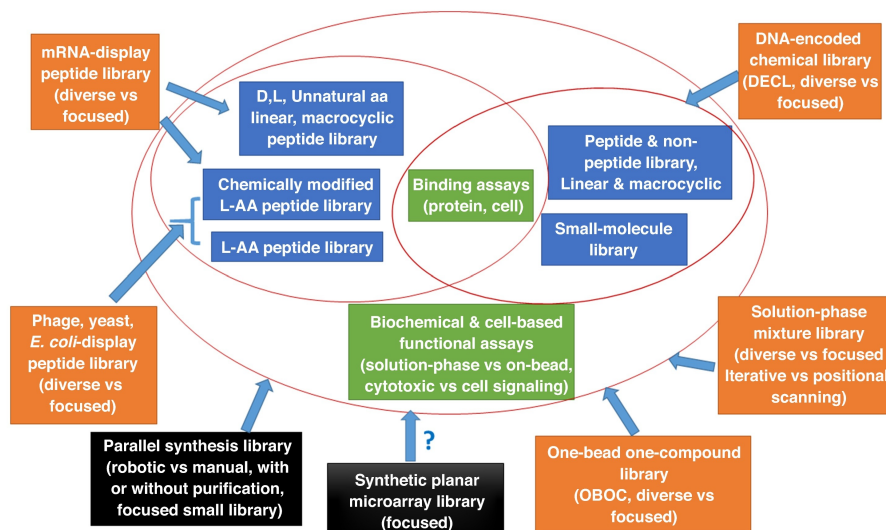


Figure 31. Overview of the different types of combinatorial chemistry libraries. Orange/black boxes show the combinatorial technique, blue indicates the type of the chemical library and green shows the screening assay type.^[104] Reprinted with permission.^a

Figure 31 shows an overview of the different types of combinatorial libraries. While most library types have the potential to generate compound collections of over a million members, parallel synthesis libraries and microarray libraries are much more focused (black boxes) and usually smaller. In the maturation process of the combinatorial chemistry technique, computational

^a Elsevier, License Number: 4526970559050, 13.02.2019

chemistry evolved as an essential tool for the *in silico* creation and analysis of combinatorial libraries as well as for the handling of large screening data sets.^[104] Computer-assisted library design contributes to the improvement of the chemical structure of building blocks and library members to cover a broader chemical space^[110] or to create more specific variants of a known target binder.^[111] Although combinatorial chemistry has opened a vast potential for the development of new drug substances, the drawbacks of this technology soon emerged. New strategies for library design were required for the identification of single compounds in large mixtures. Moreover, to create one member of every theoretically possible molecule with a molecular mass below 500 Da only using the common atoms of organic structures would exceed the number of atoms available in the universe.^[103] After the initial enthusiasm about the new technologies, combinatorial chemistry was relegated for a long time to the creation of focused libraries for hit optimization whereas high-throughput screening (HTS) of compound collections was employed for the *de novo* discovery of drug compounds. Advances in combinatorial chemistry, especially the successful encoding of library members, led to the revival of these technologies and is nowadays re-employed in drug discovery and optimization by academic and industrial researchers.

In the early days, companies performed HTS with randomly assembled compound collections from previous activities in dye and fine chemical synthesis.^[112] Combinatorial methodologies enlarged those compound collections, using a few simple chemical reactions without paying too much attention to diversity issues. Owing to the huge chemical space, one might think that testing large numbers of molecules should increase the chance of finding new hits. However, it is more important to screen compound collections with a high diversity rather than large numbers.^[113] An improved hit rate is often found with structurally diverse libraries in comparison to large, structurally more similar compound collections. This factor is why large pharmaceutical companies have invested a lot of money in maintaining and improving their libraries, with a focus on variety rather than size. Such collections typically consist of several hundred thousand to a few million compounds. HTS almost never delivers high affinity hits that fulfill all the criteria for drug market release in first instance.^[113] Usually, the found hits become lead structures, which are further optimized by computational and combinatorial methodologies.^[112] The limiting aspect of HTS screening libraries is that every library member needs to be acquired or synthesized, stored, dispensed and screened individually. These operations require large financial investments, equipment, space and people, which define the boundaries of the screening assay. With the technological progress in process automation, screening campaigns have become automated.^[114] However, a screening campaign with about 1 million compounds lasts up to several months. Most often several rounds of screening are required to minimize the rate of false positive and false negative results. Biochemical methods are routinely used in *in vitro* HTS but cellular screening assays are also employed for lead discovery and optimization due to the higher physiological relevance. Fluorescence, luminescence and absorbance measurements are the most common detection methods for those operations. The measurable signal is either generated by direct enzyme of interest targeting (and thereby start or stop the generation of a luminescence signal) or by targeting the enzyme/protein to indirectly (de)activate a secondary process for the termination or release of a measurable signal by a second, unaffected enzyme.^[114] Luciferase^[115] and GFP^[116] proteins are often found in these kinds of assays.^[113]

6.2 Phage Display Libraries

In 1985, Smith^[108] described, for the first time, a system in which a filamentous bacteriophage (a virus that can infect bacteria) was used for the expression of peptides and proteins on its surface, induced by a previously cloned vector.^[117] This method enabled the construction of phage libraries, in which every virus expressed a different peptide or protein. Most commonly filamentous *E. coli* bacteriophages (f1, fd, M13) are used for the generation of such libraries. **Figure 32a** shows the schematic setting of a phage display library. The expressed protein/peptide is usually displayed as a fusion protein with the phage pIII protein. The phage pVIII protein is alternatively used for the generation of such fusion proteins, however, only short peptides may be expressed with this system in order to retain phage functionality.^[118] The expressed protein on the surface (phenotype) is physically connected to the viral DNA (genotype) inside the phage capsule.

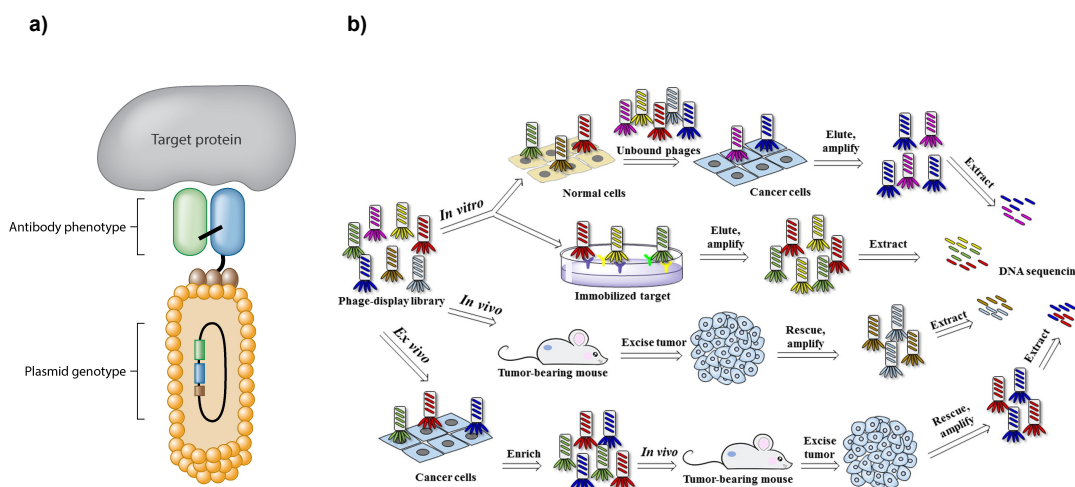


Figure 32. a) Schematic representation of a phage display library with an example of a displayed antibody. The library is divided into a phenotype part (antibody or protein/peptide) and a genotype part (DNA strand inside the viral cage).^[119] b) Schematic description of the selection procedure (biopanning) with phage display libraries *in vitro*, *in vivo* and *ex vivo*. Several rounds of selections are usually conducted to find high affinity ligands.^[120] Reprinted with permission.^{bc}

This technique allows the construction of libraries containing up to 10^{10} different variants, which are simultaneously screened against the target. DNA sequencing identifies all binders obtained after a selection round (biopanning) due to the genetic code inside the phage capsule. This is a big contrast to the libraries described in the previous section, where no compound encoding was applied. Phage display technology improves studies of protein-ligand interactions, receptor and antibody-binding site characterizations, screening of cloned antibody collections, search for enzyme substrates and display of epitopes for monoclonal antibodies. There are fewer limitations in terms of length and size of the expressed peptides and proteins compared to conventional synthetic peptide libraries.^[120] Phage display libraries are highly efficient, inexpensive and even commercially available. *In vitro*, as well as *in vivo* assays, have been conducted with this library type (**Figure 32b**). Since bacteriophages can only express linear peptides, Heinis *et al.* successfully introduced post-translational bridging moieties to generate macrocyclic peptide libraries for an improved scaffold diversity.^[121,122] The technology's main drawbacks are its almost exclusive applicability to binding assays and the expressed peptides and proteins only consist of the proteinogenic *L*-amino acids. Typical screening procedures for ligands with this library type are illustrated in **Figure 32b**. Biopanning starts with a first round of selections of the library against the targets. Weak or non-binding members are simply washed away and the remaining

^b Annual Reviews, License Number: 4527170686357, 13.02.2019

^c Elsevier, License Number: 4527140523283, 13.02.2019

binders are eluted and amplified in bacteria. *In vitro*, *in vivo* or *ex vivo* procedures, as well as combinations thereof, are possible depending on the assay requirements. Several rounds of target testing are often conducted to increase the enrichment of tight binders. The development of human monoclonal antibodies is a field in which phage display technology has been successfully applied.^[117]

6.3 DNA-Encoded Chemical Libraries

Inspired by the success of monoclonal antibody development with phage display libraries, Brenner and Lerner proposed an encoded chemical library technique in 1992.^[109] Their system used a solid support (beads) to immobilize each screening compound along with an encoded DNA strand on the same bead. This assembly allowed the synthesis of identifiable small molecules in a combinatorial way, giving access to very large libraries. The attached DNA strands carry unique "barcodes" for the different molecules, which allows testing the complete set of encoded molecules at once. This strategy was considered as a combination of the traditional high-throughput screening of compound collections and the encoded screening from phage display.^[109,119,123] Only one year later, two groups published reports on the synthesis of bead-immobilized peptide libraries with attached "barcode" oligonucleotide strands for peptide sequence encoding.^[124,125] A few years later, three groups reported new concepts for the design of encoded small molecule libraries without the need of a solid support.^[126–128] These findings led to the development of several different construction strategies for DNA-encoded chemical libraries (DECL).^[126,127] This technology opened up the possibility to simultaneously screen very large numbers (up to billions) of small molecules in a time- and cost-efficient manner, which makes it affordable for smaller companies and academia.

Types of DNA-Encoded Chemical Libraries

There are two main DECL types that differ in the type of small molecule attachment. Single-pharmacophore libraries (**Figure 33a**) consist of a double-stranded DNA with the small molecule attached to one of the strands (usually at the 5' end). This is the most frequently found library type in literature. Single-pharmacophore libraries are further subdivided in different types, depending on their encoding scheme (see next section).^[129] The second, less commonly used library type is the dual-pharmacophore library (**Figure 33b**). In this setup, two pools of encoded small molecules are combined to form the final library. Both pools contain single-stranded DNAs that are complementary to each other. The small molecules are attached to the 5'- or the 3'- end of the oligonucleotides. Upon hybridization of the sub-libraries, a large combinatorial diversity is achieved and the ligand pairs end up at the same extremity of the duplex DNA strand. This library design is often referred to as encoded self-assembling chemical (ESAC) library technology and was introduced by researchers from the Neri group in 2004.^[127]

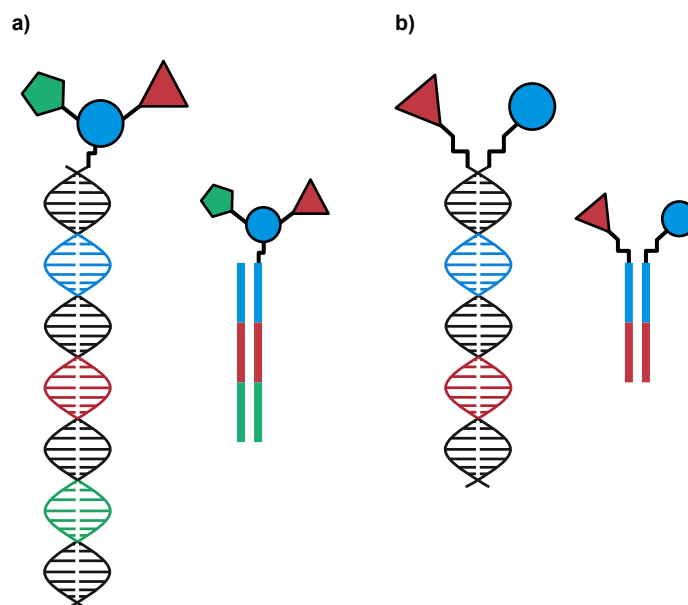


Figure 33. Schematic design of **a)** single-pharmacophore libraries and **b)** dual-pharmacophore libraries. The library types are drawn in a double helix DNA structure and in a linear structure.^[119]

Single-Pharmacophore Libraries: DNA-Templated Synthesis

One variant of single-pharmacophore libraries was pioneered by the Liu group as early as 2001.^[130] In this approach, a preformed library of long DNA strands (templates), which carry the molecular scaffold (and/or the first building blocks) subsequently reacts with a set of building blocks, encoded by complementary DNA strands. Prior to the synthesis of the actual pharmacophore, the DNA template library is preassembled by a split-and-pool technology.^[131] The added oligonucleotides carrying the new building blocks must be highly specific for annealing with certain regions of the longer DNA strand. The high specificity of the encoded building blocks is essential to correctly bring them in close proximity to its reaction partner. This assembly increases the local molarity of the reagents, which is very advantageous for the synthesis of encoded libraries. After successful coupling of the new building block with the scaffold moiety, the linker between the building blocks and their directing DNA strands are cleaved and the redundant DNA is removed to prepare the library for the next round of synthesis (**Figure 34a**).

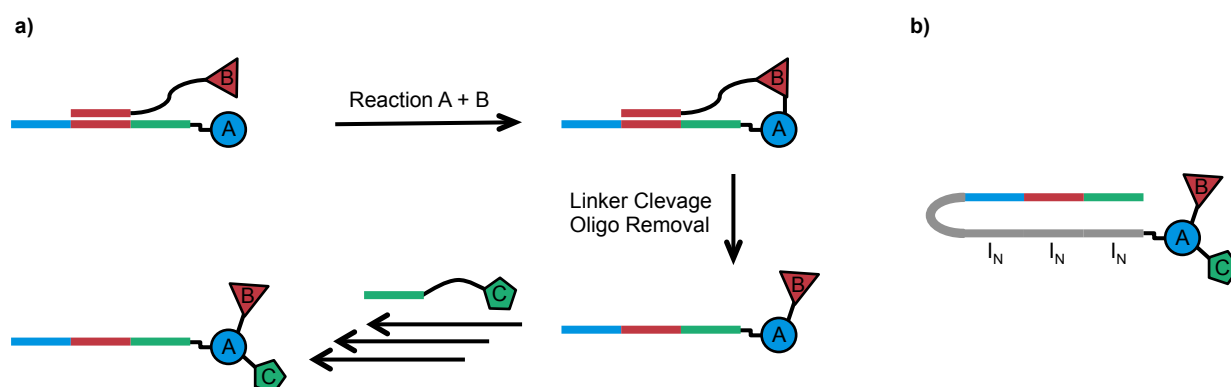


Figure 34. **a)** Schematic representation of the assembly of a DNA-templated library. The encoded building blocks specifically anneal with the template DNA strand to enable the chemical reaction. After the chemical reaction, linker cleavage and oligonucleotide removal are necessary to enable the next round of synthesis. **b)** A variation of the DNA-templated synthesis method. The template DNA strand is replaced with a DNA strand carrying poly inosine (I_N) stretches instead of predefined codons.

With that method, the Liu group has created several different DNA-templated macrocycle libraries.^[132–134] Li and co-workers reported a variation of this technology in 2013. A simple DNA strand containing poly-inosine stretches instead of predefined codons was used as the template DNA strand (**Figure 34b**).^[135] Since poly-inosine stretches can undergo base-pairing with any codon segments, the combinatorial pre-assembly of the template DNA strands is unnecessary. Further variations have been developed, such as the Yocto reactor from the Danish company Vipergen.^[136] This technology uses hairpin loops with attached building blocks that are annealed and ligated, so the reaction can be carried out in the thereby formed yoctoliter reactor. Repetition of this assembly yields a library of compounds with attached double-stranded DNAs that carry all necessary information. DNA routing is another type of DNA-templated synthesis and was reported by the Harbury group in 2004.^[128] Long, preformed DNA strands are annealed to complementary strands that are attached to different solid supports. Only the oligonucleotides with the correct annealing site are retained on the support by hybridization, the excess DNA is washed away and transferred to another solid support with a different annealing oligonucleotide. Every solid support is then used for the reaction with a specific building block. This setup allows a library construction in a templated fashion, using several rounds of hybridization to the solid supports and chemical reactions with the building blocks. The technique is strongly dependent on a high fidelity in the hybridization step and, due to the complicated setting for library generation, may quickly turn out to be impractical for the synthesis of big and complex DECLs.

Single-Pharmacophore Libraries: DNA-Recorded Synthesis

DNA-recorded synthesis refers to an encoding strategy in which newly attached building blocks are subsequently encoded by DNA fragments bearing the codons for those building blocks (**Figure 35a**). This leads to a continuous elongation of the DNA chain while "recording" the introduced chemical moieties by DNA codons. This technique is presently the most popular encoding strategy.^[119]

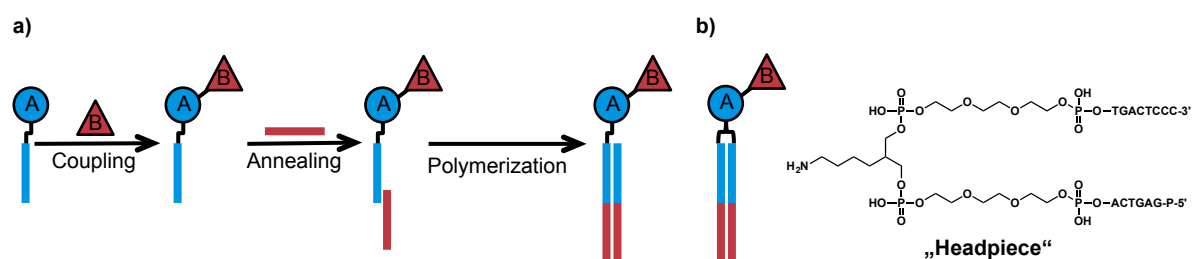


Figure 35. a) Schematic representation of a DNA-recorded library. Encoded building blocks are modified with new chemical moieties, followed by an encoding step with another DNA fragment. In the shown case a single-stranded sub-library is chemically modified, followed by encoding with a second, partially complementary single-stranded DNA. This yields a duplex DNA after enzymatic polymerization. The same type of library can also be obtained by sequential encoding with duplex DNA fragments. b) Representation of a library using duplex DNA fragments encoding whereby the "headpiece" covalently connects the forward and reverse DNA strands.

As outlined in **Figure 35a**, the construction of this library type begins with a set of encoded chemical moieties for which either single- or double-stranded DNAs can be used. The encoding of a single-stranded sub-library often proceeds by splint ligation and Klenow fill-in techniques.^[137] During this process, the single-stranded DNA is converted into double-stranded oligonucleotides which bear the codon and anticodon regions. There is no limit to the number of diversity elements that can be introduced since several consecutive splint ligation steps or repetitive restriction/ligation procedures with the duplex DNA can be performed.^[138] Another encoding possibility is to directly start from a duplex DNA fragment, which is chemically modified and continuously encoded by duplex DNA ligations. Researchers from GlaxoSmithKline (GSK) developed a very interesting way for duplex DNA encoding by covalently connecting the two DNA

strands with each other using the so-called “headpiece” (**Figure 35b**). This part consists of a phosphorylated and PEGylated branched amine, which carries complementary short DNA strands (six and eight bases) at each end of the two PEG linker chains. This assembly creates a 3' two base overhang that serves as the “sticky end” for consecutive duplex DNA encoding with short duplex DNA fragments.^[139] The GSK researchers created a library of several hundred million members, applying four cycles of chemical modifications and subsequent encoding.

Dual-Pharmacophore Libraries: Encoded Self-Assembling Chemical (ESAC) Libraries

Dual-Pharmacophore libraries consist of complementary DNA strand pairs, which possess chemical moieties at their extremity.^[140] This type of library, designed in 2004 by the Neri group,^[127] was initially used for decoding with oligonucleotide microarrays.^[141] A modified approach was introduced in 2015, which allowed for decoding of the library after affinity selections by high-throughput sequencing.^[142] The first set of building blocks is coupled to a non-coding oligonucleotide containing an abasic stretch that is further encoded by splint ligation to generate sub-library A. This library contains DNA strands with the codons for all the used building blocks along with the abasic site that is necessary for hybridization with the second sub-library. Sub-library B consists of building blocks that are directly coupled to their encoding DNA strands.

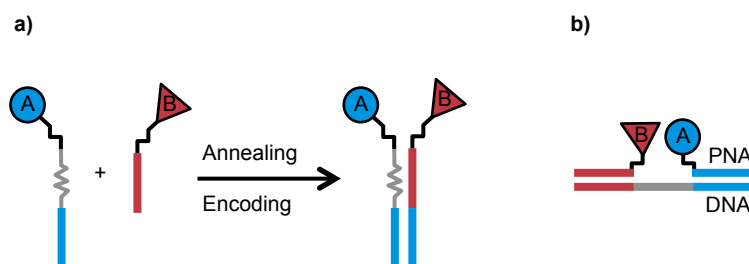


Figure 36. a) Schematic assembly of an ESAC library. Two sets of encoded building blocks containing complementary DNA regions are annealed and subsequently encoded by polymerization. One set (building blocks A) contains an abasic site (grey zig-zag line), which can anneal with any codon from the second set of building blocks. b) A variation of the technique uses DNA strands as templates, which can anneal with PNA-encoded sub-libraries.

Hybridization of the two sub-libraries yields all possible combinations in a double-stranded format. Subsequent polymerization copies the code from sub-library A onto the sub-library B DNA strand (**Figure 36a**). This arrangement generates one DNA strand with the codons for the building block pair, which is analyzable by high-throughput DNA sequencing. The chemical building blocks need to end up at the same side of the DNA duplex for application in protein affinity selections. Therefore, the building blocks must be connected to the 3'-end of one sub-library and to the 5'-end of the other sub-library.

Researchers from the Winssinger lab presented a similar approach, using peptide nucleic acids (PNAs) as encoding tags.^[143] The PNA-encoded sub-libraries were hybridized with a pool of complementary DNA template strands to generate the dual-pharmacophore library (**Figure 36b**). PNAs were found to be more stable towards certain chemical reaction conditions compared to DNA, which creates the possibility to use chemical modifications of the building blocks, unknown to DNA-encoded library synthesis. Although PNAs can be chemically encoded during library assembly, an amplification procedure, such as PCR, known from DECLs, is not possible. That prevents an amplification of the PNA code for protein binders and renders decoding procedures very difficult. This is the reason why the PNA-encoded sub-libraries are hybridized with complementary DNA strands containing the building block codes in every possible combination.

Another very interesting approach is the encoded dynamic combinatorial chemical libraries (EDCCL) methodology. This technology uses dual-pharmacophore libraries that are thermodynamically less stable.^[144] A lowered DNA melting temperature, caused by short

annealing stretches, induces this instability. Upon strong binding of a building block pair to the protein substrate, the thermodynamic equilibrium is shifted such as to generate higher amounts of the strong binder pair due to the dynamic recombination of the library. Repeating the procedure results in a high enrichment of the potent building block pair.

6.4 Chemical Requirements and Hit Validation

The main advantage of DECLs, compared to conventional libraries (compound collections), is that every single member is linked to a unique DNA codon. This phenotype (small-molecule) - genotype (DNA tag) connection is the key for successful hit identification, even though the biologically more interesting part consists of the small-molecules collection. Since most libraries are assembled in a split-and-pool synthesis manner,^[131] chemical reactions need to be conducted in the presence of the DNA tag. For this purpose, the applied chemical reactions need to be oligonucleotide-compatible.^[145] As we know from the research of Watson and Crick in the 1950s, DNA consists of a phosphoribose diester backbone with four different nucleobases (two purine and two pyrimidine bases).^[6] This setup shows many potential modification sites for reactive chemicals. Predominantly, the phosphate backbone and the bases are susceptible to chemical modifications. Such alterations of the DNA tag in a DECL bear the risk that the DNA cannot be correctly amplified by PCR. Furthermore, lack of amplifiable DNA or read-out errors during oligonucleotide sequencing might result in the loss of potential target hits.

Most standard chemical reactions are conducted at relatively high concentrations (10 mM - 1 M concentrations). For modifications of chemical entities attached to a DNA strand, however, the reactions need to be performed at much lower concentrations (1 μ M - 1 mM concentrations). Stoichiometric increase of reagents (up to several thousand equivalents) forces reactions to proceed under pseudo-first-order conditions. This is why reactions that usually require millimolar concentrations are possible to be carried out with one reagent (such as the DNA-tagged building block) present at micromolar concentrations. The risk of random modifications to the DNA tag increases with the amount of added reagents. Additionally, only reactions and reagents compatible with the aqueous conditions, necessary for solubilizing the DNA, are useful for DECL construction. Therefore, highly reactive chemicals, such as organo-lithium compounds, most hydrides, anhydrides, acyl chlorides or very sensitive transition metal catalysts, may not be used since they react (violently) with water. Routinely-used reactions for the construction of DECLs include amide bond formation^[146,147], copper-catalyzed click reaction^[137,144], reductive amination^[146,148], Suzuki reaction^[42,145,148], Wittig-type reactions^[132,134], Diels-Alder reaction^[138,149,150], Staudinger reduction^[137] and certain transition metal-catalyzed reactions.^[151-155] Even though there is a good set of chemical transformations that have been successfully applied for DECL construction, there are still major limitations to the methods due to the dependence of certain building blocks on special reaction conditions (organic solvents, high temperatures, additives, acidic pH). Immobilizing the DNA on a solid support helps overcome the DNA incompatibility with high levels of organic solvents. Although such a technique allows the use of special chemical transformations, the immobilizing strategy also complicates library synthesis because of extensive solid support handling and might lead to undesired DNA damage.^[138] A recent publication from Paegel investigated the effects of a series of reactions on the amplification performance of a DNA strand after the chemical reaction.^[145]

In general, double-stranded DNA is less susceptible to undesired modifications due to the shielding properties of the Watson-Crick base pairing (also see results in Chapter 8.3). This is why researchers from GSK preferred to develop the "headpiece" duplex DNA-encoding system for the generation of their 800 million member library (description in Chapter 6.3).^[139]

Such limitations clearly specify the need for new DNA-compatible reactions to generate new and more diverse libraries by reactions currently unknown in DECL assembly.^[156-158]

For the construction of a successful library, there are several properties that need to be fulfilled by the building blocks themselves. The DNA-compatibility and the reactivity under the aqueous conditions were mentioned before, but the size and shape of the building blocks also need to be considered. Depending on the number of used building blocks, the molecular weight of the individual blocks is limited. Additionally, functional groups, heteroatoms and chiral centers play a very important role since they define most of the properties of the final compound library. Creating drug-like molecules is only possible by taking the proposed physicochemical guidelines, such as the rule-of-five defined by Lipinski, into account.^[159] For a detailed discussion on the properties of conventional small molecules and macrocycles, see Chapter 8.1.

In the simplest implementation, chemical building blocks with one reactive functional group (most often amines or carboxylic acids) are coupled in a single step to the encoding DNA strand. Such encoded compound collections (sub-libraries) are commonly used in dual-pharmacophore libraries (see previous section). More complicated libraries are created by the subsequent attachment of two or more building blocks to the same encoding oligonucleotide (most often employed in single-pharmacophore libraries). For this purpose, the building blocks and/or the library scaffold require at least two sites with reactive functional groups to conduct the building block connections. For linear compounds, amino acids (such as alanine, proline or serine) and amino acid derivatives are very well suited. It is often desirable to also obtain branched or cyclic molecules, which demand a combination of building blocks with two or three modification sites (e.g. lysine, cysteine or aromatic derivatives).

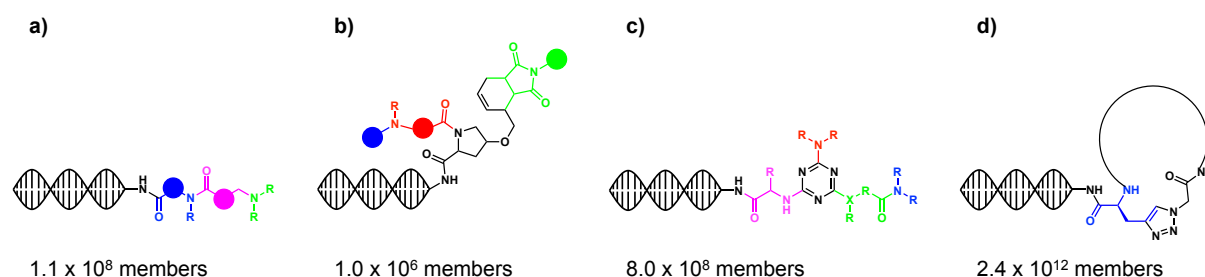


Figure 37. Representation of selected DECL examples with various types of building blocks. The exact encoding scheme with codon coloring was neglected for simplicity. **a)** A linear encoded library consisting of amino acids (blue), formyl acids (pink) and amines (green).^[160] **b)** A branched DECL using a trifunctional linker basic scaffold (black). The diversity was introduced with amino acids (red), amine modifiers (like carboxylic acids or sulfonyl chlorides, blue) and maleimides (green).^[138] **c)** A four building blocks library using 1,3,5-triazine as the central scaffold trifunctional linker. Diversity was introduced with amino acids (pink), bifunctional acids (green) and amines (red and blue).^[139] **d)** A macrocytic library consisting of coupled amino acids. The central scaffold (blue) consisted of an amino acid with an alkyne sidechain for click macrocyclization.^[161]

Figure 37a shows a typical example for a library that was designed in a linear fashion. The DNA strand was modified with a set of amino acids that were further coupled to a collection of formyl acid building blocks. The sub-library was finally modified with a series of amines by reductive amination to yield a 110 million member DECL.^[160] A branched approach was shown by the Neri group for the discovery of carbonic anhydrase inhibitors from a one million compound library.^[138] They used a modified hydroxyproline central scaffold, which connected the DNA strand with three different sets of building blocks (**Figure 37b**). Aromatic (or heteroaromatic) rings are very well suited for the construction of non-linear encoded libraries. In the previously mentioned GSK 800 million members library^[139], 1,3,5-triazine (from cyanuric chloride) was used as the central scaffold to connect the DNA codons with the building blocks (**Figure 37c**). In their approach, they used bifunctional building blocks (amino acids and bifunctional acids), as well as monofunctional building blocks (two sets of amines) for library construction. Another approach from GSK researchers was reported in 2018, showing a collection of 2.4 trillion macrocyclic peptides (**Figure 37d**).^[161] They assembled the library using amino acid couplings in a linear fashion,

which was macrocyclized in the last step using the Click reaction. For this purpose, an amino acid-derived trifunctional linker with an alkyne sidechain was used.

The selection of a DNA-encoded chemical library against the target proteins of interest represents the most important step for the discovery of strong protein binders or enzyme inhibitors. In contrast to conventional high-throughput screening, no selection-related absorption or luminescence events are necessary to detect binding of the ligand to the target. The most common approach used for these selections is to immobilize the target on a solid support and perform an affinity screen with the DECL. Commonly used solid supports include functionalized magnetic beads^[162] or cyanogenbromide-modified sepharose.^[163] Such particles, especially magnetic beads, allow very simple bind, wash and elute procedures. Furthermore, they enable automation of the affinity screenings, which results in a faster screening of multiple target selections in parallel.^[164] qPCR was shown to be a reliable tool for selection quality estimations before DNA sequencing. With this relatively simple quality control, potentially tricky selections are optimized to improve the enrichment factor of the binding molecules.^[165]

Selection strategies that do not make use of solid support immobilization have also been proposed. Capillary electrophoresis might be applied to separate protein-ligand complexes from the remaining library with subsequent DNA sequencing.^[166,167] Interaction-dependent PCR is another option for affinity selections. A library of DNA-tagged proteins is screened with a DECL in the presence of a polymerase. In the event of a ligand binding to the protein, the DNA-tag from the protein anneals with the ligand DNA strand and serves as primer for the polymerase DNA chain extension (PCR).^[168] Owing to this PCR step, the code from the protein is copied to the ligand DNA strand, which is further analyzed by DNA sequencing. The technique allows the screening of protein collections with a DECL in a single step. The technology has been further improved by using interaction dependent PCR in unpurified protein mixtures from cell lysates.^[169] Direct crosslinking of the encoded ligand and the target protein enables the selection of strong and weaker ligands. Upon binding of the ligand to the protein target, a crosslinking unit attaches the encoded compound to the protein. The crosslink moiety is separately introduced into the selection mixture as a DNA-tagged reactive molecule, which undergoes a hybridization with the DECL to trigger the crosslinking of the ligand with the protein. This procedure stabilizes the protein-ligand complex and helps to enrich the binders more strongly.^[170] A similar variation of this method has also been applied to generate a covalent link between the target protein and the DNA codon of the ligand by a photo cross-linking method.^[171]

High-throughput sequencing has revolutionized the use of encoded chemical libraries. Decoding the eluted binders from the affinity selections identifies the molecules and yields a relative quantification by comparing the presence of the found molecules in the library before and after protein selection (enrichment factor).^[172] For the determination of the absolute binding properties, the ligands identified need to be individually tested. This is particularly important for the exclusion of false positive results, which often result from unspecific (random) binding to the protein, the solid support, or from systematic difficulties.

A series of hits is frequently chemically resynthesized and tested against the protein target with biochemical or biophysical methods.^[179] Isothermal titration calorimetry (ITC) is very well suited for this task since the binding constant (K_D) of the ligand to the protein target is directly determined from the measurement. Nevertheless, this methodology exhibits several limitations, most notably the requirement for a certain minimal amount of chemical and protein. This is the limiting factor for proteins, which are often only available in very low quantities. Furthermore, the solubility of the tested ligands might cause problems during affinity measurements. ITC accepts a certain percentage of organic solvents but the protein must be stable under these conditions. In this prospect, lipophilic compounds most often cannot be tested with this method.

For difficult cases, it may be beneficial to resynthesize the compounds with a fluorophore tag to perform fluorescence polarization measurements.^[172] On-DNA resynthesis ensures the solubility of the compound in the aqueous environment, which is especially advantageous for lipophilic molecules. Additionally, the synthesis of the compound is simplified due to the straightforward DNA purification techniques (such as ethanol precipitation). Researchers from the Neri group have recently shown methodologies to quickly assess the binding properties of DNA-tagged ligands by fluorescence and chemoluminescence methods.^[173] However, care must still be taken in the application of such methods, since the DNA strand can potentially have an influence on the binding affinity of the small molecule to the protein target. In DECL technology, there is a certain chance to miss potentially good target binders due to the changes in binding, induced by the oligonucleotide tag.

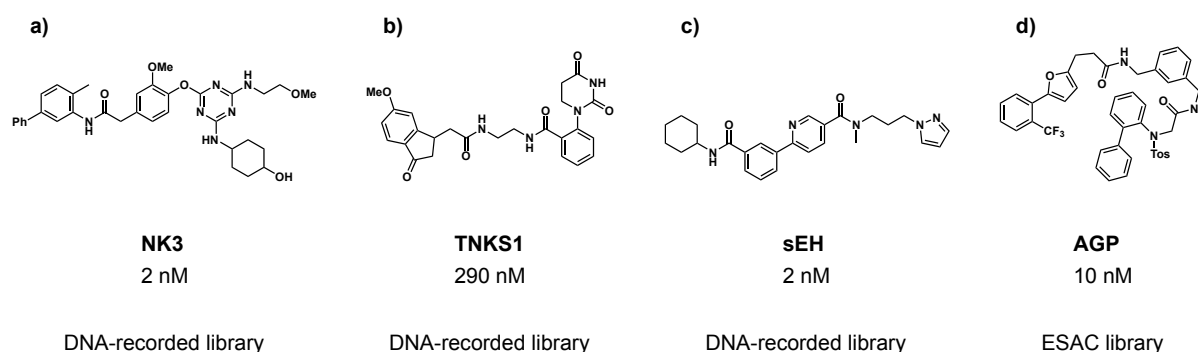


Figure 38. Selected examples for tight binders against various protein targets. **a)** A binder from GSK against neurokinin 3 (NK3) based on their "headpiece" encoded single-pharmacophore library technique with 41 million library members.^[174] **b)** A binder from the Neri research group against tankyrase-1 (TNKS-1) from a two building blocks recorded library consisting of 103'200 library members.^[175] **c)** A 334 million membered library from a recorded library with triazole connected oligonucleotide codons. The library was tested against soluble epoxide hydrolase (sEH) and revealed the shown binder.^[176] **d)** The shown α -1-acid glycoprotein (AGP) binder was identified from a 235'400 members dual-display library, developed by the Neri lab.^[177]

A few binders found against different target proteins are shown in **Figure 38**. All of these hits were discovered using the encoded chemical libraries technology with various types of library-encoding procedures. All of these binders showed very tight binding properties with binding affinities in the nanomolar range. The AGP binder (**Figure 38d**) was identified from an ESAC library, whereas all other shown binders were found from single-pharmacophore libraries. The soluble epoxide hydrolase (sEH) binder in **Figure 38c** was generated with a library approach that used copper-catalyzed click reactions to connect the encoding DNA fragments by triazole units. No enzymatic encoding was necessary for the generation of this library and the codon connecting units were well tolerated by the polymerases for affinity selection analyses.^[176]

6.5 Description of our Natural Product-like DNA-Encoded Macrocycle Library

The number of publications about DNA-encoded small molecule libraries has vastly increased in the last 15 years. Due to new encoding strategies, along with high-throughput DNA sequencing, the technology has become very interesting for industry and academia. Most of the reported libraries use linear or branched chemical moieties, often consisting of peptidic fragments. The number of encoded macrocycle libraries, however, is low; predominantly the Liu research group has developed DNA-encoded macrocycle libraries with their DNA-templated library synthesis approach. Their earliest report from 2004 generated a DNA-templated library with 65 macrocycles.^[132] A few years later, they reported the development of a new macrocycle library consisting of 13'000 members.^[133] From this library, several protein inhibitors were found.^[178,179] In 2018, they published the synthesis of a new library with 256'000 macrocyclic members.^[134] Two other independent reports were also published about the generation of DNA-encoded

macrocycle libraries. GSK showed the assembly of a peptide macrocycle library with 2.4×10^{12} members. The Neri group published a 35 million macrocycle-containing library, based on a peptide macrocycle scaffold with large sidechain diversity (**Figure 40d**). Macrocycles, as a structural class, are still under-represented in drug discovery despite their special properties concerning oral bioavailability (see Chapter 8.1 for a detailed description of the physicochemical properties of macrocycles).^[180]

We envisioned the design of a new DNA-encoded macrocycle library (DEML) comprising of mixed peptide-polyketide and polyene moieties that resemble natural product macrocycles.

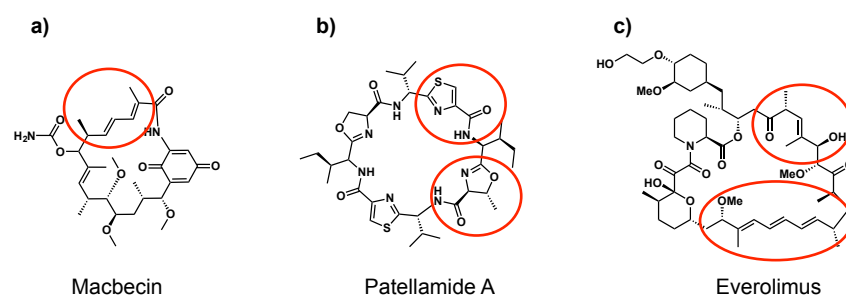


Figure 39. Structures of macrocyclic natural-products and drugs. Polyene, polyketide and special peptidic moieties are highlighted. **a)** Macbecin an antitumor antibiotic isolated from *Nocardia* bacteria.^[181] **b)** Patellamide A an anticancer therapeutic agent isolated from *Lissoclinum patella*, an ascidian.^[182] **c)** Everolimus is an immunosuppressant agent used in transplantation. It is also known as Afinitor from Novartis.^[183]

A few examples of such natural product-like macrocycles are shown in **Figure 39**. Our main focus during library development was the structural diversity of our macrocyclic scaffold. We included natural product-like moieties at two distinct positions in the ring, along with other hydrophobic elements to increase the number of different ring sizes and shapes. **Figure 40a** shows the schematic construction of our library. We used an ethylene diamine-modified terephthalic acid as the core structure for the further attachment of the building blocks and the encoding DNA strand.

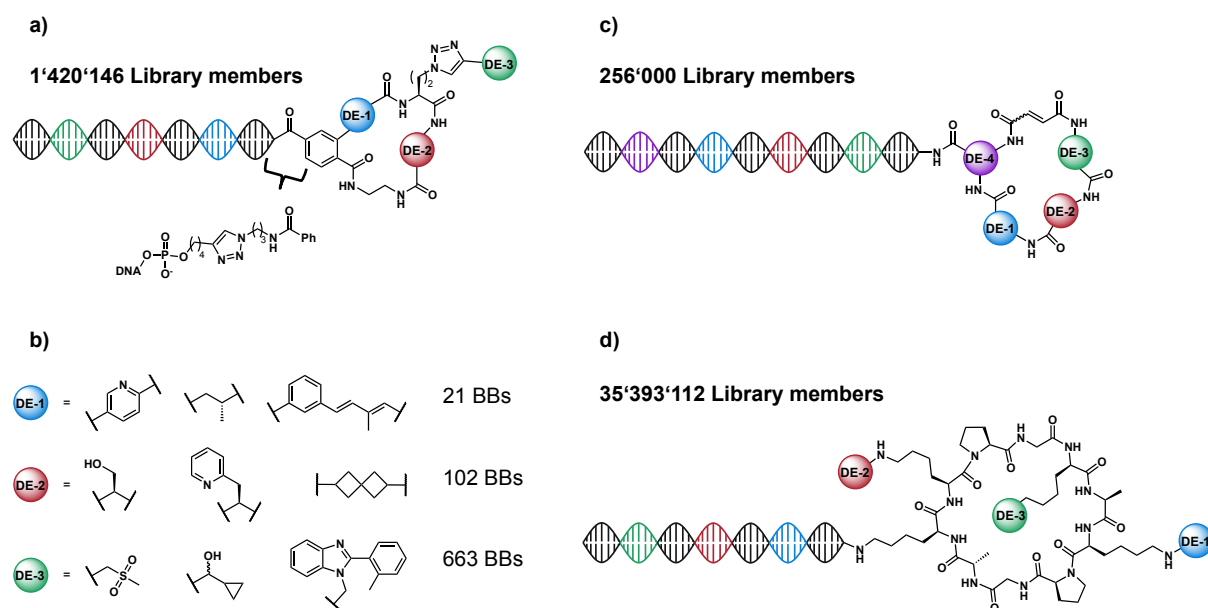


Figure 40 **a)** Schematic representation of the macrocycle library presented in this work. The library contains three building blocks, which are encoded by the attached DNA strand. A total of approximately 1.4×10^6 library members were generated during the assembly of the DEML. The structure part underneath displays the exact linkage between DNA strand and the macrocycle. **b)** Representative building block examples from the three diversity elements. **c)** The reported DNA-templated macrocycle library from the Liu lab.^[134] It contains four diversity elements, which generate high scaffold diversity. **d)** The recently reported library from the Neri group.^[137] The basic scaffold is decorated with three sets of diversity elements to generate deep sidechain variation.

Two sets of building blocks were incorporated within the ring scaffold for ring diversity enhancement (giving up to 2^{142} macrocycle scaffolds) and a third set was attached at the ring periphery to generate a high number of library members (approx. 1.4×10^6 members). The building blocks of the first diversity element were designed to mainly incorporate natural product-like moieties and comprised of alkenes, dienes and alkyl chains, but also aromatic, heteroaromatic and etheral moieties were introduced (**Figure 40b**).

The second diversity element was mainly derived from amino acids. We used *L*- and *D*- amino acids, along with several unnatural modifications thereof. A bunch of elements were specifically designed and synthesized to further incorporate natural product elements and to vary the size and rigidity of the macrocyclic ring.

The last diversifying element was a large series of terminal alkynes, which were attached after macrocyclization. This set diversified the library with a broad range of functional groups and structural elements, highly varying in size, shape and lipophilicity.

The reported macrocycle library from the Liu group (**Figure 40c**) showed a high diversity of the ring scaffold but contained mostly peptidic moieties. In comparison to our work, they did not include larger hydrophobic moieties based on pure carbon frameworks. Furthermore, the peripheric diversity of their macrocycles was much lower compared to other macrocycle libraries. The setup of the Neri library showed a constant peptide macrocycle scaffold with three diversification sites at the periphery (**Figure 40d**).

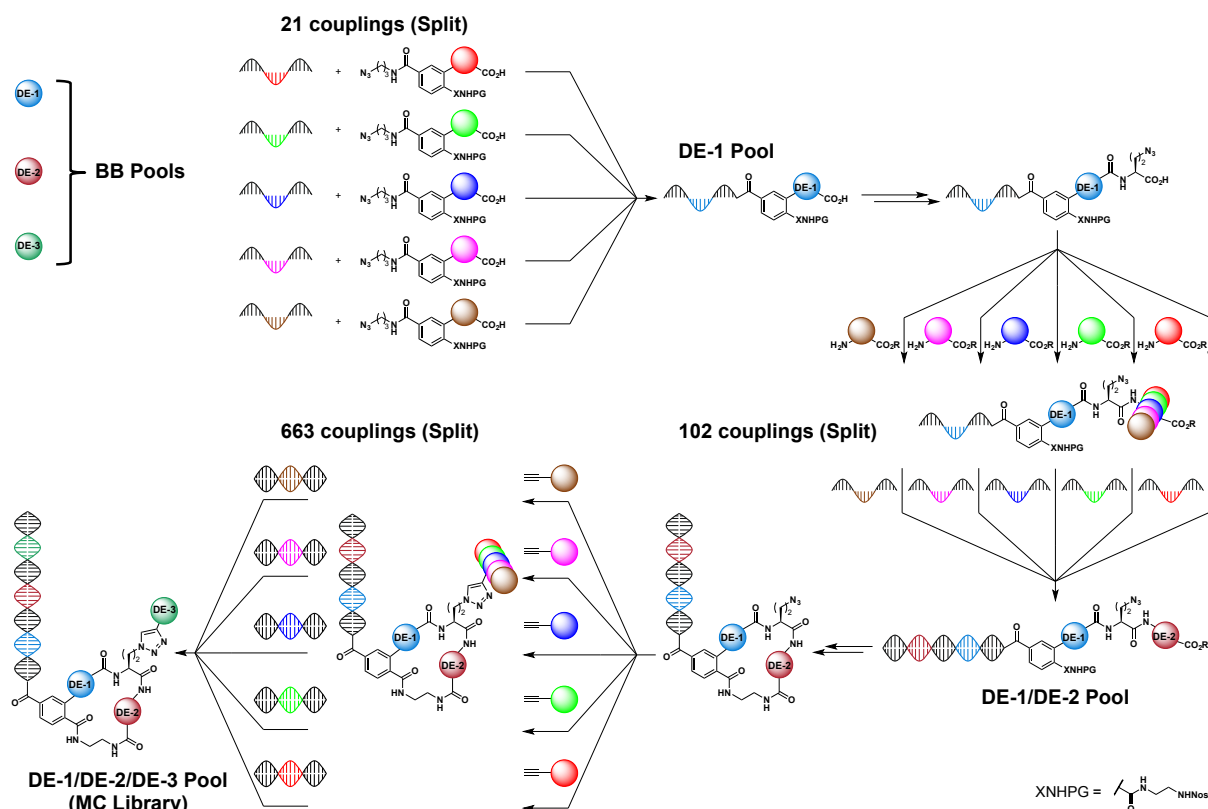


Figure 41. Split-Pool synthesis scheme. For combinatorial library assembly the compounds pool is divided into several vessels (split), coupled with the next building blocks, encoded with a DNA codon and finally pooled to get to the next step of the library construction. Three split-and-pool steps were necessary for the synthesis of the DEML, whereby the DE-1 introduction did not proceed in a parallel manner, as was it the case for DE-2 and DE-3 couplings, but in single steps with HPLC purification. Labeled spheres represent the complete building block pool of the corresponding diversity element. Unlabeled spheres represent single building blocks.

For the construction of our macrocycle library, we used the split-and-pool technique to synthesize our compound collection in a combinatorial manner.^[131] The first diversity elements (21 building blocks) were introduced off-DNA and were subsequently coupled to the DNA codon, and followed by HPLC purification. The collection of these encoded compounds was pooled and used for the simultaneous introduction of the trifunctional linker. For the modification with the amino acid sub-library, we split the DE-1 pool into 102 vessels, performed the reactions with all DE-2 building blocks, followed by encoding and pooling again (second split-pool event in **Figure 41**). This sub-library consisting of all DE-1/DE-2 combinations was cyclized to generate the first macrocycle sub-library (2'142 macrocycles). It was then divided into 663 vessels and reacted with the alkyne building blocks (DE-3). After encoding and pooling, we obtained the final macrocycle library comprised of 1'420'146 members.

We then tested our DEML against human serum albumin (HSA), α -1-acid glycoprotein (AGP) and carbonic anhydrase 9 (CA9). HSA is the most abundant protein in the human blood serum and carries mostly acidic ligands, which are otherwise insoluble in blood serum.^[184,185] Many commercial drugs (e.g. warfarin or ibuprofen) are transported by HSA^[186,187]. AGP is a second, less abundant transport protein in the human blood serum, which carries mainly neutral and basic (drug) molecules, such as the macrocyclic polymyxins (e.g. Colistin).^[188] These plasma proteins are of major importance for the development of new drug candidates, which might be bound and transported by those carrier proteins. CA9 is one of the isoforms of the zinc-containing transmembrane enzyme, carbonic anhydrase.^[189] It is involved in tumor acidification by hydration of carbon dioxide to bicarbonate and protons. Due to its very high catalytic activity, it is often used

as a marker for tumor hypoxia and acts as a prognostic factor for several human cancers.^[189] Owing to these properties, CA9 has gained a lot of attention in drug research.

The first version of the DNA-encoded macrocycle library (see next Chapter) was initially developed in a Sinergia project (Engineering the targeted drugs of the future: A general approach, number: 160699) in collaboration with the Schneider and Neri research groups from the ETH Zürich. From this work, two macrocycle libraries were published, each from the Neri^[137] and the Gillingham group.^[190] The Schneider research group focused on the development of peptide cytotoxins and assisted with *in-silico* predictions of potential drug candidates for selected proteins.

7. Assembly of a DNA-Encoded Macrocyclic Library

7.1 Chemical Library Designs

First Design of the DNA-Encoded Macrocyclic Library

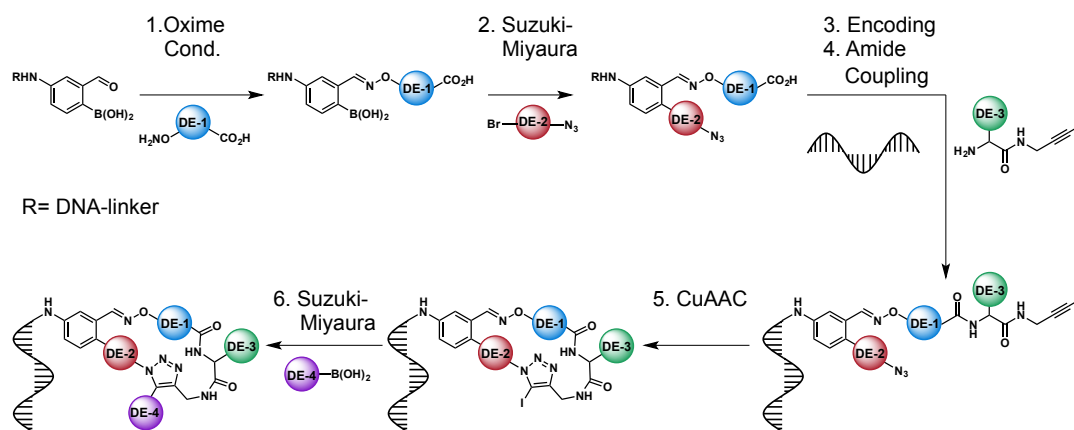


Figure 42. Synthetic scheme of the first library design. The initial scaffold was modified by boron-assisted oxime condensations, followed by Suzuki-Miyaura cross-coupling with a set of bromoazides. After DNA linkage, the third diversity element consisting of iodopropargyl amino acids was introduced, continued by copper-catalyzed click macrocyclization. The final diversification was achieved by Suzuki-Miyaura cross-coupling with a boronic acids sub-library. Detailed DNA encoding is omitted for clarity. The test macrocycle synthesis was performed off-DNA. Estimated library size: 4.3×10^7 members.

The first attempt to build a macrocycle library was based on the subsequent modification of a linker-modified 4-amino-2-formylphenylboronic acid subunit. Inspired by the very successful boron-assisted oxime condensation methodology (described in Chapter 2), we initially envisioned the synthesis of several oximes with a selection of hydroxylamines. Subsequent modification of the oxime boronates with a set of commercial or synthesized bromoazides in Suzuki-Miyaura reactions would deliver the azido-acids sub-library. **Figure 42** shows an overview of the library design. The azido-acids sub-library would be treated like a single diversity element since we anticipated synthesizing all members individually with separate purification steps. After DNA-encoding, the third diversity element consisting of iodopropargyl amino acids would be introduced by amide couplings, followed by CuAAC-mediated macrocyclization and a final diversification step with a large set of boronic acids. Given the high synthetic effort to generate the azido-acids and the modified amino acids sub-libraries, we sought to end up with a forty- to fifty-million macrocycle library.

The actual synthesis of this library turned out to be tricky. We tested the chemistry without the introduction of any DNA strands to be able to use all analytical methods for small molecule characterization and we focused on reaction conditions that were potentially compatible to on-DNA synthesis. Oxime formation (with aminoxyacetic acid) showed to be the simplest step of all and the reaction proceeded with full conversion. The Suzuki-Miyaura cross-coupling revealed a high intolerance towards the oxime. Only very low quantities of the desired azido-acid product could be generated, with mostly deborylation products found. A swap in assembling order yielded a much higher amount of product. The Suzuki-Miyaura reaction of the azido-bromide (4-azidophenylbromide was used as test substrate) with the aldehyde subunit yielded >60% product. Oxime generation was then achieved by acid-catalyzed oxime condensation with near quantitative yields. Due to the swap of the reaction order, the advantageous boronic acid-assisted oxime condensation was not possible anymore. In the fourth step (the DNA encoding was skipped), we successfully introduced iodopropargyl-modified *L*-alanine in high yield (>70%).

whereby the amino acid was modified beforehand according to a literature procedure.^[191] The fifth step of the trial macrocycle assembly turned out to be the final step of the synthesis. Even though reports of very successful copper-catalyzed iodoalkyne-azide click reactions were published,^[192] we could not achieve the desired macrocyclization along with a conserved iodo-triazole moiety. The main side products found, were alkyne and triazole deiodinations. Linear, dimeric structures with intact iodotriazoles could also be identified. Swapping the reaction order, like in the case of the first two steps, was not possible either, because the click reaction between the azido-acid and the alkyne-amino acid yielded very low amounts of the target iodotriazole.

Second Design of the DNA-Encoded Macrocycle Library

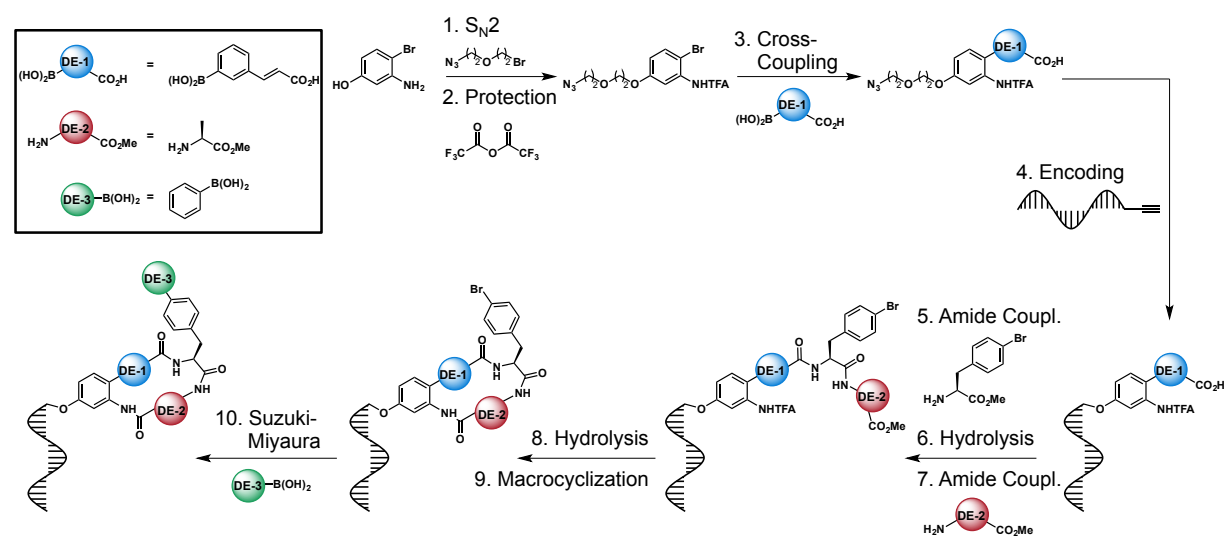


Figure 43. Synthetic strategy of the second macrocycle library attempt. The building blocks displayed in the box were used for strategy validation on-DNA and off-DNA. The strategy included three amide coupling steps, two cross-couplings, one CuAAC and one S_N2 reaction along with protection-deprotection chemistry. The exact DNA-small molecule linkage and the detailed enzymatic encoding are omitted for clarity.

Due to the synthetic difficulties of the first library strategy, we developed a new library assembly procedure. We only included chemical transformations that were known for DECL compatibility, along with readily available building block collections. The strategy is outlined in **Figure 43**. We started with 2-bromo-5-hydroxyaniline as modular scaffold with three diversification sites. The hydroxy group was used for the attachment of an azido linker by nucleophilic substitution for the later connection of the small molecule to the DNA strand. Under common S_N2 reaction conditions at elevated temperatures, the substitution selectively occurred with the aromatic alcohol. For the protection of the aromatic amine, we considered TFA as suitable for our needs due to the simple installment with trifluoroacetic anhydride and its stability during library assembly.^[193]

The first diversity element was introduced by Suzuki-Miyaura cross-coupling in an aqueous environment. We used 3-(2-carboxyvinyl)phenylboronic acid (see the black box in **Figure 43**) as building block for the strategy validation. The encoding step was skipped in the first attempt to prove the suitability of the chemistry off-DNA. For a trifunctional linker, we chose 4-bromophenylalanine, which connected the two ring diversity elements and provided the coupling site for the final peripheral diversification. We also considered 3-bromo-5-aminobenzoic acid as a potential linker, which would keep the third diversity element closer to the ring scaffold, but this molecule turned out to be very inefficient in standard amide couplings. Subsequent ester hydrolysis was performed using lithium hydroxide. It was important to use a final concentration of 5 mM LiOH to ensure efficient ester hydrolysis, while keeping the TFA protecting group untouched. This procedure also worked at very dilute concentrations (< 1 mM). We continued the

test macrocycle assembly by the introduction of *L*-alanine under standard amide coupling conditions. The following deprotection step (ester and trifluoroacetamide) was achieved by an increased LiOH concentration (50 mM) at slightly elevated temperatures (45°C). The macrocyclization step again turned out to be the bottleneck of this synthetic approach. We tested several conditions by exchanging coupling reagents and solvents/bases to achieve macrocyclization, but at best, we found 8% conversion of starting material to the desired macrocycle. We tried to determine whether the geometry or flexibility of the macrocycle was unfavorable for the cyclization to proceed. For this purpose, we attached glycine after the second diversity element, which should act as a very flexible three-atom spacer. While the attachment of the glycine was successfully achieved, the macrocyclization failed again. Incorporation of a three-atom spacer, containing a more reactive primary alkylamine (2-bromoethylamine), by nucleophilic substitution with the aniline nitrogen was achieved in very low amounts.

The described synthesis was repeated on-DNA as well, which was successfully conducted up to the macrocyclization step (amide coupling conditions were adopted from literature procedures).^[147] We obtained the desired macrocycles neither on-DNA nor as small molecule. With the linear on-DNA macrocycle precursor, we also tested Suzuki-Miyaura cross-couplings to validate this diversification step under DNA-compatible conditions. The desired cross-coupled product was not clearly identified, but we saw a strong peak broadening, which was a hint that potential DNA damage had occurred.

Concluding, we realized that the presented strategy was in principle suitable for synthesis on DNA and the amide and protection/deprotection steps worked with the developed conditions. The reactivity of the aniline moiety was the limiting factor for the synthesis. Alternative strategies with acyl chlorides, that should be more reactive towards amide bond formation with anilines, were not tested since acyl chlorides are incompatible with the aqueous conditions required for on-DNA synthesis.

Final Design of the DNA-Encoded Macrocycle Library

Based on the results from the design of the second library, we developed a new but similar strategy to achieve successful macrocyclization. We excluded the set of boronic acids as DE-3 building blocks due to difficult transformations and the potential for DNA damage by the palladium catalysts. We instead considered copper-catalyzed click reactions (CuAAC) with terminal alkynes more suitable for DE-3 incorporation. These reactions are routinely used in biological environments since they work at low concentrations with high efficiencies.^[97,194–196] The development of a new central scaffold with a more reactive amine for the macrocyclization step was the main focus of this new approach. We considered 2-bromo-5-hydroxybenzaldehyde (**A**) as a valuable candidate for the generation of the natural product-like DEML. The aldehyde gives the opportunity to incorporate chemically interesting moieties via a set of different chemical transformations. Most importantly, Wittig-type reactions were selected for the generation of polyolefin structures. Unfortunately, modifications of this compound turned out to be tedious and impractical for the synthesis of a larger set of scaffold-DE-1 blocks.

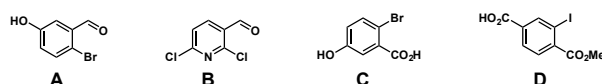


Figure 44. Four commercially available central scaffold units that were evaluated for macrocycle library synthesis.

B could be modified with 3-azidopropanol as the DNA linker in a nucleophilic aromatic substitution, but the aldehyde needed acetal protection. The isolated product was a mixture of *para* (desired) and *ortho* (undesired) substitution to the aldehyde in a 2:1 ratio. The most similar approach to the second library design (previous section) was evaluated with **C**. We replaced the

amine with a carboxylic acid that was further modified with Nosyl-protected ethylenediamine (**2-3**). The limiting step in this approach was the introduction of a linker for DNA attachment (we used the same linker from step 1 in **Figure 43**). The hydroxy group and the sulfonamide of the Nosyl protecting group were both modified by nucleophilic substitution. This led to the development of a strategy, based on a 2-iodoterephthalic acid core structure (**D**). 3-Azidopropylamine **2-1** was coupled with the terephthalate derivative **D** and further modification with the Nosyl-protected diamine **2-3** yielded the basic structure for the assembly of the first building block set (**2-4** in **Figure 45**). The diversification element was introduced by cross-coupling reactions (see Chapter 7.4 for further details).

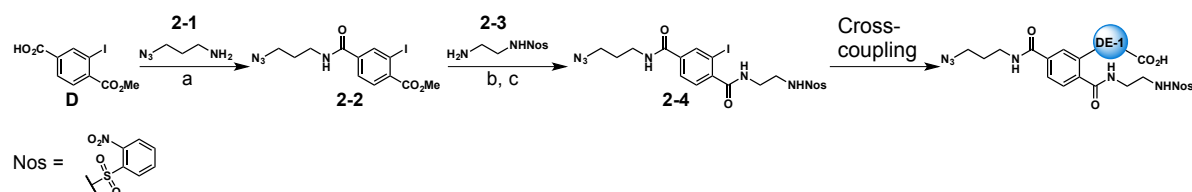


Figure 45. Synthetic scheme for the synthesis of the diversity elements 1 sub-library. **a)** HATU, DIPEA, THF/DMF, RT, 2 h, **95%**. **b)** LiOH, H₂O/MeCN 1:1, RT, 4 h, **98%**. **c)** HATU, DIPEA, DMF, RT, 3 h, **92%**.

We synthesized a set of 21 building blocks, which were further used for the assembly of the DNA-encoded macrocycle library. A detailed synthetic outline is shown in **Figure 46**. The first set of building blocks was encoded with a short DNA strand, pooled and modified with the trifunctional linker **2-5**. The second diversity element (102 amino acid building blocks) was introduced in a split synthesis step by amide coupling (compare **Figure 41** for details about the split- and pool synthetic steps). These building blocks were encoded and combined to yield the DE-1/DE-2 pool. After full deprotection of the ester and the Nosyl group, the DE-1/DE-2 pool was macrocyclized to yield the DEML sub-library (smDEML), which was modified by click reaction with 663 terminal alkyne building blocks in a split synthesis step. Encoding and pooling of the vessels yielded the final encoded macrocycle library. Reaction details and optimizations are discussed in Chapter 7.4.

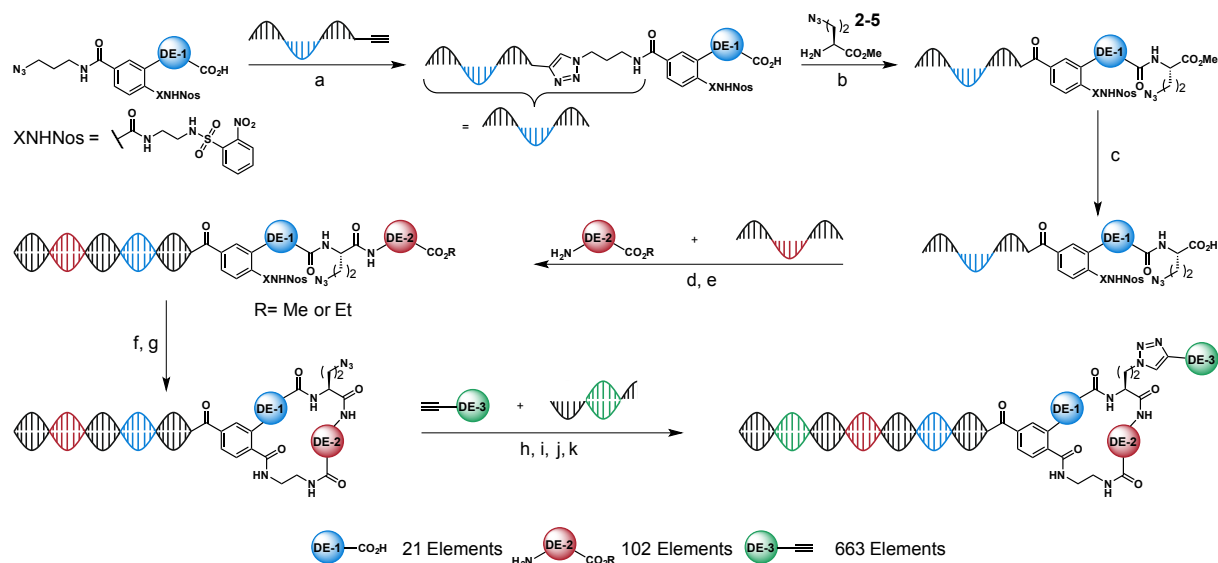


Figure 46. DEML synthesis overview. The individual steps are given with optimized reaction conditions. Split- and pool steps are not especially marked. **a)** Cu-TBTA, NaOAsc, TEAA buffer pH 7.2, DMSO, RT, 20 h. **b)** DMTMM-BF₄, NMM, MOPS buffer pH 8.2, DMSO, RT, 24 h. **c)** LiOH, H₂O/MeCN 3:1, RT, 2.5 h. **d)** DMTMM-BF₄, NMM, MOPS buffer pH 8.2, DMSO, RT, 20 h or EDC·HCl, HOAt, DIPEA, MOPS buffer pH 8.2, DMSO, RT, 20 h. **e)** Klenow Polymerase, NEBuffer 2, dNTPs, 25°C, 30 min. **f)** 2-Mercaptoethanol, DBU, MOPS buffer pH 8.2, DMSO, RT, 18 h. **h)** BamHI-HF, CutSmart buffer, 37°C, 30 min. **i)** Cu-TBTA, NaOAsc, TEAA buffer pH 7.2, DMSO, RT, 16 h. **j)** T4 DNA ligase, NEB ligase buffer, ATP, 16°C, 16 h. **k)** Klenow Polymerase, NEB ligase Buffer, dNTPs, 25°C, 30 min.

The enzymatic encoding of our library was the key for successful evaluation of potent protein binders. Every building block from each diversity element was encoded by a unique DNA coding sequence (codon). We decided to encode our library by a Klenow polymerization-restriction-ligation methodology. Since the first diversity elements were encoded individually and then pooled, the schematic representation of the encoding in **Figure 47** starts after the introduction of the DE-2 elements. A single-stranded DNA with the according DE-2 codon was annealed to the DNA strands of the former DE-1 pool. The formed partially double-stranded DNA was filled in with the Klenow fragment polymerase to yield the coding DNA in duplex format. After the macrocyclization step, the duplex DNA was cut at two specific sites (restriction) on opposite sides of the DNA duplex to yield a 5'- four base overhang (sticky end). This overhang was used for encoding of the last chemical step. We introduced a pre-annealed, partially double-stranded DNA fragment by T4 ligation using the sticky ends for annealing. The final stage consisted of another Klenow fill-in step to generate a DEML with a fully duplex oligonucleotide strand. The presented encoding format has been applied before for the generation of DNA-recorded chemical libraries.^[138]

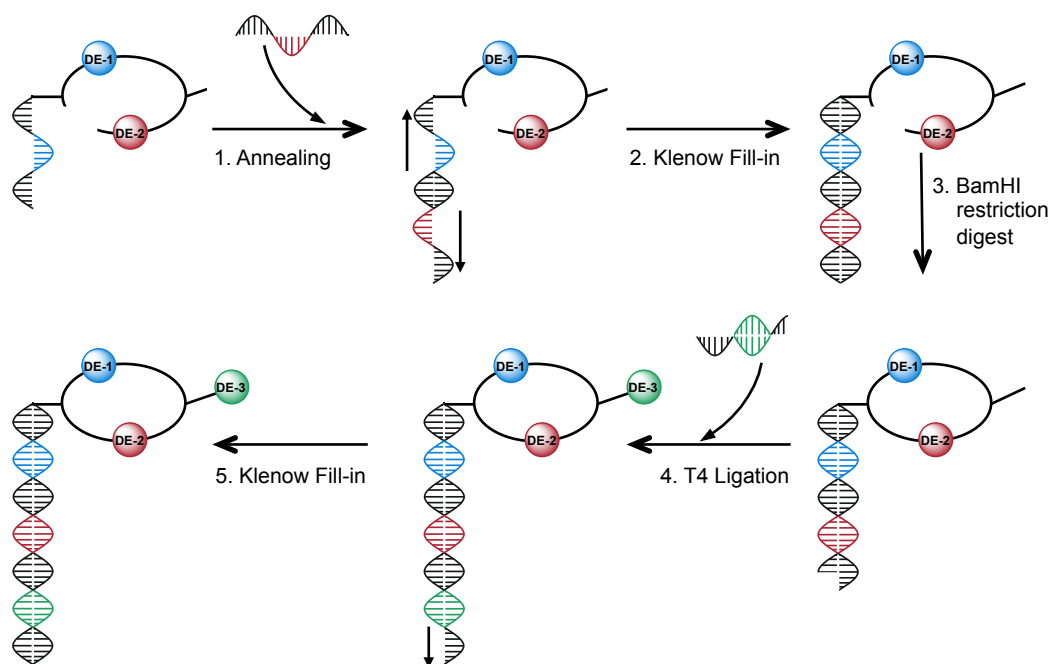


Figure 47. Schematic representation of the library encoding. The scheme was adapted from the overview **Figure 46**. Chemical transformation steps are omitted for clarity. A Klenow polymerization-restriction-ligation procedure was used for library encoding. The shown encoding steps start at the second split synthesis step (amino acids coupling).

7.2 Library Validation off-DNA

Before we started the library synthesis with DNA-encoding, we synthesized a model macrocycle to make sure that our chemical strategy was valuable. We used 3-(2-carboxyvinyl)phenylboronic acid in a Suzuki-Miyaura cross-coupling to generate **NP01**, which was further modified by the trifunctional linker **2-5** (**Figure 48**). After basic hydrolysis of **2-6**, the free acid **2-7** was modified with *L*-alanine methyl ester (**AA001**) to obtain the linear macrocycle precursor **2-8**. This compound was transformed to its free "amino-acid" form **2-9**, which was used for subsequent macrocyclization, followed by peripheral modification with ethyl propiolate to the final macrocycle **2-11**. Since macrocycle **2-10** contained two azide groups at different positions, the final macrocycle contained two triazole units. In the on-DNA synthesis, the propylazide linker will be used for DNA attachment prior to incorporation of the trifunctional azide linker **2-5**. The success of this synthesis with acceptable yields (>60% per step) laid the foundation for further development of the DEML.

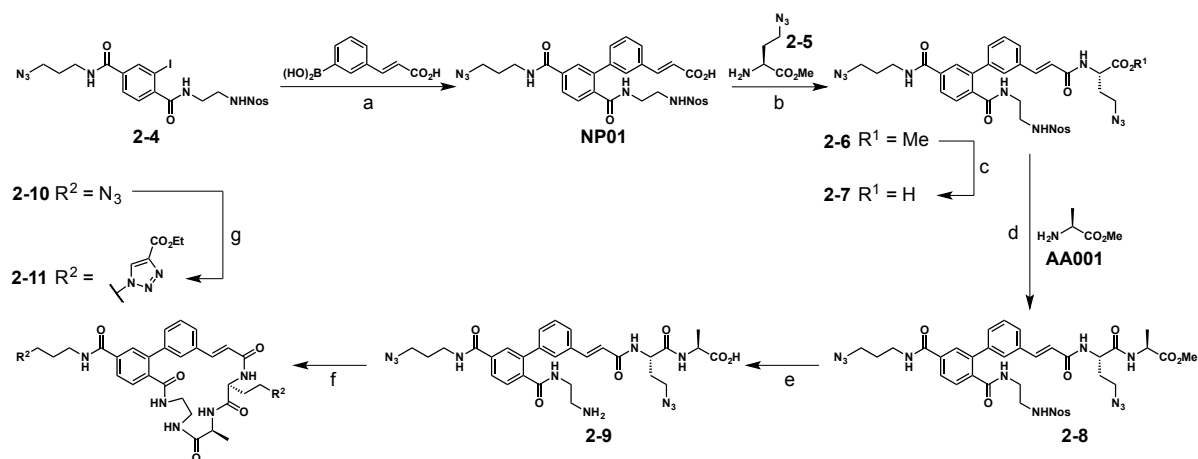


Figure 48. Synthetic scheme of the off-DNA macrocycle synthesis. **a**) Pd (π -cinnamyl) chloride dimer, K_3PO_4 , $\text{EtOH}/\text{H}_2\text{O}$ 2:1, 50°C , 1 h, **63%**. **b**) HATU, DIPEA, DMF, RT, 1 h, **75%**. **c**) LiOH, $\text{MeCN}/\text{H}_2\text{O}$ 1:1, RT, 2 h, **74%**. **d**) HATU, DIPEA, DMF, RT, 2 h, **62%**. **e**) PhSH, DIPEA, DMF, 40°C , 3 h, then LiOH, H_2O , 30 min, RT, **78%**. **f**) HATU, DIPEA, DMF, RT, 1 h, **63%**. **g**) Ethyl propiolate, Cu-TBTA, NaOAsc, DMSO, H_2O , MeCN, RT, 2 h, **68%**.

7.3 Library Validation on-DNA

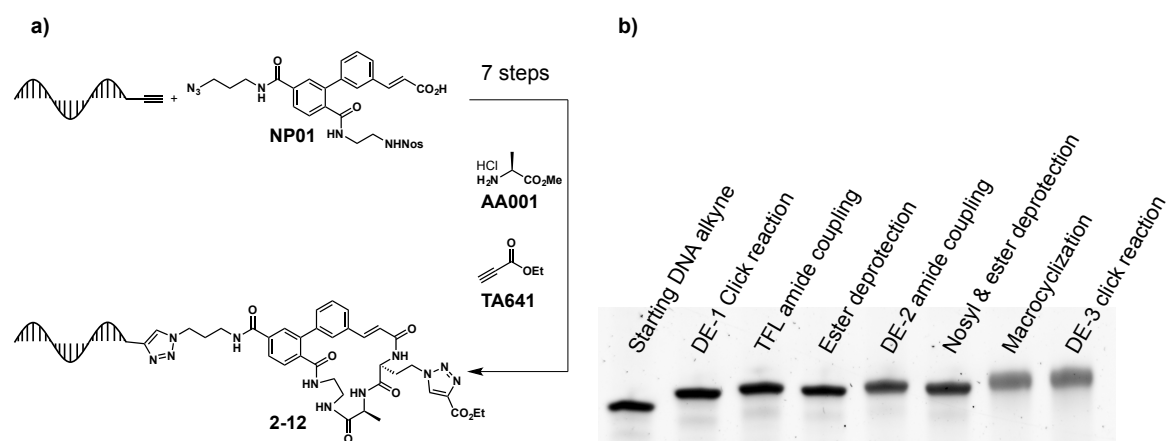


Figure 49. **a**) Schematic representation of the macrocycle synthesis on DNA in seven steps. Encoding of DE-2 and DE-3 was not performed. Reaction steps and conditions are outlined in detail in **Figure 46**. **b**) Denaturing PAGE gel of the synthetic steps during on-DNA macrocycle synthesis validation.

After successful synthesis of the small-molecule macrocycle, we also validated and optimized the strategy in the presence of a DNA strand. **Figure 49a** shows the simplified synthetic procedure in which we started from an alkyne-modified DNA strand to synthesize the final encoded macrocycle **2-12** in seven synthetic steps (for a detailed synthetic outline, see **Figure 46**). We included the diversity elements **NP01**, **AA001** and **TA641**, but skipped the encoding of these elements for simplicity. At a later stage, we saw that the encoding steps and the chemical transformations did not affect each other. The effect of the chemical synthesis is shown in **Figure 49b**. The mass increase (or decrease) during the synthesis along with the temporary generation of ionizable groups was observed by slight band shifts over the course of the synthesis. The last two bands with the attached macrocycles are less pronounced compared to the linear precursors. Since the assembly contains a big hydrophilic DNA part with a partially polar, but rather hydrophobic small-molecule attached, we assumed that the exceptional physicochemical properties of macrocycles cause interactions with the DNA. Polyacrylamide effects on the DNA-macrocycle system and/or accumulated DNA damage might also cause some blurriness of the DNA band.

7.4 Assembly of the Macrocyclic Library

Diversity Element 1 Sub-Library

The most challenging part for the development of a natural product-like macrocycle library was the incorporation of typical elements, such as polyketides, polyolefins or heteroaromatic building blocks. The first set of diversity elements was the ideal position for mostly carbon framework building blocks. The members of this diversity element had to be created individually from the common scaffold **2-4**.

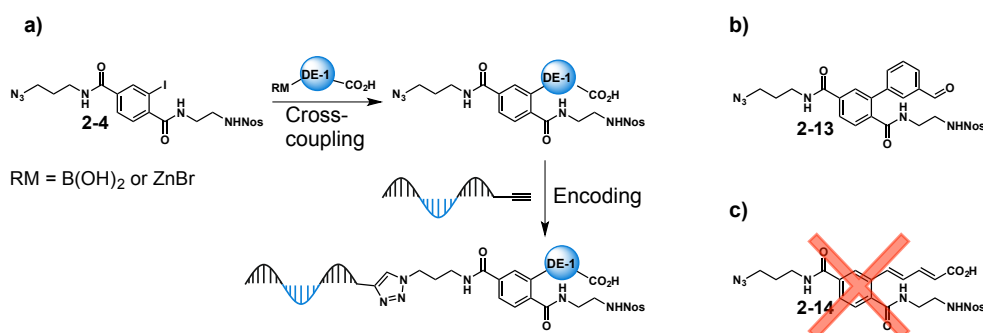


Figure 50. a) General assembly of the DE-1 elements with subsequent DNA encoding. b) The aldehyde intermediate for the synthesis of polyolefin building blocks. c) Desired structural elements, which were not accessible.

Palladium-catalyzed cross-coupling reactions such as the Suzuki-Miyaura and Negishi couplings turned out to be ideal for this task (**Figure 50**). We found a set of commercially available boronic acids and zinc reagents that were valuable for the construction of these building blocks. Due to the high synthetic effort to create such elements, we limited the number of elements to about 20 (in the end 21!). With an operationally simple procedure for the Suzuki cross-coupling, we could rapidly create nine building blocks from commercial sources with substituted aromatic and heteroaromatic moieties (**NP01-NP09**, **Table 5**). Furthermore, commercially available zinc reagents allowed the incorporation of alkyl moieties and a furanyl group (**NP12-NP16**). All cross-couplings were carried out with the palladium (π -cinnamyl) chloride dimer catalyst. Due to the presence of an azido group, the use of phosphine-based catalysts was not possible because of potential Staudinger reduction. The employed catalyst showed good compatibility with the functional groups, no need of highly inert reaction conditions and it yielded enough material (typically 50-90% isolated yield) to proceed with the library synthesis.

For the generation of polyolefin moieties (**NP17-NP21**) or non-commercial *cis*-olefins (**NP11**), we used Wittig-type reactions (Horner-Wadsworth-Emmons reaction and Still-Gennari variant) in combination with the successfully obtained aldehyde **2-13**. The majority of the shown building blocks were comprised of biaryl structures. The direct incorporation of polyolefins without the need for biaryl moieties (**Figure 50c**) would have been very desirable, but we could not achieve this goal. The direct coupling of olefinic boronic acids, carbonylation or Negishi couplings with an acetal-protected aldehyde zinc reagent with precursor **2-4** for subsequent Wittig-olefination were unsuccessful. Neither the generation of a Grignard species with subsequent quenching with aldehyde precursors, such as the Vilsmeier reagent, DMF, paraformaldehyde or triethylorthoformate, nor the introduction of a vinyl group for olefin metathesis were successful. All the mentioned attempts were repeated with precursors of **2-4** or with a different protecting scheme, however, the desired products could not be obtained.

The difficulty of the described carbonyl-introducing reactions became clear after the synthesis of the benzylic alcohol **2-16** (**Figure 51**). Lactone **2-15** was opened with nosylethylenediamine **2-3** and in a further oxidation step with common mild oxidizing agents, the desired aldehyde **2-17** was

observed. However, the aldehyde underwent a further cyclization step with the sulfonamide group (**2-18**). This cyclic hemiaminal could neither be transformed back to the aldehyde nor directly used in Wittig-type reactions.



Figure 51. Aldehyde synthesis from lactone **2-15**. Reaction conditions: **a)** K_2CO_3 , THF, 65°C, 23 h, **62%**. **b)** DMP or IBX, (additives), THF, RT, 15 min, **0%**.

Due to these difficulties, we continued with the shown building blocks in **Table 5**. New strategies for the incorporation of polyolefin moieties without the need of an aryl core would be highly desirable. Furthermore, polyketide moieties should be taken into account, even though the synthetic effort for this complicated building block type will limit the number of obtainable compounds to only a few members.

Table 5. Structures of the set of diversity element 1 building blocks. R indicates the constant terephthalate scaffold.

No.	Structure	No.	Structure	No.	Structure
NP01 ^[a]		NP08 ^[a]		NP15 ^[b]	
NP02 ^[a]		NP09 ^[a]		NP16 ^[b]	
NP03 ^[a]		NP10		NP17	
NP04 ^[a]		NP11		NP18	
NP05 ^[a]		NP12 ^[b]		NP19	
NP06 ^[a]		NP13 ^[b]		NP20	
NP07 ^[a]		NP14 ^[b]		NP21	

[a] Commercially available as boronic acid derivatives. [b] These elements were synthesized by Negishi cross-coupling with the corresponding commercial zinc reagents.

For the generation of the DE-1 sub-library, all building blocks depicted in **Table 5** were coupled by the copper-catalyzed click reaction (CuAAC) to the corresponding 5'-alkyne modified 39-mer oligonucleotides. The DNA tags were comprised of two primer annealing sites (for all building blocks the same) and individual three base codons in the middle of the strands. These transformations were carried out with high yields (approx. 80%), even though some part of the DNA was always found to be unreacted, originating from alkyne oxidations during DNA synthesis. The DNA side products were successfully separated from the DE-1 DNAs by reverse-phase preparative HPLC purification.

Trifunctional Linker Synthesis and Coupling with the DEML

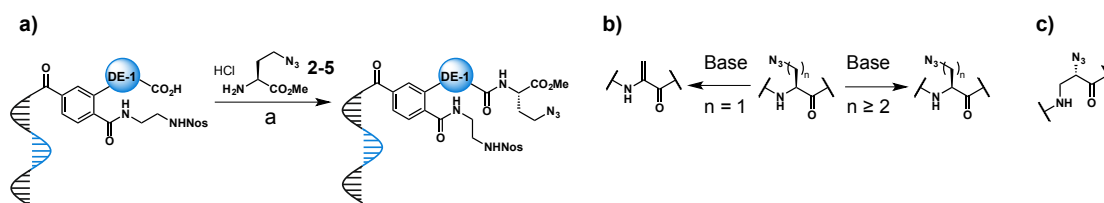


Figure 52. a) Scheme for the trifunctional linker coupling with the DE-1 sub-library pool. Reaction conditions: a) DMTMM-BF₄, NMM, MOPS buffer pH 8.2, DMSO, RT, 24 h. b) The influence of methylene spacers in β - and γ -azido amino acid linkers under basic conditions. c) Structure of the unstable α -azido- β -amino acid.

The incorporation of a trifunctional linker (TFL) into the library was a very important task because it connects the three diversity elements. The TFL strongly influences the final structure of the macrocycle, and it may influence the interplay between the diversity elements. While the geometry of the macrocyclic scaffold is largely dependent on the structures of DE-1 and DE-2, the trifunctional linker defines the proximity of the peripheral diversification site. Depending on the interaction mode of a macrocycle with a certain protein, it may be beneficial to keep the overall structure compact and get the building blocks in close relation with each other, so that the macrocycle may fit into the binding pocket of the protein. Alternatively, large peripheral groups might improve binding to several locations on the protein surface, which also generates high affinity. We decided to keep the peripheral modification close to the macrocyclic scaffold.

At first, we tried to synthesize a β -amino acid with an azido group attached to the α -carbon (**Figure 52c**). This compound would generate the shortest possible triazole linkage to the DE-3 elements. The synthesis of this compound turned out to be possible up to its Fmoc-protected precursor. Upon protecting group removal, the desired compound could be detected in solution, but it rapidly decomposed upon isolation.

In further attempts, we turned to azido-modified α -amino acids (**Figure 52b**), which are commercially available or could be synthesized in a one-step procedure by copper-catalyzed amine-azide transfer from commercial precursors. The simplest representative is *L*-azido-alanine ($n=1$), which we could successfully introduce into DNA-encoded macrocycle precursors. While amide couplings with this azido-amino acid worked relatively well, treatment with strong basic conditions (such as those used for ester hydrolysis) led to a substantial amount of azide elimination to form dehydroalanine (see **Figure 52b**). This elimination product cannot be used for the incorporation of the third diversity element, but dehydroalanine is an interesting building block for future encoded-libraries. Finally, we ended up using a trifunctional linker with two methylene spacers (**2-5**, $n=2$), which was resistant to azide elimination.

The amide coupling with encoded DE-1 elements was found to be possible with EDC/HOAt or DMTMM-BF₄ as coupling reagents with similarly high efficiencies. For library construction with the complete pool, we decided to use the DMTMM conditions because we found a decreased amount of DNA-damage under these conditions (see Chapter 8.3 for qPCR experiments). Test assays with eight representative DE-1 building blocks showed full conversion to the desired product, except for **NP07**. We also carried out LC-MS analysis of the coupled pool (20 BBs) and identified all coupling products while no traces of the starting materials were observed. **NP07** was treated twice with the coupling solution to achieve full conversion and was then combined with the rest of the DE-1 pool.

Ester Deprotection

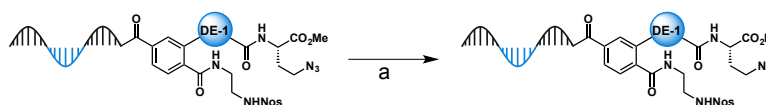


Figure 53. Ester deprotection of the library sub-pool. Reaction conditions: **a)** LiOH monohydrate, H₂O/MeCN 2:1, RT, 2.5 h.

For the hydrolysis of the ester, lithium hydroxide was used in the presence of about 30% organic solvent (usually acetonitrile). Unlike in amide coupling, where at least 1000 equivalents of the reagents were necessary, ester hydrolysis always proceeded within a few hours with only 100-150 equivalents of LiOH. Except for elimination reactions with azido-alanine, we never came across any undesired side reactions in this deprotection step.

DE-2 Screening and Coupling with the DE-1 Pool

The second diversity element consisted of a collection of amino acids. We included *L*- and *D*-amino acids, modified amino acids, unnatural amino acids with pure alkyl sidechains, but also with various functional groups (hydroxy, pyridine, thiazole, quinoline, sulfoxide). We also selected cyclic, spirocyclic, bridged and strained amino acids to introduce some rigidity into the macrocycle scaffold. The length and sizes of the building blocks were very diverse, ranging from small glycine up to a very long mini-PEG or the very bulky adamantyl-containing amino acid. High commercial availability along with easy synthetic procedures rendered this diversity element ideal to expand the diversity of our ring scaffold. We synthesized several olefinic (alkene and diene) amino acids to expand the number of natural product-like building blocks in our library.

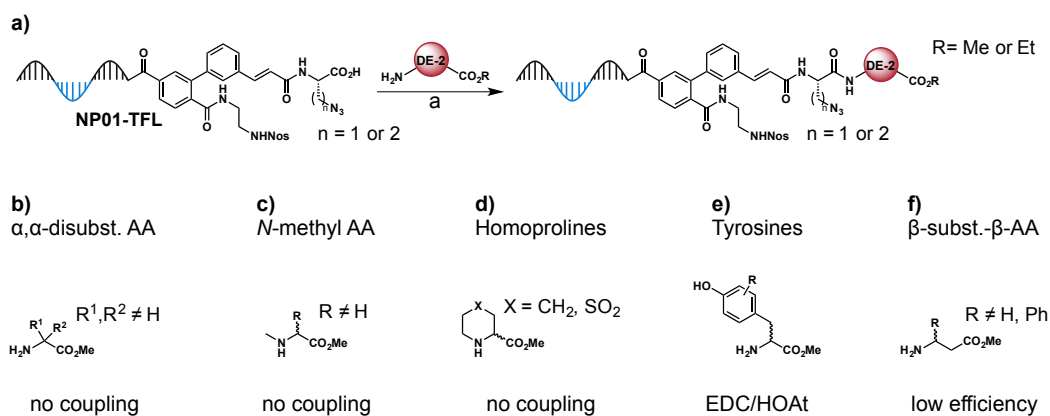


Figure 54. **a)** Schematic representation of the DE-2 test assay with 126 amino acids. We used trifunctional linkers with one or two methylene spacers. Reaction conditions: **a)** DMTMM-BF₄, NMM, MOPS buffer, DMSO, RT, 24 h or EDC·HCl, HOAt, DIPEA, MOPS buffer, DMSO, RT, 24 h. **b)** α,α-disubstituted amino acids showed no coupling. **c)** N-methylated amino acids showed no coupling. **d)** Homoprolines showed no coupling. **e)** Tyrosine and its derivatives showed good couplings with EDC/HOAt conditions. **f)** β-substituted β-amino acids showed very low efficiencies with non-aromatic sidechains.

Prior to the introduction of the second diversity element, we screened our collection of 126 amino acids to explore the reactivity of these compounds towards amide coupling with the DE-1-TFL pool. For this purpose, we took a model precursor (**NP01-TFL**), screened it against all amino acids (DMTMM coupling conditions) and analyzed the results with LC-MS detection. In the case of unsuccessful couplings, we repeated the reaction under EDC/HOAt conditions to find potential improvements in coupling efficiency. We set the limit of usefulness for library construction at >80% conversion of the starting material and ≥50% product purity.

We found several trends in reactivity depending on the amino acid structure. In general, α,α -disubstituted amino acids (**Figure 54b**) did not couple (**AA093** in combination with the EDC/HOAt/DIPEA system was the exception) and neither did *N*-methylated amino acids (**Figure 54c**). Sarcosine (*N*-methylglycine) was the only exception. Cyclic amino acids also showed a reactivity trend depending on the nature of the ring scaffold. The 4-membered ring amino acid **AA094**, as well as proline derivatives, worked well under the tested reaction conditions, whereas homoprolines (**Figure 54d**) did not work at all. Interestingly, β - and γ -homoprolines yielded a good amount of the desired product, which led to the conclusion that the arrangement of the functional groups in cyclic amino acids has a big influence on their reactivity. A very important, but difficult, group of compounds were linear β -amino acids. In the case of α -substituted β -amino acids, the couplings proceeded with good conversions and purities, whereas with β -substituted β -amino acids (**Figure 54f**), only the members with flat, aromatic sidechains gave acceptable yields. The special case amino acids with electron-withdrawing substituents, good leaving groups, a hydroxylamine or a thiazoline ring scaffold generally showed no conversion to the desired products at all. Due to the reactivity of aromatic alcohols with DMTMM-BF₄, tyrosine derivatives (**Figure 54e**) were only successfully coupled under EDC/HOAt/DIPEA conditions. Apart from the previously mentioned dehydroalanine side product (we included successful coupling to this moiety in the coupling efficiency), we also identified the two major side products that we observed during the test assay by LC-MS (**Figure 55**). The complete list of amino acids with coupling efficiency results are found in **Table 12** in the experimental. A more detailed discussion about the occurrence of the identified side products will be held in the section about the Nosyl deprotection (Page 90).

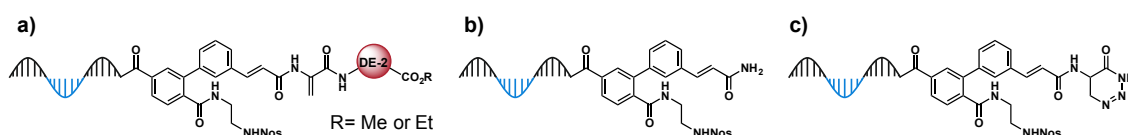


Figure 55. Proposed structure of found common side products. **a)** Elimination product to the dehydroalanine species. **b)** Terminal amide product. **c)** Dihydrotriazinone product.

The results from this screening were used for the assembly of the final library. Taking into account our cut-off limits (>80% conversion, \geq 50% product purity), we included 102 amino acids for the actual synthesis. This was performed in a split procedure, whereby the DE-1-TFL pool was split into 102 vessels and every vessel was reacted with one amino acid under the elaborated coupling conditions.

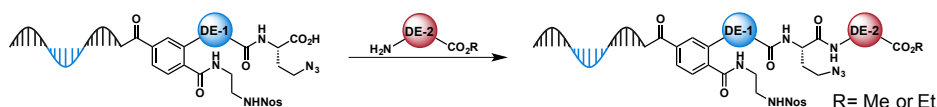


Figure 56. DE-2 coupling of the DE-1-TFL pool with the selected 102 amino acids. Coupling conditions matched the conditions described in **Figure 54a**.

Encoding of the Second Diversity Elements

After the introduction of the DE-2 sub-library, we encoded the new building blocks with a 40-mer DNA strand (**Figure 57a**). This DNA strand consisted of an 18-base annealing site, complementary to the DE-1 coding DNA, the four bases codon for the specific DE-2 building block and the BamHI restriction site 5'-GGATCC-3'.

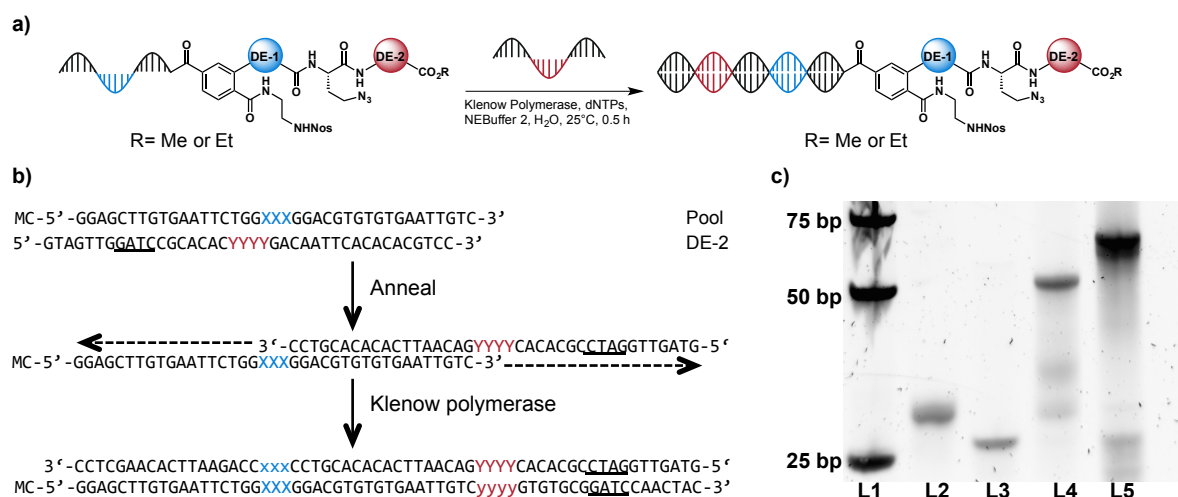


Figure 57. **a)** Schematic representation of the DE-2 encoding. **b)** Exact sequences for the annealing and Klenow fill-in during encoding. **c)** 12% DNA polyacrylamide gel, SYBR gold staining. **L1:** Low MW DNA ladder, **L2:** DNA-DE-1 pool, **L3:** DE-2 coding DNA strand, **L4:** Annealing of DNA-DE-1 pool and DE-2 coding DNA strand, **L5:** Klenow extension.

The exact DNA sequence for the encoding of the amino acids is shown in **Figure 57b**. The codons are represented as XXX (for DE-1) and YYYYY (for DE-2). Underlined bases indicate the BamHI restriction site. Annealing of the coding DNA created an 18-base duplex strand with long 5' overhangs. The large Klenow fragment subsequently polymerized the two strands to create a 61-base pair long duplex DNA. During this procedure, the reverse complementary codons of the DE-1 and DE-2 building blocks were incorporated into the complementary oligonucleotide strands. The conditions for the encoding step were adapted from the manufacturer's published procedure. We did not see any undesired side product formation, nor the truncation of DNA strands due to exonuclease activity of the Klenow fragment. The detailed annealing and encoding results are presented in the gel picture (**Figure 57c**). After the encoding of all building blocks, the vessels were combined (pooled) and used for further library assembly.

Nosyl and Ester Deprotection Screening and Application with the DE-1/DE-2 Pool

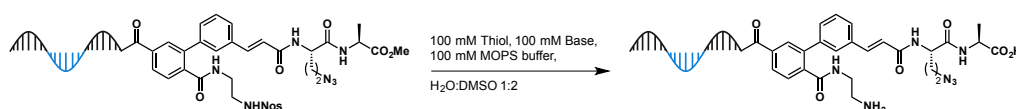


Table 6. Nosyl deprotection screening and optimization.

Entry	Thiol	Base	Temperature	Time	Nosyl Deprotection	Ester Hydrolysis
1	PhSH	DIPEA	60°C	6 h	Yes	No
2	BME	DBU	60°C	6 h	Yes	Yes
3	MPAA	DIPEA	60°C	6 h	Incomplete	No
4	NaTG ^[a]	DIPEA	60°C	6 h	No	No
5	PhSH	DBU	60°C	2 h	Incomplete	Incomplete
6	PhSH	DBU	60°C	4 h	Yes	Yes
7	PhSH	DBU	60°C	6 h	Yes	Yes
8	BME	DBU	RT	2 h	Yes	Yes
9	BME	DBU	RT	4 h	Yes	Yes
10	BME	DBU	RT	6 h	Yes	Yes
11 ^[b]	MPAA	DBU	60°C	6 h	Incomplete	Incomplete
12	BME	DIPEA	RT	6 h	Incomplete	No
13	BME	NMM	RT	6 h	No	No

[a] Stock solution 1:1 H₂O:DMSO. [b] Reagent concentrations: 200 mM.

Prior to the macrocyclization step, we needed to generate the free amine and the free carboxylic acid by Nosyl and ester deprotection. This procedure could be done in a tandem reaction or in two consecutive steps. We first analyzed the influence of the sulfur nucleophile in combination with different bases at variable temperatures on product and side product (mainly terminal amide, see **Figure 55b**) formation. Only thiophenol (PhSH) and β -mercaptoethanol (BME) completely removed the Nosyl group. Ester deprotection was uniquely achieved in combination with DBU. The mildest conditions with the least side product formation were BME/DBU at room temperature (**Table 6**, Entry 10). These findings were well in accordance to a published procedure on Nosyl deprotection with DNA-encoded amino acids on a solid support.^[197] Ester hydrolysis (with LiOH) prior to Nosyl removal made the latter reaction very inefficient and led to a high amount of the undesired side product.

We further investigated the influence of the diversity element 2 structure on deprotection efficiencies. For this purpose, we synthesized a series of encoded macrocycle precursors with constant DE-1 (**NP19**) and 20 diverse DE-2 elements (**Figure 58a**). The best protecting group removal was achieved with proline derivatives (**AA049/053/106**), however the common proteinogenic amino acids Gly, Val and Phe (**AA009/010/012/056/058**) also showed good deprotection efficiencies (**Figure 58c**). The size and bulkiness of the amino acids seemed to have a big influence on the deprotection, because the sterically hindered dicyclohexylalanine (**AA008**) and the adamantyl amino acid (**AA101**) showed little to no conversion to the desired product, but the formation of a substantial amount of side product was observed. Despite relatively good conversions to the desired products, sarcosine (**AA013**), methylhistidine (**AA018**)

and serine (**AA025**) were amongst the members with the highest side product formation of all tested compounds. *L*-Asparagine (**AA047**), with its nucleophilic amide sidechain, underwent an intramolecular cyclization reaction with the adjacent ester to form the corresponding cyclic imide (**Figure 58b**).

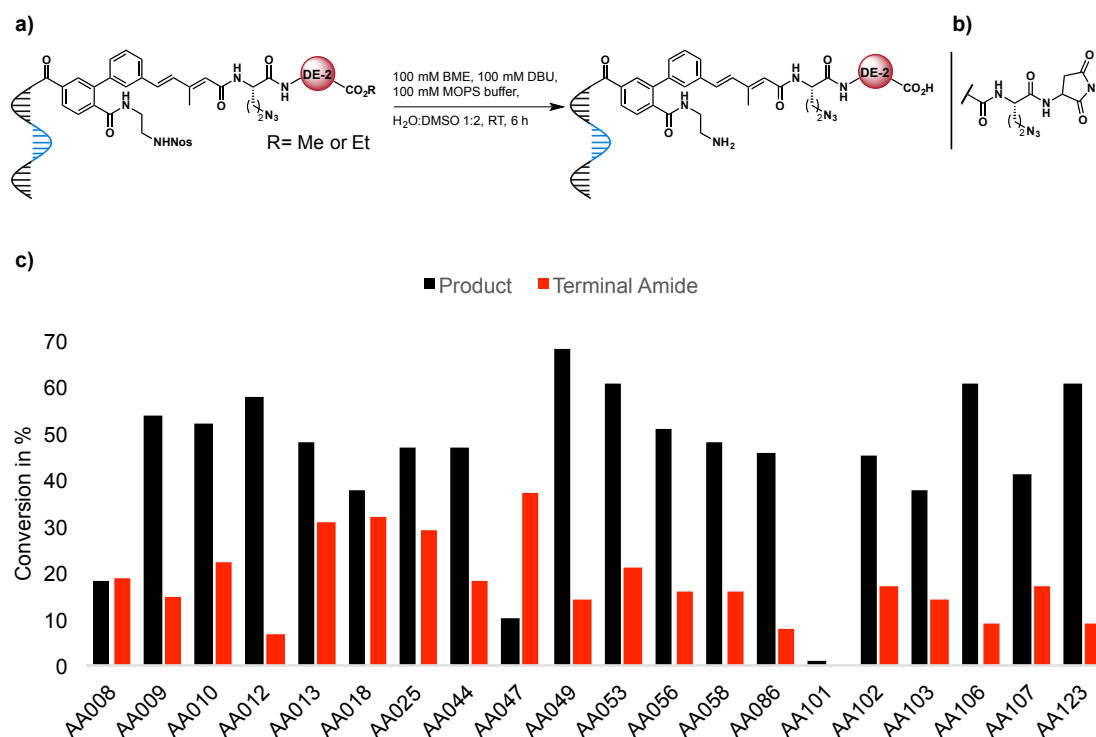


Figure 58. a) Schematic representation of the deprotection test assay. b) The found cyclic imide side product with *L*-asparagine (**AA047**). c) Nosyl deprotection results with 20 representative DE-2 building blocks, including the most special amino acids in terms of size and structure. The amounts of desired products are shown with black pillars, the undesired terminal amide side product is represented in red.^[198]

The optimized reaction conditions were then applied to the DE-1/DE-2 sub-library to generate the pool of free "amino acids", ready for macrocyclization. In a future attempt, several building blocks, including **AA008**, **AA047** and **AA101**, should be excluded from the library assembly due to the high amounts of side product and the low reactivity in macrocyclization.

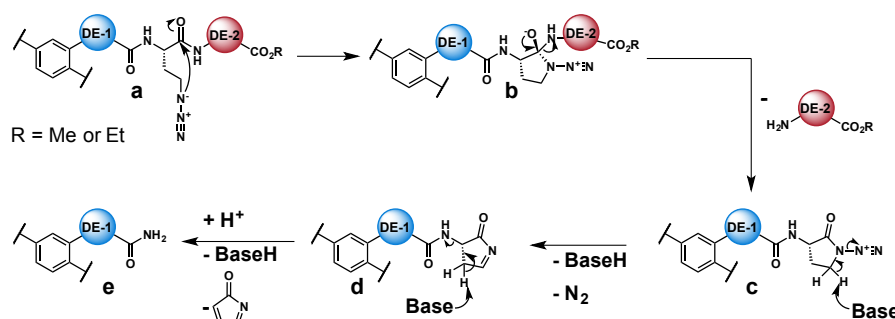


Figure 59. Proposed mechanism for the generation of the terminal amide side product during amide coupling and Nosyl deprotection. In the first step the azide attacks the amide bond of the trifunctional linker and the DE-2 (**a**). The formed five-membered intermediate **b** eliminates the DE-2 element followed by a base-induced nitrogen loss in **c**. In the final step *2H*-pyrrol-2-one is generated to yield the final terminal amide side product (**e**).

During the validation of the reactions and the compatibility of the building blocks, we often came across a certain side product with a molecular weight that varied with the nature of the DE-1

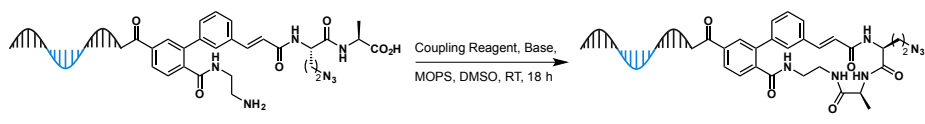
building blocks. From the mass spectrum, we proposed a DE-1 terminal amide structure (an example is shown in **Figure 55b** and a general structure is shown in **Figure 59**). We came across this particular side product after DE-2 amide coupling and Nosyl deprotection. We could demonstrate that the azide from the trifunctional linker is involved in terminal amide generation, because the resynthesis of two test molecules with an alkyne trifunctional linker in combination with **AA008** and **AA018**, respectively, showed no terminal amide formation. We also determined that the structure of the DE-2 element dictated the amount of the formed terminal amide (**Figure 58c**). A dependency on the structure of the DE-1 moiety was not observed.

We proposed that the azido group attacks the amide bond between the trifunctional linker and the DE-2 building block to form a five-membered ring intermediate (**b** in **Figure 59**), which further eliminates the DE-2 amino acid. This proposed step is supported by the fact that five-membered structures are kinetically favored. Moreover, the influence of the DE-2 structure supports such an assumption, since very bulky amino acid sidechains showed reduced reactivity in the deprotection, as well as side product formation (see **Figure 58**, **AA008** and **AA101**). It may be assumed that the large substituents shielded the adjacent groups, resulting in lowered terminal amide formation. Compound **c** can further lose a nitrogen molecule, which is likely the driving force for this side reaction. The dihydro-2*H*-pyrrole-2-one intermediate further undergoes the elimination of a 2*H*-pyrrol-2-one molecule, however, this process could also stem from the compound ionization in the ESI-MS. It is still unclear which compound we kept in solution after the reaction (compound **d** or **e**), but we could only identify the terminal amide **e** in the mass analysis. An alternative mechanism, involving going over a six-membered ring intermediate in the first step (reaction with the DE-1 amide), is also possible, but we considered it to be less likely, because it did not explain the influence of the DE-2 sidechain structure. However, the proposed mechanism does not explain why this side reaction only occurs during amide coupling and Nosyl deprotection, whereas ester hydrolysis with lithium hydroxide did not yield any side products. Perhaps, activating species, such as the coupling reagents for amide couplings, promote the generation of the undesired side product, although their direct influence is unclear.

Macrocyclization of the DE-1/DE-2 Sub-library

We tested the influence of various coupling reagents on the efficiency of the macrocyclization reaction. DMTMM-BF₄ was superior to all other reagents tested (see **Table 7**). In contrast to intermolecular amino acid couplings, the macrocyclizations performed well with fewer equivalents of the reagents (250-500 eq.). We could reduce the coupling reagent amount to as low as 100 eq. for the cyclization reaction to occur within the tested time frame, albeit with more side products. We think this is a result of slower kinetics in the cyclization reaction, which consequently increased side product formation. Optimal conditions were found with 250-500 equivalents of coupling reagent and 1000 equivalents of base. The DNA concentration in the final mixture was set to 10 μM to avoid dimerization products, but also concentrations down to 1 μM still gave efficient couplings.

Table 7. Efficiency of the macrocyclization with different coupling reagents and amounts. A model compound was used for the optimization (Scheme).



Entry	Coupling Reagent	Equivalents	Base	Macrocyclization
1	DMTMM-BF ₄	500	NMM	Yes
2	EEDQ	500	-	No
3	EDC*HCl/sulfo-NHS	500	DIPEA	very little
4	DMTMM-BF ₄	250	NMM	Yes
5 ^[a]	DMTMM-BF ₄	100	NMM	Yes

[a] Larger amount of side product formation.

We further investigated the influence of the DE-2 elements for macrocyclization reactions. As shown in **Figure 60**, the majority of the tested compounds showed cyclization ratios of >60%. Only amino acids that also caused difficulties during the preceding Nosyl deprotection (**AA008**, **AA047** and **AA101**) did not macrocyclize with this method. For **AA008** and **AA101**, the bulky and sterically hindered sidechains influenced the deprotection reaction, giving a small amount or even no unprotected species. Moreover, the bulky groups of these members potentially inhibited the activation of the carboxylic acid and/or prevented the amine from the nucleophilic attack. The asparagine macrocycle (**AA047**) was not accessible because the asparagine sidechain fully underwent the imide side product formation (**Figure 58b**).

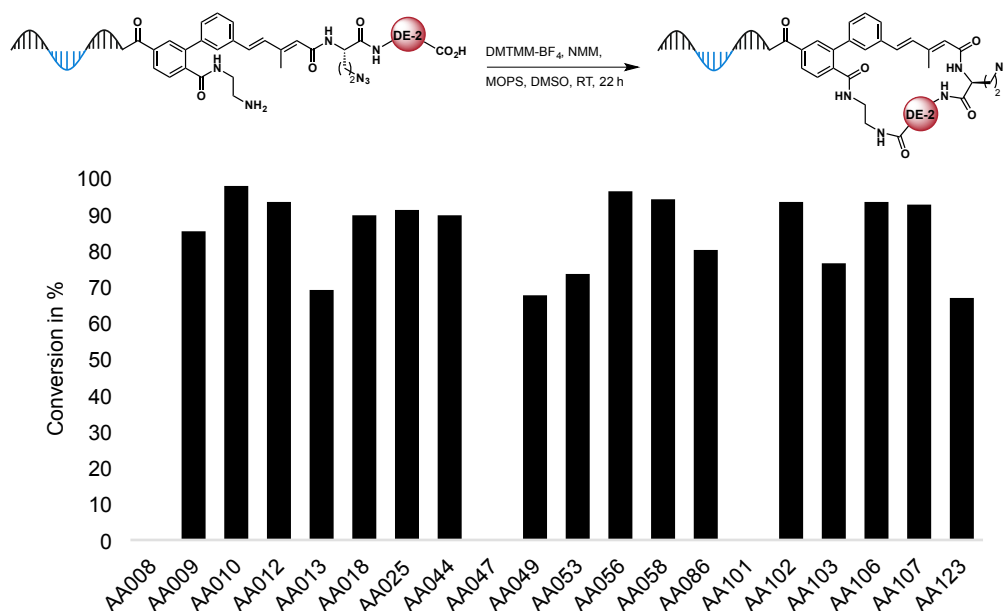


Figure 60. Schematic representation of the efficiency evaluation during macrocycle formation. We used the same precursors as for the full deprotection.^[198]

For the test macrocyclization (as was the case for the previous deprotection), we chose the most difficult and representative amino acid building blocks in combination with a rather rigid DE-1 moiety. This setup should ensure that we would find potential difficulties with those building blocks, which are rich in functional sidechain groups, rigidity and sterical demand.

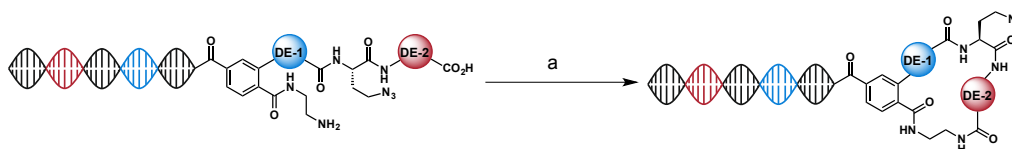


Figure 61. Schematic representation of the library macrocyclization. Reaction conditions: **a)** DMTMM-BF₄, NMM, MOPS buffer, DMSO, RT, 18 h.

With the certainty that the macrocyclizations should work well, we treated our precursor library with the established DMTMM conditions (300 eq.). Careful analysis of a reference sample indicated incomplete macrocyclization. We treated the library and the test sample with another portion of coupling reagent and were happy to find no additional side reactions, along with full conversion to the macrocyclic products. The macrocycle library (smDEML) was purified by reversed-phase chromatography. Some residual DE-2 encoding DNA could not be removed due to co-elution with the product. Nevertheless, the purity was high (>80% estimated from gel analysis, **Figure 64c**) and some of this library was kept for protein affinity testing. Comparison of the protein affinity results from this batch and the final library will give an idea about which part(s) of the molecules bound to the protein. If a macrocycle (without DE-3) and its DE-3 modification(s) both show high affinity, then the macrocyclic core is responsible for the protein-small molecule interactions. If the binder only shows up in the final macrocycle library, but not in the smDEML, then it is almost certain that the contribution to binding was due to the peripheral moiety and not due to the macrocyclic core.

Restriction Digest

For the last encoding step, we needed to generate 5' overhangs for the ligation with the DE-3 coding DNA strands. We chose to achieve this by BamHI restriction digest since we deliberately included the corresponding restriction site (see **Figure 47** for the schematic representation of the library encoding). In order to efficiently restrict the library, we preferred to perform this transformation in a pooled step, rather than in a split synthesis step with more than 600 vessels.

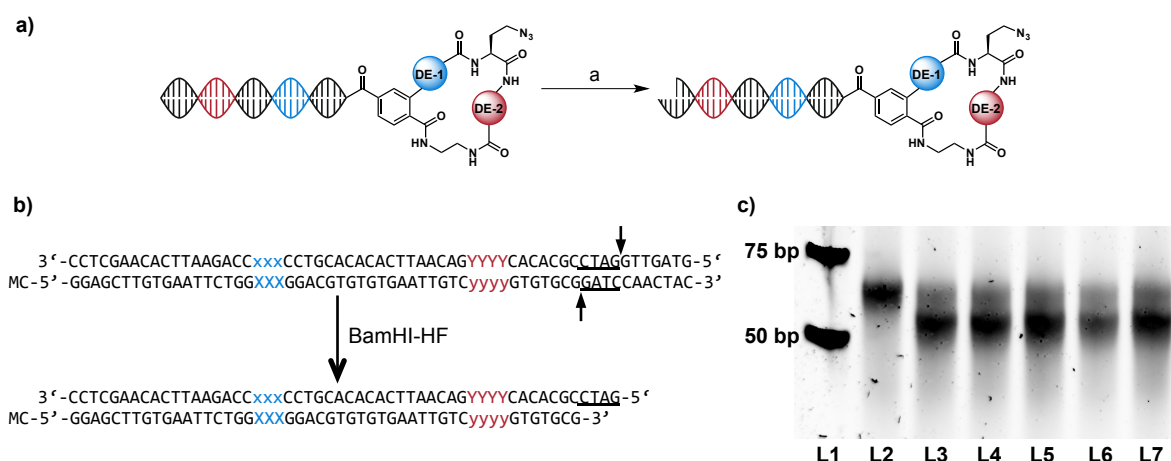


Figure 62. **a)** Schematic representation of the restriction digest. Reaction conditions: **a)** BamHI-HF, CutSmart buffer, 37°C, 30 min. **b)** Restriction scheme. Underlined bases indicate the recognition site for the BamHI-HF enzyme, the arrows show the restriction sites. **c)** Native PAGE gel of several restriction samples. **L1:** Low MW DNA ladder, **L2:** small DEML before restriction, **L3-L7:** Restricted samples.

The BamHI restriction enzyme cuts double-stranded oligonucleotides between the two guanines of the palindromic GGATCC restriction site to generate a four base 5' overhang, a so-called sticky end (**Figure 62b**). We performed a deep evaluation of potential reaction condition improvements (amount of enzyme, duration, double restriction) to find that the restriction worked very reliably,

but always with only about 80% overall restriction. Approximately 20% of the DNA stayed in its undigested form. Reference tests with unmodified duplex DNA strands with different lengths showed the same results. Therefore, we conducted the restriction digest with the library, taking into account that a certain percentage of our library would remain uncut. **Figure 62b** shows the successful restriction of the library. The loss in DNA length is displayed along with the remaining native DNA strands (grey blurry band above the restricted DNA band).

Introduction of the DE-3 Sub-Library by Click Reaction

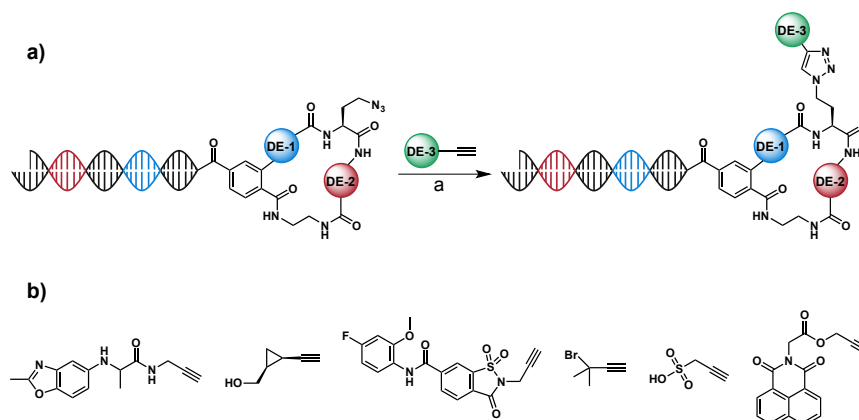


Figure 63. a) Schematic representation of the DE-3 coupling. Reaction conditions: a) Cu-TBTA, NaOAsc, TEAA, DMSO, RT, 16 h. b) A few representative examples of the diverse set of terminal alkyne building blocks.

The introduction of a 663-member terminal alkyne sub-library by copper-catalyzed click reaction represented the last chemical diversification step of the DEML (**Figure 63a**). The majority of the building blocks were commercially available and only a few very specific members had to be specifically synthesized. We tried to include a very diverse set of building blocks, covering a broad range of chemical space. We incorporated a vast set of functional groups such as amines, amides, ureas, sulfonamides, sulfonic acids, alcohols, esters, imides, halides and many more (**Figure 63b**). We also chose members with very different structures, shapes and stereochemical arrangements. Short linear aliphatic alkynes, branched functionalized or small cyclic members were included as well as large aromatic, heteroaromatic and complex alkyl-aryl framework structures. We did not assay the coupling efficiencies of each building block as was done for the DE-2 amino acids, but random sampling demonstrated that the CuAAC click reaction was very tolerant towards the incorporation of all building blocks tested in the course of library assembly and validation. Therefore, we considered this reaction to be the most reliable and least building-block dependent transformation of all chemical steps used during the DEML synthesis.

The only limitation of this chemical step turned out to be the reactants' concentration. While for the assembly of the encoded DE-1 sub-library only a few equivalents of the reactants and catalysts with a high DNA concentration were necessary, in this step, it was vital to perform the reaction at a single digit micromolar concentration. For this purpose a pseudo-first-order reaction was required. We discovered that the reaction performed very well with a few hundred equivalents of the reactants and catalyst, which is in sharp contrast to the amide couplings, where thousands of equivalents are necessary for a clean reaction.

We split the previously restricted smDEML pool into 663 vessels and carried out the click reactions with the complete set of terminal alkynes. Ethanol precipitation was utilized for cleanup and the DNA was encoded in a next step.

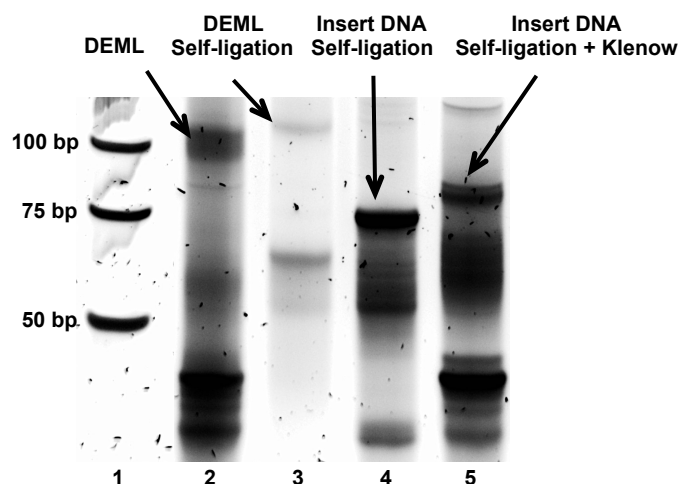


Figure 65. The purified DEML in comparison with self-ligated library or insert DNA strands. **L1:** Low MW DNA ladder, **L2:** Purified DEML, **L3:** DEML self-ligation assay, **L4:** Insert DNA self-ligation assay, **L5:** Insert DNA self-ligation assay after Klenow extension.

The palindromic nature of the BamHI binding and restriction site made it possible for the restricted oligonucleotides with "sticky ends" to undergo ligation reactions with each other or with itself (self-ligation). The self-ligation of the DEML was particularly critical because, in that case, the last diversity element was not correctly encoded for and both DNA strands of the duplex would carry the macrocycle. To overcome this issue, we used an excess of insert DNA to prevent DEML self-ligation. Comparison of lanes 2 and 3 in **Figure 65** led to the conclusion that DEML self-ligation did not occur during DE-3 encoding. On the other hand, the insert DNA formed its self-ligation product in good amount due to the excess in the mixture. Lanes 4 and 5 show the behavior of this excess during the DEML encoding steps. Surprisingly, the Klenow fragment seemed to truncate the self-ligated DNA to some part (new band at approx. 45 bp in lanes 2 and 5) rather than just turn it into a full duplex DNA. Completely double-stranded insert DNA (that was not ligated before) might appear at the same position, however, the intensity of this band is too high to completely stem from that reaction. These short oligonucleotide fragments were the main DNA impurity in our library, which we could not completely remove.

DEML purification was conducted by semi-preparative HPLC (**Figure 94a**) to remove enzyme, buffer and excess DNA leftovers. Since the purification step was not completely successful, we still see bands for short DNA fragments in the final PAGE analysis (**Figure 94b**). However, we considered the library useful without further cleanup because the HPLC chromatogram (**Figure 94b**) showed no potential purification improvement by the applied methods. The short DNA strands should not contain any attached small molecules and were not coding for the complete library. These excess oligonucleotides would finally be washed away during protein selections.

8. Properties Evaluation and DNA Damage Analysis of the DEML

8.1 Introduction to Physicochemical Properties of Macrocycles

Lipinski and Veber Rules for the Bioavailability of Small Molecules

The discovery of new, drug-like molecules is a very important task and can be achieved through the screening methods described in Chapter 6. Although theoretical chemical space is huge, only a tiny fraction of chemical structures possess drug-like activities. This subset of molecules is confined by suitable ADME (absorption, distribution, metabolism, excretion) and toxicity properties, enabling them to pass human phase 1 clinical trials.^[159] Drugs and their targets are spread out over chemical space, making their discovery difficult and laborious.

In a thought experiment, Lipinski defined the success of a truly diverse chemical library in finding a protein target hit as almost zero, taking into account certain assumptions. If a 100 kg human was completely composed of drug targets with molecular weights >500 Da, there would be about 10^{26} different druggable molecules. With the lower estimate of 10^{40} different compounds in chemical space, the chance of finding a drug against a target would require a diverse library with at least 10^{14} members. In reality, most drug receptors are much bigger than 500 Da and a human does not entirely consist of druggable moieties. For that reason, the number of drug targets is much smaller than 10^{26} and along with the fact that the chemical space must be assumed to be much larger (estimations range from 10^{40} to 10^{100}), the chance of finding drug molecules is even much smaller.^[159]

Very large, diverse libraries (10^{14} or more members) are theoretically possible with encoded library technologies, but are neither practical nor reliable because they suffer from combinatorial and synthetic deficiencies (e.g. a too low number of molecules per compound structure, building block availability or incompatibility of the building blocks with the reaction conditions). As presented in Chapter 6, many drug candidates with high affinity have been found with much smaller libraries containing several thousand up to a few million members. This implies that drug candidates are not distributed randomly throughout chemical space but are rather located within a well-defined (but still large) area. In addition, many drug targets might bind several different molecules with high affinity, which simplifies the search for good drug candidates. From this fact, libraries that cover well-defined areas of chemical space are preferred (so-called focused libraries that were specifically designed with respect to physicochemical properties or taking into account certain drug target specifications). Truly diverse libraries, as mentioned at the beginning of this section, can be created *in silico*, but in an experimental setup are not practical due to the limitations of chemical library assembly.^[159]

In addition to these limitations, drug candidates not only need to be active in target binding but they also need to possess the correct physicochemical properties to make them orally bioavailable.^[199] Compounds with many very polar groups may display extraordinary solubility properties in a biological environment but membrane permeability will decrease with higher polarity. In contrast, high lipophilicity makes a compound more permeable across the membrane of a cell, but decreases aqueous solubility. These competing factors lead to difficulties for the administration of the drug and might influence the drug uptake into the cell. Based on these findings and in comparison with successfully developed remedies, Lipinski defined a catalogue of physicochemical properties that small molecule drug candidates should fulfill to be considered as lead compounds in pharmaceutical medicine discovery. This property-set is called the (Lipinski) rule-of-five (Ro5) because the cutoffs for all parameters were close to five or a multiple of five. He defined that the molecular weight of a drug candidate should be ≤ 500 Da, the partition coefficient logP should be ≤ 5 , the number of hydrogen bond donors should not exceed five and the number

of hydrogen bond acceptors should be maximum 10 (see **Table 8**).^[159,200] However, the Ro5 are not strict. Violation of one of the parameters might still deliver reasonable drug candidates but upon violation of two or more parameters the compound becomes very unlikely to be drug-like.^[199] Nevertheless, oral bioavailability measurements with more than a thousand compounds on a rat model showed that the proposed properties of the Ro5 are insufficient to fully describe the drug-likeness of a molecule. The rigidity of a compound, measured by the number of rotatable bonds as well as the polar surface area must also be included in the physicochemical properties analyses. LogP and molecular mass are usually used for the analysis of a compound's lipophilicity, however these descriptors do not take into account structure-specific properties that influence the lipophilicity of the drug candidate. The number of rotatable bonds in combination with the polar surface area complements these properties. Veber *et al.* proposed to include cutoffs for the polar surface area at $\leq 140 \text{ \AA}^2$ and for the number of rotatable bonds at ≤ 10 .^[201] Despite these guidelines, supported by oral bioavailability measurements and calculated physicochemical properties with a large set of compounds, several molecule classes violate these rules, yet still are orally bioavailable.^[202,203] Among these classes are antibiotics, antifungals, vitamins and macrocycles.^[199,204]

Macrocycle Properties and Features Defined by Whitty

As described in the previous section, macrocycles often violate the Lipinski and Veber rules but still operate as bioavailable, potent drugs. Macrocycles are defined as cyclic molecules consisting of 12 or more ring atoms, which provides the compound a certain degree of structural pre-organization.^[204] To address the binding features to proteins of this compound class, Whitty *et al.* analyzed the binding modes and the structural conformations of 22 X-ray protein-macrocycle complexes consisting of 19 distinct macrocycles and 13 different proteins.^[205]

It was found that macrocycles interact with proteins in fundamentally different binding modes compared to linear small molecules. Macrocycles can interact with the protein in a face-on binding mode, which means that the core ring structure is in contact with the protein surface along with a set of the ring substituents. The edge-on binding mode describes a macrocycle sitting almost perpendicular on the protein, whereby only one side of the ring is in direct protein contact while the outer part of the ring scaffold is exposed to the solvent. Macrocycles with the edge-on binding mode usually adopt a flattened conformation that allows additional interactions of substituents located on the solvent-exposed part with the protein. Some macrocycles also adopt very compact, almost globular geometries to interact with the protein in a cleft or a pronounced depression.

The Whitty group could also show that the peripheral atoms (atoms connected with one covalent bond to the ring scaffold, e.g. methyl groups, carbonyl or hydroxy groups) interacted the most with the protein. Larger substituent interactions were still pronounced whereas the actual ring atoms showed the least interactions with the protein target. Due to the, in general, larger structure of macrocycles, they can interact with more and farther distributed binding spots on the protein than is the case with conventional small molecules. Even if the binding energy of a distinct spot on the protein with the macrocyclic ligand is much lower compared to small molecules, the overall binding energy may be higher because often a larger number of such spots can be simultaneously occupied by the macrocycle. This feature renders macrocycles capable of drugging proteins that were considered undruggable with conventional small molecules. Deriving from the results of this protein-macrocycle binding evaluation, the authors proposed a set of physicochemical properties guidelines in analogy to the Lipinski rules (see **Table 8**). In addition to good binding properties, macrocycles also need to show high metabolic stability as well as good solubility and membrane permeability, which are key features for orally bioavailable drugs.^[205]

A conventional small molecule drug with MW <500 Da, for example, contains about 25% polar heavy atoms to maintain aqueous solubility along with good membrane permeability. Such a molecule usually fulfills the physicochemical properties guidelines from Lipinski and Veber. For a compound with MW = 1000 Da, about 25% polar heavy atoms are also required for water-solubility. Consequently, the small and the big molecule will have the same octanol-water partition coefficient (logP). However, the larger compound will most probably not be orally bioavailable due to the violation of the Ro5 in terms of mass, hydrogen bond donors, hydrogen bond acceptors and polar surface area. The energetic penalty of desolvation and transfer into a nonpolar environment increases with the absolute number of polar groups. With increasing molecular weight it becomes more difficult for a compound to show high solubility in polar and nonpolar environments.^[206]

Large molecules (such as certain macrocycles) possess "chameleonic" properties to achieve good oral bioavailability.^[206,207] These properties result from conformational changes dependent on the environment. In an aqueous environment the macrocycle is in equilibrium between an "open" conformation whereby the polar groups are exposed into the aqueous environment and a "closed" conformation with partially shielded polar groups within the compound along with intramolecular hydrogen bonds (**Figure 66**). Upon permeation through the membrane of a cell, only the "closed" form is present, shielding the polar groups within the compound and exposing the nonpolar groups to ensure sufficiently high lipophilicity. When the macrocycle reaches the inside of the cell it readopts the conformational equilibrium for the aqueous environment. Such equilibria render strong solvating interactions with the water low to reduce the energetic cost upon desolvation and membrane permeation while maintaining aqueous solubility. For high molecular weight compounds it is a prerequisite to show a certain degree of "chameleonic" behavior in order to be orally bioavailable.^[206,207]

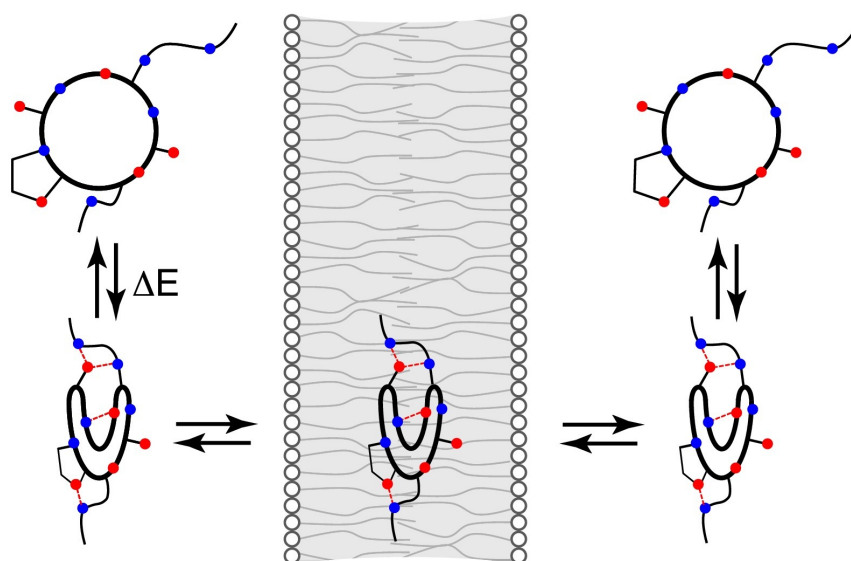


Figure 66. Cartoon explanation of chameleonic properties. Macrocycles in aqueous solution are in equilibrium between an "open" (upper structure) and "closed" (lower structure) form. Upon membrane permeation only the "closed" conformation is adopted to shield the polar groups within the molecule. Upon resolution in the aqueous environment the polar groups are exposed to the solvent again and the macrocycle readopts the conformational equilibrium.^[206] Reprinted with permission.^d

^d Elsevier, License Number: 4537081008616, 27.02.2019.

Macrocycle Properties Requirements by Kihlberg

Kihlberg *et al.* reported another set of macrocycle design guidelines in 2016, which were based on experimentally determined physicochemical properties.^[208] They selected a set of 214 non-peptidic macrocycles from a >100'000 compounds high-throughput screening library. This set consisted of three main types (different macrocyclization reactions) and a few miscellaneous macrocycles. These substances were used for solubility, lipophilicity, pK_a and permeability measurements across Caco-2 cell monolayers. From these empirical results Kihlberg *et al.* identified structural motifs that influenced cell permeability. Phenyl, pyridyl, isoxazole and tertiary amines were found to improve cell permeability whereas ureas, carbonyl groups, sulfonamides and secondary amines were unfavorable. Furthermore, stereo- and regiochemistry was shown to potentially have a big influence on cell permeability, while no substantial effects on aqueous solubility were observed. They further demonstrated that unfavorable functional groups in combination with favorable groups might exhibit synergistic cell permeability improvements. The opposite was observed as well, whereby antagonistic effects even further reduced cell uptake than what was expected for the functional groups alone. With their assay, they showed that predictions for oral bioavailability are achieved in a more reliable form, if the three-dimensional compound structure and certain higher energy conformations are also taken into account. From these results they generated a set of physicochemical properties guidelines,^[209] which are depicted in **Table 8**.

8.2 Evaluation of the Physicochemical Properties of the DEML

The building blocks that were used for the construction of the DEML had all been manually selected with an emphasis on including natural product-like moieties and covering a broad range of chemical space in terms of functional groups, size and structure. We did this without the use of informatics tools that could predict the suitability of building blocks for the construction of drug-like molecules. For that reason, we analyzed the physicochemical properties of the final macrocycle library to get an idea of whether our macrocycle collection occupied a favorable chemical space for oral bioavailability. Based on this analysis, we will be able to optimize the building blocks for future libraries. For this purpose we developed an in-house software to construct the full library from its fragments (see Chapter 13.15) and subsequently calculate chemical properties with open-source software.^[210] The software suite is available on <https://github.com/Gillingham-Lab/DECL-Gen> and was developed in Python 3.6 with the following packages: Biopython^[211], RDKit^[212], NumPy^[213], SciPy^[214], pandas^[215] and Matplotlib.^[216]

The following properties were calculated with RDKit for the macrocycle library: Molecular weight, TPSA (calculated surface sum of all polar atoms), AlogP (logarithmic value of an atom-based calculation of a compound's partition coefficient between a polar/nonpolar mixture), number of hydrogen bond donors, number of hydrogen bond acceptors, number of rotatable bonds and size of the biggest ring.^e We compared these calculated results with the guidelines from Whitty^[205] and Kihlberg^[209], which gave us an estimate of the potential (oral) bioavailability of our macrocycles.

^e We also calculated the quantitative estimation of drug-likeness (QED), number of heteroatoms, number of N,O,NH,OH, number of rings, number of sp³ carbons and number of heavy atoms.

Table 8. Comparison of the proposed guidelines from Whitty, Kihlberg and Lipinski. MW = molecular weight, AlogP = atom based partition coefficient (logP) calculation, TPSA = topological polar surface area, HBDs = hydrogen bond donors, HBAs = hydrogen bond acceptors, RotBs = rotatable bonds.

	MW [g/mol]	AlogP	TPSA [\AA^2]	No. HBDs	No. HBAs	No. RotBs
Whitty ^[a]	$600 \leq x \leq 1200$	$-2 \leq x \leq 6$	$180 \leq x \leq 320$	≤ 12	$12 \leq x \leq 16$	≤ 15
Kihlberg	≤ 1000	$-2 \leq x \leq 10$	≤ 250	≤ 6	≤ 15	≤ 20
Lipinski	≤ 500	≤ 5	≤ 140 ^[b]	≤ 5	≤ 10	≤ 10 ^[b]

[a] Proposed physicochemical guidelines for oral macrocyclic drugs. [b] From the enhanced rules for conventional small molecule drugs by Veber.^[201]

As previously explained, the common rule-of-five (Ro5) from Lipinski^{[159][200]} cannot be applied to macrocycles since their occupied chemical space mostly lies beyond the Ro5 region. In that sense these compounds are pharmacologically uninteresting from a common small molecule perspective. However, as discussed in Chapter 8.1, macrocycles behave very differently than small molecules, making them suitable drug candidates despite their Ro5 violation. Comparing our results with the reported guidelines from **Table 8** we directly saw that our macrocycle library fit well to the Whitty and the Kihlberg rules (**Figure 67**).

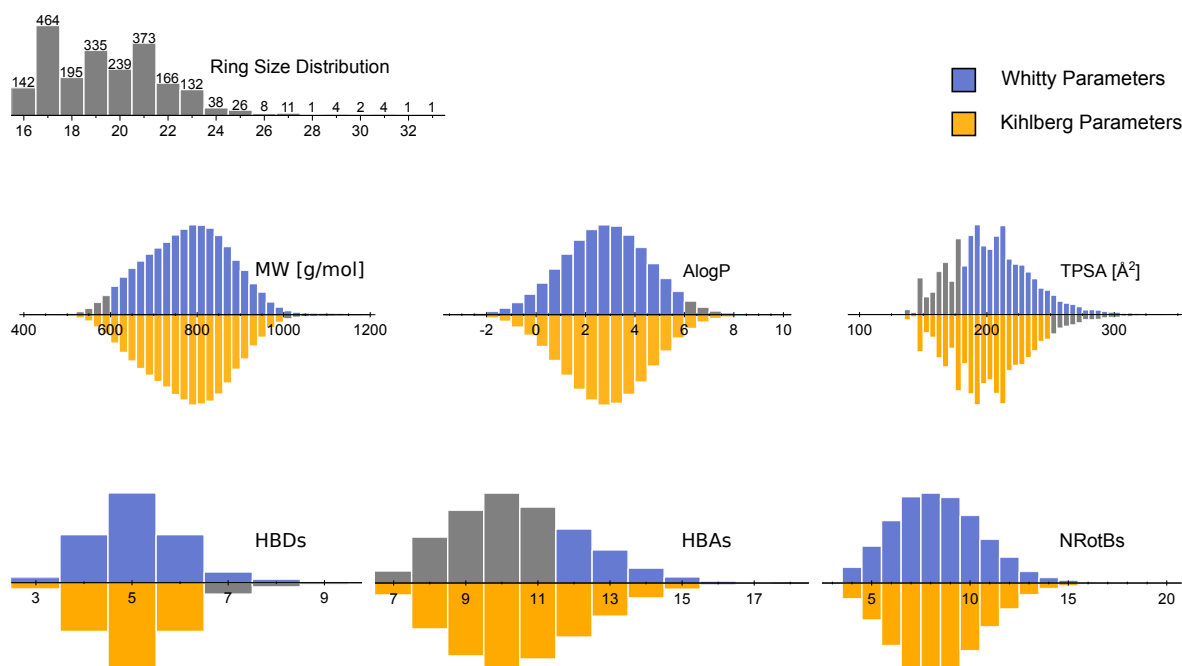


Figure 67. The most important calculated physicochemical properties of the macrocycle collection. Coloring shows the comparison of the calculated properties with the guidelines from Whitty (blue) and Kihlberg (orange) for potential macrocyclic drugs beyond the rule-of-five chemical space.^f

We found that the largest deviation from the Whitty parameters lies with the number of hydrogen bond acceptors (HBAs). The majority of our macrocycles possessed 9-12 HBAs, which was too low of a number by the Whitty definition. In contrast, the Kihlberg parameters that arose from direct macrocycle measurements state that 15 or less HBAs is suitable for a macrocyclic drug. A similar but less pronounced picture was found with the topological surface area (TPSA). For both

^f Analysis plots were generated by B. Sauter with the described software suite. Coloring and modification of the plot representation was performed by C. Stress.

guidelines a fraction of our library lay outside the ideal parameters, although both rule sets seemed to cover about 75% of our structures. We analyzed the distribution of the macrocycle ring sizes to further highlight the diversity and potential applicability of the DEML in drug discovery (top left graph in **Figure 67**). Most of the ring scaffolds contained 16 to 23 ring atoms with a few outliers having up to 33 ring atoms. From this analysis we felt confident that we had generated a diverse natural product-like macrocycle library that mostly fulfilled the reported criteria for drug-likeness.

8.3 Quantitative PCR (qPCR)

An encoded chemical library is only successful if the coding DNA strand remains intact and amplifiable. Therefore, not every chemical transformation may be conducted in the presence of a DNA strand due to incompatibility with the aqueous environment or chemical modification of the oligonucleotide. Malone and Paegel from the Scripps Research Institute Florida showed a set of DNA-compatible reactions and assayed the viability of the DNA strands by qPCR.^[145] They defined that reactions were useful for DECL synthesis if >30% viable DNA remained after the chemical transformation. For that reason, we assessed the DNA damage that occurred throughout the synthesis of our library.

We tested the potential DNA damage on a reference DNA strand without an attached small molecule. qPCR of the purified DNA strand revealed the remaining amount of fully intact DNA. Unsurprisingly, we observed that double-stranded DNA was more resistant to chemical modifications than single-stranded DNA (**Figure 68**). This implies that the Watson-Crick base pairing^[6] shields the nucleobases from the reactive species and predominantly backbone modifications occur. In the case of the click reaction, containing approximately 200 eq. of copper^(I), much more single-stranded DNA (89% damage) was damaged than double-stranded oligonucleotides (26% damage).

For amide couplings, we first tested the coupling conditions that were used for the DEML assembly. Since in our DEML assay the carboxylic acid was located on the DNA, the amine coupling partner was added in a huge excess. Good results were observed in the DMTMM couplings (DMTMM Amine) in which about 30% of the single-stranded and more than 75% of the double-stranded DNA were recovered intact. These were the conditions most often applied for amide couplings in our DEML assembly. On the other hand, we found troubling results for the EDC coupling with amines, where >99% of the DNA (single-stranded or double-stranded) was damaged. In previous reports on DNA damage during chemical synthesis, peptide coupling conditions showed a certain oligonucleotide damage potential (up to 80% damage), but never in such a huge amount.^[145] However, most DECL assemblies use the amine located on the DNA strand^[137,147] and perform the reactions with a large excess of activated carboxylic acid. Under our reaction conditions >99% of the coupling reagents and the amine did not react with the DNA-attached carboxylic acid, but were available for undesired oligonucleotide damaging reactions. This seemed to be particularly disadvantageous for the EDC coupling conditions. Luckily, we only used these conditions for a small number of amino acid couplings (e.g. tyrosine DE-2 building blocks). For comparison, we investigated the DNA damage of reported amide coupling conditions, where the amine moiety was located on the DNA strand and treated it with a huge excess of activated carboxylic acids. The DNA damage with DMTMM was similar as in the first assay (slightly less DNA damage with single-stranded DNA). In contrast to the first peptide couplings with EDC, in the control approach we obtained amplifiable DNA in a useful range (>30% for single- and double-stranded DNA).

The deprotection steps (ester hydrolysis and Nosyl) seemed to be the mildest conditions for single-stranded DNA, even though more than 50% damage was observed in both cases. Taking into account the determined DNA damage ratios, only 0.1% of our library DNA was left fully intact

(calculated from the stated percentages in **Figure 68** in respect to the synthesis scheme, **Figure 46**). Yet, this does not take into account the amount of DNA modifications that might have been removed again during the synthesis like activated phosphate backbone hydrolysis. DMTMM is known to modify DNA, but these modifications can be removed again by base treatment of the DNA.^[155] Moreover, DNA backbone modifications may not be problematic for the successful use of the DECL.^[217,218] Furthermore, some polymerases might even accept certain modifications of the DNA template for PCR replication.^[219,220]

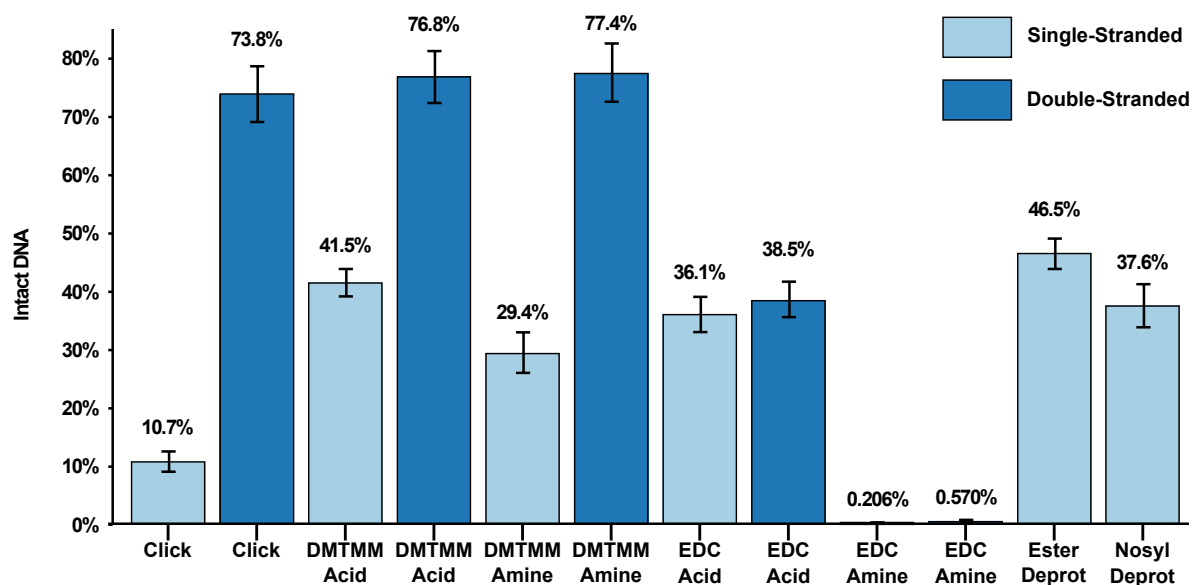


Figure 68. qPCR results from all tested reaction conditions with single- or double-stranded DNA. Not all of these conditions were applied during DEML assembly, but several of them have also been routinely used in DECL synthesis (e.g. EDC acid coupling).

In addition to the shown reaction conditions, we also investigated on the DNA damage of BIQ formation on DNA (reaction conditions from Part I, see Chapter 3). These conditions revealed a recovery rate of intact DNA of >40% for single-stranded and >60% for double-stranded DNA. In conclusion, BIQ formation on DNA may be very interesting as it fulfills all criteria for the use in DNA-encoded libraries in terms of DNA and reaction conditions compatibility as well as novelty of the introduced building blocks.

9. Protein Affinity Selections, Sequencing and Data Analysis

9.1 Protein Affinity Selections

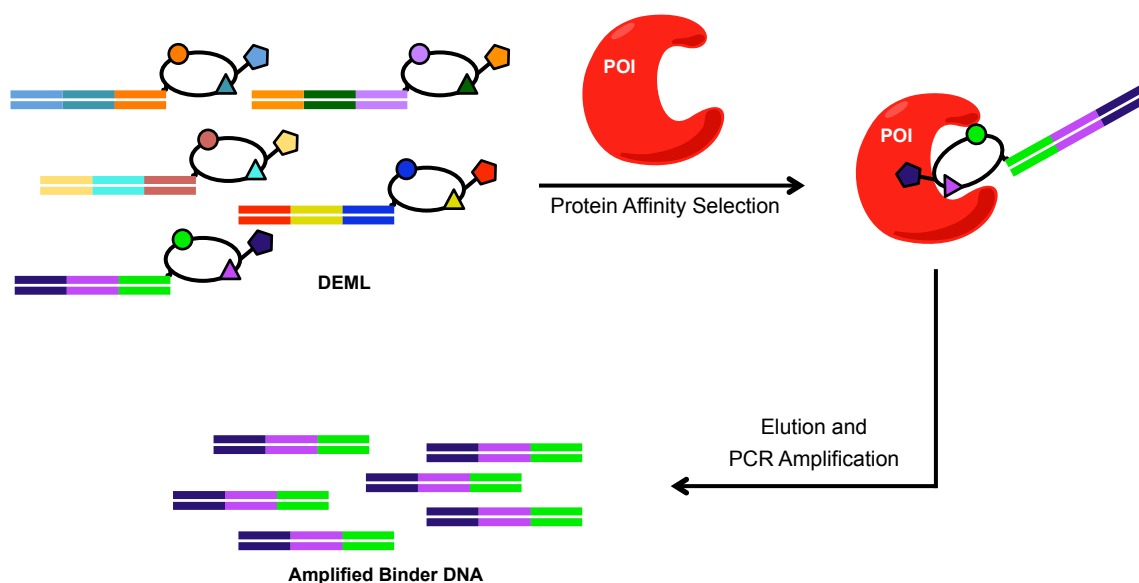


Figure 69. Schematic representation of protein affinity selections. The DEML was tested against the protein of interest (POI) to find enriched protein binders. Elution of these binders with subsequent PCR amplification delivered the binder DNA in sufficient quantities for DNA sequencing. For simplicity a binding event of a single binder is displayed. Usually series of binders are identified after sequencing with different relative abundances.

The synthesized DNA-encoded macrocycle library (DEML) and its precursor macrocycle library (smDEML, DE-3 missing) were tested for binding affinities (**Figure 69**) to human serum albumin (HSA), α -1-acid glycoprotein (AGP) and carbonic anhydrase 9 (CA9). In a first step, HSA and AGP were biotinylated according to a published procedure with an activated biotin derivative.^[164] The unspecificity of this procedure ensured a protein display on the streptavidin-coated magnetic beads in all possible orientations. The success of the biotinylation was confirmed by a band-shift assay with avidin. CA9 is commercially available with a C-terminal polyhistidine tag (His-tag). For this reason, we used magnetic nickel beads for His-tag CA9 immobilization. Blockage of the unoccupied protein binding sites was achieved with *D*-biotin (streptavidin) and imidazole (nickel beads). After protein immobilization, we incubated the two libraries with the protein beads, followed by several washing steps to remove weak and unspecific binding. Random DNA binding of the library to the beads or the proteins was minimized by the addition of sheared salmon sperm DNA, blocking unspecific DNA interaction sites. The remaining binders were eluted from the beads by heat denaturation of the proteins. For control experiments, we used dummy samples (beads with biotin-blocked or imidazole-blocked protein binding sites), which did not contain the target protein. Along with the library fingerprint, these reference selections were later used for data evaluation.

We conducted the affinity assays in duplicates with freshly prepared library solutions, whereby the same assays were performed on two consecutive days. For the CA9 assays, we included in one of the two selections a positive control spike-macrocycle with the known acetazolamide binder **TA664**. DNA amplification by PCR yielded the required oligonucleotide amounts for sequencing, whereby the selection specific identification barcodes, along with the Illumina primers were incorporated on both ends of the duplex DNA strands. We purified the PCR products with DNA column purification kits and agarose gel electrophoresis (after the 2nd PCR). Details on the selection protocols, PCR amplification and purification gels are presented in Chapters 13.7 - 13.9.

9.2 Next Generation Sequencing (NGS)

Introduction

Sequencing of the amplified DNA after protein affinity screening is the most important step for the evaluation of target binders. This is the major difference between high-throughput screening and encoded library technologies. Traditional DNA sequencing such as radioactive sequencing by Maxam and Gilbert^[221] or sequencing with fluorescently labeled dideoxynucleoside triphosphates in chain terminating reactions by Sanger^[222] have been used for many years to determine DNA sequences from genomic DNA. While Sanger sequencing is still a popular method for certain oligonucleotide sequencing applications, the Maxam-Gilbert method has mostly vanished.^[223] The main disadvantage of those methods is the rather low throughput of the sequencing, along with labor-intensive procedures.

In 2004, the first machines were released that used completely new techniques for high-throughput parallel DNA sequencing. The data output and speed outperformed state-of-the-art Sanger sequencing tremendously and thus coined the term "next generation sequencing".^[224] This improvement caused a strong decrease in sequencing costs. Since then, many different techniques have been established. All of them have in common, that they can produce large, genome-scale datasets. Many of these modern techniques are limited by read length, which requires the preparation of a so-called DNA library. Long DNA sequences are enzymatically restricted into shorter strands, which are then ligated to specific adapters, amplified and purified. After the DNA sequencing, bioinformatics tools reassemble the genome from the short oligonucleotide reads. In the case of DECLs the eluted DNA strands are usually short enough to be directly used for the adapter ligation, amplification and sequencing (see Chapter 9.1).

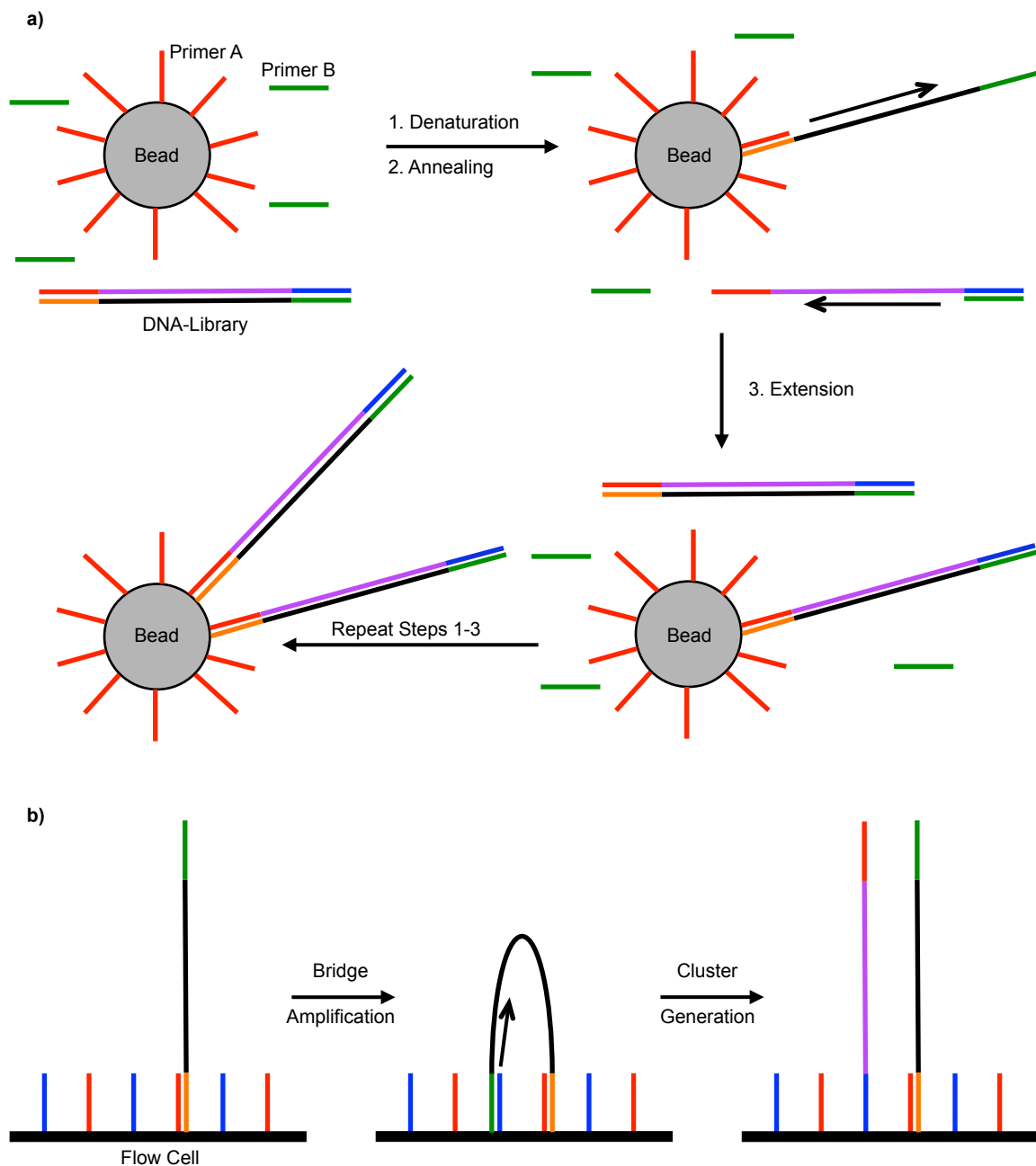


Figure 70. a) Schematic representation of the emulsion PCR amplification. A primer modified solid bead anneals with a specific DNA strand and gets amplified to immobilize the reverse complementary sequence on the bead. The process is repeated several times to cover the solid bead with a single DNA sequence. **b)** Schematic representation of the bridge amplification. The surface of a flow cell is covered with two different primers. A complementary oligonucleotide anneals and further adopts a bridge structure to anneal with the second primer. This induces the generation of the covalently attached reverse complementary sequence. Repetition of this procedure generates clusters of DNA with identical sequence.

A certain amount of identical DNA is a prerequisite to generate a detectable signal for the DNA sequencing reaction. For this purpose, the DNA library is amplified again using two major methodologies. Emulsion PCR uses tiny water droplets in oil as micro reactors (**Figure 70a**). In addition to the necessary reagents and primers for PCR amplification, every micro reactor contains one solid bead with an attached primer that is complementary to a DNA strand from the DNA library. With this method, single beads carrying identical DNA strands are generated.^[225] Disadvantages of this methodology are that the DNA distribution of one DNA strand per bead is

not always perfectly achieved. This leads to sequencing data that has to be filtered from the reliable data.

Another methodology (the so-called bridge amplification) uses immobilized DNA primers on a glass plate (**Figure 70b**). The DNA molecules from the DNA library specifically bind to these primers. PCR amplification then generates colonies of identical DNA strands. The amplification is achieved since both primers (forward and reverse) are located on the glass plate. The DNA adopts a bridged shape for priming with the complementary primer to generate colonies in which the forward and the reverse oligonucleotide strands are present.^[223]

The actual sequencing is performed after the amplification by several different procedures. In the Roche 454 pyrosequencing the generated DNA beads (from emulsion PCR with biotinylated primers) are placed in a picotiter-plate. Non-biotinylated complementary DNA strands are washed away and the nucleotide building blocks are washed over the plate one at a time. The sequencing method is based on the DNA synthesis from the single-stranded oligonucleotides attached to the beads in the plate. The incorporation of a nucleotide generates a molecule of pyrophosphate, which is used for the generation of ATP. In a coupled luciferin-luciferase reaction the generated ATP induces the emission of light, which is recorded by a camera. The amount of emitted light is proportional to the number of incorporated nucleotides.^[224]

The SOLiD (Sequencing by Oligonucleotide Ligation and Detection) system uses ligation reactions for DNA sequencing. After emulsion PCR the DNA strands are treated with a primer, followed by the ligation of different fluorescently labeled di-nucleotides (pre-defined nucleotide combinations). The fluorophore is cleaved off, which generates the detection signal and the next round of di-nucleotides is incorporated. After completion of the first cycle of ligation, dye cleavage and detection the generated complementary strand is removed and a new cycle starts with a one base longer primer. After six cycles of this ligation series, all nucleotides have been read in duplicates. The detection signals of the dyes can then be transformed into the DNA sequence with bioinformatics tools.^[224]

Ion Torrent technology uses semiconductors to detect tiniest pH changes. During the synthesis of a DNA strand, protons are released, which induces measurable acidity changes. The number of released protons is proportional to the number of incorporated nucleotides. After emulsion PCR the beads are placed in microwells and then stepwise flooded with single nucleotide solutions. This technology generates large datasets within short time and low cost. However, the accuracy of the method is only at about 98%.^[224]

The Illumina platform uses the described bridge amplification for DNA immobilization and generation of the required clusters. The sequencing is performed by DNA synthesis with reversibly blocked dye nucleotides. The nucleotides are stepwise incorporated, the dye cleaved off and detected by a camera. This enables the next cycle of nucleotide incorporation. With this technology all DNA strands are elongated simultaneously one nucleotide per step. This method generates a high output per run at a reduced price.^[224]

Further techniques have evolved over time and there is a lot of research focusing on new procedures for DNA sequencing. An example is *in situ* sequencing (ISS) that enables the sequencing of the genome and transcriptome of whole tissue samples.^[226]

Sequencing of the DEMs

We equipped our eluted binders from the protein affinity assays during the two PCR steps with the necessary primers for Illumina sequencing. Each sample was given a unique index primer sequence and all samples were pooled together. The sequencing of that pool was performed with a target read count of 240 million reads and paired-end 150 sequencing (reading 150 bases from

both ends of the DNA) by Novogene. From the generated sequencing data, we analyzed the results with the DECL-Gen.^[227] The theoretical DNA strands were aligned with the obtained sequencing results and the codons were extracted. DNA strands with damaged or missing codons were discarded, as well as reads where the codon on both strands was read differently. Sequences that displayed gaps or errors in the constant regions of the DNA strand with intact codons were also included to improve the number and quality of the data set. Detailed statistics about the sequencing data are displayed in **Table 13**.

For the smDEML assays we had a very high coverage of several thousand reads per codon, whereas for the DEML we achieved coverage of 5-7 reads per codon. That is explicable by the much bigger library size of the DEML versus the smDEML at a comparable sequencing depth. The coverage gives an estimate of the expected number of reads per codon in an equally distributed sample. This is never the case in enrichment assays where the highly enriched sequences appear much more frequently than those of the weak binders. The invalid codon content was in all samples below 7% with a trend that the smDEML showed significantly less invalid codons than the DEML. This indicated a lower purity of the DEML codons, which may stem from the last encoding step or the additional chemical transformation, leading to more DNA damage. We were happy to find 100% of all codons in the smDEML fingerprint and about 93% of all codons in the DEML. This stated that above 90% of all codon combinations had been generated during library assembly.

9.3 NGS Results Analysis

Hit Identification (smDEML)

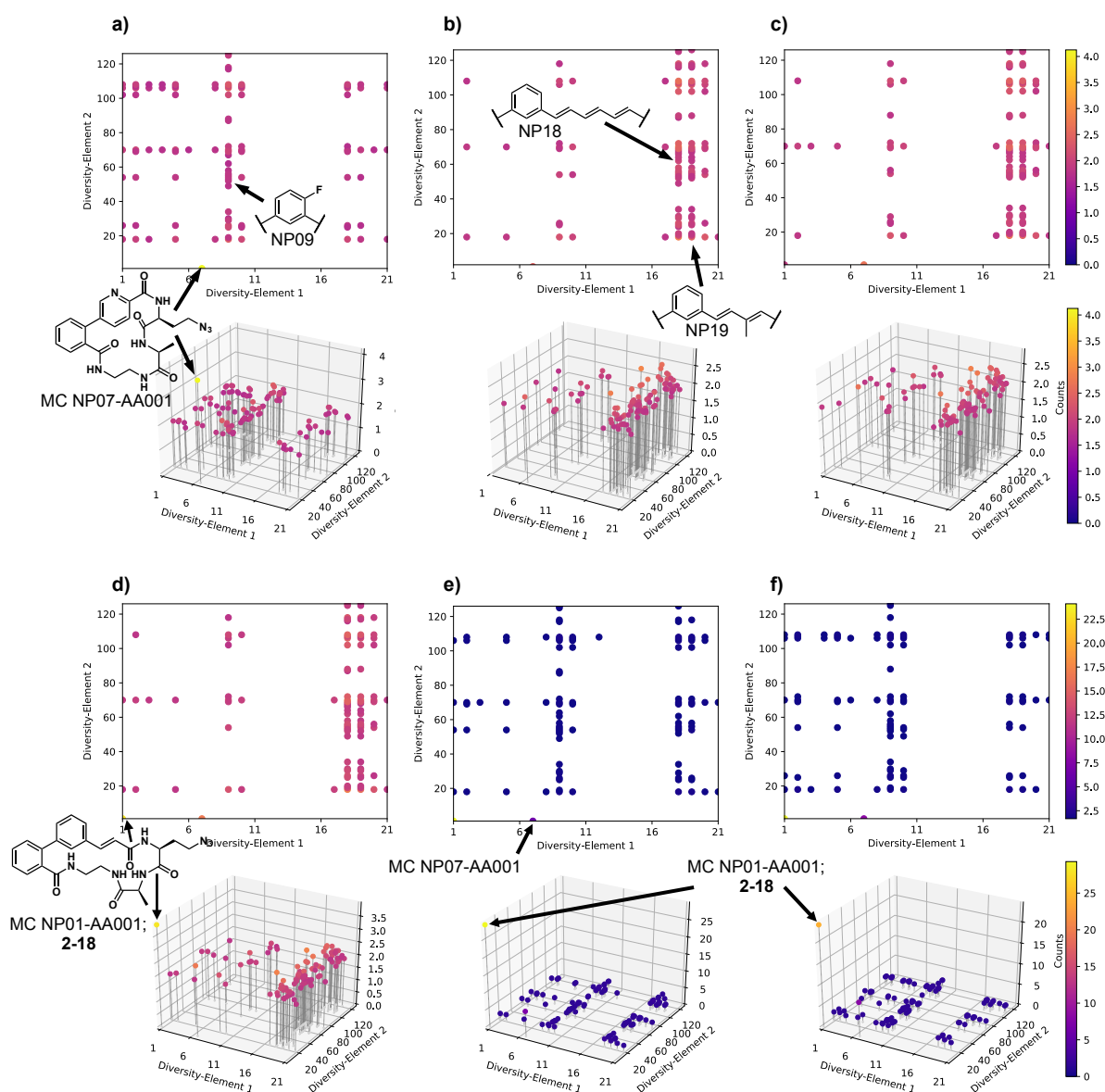


Figure 71. Scatterplots for the selections of the smDEML, presented in 2D (upper) and 3D (lower) plots. For clarity, only the 100 most abundant compounds are shown. **a)** smDEML fingerprint. **b)** smDEML dummy selection (streptavidin beads only). **c)** smDEML selection against AGP. **d)** smDEML selection against HSA. Coloring and scale bar according to a-c. **e)** smDEML CA9 dummy selection (nickel beads only). Coloring and scale bar according to f. **f)** CA9 selection.^[227]

Potential binders were identified by comparison of the found hits in the protein selections to the dummy selections and the fingerprint. Hits containing the DE-1 elements **NP09**, **NP18**, **NP19** and **NP20** were directly excluded, since these compounds were highly overrepresented in the fingerprint and/or the dummy selections. Those moieties seemed to favor binding to the magnetic beads (perhaps streptavidin or nickel), which led to the conclusion that these compounds were false positives. Macrocyclic **NP07-AA001** was found to be the most abundant compound in the fingerprint and reappeared in high amounts in all other selections, which required its exclusion from the analysis (**Figure 71**). The AGP selection did not reveal any potentially interesting protein binders. For HSA we found only one enriched compound (MC NP01-AA001, **2-18**) that seemed to be an HSA binder. However, in addition to the previously stated, overrepresented building

blocks that also appeared in the CA9 selections, we found macrocycle **2-18** in these selections as well, but this time highly enriched in the dummy and the protein selections (**Figure 71e** and **f**). We speculated that this compound might also be a good binder for nickel. We created a list of the remaining potential macrocyclic protein binders (see **Table 9**) from which we chose the compounds for synthesis and binding affinity measurements. Only the top hit in the HSA selection was finally considered valuable for binding affinity tests.

Hit Identification (DEML)

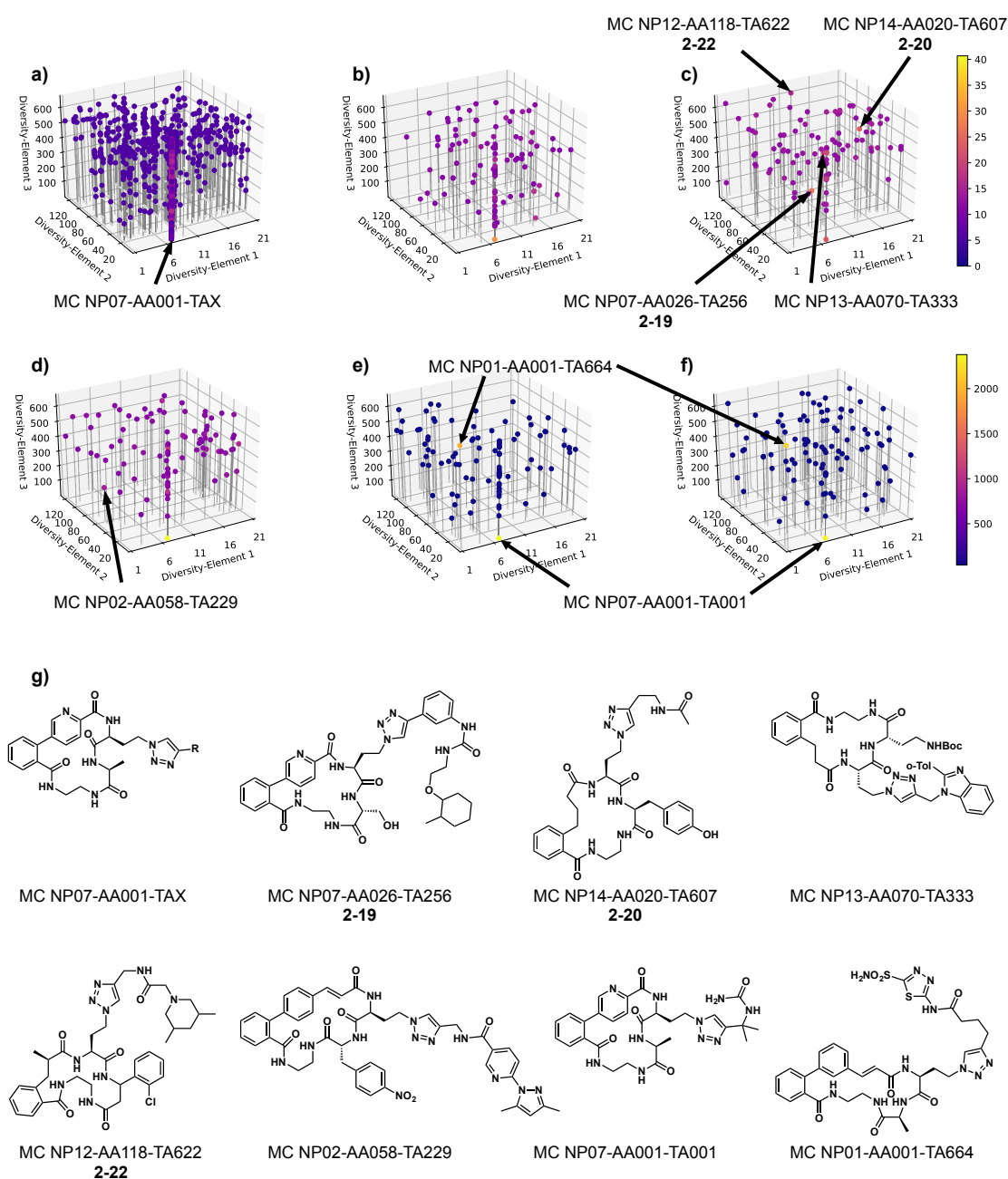


Figure 72. Scatterplots for the selections of the DEML in 3D. For clarity, only the 100 most abundant compounds are shown for the selections, 1000 compounds for the fingerprint plot. **a)** DEML fingerprint. **b)** DEML dummy selection (streptavidin beads only). **c)** DEML selection against AGP. **d)** DEML selection against HSA. Scale bar and coloring according to a-c. **e)** DEML CA9 dummy selection (nickel beads only). Coloring and scale bar according to f. **f)** CA9 selection. **g)** Structures of the indicated compounds (^[227]).

As with the smDEML, potential binders were identified by comparison of the found hits in the protein selections to the dummy selections and the fingerprint. Hits containing the DE-1 elements **NP09**, **NP18**, **NP19** and **NP20** were excluded again, since these compounds were highly overrepresented in the fingerprint and dummy selections. The compound group **MC NP07-AA001-TAX** was highly overrepresented in the fingerprint, resulting from the same high abundance of macrocycle **MC NP07-AA001** (compare to smDEML selection analysis). For the carbonic anhydrase 9 selections we found **MC NP07-AA001-TA001** to be even more enriched than in the HSA and AGP selections. We proposed that the pyridyl moiety in combination with a triazole and a urea could form complexes with the nickel (or other metals) from the beads. As a positive control we spiked the library for the CA9 selections with the individually synthesized macrocycle **NP01-AA001-TA664** that contained the azolamide building block, a known carbonic anhydrase 9 binder.^[228] However, an enrichment compared to the dummy selection was not observed. Either the spike concentration was too high in the assays, so that we could not see enrichment differences between the sample and dummy, or no protein binding took place (which might be an indication of unsuccessful protein immobilization).

A summary of the top hits of these selections is shown in **Table 9**. We chose four compounds from the AGP selection and one macrocycle from the HSA selection (see **Figure 72** and **Table 9** for the details) for protein binding affinity evaluations. In the case of macrocycles **MC NP02-AA058-TA229** and **MC NP13-AA070-TA333** the first diversity element was changed to **NP01** and **NP12** respectively. These changes were done due to the high similarity of these elements and to simplify the macrocycle resyntheses (see the structures of **2-21** and **2-23**). From the CA9 results we did not choose any compounds for resynthesis because of the high uncertainty level in the selection data.

Table 9. Results for the different library-protein selections. The compounds were ranked after comparison of the protein selection with the dummy selection experiment and the fingerprint. Only the top 6 compounds are listed, the rest was considered too little enriched. Compounds in **bold face** and grey coloring were chosen for resynthesis and binding affinity measurements.

Small DEML selection against AGP				DEML selection against AGP				
Rank	DE-1	DE-2	Number	Rank	DE-1	DE-2	DE-3	Number
1	NP20	AA069	NP20-AA069	1	NP07	AA026	TA256	NP07-AA026-TA256 2-19
2	NP19	AA088	NP19-AA088	2	NP14	AA020	TA607	NP14-AA020-TA607 2-20
3	NP19	AA117	NP19-AA117	3	NP13 ^[b]	AA070	TA333	NP13-AA070-TA333 2-21
4	NP18	AA053	NP18-AA053	4	NP12	AA118	TA622	NP12-AA118-TA622 2-22
				5	NP21	AA125	TA412	NP21/AA125/TA412
				6	NP09	AA029	TA443	NP09/AA029/TA443
Small DEML selection against HSA				DEML selection against HSA				
Rank	DE-1	DE-2	Number	Rank	DE-1	DE-2	DE-3	Number
1	NP01	AA001	NP01-AA001 2-18	1	NP21	AA026	TA423	NP21-AA026-TA423
2	NP18	AA034	NP18-AA034	2	NP14	AA086	TA621	NP14-AA086-TA621
3	NP17	AA070	NP17-AA070	3	NP02 ^[a]	AA058	TA229	NP02-AA058-TA229 2-23
4	NP20	AA069	NP20-AA069	4	NP15	AA020	TA459	NP15-AA020-TA459
5	NP19	AA029	NP19-AA029	5	NP19	AA039	TA357	NP19-AA039-TA357
6	NP20	AA072	NP20-AA072	6	NP12	AA029	TA553	NP12-AA029-TA553
Small DEML selection against CA9				DEML selection against CA9				
Rank	DE-1	DE-2	Number	Rank	DE-1	DE-2	DE-3	Number
1	NP10	AA026	NP10-AA026	1	NP01	AA006	TA316	NP01-AA006-TA316
2	NP02	AA069	NP02-AA069	2	NP08	AA042	TA519	NP08-AA042-TA519
3	NP05	AA026	NP05-AA026	3	NP08	AA018	TA410	NP08-AA018-TA410
4	NP09	AA019	NP09-AA019	4	NP06	AA072	TA589	NP06-AA072-TA589
5	NP10	AA107	NP10-AA107	5	NP19	AA102	TA429	NP19-AA102-TA429
6				6	NP05	AA032	TA430	NP05-AA032-TA430

[a] Due to the high similarity and ease of chemical synthesis, NP02 was replaced by NP01 for off-DNA macrocycle synthesis. [b] Due to the high similarity and ease of chemical synthesis, NP13 was replaced by NP12 for off-DNA macrocycle synthesis.

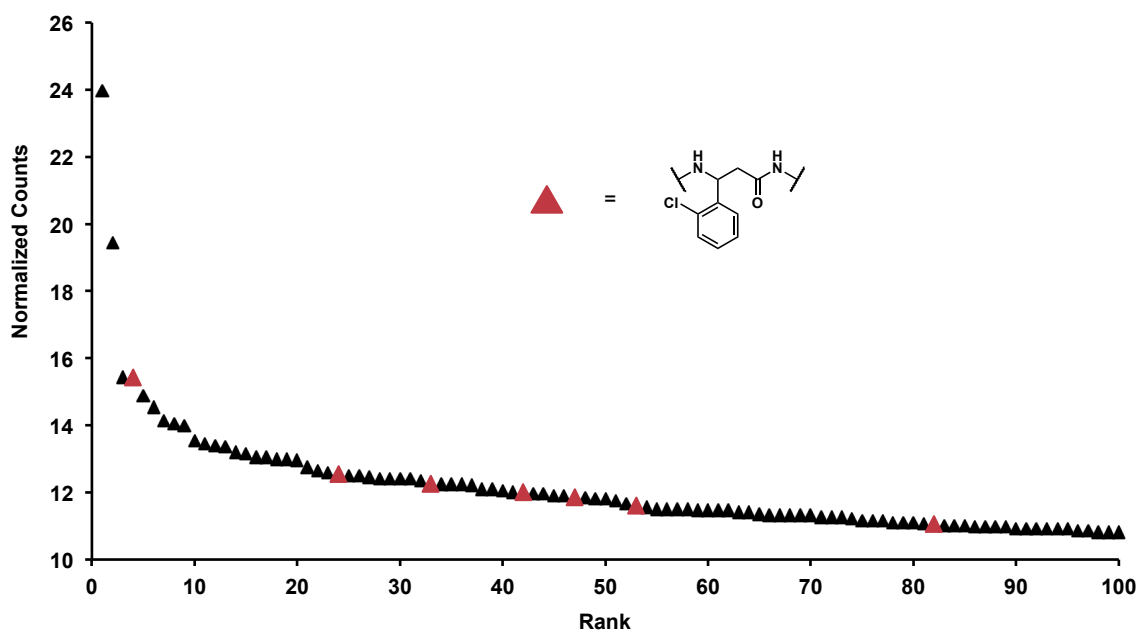


Figure 73. Top 100 AGP binders ranking. Molecules with the 2-chlorophenyl β -amino acid are highlighted in red.

From the results of the smDEML and DEML affinity selections (**Figure 71** & **Figure 72**) we could not directly see any trends concerning enrichment of diversity elements or diversity element combinations. The found top hits all seemed to be singleton hits except for the bead binders, which all showed the same or very similar diversity elements. For that reason we looked at the distribution of the second diversity elements among the top 100 hits in the DEML versus AGP selections. The most abundant element that we found was building block **AA001** (*L*-alanine), followed by the 2-chlorophenyl- β -amino acid (**AA118**) element. However, **AA001** was also the most abundant building block in the dummy samples, whereas the **AA118** did not show up among the top 10 binders in the dummy selections. This comparison showed a clear enrichment of building block **AA118** in the selections against AGP. We noticed similar behavior for only one other building block. **AA029** (*L*-tryptophan) was also slightly enriched but to a lesser extent than **AA118** (5 appearances in the top 100 hits). **Figure 73** displays the locations of the **AA118** containing macrocycles among the top 100 AGP binders. We found it as the fourth highest enriched compound along with another five times between ranks 20 and 60.

10. Macrocycle Resynthesis and Protein Binding Affinity Measurements

10.1 Resynthesis of the Macrocycles

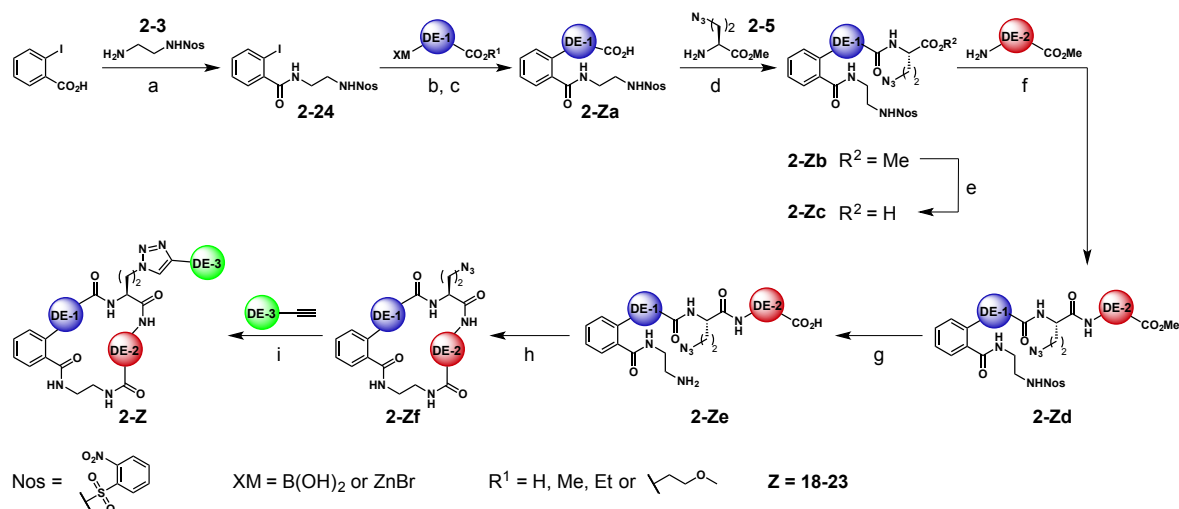


Figure 74. Schematic description of the macrocycle resynthesis for protein binding affinity assays. **Z** is the place holder for the macrocycles defined in **Table 9**. **DE-1**: **NP01**, **NP07**, **NP12**, **NP14**. **DE-2**: **AA001**, **AA020**, **AA026**, **AA058**, **AA070**, **AA118**. **DE-3**: **TA229**, **TA256**, **TA333**, **TA607**, **TA622**. Reaction conditions: **a**) HBTU, DIPEA, THF, RT, 2 h, **94%**. **b**) Pd (π -cinnamyl) chloride dimer, K₃PO₄, EtOH, H₂O, 50°C 1 h or Pd(dppf)Cl₂, NEt₃, methoxyethanol, 100°C, 4 h or Pd (π -cinnamyl) chloride dimer, THF, 0°C \rightarrow RT, 1 h, **53-82%**. **c**) LiOH * H₂O, MeCN, H₂O, RT, 2.5 h, **49-99%**. **d**) HBTU, DIPEA, THF, RT, 1.5 h, **96-99%**. **e**) LiOH*H₂O, MeCN, H₂O, RT, 1.5 h, **73-99%**. **f**) HATU, DIPEA, THF, RT, 1 h, **66-94%**. **g**) 1. PhSH, DIPEA, MeCN, RT, 2 h; 2. LiOH*H₂O, MeCN, H₂O, RT, 1.5 h, **76-99%**. **h**) HATU, DIPEA, THF, RT, 2 h, **5-89%**. **i**) CuSO₄, NaOAsc, DMSO, H₂O, RT, 2-14 h, **35-93%**.

In order to perform quantitative protein affinity measurements we synthesized the most promising macrocycles from the DEML enrichments (see **Table 9**) as small molecules. In comparison to the macrocycle validation synthesis (see **Figure 48**) we eliminated the DNA attachment site. We started with the nosylethylenediamine (**2-3**) coupling with iodobenzoic acid and subsequent DE-1 cross coupling reactions (Suzuki-Miyaura and Negishi couplings). Following the established deprotection and amide coupling procedures we introduced the trifunctional linker **2-5** along with a set of six different amino acid building blocks. During these amide couplings we encountered 10-20% epimerization of the azidoethyl sidechain of the trifunctional linker. In the case of **AA118** (compound **2-22d**) we found a mixture of four diastereomers in a 7:36:42:15 ratio. **AA118** was not available as an enantiomerically pure compound, just as a racemate, which explains the generation of four diastereomers at an estimated epimerization of approximately 10%.

The Nosyl deprotection was performed with thiophenol and subsequent ester hydrolysis. The macrocyclization turned out to be the most difficult step and seemed dependent on the rigidity of the precursor. The employed DE-2 elements were all very similar in structure (α - and β -amino acids), so the differences mostly arose from the DE-1 building block shapes and the linear precursor chemical properties (e.g. solubility). While the flexible alkyl based DE-1 elements (**NP12** and **NP14**) gave medium yields of 40-70%, the more rigid aryl based DE-1 elements (**NP01** and **NP07**) delivered rather low yields (<35%). Dimerization was not a substantial problem since the reactions were performed at concentrations of 1-2 mM. However, macrocycles **2-18** and **2-23f** in particular (both contained **NP01**) caused separation problems in combination with reduced solubility. **2-18** had to be purified by reversed phase and normal phase column separations to obtain pure material. **2-23f** reacted under the same conditions as the rest of the macrocycles to give a larger set of (side) products. Isolation in pure form was not possible at this step, but could be achieved after the subsequent click reaction. In order to generate pure material for binding assays, we had to resynthesize **2-23f** in a second batch from leftover material of the

precursor (**2-23e**) on a small scale with slightly altered reaction conditions (89% product were isolated with this approach).

Click reactions generally performed well with medium to high yields. Only the synthesis of **2-23** was low yielding (35%) due to the removal of a high amount of undesired side products from the previous macrocyclization step. All six chosen macrocycles were successfully synthesized along with their macrocyclic precursors, yielding a total of 11 macrocycles for binding affinity measurements. Every compound (except from **2-18** and **2-23**; isolated yield: 5 and 37 mg, respectively) was obtained in amounts ranging from 60 to several hundred milligrams of pure material.

10.2 Differential Scanning Fluorimetry (DSF)

Introduction

For the quantitative measurement of ligand-protein binding interactions, biophysical methods such as isothermal titration calorimetry (ITC) are inevitable. However, some of these methods may become very laborious and time-consuming for larger numbers of ligands and proteins that need to be assayed. Differential scanning fluorimetry (DSF) is a biophysical method that is based on the thermal denaturation of a protein. Protein binders are assumed to stabilize the three-dimensional structure of the protein, which results in a shift of the melting temperature of the protein (thermal shift) compared to its unbound state. The melting temperature is defined as the halfway transition point during protein denaturation.^[229]

DSF is a semi-quantitative method that usually does not deliver exact binding constants, but depending on the ligand affinity, different thermal shifts can be observed within a series of ligands versus a single protein. In general, the higher the ligand affinity, the more intensive the thermal shift. The most important prerequisite for DSF measurements is, that ligands still show good affinity at the melting temperature of the unbound protein. Ligands that stabilize the folded state of the protein give shifts to higher melting temperatures, whereas ligands that preferentially bind to unfolded states (destabilize) yield negative thermal shifts. Nonetheless, destabilizing ligands are still interesting because they might also bind to the folded form of the protein or a pseudo-unfolded form (exposure of some hydrophobic protein areas), which are masked by the higher (negative) thermal shift at the melting point.

In practice, this technique is simple and allows high-throughput parallel screening with a low consumption of resources. The thermal shifts are measured by the detection of fluorescence emission changes of the protein. This can be either UV excitation of aromatic amino acids (Trp, Tyr, Phe) enabling the fluorescence detection, or alternatively the addition of a fluorescent dye that binds to the protein. In any case, the structural rearrangement of the protein during denaturation exposes hydrophobic regions to the aqueous environment and induces a change in the fluorescence properties of the amino acids or the dye.^[229] This technique has been successfully applied to the parallel screening of ligand collections to a set of proteins^[230,231] with even in-cell measurements being possible.^[232]

DSF was chosen to be the first step for the quantitative binding assessment of our identified potential ligands from **Table 9**. To the best of our knowledge, this is the first time that DSF has been used for ligand binding assays for molecules, identified from DECL-protein assays.

Results

The six potential hits and their macrocyclic precursors (missing DE-3, five compounds) were tested for thermal shifts with AGP and HSA. In contrast to the results from the DEML selections (two HSA binders and four AGP binders) we assayed all macrocycles versus both proteins in triplicate measurements.

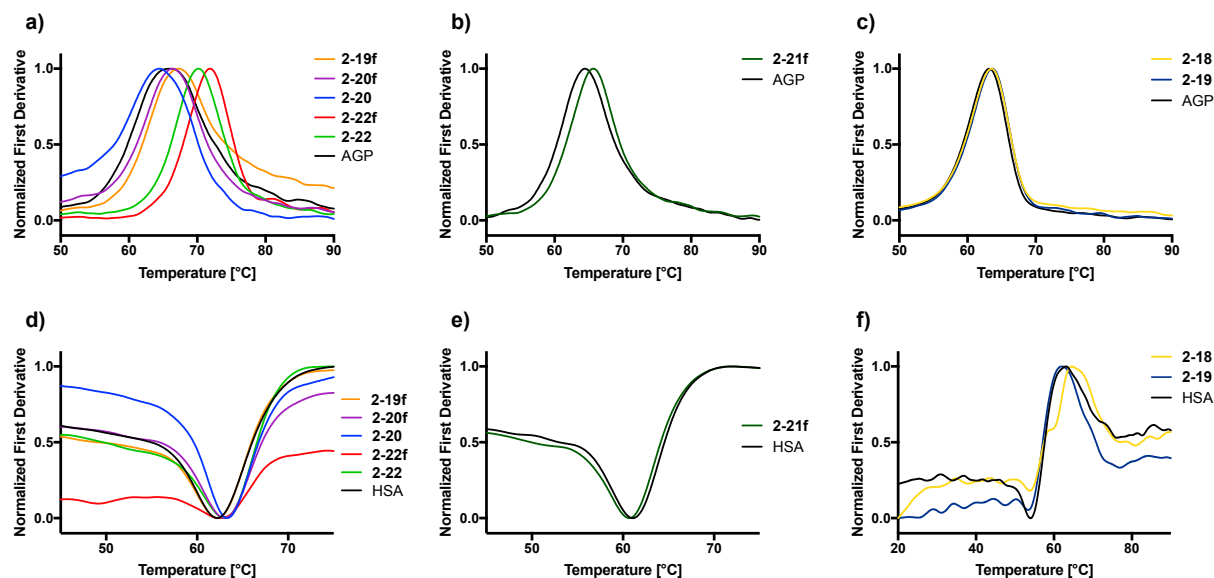


Figure 75. Normalized first derivatives of the DSF measurements. Compounds **2-23** and **2-23f** are not shown as they yielded no thermal shifts because of high insolubility. **a)** Measurements of AGP with MC ligands in PBS buffer. **b)** Measurement of AGP with MC ligand **2-21f** in 5% DMSO in PBS. **c)** Measurements of AGP with MC ligands in 25% DMSO in PBS. **d)** Measurements of HSA with MC ligands in PBS buffer. **e)** Measurement of HSA with MC ligand **2-21f** in 5% DMSO in PBS. **f)** Measurements of AGP with MC ligands in 25% DMSO in PBS.⁹

Thermal shift data for our synthesized macrocycles is presented in **Figure 75** as normalized first derivative curves. The protein denaturation halfway transition point was determined by the first derivative of the melting curve. Minima/maxima of these first derivatives show the melting temperature of the protein-ligand complexes. Five macrocycles showed good solubility in PBS and could be measured without the need for organic solvents. In the AGP assays especially MC **2-22** and its precursor **2-22f** showed large thermal shifts. The other potential AGP binders **2-19**, **2-20** as well as **2-21f** showed much smaller shifts, as was the case for the corresponding precursors (see **Figure 75** and **Table 10**). Macrocycle **2-21** could not be measured due to its very fluorescent DE-3 moiety that interfered with the thermal shift assay. Interestingly, **2-20** displayed a negative thermal shift of -1.2°C that implies a protein destabilization (preferred binding to the unfolded or a pseudo-unfolded protein form). AGP without a ligand in pure PBS showed a broadened melting curve first derivative. Addition of 5% DMSO improved the curve shape, however, there were no significant changes in the measured thermal shifts (**Table 10**). As expected from the DEML affinity assays macrocycle **2-18** showed no binding to AGP.

In the HSA binding assays none of the tested macrocycles showed strong binding that resulted in a large thermal shift ($>3^{\circ}\text{C}$). The potential HSA binder **2-18** showed a melting shift with a high difference (9.5°C), however, this value was not trustworthy due to the bad shape of the melting curve and its first derivative. The other binder from the DEML-HSA assay, **2-23** and its precursor **2-23f**, were not measurable because the compounds were too insoluble to achieve the necessary concentrations. As a consequence the thermal shift was not determined for these molecules.

⁹ Normalized thermal shift plots were generated by Dr. Timothy Sharpe from the Biozentrum, University of Basel.

Table 10. Measured ΔT_m shifts during macrocycle-protein binding events. Positive ΔT_m values imply stabilization of the protein structure, negative ΔT_m values imply a destabilization of the protein structure. Macrocycles **2-23f** and **2-23** are not shown in this analysis due to their poor solubility, which resulted in 0°C shift of the protein melting temperature.^h

Compound	ΔT_m for AGP [°C]	ΔT_m for HSA [°C]	% DMSO	Comments
2-18	0.32	9.47 ^[a]	25	Bad peak shape
2-19	0.32	8.65 ^[a]	25	Bad peak shape
2-19f	1.37	-0.07	0	
2-19f ^[b]	1.38	-	5	
2-20	-1.21	1.27	0	
2-20 ^[b]	-0.78	-	5	
2-20f	0.49	0.83	0	
2-20f ^[b]	0.37	-	5	
2-21f	1.16	-0.44	5	
2-22	4.32	0.67	0	AGP Binder
2-22f	5.94	N/A ^[c]	0	AGP Binder

[a] Unreliable data. Curve shape is too unsteady for accurate melting temperature shift measurements. [b] Due to a broadened peak shape of pure AGP in PBS the assay was repeated in 5% DMSO in PBS which showed a sharper melting peak. [c] Peak is too broad to determine a ΔT_m value.

From the found temperature shifts we chose **2-20**, **2-21** (no DSF measurements possible), **2-22** and **2-22f** for ITC measurements with AGP and **2-18**, **2-20** and **2-22** for ITC measurements with HSA.

10.3 Isothermal Titration Calorimetry (ITC)

Introduction

Isothermal titration calorimetry is a very important method for the direct measurement of a reaction's thermodynamic parameters in solution. It is most widely employed to study macromolecular processes such as protein-protein, DNA-protein or antibody-antigen interactions as well as affinity determinations of small molecules with macromolecular systems. In a single experiment the association constant, binding enthalpy and stoichiometry are measured.^[233] The setup of an ITC machine is based on measuring the power input changes caused by enthalpy changes during the reactions in solution while maintaining a constant reaction temperature. An ITC machine consists of two identical cells made of highly thermal conductive materials with an adiabatic jacket. Sensitive thermophile/thermocouple circuits detect any temperature changes between the two cells. One cell is used as reference cell and is filled with buffer whereas the second cell contains the sample. The titrant is slowly added to the sample cell by a syringe with gentle mixing. Depending on the thermodynamics of the reaction, heating or cooling of the sample cell is required to maintain a constant temperature in both cells. The required time-dependent power input is the actual measure. For exothermic reactions the power input has to be lowered and for endothermic reactions an increased power input is necessary. The heat changes

^h Data was generated by Dr. Timothy Sharpe from the Biozentrum, University of Basel during thermal shift plot generation.

are directly proportional to the fraction of bound ligand. Over time a saturation curve is obtained that determines the thermodynamic parameters. It is important to accurately measure the protein and titrant concentrations for precise determination of the parameters. That is why reference titrations, lacking the macromolecule in the sample cell, are vital to identify the heat of dilution portion, which is then subtracted from the actual titration.^[233] ITC measurements are not ideal for high-throughput assays because of the required large amounts of protein and ligand, along with a time-consuming measurement. Moreover, the assay is very sensitive to slightest changes in buffer composition and environmental influences. In comparison to DSF, the parallelization of ITC measurements is much more difficult.

Results

From the semi-quantitative protein binding assays we first started with the exact measurement of the potential AGP binders **2-22** and **2-22f** by ITC. Both compounds directly showed a good binding affinity in the one digit micromolar range (**Figure 76**). While **2-22f** showed an exothermic binding, which revealed an enthalpy driven (negative ΔH value) interaction, the bigger macrocycle **2-22** bound the protein in an endothermic process, implying entropy change as the main driving force for the protein-ligand complexation (positive ΔH value). To obtain an appropriate signal for the differential power changes with **2-22** we lowered the measurement temperature to 10°C, whereas the assays with **2-22f** could be easily conducted at 25°C. These measurements clearly showed that the DSF assay correctly predicted good binding to the proteins with a stronger binding of **2-22f** (thermal shift 5.9°C and $K_D = 4.1 \mu\text{M}$) than **2-22** (thermal shift 4.3°C and $K_D = 7.0 \mu\text{M}$). Since both compounds contained the same ring structure, we concluded that the macrocyclic core was responsible for the protein binding, not the sidechain diversity element.

The question, why we did not find **2-22f** as a hit in the smDEML/AGP enrichment assays came up at this point. This was a clear false negative result from the smDEML enrichments. And why did we not find a series of macrocycles with the **2-22f** scaffold to show up as binders in the DEML enrichments? To answer the later question, we can only make assumptions. The largest difference between the DEML selections and the binding constant evaluations was the attached DNA strand. In the DEML selections, DNA is considered to have no or only very low interactions with the protein due to the addition of salmon sperm DNA. However, the DNA represents a much bigger part of the encoded molecule than the actual ligand. Steric interferences of the DNA with the protein might reduce the binding affinity. Furthermore, interactions of the macrocycle with the oligonucleotide tail could lead to a coiled structure that shields the less polar macrocycle from the aqueous environment by interactions with the nucleobases. This is a possible explanation why **2-22f** did not show up as a binder. In contrast, **2-22** possesses a sidechain with a tertiary amine that is protonated at physiological pH, which makes the compound more water-soluble and perhaps more accessible for protein binding. However, such an assumption does not explain why we only found singleton hits in the DEML enrichment assays.

In **Figure 73** we showed that the 2-chlorophenyl- β -amino acid was the most abundant DE-2 element in the top 100 enriched binders of the AGP assay. Therefore, we can state that this element is most probably responsible for AGP binding. However, it cannot be the only part of the molecule that causes the binding affinity to AGP, otherwise we would find a larger series of molecules containing the **AA118** element among the top hits. Supposedly, the interplay between certain diversity element building blocks combinations may influence the binding affinity. All these findings and conclusions complicate a careful discussion of the results along with suspected imperfect DEML enrichments and DNA sequencing. Finally, we also found that only about 50% of the AGP were involved in ligand binding (molar ratios in both measurements approximately 0.5). This is an indication that only certain isoforms of AGP bind the ligands.^[188]

The Neri group has published several reports about other, small molecule binders from encoded libraries that showed single-digit nanomolar binding affinities. However, their binding affinities were determined by on-DNA fluorescence polarization measurements, which usually give much stronger binding than confirmation by ITC. Their DNA-free small molecule AGP binders showed single-digit micromolar affinities, comparable to our hits. In the case of their macrocyclic binders, even on-DNA fluorescence polarization only showed single-digit micromolar binding affinities to AGP.^[137,142,177]

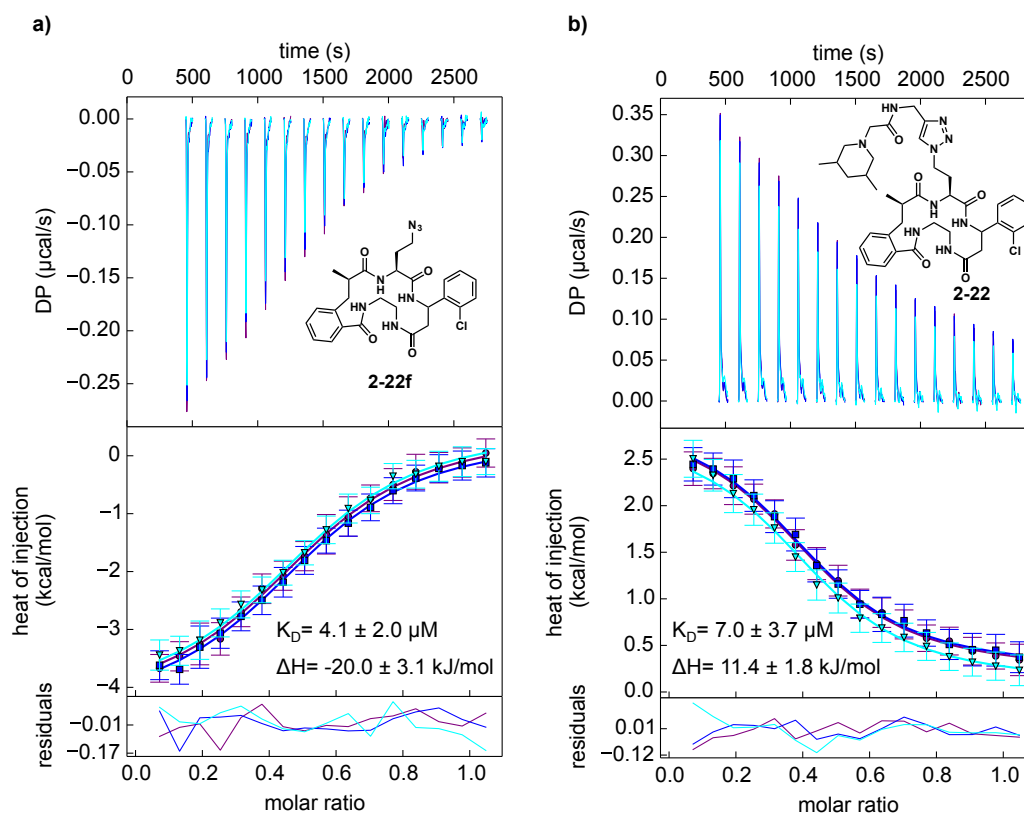


Figure 76. **a)** Overlay of the triplicate ITC measurements of **2-22f** versus AGP after data processing with heat of dilution subtraction. The binding affinity constant K_D and the binding enthalpy ΔH were calculated after data fitting with specialized software. **b)** Overlay of the triplicate ITC measurements of **2-22** versus AGP after data processing with heat of dilution subtraction. The binding affinity constant K_D and the binding enthalpy ΔH were calculated after data fitting with specialized software.ⁱ

We conducted further ITC measurements with **2-20** and **2-21** versus AGP but none of these compounds showed strong binding to the protein. We did not exactly evaluate any binding constants, but from the behavior during ITC measurements we could estimate a binding constant of $>100 \mu\text{M}$ for **2-21** and no detectable binding for **2-20** (see **Figure 104**). All assayed HSA binders showed weak binding interactions with binding constants in the several hundred micromolar range (see **Figure 105**). We did not further optimize these assays.

We clearly showed that the majority of the identified binders from the DEML enrichments were false positives that actually possessed rather weak or even no binding affinity to AGP and HSA. Only macrocycle **2-22** showed binding to AGP in all of the used affinity assays. Furthermore, we found **2-22f** to possess even better binding properties for AGP despite little enrichment in the smDEML selection, a clear false negative result.

ⁱ Graph generation, data fitting and parameter determinations were conducted by Dr. Timothy Sharpe from the Biozentrum, University of Basel.

11. Conclusions and Perspective

We have developed and assembled a DNA-encoded macrocycle library based on the incorporation of natural product-derived building blocks in a split-and-pool synthetic procedure. The library design focused on the generation of highly diverse macrocycle scaffolds that covered a broad area of chemical space in terms of structure, ring size and polarity. The scaffold ring composed of two sets of building blocks along with a trifunctional linker and the basic core, responsible for attachment of the DNA strand and macrocyclization. The first set of diversifying elements was designed and individually synthesized to incorporate natural product motifs. The majority of these elements contained alkyl chains, polyolefin moieties and heteroaromatic parts. Polyketide-like structures were not included due to the disproportional synthetic effort to create such elements.

The second set of building blocks contained highly diverse amino acids. Along with the common *D*- and *L*-amino acids, we incorporated unnatural α - and β -amino acids with alkyl, aryl and heteroaromatic substituents. For a high structural diversity we also included diverse olefinic and cyclic amino acids, which introduced a certain degree of conformational rigidity due to bridged and spirocyclic structures. Prior to the incorporation of the DE-2 building blocks we conducted a thorough compound and conditions screening in which the low yielding elements were removed from the compounds set. We generated $21 \times 102 = 2'142$ linear macrocycle precursors from those building block collections. The trifunctional linker that was integrated to connect all three diversity elements, influenced many of the reactions and required careful design. Due to synthetic difficulties as well as undesired elimination processes we were forced to include a C_2 spacing unit between the macrocycle scaffold and the DE-3 attachment site. The azide moiety of the trifunctional linker turned out to be the critical spot because it caused the formation of a terminal amide side product under certain reaction conditions.

To the best of our knowledge we are the first to successfully use the nosyl protecting group in DNA-encoded chemical library synthesis. Nosyl deprotection was a vital step to generate the reactive amine essential for macrocyclization. While nosyl deprotection efficiency was strongly dependent on the reaction conditions and the structure of the DE-2 building blocks, the macrocyclization step turned out to be very efficient with all individually tested compounds. The final diversification of the sub-library was achieved by click (CuAAC) reaction with a collection of 663 diverse terminal alkynes to generate the 1'420'146 member library (DEML). From the building block screenings, we concluded that most effort must be taken in design and evaluation of the trifunctional linker along with the amino acids collection during DE-2 amide coupling and nosyl deprotection.

The library was encoded consecutively after each diversifying chemical step. Every building block was encoded with a unique oligonucleotide "barcode". We chose a polymerization-restriction-ligation strategy to modify the DNA tag in a consecutive manner. The selected BamHI restriction process turned out to be the bottleneck in this encoding strategy, since the restriction did not proceed to completion, thereby generating some left-over DNA. Library purification turned out to be particularly tricky because of the co-elution of undesired oligonucleotide strands.

We performed affinity selections of the DEML and the precursor library (lack of DE-3, smDEML) with AGP, HSA and CA9. While we found several singleton hits for AGP and HSA that were promising for resynthesis, the CA9 selections were very unreliable due to supposed interactions of the macrocycles with the nickel beads. We finally selected two compounds for affinity measurements with HSA and four compounds for AGP. To the best of our knowledge, we were the first to use differential scanning fluorimetry (DSF) for semi-quantitative protein affinity measurements with potential binders from DECL selections. We treated both proteins with all the resynthesized compounds to screen for potential false positive and false negative hits. The

sidechain and macrocycle ring influences on protein binding were evaluated by inclusion of the precursor macrocycles (without DE-3). From the thermal shift results we identified one macrocycle and its precursor as potentially good binders to AGP. Isothermal titration calorimetry (ITC) measurements yielded binding constants (K_D) of 4.1 μM (small macrocycle) and 7.0 μM (large macrocycle). While all other tested ligands showed very weak or no binding affinity (false positive results from the DEML enrichments), discovery of the small macrocycle's high AGP binding affinity represented a clear false negative smDEML selection result. The DNA tag potentially changes or even inhibits protein binding, which might be a drawback of the encoded libraries selection procedure. The newly discovered AGP binders hold a high potential for further optimizations to increase protein affinity.

The structure of our library was developed in a way to further modify and diversify the macrocycles for future attempts. In order to improve hit rates in enrichment assays, the structures should display a high similarity to natural macrocycles. In **Figure 77** we show potential modification sites and new building block types that might be incorporated. The first set of building blocks (DE-1, ① in **Figure 77**) has the highest potential to create new macrocyclic scaffolds. Increasing the number of building blocks in this diversity element seems to be a simple task, because many boronic acids are commercially available. However, bifunctional boronic acids, displaying natural product features, are rare and their creation might require laborious procedures. Polyolefinic moieties directly attached to the basic scaffold would be highly desirable, a task that could not be achieved in the first DEML generation. Even more interesting would be the use of polyketide parts in the library (see structures in ①). Such building blocks are very common in natural macrocycles, but their synthesis in high purity and well-defined stereochemistry is extremely challenging. New methods to have quick access to such complex carbon frameworks would initiate the construction of libraries with chemical features unknown at present.^[234,235]

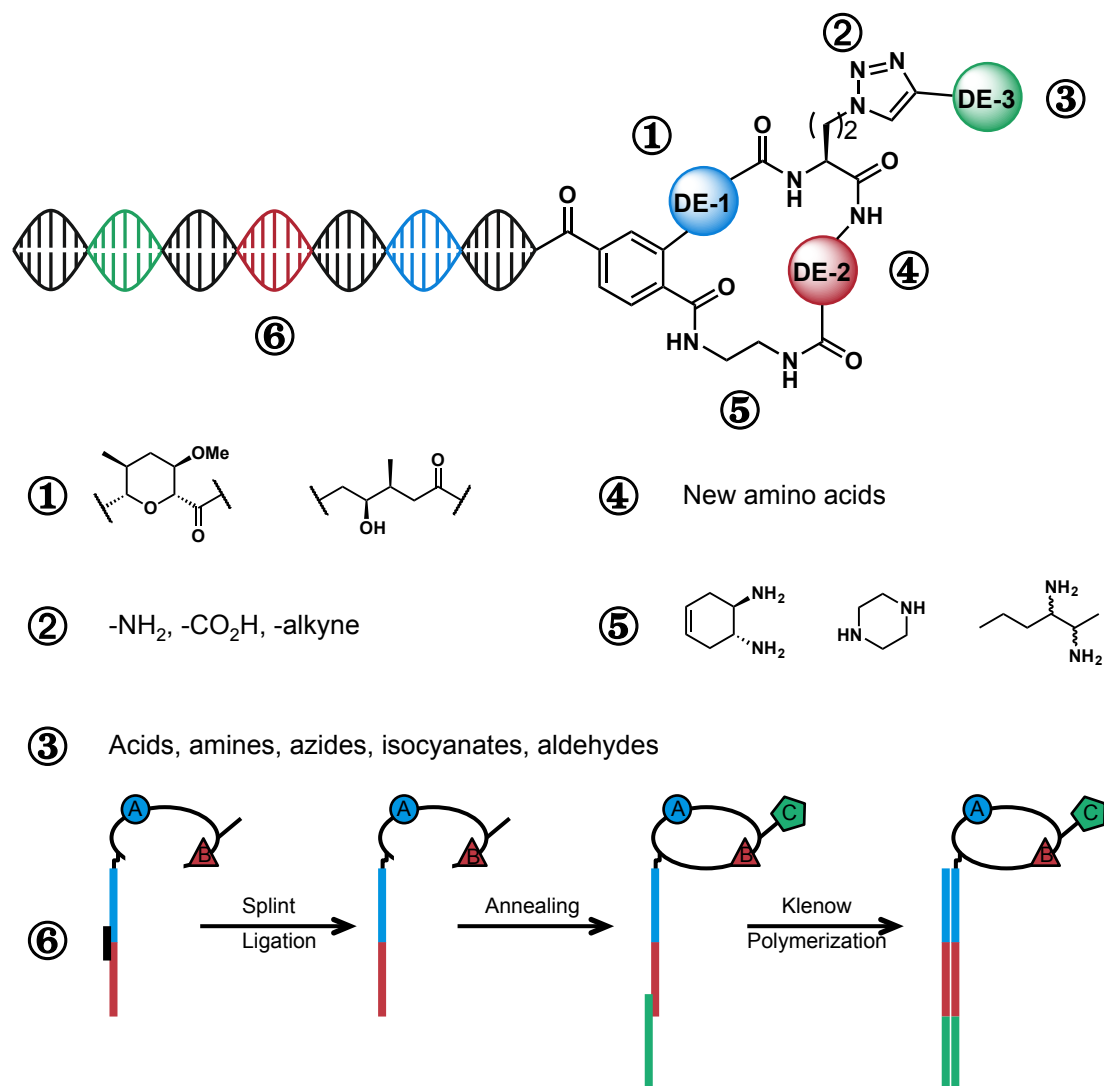


Figure 77. Schematic overview of the potential modification and improvement sites. New building blocks might be designed, synthesized and evaluated along with an improved encoding strategy.

As we have realized from the DEML construction, the trifunctional linker (2) played a very important role during the chemical synthesis of our library. The azide seemed to be responsible for the increased amount of terminal amide side product. Exchanging the azide with an alkyne overcomes this problem as we have shown in separate assays. Even though, the azide may also be considered as an amine protecting group, not as the actual functional substituent. Staudinger reduction, which is compatible with aqueous conditions, reliably yields the desired amine that is then further modified with carboxylic acids (amide coupling), isocyanates (ureas) or aldehydes (reductive amination, 3). A very simple approach might be the use of a carboxylic acid trifunctional linker (e.g. aspartic acid) to take advantage of the very large set of available amine and amino acid building blocks. In any case, the trifunctional linker is a key structure that needs a thorough consideration and design, but might also act as a diversity element itself.

Diversity element 2 (DE-2) consisted of a collection of amino acids, a type of molecule that is readily available from commercial sources in large numbers or can often be synthesized in a few steps (4). Uninteresting compounds should be exchanged for more fancy building blocks, featuring novel structural moieties (spirocyclic parts, branched or bridged structures with chiral centers) or functionalities (sulfonamides, imides, heterocycles, selenium groups). Key to the

successful incorporation of novel amino acids into the library is a thorough assessment of the coupling efficiencies of the building blocks and their influence on the succeeding chemical steps.

In our library design we used an ethylene diamine spacer for macrocyclization (Ⓔ). Many diamines are commercially available such as cyclic primary and secondary diamines as well as linear compounds with chiral centers. This element could be used to generate several individual libraries in which the influence of this part of the macrocycle is investigated. However, an even more fruitful approach might be to consider this linker as a third ring diversity element. The synthetic effort to create all combinations of the 21 (or more) DE-1 elements with all diamines and subsequent coupling to the DNA strands (as we did it for the DEML synthesis with only one diamine linker) would exceed a reasonable timeframe for DECL construction. Incorporation of the diamines set would be performed in a combinatorial step with subsequent encoding. This goal is achieved if the terephthalate core remained on one site as a methyl ester during *t*Bu-protected trifunctional linker coupling. Methyl ester hydrolysis, followed by diamine incorporation and encoding, would yield the desired products (see **Figure 78**). Careful temperature and pH control during deprotection makes the terephthalate core's methyl ester and the trifunctional linker's *tert*-butyl ester orthogonal.^[193]

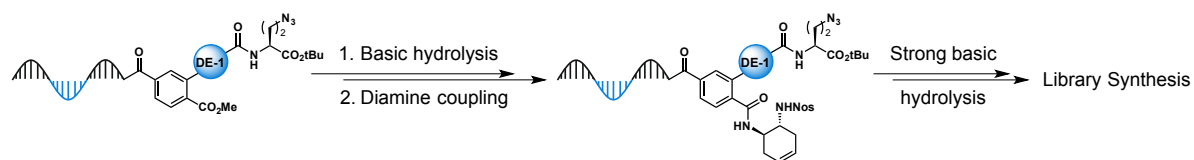


Figure 78. Proposed synthetic strategy for the incorporation of a set of diamine building blocks. Temperature and pH control make the two different esters orthogonal. Encoding steps were omitted for simplicity.

The last improvement of the DEML strategy should be applied to the encoding procedure. As we have seen, the restriction digest of double-stranded DNA yielded unsatisfactory results, which produced undesired DNA impurities in the final encoded library. We propose to use splint ligation and Klenow polymerization strategies for the encoding of future libraries (Ⓕ). Splint ligations are repeated for the encoding of several subsequent diversifications. Klenow fill-in represents the encoding step of choice for the last diversification element.

Design and validation of a DNA-encoded library are the foundation for its successful construction. Nevertheless, there are three main considerations to render a DECL valuable for drug discovery. Library size matters! Large libraries are desirable since they can cover a broad range of chemical space. However, large libraries ($>10^9$ members) come with the bias that many theoretical building block combinations are not generated during the library assembly. Furthermore, at some point the number of molecules per library member drops to a level, where an efficient enrichment with the protein cannot be achieved anymore. To overcome this issue, larger amounts of library need to be synthesized, which quickly becomes impractical and higher library concentrations are necessary for protein selections with the risk of more random DNA-protein interactions. We think that the ideal library size ranges from 10^5 to 10^6 members in focused libraries. Secondly, prior to the construction of a library or the incorporation of new building blocks, *in silico* properties calculations should be performed to check whether the generated set of compounds fulfills the required physicochemical properties (see **Table 8**).

The last consideration should include strategies for library purification. Left-over DNA influences the quality of the protein selections and may generate higher rates of false positives. Chromatographic procedures are inevitable for good library quality, especially size-exclusion based columns hold promise for excess splint or encoding DNA removal.

12. DNA-Encoded Small Molecule Library

12.1 Introduction

From the lessons learned during the design, synthesis, evaluation and testing of the encoded macrocycle library we intended to create a novel small molecule library based on the synthesized and acquired building blocks and DNA strands from the previous library. We tried to keep the library design simple, in principle a connection of two different building blocks. The novelty in this approach was to use building blocks off-the-shelf with different reactive groups that were attached to the same scaffold modification site with a versatile substituent (see purple circle in **Figure 79**).

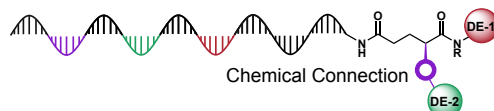


Figure 79. Schematic representation of the small molecule library design. Two diversity elements were connected by a simple trifunctional linker with the DNA strand. The purple circle stands for a series of chemical connection possibilities to attach different DE-2 sub-libraries. This connection scheme is encoded by a third DNA codon (indexing).

An azido group seemed best suited as versatile functional group for this library, since it undergoes click reactions with our alkyne building block library, and can be transformed to an amine, which reacts with several electrophilic moieties (e.g. isothiocyanates, aldehydes). We also introduced a new concept that we called reaction indexing. The notion behind this method is to use a set of DNA strands to encode different reactions with different building blocks. Conventional encoding uses X DNA strands to encode X building blocks. With our reaction indexing strategy, we use $n+X$ DNA strands to encode $n \cdot X$ ($n > 1$) building blocks. This is achieved by additional DNA incorporation after encoding of the building block (see purple DNA codon in **Figure 79**). For example, 300 DNA strands can encode sub-libraries of 300 alkynes, 300 carboxylic acids and 300 isocyanates. All these types of molecules react with an azide or the amine, generated by azide reduction. To distinguish the connection type (triazole, amide, thiourea), three additional coding DNA strands (indexing codons) are incorporated after building block encoding. The different building blocks and their connection types are unambiguously reidentified by the combination of building block codon and indexing codon. With this method, only 303 unique DNA strands are required to encode 900 building blocks that were attached via three different chemical transformations. An example of this indexing strategy is outlined in **Figure 80c** with two building block types (alkynes and carboxylic acids) at the same diversification spot.

12.2 Design and Assembly of the Small Molecule Library

Design

The detailed design and synthesis overview of the small molecule library is depicted in **Figure 80**. The scaffold was synthesized from *tert*-butyl modified glutamic acid, which was transformed to the desired azido trifunctional linker in three consecutive steps. After DNA attachment, we diversified the linker (**2-26**) with two sets of building blocks. The first set consisted of amino acids and amines (primary and secondary). We used in total 150 compounds for this diversification site. The second set of building blocks consisted of 664 terminal alkynes and of 89 carboxylic acids. The alkynes were introduced by copper-catalyzed click reaction (CuAAC) and the carboxylic acids were incorporated by amide coupling after Staudinger reduction of the azide to the amine. Encoding of the library was performed by splint mediated T4 ligation (see **Figure 77**, © **first step** for a schematic description of this encoding procedure). The two different DE-2 building blocks

sub-libraries were further encoded by indexing, to use the same codon DNA strands for different building block types. In principle the index codon represents the connection type of the DE-2 element (triazole or amide bond). The size of the final library was $150 \times 753 = 112'750$ members (**Figure 80b**). We employed an encoding strategy that resulted in a purely single-stranded DNA-encoded library. This left the possibilities of combining this library with other libraries and potential further modification or application in an interaction-dependent PCR protein screening assay (description of this technique in Chapter 6.4).

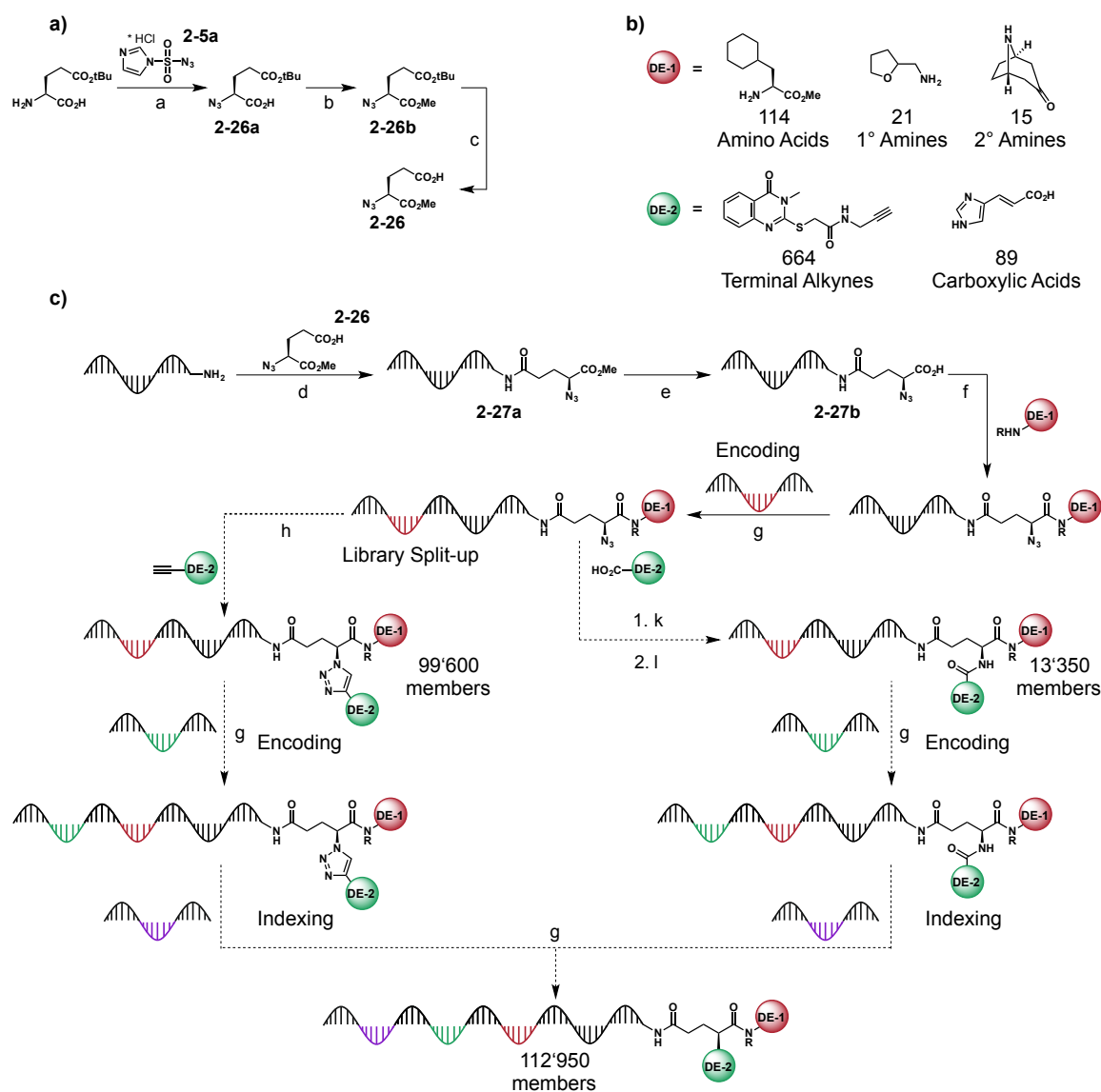


Figure 80. **a)** Chemical synthesis of the trifunctional linker to connect the DNA strand with the two diversity elements. **b)** Examples of the used building blocks. For DE-1 we employed amino acids, primary and secondary amines. For the DE-2 we used alkynes and carboxylic acids. **c)** Synthetic overview of the chemical synthesis of the library and the encoding steps. Dashed arrows indicate reaction steps that had not yet been accomplished when the author stopped working on this project. Reaction conditions: a) CuSO_4 pentahydrate, K_2CO_3 , MeOH, RT, 2 h, **98%**. b) TMS-diazomethane, MeOH, RT, 17 h, **95%**. c) Formic acid, RT, 17 h, **78%**. d) EDC hydrochloride, HOAt, DIPEA, MOPS buffer, DMSO, RT, 22 h. e) LiOH monohydrate, H_2O , MeCN, RT, 5 h. f) DMTMM- BF_4 , NMM, MOPS buffer, DMSO, RT, 20 h or EDC hydrochloride, HOAt, DIPEA, MOPS buffer, DMSO, RT, 20 h. g) Splint DNA, T4 PNK, T4 DNA ligase, 10X ligase buffer, H_2O , ATP. h) Cu-TBTA, NaOAsc, TEAA buffer, DMSO, RT, 21 h. k) TCEP, TEAA buffer, RT, 16 h. l) EDC hydrochloride, HOAt, DIPEA, MOPS buffer, DMSO, RT, 17 h.

Scaffold Synthesis and DNA Attachment

We synthesized the scaffold for the small molecule library from H-Glu(OtBu)-OH in three steps with high yields (**Figure 80a**). In the first transformation we used copper-catalyzed amine-azide transfer for the introduction of the azido group at the α -position of the glutamic ester (**2-26a**).^[236] After methyl ester formation with TMS-diazomethane (**2-26b**) the *tert*-butyl ester was selectively deprotected under acidic conditions to yield the final scaffold linker **2-26** with the free γ -carboxylic acid. A single 5'-hexylamine modified DNA strand was attached by EDC/HOAt mediated amide coupling. We designed the sequence of this DNA strand in a way to enable splint ligation encoding, but we also left a three base coding site (in this approach with an AAA sequence) in case we wanted to generate several libraries with the same building blocks but different scaffold spacers.

Library Validation

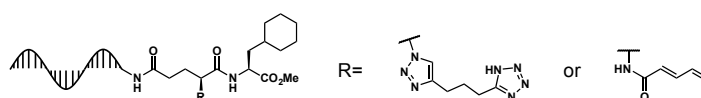


Figure 81. Validation compounds. Both DE-2 reactions, click and amide coupling, were tested prior to the assembly of the final library. Reaction conditions were according to the stated conditions in **Figure 80**.

We validated the chemistry of the small molecule library with three different building blocks. For the DE-1 modification we chose cyclohexylalanine (**AA006**) that was coupled under DMTMM coupling conditions. For the DE-2 click reaction we chose the tetrazole-alkyne **TA662** that yielded perfect conversions and high purity. We also validated that the Staudinger reduction of the azide worked well using tris(2-carboxyethyl)phosphine (TCEP) with subsequent 2,4-pentadienoic acid (**CA001**) modification under EDC/HOAt coupling conditions. In terms of DNA damaging conditions, this amide coupling step should not be more harmful to the oligonucleotide strand than any of the other reaction conditions applied (compared to our own qPCR results for DNA viability after chemical transformations in **Figure 68**).

Validation and Introduction of the DE-1 Building Blocks

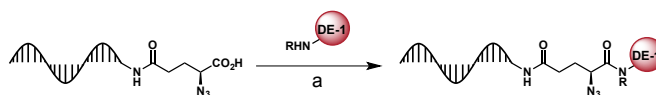


Figure 82. Introduction of the first diversity element. Reaction conditions: a) DMTMM-BF₄, NMM, MOPS buffer, DMSO, RT, 20 h or EDC hydrochloride, HOAt, DIPEA, MOPS buffer, DMSO, RT, 20 h.

After successful validation of the chemistry for the library assembly, we screened 126 amino acids, 26 primary and 26 secondary amines for their reactivity towards amide coupling with the DNA-modified linker scaffold. All building blocks were selected from existing sub-libraries and off-the-shelf compounds. The screening conditions were similar to the screening in the macrocycle library. We employed the DMTMM conditions except for the known building blocks that required the EDC/HOAt coupling conditions (tyrosine and its derivatives). We set a threshold of 80% product purity. 153 building blocks fulfilled this condition, while 25 compounds were not suitable for library assembly (**Table 14**). The majority of the building blocks showed purities of >90% (134 building blocks). Interestingly, several amino acids that were tricky or even unsuitable for the synthesis in the macrocycle library showed good to excellent conversions in this assay. In general *N*-methylated amino acids were unsuitable along with some of the special case amino acids (good leaving groups, hydroxylamine, thiazoline sidechain) known from the previous macrocycle library screening. Fortunately, homoprolines and α,α -disubstituted amino acids now showed good to excellent yields. Primary amines were generally very well suited for this

modification with the exception of *tert*-butyl amine, which appeared to be too bulky to undergo an efficient coupling. Secondary amines, on the contrary, were more difficult to couple. Cyclic secondary amines showed good conversions unless both positions *ortho* to the amine were substituted (steric hindrance, e.g. 2,6-dimethylpiperidine). Linear secondary amines were the trickiest coupling partners. Only short linear secondary amines (*N*-methyl-*N*-propargylamine, *N*-(2-hydroxyethyl)-*N*-methylamine) showed high enough conversions and purities for library build-up. From the set of suitable building blocks we chose 150 for library construction. After ester hydrolysis of the scaffold linker we introduced the building blocks by the evaluated DMTM and EDC/HOAt coupling conditions in a combinatorial split synthesis step.

Encoding of the DE-1 Building Blocks by Splint Ligation

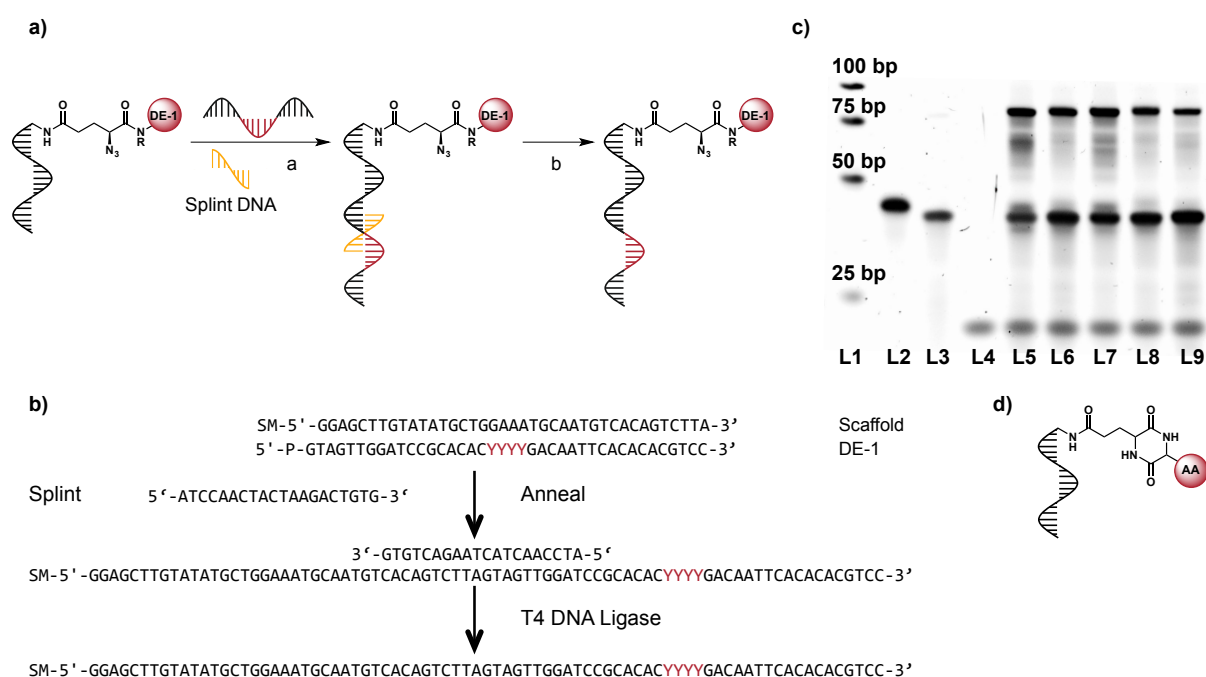


Figure 83. a) Schematic procedure of the splint ligation. Reaction conditions: a) T4 PNK, ligase buffer, ATP, 37°C, 30 min. b) T4 DNA ligase, ligase buffer, ATP, 4°C, 16 h. b) Detailed splint ligation with the exact DNA sequences. c) 12% denat. PAGE analysis, SYBR gold staining. L1: Low MW DNA ladder, L2: Scaffold DNA before DE-1 coupling & encoding, L3: sample DE-1 DNA strand, L4: Splint DNA, L5-L9: Random sampling of different encoded DE-1 reactions from the well plates. d) The identified diketopiperazine side product.

The individual amide coupling reactions were encoded by splint ligation with 5'-phosphorylated DNA strands (**Figure 83a**). We used the same oligonucleotides as we had used for the amino acid encodings in the macrocycle library. The scaffold DNA strand and the DE-1 encoding strand did not contain any complementary regions so that annealing with a third oligonucleotide (the splint) could be performed. This splint DNA was designed in a way to linearly align the two other oligonucleotides by partial annealing with both strands (for detailed DNA sequences see **Figure 83b**). The connection between the aligned coding oligonucleotides was established by DNA ligation. For that reaction, the DE-1 DNA strand required a 5'-phosphate group, which we introduced by *in situ* phosphorylation with the T4 polynucleotide kinase (PNK). We optimized the phosphorylation-ligation procedure prior to library encoding. **Figure 83c** shows the very successful encoding of a randomly selected set of amide couplings. After encoding, we pooled the 150 reactions and purified them by chromatographic separation. However, as we figured out soon after the encoding, the azido group that we needed for diversification with the terminal alkyne sub-library, had been reduced to the amine during the encoding and work-up process. Control reactions revealed that the 10 mM concentration of DTT in the commercial ligase buffer

was capable of reducing the azido group, at the α -position to the amide, to the amine, which was supported by elevated temperatures (we concentrated the library samples in the SpeedVac at 45°C or 60°C). That was in sharp contrast to the azido group of the DEML trifunctional linker, which was not located in close proximity to any functional groups. Control experiments proved that the DEML alkyl azide was not reduced under these conditions. We repeated the library construction steps and splint ligation encoding with DTT-free buffer. Optimization experiments showed that DTT was not a necessary additive for the phosphorylation or the ligation step. Finally, we ended up with two sub-libraries, one with the clickable azide and the other with an amine, that we initially intended to generate by TCEP-mediated azide reduction.

Screening of a Carboxylic Acids Sub-Library for DE-2 Introduction

For the second diversification site we aimed to introduce different types of building blocks. Next to the 664 terminal alkyne sub-library that we had used and evaluated in the macrocycle library, we envisioned to use carboxylic acids as a second set of building blocks. We assembled in total 108 diverse carboxylic acids off-the-shelf (many more would be commercially available). We used aliphatic and aromatic carboxylic acids, which allowed us to include many structural features such as branched chains, polyolefins, bridged cyclic structures and even a very large lipophilic structure with the steroid fused ring scaffold (cholic acid). We also employed many different functional groups such as alcohols, sulfonamides, ketones, substituted aromatic and heteroaromatic structures as well as nucleobase-derived carboxylic acids. For simplicity, the structures were grouped according to their structural and electronic features and 39 representative members of this collection were tested to find trends in reactivity.

Anthracene derivatives and many hydroxyl groups containing compounds were low yielding. Electron withdrawing groups (especially the nitro group) at the aromatic ring or halogenated acetic acids showed low or even no conversion. Aldehydes were unsuitable because they reacted in unwanted side reactions, most likely imine formations, which left the amine coupling partner inaccessible for the activated ester. Steric influences from large or hindered groups were observed as well, but the effect was not strongly pronounced. For example *tert*-butyl carboxylic acid and adamantyl carboxylic acid showed >80% purity. In the end we set the cut-off at >50% purity, which delivered 89 carboxylic acids. The majority of these compounds (those that were evaluated) showed >80% purity, only a small set of building blocks that gave lower yields were included. Side products were not commonly observed during the reactions and in most cases unreacted starting material was the main impurity. However, we figured out that the free amine (after azide reduction) was capable of diketopiperazine (DKP) formation with some DE-1 amino acid esters (**Figure 83d**). A side reaction that could be prevented by reduced temperatures during azide reduction. All of the DKP amount seemed to stem from the azide reduction step and not from the DE-2 introduction. The detailed list of the used building blocks is shown in **Table 15**.

Screening of Other DE-2 Building Blocks

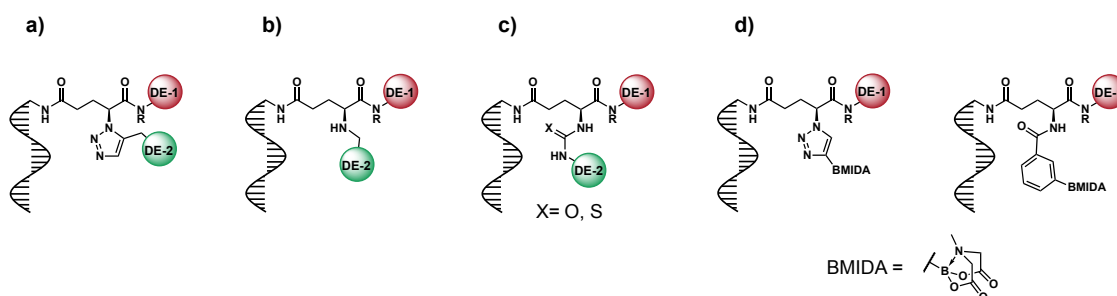


Figure 84. **a)** Nickel click reaction to yield 1,5-disubstituted triazoles. **b)** Reductive amination with aldehydes to yield secondary amines. **c)** Reaction with iso(thio)cyanates to yield (thio)ureas. **d)** Two different methods to introduce a boronic acid ester for further modifications.

The strength of our newly developed indexing strategy was the capability to incorporate different sets of building blocks using the same oligonucleotide strands. In the simplest implementation we would use our large terminal alkyne sub-library with two different connection methodologies. A report from 2017 showed that 1,5-disubstituted 1,2,3-triazoles were accessible in nickel-catalyzed cycloaddition reactions (NiAAC) under aqueous conditions.^[237] A reaction that had so far only been accessible with moisture sensitive ruthenium catalysts. Implementation of this reaction type would yield the small molecule library with the terminal alkynes attached in the conventional 1,4-disubstituted 1,2,3-triazole connection from the CuAAC as well as in the isomeric 1,5-disubstituted form from the NiAAC reaction (**Figure 84a**). We tried to implement this reaction into our library with a screening for potential condition optimizations. Unfortunately, we did not succeed in generating the desired 1,5-disubstituted products. In fact, in most attempts with various nickel precatalysts, we did not see any transformation at all. Only with the Ni-TBTA complex (analog to the Cu-TBTA complex) we found conversion of the starting material to several unidentified reaction products.

Another frequently used reaction in DECL assembly is reductive amination of aldehydes (**Figure 84b**). We tested this reaction type and found that in principle it worked, however the purities were very low. We found the desired products of the tested compounds along with a majority of unidentified side products. For this reason, we concluded that reductive amination was unsuitable for diversification of our library.

Isocyanates and isothiocyanates are not as readily available in larger collections, as is the case for the very common building blocks (amines, carboxylic acids, aldehydes, boronic acids). However, we considered this type of functional groups, because they had been used for the modification of biomolecules under aqueous conditions before (see Chapter 2.2 for an example with a protein). Unfortunately, we could not find any of the desired products, instead the diketopiperazine side product formation was strongly favored (**Figure 84c**).

From the very beginning of our research in DECL construction we envisioned to use boronic acid building blocks for the construction of encoded chemical libraries. Boronic acids are commercially available in large collections and they have been used in many reaction types for decades.^[238,239] The versatility of this functional group is unquestioned and the most important transformation still remains the Suzuki-Miyaura cross-coupling reaction. In the early stages of the DEML assembly we were not successful in employing this chemistry with the macrocycle library construction. Based on our more recent developments, we envisioned attaching a boronic acid to a DNA strand and using a set of modification reactions and compound classes to introduce diversity (**Figure 84d**). We tried to incorporate a MIDA protected boronic acid alkyne by click reaction. LC-MS analysis showed a very small amount of the desired product, however, the majority of the species was the click product with H-insertion at the boronic ester site. The copper catalyst

promoted a hydrogen insertion reaction with the carbon-boron bond. In a second attempt we explored the amide coupling with a MIDA-protected borylated benzoic acid. A small amount (approx. 10%) of the desired product was formed, whereas the remaining material underwent undesired side reactions. The incorporation of boronic acids (protected as MIDA esters) seemed to work to some extent, though the yields were not high enough for efficient library generation.

12.3 Perspective

Further construction of the small molecule library with the final diversification, encoding and subsequent protein binding assays were taken over by Basilius Sauter. For that reason the construction of this library and its application is not further discussed herein. The author showed the construction of a relatively simple library from available building blocks sets. Validation assays revealed high yields and purities with a diverse collection of compounds. Furthermore, the already available oligonucleotides were employed in a completely different encoding type compared to the macrocycle library. Only a few new DNA strands were redesigned and acquired from commercial sources. *In situ* enzymatic phosphorylation of oligonucleotides was proven to be very efficient, a factor that reduces the cost of commercial oligonucleotides. New building block types were tested, yet more reaction condition screenings and improvements are necessary to establish the use of those moieties.

Small molecule libraries with sizes of 10^5 - 10^6 members are probably the most promising attempts for drug discovery. The introduction of new building block types (such as boronic acids) will be challenging, but might change the type of libraries that will be accessible in future. Searches for new DNA-compatible reaction conditions are highly desirable to get access to a broader selection of commercially available building block collections. Our newly established index DNA encoding might vastly reduce the cost for DECL construction. Smaller numbers of oligonucleotides will be required to encode for several sets of different building blocks. In our opinion, splint ligation should be the main encoding type for the construction of this DECL type because it showed high efficiency, high versatility and low side product formation. Klenow polymerization is an additional encoding reaction of choice due to its high reliability. Traditionally, mostly double-stranded DNA-encoded libraries have been used for enrichments with protein targets. However, the development of new screening assays, such as the interaction-dependent PCR^[168], will change the way of library construction. Several purely single-stranded DNA-encoded libraries may be screened simultaneously against a collection of different DNA-tagged proteins. These libraries use the same coding DNA strands but the index codons uniquely identify every single building block from each library. After interaction-dependent PCR, binders will contain all the information about chemical structure, connection type of the building blocks (depicted by the indexing codon) and bound protein on the same oligonucleotide strand. This methodology combines the advantages of smaller libraries (high purity) with the advantages of large libraries (high diversity) in an experimentally simple assay.

13. Experimental Part DNA-Encoded Libraries

13.1 General Information

Reagents, Solvents, Oligonucleotides

See Chapter 5.1 for information about reagents and solvents. Enzymes and their buffers were purchased from New England Biolabs (NEB). Oligonucleotides were purchased from Microsynth in desalted or HPLC purified form as lyophilized material or dissolved in H₂O as 100 μM stock solutions. MOPS buffer refers to a solution of 50 mM MOPS and 500 mM NaCl in H₂O at pH 8.2.

Chromatographic Purification and Isolation

Flash chromatography was performed on SilicaFlash® gel P60 40-63 μm (230-400 mesh, SiliCycle, Quebec) according to Still^[95] or on a Biotage Isolera four with SilicaFlash® gel packed cartridges. Reversed phase flash chromatographies were run on a Biotage Isolera four with self-packed columns (LiChroprep RP-18, 40-63 μm silica from Merck). Preparative RP-HPLC was carried out on a Shimadzu Prominence UFLC Preparative Liquid Chromatograph.

Method A: Gemini NX-C₁₈, 5 μm, 110 Å, 21.2 x 250 mm from Phenomenex with a flow rate of 20 mL/min, gradient: 1% (3 min)-99% (25 min)-99% (3 min) (B), monitoring and collecting the products at 254 nm. Buffer (A): 0.1% TFA (v/v) in H₂O, Buffer (B): 0.1% TFA (v/v) in MeCN.

Method B: Gemini NX-C₁₈, 5 μm, 110 Å, 21.2 x 250 mm from Phenomenex with a flow rate of 20 mL/min, gradient: 1% (3 min)-80% (25 min)-99% (0.1 min)-99% (3 min) (B), monitoring and collecting the products at 254 nm. Buffer (A): 50 mM TEAA in H₂O, pH 7.2, Buffer (B): MeCN.

Method C: Jupiter C₄, 5 μm, 300 Å, 10 x 250 mm from Phenomenex with a flow rate of 10 mL/min, gradient: 0% (3 min)-30% (17 min)-30% (2 min) (B), monitoring and collecting the products at 254 nm. Buffer (A): 50 mM TEAA in H₂O, pH 7.2, Buffer (B): MeCN.

The crude compound mixtures were injected as H₂O, MeCN, MeOH or DMSO solutions. Buffers and HPLC eluents were prepared with nanopure water (resistivity 18.2 MΩ). Concentration under reduced pressure was performed by rotatory evaporation at 40°C water bath temperature. Aqueous product fractions were frozen in liquid N₂ and lyophilized on a Christ Alpha 2-4 LDplus flask lyophilizer at 0.3 mbar or below.

Chromatographic Analysis

Analytical TLC was performed on Silica gel 60 F254, 0.25 mm pre-coated glass plates (Merck) and visualized by fluorescence quenching under UV light at 254 nm and subsequent KMnO₄ or ninhydrin staining.

HPLC analysis was performed on an Agilent 1100 system equipped with Jupiter C₄, 5 μm, 300 Å, 2 x 50 mm or 2 x 150 mm columns with a flow rate of 0.6 ml/min. Gradients: 0% (1.8 min)-30% (3.2 min)-90% (2.2 min)-90% (1.8 min) (B) or 0% (1.8 min)-30% (12.2 min)-90% (4 min)-90% (2 min) (B), Buffer (A): 50 mM NH₄OAc in H₂O, pH 7.2, Buffer (B): MeCN.

ESI-MS and LC-MS

For information about ESI-MS and UPLC-MS measurements, see Chapter 5.1.

LC-MS spectra were recorded on a hyphenated system, consisting of the previously described Agilent 1100 HPLC and the Bruker Esquire3000 ESI-MS with a direct connection tube between the devices (no flow splitter). The ESI-MS was run at 350°C with a N₂ flow of 10.5 L/min and

35 psi pressure in positive ionization mode. Tuning ranges were 500 - 1400 or 1000 - 1800 m/z. Control software for the ESI-MS was Esquire Control and for the hyphenated system HyStar 3.1 was used.

Nuclear Magnetic Resonance (NMR)

See Chapter 5.1 for detailed information.

DNA Purification/Handling

0.2, 0.5, 1.5 and 2.0 ml tubes were centrifuged in an Eppendorf Centrifuge 5418R with a FA-45-18-11 rotor at max. speed (14'000 rpm). 5, 15 and 50 ml tubes were centrifuged in an Eppendorf Centrifuge 5804R with the S-4-72 rotor at max. speed (4'200 rpm). 96-well plates were centrifuged with an A-2-DWP rotor at max. speed (3'700 rpm). DNA and protein sample heating or cooling was performed with a BIOER Mixing Block MB-102. Vacuum centrifugation was performed with an Eppendorf Concentrator 5301. DNA and protein concentrations were measured on a Nanodrop 2000 from Thermo Scientific via absorption at 260 nm and 280 nm, respectively.

Gel Electrophoresis

Gel electrophoresis was performed with a Bio-Rad PowerPac HV high-voltage power supply. Gels were prepared with an area of 83 x 83 mm and a thickness of 1.0 or 1.5 mm with 10 or 15 wells.

Native DNA polyacrylamide gels were prepared from 40% acrylamide/bis-acrylamide 19:1 (Fisher Scientific) with TBE (TRIS-Borate-EDTA) buffer with 0.1% (v/v) TEMED (*N,N,N',N'*-Tetramethylethylene-1,2-diamine) and 0.1% (v/v) APS (25% ammonium persulfate in H₂O solution). Loading dye: Gel loading dye purple (6X), no SDS from NEB. To improve loading, the dye was used as 3X.

For denaturing DNA polyacrylamide gels the same recipe was used with the addition of urea (final concentration: 7 M). The DNA sample was treated with 2X formamide loading dye (95% formamide, 5 mM EDTA, pH 8, 0.025% (m/v) bromophenol blue) and denatured at 95°C for 2 min prior to loading onto the gel. Visualization was achieved with SYBR gold dye staining and blue LED fluorescence with a 530/28 filter.

DNA agarose gels were prepared by heat dissolving agarose (Fisher Scientific) in 1X TBE (50 ml per gel) and cooling to room temperature. SYBR gold dye for visualization was directly mixed with the warm agarose solution. Loading dye: 0.025% (m/v) bromophenol blue in 30% (v/v) glycerol in TE buffer, pH 8.

Denaturing protein gels were prepared with a 5% stacking gel (approx. 2 cm high) and a resolving gel of different percentages (approx. 6 cm). The stacking gel was prepared from 40% acrylamide/bis-acrylamide 37.5:1 (Fisher Scientific) with 0.1% (m/v) SDS, 0.2% (v/v) TEMED and 0.4% (v/v) APS (10% APS solution in H₂O) in 125 mM TRIS, pH 6.8. The resolving gel was prepared from 40% acrylamide/bis-acrylamide 37.5:1 (Fisher Scientific) with 0.1% (m/v) SDS, 0.1% (v/v) TEMED and 0.3% (v/v) APS (10% APS solution in H₂O) in 375 mM TRIS, pH 8.8. Protein samples were treated with 2X loading dye (66 mM TRIS pH 6.8, 2% (m/v) SDS, 0.01% (m/v) bromophenol blue, 30% (v/v) glycerol) and denatured at 95°C for 5 min prior to gel loading. Visualization was achieved with Coomassie Blue staining.

Running buffer for DNA gels: 1X TBE. Running buffer for protein gels: 193 mM glycine, 25 mM TRIS, 0.1% (m/v) SDS.

Gel imaging and analysis was performed on a Bio-Rad ChemiDoc MP system.

PCR and qPCR

PCR was performed in 0.2 ml PCR tubes in a Bio-Rad T100 Thermal cycler. qPCR was performed in a StepOnePlus real-time PCR system from Applied Biosystems using StepOne v2.3 software. qPCR samples were set up in 96-well plates.

Differential Scanning Fluorimetry (DSF) and Isothermal Titration Calorimetry (ITC)

Thermal denaturation experiments were performed with a Nanotemper Prometheus NT.48 instrument using standard quality capillaries. All samples were subjected to continuous ramping from 20-95°C at a rate of 1.5°C/min. The thermal denaturation was monitored by the intrinsic fluorescence emission at 330/350 nm after excitation at 285 nm. The data was processed using a beta-version of the Nanotemper PR.Analysis software and Graphpad Prism v7.

ITC experiments were performed on an ITC200 instrument from Malvern Panalytic. Protein samples were degassed prior to ligand binding assays. The ligand solution in the syringe was added stepwise to the protein solution in the sample cell at 10 or 25°C. 300 s initial delay with stirring was followed by 17 syringe injections (1 x 0.5 µL, 16 x 2.3 µL). The baseline subtraction and integration of the differential power vs. time was performed using NITPIC^[240] and data fitting was performed using Sedphat.^[241]

DNA Ethanol Precipitation

DNA samples were treated with 3 M NaOAc pH 5.2 buffer (10% of DNA sample volume) and mixed with EtOH (300-400% of DNA sample volume). The mixture was kept on ice for 2 h (unless otherwise stated). The DNA suspension was centrifuged (4°C, max. speed, 30 min) and the supernatant was discarded. The obtained pellet was washed twice with cold EtOH and centrifuged again (4°C, max. speed, 2 x 15 min). The supernatants were discarded and the washed pellet was dried in the air for 30 min. Purified pellets were redissolved in H₂O or the appropriate buffer for continued synthesis.

Cu(II)-TBTA Stock Solution, 10.0 mM in 55% DMSO

The Cu(II)-TBTA stock solution was prepared according to a published procedure for click modifications of oligonucleotides found on the Lumiprobe webpage (<https://www.lumiprobe.com/protocols/click-chemistry-dna-labeling>).

A solution of copper(II) sulfate pentahydrate (50.0 mg, 200.0 µmol 1.0 eq.) in distilled H₂O (10 mL) was mixed with a solution of TBTA (116.0 mg, 219.0 µmol, 1.1 eq.) in DMSO (11 mL). The dark blue solution was stored at room temperature for months without any observed loss of catalytic activity in copper-catalyzed click reactions.

13.2 LC-MS Analysis of DNA-tagged Small Molecules

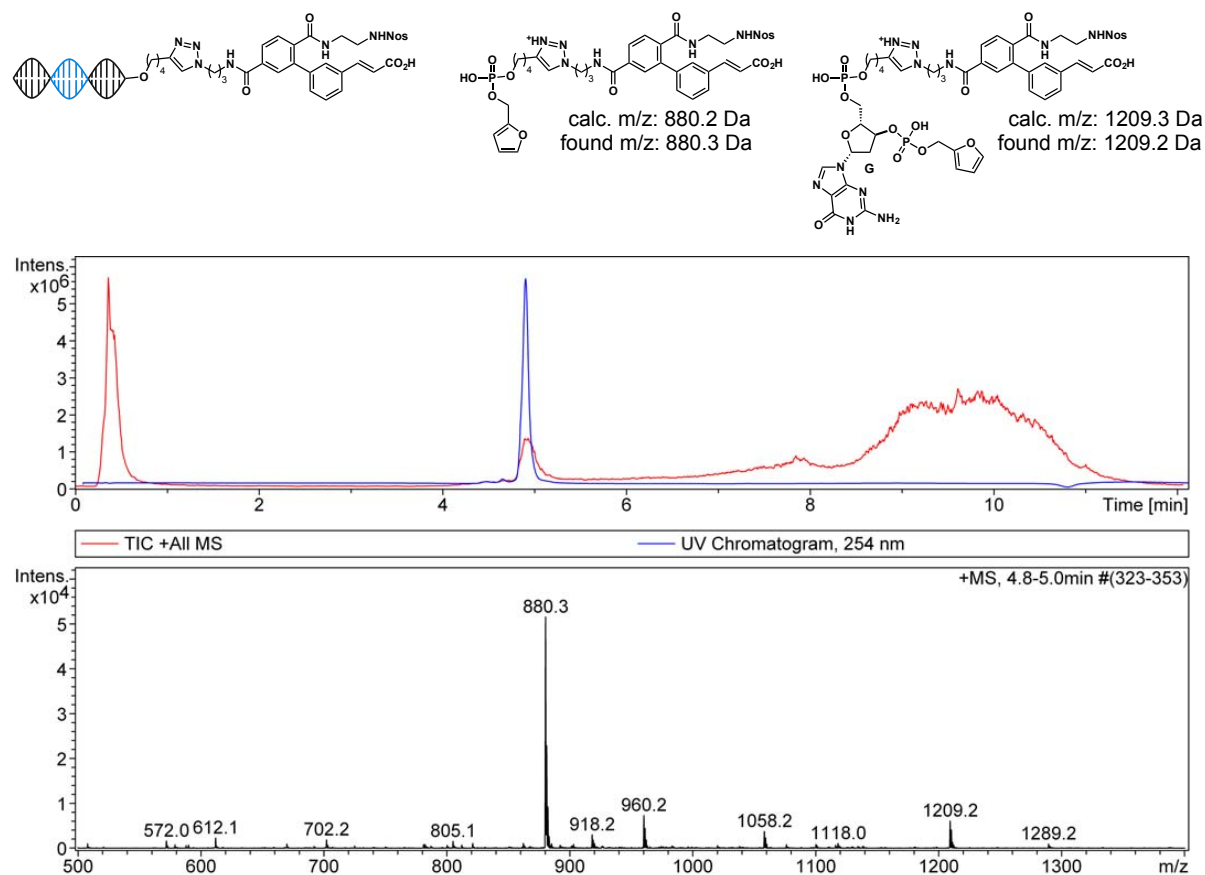
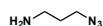


Figure 85. Example of an encoded macrocycle precursor LC-MS analysis with the found ion fragments.^[242]

Chemical modifications of DNA-tagged small molecules were analyzed with the previously described LC-MS system. DNA-encoded compounds could be analyzed due to the specific fragmentation of the molecules. Therefore, no multiply charged species needed to be deconvoluted, which made it possible to easily detect 1 Da modifications of the attached small molecule. The accuracy of this method was determined to be ≤ 0.3 Da. In **Figure 85** an example of an encoded macrocycle precursor is shown. The MS trace analysis clearly shows the proposed fragments. The major fragment consists of the small molecule with the phosphate and the ribose (now as furyl group) of the first nucleotide. The rest of the DNA strand has been eliminated during ionization.^[242] A second, less abundant fragment has the same structure, but the elimination has occurred at the second nucleotide whereby the first nucleotide (in our case G) stayed intact. Sodium ($M + Na^+$) and potassium ($M + K^+$) adducts were often found along with the protonated species.

13.3 Off-DNA Macrocycle Synthesis

Synthesis of 3-azidopropan-1-amine 2-1

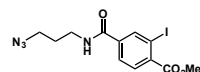


3-Bromopropan-1-amine hydrobromide (5.0 g, 22.8 mmol, 1.0 eq.) and sodium azide (4.5 g, 68.5 mmol, 3.0 eq.) were dissolved in H_2O (20 mL) and heated to $80^\circ C$ for 22 h. After cooling a solution of KOH (6.0 g, 106.9 mmol, 4.7 eq.) in H_2O (10 mL) was added while cooling in an ice bath. The solution was extracted with DCM (4 x 50 mL) and the combined organic layers were washed with brine (50 mL) and dried over Na_2SO_4 . The solvent was removed *in vacuo* ($40^\circ C$,

250 mbar) to yield the desired product as slightly yellow liquid (2.0 g, **87.4%**). Analytical data was in agreement with reported data.^[243]

¹H NMR (400 MHz, CDCl₃) δ/ppm: 3.42 (t, *J* = 6.6 Hz, 2H), 2.89 (t, *J* = 6.8 Hz, 2H), 2.79 (s, 2H), 1.81 (p, *J* = 6.8 Hz, 2H).

Synthesis of methyl 4-((3-azidopropyl)carbamoyl)-2-iodobenzoate **2-2**



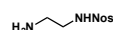
3-Iodo-4-methoxycarbonylbenzoic acid (1.0 g, 3.3 mmol, 1.0 eq.), **2-1** (491.0 mg, 4.9 mmol, 1.5 eq.), HATU (1.9 g, 4.9 mmol, 1.5 eq.) and DIPEA (1.7 mL, 9.8 mmol, 3.0 eq.) were dissolved in THF (30 mL) and DMF (7 mL). The mixture was stirred at RT for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The volatiles were removed by rotary evaporation, the residue was dissolved in EtOAc (100 mL) and washed with half-saturated NH₄Cl solution (3 x 80 mL) and H₂O (80 mL). The combined aqueous layers were extracted with EtOAc (100 mL). The combined organic layers were washed with brine (50 mL), dried over Na₂SO₄ and concentrated *in vacuo*. The crude product was purified by flash column chromatography (Silica, 150 g, cyclohexane:EtOAc 2:1, R_f = 0.28, UV). Product-containing fractions were combined and concentrated *in vacuo* to yield the desired product as yellowish oil (1.2 g, **95%**).

¹H NMR (400 MHz, CDCl₃) δ/ppm: 8.34 (d, *J* = 1.7 Hz, 1H), 7.84 (d, *J* = 8.0 Hz, 1H), 7.77 (dd, *J* = 8.1, 1.7 Hz, 1H), 6.41 (s, 1H), 3.95 (s, 3H), 3.56 (td, *J* = 6.6, 5.8 Hz, 2H), 3.47 (t, *J* = 6.4 Hz, 2H), 1.92 (p, *J* = 6.6 Hz, 2H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 166.51, 165.27, 139.77, 138.11, 137.85, 131.11, 126.49, 94.24, 52.91, 49.78, 38.31, 28.77.

HRMS (ESI): C₁₂H₁₄IN₄O₃⁺ *calcd*: 389.0105, *found*: 389.0109.

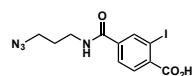
N-(2-aminoethyl)-2-nitrobenzenesulfonamide **2-3**



A solution of 2-nitrobenzenesulfonyl chloride (25.0 g, 113.0 mmol, 1.0 eq.) in DCM (250 mL) was added dropwise to a cooled solution of ethylenediamine (75.4 mL, 1.1 mol, 10.0 eq.) in DCM (250 mL) over 1.5 h. The mixture was stirred at RT for 3 h. The yellow mixture was washed with H₂O (3 x 200 mL) and brine (150 mL), the organic layer was dried over Na₂SO₄ and concentrated (residual ethylenediamine was removed by high vacuum drying) to yield the desired product as a thick yellow oil (12.2 g, **44%**). The material was used without further purification for the next step. Analytical data was in agreement with reported data.^[244]

¹H NMR (400 MHz, CDCl₃) δ/ppm: 8.17 – 8.12 (m, 1H), 7.88 – 7.84 (m, 1H), 7.75 (m, 2H), 3.12 (dd, *J* = 6.5, 5.0 Hz, 2H), 2.87 (dd, *J* = 6.5, 4.9 Hz, 2H).

Synthesis of 4-((3-azidopropyl)carbamoyl)-2-iodobenzoic acid **2-4a**



2-2 (1.2 g, 3.1 mmol, 1.0 eq.) was dissolved in MeCN (10 mL) and a solution of LiOH hydrate (646.0 mg, 15.4 mmol, 5.0 eq.) in H₂O (10 mL) was added. The solution was stirred at RT for 4 h, after which UPLC-MS analysis showed complete conversion to the desired product. The reaction mixture was acidified with TFA to pH 2-3 and directly purified by reversed phase flash column

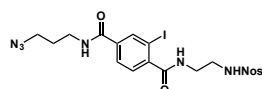
chromatography on the ISOLERA (RP-Silica, 100 g, H₂O/MeCN + 0.1% TFA, UV, 2 runs). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (1.1 g, **98%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 13.54 (s, 1H), 8.72 (t, *J* = 5.6 Hz, 1H), 8.38 (d, *J* = 1.6 Hz, 1H), 7.90 (dd, *J* = 8.0, 1.6 Hz, 1H), 7.76 (d, *J* = 8.0 Hz, 1H), 3.41 (t, *J* = 6.7 Hz, 2H), 3.32 (q, *J* = 6.5 Hz, 2H), 1.78 (p, *J* = 6.8 Hz, 2H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 168.02, 164.17, 138.70, 137.16, 129.56, 126.94, 93.82, 48.50, 36.80, 28.22.

HRMS (ESI): C₁₁H₁₁IN₄NaO₃⁺ *calcd*: 396.9768, *found*: 396.9769.

Synthesis of (Nosyl)ethyl-4-((3-azidopropyl)carbamoyl)-2-iodobenzoic amide **2-4**



2-4a (1.1 g, 3.0 mmol, 1.0 eq.), **2-3** (1.1 g, 4.5 mmol, 1.5 eq.), HATU (1.4 g, 3.6 mmol, 1.2 eq.) and DIPEA (1.6 mL, 9.0 mmol, 3.0 eq.) were dissolved in DMF (12 mL) and stirred at RT for 3 h, after which UPLC-MS analysis showed complete conversion of the starting material. The volatiles were removed *in vacuo* and the residue was dissolved in EtOAc (90 mL). The organic layer was washed with saturated NH₄Cl (3 x 80 mL) and H₂O (80 mL) and the combined aqueous layers were extracted with EtOAc (100 mL). The combined organic layers were washed with brine (50 mL), dried over Na₂SO₄ and concentrated by rotary evaporation. The crude was purified by flash column chromatography (Silica, 180 g, 1:4 cyclohexane:EtOAc, R_f = 0.2, UV) to yield the desired product as a yellow solid (1.6 g, **92%**).

¹H NMR (400 MHz, CD₃CN) δ/ppm: 8.25 (d, *J* = 1.6 Hz, 1H), 8.08 (dd, *J* = 5.9, 3.4 Hz, 1H), 7.89 – 7.85 (m, 1H), 7.84 – 7.76 (m, 3H), 7.37 (d, *J* = 7.9 Hz, 1H), 7.16 (s, 1H), 6.95 (s, 1H), 6.19 (s, 1H), 3.49 – 3.43 (m, 2H), 3.43 – 3.36 (m, 4H), 3.27 (t, *J* = 6.0 Hz, 2H), 1.83 (t, *J* = 6.7 Hz, 2H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 169.91, 166.24, 148.14, 144.06, 138.37, 136.61, 133.90, 133.63, 133.06, 131.15, 127.77, 126.70, 92.40, 49.58, 43.39, 40.13, 39.11, 28.75.

HRMS (ESI): C₁₉H₂₀IN₇NaO₆S⁺ *calcd*: 624.0133, *found*: 624.0144.

Imidazole-1-sulfonyl azide hydrochloride **2-5a**

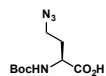


Synthesis according to a literature reported procedure.^[236]

Under an inert atmosphere NaN₃ (3.9 g, 60.0 mmol, 1.0 eq.) was suspended in MeCN (60 mL) and cooled to 0°C in an ice bath. Sulfuryl chloride (4.9 mL, 60.0 mmol, 1.0 eq.) was added dropwise over 10 min and the mixture was stirred at RT for 18 h. After that, the mixture was cooled in an ice-bath and imidazole (7.8 g, 114.0 mmol, 1.9 eq.) was added portionwise. Hereafter, the mixture was stirred at RT for another 3 h. EtOAc (120 mL) was added, the mixture was washed with H₂O (2 x 120 mL) and saturated NaHCO₃ (2 x 120 mL), dried over Na₂SO₄ and filtered. A solution of HCl in EtOH (prepared by the addition of AcCl (6.42 mL, 90 mmol) to dry EtOH (22.5 mL) while cooling) was added dropwise to the filtrate. The mixture was cooled in an ice bath and stirred for 1 h. The formed solid was filtered off, washed with EtOAc (3 x 10 mL) and dried at high vacuum to yield the desired product as white solid (9.1 g, **72%**). Analytical data was in agreement with reported data.

^1H NMR (400 MHz, D_2O) δ /ppm: 9.44 (t, $J = 1.4$ Hz, 1H), 8.06 (dd, $J = 2.2, 1.6$ Hz, 1H), 7.65 (dd, $J = 2.2, 1.2$ Hz, 1H).

Boc-Dab(N_3)-OH **2-5b**



The compound was synthesized according to a modified procedure by E.D. Goddard-Borger and R.V. Stick.^[236] Compound is also commercially available as CHA salt.

Under an inert atmosphere Boc-Dab-OH (5.9 g, 27.2 mmol, 1.0 eq.) and **2-5a** (6.8 g, 32.7 mmol, 1.2 eq) were dissolved in MeOH (140 mL) and potassium carbonate (10.2 g, 73.5 mmol, 2.7 eq.) was added, followed by CuSO_4 pentahydrate (68.0 mg, 272 μmol , 1 mol%). The blue mixture was stirred at RT for 2 h, after which TLC (5% MeOH in DCM, Ninhydrin) showed complete consumption of the starting material. The solvent was removed and the crude was dissolved in EtOAc (150 mL) and H_2O (150 mL). The biphasic mixture was acidified with concentrated HCl to pH 1-2. The biphasic mixture was separated and the aqueous layer was extracted with EtOAc (2 x 150 mL) and the combined organic layers were washed with H_2O (2 x 150 mL) and brine (150 mL), dried over Na_2SO_4 and concentrated by rotary evaporation to yield the desired product as slightly yellow oil (7.9 g, >100%). NMR analysis showed some unidentified peaks. The material was used as obtained without further purification.

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm 12.60 (s, 1H), 3.96 (ddd, $J = 9.9, 8.2, 4.5$ Hz, 1H), 3.42 (ddd, $J = 12.4, 7.2, 5.4$ Hz, 1H), 3.34 (ddd, $J = 12.3, 8.0, 6.5$ Hz, 1H), 1.97 – 1.84 (m, 1H), 1.85 – 1.71 (m, 1H), 1.38 (s, 9H).

H-Dab(N_3)-OMe **2-5**



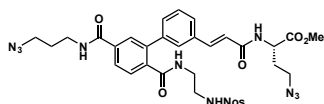
Under an inert atmosphere **2-5b** (7.9 g, 32.5 mmol, 1.0 eq.) was dissolved in MeOH (150 mL) and cooled in an ice bath. Thionyl chloride (19.0 mL, 260.0 mmol, 8.0 eq.) was added dropwise and the solution was stirred at 0°C for 15 min. The cooling bath was removed and the solution was stirred at RT for 24 h, after which TLC (BAW, Ninhydrin, $R_f = 0.58$) and NMR showed complete conversion to the desired product. The solvent was removed by rotary evaporation and the residue was dissolved in half-saturated NaHCO_3 solution (200 mL). The aqueous solution was extracted with EtOAc (10 x 100 mL), the combined organic layers were dried over Na_2SO_4 and concentrated to yield the desired product as a yellow liquid, which solidified over time in the fridge (3.4 g, 66%).

^1H NMR (400 MHz, CD_3OD) δ /ppm: 3.74 (s, 3H), 3.55 (dd, $J = 7.5, 5.6$ Hz, 1H), 3.45 (td, $J = 6.6, 1.5$ Hz, 2H), 1.98 (dddd, $J = 14.0, 7.3, 6.8, 5.6$ Hz, 1H), 1.81 (ddt, $J = 14.0, 7.5, 6.4$ Hz, 1H).

^{13}C NMR (101 MHz, CD_3OD) δ /ppm: 176.65, 52.75, 52.60, 49.85, 34.47.

HRMS (ESI): $\text{C}_5\text{H}_{11}\text{N}_4\text{O}_2^+$ *calcd*: 159.0877, *found*: 159.0876.

Synthesis of Nos-Sc(N₃)-NP01-TFL(N₃)-OMe 2-6



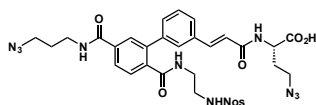
In a 2 mL Eppendorf tube **NP01** (24.0 mg, 38.6 μmol , 1.0 eq.), TFL **2-5** (9.2 mg, 57.9 μmol , 1.5 eq.), HATU (22.0 mg, 57.9 μmol , 1.5 eq.) and DIPEA (33.6 μL , 193.0 μmol , 5.0 eq.) were dissolved in DMF (300 μL) and agitated at RT for 1 h, after which UPLC-MS analysis showed complete conversion to the desired product. The mixture was purified by reversed-phase preparative HPLC (Method A) to yield the desired product as white solid (22 mg, **75%**).

¹H NMR (500 MHz, CDCl₃) δ /ppm: 8.06 – 8.00 (m, 1H), 7.86 – 7.81 (m, 1H), 7.77 – 7.73 (m, 2H), 7.72 – 7.68 (m, 1H), 7.66 – 7.60 (m, 1H), 7.54 (d, J = 15.6 Hz, 1H), 7.42 (d, J = 8.1 Hz, 2H), 7.40 – 7.28 (m, 5H), 6.60 (s, 1H), 6.47 (d, J = 15.7 Hz, 1H), 5.89 (s, 1H), 4.80 (td, J = 7.7, 5.0 Hz, 1H), 3.75 (s, 3H), 3.58 (q, J = 6.5 Hz, 2H), 3.45 (t, J = 6.6 Hz, 4H), 3.35 – 3.21 (m, 2H), 3.00 (d, J = 5.8 Hz, 2H), 2.20 (dtd, J = 14.0, 6.9, 5.0 Hz, 1H), 2.08 (dq, J = 14.1, 6.9 Hz, 1H), 1.94 (p, J = 6.6 Hz, 2H).

¹³C NMR (500 MHz, 2D NMR, CDCl₃) δ /ppm: 172.03, 169.77, 167.27, 166.09, 147.77, 141.10, 139.42, 139.20, 139.10, 137.69, 135.68, 135.00, 133.86, 132.89, 130.94, 130.10, 129.17, 129.12, 129.06, 128.48, 126.32, 125.99, 125.33, 120.88, 52.62, 50.31, 49.33, 47.70, 42.67, 39.65, 37.96, 31.31, 28.52.

HRMS (ESI): C₃₃H₃₅N₁₁NaO₉S⁺ *calcd*: 784.2232, *found*: 784.2237.

Synthesis of Nos-Sc(N₃)-NP01-TFL(N₃)-OH 2-7



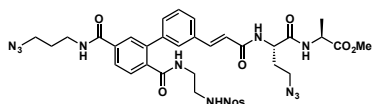
A solution of **2-6** (22.0 mg, 28.9 μmol , 1.0 eq.) in MeCN (578 μL) and a solution of LiOH monohydrate (2.9 mg, 69.4 μmol , 2.4 eq.) in H₂O (678 μL) were mixed in an Eppendorf tube and the reaction was agitated at RT for 2 h. The mixture was purified by reversed-phase preparative HPLC (Method A) to yield the desired product as white solid (16 mg, **74%**).

¹H NMR (500 MHz, CDCl₃) δ /ppm: 7.95 – 7.89 (m, 1H), 7.77 – 7.72 (m, 2H), 7.72 – 7.69 (m, 1H), 7.70 – 7.63 (m, 2H), 7.54 – 7.48 (m, 2H), 7.45 (d, J = 15.7 Hz, 1H), 7.36 (ddd, J = 5.1, 3.9, 1.7 Hz, 1H), 7.35 – 7.30 (m, 2H), 6.49 (d, J = 15.5 Hz, 1H), 4.61 (dt, J = 7.9, 3.6 Hz, 1H), 3.41 (t, J = 6.8 Hz, 2H), 3.39 – 3.30 (m, 4H), 3.20 (dd, J = 6.6, 5.0 Hz, 2H), 2.91 – 2.83 (m, 2H), 2.13 (dtd, J = 14.0, 7.1, 5.0 Hz, 1H), 1.98 – 1.89 (m, 1H), 1.82 (p, J = 6.7 Hz, 2H).

¹³C NMR (500 MHz, 2D NMR, CDCl₃) δ /ppm: 173.20, 170.33, 167.46, 166.22, 147.60, 140.67, 139.72, 139.13, 138.00, 135.72, 134.75, 133.67, 132.90, 132.59, 130.61, 129.77, 128.97, 128.77, 128.08, 127.99, 127.08, 126.06, 125.03, 120.65, 49.92, 49.02, 47.66, 42.13, 39.44, 37.35, 30.91, 28.31.

HRMS (ESI): C₃₂H₃₂N₁₁O₉S⁻ *calcd*: 746.2111, *found*: 746.2118.

Synthesis of Nos-Sc(N₃)-NP01-TFL(N₃)-AA001-OMe 2-8



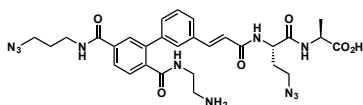
Precursor 2-6 (16.0 mg, 21.4 μmol , 1.0 eq.), AA001 (4.5 mg, 32.1 μmol , 1.5 eq.), HATU (12.2 mg, 32.1 μmol , 1.5 eq.) and DIPEA (18.6 μL , 107.0 μmol , 5.0 eq.) were dissolved in DMF (700 μL) and stirred at RT for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The mixture was purified by reversed-phase preparative HPLC (Method A) to yield the desired product as white solid (11 mg, **62%**).

¹H NMR (500 MHz, CD₃CN) δ /ppm: 8.03 – 7.95 (m, 1H), 7.88 – 7.84 (m, 1H), 7.84 – 7.78 (m, 4H), 7.57 (s, 1H), 7.55 – 7.46 (m, 3H), 7.42 – 7.34 (m, 2H), 7.32 (t, J = 5.9 Hz, 1H), 7.15 (d, J = 7.1 Hz, 1H), 7.09 (d, J = 8.0 Hz, 1H), 6.72 (t, J = 6.0 Hz, 1H), 6.62 (d, J = 15.7 Hz, 1H), 6.06 (t, J = 6.0 Hz, 1H), 4.57 (td, J = 8.3, 5.4 Hz, 1H), 4.37 (p, J = 7.2 Hz, 1H), 3.66 (s, 3H), 3.47 – 3.38 (m, 6H), 3.23 (q, J = 6.2 Hz, 2H), 2.95 (q, J = 6.2 Hz, 2H), 2.07 (dtd, J = 14.3, 7.3, 5.3 Hz, 1H), 1.87 (dp, J = 29.7, 6.7 Hz, 3H), 1.32 (d, J = 7.3 Hz, 3H).

¹³C NMR (500 MHz, 2D NMR, CD₃CN) δ /ppm: 173.82, 171.86, 170.23, 167.04, 166.24, 148.79, 141.19, 140.75, 140.17, 139.59, 137.00, 135.98, 135.14, 133.87, 133.50, 131.40, 130.84, 129.83, 129.56, 129.03, 128.66, 128.25, 127.23, 126.01, 122.47, 52.69, 51.56, 49.79, 49.05, 48.58, 43.51, 40.00, 37.84, 32.15, 29.29, 17.40

HRMS (ESI): C₃₆H₄₀N₁₂NaO₁₀S⁺ *calcd*: 855.2603, *found*: 855.2607.

Synthesis of NH₂-Sc(N₃)-NP01-TFL(N₃)-AA001-OH 2-9



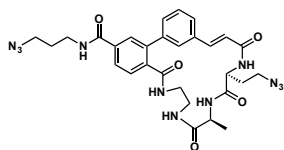
Precursor **2-8** (11.0 mg, 13.2 μmol , 1.0 eq.) was dissolved in DMF (400 μL) and rendered inert with N₂. DIPEA (23.0 μL , 132.0 μmol , 10.0 eq.) and thiophenol (6.7 μL , 66 μmol , 5.0 eq.) were added subsequently and the solution was stirred at 40°C for 3 h, after which UPLC-MS analysis showed full conversion to the Nosyl-deprotected intermediate. The yellow solution was cooled to RT and a solution of LiOH monohydrate (5.5 mg, 132.0 μmol , 10.0 eq.) in H₂O (264 μL) was added. The mixture was stirred at RT for 30 min, after which UPLC-MS analysis showed complete conversion to the desired product. The mixture was purified by reversed-phase preparative HPLC (Method A) to yield the desired product as white solid (7 mg, **78%**).

¹H NMR (500 MHz, CD₃OD) δ /ppm: 7.96 – 7.89 (m, 2H), 7.68 (dd, J = 7.7, 0.7 Hz, 1H), 7.66 – 7.62 (m, 2H), 7.60 (d, J = 15.8 Hz, 1H), 7.53 – 7.47 (m, 2H), 6.74 (d, J = 15.8 Hz, 1H), 4.62 (dd, J = 8.5, 5.6 Hz, 1H), 4.37 (q, J = 7.3 Hz, 1H), 3.54 – 3.45 (m, 4H), 3.44 – 3.38 (m, 4H), 2.88 (td, J = 6.8, 1.4 Hz, 2H), 2.13 (dtd, J = 13.0, 7.3, 5.6 Hz, 1H), 2.01 – 1.93 (m, 1H), 1.90 (p, J = 6.7 Hz, 2H), 1.43 (d, J = 7.3 Hz, 3H).

¹³C NMR (500 MHz, 2D NMR, CD₃OD) δ /ppm: 175.74, 172.90, 172.69, 168.76, 168.00, 141.66, 141.27, 140.62, 139.41, 137.15, 136.07, 131.01, 130.08, 129.81, 129.07, 129.04, 128.13, 127.38, 121.88, 52.12, 49.98, 49.42, 48.74, 39.72, 38.31, 38.21, 32.30, 29.42, 17.28.

HRMS (ESI): C₂₉H₃₆N₁₁O₆⁺ *calcd*: 634.2845, *found*: 634.2854.

Synthesis of MC(N₃)-NP01-AA001 2-10



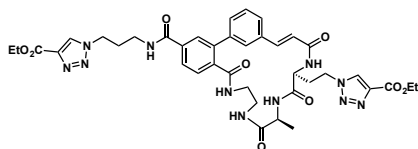
Under an inert atmosphere precursor **2-9** (6.5 mg, 10.3 μmol , 1.0 eq.) was dissolved in DMF (6.5 mL) and HATU (5.9 mg, 15.4 μmol , 1.5 eq.) and DIPEA (8.9 μL , 51.3 μmol , 5.0 eq.) were added successively. The yellow solution was stirred at RT for 1 h, after which UPLC-MS analysis showed complete conversion to the desired product. The solvent was removed by rotary evaporation and the residue was purified by reversed-phase preparative HPLC (Method A) to yield the desired product as white solid (4 mg, **63%**).

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 8.69 (t, *J* = 5.6 Hz, 1H), 8.56 (t, *J* = 5.4 Hz, 1H), 8.36 (d, *J* = 7.6 Hz, 1H), 8.00 (d, *J* = 1.8 Hz, 1H), 7.88 (dd, *J* = 7.9, 1.7 Hz, 1H), 7.71 (t, *J* = 6.0 Hz, 1H), 7.68 (d, *J* = 8.7 Hz, 1H), 7.67 – 7.64 (m, 1H), 7.60 – 7.57 (m, 1H), 7.55 (d, *J* = 7.4 Hz, 1H), 7.51 (d, *J* = 7.9 Hz, 1H), 7.47 (d, *J* = 1.7 Hz, 1H), 7.41 (d, *J* = 15.8 Hz, 1H), 6.55 (d, *J* = 15.8 Hz, 1H), 4.27 – 4.19 (m, 2H), 3.51 – 3.46 (m, 1H), 3.44 – 3.42 (m, 2H), 3.40 – 3.35 (m, 2H), 3.30 – 3.26 (m, 1H), 3.15 – 3.08 (m, 2H), 2.92 – 2.84 (m, 1H), 1.90 (dq, *J* = 13.8, 6.9 Hz, 2H), 1.81 (p, *J* = 6.8 Hz, 3H), 1.18 (d, *J* = 7.0 Hz, 3H).

¹³C NMR (500 MHz, 2D NMR, DMSO-*d*₆) δ /ppm: 172.28, 170.91, 169.52, 166.12, 165.58, 140.32, 139.73, 138.25, 135.52, 135.20, 129.38, 129.20, 129.11, 129.08, 127.60, 127.14, 126.19, 124.54, 123.01, 53.09, 48.64, 48.62, 47.68, 39.28, 38.38, 36.51, 29.72, 28.09, 17.34.

HRMS (ESI): C₂₉H₃₄N₁₁O₅⁺ calcd: 616.2739, found: 616.2737.

Synthesis of MC(N₃)-NP01-AA001-TA641 2-11^[198]



Under a N₂ atmosphere MC **2-10** (19.0 mg, 30.9 μmol , 1.0 eq.), NaOAsc (3.7 mg, 18.5 μmol , 0.6 eq.) and ethyl propiolate **TA641** (6.6 μL , 64.8 μmol , 2.1 eq.) were dissolved in DMSO (3 mL), H₂O (550 μL) and MeCN (550 μL). Cu(II)-TBTA complex solution (926.0 μL , 9.3 μmol , 0.3 eq.) was added and the mixture was stirred at RT for 2 h. The mixture was purified by reversed-phase preparative HPLC (Method A) to yield the desired product as a white solid (17 mg, **68%**).

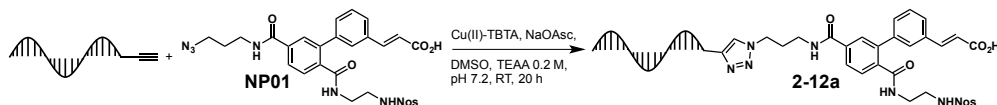
¹H-NMR (600 MHz, DMSO-*d*₆) δ /ppm: 8.83 (s, 1H), 8.80 (s, 1H), 8.72 (t, *J* = 5.7 Hz, 1H), 8.57 (t, *J* = 5.7 Hz, 1H), 8.44 (d, *J* = 7.5 Hz, 1H), 8.01 – 7.98 (m, 1H), 7.89 – 7.86 (m, 1H), 7.72 – 7.69 (m, 2H), 7.67 – 7.64 (m, 1H), 7.60 – 7.58 (m, 1H), 7.57 – 7.55 (m, 1H), 7.52 – 7.50 (m, 1H), 7.48 – 7.45 (m, 1H), 7.41 (d, *J* = 15.7 Hz, 1H), 6.54 (d, *J* = 15.7 Hz, 1H), 4.55 (t, *J* = 7.1 Hz, 2H), 4.50 (t, *J* = 6.9 Hz, 2H), 4.34 – 4.27 (m, 4H), 4.24 – 4.19 (m, 1H), 4.10 – 4.04 (m, 1H), 3.33 – 3.28 (m, 2H), 3.14 – 3.08 (m, 2H), 2.92 – 2.85 (m, 1H), 2.34 – 2.28 (m, 1H), 2.28 – 2.20 (m, 1H), 2.18 – 2.12 (p, *J* = 7.1 Hz, 2H), 1.30 (q, *J* = 6.8 Hz, 6H), 1.17 (d, *J* = 7.0 Hz, 3H).

¹³C-NMR (151 MHz DMSO-*d*₆) δ /ppm: 172.2, 170.5, 169.3, 166.0, 165.6, 160.3, 160.3, 140.0, 139.9, 139.9, 138.7, 138.7, 138.1, 135.2, 135.0, 129.4, 129.4, 129.3, 129.2, 129.2, 127.8, 127.2, 126.4, 124.8, 123.1, 60.5, 60.5, 53.1, 48.7, 47.8, 46.9, 39.6, 38.5, 36.5, 30.9, 29.5, 17.5, 14.2 (2x).

HR-MS: m/z for C₃₉H₄₅N₁₁O₉Na⁺ calcd. 834.3294, found: 834.3306.

13.4 Test Macrocycle Synthesis on DNA

Synthesis of the Encoded Macrocycle Precursor **2-12a** by Click Reaction with Alkyne-Modified DNA



In a 1.5 mL Eppendorf tube a 5'-alkyne-DNA strand (1.0 mM in H₂O, 44.0 μ L, 1.0 eq.) of the sequence 5'-hexyne-GGA GCT TGT GAA TTC TGG ATG GGA CGT GTG TGA ATT GTC-3', DE-1 azide **NP01** (10.0 mM in DMSO, 8.8 μ L, 2.0 eq.), sodium ascorbate (5.0 mM in H₂O, 44.0 μ L, 5.0 eq.), TEAA buffer (1.0 M, pH 7.2, 88.0 μ L), DMSO (198 μ L) and H₂O (35 μ L) were mixed. N₂ was bubbled through the solution for 30 s. The Cu(II)-TBTA solution (10.0 mM in 55% DMSO, 22.0 μ L, 5.0 eq.) was added and the solution was degassed again for 30 s. The mixture was agitated at RT for 20 h, after which HPLC showed 88% conversion to the desired product. The mixture was purified by preparative HPLC (Method B). The product containing fractions were combined and lyophilized to yield the desired product **2-12a** as a white solid. The DNA was dissolved in MOPS buffer (161 μ L) to yield a clear solution (140 μ M, **51%**).

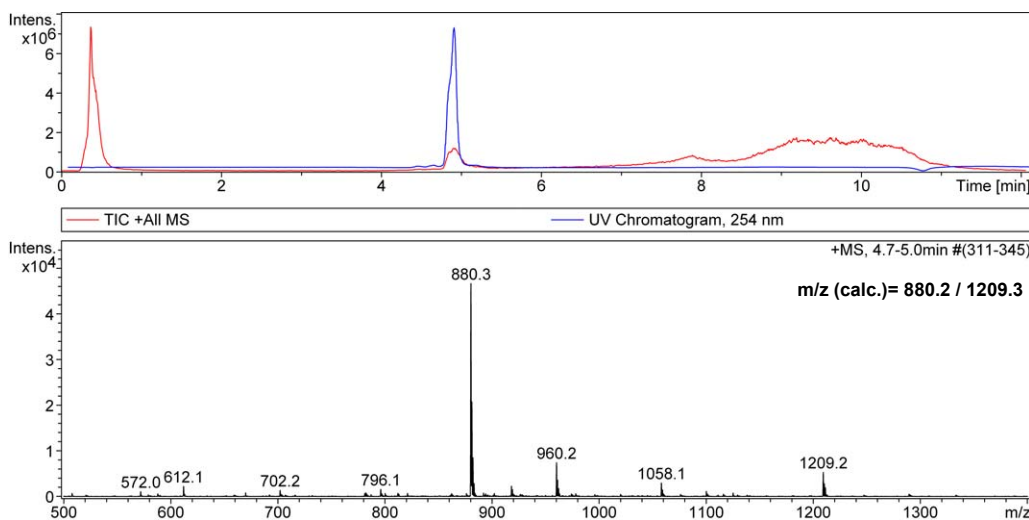
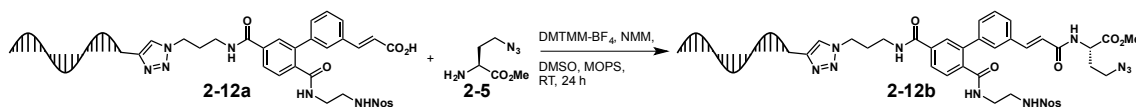


Figure 86. LC-MS chromatogram of the purified encoded compound **2-12a**.

TFL **2-5** Amide Coupling with **2-12a**



In a 1.5 mL Eppendorf tube the DNA-carboxylic acid **2-12a** (140.0 μ M in MOPS buffer, 21.4 μ L, 1.0 eq.) was mixed with the trifunctional linker **2-5** (0.5 M in DMSO, 6.0 μ L, 1000.0 eq.), DMTMM-BF₄ (400.0 mM in DMSO, 7.5 μ L, 1000.0 eq.) and NMM (400.0 mM in DMSO, 15.7 μ L, 2100.0 eq.). The mixture was agitated at RT for 24 h. The reaction was purified by EtOH precipitation and the clean DNA pellet was dissolved in H₂O (25 μ L) to yield the desired product **2-12b** as a clear solution (130 μ M, **quant.**).

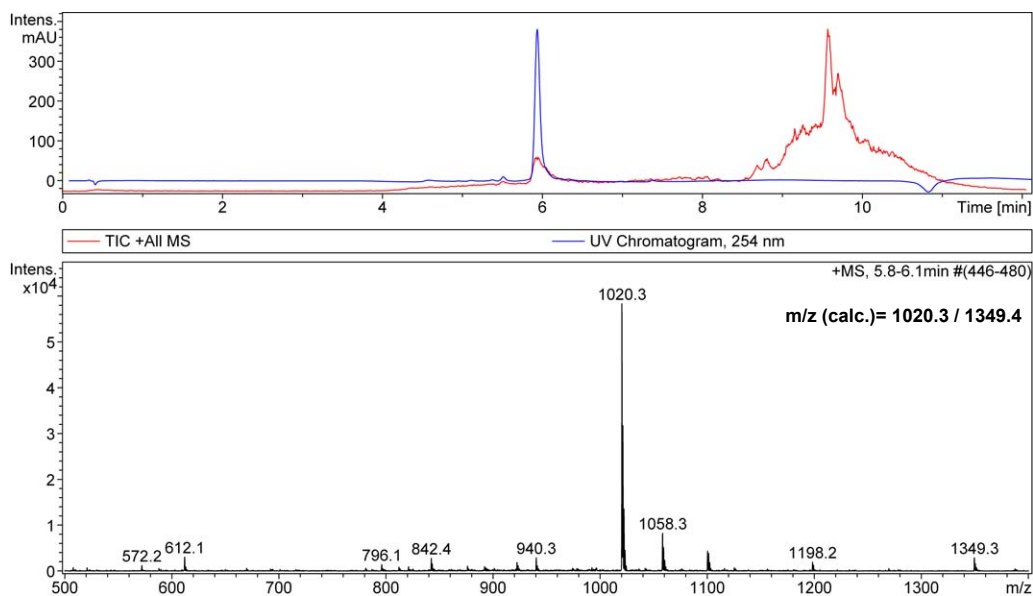
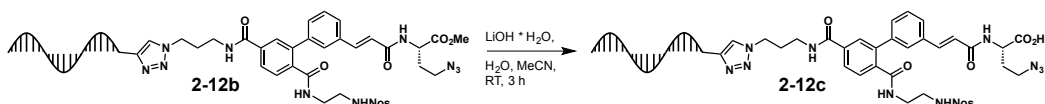


Figure 87. LC-MS chromatogram of the purified encoded compound **2-12b**.

Ester Hydrolysis of **2-12b**



In a 1.5 mL Eppendorf tube **2-12b** (130.0 μM in H_2O , 25.0 μL , 1.0 eq.), LiOH monohydrate (100.0 mM in H_2O , 4.9 μL , 150.0 eq.) and MeCN (15 μL) were mixed and agitated at RT for 3 h. The reaction was purified by EtOH precipitation and the clean DNA pellet was dissolved in MOPS buffer (28 μL) to yield the desired product **2-12c** as a clear solution (100 μM , **86%**).

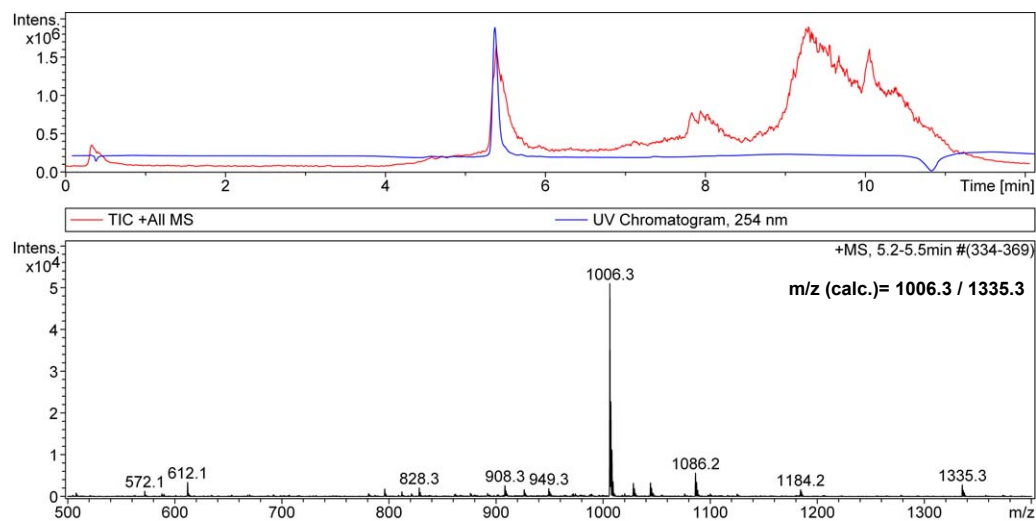
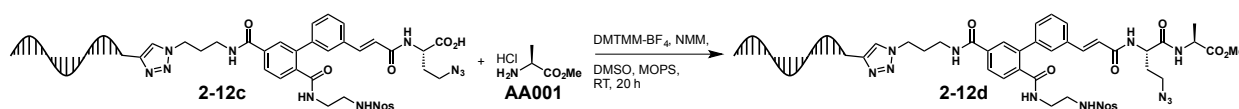


Figure 88. LC-MS chromatogram of the purified encoded compound **2-12c**.

Amide Coupling with **AA001**, Introduction of the 2nd Diversity Element



In a 1.5 mL Eppendorf tube **2-12c** (90.0 μ M in MOPS buffer, 29.0 μ L, 1.0 eq.) was mixed with *L*-alanine methyl ester hydrochloride **AA001** (500.0 mM in DMSO, 5.2 μ L, 1000.0 eq.), DMTMM-BF₄ (150.0 mM in DMSO, 17.4 μ L, 1000.0 eq.) and NMM (250.0 mM in DMSO, 36.5 μ L, 3500.0 eq.) and the solution was agitated at RT for 20 h. The mixture was purified by EtOH precipitation and the clean DNA-pellet was dissolved in H₂O (135 μ L) to yield the desired product **2-12d** as a clear solution (20 μ M, **quant.**).

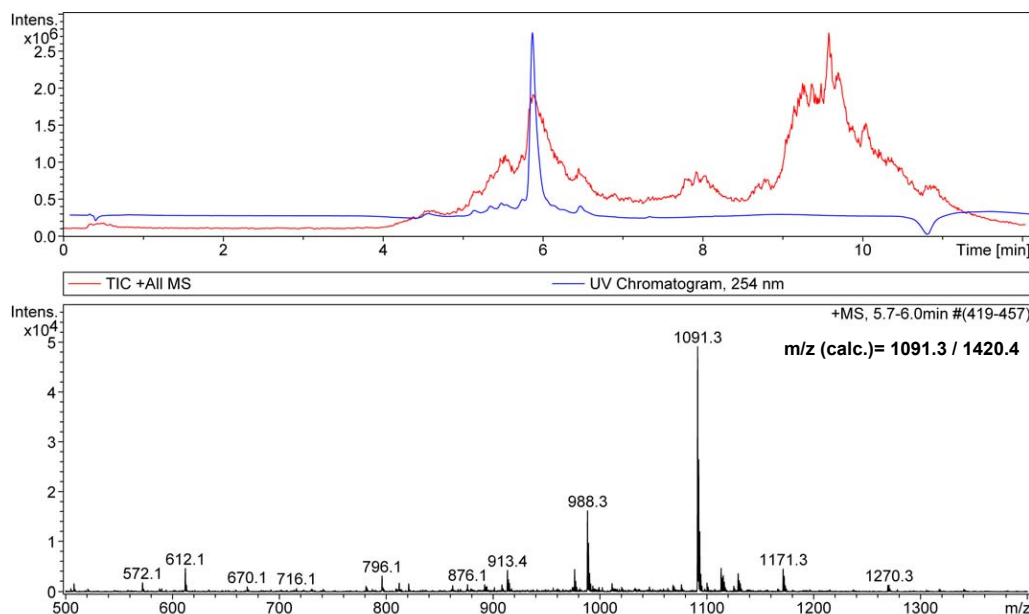
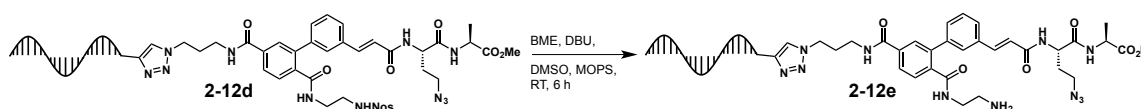


Figure 89. LC-MS chromatogram of the purified encoded compound **2-12d**.

Ester Hydrolysis and Nosyl Deprotection of **2-12d**



In a 0.5 mL Eppendorf tube **2-12d** (50.0 μ M in 300.0 mM MOPS/500.0 mM NaCl pH 8.2 buffer, 15.0 μ L, 1.0 eq.) was mixed with DBU (300.0 mM in DMSO, 15.0 μ L, 6000.0 eq.) and BME (300.0 mM in DMSO, 15.0 μ L, 6000.0 eq.). The mixture was degassed with N₂ for 30 s and agitated at RT for 6 h. The reaction was purified by EtOH precipitation and the clean DNA pellet was dissolved in MOPS buffer (20 μ L) to yield the desired product **2-12e** as a clear solution (30 μ M, **80%**).

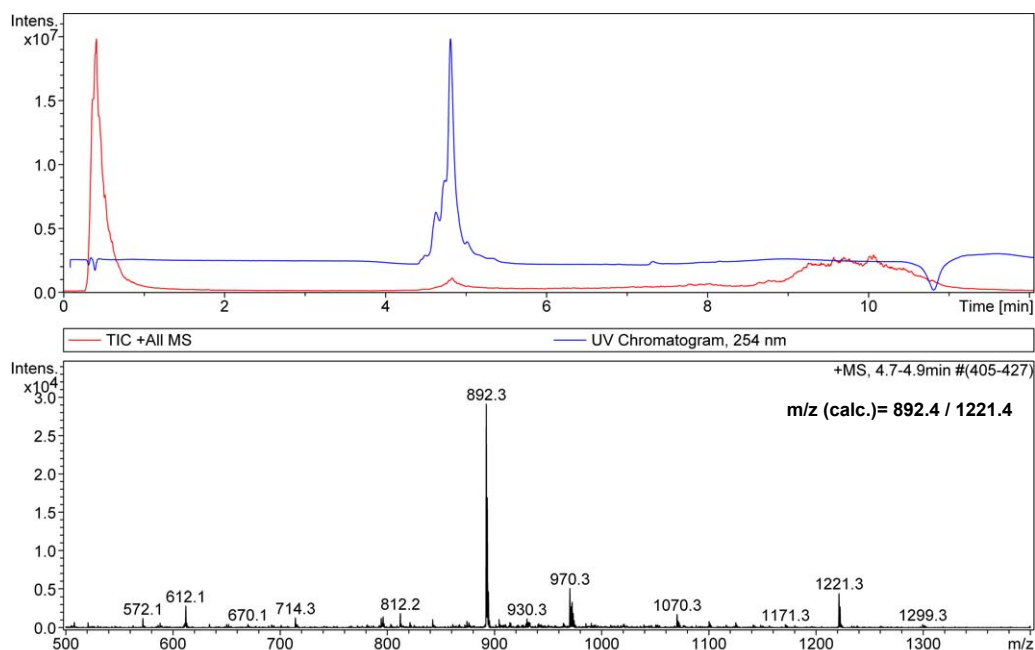
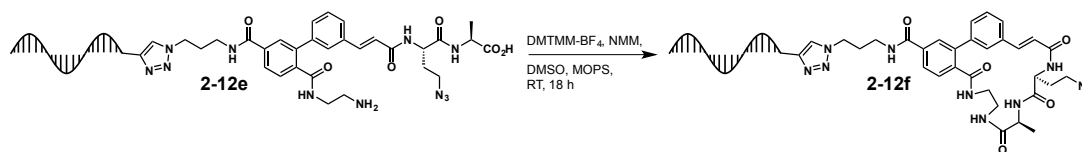


Figure 90. LC-MS chromatogram of the purified encoded compound **2-12e**.

Macrocyclization of **2-12e**



In a 0.5 mL Eppendorf tube **2-12e** (25.0 μM in MOPS buffer, 13.0 μL , 1.0 eq.) was mixed with MOPS buffer (2.0 μL), DMSO (4.6 μL), DMTMM- BF_4 (25.0 mM in DMSO, 3.9 μL , 300.0 eq.) and NMM (50.0 mM in DMSO, 6.5 μL , 1000.0 eq.). The mixture was agitated at RT for 18 h and the reaction was purified by EtOH precipitation. The reaction was set up again with above stated reagent amounts. After EtOH precipitation the clean DNA pellet was dissolved in H_2O (15 μL) to yield the desired product **2-12f** as a clear solution (24 μM , **quant.**). **Note:** The macrocyclization was repeated once because LC-MS showed incomplete conversion. Increase of the coupling reagent amount (up to 500 eq.) and DMSO amount (>55%) should yield full macrocyclization in one step.

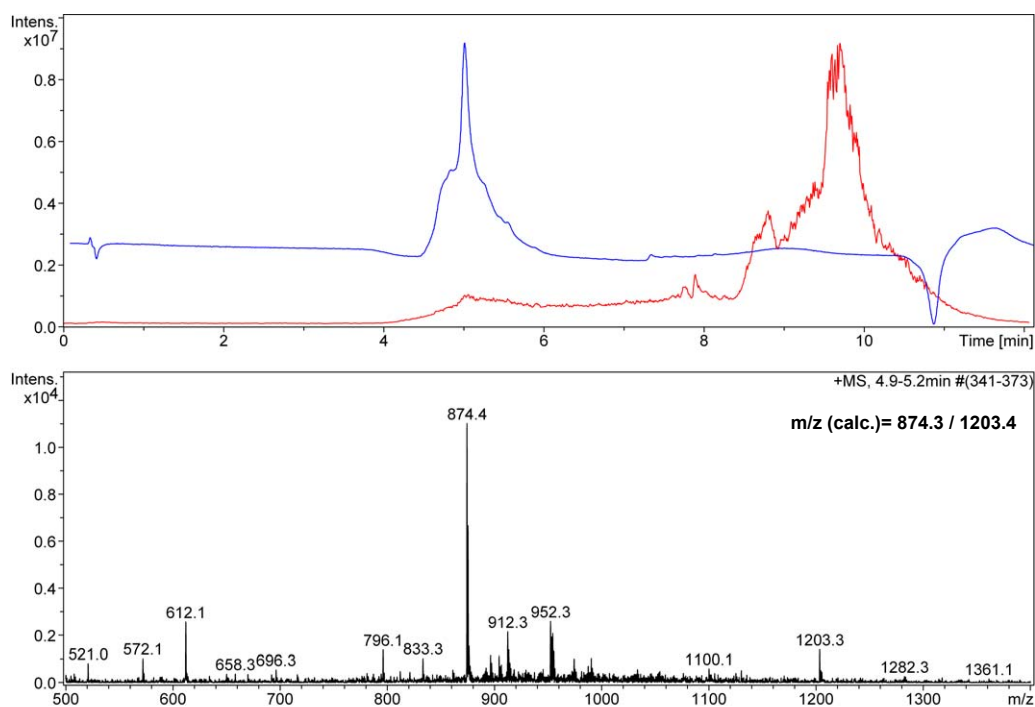
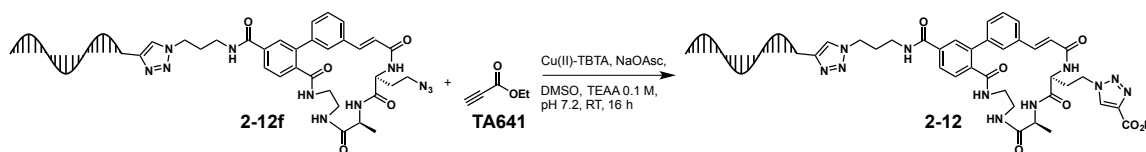


Figure 91. LC-MS chromatogram of the purified encoded compound **2-12f**.

Diversity Element 3 Introduction by Click Reaction



In a 0.5 mL Eppendorf tube, **2-12f** (10.0 μM in 500.0 mM TEAA pH 7.2, 26.4 μL , 1.0 eq.) was mixed with ethyl propiolate **TA641** (1.0 mM in DMSO, 26.4 μL , 100.0 eq.), sodium ascorbate (3.0 mM in H_2O , 17.6 μL , 200.0 eq.) and DMSO (17.5 μL) and the mixture was degassed with N_2 for 30 s. Copper(II)-TBTA complex (3.0 mM in 55% DMSO, 17.6 μL , 200.0 eq.) was added and the solution was degassed with N_2 again. The solution was agitated at RT for 16 h. The reaction was purified by EtOH precipitation and the clean DNA pellet was dissolved in H_2O (15 μL) to yield the final product **2-12** as a clear solution (14 μM , **80%**).

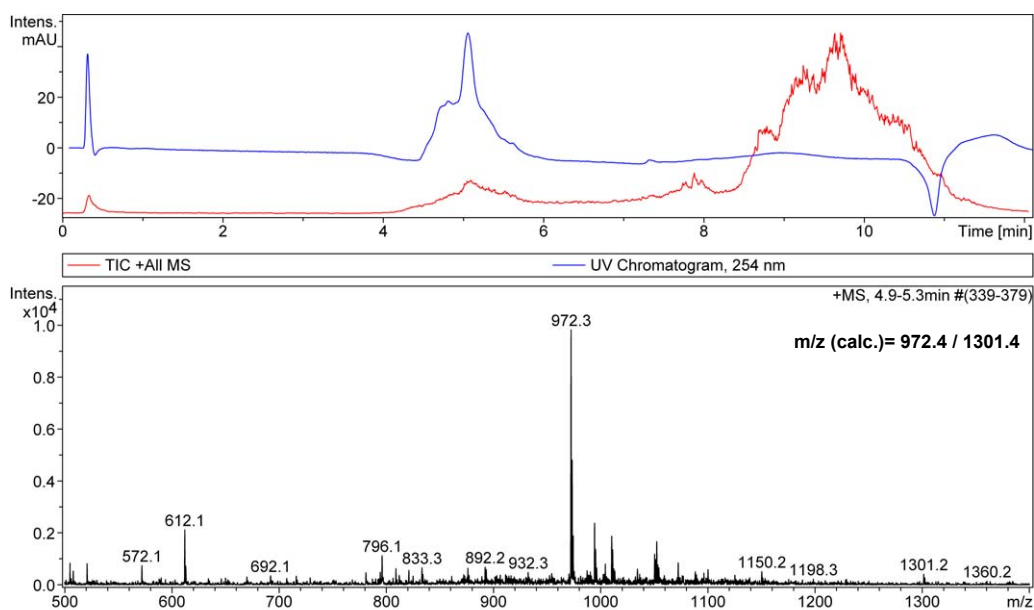
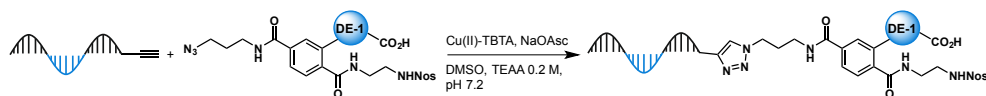


Figure 92. LC-MS chromatogram of the purified encoded compound **2-12**.

13.5 DNA-Encoded Library Assembly, Synthetic Procedures

Encoding of DE-1 with Alkyne-Modified DNA (21 elements)

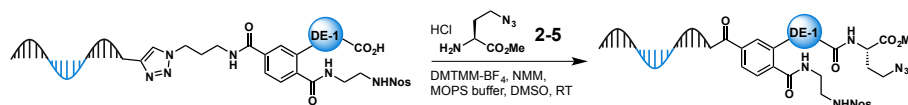


In a 1.5 mL Eppendorf tube the alkyne-DNA strand (1.0 mM in H₂O, 1.0 eq.) of the sequence 5'-hexyne-GGAGCTTGTGAATTCTGGXXXGGACGTGTGTGAATTGTC-3', DE-1 azide (**NP01 - 21**, 10.0 mM in DMSO, 2.0 eq.), sodium ascorbate (5.0 mM in H₂O, 5.0 eq.), TEAA buffer (1.0 M, pH 7.2), DMSO and H₂O were mixed. N₂ was bubbled through the solution for 30 s. Cu(II)-TBTA complex (10.0 mM in 55% DMSO, 5.0 eq.) was added, N₂ was bubbled through the solution again and the mixture was shaken at RT overnight (20 h). The final concentrations of the reagents were as follows: 100.0 μM DNA, 200.0 μM DE-1 azide, 500.0 μM Cu(II)-TBTA complex, 500.0 μM sodium ascorbate, 200.0 mM TEAA, 50% DMSO. Typical DNA amounts: 30 nmol. Conversion and purity of the product was measured by HPLC. The mixture was directly purified by prep. HPLC (Method B). Product-containing fractions were combined and lyophilized. The isolated product was dissolved in MOPS buffer to yield a 140 μM solution. Typical conversions: 80-90%. XXX represents the three-base codon for the diversity element 1. This procedure was repeated for all DE-1 building blocks (**Table 11**).

Table 11. Masses found from the DE-1 encoding by LC-MS.

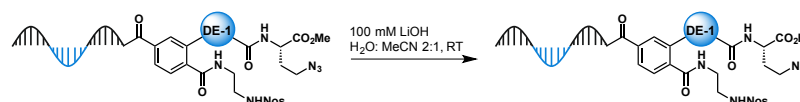
No.	Mass calc. Mass found	No.	Mass calc. Mass found	No.	Mass calc. Mass found
NP01	880.2/1209.3 880.3/1209.2	NP08	922.2/1251.3 922.2/1251.1	NP15	844.2/1173.3 844.5/1173.5
NP02	880.2/1209.3 880.3/1209.2	NP09	872.2/1201.3 872.2/1201.1	NP16	848.3/1177.3 848.6/1177.6
NP03	882.3/1211.3 882.3/1211.2	NP10	928.3/1257.3 928.3/1257.2	NP17	906.3/1235.3 906.6/1235.6
NP04	882.3/1211.3 882.3/1211.2	NP11	880.2/1209.3 880.6/1209.6	NP18	932.3/1261.3 932.6/1261.6
NP05	854.2/1183.3 854.3/1183.2	NP12	820.2/1149.3 820.6/1149.6	NP19	920.3/1249.3 920.6/1249.6
NP06	860.2/1189.2 860.2/1189.1	NP13	806.2/1135.3 806.5/1135.5	NP20	920.3/1249.3 920.6/1249.6
NP07	855.2/1184.3 855.5/1184.5	NP14	820.2/1149.3 820.5/1149.5	NP21	920.3/1249.3 920.3/1249.2

Trifunctional Linker Coupling with DNA-DE-1 Pool



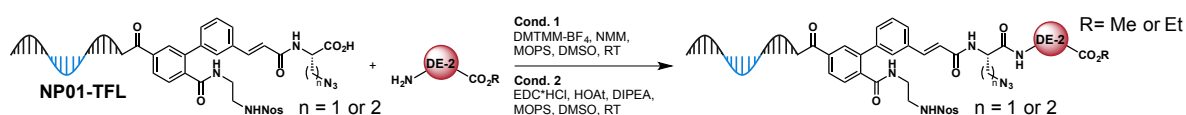
20 encoded DE-1 DNA strands (**NP01-06** and **NP08-21**; 21.5 μ L, 3.0 nmol each, 50.0 mM MOPS buffer) were pooled in a LowBind 1.5 mL Eppendorf tube and NMM (315.0 μ L, 400.0 mM in DMSO, 2100.0 eq.) was added. The trifunctional linker (**2-5**, 120.0 μ L, 500.0 mM in DMSO, 1000.0 eq.) was added, followed by DMTMM-BF₄ (150.0 μ L, 400.0 mM in DMSO, 1000.0 eq.). The solution was agitated at RT for 24 h. The mixture was purified by ethanol precipitation and the pool was dissolved in H₂O (540 μ L) to yield a 130 μ M solution. Success of the coupling was checked by HPLC analysis. Building block **NP07** was coupled independently since 2 coupling steps were necessary to achieve full conversion to the desired product. The procedure was conducted at the same reagent concentrations as with the pooled elements. Element **NP07** was pooled with the other 20 elements after successful modification to form the DNA-DE-1 pool.

Ester Hydrolysis of the DNA-DE-1 Pool



In a LowBind 1.5 mL Eppendorf tube the DNA-DE-1 pool (130.0 μ M in H₂O, 564.0 μ L, 1.0 eq.) was mixed with MeCN (340 μ L) and LiOH monohydrate (100.0 mM in H₂O, 110.0 μ L, 150.0 eq.). The solution was shaken at RT for 2.5 h, followed by ethanol precipitation. HPLC analysis showed full conversion of the starting material (peak shift). The pool was dissolved in MOPS buffer (648 μ L) to yield a 100 μ M solution.

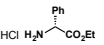
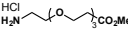
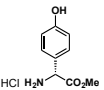

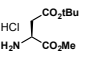

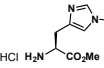

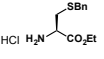

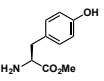

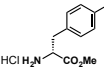

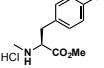

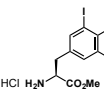

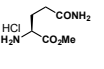

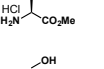

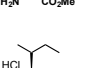

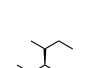

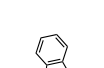

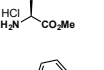

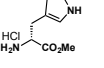

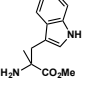

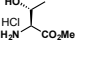

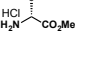

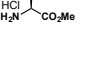

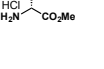

Amino Acid Test Couplings for Usage in the DEML Synthesis



In a PCR tube the encoded **NP01-TFL** (74.0 μ M, MOPS buffer, 2.0 μ L 1.0 eq.), NMM (250.0 mM, DMSO, 2.0 μ L, 3378.0 eq.), the amino acid (500.0 mM, DMSO, 0.3 μ L, 1014.0 eq.) and DMTMM-BF₄ (75.0 mM, DMSO, 2.0 μ L, 1014.0 eq.) were mixed. The reaction was left standing at RT for 19 h and then purified by EtOH precipitation. The DNA was dissolved in H₂O (6 μ L) and analyzed by HPLC (2 μ L) and LC-MS (4 μ L). 126 amino acids were tested with these conditions. Low yielding or non-converting building blocks were repeated with EDC/HOAt conditions (NMM was replaced by DIPEA, DMTMM-BF₄ was replaced by EDC hydrochloride (150.0 mM, 1.0 μ L) and HOAt (150.0 mM, 1.0 μ L)). In the case of the **TFL** with $n = 1$ the dehydroalanine side product (see **Figure 55a**) was found along with the desired products. These eliminated species were included in the conversion and purity yields of the amide coupling reaction.

Table 12. DE-2 amino acids screening results. 126 building blocks were tested for their reactivity in amide coupling. Unless stated otherwise the screening was performed with the short trifunctional linker ($n = 1$). Red coloring indicates the building blocks that were excluded from the final DEML assembly due to insufficient conversion and purity. Cutoff: >80% conversion and \geq 50% purity.

No.	Structure	Calc. Mass [m/z]	Obs. Mass [m/z]	Conversion (Purity)	No.	Structure	Calc. Mass [m/z]	Obs. Mass [m/z]	Conversion (Purity)
AA001		1077.3	1077.3	100% (70%)	AA064		1231.3 1233.3	1231.3 1233.3	100% (70%)
AA002		1077.3	1077.3	100% (92%)	AA065		1119.4	1119.4	100% (76%)
AA003		1077.3	1077.3	100% (66%)	AA066		1119.4	1119.4	100% (71%)
AA004		1111.3	1032.2 ^[a]	100% (0%)	AA067		1105.4	1105.4	100% (80%)
AA005		1091.3	879.3 ^[b] 976.3 ^[c]	100% (0%) 100% (0%) ^[d]	AA068		1192.4	1092.4 ^[g]	94% (69%)
AA006		1159.4	1159.4	100% (75%)	AA069		1192.4	1092.4 ^[g]	100% (66%)
AA007		1154.4	1154.3 1168.3 ^[e]	62% (34%) 100% (89%) ^[d]	AA070		1206.4	1106.4 ^[g]	95% (72%)
AA008		1241.5	1241.4	100% (90%)	AA071		1091.3	879.3 ^[b]	100% (<5%)
AA009		1105.4	1105.3	100% (83%)	AA072		1091.3	1091.4	100% (72%)
AA010		1105.4	1105.3	100% (82%)	AA073		1101.3	1101.4	100% (66%)
AA011		1119.4	1074.3 ^[f]	100% (0%)	AA074		1119.4	1119.3	90% (60%)
AA012		1063.3	1063.3	100% (92%)	AA075		1091.3	1091.4	100% (72%)
AA013		1077.3	1077.3	100% (82%)	AA076		1119.4	1119.4	94% (79%)
AA014		1139.3	1139.3	100% (75%)	AA077		1079.3	879 ^[b] 1024.3 ^[f]	100% (0%)

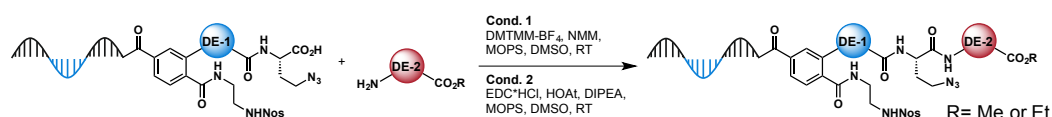
AA015		1153.4	1153.3	100% (79%)	AA078		1209.4	1209.4	95% (73%)
AA016		1155.3	1155.3	63% (39%) 100% (78%) ^[d]	AA079		1147.4	1147.4 976.3 ^[c]	95% (42%) 100% (32%) ^[d]
AA017		1177.4	1177.3	100% (73%)	AA080		1119.4	1119.4	100% (73%)
AA018		1157.4	1157.3	100% (91%)	AA081		1133.4	1133.4	95% (32%)
AA019		1213.4	1213.3	95% (73%)	AA082		1147.4	1147.4	90% (31%)
AA020		1169.4	1169.3 1183.3 ^[e]	63% (39%) 100% (62%) ^[d]	AA083		1102.3	1031.3 ^[f]	100% (0%)
AA021		1169.4	1169.3	71% (54%) 100% (71%) ^[d]	AA084		1170.3	1170.3	100% (69%)
AA022		1183.4	1138.3	100% (0%)	AA085		1159.3	879.3 ^[b]	100% (0%)
AA023		1421.2	1435.2 ^[e]	100% (35%) 100% (63%) ^[d]	AA086		1087.3	1087.4	93% (72%)
AA024		1134.4	1134.3	100% (89%)	AA087		1089.3	1089.4	100% (69%)
AA025		1093.3	1093.3	100% (86%)	AA088		1089.3	1089.4	100% (75%)
AA026		1093.3	1093.3	100% (89%)	AA089		1129.4	1129.4	100% (75%)
AA027		1119.4	1119.4	100% (72%)	AA090		1117.4	1117.4	97% (54%)
AA028		1133.4	1088.3 ^[f]	100% (0%)	AA091		1131.4	1131.4	92% (78%)
AA029		1192.4	1192.3	100% (71%)	AA092		1131.4	1131.4	88% (65%)
AA030		1192.4	1192.3	100% (78%)	AA093		1103.3	1103.4 1117.4 ^[e]	100% (45%) 100% (70%) ^[d]
AA031		1206.4	879.3 ^[b]	100% (<5%)	AA094		1089.3	1089.3	36% (36%) 100% (86%) ^[d]
AA032		1107.3	1107.3	100% (88%)	AA095		1159.4	879.3 ^[b]	100% (0%)
AA033		1107.3	1107.3	100% (90%)	AA096		1103.3	1103.4	90% (37%) 100% (24%) ^[d]
AA034		1107.3	1107.3	100% (90%)	AA097		1143.4	1143.3	98% (58%)
AA035		1107.3	1107.3	100% (90%)	AA098		1145.4	1145.4 1033.3 ^[f]	100% (<5%)

AA036		1207.4	1207.4	100% (86%)	AA099		1193.4	1193.3	100% (71%)
AA037		1207.4	1207.4	100% (89%)	AA100		1181.3	949.3 ^[a]	56% (0%)
AA038		1220.4	1120.4 ^[g]	94% (74%)	AA101		1225.5	1225.4	95% (85%)
AA039		1119.4	1119.3	100% (72%)	AA102		1143.4	1143.4	100% (60%)
AA040		1119.4	1119.4	100% (71%)	AA103		1159.4	1159.4	89% (43%)
AA041		1234.4	1134.4 ^[g]	98% (75%)	AA104		1148.3	910.2 ^[f]	95% (0%)
AA042		1268.4	1268.4	100% (85%)	AA105		1167.3	879.3 ^[b]	100% (0%)
AA043		1234.4	1134.4 ^[g]	23% (13%)	AA106		1240.4	1240.4	93% (76%)
AA044		1137.3	1137.3	100% (76%)	AA107		1160.3	1160.3	93% (66%)
AA045		1137.3	1137.3	100% (78%)	AA108		1204.4	1204.4	100% (71%)
AA046		1191.4	1191.4 ^[h]	100% (73%)	AA109		1251.4	1251.4	100% (73%)
AA047		1120.3	1120.3	100% (80%)	AA110		1203.4	1203.4	87% (71%)
AA048		1103.3	1103.3	95% (72%)	AA111		1167.4	1167.4	100% (63%)
AA049		1103.3	1103.3	100% (82%)	AA112		1167.4	1167.4	95% (52%)
AA050		1117.4	1117.3 879.3 ^[b] 976.3 ^[c]	100% (<5%) 100% (0%) ^[d]	AA113		1185.4	1185.4	95% (60%)
AA051		1117.4	1117.3	100% (<5%)	AA114		1179.4	1179.4	100% (50%)
AA052		1119.3	1119.3	100% (75%)	AA115		1197.4	1197.4	95% (73%)
AA053		1119.3	1119.3	100% (83%)	AA116		1181.4	1181.4	93% (51%)
AA054		1101.3	1101.3	100% (82%)	AA117		1199.4	1199.4	100% (60%)

AA055		1153.4	1153.3	100% (75%)	AA118		1187.3	1187.4	93% (50%)
AA056		1153.4	1153.3	100% (78%)	AA119		1229.4	1229.4	94% (58%)
AA057		1167.4	1167.3 879.3 ^[b] 976.3 ^[c]	100% (<5%) 100% (0%) ^[d]	AA120		1217.4	1217.4	94% (65%)
AA058		1198.3	1198.3	100% (74%)	AA121		1145.3	1145.3	95% (64%)
AA059		1198.3	1198.3	100% (57%)	AA122		1143.4	1143.3	100% (60%)
AA060		1167.4	879.3 ^[b] 1122.3 ^[f]	100% (<5%)	AA123		1247.4	1247.4	100% (76%)
AA061		1185.4	1185.4	100% (79%)	AA124		1143.4	1143.4	100% (74%)
AA062		1171.4	1171.4	100% (76%)	AA125		1129.4 ^[e]	1129.3	100% (69%)
AA063		1171.4	1171.4	100% (76%)	AA126		1174.3 ^[e]	1174.3	100% (77%)

[a] Elimination reaction(s) to dehydroalanine species. [b] Terminal amide side product. [c] Dihydrotriazinone. [d] EDC/HOAt/DIPEA reaction conditions. [e] Trifunctional linker with $n = 2$ [f] unknown species [g] Boc removal under ESI-MS conditions [h] *t*Bu ester hydrolysis in ESI-MS also observed

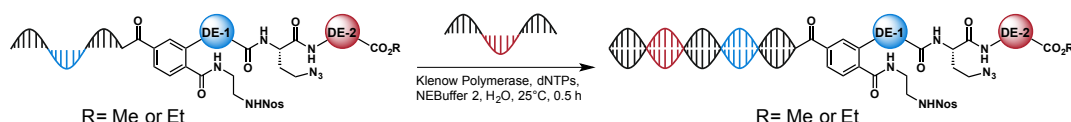
Amino Acid Couplings, DE-2 Attachment by Split Synthesis



The amide coupling reactions were performed in U-shaped 96 well-plates (Eppendorf Microplate 96/U-PP, white border, PCR clean, 250 μ L well-volume). The stated reagent amounts apply per well.

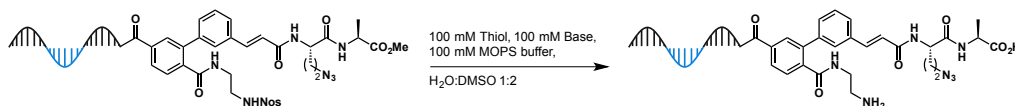
The DNA-DE-1 pool (90.0 μ M in MOPS buffer, 6.5 μ L, 1.0 eq.) was distributed in 102 wells on 2 plates. NMM (250.0 mM in DMSO, 8.2 μ L, 3500.0 eq.) was added, followed by the 102 amino acids (500.0 mM in DMSO, 1.2 μ L, 1000.0 eq.). DMTMM-BF₄ (150.0 mM in DMSO, 3.9 μ L, 1000.0 eq.) was added and the solutions were thoroughly mixed. The plates were slowly agitated at RT for 20 h. NaOAc buffer (3.0 M, pH 5.20, 2.5 μ L) was added, followed by EtOH (75 μ L). The plates were placed in the fridge overnight. The plates were centrifuged (3700 rpm, 4°C, 60 min) and the supernatant was discarded. The DNA pellets were washed with EtOH (75%, 2 x 50 μ L) and the plates were centrifuged (3700 rpm, 4°C, 2 x 40 min). The purified pellets were air-dried for 20 min. The DNA pellets were dissolved in H₂O (20 μ L) to yield 18-20 μ M solutions (average of 9 measurements). For Amino Acids **AA007**, **AA016**, **AA020**, **AA021**, **AA023**, **AA093** and **AA094** NMM was replaced by DIPEA (250.0 mM in DMSO, 8.2 μ L, 3500.0 eq.) and the coupling reagents were changed to EDC hydrochloride (300.0 mM in DMSO, 2.0 μ L, 1000.0 eq.) and HOAt (300.0 mM in DMSO, 2.0 μ L, 1000.0 eq.).

Klenow Encoding of the Coupled DE-2 Building Blocks



The encodings were performed in PCR tube 96 well-plates (Eppendorf twin.tec PCR Plate 96 LoBind, semi-skirted, 250 μ L well-volume). The amino acid coupling products (20.0 μ M in H₂O, 20.0 μ L) were transferred to the PCR 96 well-plates and diluted with H₂O (146 μ L) and 10X NEBuffer 2 (20 μ L). The DE-2-encoding DNA strand (100.0 μ M in H₂O, 8.0 μ L, 2.0 eq.) with the general sequence 5'-GTAGTTGGATCCGCACACACACACACATTCACACACACGTCC-3' was added and the solution was annealed by heating to 65°C for 5 min. YYY represents the coding sequence for the DE-2 amino acids and the underline shows the BamHI restriction site. The used annealing gradient was: 65°C(5 min)-50°C(0.5 min)-40°C(0.5 min)-25°C. The dNTPs mixture (10.0 mM in H₂O each dNTP, 4.0 μ L, 100.0 eq.) was added, followed by the Klenow polymerase (5000 U/ml, 2.0 μ L, 10.0 Units). The reaction was incubated at 25°C for 30 min while shaking (150 rpm). To stop the reaction, EDTA (500.0 mM, 5.0 μ L) was added and the mixture was incubated at 75°C for 20 min. The 102 encoded wells were combined in Eppendorf tubes (2 ml tubes, 14 pcs) and were concentrated to 8 ml by SpeedVac (45°C, 6 h). NaOAc buffer (3.0 M, pH 5.2, 960.0 μ L) was added, followed by EtOH (28.8 ml). The mixture was left in the fridge (4°C) overnight. Hereafter, the suspension was centrifuged (4200 rpm, 4°C, 60 min) and the supernatant was discarded. The pellet was suspended in EtOH (75%, 2 x 8 ml) and centrifuged (4200 rpm, 2 x 30 min, 4°C). The clean pellet was air-dried for 20 min. The DNA was dissolved in MOPS buffer (300.0 mM, 500.0 mM NaCl, pH 8.2, 800.0 μ L) to yield a 200 μ M solution. Success of the encoding was checked by native DNA polyacrylamide gel electrophoresis (12%, TBE, 150 V, 75 min, SYBR Gold staining).

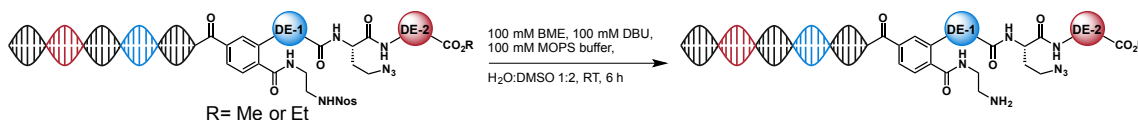
Optimization of the Ester and Nosyl Deprotection



Stock solutions: DNA test sample (87.0 μ M in 300.0 mM MOPS 500.0 mM NaCl buffer, 2.0 μ L, 1.0 eq.), thiol (300.0 mM in DMSO, 2.0 μ L, 3448.0 eq.), base (300.0 mM in DMSO, 2.0 μ L, 3448.0 eq.).

In a PCR tube all stock solutions were mixed and the solution was degassed with N₂ for 30 s, followed by incubation according to **Table 6**. The sample was purified by ethanol precipitation, dissolved in H₂O (6 μ L) and further analyzed by HPLC and LC-MS.

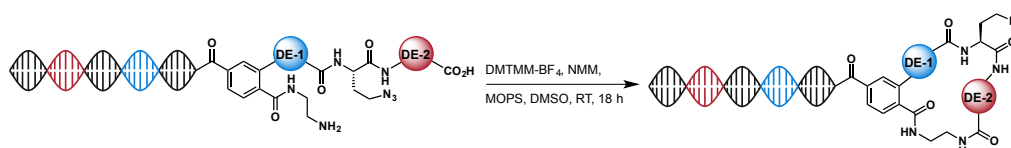
Ester and Deprotection of the DE-1/DE-2 sub-library



In a 5 mL Eppendorf tube the encoded library (800.0 μ L in 300.0 mM MOPS, 500.0 mM NaCl buffer), DBU (300.0 mM in DMSO, 800.0 μ L) and BME (300.0 mM in DMSO, 800.0 μ L) were mixed. The mixture was degassed with N₂ for 30 s, followed by agitating at RT for 6 h. A white precipitate formed after BME addition. The suspension was treated with NaOAc buffer (3.0 M, pH 5.2, 250.0 μ L) and EtOH (8.1 ml) and was placed in the fridge (4°C) overnight. The mixture was centrifuged (4200 rpm, 4°C, 60 min) and the supernatant was discarded. The pellet was washed with EtOH (75%, 2 x 1 ml) and centrifuged (4200 rpm, 4°C, 2 x 30 min), then dried in the

air for 30 min. The DEML was dissolved in MOPS buffer (700 μL) to yield a 79 μM solution. Success of the transformation was checked by HPLC (peak shift).

Macrocyclization of the Encoded DE-1/DE-2 Pool



In a 5 mL Eppendorf tube the DEML (900.0 μL , 56.0 μM in MOPS buffer) was diluted with MOPS buffer (1.4 mL) and DMSO (637 μL). NMM (50.0 mM in DMSO, 1.0 mL, 1000.0 eq.) and DMTMM-BF₄ (25.0 mM in DMSO, 605.0 μL , 300.0 eq.) were mixed and agitated at RT for 18 h. NaOAc buffer (3.0 M, pH 5.2, 450.0 μL) was added, followed by EtOH (15 mL). The suspension was placed on ice for 2 h, followed by centrifugation (4200 rpm, 4°C, 60 min) and the supernatant was discarded. The DNA pellet was washed with EtOH (75%, 2 x 4 mL), centrifuged (4200 rpm, 4°C, 2 x 30 min) and air-dried for 30 min. The DEML pellet was dissolved in H₂O (800 μL). The coupling was repeated once more. The DEML was purified by ssDNA digestion in 8 batches, followed by reversed phase preparative column chromatography. The library (100 μL) was diluted with nuclease-free H₂O (770 μL) and exonuclease 1 buffer (100 μL) was added, followed by the exonuclease 1 (*E. coli*) enzyme (20000 U/mL, 30 μL , 600.0 U). The solution was incubated at 37°C for 40 min and the enzyme was deactivated by heating to 80°C for 20 min. The crude material was directly purified by preparative reversed phase HPLC (Method B). Product-containing fractions were combined, lyophilized and dissolved in nuclease-free H₂O (1.2 mL, 26 μM). 100 μL of the DEML solution was diluted with TRIS (10.0 mM, pH 8.0) to 1.0 μM final concentration. The library was aliquoted (10 μL aliquots) and stored in the freezer (-20°C) for protein selection experiments (small DEML).

Note: The macrocyclization was repeated once, because the reference sample, which was synthesized simultaneously, showed incomplete conversion. Increase of the coupling reagent amount (up to 500 eq.) and DMSO amount (>55%) should yield full macrocyclization in a single step.

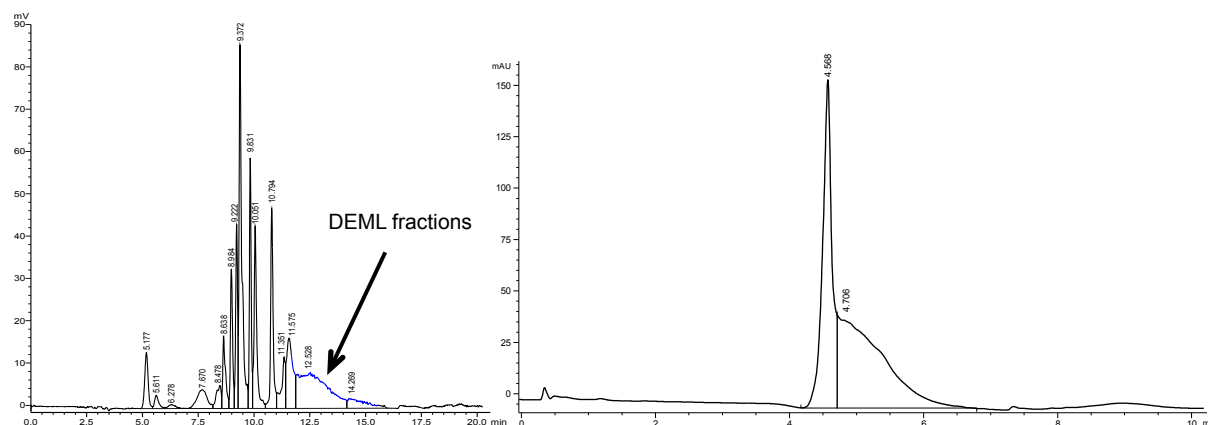
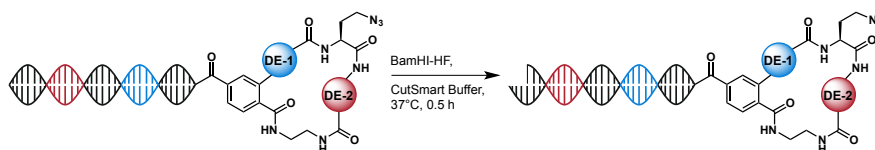


Figure 93. **Left:** Preparative HPLC chromatogram at 254 nm for the purification of the DEML. Blue indication shows the DEML containing fractions. **Right:** HPLC chromatogram at 254 nm of the purified small DEML. The large peak at 4.568 min contained some residual DE-2 coding DNA strands from the last encoding step.

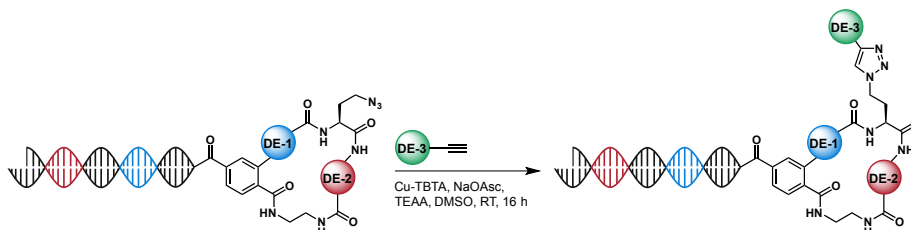
BamHI-HF Restriction Digest



The restriction digest of the DEML was performed in 20 batches; amounts per batch.

In an Eppendorf tube, DEML (55.0 μL , 26.0 μM in H_2O) was mixed with 10X CutSmart buffer (75 μL) and diluted with H_2O (565 μL). BamHI-HF (55.0 μL , 1100.0 U, 20000 U/mL) was added and incubated at 37°C for 30 min. The success of the reaction was checked by native polyacrylamide gel electrophoresis (12%, TBE, 150 V, 70 min). A solution of phenol:chloroform:isoamylalcohol 24:24:1 (740 μL) was added to the reaction and the mixture was vortexed for 2 min (phenol-chloroform extraction). The biphasic mixture was centrifuged (16900 g, RT, 5 min) and the aqueous layer was carefully removed. The aqueous phase was washed with chloroform (740 μL), vortexed for 2 min and centrifuged (16900 g, RT, 5 min). The aqueous layer was carefully removed and concentrated in the SpeedVac (45°C, 4 h). The 20 combined samples were diluted with H_2O to 4 mL total volume. NaOAc buffer (3.0 M, pH 5.2, 400.0 μL) was added, followed by EtOH (13.2 mL). The mixture was placed on ice for 1 h and in the freezer at -80°C for 1 h. The suspension was centrifuged (4200 rpm, 4°C, 1 h) and the supernatant was removed. The pellet was washed with EtOH (75%, 2 x 4 mL) and centrifuged (4200 rpm, 4°C, 2 x 20 min). The restricted DNA was dried in the air for 30 min. The DNA pellet was dissolved in TEAA buffer (500.0 mM, pH 7.2, 2.3 mL) to yield a 10.5 μM solution. The restricted library was used as obtained for further experiments.

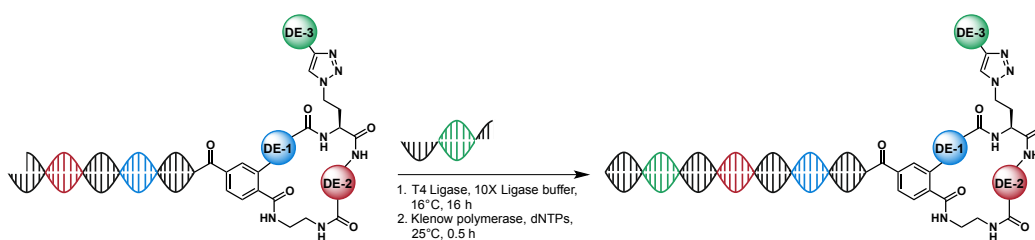
Click Reactions, DE-3 Incorporation by Split Synthesis



The reactions were performed in U-shaped 96 well-plates (Eppendorf Microplate 96/U-PP, white border, PCR clean, 250 μL well-volume). Columns 1 and 12 were left empty, 80 reactions per plate. 663 reactions were performed in 9 plates. The stated reagent amounts are given per well. Addition of the catalyst and the reducing agent was performed in a glove bag under an inert atmosphere (N_2).

The restricted DEML (3.5 μL , 10.5 μM in 500.0 mM TEAA buffer, pH 7.2) was mixed with the DE-3 alkyne (3.7 μL , 1.0 mM in DMSO, 100.0 eq.) and DMSO (2.5 μL). Under an inert atmosphere degassed NaOAsc solution (2.5 μL , 3.0 mM in H_2O , 200.0 eq.) was added, followed by Cu(II)-TBTA complex solution (2.5 μL , 3.0 mM in 55% DMSO/ H_2O , 200.0 eq.). The plates were sealed (Starlab StarSeal sealing tape polyolefin) and left standing in the glove bag at RT for 16 h. NaOAc buffer (3.0 M, pH 5.2, 2.5 μL) was added, followed by EtOH (65 μL). The plates were placed in the fridge over the weekend. The 96-well plates were centrifuged (3700 rpm, 4°C, 60 min) and the supernatant was removed. The DNA was washed with EtOH (75%, 2 x 50 μL) and centrifuged (3700 rpm, 4°C, 2 x 30 min). The DNA pellet was dissolved in nuclease-free H_2O (10 μL) for the following encoding reaction.

Encoding of DE-3 by T4 Ligation and Klenow Fill-in



The encoding was performed in V-shaped 96 well-plates (Eppendorf twin.tec PCR Plate 96 LoBind, semi-skirted, 250 μ L well-volume). The given amounts and volumes are valid per well. DE-3 encoding DNA strands (called insert DNA) consisted of a partially double-stranded DNA piece, that was preformed by the annealing of the coding strand with the sequence 5'-GTTCAAGCCACTTACCTZZZZZTGTATGCCTACCTATGAGA-3' and a 5'-phosphorylated strand with the sequence 5'-P-GATCCAAGTTCGGTGAATGGA-3'. ZZZZZ stands for the coding sequence of the DE-3 building blocks.

The DEML (10.0 μ L, approx. 3.7 μ M) was diluted with nuclease-free H₂O (15.3 μ L) and mixed with the insert DNA (2.2 μ L, 50.0 μ M in H₂O, 3.0 eq.). 10X ligase buffer (3.5 μ L) and T4 DNA ligase (4.0 μ L, 80.0 U, 20000 U/mL) were added. The 96-well plates were incubated at 16°C for 16 h and hereafter the reaction was stopped by heating to 65°C for 10 min. dNTPs solution (3.7 μ L, 1.0 mM in H₂O, 100.0 eq.) was added, followed by Klenow polymerase (4.0 μ L, 2.0 U, 500 U/mL) and incubated at 25°C for 30 min. The reaction was stopped by the addition of EDTA (125.0 mM in H₂O, 4.0 μ L) and heating to 75°C for 20 min. All 663 reactions were pooled in 2.0 mL Eppendorf tubes and were concentrated in the SpeedVac at 60°C. The concentrates were mixed with NaOAc buffer (3.0 M, pH 5.2, 840 μ L), followed by EtOH (28 mL). The mixture was put on ice for 3 h and kept in the fridge overnight. The suspension was centrifuged (4200 rpm, 4°C, 60 min) and the supernatant was discarded. The precipitate was washed with EtOH (75%, 2 x 8 mL) and centrifuged (4200 rpm, 4°C, 2 x 20 min). The DNA pellet was air-dried for 1 h and was dissolved in H₂O (1 mL). The encoded macrocycle library was purified by semi-preparative reversed phase HPLC (Method C). The product-containing fractions were combined, lyophilized and dissolved in 10 mM TRIS buffer pH 7.42 to a final concentration of 9.4 μ M (3890 μ L). The library was aliquoted (10 μ L) and stored in the freezer (-20°C).

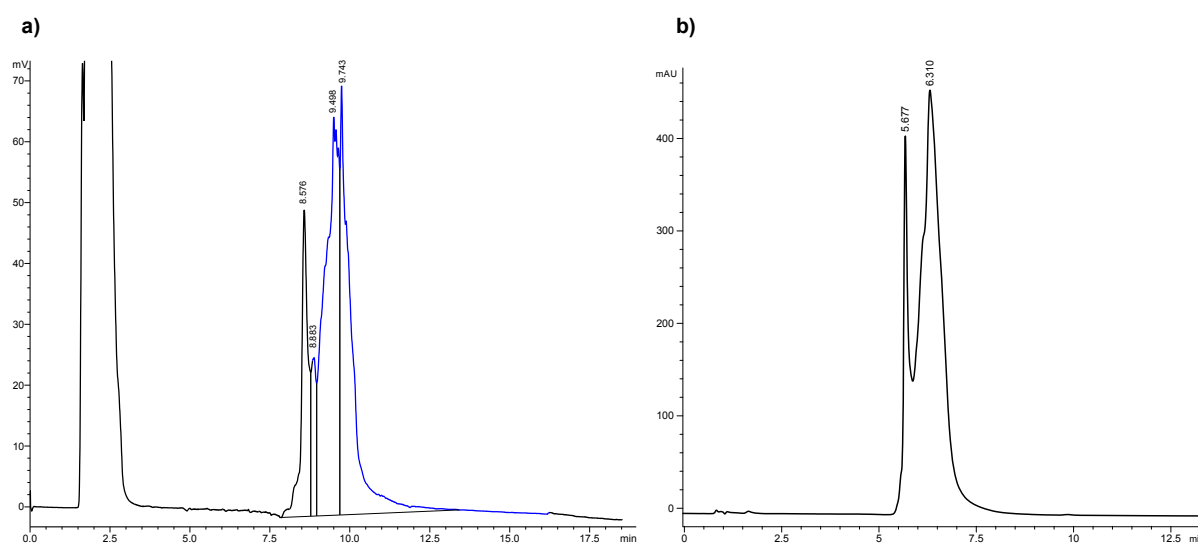


Figure 94. a) DEML purification on the semi-preparative column. Blue indication shows the library containing fractions. b) HPLC analysis of the purified final library. DEML and leftover DNA strands from the encoding steps overlap.

13.6. Quantitative PCR (qPCR) of Chemical Reactions on DNA

General Procedure

DNA damage evaluations were made with representative DNA strands, based on the same structure as the coding DNA strands of the DEML. For the evaluation of chemical modifications on a single-stranded DNA, we used DNA with the sequence 5'-GGAGCTTGTGAATTCTGGATGGACGTGTGTGAATTGTCTTTTGTGTGCGGATCCAAGTTCGGTGAATGGATTTTTTACTACG GATGGATACTCT-3' (**DNA1**).

Chemical modifications on a double-stranded system were evaluated on a DNA pair consisting of **DNA1** annealed to its reverse complementary strand with the sequence 5'-AGAGTATCCATCCG TAGTAAAAAATCCATTACCGAACTTGGATCCGCACACAAAAGACAATTCACACACGTCCCAT CCAGAATTCACAAGCTCC-3' (**DNA2**).

The annealing of **DNA1** and **DNA2** was performed with the following gradient: 95°C(1 min)-70°C(2 min)-45°C(2 min) to yield the double-stranded **DNA3**. **DNA1** and **DNA3** were diluted with H₂O to a final concentration of 0.45 ng/μL. **Primer1** had the following sequence: 5'-AGAGTATCCATCCGTAGT-3' and **Primer2**: 5'-GGAGCTTGTGAATTCTGG-3'.

For the qPCR assay the following solutions were mixed in a 96-well plate: **DNA1** or **DNA3** (3.0 μL, 0.45 ng/μL in H₂O), **Primer1** (0.5 μL, 2.5 μM in H₂O), **Primer2** (0.5 μL, 2.5 μM in H₂O), H₂O (3.5 μL) and SYBR Green Master Mix (5.0 μL). After mixing and centrifugation of the plate, the PCR was run according to the following program: 95°C(2 min)-40 x (95°C(15 s)-60°C(1 min)). The progress of the DNA amplification was observed by fluorescence emission of the DNA-SYBR complex formed during the PCR. All assays were performed with three or more valid replicates that were further processed. Standard curves with unmodified **DNA1** or **DNA3** at four concentrations (1X, 10X, 100X, 1000X dilutions) were generated as well as dummy samples (no primers and primer only). The quantity of remaining intact DNA could be calculated by the number of PCR cycles necessary for complete consumption of the primers of each assay relative to the generated standard curves. For this purpose the standard curves were fitted with logarithmic trendlines in Excel. For single-stranded assays we calculated the following trendline: $y = -4.384 \ln(x) + 7.4961, R^2 = 0.99126$, with y = the number of PCR cycles and x = the remaining intact DNA quantity. For double-stranded DNA we calculated the trendline: $y = -3.421 \ln(x) + 11.458, R^2 = 0.98647$.

Ester Deprotection Conditions

In a PCR tube **DNA1** (3.0 μL, 110.0 μM in H₂O, 1.0 eq.) was mixed with H₂O (2 μL), MeCN (3.5 μL) and a solution of LiOH monohydrate (1.65 μL, 100.0 mM in H₂O, 500.0 eq.). The mixture was left standing at RT for 16 h, followed by EtOH purification. The washed DNA pellet was dissolved in H₂O (30 μL) and further diluted to the desired concentration for qPCR.

Click Reaction Conditions

In a PCR tube **DNA1** (3.0 μL, 110.0 μM in H₂O, 1.0 eq.) was mixed with TEAA (1.0 M, pH 7.2, 3.0 μL), DMSO (7 μL), a solution of TA662 (3.3 μL, 10.0 mM in DMSO, 100.0 eq.) and a solution of sodium ascorbate (3.3 μL, 20.0 mM in H₂O, 200.0 eq.). The mixture was degassed for 30 s by bubbling a stream of N₂ gas through the solution. Cu(II)-TBTA complex solution (6.6 μL, 10.0 mM in 55% DMSO, 200.0 eq.) was added and the mixture was degassed again for 30 s. The mixture was left standing at RT for 16 h, followed by EtOH purification. The washed DNA pellet was dissolved in H₂O (30 μL) and was further diluted to the desired concentration for qPCR. For the double-stranded assay **DNA1** was replaced by **DNA3** (6.0 μL, 50.0 μM, 1.0 eq.)

DMTMM Coupling with Excess Amine Conditions

In a PCR tube **DNA1** (3.0 μ L, 110.0 μ M in H₂O, 1.0 eq.) was mixed with MOPS buffer (300.0 mM, pH 8.2, 0.5 μ L), NMM (2.2 μ L, 300.0 mM in DMSO, 2000.0 eq.), AA001 (3.3 μ L, 100.0 mM in DMSO, 1000.0 eq.) and DMTMM-BF₄ (1.7 μ L, 200.0 mM in DMSO, 1000.0 eq.). The mixture was left standing at RT for 16 h, followed by EtOH purification. The washed DNA pellet was dissolved in H₂O (30 μ L) and further diluted to the desired concentration for qPCR. For the double-stranded assay **DNA1** was replaced by **DNA3** (6.0 μ L, 50.0 μ M, 1.0 eq.)

DMTMM Coupling with Excess Acid Conditions

In a PCR tube **DNA1** (3.0 μ L, 110.0 μ M in H₂O, 1.0 eq.) was mixed with MOPS buffer (300.0 mM, pH 8.2, 0.5 μ L), NMM (2.2 μ L, 300.0 mM in DMSO, 2000.0 eq.), 1-*H*-indazole-3-carboxylic acid (3.3 μ L, 100.0 mM in DMSO, 1000.0 eq.) and DMTMM-BF₄ (1.7 μ L, 200.0 mM in DMSO, 1000.0 eq.). The mixture was left standing at RT for 16 h, followed by EtOH purification. The washed DNA pellet was dissolved in H₂O (30 μ L) and further diluted to the desired concentration for qPCR. For the double-stranded assay **DNA1** was replaced by **DNA3** (6.0 μ L, 50.0 μ M, 1.0 eq.)

EDC/HOAt Coupling with Excess Amine Conditions

In a PCR tube **DNA1** (3.0 μ L, 110.0 μ M in H₂O, 1.0 eq.) was mixed with MOPS buffer (300.0 mM, pH 8.2, 0.5 μ L), DIPEA (3.7 μ L, 180.0 mM in DMSO, 2000.0 eq.), AA001 (3.3 μ L, 100.0 mM in DMSO, 1000.0 eq.), EDC hydrochloride (1.7 μ L, 200.0 mM in DMSO, 1000.0 eq.) and HOAt (1.7 μ L, 200.0 mM in DMSO, 1000.0 eq.). The mixture was left standing at RT for 16 h, followed by EtOH purification. The washed DNA pellet was dissolved in H₂O (30 μ L) and further diluted to the desired concentration for qPCR. For the double-stranded assay **DNA1** was replaced by **DNA3** (6.0 μ L, 50.0 μ M, 1.0 eq.)

EDC/HOAt Coupling with Excess Acid Conditions

In a PCR tube **DNA1** (3.0 μ L, 110.0 μ M in H₂O, 1.0 eq.) was mixed with MOPS buffer (300.0 mM, pH 8.2, 0.5 μ L), DIPEA (3.7 μ L, 180.0 mM in DMSO, 2000.0 eq.), 1-*H*-indazole-3-carboxylic acid (3.3 μ L, 100.0 mM in DMSO, 1000.0 eq.), EDC hydrochloride (1.7 μ L, 200.0 mM in DMSO, 1000.0 eq.) and HOAt (1.7 μ L, 200.0 mM in DMSO, 1000.0 eq.). The mixture was left standing at RT for 16 h, followed by EtOH purification. The washed DNA pellet was dissolved in H₂O (30 μ L) and further diluted to the desired concentration for qPCR. For the double-stranded assay **DNA1** was replaced by **DNA3** (6.0 μ L, 50.0 μ M, 1.0 eq.)

Nosyl Deprotection Conditions

In a PCR tube **DNA1** (3.0 μ L, 110.0 μ M in H₂O, 1.0 eq.) was mixed with MOPS buffer (600.0 mM, pH 8.2, 3.0 μ L), DBU (6.0 μ L, 300.0 mM in DMSO, 5455.0 eq.) and BME (6.0 μ L, 300.0 mM in DMSO, 5455.0 eq.). The mixture was degassed by bubbling N₂ through the solution for 30 s and then left standing at RT for 16 h, followed by EtOH purification. The washed DNA pellet was dissolved in H₂O (30 μ L) and further diluted to the desired concentration for qPCR.

13.7 Protein Biotinylation of HSA and AGP

Protein biotinylation was performed according to a modified procedure from Scheuermann et al.^[164] The given procedure below was applied for HSA and AGP modifications.

In a 1.5 mL Eppendorf tube the protein (1.0 mg/ml in PBS buffer, 1.0 ml) was mixed with NHS-LC-Biotin (10.0 mM in DMSO, 20.0 eq.) and the solution was incubated at 4°C for 3 h (Thermo Block, 500 rpm shaking). The reaction was quenched by the addition of TRIS hydrochloride buffer (1.0 M, pH 7.42, 66.0 eq.) and was incubated at 4°C for another 1 h. PBS buffer (1 mL) was added and the reaction buffer was exchanged using a Sartorius Vivaspin 2 10000 MWCO CTA column (4200 rpm, 4°C, 4 x 20 min, 3 x 2 mL PBS). The protein solution was collected and PBS was added to get a final protein concentration of 10 µM. The protein solution was aliquoted (20 µL aliquots) and stored in the freezer at -20°C. To check the success of the modification, a biotinylation band shift assay was performed.

The modified protein (7.6 µL) was diluted with PBS (37.4 µL) and mixed with an equimolar amount of avidin (1.0 mg/ml in PBS, 5.0 µL). The solution was left standing at RT for 5 min and then placed on ice for 1 h. The result of the experiment was controlled by SDS-PAGE (coomassie stain). Only the indicated control samples were treated by heat denaturation prior to loading.

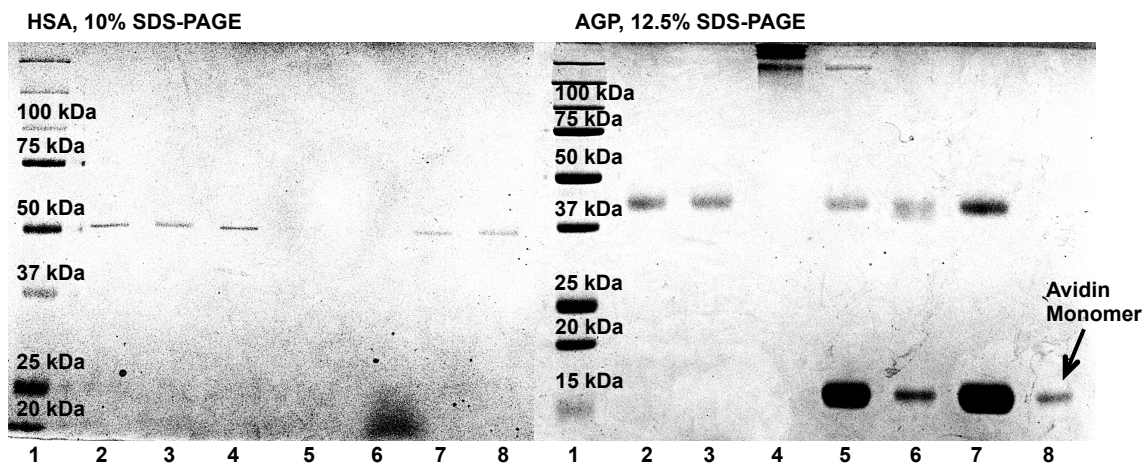


Figure 95. Protein biotinylation band shift assays, coomassie stain. **Left:** HSA band shift, **L1:** Protein standards, **L2:** HSA unmodified, **L3:** HSA unmodified + avidin, **L4:** HSA modified, **L5:** HSA modified + avidin, **L6:** Avidin, **L7:** HSA unmodified + avidin (95°C), **L8:** HSA modified + avidin (95°C). **Right:** AGP band shift, **L1:** Protein standards, **L2:** AGP unmodified, **L3:** AGP modified, **L4:** AGP modified + avidin, **L5:** AGP modified + avidin (95°C), **L6:** AGP unmodified + avidin, **L7:** AGP unmodified + avidin (95°C), **L8:** Avidin. **Note:** Avidin tetramer is not visible since it stuck to the stacking gel.

13.8 DEML Protein Target Selections

HSA and AGP Selections

Stock solutions:

- **Biotin solution:** 200 mM *D*-biotin in DMSO (2 mL)
- **PBST:** 0.05% Tween-20 in PBS (10 mL)
- **PBST-Biotin:** 1:1 mixture of biotin stock solution + 50% DMSO in PBST (2 mL)→ do not put on ice!
- **Salmon Sperm DNA:** 2 mg/mL in PBS
- **PBST-SS:** PBST containing 200 µg/mL salmon sperm DNA. Mix 1.8 mL PBST with 200 µL salmon sperm DNA stock solution (2 mL).
- **DEML-WS:** DEML working solution. DEML stock (9.4 µM, 10.0 µL) was mixed with PBST-SS (50 µL) and diluted with PBST (940 µL) to yield the working solution (1.0 mL, 94.0 nM for 10 assays).
- **Protein solution:** Biotinylated protein sample (10.0 µL, 10.0 µM) was diluted with PBS (90 µL) to the final concentration (1 µM, 100.0 µL)

All stock solutions were prepared as stated above. The DEML-WS and the protein solution were kept on ice. 100 µg hydrophilic streptavidin magnetic beads (25.0 µL, NEB S1421S) were placed in a magnetic rack and the buffer was discarded. The beads were washed with PBS (3 x 1 mL). The beads were kept in 1 mL PBS on ice until usage.

The buffer was removed from the beads and the protein solution (100.0 µL, 100.0 pmol) was added. The mixture was incubated at 4°C for 30 min while gently mixing by rotation. The tube was placed in the magnetic rack and the buffer was discarded. The beads were suspended in PBST-Biotin (2 x 200 µL) and the supernatants were removed. The beads were washed with PBST (200 µL), 10% PBST in DMSO (200 µL, to remove biotin precipitate) and PBST (2 x 200 µL). The beads were transferred to a fresh tube and DEML-WS (100.0 µL, 9.4 pmol DEML) was added. The mixture was incubated at 4°C for 1 h while gently mixing by rotation. The tube was transferred to the magnetic rack and the supernatant was discarded. The beads were washed with PBST (5 x 200 µL), exchanging the tube after 2 washing steps (reduce carryover of unbound DNA). The washed beads were suspended in TRIS buffer (10.0 mM, pH 8.5, 100.0 µL) and the bound DNA was eluted from the protein by heat denaturation (95°C for 10 min). After cooling, the magnetic beads were removed and the buffer solution was stored in a 1.5 mL Eppendorf tube at -20°C. PCR amplification and high-throughput sequencing were used for analyzing the target binding results.

CA9 Selections

Stock solutions:

- **PBST:** 0.05% Tween-20 in PBS (10 mL)
- **PBST-Im:** PBST with 10 mM imidazole
- **Salmon Sperm DNA:** 2 mg/mL in PBS
- **PBST-SS:** PBST containing 200 µg/mL salmon sperm DNA. Mix 1.8 mL PBST with 200 µL salmon sperm DNA stock solution (2 mL).
- **DEML-WS:** DEML working solution. DEML stock (9.4 µM, 10.0 µL) was mixed with PBST-SS (50 µL) and diluted with PBST (940 µL) to yield the working solution (1.0 mL, 94.0 nM for 10 assays).
- **Protein solution:** His-tagged CA9 (2.5 µg, 50.0 pmol) was reconstituted with sterile H₂O (17.5 µL) and was diluted with PBST (32.5 µL) to yield the final concentration (1.0 µM, 50 µL).

All stock solutions were prepared as stated above. The DEML-WS and the protein solution were kept on ice. 250 µg His magnetic beads (6.25 µL, Invitrogen Dynabeads, 10103D) were placed in a magnetic rack and the buffer was discarded. The beads were washed with PBS (3 x 1 mL). The beads were kept in 1 mL PBS on ice until usage.

The buffer was removed from the beads and the protein solution (50.0 µL, 50.0 pmol) was added. The mixture was incubated at 4°C for 30 min while gently mixing by rotation. The tube was placed in the magnetic rack and the buffer was discarded. The beads were suspended in PBST-Im (2 x 100 µL) and the supernatants were removed. The beads were washed with PBST (3 x 200 µL). The beads were transferred to a fresh tube and DEML-WS (100.0 µL, 9.4 pmol DEML) was added. The mixture was incubated at 4°C for 1 h while gently mixing by rotation. The tube was transferred to the magnetic rack and the supernatant was discarded. The beads were washed with PBST (5 x 200 µL), exchanging the tube after 2 washing steps (reduce carryover of unbound DNA). The washed beads were suspended in TRIS buffer (10.0 mM, pH 8.5, 100.0 µL) and the bound DNA was eluted from the protein by heat denaturation (95°C for 10 min). After cooling, the magnetic beads were removed and the buffer solution was stored in a 1.5 mL Eppendorf tube at -20°C. PCR amplification and high-throughput sequencing were used for analyzing the target binding results.

Note: For the smDEML the procedure was analog to the described procedure. Due to the higher purity and smaller number of individual members the small DEML-WS was prepared from the 1.0 µM small DEML solution, giving a final work solution concentration of 10.0 nM. Target binding assays were performed in duplicates with two freshly prepared working solutions that were used on two consecutive days. In the CA9 DEML-WS we spiked in a known CA9 binder in a tenfold excess. The positive binder was synthesized according to the established procedures (see Chapter 13.4) and was composed of building blocks **NP01**, **AA001** and **TA664**. Spiking was only performed with one selection (protein + dummy) of the DEML. Performed assays with AGP and HSA: 2 x smDEML, 2 x DEML, 4 x dummy selections (beads only, no bound target proteins). In total: 12 selection assays. Selections performed with CA9: 2 x smDEML, 2 x DEML, 4 x dummy selections. In total 8 assays.

13.9 PCR Amplification of Eluted DNA

PCR 1

Eluted DNA from selection:	5 µL	UniPrimAdapt (7.5 µM):	4 µL
Phusion HF buffer (5x):	10 µL	IndexPrimAdapt (10.0 µM):	3 µL
Phusion MgCl₂ (50.0 mM):	2 µL	Phusion (2U/µL):	0.25 µL
dNTPs (10.0 mM):	1.25 µL	H₂O:	24.5 µL

In a PCR vial all stock solutions were mixed and kept on ice. The polymerase was added shortly before the PCR program started. 22 PCR cycles were performed. PCR cycle program: 98°C for 2 min 15 s then 22 x 98°C (45 s) → 69°C (45 s) → 72°C (45 s), 72°C (5 min). The progress of the reactions was controlled by native PAGE analysis (10%, TBE, 150 V, 45 min, SYBR Gold staining). The samples were purified with the PCR clean up kit (Macherey&Nagel). Two washing steps were included and the purified PCR products were eluted with NE buffer (2 x 20 µl). The PCR products were diluted with NE buffer to a final concentration of 100 nM.

For the fingerprint sample, an aliquot of the DEML (small DEML) working solution was diluted 1:10 and processed as stated above.

UniPrimAdapt: 5'-ACACGACGCTCTCCGATCTGTAGTTGGATCCGCAC-3' (small DEML)

5'-ACACGACGCTCTCCGATCTAGAGTATCCATCCGTA-3' (DEML)

IndexPrimAdapt: 5'-GACGTGTGCTCTCCGATCTGGAGCTTGTGAATTCT-3' (both libraries the same)

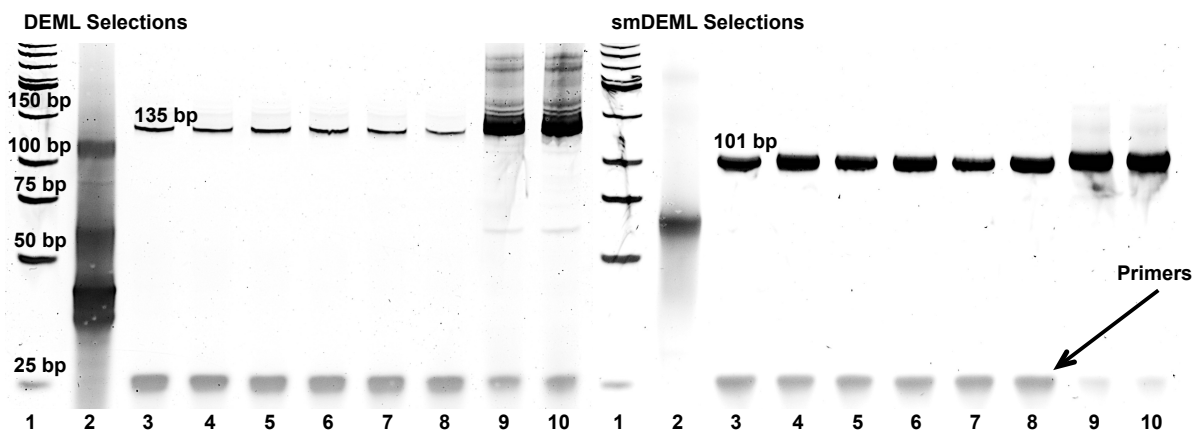


Figure 96. 10% Native PAGE gels after PCR1 with the DEML selections (**left**) and the small DEML selections (**right**) with HSA and AGP, SYBR gold staining. **L1:** Low MW DNA ladder, **L2:** (small) DEML, **L3 & L4:** HSA selections, **L5 & 6:** Dummy selections, **L7 & 8:** AGP selections, **L9 & 10:** Library fingerprints.

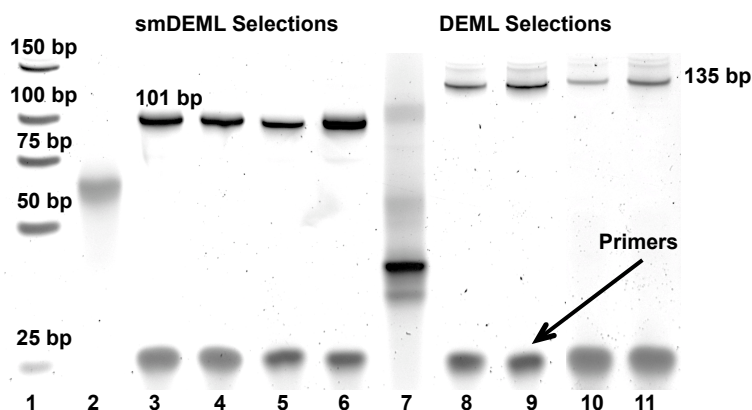


Figure 97. 10% Native PAGE gels after PCR1 with the smDEML selections (**left**) and the DEML selections (**right**) with CA9, SYBR gold staining. **L1:** Low MW DNA ladder, **L2:** smDEML, **L3 & L5:** CA9 selections, **L4 & L6:** Dummy selections, **L7:** DEML, **L8 & L10:** CA9 selections, **L9 & L11:** Dummy selections.

PCR2

PCR products from PCR1 were diluted 1:10 with NE buffer prior to PCR2

PCR 1 DNA (10.0 nM):	10 μ L	Universal Primer (10.0 μM):	3 μ L
Phusion HF buffer (5x):	20 μ L	Index Primer (10.0 μM):	3 μ L
Phusion MgCl₂ (50.0 mM):	4 μ L	Phusion (2U/μL):	0.5 μ L
dNTPs (10.0 mM):	2.5 μ L	H₂O:	57 μ L

In a PCR vial all stock solutions were mixed and kept on ice. The polymerase was added shortly before the PCR program started. 15 PCR cycles were performed. PCR cycle program: 98°C for 2 min 15 s then 15 x 98°C (45 s) \rightarrow 69°C (45 s) \rightarrow 72°C (45 s), 72°C (5 min). The progress of the reactions was controlled by native PAGE analysis (8%, TBE, 150 V, 45 min, SYBR Gold staining). The samples were purified by agarose gel electrophoresis (2%, TBE, 100 V, 45 min) and the desired bands were cut out of the gel with subsequent DNA isolation with the PCR clean up kit (Macherey&Nagel). Two washing steps were included and the purified PCR products were eluted with NE buffer (2 x 20 μ l). After ethanol precipitation the amplified and indexed products were diluted with NE buffer to a final concentration of 100 nM.

Primer sequences for Illumina Sequencing (NEBNext Multiplex Oligos for Illumina, Sets 1 and 3):

UniPrimer: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'

IndexPrimers: 5'-
CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'

NNNNNN represents the indexing codons (selection ID) defined by the NEB Kits.

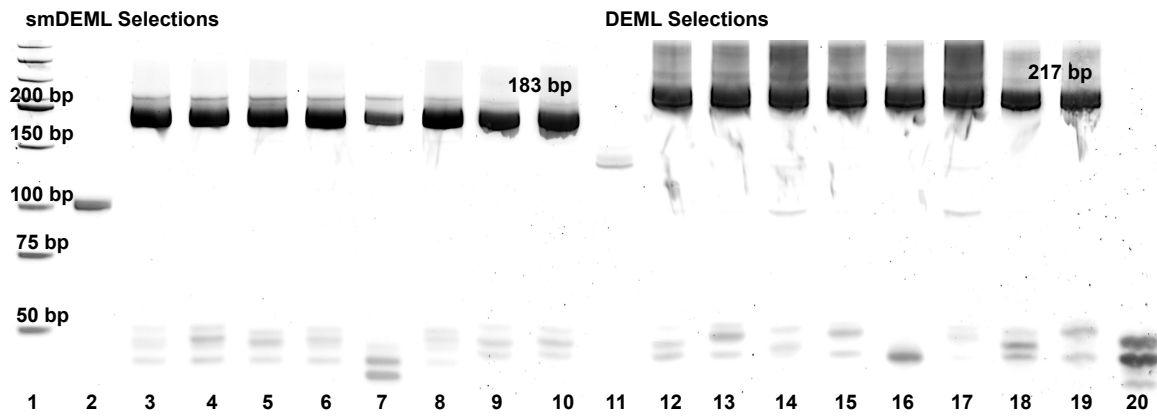


Figure 98. 8% Native PAGE gels after PCR2 amplification of HSA and AGP selections. L1: Low MW DNA ladder, L2 & L11: PCR1 (HSA selection), L3, L6, L12, L15: HSA selections, L4, L7, L13, L16: Dummy selections, L5, L8, L14, L17: Fingerprints, L9, L10, L18, L19: AGP selections, L20: Primers.

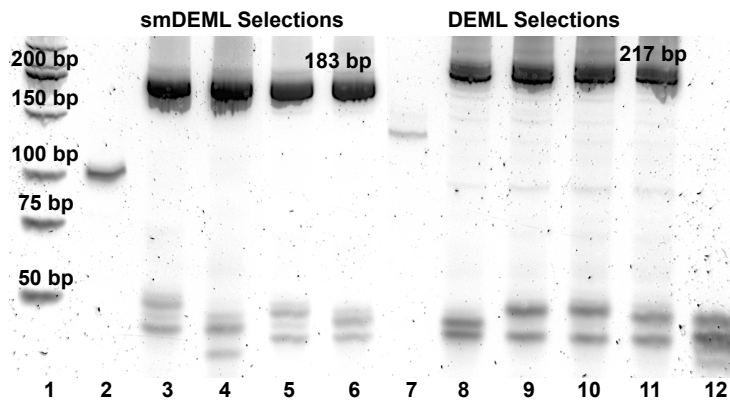


Figure 99. 8% Native PAGE gels after PCR2 amplification of CA9 selections. L1: Low MW DNA ladder, L2 & L7: PCR1, L3, L5, L8, L10: CA9 selections, L4, L6, L9, L11: Dummy selections, L12: Primers.

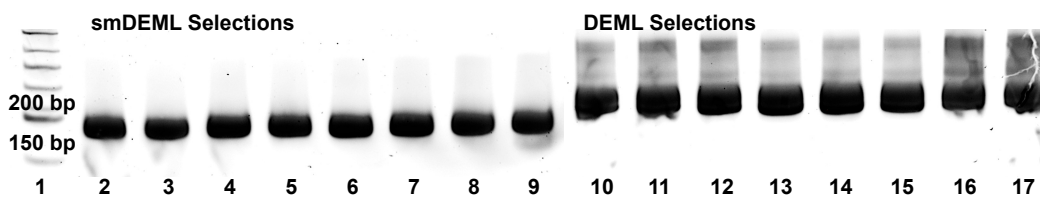


Figure 100. 8% Native PAGE gels after PCR2 purification of HSA and AGP selections. L1: Low MW DNA ladder, L2, L3, L10, L11: HSA selections, L4, L5, L12, L13: Dummy selections, L6, L7, L14, L15: AGP selections, L8, L9, L16, L17: Fingerprints.

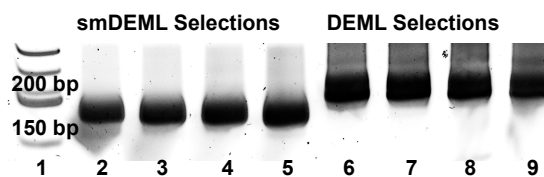


Figure 101. 8% Native PAGE gel after PCR2 purification of CA9 selections. L1: Low MW DNA ladder, L2, L3, L6, L7: CA9 selections, L4, L5, L8, L9: Dummy selections.

13.10 Next Generation Sequencing (NGS)^[227]

In order to pool the individual selections, the DNA concentration of each sample was measured to determine its concentration. The separate selections were then pooled in equal amounts to form two pools, one with the small library and one with the full library. These two libraries were pooled in equal amounts again and sent to Novogene for next generation sequencing with a targeted read count of 240 million reads and paired-end 150 sequencing.

We extracted the codon hit counts from next generation sequencing data files (.fq) with our DECL-Gen software suite on the University of Basel computational cluster (SciCore). The script aligns the theoretical DNA strand to the read with the pairwise2 algorithm provided by the Biopython package, and then extracts the codons from the aligned read. This is done for the mate pair, too, and the two codon sets are controlled for identity, discarding the read if they do not match. After extraction, the counts for both replicates have been averaged and the list was cleaned from non-existing codon combinations deriving from random mutations during PCR or from sequencing errors. The results for each data set are summarized in **Table 13**.

Examples

Expected correct read (black), found (green); the Illumina adapter is marked with underline, the codons are in bold. The missing three bases in the found read is due to the read limit of 150 bp.

```
GGAGCTTGTGAATTCTGGxxxGGACGTGTGTGAATTGTCyyyyGTGTGCGGATCCAAGTTCGGTGAATGGAzzzzzzACTACGGATG
GGAGCTTGTGAATTCTGGAGCGGACGTGTGTGAATTGTCACAAGTGTGCGGATCCAAGTTCGGTGAATGGATTACGATACTACGGATG
GATACTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT
GATACTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCTGTGGTCGCCGTATC
```

Expected correct read (black), found with mistakes (red). Same markings apply as above. As the third codon is only partially preserved, this read cannot be used for evaluation and is discarded.

```
GGAGCTTGTGAATTCTGGxxxGGACGTGTGTGAATTGTCyyyyGTGTGCGGATCCAAGTTCGGTGAATGGAzzzzzzACTACGGATG
GGAGCTTGTGAATTCTGGAAGGGACGTGTGTGAATTGTCCAATTTGTGCGGATCCAAGTTCGGTGAATGGAAA_____ACGGATG
GATACTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT
GATACTCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATAAA
```

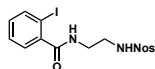
Table 13. Summary of the enrichment assays that were analyzed by NGS. The counts of the different enrichments were averaged by replicate and then normalized by their mean.

No	Target	Replicate	Library Size	Reads ^[a]	Matching Codon Pairs ^[b]	Coverage ^[c]	Invalid Codon Content ^[d]	Found Codons ^[e]
1	Fingerprint	1	2'142	15.77	13.41	6260	2.6%	2142 (100%)
2		2	2'142	20.03	17.38	8114		
3	Beads	1	2'142	17.23	15.17	7082	2.1%	2142 (100%)
4		2	2'142	15.14	13.20	6160		
5	HSA	1	2'142	15.81	13.83	6457	2.1%	2142 (100%)
6		2	2'142	15.21	13.28	6200		
7	AGP	1	2'142	17.43	15.22	7106	2.1%	2142 (100%)
8		2	2'142	16.76	14.47	6755		
9	Ni-Beads	1	2'142	17.5	15.4	7190	2.3%	2142 (100%)
10		2	2'142	19.4	16.6	7750		
11	CA9	1	2'142	15.1	13.1	6116	2.4%	2142 (100%)
12		2	2'142	17.9	15.6	7283		
13	Fingerprint	1	1'420'146	13.07	8.55	6.02	6.7%	1'318'952 (92.9%)
14		2	1'420'146	11.95	7.87	5.54		
15	Beads	1	1'420'146	12.02	8.87	6.25	4.2%	819'242 (57.7%)
16		2	1'420'146	12.30	8.97	6.32		
17	HSA	1	1'420'146	12.33	9.06	6.38	4.3%	838'791 (59.1%)
18		2	1'420'146	13.28	9.77	6.88		
19	AGP	1	1'420'146	10.82	8.02	5.65	4.0%	788'545 (55.5%)
20		2	1'420'146	11.54	8.54	6.01		
21	Ni-Beads	1	1'420'146	12.9	8.66	6.10	6.3%	756'037 (53.2%)
22		2	1'420'147	10.9	7.55	5.32		
23	CA9	1	1'420'146	10.4	6.85	4.82	6.1%	655'821 (46.2%)
24		2	1'420'147	11.8	8.09	5.70		

[a] Read counts are given in million mate reads. [b] In million reads. Only reads where both read mates have the same codons were considered. [c] Coverage is an indication of how many reads per codon can be expected if all compounds are distributed equally in the sample. It was calculated by the following equation: "Coverage" = "Matching codon pairs" / "Library Size". [d] This number indicates how many codons were found that did not encode for a compound within a library. [e] Depending on sequencing depth, a part of the library population never gets sequenced in both replicates even if they are equally distributed. A lower number of found codons indicated a strong deviation in the population distribution due to enrichment.

13.11 Macrocycle Resynthesis Procedures

2-Iodo-N-(2-((2-nitrophenyl)sulfonamido)ethyl)benzamide **2-24** (Nos-Sc-I)



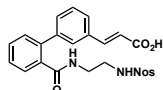
In a 250 mL flask 2-iodobenzoic acid (8.2 g, 33.1 mmol, 1.0 eq.), **2-3** (12.2 g, 49.6 mmol, 1.5 eq.), HBTU (15.0 g, 39.7 mmol, 1.2 eq.) and DIPEA (17.3 mL, 99.2 mmol, 3.0 eq.) were dissolved in THF (150 mL) and stirred at RT for 2 h, after which UPLC-MS analysis showed 93% conversion to the desired product (HBTU dissolved after about 30 min reaction time). The solvent was removed *in vacuo* and the residue was dissolved in EtOAc (250 mL). The organic layer was washed with diluted HCl (20 mM, 3 x 100 mL), half-saturated NaHCO₃ (3 x 100 mL) and brine (100 mL). The organic layer was dried over MgSO₄ and concentrated. The crude was purified by flash column chromatography on the ISOLERA (Silica, 2 x 340 g, cyclohexane:EtOAc 20%→70% EtOAc, UV). The product containing fractions were combined and concentrated. The obtained solid was dissolved in MeCN (100 mL), filtered over a G4 glass sintered funnel and concentrated to yield the desired product as a yellow solid (14.7 g, **94%**).

¹H NMR (400 MHz, CD₃CN) δ/ppm: 8.10 – 8.03 (m, 1H), 7.91 – 7.83 (m, 2H), 7.83 – 7.76 (m, 2H), 7.41 (ddd, *J* = 7.5, 7.5, 1.1 Hz, 1H), 7.32 (dd, *J* = 7.6, 1.7 Hz, 1H), 7.14 (ddd, *J* = 7.9, 7.4, 1.8 Hz, 1H), 6.90 (brs, 1H), 6.21 (t, *J* = 5.9 Hz, 1H), 3.44 (dt, *J* = 5.8, 6.0 Hz, 2H), 3.30 – 3.22 (dt, 6.0, 6.1 Hz, 2H).

¹³C NMR (101 MHz, CD₃CN) δ/ppm: 170.75, 143.29, 140.57, 135.19, 135.10, 133.88, 133.80, 131.97, 131.40, 129.15, 128.88, 126.02, 93.03, 43.98, 40.27.

HRMS (ESI): C₁₅H₁₄IN₃NaO₅S⁺ *calcd*: 497.9591, *found*: 497.9597.

Nos-Sc-NP01-OH **2-18a**



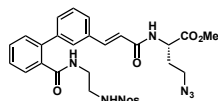
In a 250 mL flask **2-24** (3.0 g, 6.3 mmol, 1.0 eq.), 3-(2-carboxyvinyl)benzeneboronic acid (1.8 g, 9.5 mmol, 1.5 eq.), Pd (π-cinnamyl) chloride dimer (327.0 mg, 631.0 μmol, 10 mol%) and K₃PO₄ (4.0 g, 18.9 mmol, 3.0 eq.) were dissolved in EtOH (56 mL) and H₂O (24 mL). The solution was stirred at 50°C for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed and the residue was taken up in EtOAc (200 mL) and H₂O (150 mL). The mixture was acidified with concentrated HCl to pH 1. The biphasic mixture was separated and the aqueous layer was extracted with EtOAc (2 x 100 mL, 1 x 50 mL). The combined organic layers were washed with brine (100 mL), dried over Na₂SO₄ and were concentrated by rotary evaporation. The crude was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, H₂O/MeCN + 0.1% TFA, UV). The product-containing fractions were freeze dried to yield the desired product as off-white powder (2.2 g, **70%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 8.22 (t, *J* = 5.8 Hz, 1H), 8.11 (t, *J* = 5.9 Hz, 1H), 7.99 (dd, *J* = 5.9, 3.4 Hz, 1H), 7.96 – 7.92 (m, 1H), 7.91 – 7.86 (m, 2H), 7.64 – 7.54 (m, 3H), 7.54 - 7.48 (m, 1H), 7.47 – 7.40 (m, 3H), 7.38 – 7.30 (m, 2H), 6.50 (d, *J* = 16.0 Hz, 1H), 3.14 (dt, *J* = 7.6, 6.0 Hz, 2H), 2.77 (dt, *J* = 7.6, 6.2 Hz, 2H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 169.19, 167.48, 147.70, 143.68, 140.77, 138.44, 136.94, 134.10, 132.71, 132.54, 130.14, 129.81, 129.51, 129.38, 128.63, 127.94, 127.63, 127.34, 126.99, 124.57, 119.46, 41.56, 38.78.

HRMS (ESI): C₂₄H₂₁N₃NaO₇S⁺ *calcd*: 518.0992, *found*: 518.0991.

Nos-Sc-NP01-TFL(N₃)-OMe 2-18b



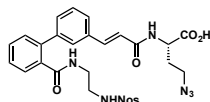
In a 500 mL flask **2-18a** (2.2 g, 4.4 mmol, 1.0 eq.), **2-5** (841.0 mg, 5.3 mmol, 1.2 eq.), HBTU (2.0 g, 5.3 mmol, 1.2 eq.) and DIPEA (2.3 mL, 13.3 mmol, 3.0 eq.) were dissolved in THF (40 mL) and the mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the crude was purified by flash column chromatography on the ISOLERA (Silica, 340 g, DCM/MeOH, UV) to yield the desired product as a thick, yellow oil (3.7 g, >100%). NMR showed co-eluted byproducts from the coupling reagents. The material was used without further purification for the next synthetic step.

¹H NMR (400 MHz, CDCl₃) δ/ppm: 8.01 – 7.93 (m, 1H), 7.80 – 7.73 (m, 1H), 7.73 – 7.65 (m, 2H), 7.58 – 7.25 (m, 9H), 7.00 (d, *J* = 7.8 Hz, 1H), 6.54 (d, *J* = 15.7 Hz, 1H), 6.50 (t, *J* = 5.8 Hz, 1H), 5.92 (t, *J* = 5.9 Hz, 1H), 4.70 (td, *J* = 7.9, 5.0 Hz, 1H), 3.70 (s, 3H), 3.37 (t, *J* = 6.7 Hz, 2H), 3.33 – 3.25 (m, 2H), 3.01 – 2.93 (m, 2H), 2.18 – 2.07 (m, 1H), 2.03 – 1.90 (m, 1H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 172.26, 170.70, 166.13, 147.93, 140.97, 140.89, 139.09, 135.47, 135.00, 133.92, 133.07, 132.98, 130.90, 130.44, 130.25, 130.11, 129.18, 128.23, 127.87, 127.75, 127.61, 125.30, 121.03, 52.68, 50.34, 47.81, 43.06, 39.85, 31.19.

HRMS (ESI): C₂₉H₂₉N₇NaO₈S⁺ *calcd*: 658.1691, *found*: 658.1693.

Nos-Sc-NP01-TFL(N₃)-OH 2-18c



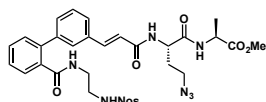
In a 100 mL flask **2-18b** (3.6 g, 5.7 mmol, 1.0 eq.) was dissolved in MeCN (34 mL) and a solution of LiOH monohydrate (1.2 g, 28.5 mmol, 5.0 eq.) in H₂O (17 mL) was added. The mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed full conversion to the desired product. Acidifying the mixture to pH 5-6 with concentrated HCl quenched the reaction and the volatiles were removed *in vacuo*. The residue was taken up in EtOAc (100 mL) and H₂O (100 mL). The biphasic mixture was acidified to pH 1 with concentrated HCl and the layers were separated. The aqueous layer was extracted with EtOAc (2 x 100 mL) and the combined organic layers were washed with half-saturated brine (2 x 100 mL) and brine (100 mL), dried over MgSO₄ and concentrated by rotary evaporation to yield the desired product as a yellowish solid (2.8 g, 79%). NMR showed some residual tetramethylurea from the HBTU coupling step. The material was used without further purification for the next synthetic step.

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 12.81 (s, 1H), 8.46 (d, *J* = 7.9 Hz, 1H), 8.18 (t, *J* = 5.8 Hz, 1H), 8.09 (t, *J* = 6.0 Hz, 1H), 8.02 – 7.97 (m, 1H), 7.96 – 7.92 (m, 1H), 7.92 – 7.87 (m, 2H), 7.55 – 7.49 (m, 2H), 7.48 – 7.40 (m, 5H), 7.37 – 7.29 (m, 2H), 6.73 (d, *J* = 15.8 Hz, 1H), 4.44 (ddd, *J* = 9.2, 7.9, 4.8 Hz, 1H), 3.51 – 3.38 (m, 2H), 3.14 (dt, *J* = 7.9, 6.1 Hz, 2H), 2.79 – 2.72 (m, 2H), 2.05 (dtd, *J* = 13.9, 7.4, 4.8 Hz, 1H), 1.95 – 1.85 (m, 1H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 173.06, 169.12, 165.00, 147.70, 140.79, 139.12, 138.57, 136.90, 134.63, 134.10, 132.73, 132.48, 129.72, 129.56, 129.52, 129.39, 128.62, 127.65, 127.60, 127.33, 126.27, 124.54, 121.81, 49.63, 47.53, 41.56, 38.75, 30.40.

HRMS (ESI): C₂₈H₂₇N₇NaO₈S⁺ *calcd*: 644.1534, *found*: 644.1537.

Nos-Sc-NP01-TFL(N₃)-AA001-OMe **2-18d**



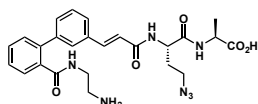
In a 50 mL flask **2-18c** (1.1 g, 1.8 mmol, 1.0 eq.) was dissolved in THF (16 mL) and H-Ala-OMe hydrochloride (298.0 mg, 2.1 mmol, 1.2 eq.) was added, followed by DIPEA (931.0 μ L, 5.3 mmol, 3.0 eq.) and HATU (810.0 mg, 2.1 mmol, 1.2 eq.). The mixture was stirred at RT for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the residue was taken up in EtOAc (100 mL) and H₂O (100 mL). The mixture was acidified with concentrated HCl to pH 1. The layers were separated and the aqueous phase was extracted with EtOAc (2 x 100 mL) and the combined organic layers were washed with diluted HCl solution (1%, 2 x 100 mL), brine (100 mL), dried over Na₂SO₄ and concentrated by rotary evaporation. The crude was purified by flash column chromatography on the ISOLERA (Silica, 120 g, DCM/MeOH, UV) to yield the desired product as a slightly yellow solid (1.0 g, **81%**).

¹H NMR (400 MHz, CD₃CN) δ /ppm: 8.02 – 7.95 (m, 1H), 7.88 – 7.83 (m, 1H), 7.83 – 7.78 (m, 2H), 7.58 – 7.51 (m, 2H), 7.51 – 7.45 (m, 3H), 7.44 – 7.38 (m, 2H), 7.37 – 7.33 (m, 2H), 7.19 (d, *J* = 7.0 Hz, 1H), 7.05 (d, *J* = 8.0 Hz, 1H), 6.70 – 6.55 (m, 2H), 6.12 (t, *J* = 5.9 Hz, 1H), 4.58 (td, *J* = 8.2, 5.4 Hz, 1H), 4.42 – 4.32 (m, 1H), 3.66 (s, 3H), 3.42 (ddd, *J* = 9.9, 5.0, 2.4 Hz, 2H), 3.22 (q, *J* = 6.2 Hz, 2H), 2.93 (q, *J* = 6.1 Hz, 2H), 2.05 (ddd, *J* = 14.3, 7.1, 5.5 Hz, 1H), 1.92 – 1.83 (m, 1H), 1.32 (d, *J* = 7.3 Hz, 3H).

¹³C NMR (101 MHz, CD₃CN) δ /ppm: 173.90, 171.89, 170.91, 166.40, 148.89, 142.01, 140.90, 140.03, 137.31, 135.94, 135.11, 133.87, 133.65, 131.40, 131.01, 130.93, 130.88, 129.79, 128.81, 128.76, 128.55, 127.87, 126.02, 122.39, 51.64, 49.14, 48.67, 43.67, 40.08, 38.82, 32.29, 17.53.

HRMS (ESI): C₃₂H₃₄N₈NaO₉S⁺ *calcd*: 729.2062, *found*: 729.2060.

NH₂-Sc-NP01-TFL(N₃)-AA001-OH **2-18e**



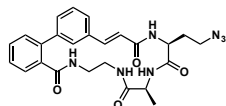
Under an inert (N₂) atmosphere **2-18d** (1.0 g, 1.4 mmol, 1.0 eq.) was dissolved in MeCN (12 mL) and DIPEA (2.5 mL, 14.2 mmol, 10.0 eq.) was added, followed by thiophenol (725.0 μ L, 7.1 mmol, 5.0 eq.). The yellow solution was stirred at RT for 2 h, after which UPLC-MS analysis showed complete Nosyl deprotection. A solution of LiOH monohydrate (468.0 mg, 11.2 mmol, 8.0 eq.) in H₂O (4 mL) was added and the mixture was stirred for 1.5 h, after which UPLC-MS analysis showed complete ester deprotection. The mixture was acidified to pH 5-6 by the addition of concentrated HCl. The volatiles were removed by rotary evaporation and the crude was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, H₂O/MeCN + 0.1% TFA, UV (254/280 nm)). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (810 mg, **>100%**). Co-elution of unknown side products. NMR and LC-MS analysis showed 18% epimerized product.

¹H NMR (400 MHz, DMSO-*d*₆) δ /ppm: 8.44 (d, *J* = 7.1 Hz, 1H), 8.38 – 8.28 (m, 2H), 7.81 (brs, 3H), 7.59 – 7.51 (m, 4H), 7.49 – 7.41 (m, 4H), 7.37 (dt, *J* = 7.6, 1.5 Hz, 1H), 6.81 (d, *J* = 15.8 Hz, 1H), 4.54 (td, *J* = 8.2, 5.2 Hz, 1H), 4.27 – 4.16 (m, 1H), 3.40 (tdd, *J* = 10.1, 8.8, 5.7 Hz, 2H), 3.28 (q, *J* = 6.5 Hz, 2H), 2.74 (q, *J* = 6.3 Hz, 2H), 2.02 – 1.90 (m, 1H), 1.83 (dtd, *J* = 13.8, 7.9, 6.1 Hz, 1H), 1.30 (d, *J* = 7.2 Hz, 3H).

^{13}C NMR (101 MHz, $\text{DMSO-}d_6$) δ /ppm: 173.94, 170.77, 169.58, 164.77, 140.82, 138.69, 136.50, 134.82, 132.39, 129.84, 129.76, 129.49, 128.77, 127.87, 127.64, 127.36, 126.33, 122.24, 49.97, 47.57, 47.45, 38.11, 36.74, 31.59, 16.89.

HRMS (ESI): $\text{C}_{25}\text{H}_{30}\text{N}_7\text{O}_5^+$ *calcd*: 508.2303, *found*: 508.2297.

MC NP01-AA001 2-18



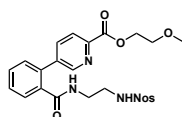
In a 1 L flask **2-18e** (764.0 mg, 1.5 mmol, 1.0 eq.) was dissolved in THF (700 mL) and DIPEA (787.0 μL , 4.5 mmol, 3.0 eq.) was added, followed by HATU (1.0 g, 2.7 mmol, 1.8 eq.). The mixture was stirred at RT for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The mixture was concentrated by rotary evaporation and purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, $\text{H}_2\text{O}:\text{MeCN} + 0.1\%$ TFA, UV 254/280 nm). Further purification was necessary due to co-eluted impurities. The impure material was purified by reversed phase preparative HPLC (Method A) followed by flash column chromatography on the ISOLERA (Silica, $\text{DCM}:\text{MeOH}$, UV). Product-containing fractions were combined and concentrated by rotary evaporation. The oily product was dissolved in $\text{MeCN}/\text{H}_2\text{O}$ and lyophilized to yield the desired product as a white solid (37.0 mg, **5%**). Mixture of diastereomers with the ratio 87:13 (UPLC-MS analysis).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ /ppm: 8.47 (t, $J = 5.4$ Hz, 1H), 8.37 (d, $J = 7.6$ Hz, 1H), 7.74 – 7.66 (m, 2H), 7.61 (dt, $J = 7.4, 1.7$ Hz, 1H), 7.54 – 7.52 (m, 2H), 7.51 – 7.48 (m, 2H), 7.48 – 7.45 (m, 2H), 7.44 – 7.41 (m, 2H), 6.55 (d, $J = 15.8$ Hz, 1H), 4.22 (ddd, $J = 13.8, 8.9, 6.7$ Hz, 2H), 3.51 – 3.39 (m, 2H), 3.31 – 3.25 (m, 1H), 3.12 – 3.07 (m, 2H), 2.90 – 2.82 (m, 1H), 1.93 – 1.86 (m, 2H), 1.18 (d, $J = 7.0$ Hz, 3H).

^{13}C NMR (101 MHz, $\text{DMSO-}d_6$) δ /ppm: 172.30, 170.85, 169.91, 166.04, 140.50, 140.02, 138.27, 137.81, 135.00, 129.49, 129.44, 129.30, 129.14, 129.05, 127.40, 127.12, 124.49, 123.14, 53.33, 48.61, 47.59, 39.60, 38.62, 29.99, 17.61.

HRMS (ESI): $\text{C}_{25}\text{H}_{27}\text{N}_7\text{NaO}_4^+$ *calcd*: 512.2017, *found*: 512.2011.

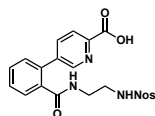
Nos-Sc-NP07-OEtOMe 2-19a*



In a 100 mL flask **2-24** (1.5 g, 3.2 mmol, 1.0 eq.), 2-(methylcarboxy)pyridine-5-boronic acid pinacol ester (1.0 g, 3.8 mmol, 1.2 eq.), $\text{Pd}(\text{dppf})\text{Cl}_2$ (348.0 mg, 475.0 μmol , 15 mol%) and triethylamine (883.0 μL , 6.3 mmol, 2.0 eq.) were dissolved in methoxyethanol (40 mL) and the mixture was stirred at 100°C for 4 h, after which UPLC-MS analysis showed full conversion to the desired product as methoxyethylester. The solvent was removed *in vacuo* and the crude was taken up in EtOAc (100 mL). The organic layer was washed with diluted HCl (pH 1, 3 x 100 mL). The combined aqueous layers were extracted with EtOAc (3 x 100 mL) and the combined organic layers were washed with brine (100 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The crude was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, $\text{H}_2\text{O}/\text{MeCN} + 0.1\%$ TFA, UV). The product-containing fractions were lyophilized to yield the desired product as dark brown waxy solid (884 mg, **53%**). NMR analysis showed co-elution of some side products. The product was used without further purification for

the next steps. Characterization of the compound was performed after ester saponification to compound **2-19a**.

Nos-Sc-NP07-OH **2-19a**



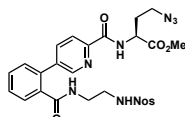
In a 100 mL flask **2-19a*** (842 mg, 1.6 mmol, 1.0 eq.) was dissolved in MeCN (10 mL) and a solution of LiOH monohydrate (334 mg, 8.0 mmol, 5.0 eq.) in H₂O (5 mL) was added. The mixture was stirred at RT for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. Acidifying with concentrated HCl to pH 2 quenched the reaction. The volatiles were removed *in vacuo* and the aqueous residue was taken up in EtOAc (50 mL) and H₂O (50 mL). The mixture was acidified to pH 1 with concentrated HCl. The layers were separated (addition of brine for good separation) and the aqueous layer was extracted with EtOAc (2 x 50 mL). The combined organic layers were washed with half-saturated brine (2 x 50 mL) and brine (50 mL), dried over Na₂SO₄ and were concentrated by rotary evaporation. The crude was purified by reversed phase flash column chromatography (RP-Silica, 340 g, 0.1% TFA in H₂O/MeCN). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (366 mg, **49%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 8.65 (dd, *J* = 2.3, 0.8 Hz, 1H), 8.46 (t, *J* = 5.7 Hz, 1H), 8.15 (t, *J* = 6.0 Hz, 1H), 8.03 (dd, *J* = 8.1, 0.8 Hz, 1H), 8.02 – 7.94 (m, 2H), 7.92 (dd, *J* = 8.0, 2.3 Hz, 1H), 7.90 – 7.83 (m, 2H), 7.65 – 7.45 (m, 4H), 3.23 – 3.13 (m, 2H), 2.96 – 2.85 (m, 2H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 168.59, 165.94, 148.46, 147.68, 146.72, 139.05, 136.93, 136.76, 135.18, 134.13, 132.76, 132.58, 130.18, 130.06, 129.46, 128.41, 128.12, 124.55, 124.18, 41.77, 39.02.

HRMS (ESI): C₂₁H₁₉N₄O₇S⁺ *calcd*: 471.0969, *found*: 471.0966.

Nos-Sc-NP07-TFL(N₃)-OMe **2-19b**



In a 50 mL flask **2-19a** (315.0 mg, 670.0 μmol, 1.0 eq.), **2-5** (127.0 mg, 803.0 μmol, 1.2 eq.), HBTU (305.0 g, 803.0 μmol, 1.2 eq.) and DIPEA (350.0 μL, 2.0 mmol, 3.0 eq.) were dissolved in THF (5 mL) and DMF (1 mL) and the mixture was stirred at RT for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the residue was dissolved in EtOAc (50 mL) and H₂O (50 mL). The mixture was acidified to pH 1 by addition of concentrated HCl. The layers were separated and the aqueous phase was extracted with EtOAc (2 x 50 mL). The combined organic layers were washed with H₂O (50 mL), brine (50 mL), dried over MgSO₄ and concentrated by rotary evaporation. The crude material was purified by flash column chromatography on the ISOLERA (Silica, 120 g, DCM/MeOH, UV) to yield the desired product as a yellowish solid (397 mg, **97%**).

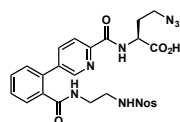
¹H NMR (400 MHz, CDCl₃) δ/ppm: 8.67 (d, *J* = 8.3 Hz, 1H), 8.65 (dd, *J* = 2.3, 0.9 Hz, 1H), 8.19 (dd, *J* = 8.0, 0.8 Hz, 1H), 8.09 – 8.04 (m, 1H), 7.93 (dd, *J* = 8.0, 2.2 Hz, 1H), 7.87 – 7.79 (m, 1H), 7.80 – 7.68 (m, 2H), 7.64 (dd, *J* = 7.6, 1.2 Hz, 1H), 7.54 (td, *J* = 7.6, 1.5 Hz, 1H), 7.47 (td, *J* = 7.5, 1.4 Hz, 1H), 7.37 (dd, *J* = 7.6, 0.9 Hz, 1H), 6.51 (t, *J* = 5.8 Hz, 1H), 5.87 (t, *J* = 6.1 Hz, 1H),

4.89 (td, $J = 7.9, 5.0$ Hz, 1H), 3.79 (s, 3H), 3.50 – 3.40 (m, 4H), 3.18 – 3.11 (m, 2H), 2.28 (dtd, $J = 14.1, 6.9, 5.0$ Hz, 1H), 2.12 (ddt, $J = 14.3, 7.9, 6.5$ Hz, 1H).

^{13}C NMR (101 MHz, CDCl_3) δ /ppm: 171.62, 169.34, 161.89, 148.13, 145.35, 140.99, 140.62, 135.96, 134.43, 133.87, 133.12, 133.02, 131.11, 130.98, 130.59, 129.64, 128.70, 125.32, 124.15, 52.85, 50.86, 47.89, 42.95, 39.92, 31.24.

HRMS (ESI): $\text{C}_{26}\text{H}_{26}\text{N}_8\text{NaO}_8\text{S}^+$ *calcd*: 633.1487, *found*: 633.1495.

Nos-Sc-NP07-TFL(N_3)-OH 2-19c



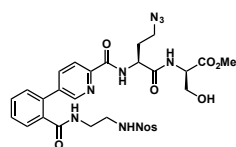
In a 50 mL flask **2-19b** (367 mg, 601.0 μmol , 1.0 eq.) was dissolved in MeCN (5 mL) and a solution of LiOH monohydrate (126 mg, 3.0 mmol, 5.0 eq.) in H_2O (2.5 mL) was added. The mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed full conversion to the desired product. Acidifying the mixture to pH 5-6 with concentrated HCl quenched the reaction and the volatiles were removed *in vacuo*. The residue was taken up in EtOAc (50 mL) and H_2O (50 mL). The biphasic mixture was acidified to pH 1 with concentrated HCl and the layers were separated. The aqueous layer was extracted with EtOAc (2 x 50 mL) and the combined organic layers were washed with half-saturated brine (2 x 50 mL) and brine (50 mL), dried over MgSO_4 and concentrated by rotary evaporation to yield the desired product as a thick, waxy oil (400 mg, >100%). NMR showed some residual tetramethylurea from the HBTU coupling step. The material was used without further purification for the next synthetic step.

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 12.92 (s, 1H), 8.98 (d, $J = 8.3$ Hz, 1H), 8.62 (dd, $J = 2.2, 0.9$ Hz, 1H), 8.47 (t, $J = 5.7$ Hz, 1H), 8.15 (t, $J = 5.9$ Hz, 1H), 8.02 (dd, $J = 8.1, 0.8$ Hz, 1H), 8.00 – 7.96 (m, 2H), 7.93 (dd, $J = 8.1, 2.2$ Hz, 1H), 7.91 – 7.81 (m, 2H), 7.65 – 7.46 (m, 4H), 4.58 (dt, $J = 8.2, 6.9$ Hz, 1H), 3.48 (dt, $J = 12.4, 6.2$ Hz, 1H), 3.39 (dt, $J = 12.5, 7.2$ Hz, 1H), 3.19 (q, $J = 6.4$ Hz, 2H), 2.94 (q, $J = 6.5$ Hz, 2H), 2.15 (q, $J = 6.8$ Hz, 2H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ /ppm: 172.76, 168.56, 163.86, 147.94, 147.66, 147.57, 138.80, 137.16, 136.72, 135.31, 134.10, 132.71, 132.57, 130.18, 130.01, 129.42, 128.29, 128.10, 124.49, 121.46, 49.77, 47.76, 41.81, 38.98, 30.03.

HRMS (ESI): $\text{C}_{25}\text{H}_{24}\text{N}_8\text{NaO}_8\text{S}^+$ *calcd*: 619.1330, *found*: 619.1332.

Nos-Sc-NP07-TFL(N_3)-AA026-OMe 2-19d



In a 50 mL flask **2-19c** (345.0 mg, 578.0 μmol , 1.0 eq.) was dissolved in THF (5 mL) and H-D-Ser-OMe hydrochloride (108.0 mg, 694.0 μmol , 1.2 eq.) was added, followed by DIPEA (302.0 μL , 1.7 mmol, 3.0 eq.) and HATU (263.0 mg, 692.0 μmol , 1.2 eq.). The mixture was stirred at RT for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the residue was taken up in EtOAc (50 mL) and H_2O (50 mL). The mixture was acidified with concentrated HCl to pH 1. The layers were separated and the aqueous phase was extracted with EtOAc (2 x 100 mL) and the combined organic layers were washed with diluted HCl solution (1%, 2 x 50 mL), brine (50 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The crude was purified by flash column chromatography on

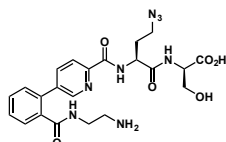
the ISOLERA (Silica, 120 g, DCM/MeOH, UV) to yield the desired product as a waxy solid (403.0 mg, **94%**).

^1H NMR (500 MHz, CD_3CN) δ /ppm: 8.62 (dd, $J = 2.3, 0.8$ Hz, 1H), 8.57 (d, $J = 8.2$ Hz, 1H), 8.08 (dd, $J = 8.0, 0.8$ Hz, 1H), 8.05 – 8.00 (m, 1H), 7.92 (dd, $J = 8.1, 2.2$ Hz, 1H), 7.89 – 7.84 (m, 1H), 7.81 (dd, $J = 5.9, 3.4$ Hz, 2H), 7.61 – 7.55 (m, 2H), 7.53 – 7.49 (m, 1H), 7.48 (ddd, $J = 7.5, 1.4, 0.6$ Hz, 1H), 7.25 (d, $J = 7.8$ Hz, 1H), 6.89 (t, $J = 6.0$ Hz, 1H), 6.12 (s, 1H), 4.74 (td, $J = 8.3, 5.1$ Hz, 1H), 4.48 (dt, $J = 8.1, 4.1$ Hz, 1H), 3.89 – 3.82 (m, 1H), 3.76 (ddd, $J = 11.2, 5.5, 3.8$ Hz, 1H), 3.66 (s, 3H), 3.46 (ddd, $J = 7.5, 6.2, 3.1$ Hz, 2H), 3.29 (q, $J = 6.0$ Hz, 2H), 3.22 (t, $J = 6.0$ Hz, 1H), 3.08 (q, $J = 5.7$ Hz, 2H), 2.19 (ddd, $J = 14.5, 7.3, 2.2$ Hz, 1H), 2.05 (ddt, $J = 14.5, 8.4, 6.3$ Hz, 1H).

^{13}C NMR (126 MHz, CD_3CN) δ /ppm: 171.88, 171.56, 170.37, 164.98, 148.92, 140.13, 138.27, 137.39, 136.61, 135.13, 133.82, 133.67, 131.38, 131.35, 131.29, 129.55, 129.52, 128.99, 125.99, 122.42, 62.49, 55.69, 52.79, 51.85, 48.86, 43.76, 40.06, 32.30.

HRMS (ESI): $\text{C}_{29}\text{H}_{31}\text{N}_9\text{NaO}_{10}\text{S}^+$ *calcd*: 720.1807, *found*: 720.1813.

NH_2 -Sc-NP07-TFL(N_3)-AA026-OH **2-19e**



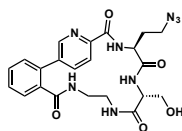
Under an inert (N_2) atmosphere **2-19d** (374.0 mg, 536.0 μmol , 1.0 eq.) was dissolved in MeCN (5 mL) and DIPEA (934.0 μL , 5.4 mmol, 10.0 eq.) was added, followed by thiophenol (273.0 μL , 2.7 mmol, 5.0 eq.). The yellow solution was stirred at RT for 2 h, after which UPLC-MS analysis showed complete Nosyl deprotection. A solution of LiOH monohydrate (170.0 mg, 4.1 mmol, 8.0 eq.) in H_2O (1.5 mL) was added and the mixture was stirred for 1.5 h, after which UPLC-MS analysis showed complete ester deprotection. The mixture was acidified to pH 5-6 by the addition of concentrated HCl. The volatiles were removed by rotary evaporation and the crude was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 120 g, $\text{H}_2\text{O}/\text{MeCN} + 0.1\%$ TFA, UV (254/280 nm)). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (292 mg, **>100%**). Co-elution of unknown side products. NMR and LC-MS analysis showed 19% epimerized product.

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 8.85 (d, $J = 8.6$ Hz, 1H), 8.65 (dd, $J = 2.3, 0.9$ Hz, 1H), 8.56 (t, $J = 5.6$ Hz, 1H), 8.47 (d, $J = 7.9$ Hz, 1H), 8.08 (dd, $J = 8.1, 0.8$ Hz, 1H), 7.97 (dd, $J = 8.1, 2.2$ Hz, 1H), 7.80 (brs, 3H), 7.67 (dd, $J = 7.5, 1.5$ Hz, 1H), 7.62 (td, $J = 7.5, 1.5$ Hz, 1H), 7.60 – 7.49 (m, 2H), 4.77 (td, $J = 8.4, 5.1$ Hz, 1H), 4.31 (ddd, $J = 7.8, 5.3, 4.1$ Hz, 1H), 3.72 (dd, $J = 10.8, 5.4$ Hz, 1H), 3.66 (dd, $J = 10.9, 4.2$ Hz, 1H), 3.49 – 3.26 (m, 4H), 2.92 – 2.76 (m, 2H), 2.06 (dddd, $J = 19.8, 13.7, 9.2, 5.5$ Hz, 2H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ /ppm: 171.61, 170.66, 168.93, 163.24, 147.86, 147.68, 138.89, 137.28, 136.28, 135.45, 132.35, 130.31, 130.26, 128.24, 121.48, 61.17, 54.78, 50.37, 47.60, 38.22, 36.88, 31.99.

HRMS (ESI): $\text{C}_{22}\text{H}_{27}\text{N}_8\text{O}_6^+$ *calcd*: 499.2048, *found*: 499.2041.

MC-NP07-AA026 2-19f



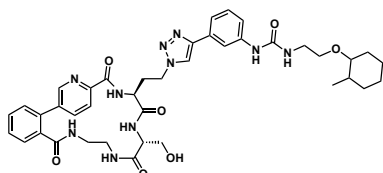
In a 500 mL flask **2-19e** (261.0 mg, 524.0 μmol , 1.0 eq.) was dissolved in THF (260 mL) and DIPEA (319.2 μL , 1.8 mmol, 3.5 eq.) was added, followed by HATU (398 mg, 1.1 mmol, 2.0 eq.). The mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed complete conversion to the desired product. The mixture was concentrated by rotary evaporation and purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, $\text{H}_2\text{O}:\text{MeCN} +0.1\%$ TFA, UV 254/280 nm) to yield the desired product as a white solid (79.0 mg, **31%**). 18 mg of the product were purified again by preparative RP-HPLC (Method A) for final analytics and protein binding assays. Mixture of diastereomers with the ratio 43:57 (UPLC-MS analysis).

^1H NMR (500 MHz, $\text{DMSO}-d_6$) δ /ppm: 9.30 (d, $J = 9.6$ Hz, 1H), 8.56 (d, $J = 2.6$ Hz, 1H), 8.17 – 8.14 (m, 1H), 8.09 (dd, $J = 8.0, 2.2$ Hz, 1H), 7.67 (t, $J = 4.9$ Hz, 1H), 7.66 – 7.63 (m, 1H), 7.62 – 7.59 (m, 1H), 7.53 – 7.51 (m, 2H), 7.24 (d, $J = 8.9$ Hz, 1H), 6.92 (t, $J = 5.3$ Hz, 1H), 4.62 (td, $J = 9.6, 5.1$ Hz, 1H), 4.19 – 4.15 (m, 1H), 3.58 – 3.54 (m, 1H), 3.53 – 3.46 (m, 3H), 3.21 – 3.13 (m, 2H), 2.91 – 2.81 (m, 2H), 2.23 – 2.15 (m, 1H), 2.10 – 2.02 (m, 1H).

^{13}C NMR (500 MHz, 2D NMR, $\text{DMSO}-d_6$) δ /ppm: 170.6, 17.40, 169.24, 165.10, 147.99, 147.43, 137.54, 137.24, 134.38, 134.20, 129.64, 128.90, 128.30, 127.87, 121.79, 60.80, 54.84, 50.42, 47.73, 40.32, 39.07, 28.51.

HRMS (ESI): $\text{C}_{22}\text{H}_{24}\text{N}_8\text{NaO}_5^+$ *calcd*: 503.1762, *found*: 503.1764.

MC NP07-AA026-TA256 2-19



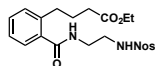
In a 2.0 mL Eppendorf tube **2-19f** (45.0 mg, 93.7 μmol , 1.0 eq.) and **TA256** (28.1 mg, 93.7 μmol , 1.0 eq.) were dissolved in DMSO (990 μL) and a solution of sodium ascorbate (11.1 mg, 56.2 μmol , 0.6 eq.) in H_2O (187 μL) was added. The solution was degassed by bubbling N_2 through the solution for 1 min. A solution of copper(II) sulfate pentahydrate (7.0 mg, 28.1 μmol , 0.3 eq.) in H_2O (140 μL) was added and the mixture was degassed again for 1 min. The reaction was agitated at RT for 2 h, after which UPLC-MS analysis showed full conversion to the desired product. The solution was directly purified by preparative RP-HPLC (Method A) and the product-containing fractions were lyophilized to yield the desired product as a white solid (60 mg, **82%**).

^1H NMR (500 MHz, $\text{DMSO}-d_6$) δ /ppm: 9.40 (d, $J = 9.5$ Hz, 1H), 8.73 (s, 1H), 8.57 (d, $J = 8.5$ Hz, 1H), 8.47 (t, $J = 5.5$ Hz, 1H), 8.17 (d, $J = 8.0$ Hz, 1H), 8.09 (dd, $J = 8.0, 2.2$ Hz, 1H), 7.94 – 7.88 (m, 1H), 7.70 – 7.65 (m, 1H), 7.64 (s, 1H), 7.63 – 7.54 (m, 1H), 7.53 (d, $J = 4.2$ Hz, 1H), 7.51 – 7.45 (m, 1H), 7.39 – 7.31 (m, 2H), 7.31 – 7.28 (m, 1H), 7.28 – 7.21 (m, 1H), 6.92 (t, $J = 5.4$ Hz, 1H), 6.13 (t, $J = 6.0$ Hz, 1H), 4.58 – 4.48 (m, 2H), 4.21 – 4.10 (m, 1H), 3.59 – 3.53 (m, 2H), 3.46 (dd, $J = 10.8, 4.9$ Hz, 1H), 3.42 – 3.29 (m, 2H), 3.28 – 3.05 (m, 4H), 2.91 – 2.81 (m, 1H), 2.81 – 2.73 (m, 1H), 2.66 – 2.54 (m, 1H), 2.49 – 2.37 (m, 1H), 2.34 – 2.15 (m, 1H), 2.01 (dt, $J = 12.6, 4.2$ Hz, 1H), 1.73 – 1.67 (m, 1H), 1.67 – 1.61 (m, 1H), 1.58 – 1.49 (m, 1H), 1.40 – 1.29 (m, 1H), 1.24 – 1.11 (m, 2H), 1.11 – 1.02 (m, 1H), 0.96 (d, $J = 6.5$ Hz, 3H).

^{13}C NMR (500 MHz, 2D NMR, $\text{DMSO-}d_6$) δ/ppm : 172.64, 170.34, 169.19, 165.29, 154.92, 150.59, 148.03, 147.49, 146.19, 140.86, 137.55, 137.27, 136.26, 134.28, 131.03, 129.65, 129.03, 128.91, 128.32, 121.82, 121.43, 117.91, 116.83, 113.99, 82.89, 67.07, 60.77, 54.87, 50.39, 46.56, 40.37, 39.40, 39.09, 37.57, 33.05, 30.66, 29.84, 24.86, 24.22, 18.49.

HRMS (ESI): $\text{C}_{40}\text{H}_{48}\text{N}_{10}\text{NaO}_7^+$ *calcd*: 803.3600, *found*: 803.3585.

Nos-Sc-NP14-OEt 2-20a*



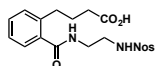
Under an inert atmosphere **2-24** (3.0 g, 6.3 mmol, 1.0 eq.) and Pd (π -cinnamyl) chloride dimer (327.0 mg, 631.0 μmol , 10 mol%) were dissolved in THF (60 mL) and the solution was cooled to 0°C. 4-Ethoxy-4-oxobutylzinc bromide solution (0.5 M in THF, 18.9 mL, 9.5 mmol, 1.5 eq.) was slowly added and the solution was stirred at 0°C for 15 min. The cooling bath was removed and the mixture was stirred at RT for 45 min, after which UPLC-MS analysis showed full consumption of the starting material. The reaction was stopped by the addition of H_2O (5 mL). The mixture was filtered and the volatiles were removed *in vacuo*. The residue was dissolved in EtOAc (150 mL) and washed with half-saturated NH_4Cl (100 mL) and H_2O (2 x 100 mL). The combined aqueous layers were extracted with EtOAc (2 x 100 mL) and the combined organic layers were treated with brine (100 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The crude was purified by flash column chromatography on the ISOLERA (Silica, 340g, cyclohexane:EtOAc, UV) to yield the desired product as a thick, slightly yellow oil (2.4 g, **82%**).

^1H NMR (400 MHz, CD_3CN) δ/ppm : 8.09 – 8.03 (m, 1H), 7.88 – 7.83 (m, 1H), 7.82 – 7.76 (m, 2H), 7.35 (td, $J = 7.5, 1.5$ Hz, 1H), 7.30 (dd, $J = 7.6, 1.4$ Hz, 1H), 7.25 (ddd, $J = 7.6, 1.3, 0.6$ Hz, 1H), 7.21 (td, $J = 7.4, 1.4$ Hz, 1H), 6.91 (s, 1H), 6.31 (t, $J = 5.7$ Hz, 1H), 4.04 (q, $J = 7.1$ Hz, 2H), 3.44 (q, $J = 5.9$ Hz, 2H), 3.27 - 3.21 (m, 2H), 2.76 – 2.68 (m, 2H), 2.24 (t, $J = 7.4$ Hz, 2H), 1.88 – 1.78 (m, 2H), 1.19 (t, $J = 7.1$ Hz, 3H).

^{13}C NMR (101 MHz, CD_3CN) δ/ppm : 174.10, 171.32, 140.76, 137.42, 135.06, 133.82, 131.40, 131.01, 130.71, 129.37, 127.99, 127.94, 126.87, 125.95, 60.89, 44.21, 40.16, 34.38, 32.98, 27.52, 14.54.

HRMS (ESI): $\text{C}_{21}\text{H}_{25}\text{N}_3\text{NaO}_7\text{S}^+$ *calcd*: 486.1305, *found*: 486.1309.

Nos-Sc-NP14-OH 2-20a



In a 100 mL flask **2-20a*** (2.3 g, 4.9 mmol, 1.0 eq.) was dissolved in MeCN (30 mL) and a solution of LiOH monohydrate (1.0 g, 24.6 mmol, 5.0 eq.) in H_2O (15 mL) was added. The mixture was stirred at RT for 2.5 h, after which UPLC-MS analysis showed complete conversion to the desired product. Acidifying with concentrated HCl to pH 5-6 quenched the reaction. The volatiles were removed and the aqueous residue was taken up in EtOAc (150 mL) and H_2O (100 mL). The mixture was acidified to pH 1 with concentrated HCl. The layers were separated (addition of brine for good separation) and the aqueous layer was extracted with EtOAc (2 x 100 mL). The combined organic layers were washed with half-saturated brine (2 x 100 mL) and brine (100 mL), dried over Na_2SO_4 and concentrated by rotary evaporation to yield the desired product as yellow, waxy solid (2.1 g, **quant. 73% HPLC purity**).

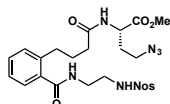
^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ/ppm : 12.00 (s, 1H), 8.30 (t, $J = 5.7$ Hz, 1H), 8.17 (t, $J = 5.9$ Hz, 1H), 8.05 – 8.01 (m, 1H), 8.00 – 7.97 (m, 1H), 7.89 – 7.85 (m, 2H), 7.38 – 7.29 (m, 2H), 7.28 –

7.17 (m, 2H), 3.34 – 3.28 (m, 2H), 3.11 – 3.05 (m, 2H), 2.74 – 2.58 (m, 2H), 2.17 (t, $J = 7.5$ Hz, 2H), 1.82 – 1.65 (m, 2H).

^{13}C NMR (101 MHz, DMSO- d_6) δ /ppm: 174.24, 169.35, 147.71, 139.37, 136.80, 134.07, 132.72, 132.68, 129.48, 129.35, 128.22, 127.22, 125.66, 124.49, 42.09, 38.97, 33.35, 31.85, 26.32.

HRMS (ESI): $\text{C}_{19}\text{H}_{22}\text{N}_3\text{O}_7\text{S}^+$ *calcd*: 436.1173, *found*: 436.1166.

Nos-Sc-NP14-TFL(N_3)-OMe **2-20b**



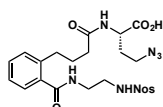
In a 100 mL flask **2-20a** (2.1 g, 4.9 mmol, 1.0 eq.), **2-5** (925.0 mg, 5.9 mmol, 1.2 eq.), HBTU (2.2 g, 5.9 mmol, 1.2 eq.) and DIPEA (2.5 mL, 14.6 mmol, 3.0 eq.) were dissolved in THF (42 mL) and the mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the residue was dissolved in EtOAc (100 mL) and H_2O (100 mL). The mixture was acidified to pH 1 by addition of concentrated HCl. The layers were separated and the aqueous phase was extracted with EtOAc (2 x 100 mL). The combined organic layers were washed with H_2O (100 mL), brine (100 mL), dried over MgSO_4 and concentrated by rotary evaporation. The crude material was purified by flash column chromatography on the ISOLERA (Silica, 340 g, DCM/MeOH, UV) to yield the desired product as a yellow oil (2.7 g, **96%**).

^1H NMR (400 MHz, CDCl_3) δ /ppm: 8.13 – 8.07 (m, 1H), 7.81 – 7.76 (m, 1H), 7.72 – 7.69 (m, 2H), 7.33 – 7.27 (m, 2H), 7.21 – 7.13 (m, 2H), 6.84 (t, $J = 5.9$ Hz, 1H), 6.74 (d, $J = 7.8$ Hz, 1H), 6.60 (t, $J = 5.8$ Hz, 1H), 4.59 (td, $J = 7.7, 5.1$ Hz, 1H), 3.71 (s, 3H), 3.62 – 3.55 (m, 2H), 3.37 – 3.30 (m, 4H), 2.83 – 2.75 (m, 2H), 2.29 (t, $J = 7.0$ Hz, 2H), 2.08 (dtd, $J = 14.0, 6.9, 5.1$ Hz, 1H), 1.98 – 1.89 (m, 3H).

^{13}C NMR (101 MHz, CDCl_3) δ /ppm: 173.52, 172.40, 171.16, 148.11, 139.75, 135.81, 133.46, 132.81, 130.98, 130.38, 130.28, 127.30, 127.17, 126.24, 125.21, 52.67, 50.15, 47.90, 43.44, 39.80, 35.67, 32.65, 31.25, 27.06.

HRMS (ESI): $\text{C}_{24}\text{H}_{29}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 598.1691, *found*: 598.1701.

Nos-Sc-NP14-TFL(N_3)-OH **2-20c**



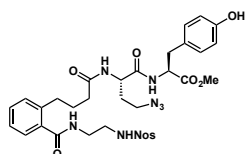
In a 100 mL flask **2-20b** (2.6 g, 4.5 mmol, 1.0 eq.) was dissolved in MeCN (25 mL) and a solution of LiOH monohydrate (943.0 mg, 22.5 mmol, 5.0 eq.) in H_2O (12.5 mL) was added. The mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed full conversion to the desired product. Acidifying the mixture to pH 5-6 with concentrated HCl quenched the reaction and the volatiles were removed *in vacuo*. The residue was taken up in EtOAc (100 mL) and H_2O (100 mL). The biphasic mixture was acidified to pH 1 with concentrated HCl and the layers were separated. The aqueous layer was extracted with EtOAc (2 x 100 mL) and the combined organic layers were washed with half-saturated brine (2 x 100 mL) and brine (100 mL), dried over MgSO_4 and concentrated by rotary evaporation to yield the desired product as a yellowish solid (2.5 g, **98%**). NMR showed some residual tetramethylurea from the HBTU coupling step. The material was used without further purification for the next synthetic step.

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ /ppm: 12.65 (s, 1H), 8.29 (t, $J = 5.7$ Hz, 1H), 8.19 (t, $J = 5.8$ Hz, 1H), 8.11 (d, $J = 7.9$ Hz, 1H), 8.06 – 8.01 (m, 1H), 8.01 – 7.96 (m, 1H), 7.90 – 7.84 (m, 2H), 7.38 – 7.28 (m, 2H), 7.28 – 7.17 (m, 2H), 4.27 (ddd, $J = 9.6, 7.9, 4.6$ Hz, 1H), 3.46 – 3.34 (m, 2H), 3.34 – 3.27 (m, 2H), 3.12 – 3.05 (m, 2H), 2.65 (dd, $J = 8.8, 6.7$ Hz, 2H), 2.11 (t, $J = 7.5$ Hz, 2H), 1.98 – 1.88 (m, 1H), 1.88 – 1.79 (m, 1H), 1.79 – 1.70 (m, 2H).

^{13}C NMR (101 MHz, $\text{DMSO-}d_6$) δ /ppm: 173.21, 172.21, 169.35, 147.70, 139.64, 136.70, 134.05, 132.71, 132.70, 129.72, 129.46, 129.33, 127.23, 125.59, 124.48, 49.27, 47.57, 42.07, 39.00, 34.89, 32.06, 30.14, 27.12.

HRMS (ESI): $\text{C}_{23}\text{H}_{27}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 584.1534, *found*: 584.1546.

Nos-Sc-NP14-TFL(N_3)-AA020-OMe **2-20d**



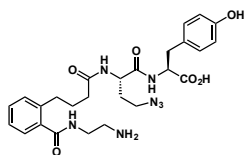
In a 50 mL flask **2-20c** (1.3 g, 2.3 mmol, 1.0 eq.) was dissolved in THF (20 mL) and H-Tyr-OMe (542.0 mg, 2.8 mmol, 1.2 eq.) was added, followed by DIPEA (1.2 mL, 6.9 mmol, 3.0 eq.) and HATU (1.1 g, 2.8 mmol, 1.2 eq.). The mixture was stirred at RT for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the residue was taken up in EtOAc (100 mL) and H_2O (100 mL). The mixture was acidified with concentrated HCl to pH 1. The layers were separated and the aqueous phase was extracted with EtOAc (2 x 100 mL) and the combined organic layers were washed with diluted HCl solution (1%, 2 x 100 mL), brine (100 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The crude was purified by flash column chromatography on the ISOLERA (Silica, 120 g, DCM/MeOH, UV) to yield the desired product as a slightly yellow solid (1.1 g, **66%**).

^1H NMR (400 MHz, CD_3CN) δ /ppm: 8.11 – 8.02 (m, 1H), 7.88 – 7.81 (m, 1H), 7.81 – 7.73 (m, 2H), 7.39 – 7.31 (m, 2H), 7.28 – 7.20 (m, 2H), 7.04 (t, $J = 6.0$ Hz, 1H), 7.00 – 6.89 (m, 3H), 6.76 (d, $J = 8.0$ Hz, 1H), 6.72 – 6.64 (m, 2H), 6.51 (t, $J = 5.8$ Hz, 1H), 4.56 (td, $J = 7.7, 5.3$ Hz, 1H), 4.36 (td, $J = 8.4, 5.5$ Hz, 1H), 3.64 (s, 3H), 3.53 – 3.41 (m, 2H), 3.38 – 3.27 (m, 2H), 3.27 – 3.21 (m, 2H), 3.00 (dd, $J = 14.0, 5.3$ Hz, 1H), 2.86 (dd, $J = 14.0, 7.7$ Hz, 1H), 2.78 – 2.65 (m, 2H), 2.17 (dt, $J = 7.2, 3.5$ Hz, 2H), 2.01 – 1.91 (m, 1H), 1.86 – 1.69 (m, 3H).

^{13}C NMR (101 MHz, CD_3CN) δ /ppm: 173.98, 172.67, 171.86, 171.76, 156.84, 148.96, 140.91, 137.25, 137.16, 135.04, 133.83, 131.34, 131.32, 131.11, 130.87, 128.42, 128.08, 126.89, 125.91, 116.07, 54.61, 52.74, 51.42, 48.78, 43.97, 40.36, 37.07, 36.15, 33.30, 31.38, 28.05.

HRMS (ESI): $\text{C}_{33}\text{H}_{38}\text{N}_8\text{NaO}_{10}\text{S}^+$ *calcd*: 761.2324, *found*: 761.2320.

NH_2 -Sc-NP14-TFL(N_3)-AA020-OH **2-20e**



Under an inert (N_2) atmosphere **2-20d** (1.1 g, 1.5 mmol, 1.0 eq.) was dissolved in MeCN (15 mL) and DIPEA (2.6 mL, 14.8 mmol, 10.0 eq.) was added, followed by thiophenol (757.0 μL , 7.4 mmol, 5.0 eq.). The yellow solution was stirred at RT for 2 h, after which UPLC-MS analysis showed complete Nosyl deprotection. A solution of LiOH monohydrate (489.0 mg, 11.7 mmol, 8.0 eq.) in H_2O (5 mL) was added and the mixture was stirred for 1.5 h, after which UPLC-MS

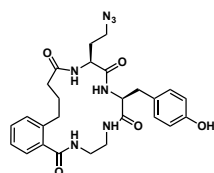
analysis showed complete ester deprotection. The mixture was acidified to pH 5-6 by the addition of concentrated HCl. The volatiles were removed by rotary evaporation and the crude was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, H₂O/MeCN + 0.1% TFA, UV (200/215 nm)). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (1.0 g, >100%). Co-elution of unknown side products. NMR and LC-MS analysis showed 16% epimerized product.

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 8.39 (t, *J* = 5.6 Hz, 1H), 8.07 (d, *J* = 7.7 Hz, 1H), 7.99 (d, *J* = 8.2 Hz, 1H), 7.90 (brs, 3H), 7.43 (dd, *J* = 7.9, 1.5 Hz, 1H), 7.37 (td, *J* = 7.5, 1.5 Hz, 1H), 7.25 (dtd, *J* = 7.5, 3.8, 1.3 Hz, 2H), 7.05 – 6.93 (m, 2H), 6.69 – 6.57 (m, 2H), 4.34 (dtd, *J* = 16.0, 8.4, 5.2 Hz, 2H), 3.46 (q, *J* = 6.2 Hz, 2H), 3.37 – 3.18 (m, 2H), 2.98 (q, *J* = 6.0 Hz, 2H), 2.92 (dd, *J* = 14.0, 5.1 Hz, 1H), 2.79 (dd, *J* = 13.9, 8.5 Hz, 1H), 2.68 (t, *J* = 7.8 Hz, 2H), 2.20 – 2.02 (m, 2H), 1.88 (dtd, *J* = 13.1, 7.6, 4.7 Hz, 1H), 1.81 – 1.66 (m, 3H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 172.83, 172.03, 171.09, 169.77, 155.97, 139.86, 136.22, 130.04, 129.87, 129.61, 127.46, 127.33, 125.62, 114.98, 53.80, 49.79, 47.55, 38.56, 36.94, 35.75, 34.96, 32.20, 31.14, 27.12.

HRMS (ESI): C₂₆H₃₄N₇O₆⁺ *calcd*: 540.2565, *found*: 540.2560.

MC NP14-AA020 2-20f



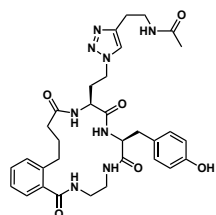
In a 1 L flask **2-20e** (967.0 mg, 1.8 mmol, 1.0 eq.) was dissolved in THF (800 mL) and DIPEA (1.1 mL, 6.3 mmol, 3.5 eq.) was added, followed by HATU (1.4 g, 3.6 mmol, 2.0 eq.). The mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed complete conversion to the desired product. The mixture was concentrated by rotary evaporation and purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, H₂O:MeCN +0.1% TFA, UV 200/215 nm) to yield the desired product as a white solid (655.0 mg, **70%**). 38 mg of the product were purified again by preparative RP-HPLC (Method A) for final analytics and protein binding assays. Mixture of diastereomers with the ratio 43:57 (UPLC-MS analysis).

¹H NMR (500 MHz, DMSO-*d*₆) δ/ppm: 8.26 (d, *J* = 4.4 Hz, 1H), 7.99 (t, *J* = 5.6 Hz, 1H), 7.49 – 7.43 (m, 2H), 7.34 (td, *J* = 7.4, 1.7 Hz, 1H), 7.28 (dd, *J* = 7.5, 1.6 Hz, 1H), 7.25 (td, *J* = 7.3, 1.2 Hz, 2H), 6.96 – 6.92 (m, 2H), 6.64 – 6.61 (m, 2H), 4.32 – 4.29 (m, 1H), 3.78 (ddd, *J* = 8.1, 6.1, 4.3 Hz, 1H), 3.44 (ddt, *J* = 13.4, 5.4, 3.3 Hz, 1H), 3.37 – 3.33 (m, 2H), 3.18 – 3.07 (m, 3H), 3.02 – 2.95 (m, 1H), 2.85 (ddd, *J* = 13.9, 8.6, 5.5 Hz, 1H), 2.73 (dd, *J* = 14.4, 10.4 Hz, 1H), 2.64 (dt, *J* = 13.8, 8.1 Hz, 1H), 2.17 (t, *J* = 7.0 Hz, 2H), 1.90 – 1.82 (m, 1H), 1.74 – 1.63 (m, 2H), 1.64 – 1.56 (m, 1H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 174.56, 171.21, 171.01, 169.89, 155.79, 138.69, 137.68, 129.75, 129.68, 129.13, 128.19, 127.27, 125.84, 114.94, 53.93, 53.14, 47.08, 39.24, 38.82, 35.31, 33.29, 31.82, 29.75, 24.78.

HRMS (ESI): C₂₆H₃₁N₇NaO₅⁺ *calcd*: 544.2279, *found*: 544.2279.

MC NP14-AA020-TA607 2-20



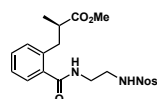
In a 10 mL flask **2-20f** (150.0 mg, 288.0 μmol , 1.0 eq.) and **TA607** (32.0 mg, 288.0 μmol , 1.0 eq.) were dissolved in DMSO (3 mL) and a solution of sodium ascorbate (34.2 mg, 173.0 μmol , 0.6 eq.) in H₂O (575 μL) was added. The solution was degassed by bubbling N₂ through the solution for 1 min. A solution of copper(II) sulfate pentahydrate (21.5 mg, 86.3 μmol , 0.3 eq.) in H₂O (431 μL) was added and the mixture was degassed again for 1 min. The reaction was stirred at RT for 2 h, after which UPLC-MS analysis showed full conversion to the desired product. The solution was directly purified by preparative RP-HPLC (Method A) and the product-containing fractions were lyophilized to yield the desired product as a white solid (90 mg, **50%**). Mixture of diastereomers in the ratio 27:23:50 (UPLC-MS).

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 8.37 (d, *J* = 4.6 Hz, 1H), 8.00 (t, *J* = 5.6 Hz, 1H), 7.93 (t, *J* = 5.7 Hz, 1H), 7.74 (s, 1H), 7.54 (d, *J* = 8.4 Hz, 1H), 7.47 (t, *J* = 5.6 Hz, 1H), 7.34 (td, *J* = 7.4, 1.8 Hz, 1H), 7.26 (dtd, *J* = 18.1, 7.5, 1.5 Hz, 3H), 6.97 – 6.92 (m, 2H), 6.64 – 6.60 (m, 2H), 4.30 (ddd, *J* = 10.2, 8.4, 4.5 Hz, 1H), 4.25 – 4.19 (m, 1H), 4.15 – 4.11 (m, 1H), 3.67 (ddd, *J* = 8.2, 6.2, 4.5 Hz, 1H), 3.44 (dtd, *J* = 13.5, 5.6, 5.2, 2.7 Hz, 1H), 3.38 – 3.33 (m, 2H), 3.30 – 3.25 (m, 2H), 3.11 (dd, *J* = 14.4, 4.4 Hz, 1H), 3.04 – 2.96 (m, 1H), 2.84 (ddd, *J* = 14.1, 8.7, 5.5 Hz, 1H), 2.78 – 2.72 (m, 3H), 2.65 (dt, *J* = 13.8, 7.8 Hz, 1H), 2.19 (t, *J* = 6.9 Hz, 2H), 2.04 – 1.94 (m, 2H), 1.86 (ddd, *J* = 13.7, 8.5, 7.1 Hz, 1H), 1.79 (s, 3H), 1.74 – 1.65 (m, 1H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ /ppm: 174.60, 171.15, 170.99, 169.88, 169.25, 155.73, 144.34, 138.71, 137.67, 129.72, 129.46, 129.14, 128.22, 127.26, 125.83, 122.43, 115.01, 54.07, 52.98, 45.97, 39.06, 38.78, 38.42, 35.26, 33.35, 31.83, 31.08, 25.58, 24.90, 22.64.

HRMS (ESI): C₃₂H₄₀N₈NaO₆⁺ *calcd*: 655.2963, *found*: 655.2964.

Nos-Sc-NP12-OMe 2-21a*



Under an inert atmosphere **2-24** (3.0 g, 6.3 mmol, 1.0 eq.) and Pd (π -cinnamyl) chloride dimer (327.0 mg, 631.0 μmol , 10 mol%) were dissolved in THF (60 mL) and the solution was cooled to 0°C. A solution of (*S*)-3-methoxy-2-methyl-3-oxopropylzinc bromide (0.5 M in THF, 18.9 mL, 9.5 mmol, 1.5 eq.) was slowly added and the solution was stirred at 0°C for 15 min. The cooling bath was removed and the mixture was stirred at RT for 45 min, after which UPLC-MS analysis showed full consumption of the starting material. The reaction was stopped by the addition of H₂O (5 mL) and the mixture was filtered over G4. The volatiles were removed *in vacuo* and the residue was dissolved in EtOAc (150 mL). The mixture was washed with half-saturated NH₄Cl (100 mL) and H₂O (2 x 100 mL). The combined aqueous layers were extracted with EtOAc (2 x 100 mL) and the combined organic layers were treated with brine (100 mL), dried over Na₂SO₄ and concentrated by rotary evaporation. The crude was purified by flash column chromatography on the ISOLERA (Silica, 340 g, cyclohexane:EtOAc, UV). The product-containing fractions were combined and concentrated to yield the desired product as a thick, slightly red oil (2.3 g, **80%**).

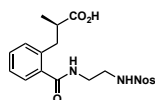
¹H NMR (400 MHz, CD₃CN) δ /ppm: 8.11 – 8.02 (m, 1H), 7.88 – 7.83 (m, 1H), 7.83 – 7.76 (m, 2H), 7.38 – 7.31 (m, 2H), 7.29 – 7.18 (m, 2H), 6.96 (s, 1H), 6.30 (t, *J* = 5.7 Hz, 1H), 3.53 (s, 3H),

3.48 – 3.38 (m, 2H), 3.25 (dt, $J = 5.7, 5.9$ Hz, 2H), 3.12 – 3.03 (m, 1H), 2.85 – 2.74 (m, 2H), 1.07 (d, $J = 6.6$ Hz, 3H).

^{13}C NMR (101 MHz, CD_3CN) δ/ppm : 177.33, 171.17, 148.94, 138.72, 137.49, 135.06, 133.82, 133.81, 131.55, 131.40, 130.65, 128.20, 127.28, 125.96, 52.04, 44.12, 41.99, 40.24, 37.43, 17.53.

HRMS (ESI): $\text{C}_{20}\text{H}_{23}\text{N}_3\text{NaO}_7\text{S}^+$ *calcd*: 472.1149, *found*: 472.1152.

Nos-Sc-NP12-OH **2-21a**



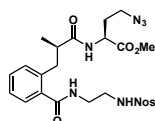
In a 100 mL flask **2-21a*** (2.2 g, 4.9 mmol, 1.0 eq.) was dissolved in MeCN (30 mL) and a solution of LiOH monohydrate (1.0 g, 24.7 mmol, 5.0 eq.) in H_2O (15 mL) was added. The mixture was stirred at RT for 2.5 h, after which UPLC-MS analysis showed complete conversion to the desired product. Acidifying with concentrated HCl to pH 5-6 quenched the reaction. The volatiles were removed *in vacuo* and the aqueous residue was taken up in EtOAc (150 mL) and H_2O (100 mL). The mixture was acidified to pH 1 with concentrated HCl. The layers were separated (addition of brine for good separation) and the aqueous layer was extracted with EtOAc (2 x 100 mL). The combined organic layers were washed with half-saturated brine (2 x 100 mL) and brine (100 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The desired product was obtained as an off-white solid (2.2 g, **quant. 93% HPLC purity**).

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ/ppm : 12.06 (s, 1H), 8.34 (t, $J = 5.7$ Hz, 1H), 8.18 (t, $J = 5.9$ Hz, 1H), 8.06 – 8.00 (m, 1H), 8.00 – 7.95 (m, 1H), 7.91 – 7.83 (m, 2H), 7.37 – 7.31 (m, 2H), 7.28 – 7.19 (m, 2H), 3.35 – 3.26 (m, 2H), 3.12 – 3.05 (m, 2H), 3.02 (dd, $J = 13.3, 6.9$ Hz, 1H), 2.72 (dd, $J = 13.3, 7.6$ Hz, 1H), 2.67 – 2.57 (m, 1H), 0.97 (d, $J = 6.9$ Hz, 3H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ/ppm : 177.00, 169.29, 147.70, 137.53, 136.89, 134.08, 132.72, 132.67, 130.29, 129.48, 129.25, 127.38, 125.99, 124.51, 42.04, 40.28, 39.00, 36.05, 16.90.

HRMS (ESI): $\text{C}_{19}\text{H}_{22}\text{N}_3\text{O}_7\text{S}^+$ *calcd*: 436.1173, *found*: 436.1167.

Nos-Sc-NP12-TFL(N_3)-OMe **2-21b**



In a 100 mL flask **2-21a** (2.2 g, 4.9 mmol, 1.0 eq.), **2-5** (937.0 mg, 5.9 mmol, 1.2 eq.), HBTU (2.2 g, 5.9 mmol, 1.2 eq.) and DIPEA (2.6 mL, 14.8 mmol, 3.0 eq.) were dissolved in THF (42 mL) and the mixture was stirred at RT for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The solvent was removed by rotary evaporation and the crude mixture was purified by flash column chromatography on the ISOLERA (Silica, 340 g, DCM/MeOH, UV). Product-containing fractions were combined and concentrated to yield the desired product as a thick, yellow oil (4.0 g, **>100%**). NMR showed co-eluted byproducts from the coupling reagent. The material was used without further purification for the next synthetic step.

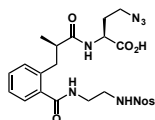
^1H NMR (400 MHz, CDCl_3) δ/ppm : 8.13 – 8.05 (m, 1H), 7.81 – 7.74 (m, 1H), 7.75 – 7.66 (m, 2H), 7.58 (t, $J = 5.8$ Hz, 1H), 7.41 – 7.28 (m, 2H), 7.25 – 7.16 (m, 2H), 6.79 (d, $J = 8.1$ Hz, 1H), 6.75 (t, $J = 6.0$ Hz, 1H), 4.53 (td, $J = 8.3, 4.6$ Hz, 1H), 3.67 (s, 3H), 3.66 – 3.61 (m, 1H), 3.54 – 3.44 (m, 1H), 3.42 – 3.30 (m, 2H), 3.19 – 3.05 (m, 2H), 2.87 – 2.80 (m, 2H), 2.71 – 2.63 (m, 1H), 1.88

(dddd, $J = 14.2, 7.7, 6.6, 4.7$ Hz, 1H), 1.66 (ddt, $J = 14.5, 8.4, 6.2$ Hz, 1H), 1.21 (d, $J = 6.7$ Hz, 3H).

^{13}C NMR (101 MHz, CDCl_3) δ /ppm: 176.43, 172.10, 171.06, 148.07, 137.35, 136.20, 133.75, 133.67, 132.81, 130.84, 130.32, 130.19, 127.94, 126.69, 125.15, 52.66, 49.60, 47.10, 43.39, 43.22, 43.04, 40.39, 31.04, 18.21.

HRMS (ESI): $\text{C}_{24}\text{H}_{29}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 598.1691, *found*: 598.1693.

Nos-Sc-NP12-TFL(N_3)-OH **2-21c**



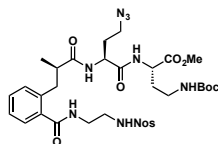
In a 100 mL flask **2-21b** (3.9 g, 6.8 mmol, 1.0 eq.) was dissolved in MeCN (40 mL) and a solution of LiOH monohydrate (1.4 g, 34.0 mmol, 5.0 eq.) in H_2O (20 mL) was added. The mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed full conversion. Acidifying the mixture to pH 5-6 with concentrated HCl quenched the reaction and the volatiles were removed *in vacuo*. The residue was taken up in EtOAc (100 mL) and H_2O (100 mL). The biphasic mixture was acidified to pH 1 with concentrated HCl and the layers were separated. The aqueous layer was extracted with EtOAc (2 x 100 mL) and the combined organic layers were washed with half-saturated brine (2 x 100 mL) and brine (100 mL), dried over MgSO_4 and concentrated to yield the desired product as a slightly yellow solid (2.8 g, **73%**). NMR showed some residual tetramethylurea from the HBTU coupling step. The material was used without further purification for the next synthetic step.

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 12.65 (s, 1H), 8.31 (t, $J = 5.7$ Hz, 1H), 8.21 (t, $J = 5.9$ Hz, 1H), 8.06 – 8.01 (m, 1H), 8.01 – 7.94 (m, 2H), 7.91 – 7.83 (m, 2H), 7.37 – 7.28 (m, 2H), 7.27 – 7.17 (m, 2H), 4.23 (ddd, $J = 9.7, 8.1, 4.4$ Hz, 1H), 3.48 – 3.20 (m, 3H), 3.18 – 2.98 (m, 4H), 2.95 – 2.83 (m, 1H), 2.76 – 2.62 (m, 1H), 1.86 (dtd, $J = 13.9, 7.7, 4.5$ Hz, 1H), 1.76 – 1.57 (m, 1H), 0.98 (d, $J = 6.4$ Hz, 3H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ /ppm: 175.30, 173.04, 169.30, 147.72, 137.80, 136.86, 134.07, 132.70, 132.67, 129.93, 129.47, 129.13, 127.32, 125.82, 124.49, 48.94, 47.25, 42.08, 41.26, 36.33, 30.13, 17.83.

HRMS (ESI): $\text{C}_{23}\text{H}_{27}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 584.1534, *found*: 584.1540.

Nos-Sc-NP12-TFL(N_3)-AA070-OMe **2-21d**



In a 50 mL flask **2-21c** (1.3 g, 2.3 mmol, 1.0 eq.) was dissolved in THF (20 mL) and H-Dab(Boc)-OMe hydrochloride (747.0 mg, 2.8 mmol, 1.2 eq.) was added, followed by DIPEA (1.2 mL, 6.9 mmol, 3.0 eq.) and HATU (1.1 g, 2.8 mmol, 1.2 eq.). The mixture was stirred at RT for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the residue was taken up in EtOAc (100 mL) and half-saturated NH_4Cl solution (100 mL). The biphasic mixture was separated and the aqueous layer was extracted with EtOAc (2 x 100 mL) and the combined organic layers were washed with saturated NH_4Cl solution (2 x 100 mL), brine (100 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The crude was purified by flash column chromatography on the ISOLERA (Silica, 120 g, DCM/MeOH,

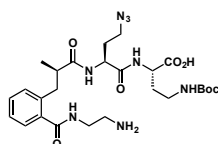
UV) to yield the desired product as a slightly yellow solid (1.8 g, >100%). NMR showed some residual tetramethylurea from the coupling step. The material was used without further purification for the next synthetic transformation.

^1H NMR (400 MHz, CD_3CN) δ /ppm: 8.11 – 8.04 (m, 1H), 7.88 – 7.82 (m, 1H), 7.82 – 7.77 (m, 2H), 7.41 – 7.31 (m, 3H), 7.31 – 7.20 (m, 2H), 7.06 (d, J = 8.5 Hz, 1H), 6.93 (d, J = 8.1 Hz, 1H), 6.89 – 6.81 (m, 1H), 5.40 (s, 1H), 4.38 (ddd, J = 9.1, 8.0, 4.8 Hz, 1H), 4.30 (ddd, J = 9.0, 8.0, 4.8 Hz, 1H), 3.64 (s, 3H), 3.63 – 3.51 (m, 1H), 3.42 – 3.34 (m, 1H), 3.33 – 3.12 (m, 3H), 3.11 – 3.02 (m, 2H), 3.01 – 2.84 (m, 3H), 2.82 – 2.76 (m, 2H), 1.82 (dtd, J = 14.0, 7.6, 4.9 Hz, 1H), 1.70 (dtd, J = 14.7, 9.1, 6.0 Hz, 1H), 1.57 (dddd, J = 14.4, 9.0, 7.1, 5.6 Hz, 1H), 1.38 (s, 9H), 1.14 (d, J = 6.5 Hz, 3H).

^{13}C NMR (101 MHz, CD_3CN) δ /ppm: 177.24, 173.07, 172.00, 171.26, 156.86, 148.97, 138.68, 137.59, 135.00, 133.97, 133.79, 131.40, 131.30, 130.70, 128.48, 127.28, 125.84, 80.09, 53.61, 52.89, 51.25, 50.77, 48.16, 44.01, 43.47, 40.60, 37.68, 32.43, 31.66, 28.60, 18.36.

HRMS (ESI): $\text{C}_{33}\text{H}_{45}\text{N}_9\text{NaO}_{11}\text{S}^+$ *calcd*: 798.2851, *found*: 798.2862.

NH_2 -Sc-NP12-TFL(N_3)-AA070-OH **2-21e**



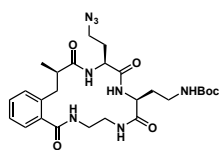
Under an inert (N_2) atmosphere **2-21d** (1.8 g, 2.3 mmol, 1.0 eq.) was dissolved in MeCN (24 mL) and DIPEA (4.0 mL, 22.8 mmol, 10.0 eq.) was added, followed by thiophenol (1.2 mL, 11.4 mmol, 5.0 eq.). The yellow solution was stirred at RT for 2 h, after which UPLC-MS analysis showed complete Nosyl deprotection. A solution of LiOH monohydrate (764.0 mg, 18.2 mmol, 8.0 eq.) in H_2O (8 mL) was added and the mixture was stirred for 1.5 h, after which UPLC-MS analysis showed complete ester deprotection. The volatiles were removed by rotary evaporation and the crude was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, $\text{H}_2\text{O}/\text{MeCN}$, UV (200/215 nm)). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (999 mg, **76%**). NMR and LC-MS analysis showed 19% epimerized product.

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 8.55 (t, J = 5.9 Hz, 1H), 8.25 (d, J = 8.1 Hz, 1H), 7.66 (d, J = 6.9 Hz, 1H), 7.48 – 7.26 (m, 2H), 7.27 – 7.02 (m, 2H), 6.66 (t, J = 5.7 Hz, 1H), 4.16 (td, J = 8.8, 8.2, 4.2 Hz, 1H), 3.73 (q, J = 6.5 Hz, 1H), 3.50 – 3.34 (m, 1H), 3.37 – 3.20 (m, 1H), 3.15 – 2.52 (m, 9H), 2.03 – 1.61 (m, 2H), 1.63 – 1.39 (m, 2H), 1.35 (s, 9H), 1.04 (d, J = 6.2 Hz, 3H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ /ppm: 175.43, 173.73, 169.99, 169.43, 155.34, 137.73, 137.19, 130.05, 129.03, 127.24, 125.79, 77.34, 52.51, 50.27, 47.31, 41.60, 39.80, 39.51, 37.31, 36.28, 33.22, 30.68, 28.26, 17.95.

HRMS (ESI): $\text{C}_{26}\text{H}_{41}\text{N}_8\text{O}_7^+$ *calcd*: 577.3093, *found*: 577.3101.

MC NP12-AA070 2-21f



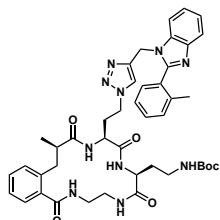
In a 1 L flask **2-21e** (955.0 mg, 1.7 mmol, 1.0 eq.) was dissolved in THF (600 mL) and DIPEA (865.0 μ L, 5.0 mmol, 3.0 eq.) was added, followed by HATU (1.1 g, 3.0 mmol, 1.8 eq.). The mixture was stirred at RT for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The mixture was concentrated by rotary evaporation and purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, H₂O:MeCN, UV 200/215 nm) to yield the desired product as a white solid (370.0 mg, **40%**). 89 mg of the product were purified again by preparative RP-HPLC (Method A, no TFA) for final analytics and protein binding assays. Mixture of diastereomers with the ratio 10:85:5 (UPLC-MS analysis).

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 8.20 (d, *J* = 5.4 Hz, 1H), 8.09 (t, *J* = 5.4 Hz, 1H), 7.74 – 7.68 (m, 1H), 7.66 (d, *J* = 8.0 Hz, 1H), 7.36 (td, *J* = 7.4, 1.6 Hz, 1H), 7.29 (d, *J* = 7.2 Hz, 1H), 7.27 (dd, *J* = 7.6, 1.6 Hz, 1H), 7.22 (td, *J* = 7.4, 1.3 Hz, 1H), 6.76 (t, *J* = 5.5 Hz, 1H), 4.05 (ddd, *J* = 10.2, 7.9, 4.3 Hz, 1H), 3.72 (dt, *J* = 8.5, 5.8 Hz, 1H), 3.53 – 3.46 (m, 1H), 3.46 – 3.37 (m, 2H), 3.29 – 3.19 (m, 3H), 3.16 – 3.09 (m, 1H), 2.98 – 2.87 (m, 2H), 2.72 – 2.62 (m, 1H), 2.56 (dd, *J* = 13.9, 7.1 Hz, 1H), 2.06 – 1.99 (m, 1H), 1.99 – 1.89 (m, 2H), 1.78 – 1.69 (m, 1H), 1.36 (s, 9H), 1.18 (d, *J* = 7.0 Hz, 3H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ /ppm: 176.63, 171.27, 171.08, 169.56, 155.53, 137.72, 137.34, 129.51, 129.09, 127.47, 125.84, 77.52, 53.44, 51.25, 47.65, 42.21, 38.84, 38.72, 37.17, 35.67, 30.40, 29.33, 28.27, 17.94.

HRMS (ESI): C₂₆H₃₈N₈NaO₆⁺ *calcd*: 581.2807, *found*: 581.2808.

MC NP12-AA070-TA333 2-21



In a 10 mL flask **2-21f** (150.0 mg, 269.0 μ mol, 1.0 eq.) and **TA333** (66.1 mg, 269.0 μ mol, 1.0 eq.) were dissolved in DMSO (2.9 mL) and a solution of sodium ascorbate (31.9 mg, 161.0 μ mol, 0.6 eq.) in H₂O (537 μ L) was added. The solution was degassed by bubbling N₂ through the solution for 1 min. A solution of copper(II) sulfate pentahydrate (20.1 mg, 80.6 μ mol, 0.3 eq.) in H₂O (403 μ L) was added and the mixture was degassed again for 1 min. The reaction was stirred at RT for 2 h, after which UPLC-MS analysis showed full conversion to the desired product. The solution was directly purified by preparative RP-HPLC (Method A, no TFA) and the product-containing fractions were lyophilized to yield the desired product as a white solid (142 mg, **66%**). Mixture of diastereomers in the ratio 87:6:7 (NMR).

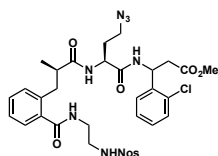
¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 8.24 (d, *J* = 5.3 Hz, 1H), 8.11 (t, *J* = 5.6 Hz, 1H), 7.91 (s, 1H), 7.72 – 7.67 (m, 2H), 7.65 (d, *J* = 8.1 Hz, 2H), 7.55 (dd, *J* = 7.6, 1.4 Hz, 1H), 7.48 (td, *J* = 7.5, 1.5 Hz, 1H), 7.41 (d, *J* = 7.6 Hz, 1H), 7.39 – 7.34 (m, 2H), 7.33 – 7.29 (m, 2H), 7.28 – 7.25 (m, 2H), 7.22 (td, *J* = 7.4, 1.2 Hz, 1H), 6.76 (t, *J* = 5.7 Hz, 1H), 5.33 (s, 2H), 4.36 (t, *J* = 7.2 Hz, 2H), 4.06 (ddd, *J* = 10.5, 8.0, 4.3 Hz, 1H), 3.61 (dt, *J* = 8.8, 5.8 Hz, 1H), 3.49 – 3.46 (m, 1H), 3.30 – 3.23 (m, 3H), 3.23 – 3.17 (m, 1H), 3.16 – 3.10 (m, 1H), 2.91 (ddt, *J* = 20.5, 13.6, 5.9 Hz, 2H), 2.68

– 2.57 (m, 1H), 2.27 – 2.16 (m, 2H), 2.15 (s, 3H), 2.04 – 1.96 (m, 1H), 1.75 – 1.66 (m, 1H), 1.31 (s, 9H), 1.14 (d, $J = 7.0$ Hz, 3H).

^{13}C NMR (101 MHz, DMSO- d_6) δ /ppm: 176.72, 171.08, 171.02, 169.54, 155.54, 152.28, 141.80, 140.64, 138.00, 137.69, 137.29, 134.10, 130.50, 130.22, 129.51, 129.09, 128.77, 128.48, 127.48, 125.86, 125.80, 123.63, 122.90, 122.60, 118.48, 111.47, 77.52, 53.40, 51.12, 46.60, 42.25, 39.40, 38.78, 38.77, 37.10, 35.68, 30.72, 30.47, 28.24, 19.43, 17.83.

HRMS (ESI): $\text{C}_{43}\text{H}_{53}\text{N}_{10}\text{O}_6^+$ *calcd*: 805.4144, *found*: 805.4142.

Nos-Sc-NP12-TFL(N_3)-AA118-OMe 2-22d



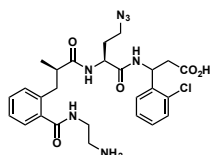
In a 25 mL flask **2-21c** (350.0 mg, 623.0 μmol , 1.0 eq.) was dissolved in THF (5 mL) and methyl 3-amino-3-(2-chlorophenyl)propanoate (160.0 mg, 748.0 μmol , 1.2 eq.) was added, followed by DIPEA (326.0 μL , 1.9 mmol, 3.0 eq.) and HATU (284.0 mg, 746.0 μmol , 1.2 eq.). The mixture was stirred at RT for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the residue was taken up in EtOAc (50 mL) and H_2O (50 mL). The mixture was acidified with concentrated HCl to pH 1. The layers were separated and the aqueous phase was extracted with EtOAc (2 x 50 mL) and the combined organic layers were washed with diluted HCl solution (1%, 2 x 50 mL), brine (50 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The crude was purified by flash column chromatography on the ISOLERA (Silica, 120 g, DCM/MeOH, UV) to yield the desired product as a slightly yellow solid (403 mg, **85%**). NMR and UPLC-MS analysis showed a mixture of four diastereomers in the ratio 7:36:42:15.

^1H NMR (500 MHz, CD_3CN) δ /ppm: 8.12 – 7.99 (m, 1H), 7.91 – 7.80 (m, 1H), 7.80 – 7.70 (m, 2H), 7.40 – 7.34 (m, 3H), 7.34 – 7.13 (m, 7H), 6.96 (t, $J = 6.0$ Hz, 1H), 6.86 (d, $J = 8.1$ Hz, 1H), 5.58 – 5.44 (m, 1H), 4.42 – 4.14 (m, 1H), 3.59 – 3.55 (m, 3H), 3.49 – 3.11 (m, 4H), 3.12 – 3.03 (m, 1H), 2.90 (dddd, $J = 14.5, 12.7, 7.3, 5.4$ Hz, 1H), 2.83 – 2.73 (m, 3H), 2.72 – 2.56 (m, 2H), 1.83 – 1.67 (m, 1H), 1.61 – 1.44 (m, 1H), 1.18 (d, $J = 6.6$ Hz, 3H).

^{13}C NMR (101 MHz, CD_3CN) δ /ppm: 177.21, 171.70, 171.23, 166.26, 149.09, 139.60, 139.45, 138.74, 137.70, 137.56, 134.93, 132.74, 131.45, 131.24, 131.13, 130.54, 129.82, 128.52, 128.42, 128.31, 127.28, 125.75, 52.29, 51.25, 48.45, 48.08, 43.80, 43.43, 40.56, 39.37, 37.52, 31.36, 18.33.

HRMS (ESI): $\text{C}_{33}\text{H}_{37}\text{ClN}_8\text{NaO}_9\text{S}^+$ *calcd*: 779.1985, *found*: 779.1992.

NH_2 -Sc-NP12-TFL(N_3)-AA118-OH 2-22e



Under an inert (N_2) atmosphere **2-22d** (406.0 mg, 536.0 μmol , 1.0 eq.) was dissolved in MeCN (5 mL) and DIPEA (934.0 μL , 5.4 mmol, 10.0 eq.) was added, followed by thiophenol (273.0 μL , 2.7 mmol, 5.0 eq.). The yellow solution was stirred at RT for 2 h, after which UPLC-MS analysis showed complete Nosyl deprotection. A solution of LiOH monohydrate (171.0 mg, 4.1 mmol, 8.0 eq.) in H_2O (1.5 mL) was added and the mixture was stirred for 1.5 h, after which UPLC-MS

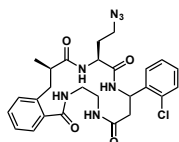
analysis showed complete ester deprotection. The mixture was acidified to pH 5-6 by the addition of concentrated HCl. The volatiles were removed by rotary evaporation and the crude was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 120 g, H₂O/MeCN + 0.1% TFA, UV (200/215 nm)). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (331 mg, >100%). Co-elution of unknown side products. NMR and LC-MS analysis showed a diastereomeric ratio of 9:36:43:12.

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 8.56 (d, *J* = 7.7 Hz, 1H), 8.39 (t, *J* = 5.4 Hz, 1H), 8.17 – 7.58 (m, 4H), 7.57 – 6.82 (m, 8H), 5.48 (dd, *J* = 8.0, 6.3 Hz, 1H), 4.38 – 4.21 (m, 1H), 3.58 – 3.17 (m, 2H), 3.09 – 2.81 (m, 4H), 2.80 – 2.54 (m, 4H), 2.48 – 2.33 (m, 1H), 1.80 – 1.44 (m, 2H), 0.96 (d, *J* = 6.3 Hz, 3H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 171.41, 171.35, 170.06, 169.71, 139.51, 139.48, 137.97, 136.41, 131.70, 131.50, 130.07, 129.41, 128.80, 127.47, 127.34, 125.84, 49.78, 47.17, 46.98, 41.23, 39.22, 38.63, 36.91, 36.48, 31.14, 17.76.

HRMS (ESI): C₂₆H₃₃ClN₇O₅⁺ *calcd*: 558.2226, *found*: 558.2218.

MC NP12-AA118 2-22f



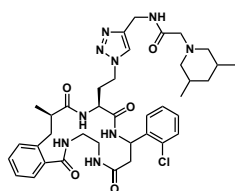
In a 500 mL flask **2-22e** (298.0 mg, 534 μmol, 1.0 eq.) was dissolved in THF (250 mL) and DIPEA (326.0 μL, 1.9 mmol, 3.5 eq.) was added, followed by HATU (406.0 mg, 1.1 mmol, 2.0 eq.). The mixture was stirred at RT for 1.5 h, after which UPLC-MS analysis showed complete conversion to the desired product. The mixture was concentrated by rotary evaporation and purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 120 g, H₂O:MeCN +0.1% TFA, UV 200/215 nm) to yield the desired product as a white solid (148.0 mg, **51%**). 32 mg of the product were purified again by preparative RP-HPLC (Method A) for final analytics and protein binding assays. Mixture of diastereomers with the ratio 50:38:12 (UPLC-MS analysis).

¹H NMR (500 MHz, DMSO-*d*₆) δ/ppm: 8.75 (dd, *J* = 7.0, 3.7 Hz, 1H), 8.39 (d, *J* = 7.1 Hz, 1H), 8.01 (d, *J* = 4.7 Hz, 1H), 7.85 (t, *J* = 5.6 Hz, 1H), 7.42 – 7.39 (m, 2H), 7.37 – 7.35 (m, 1H), 7.34 – 7.33 (m, 1H), 7.31 – 7.29 (m, 1H), 7.26 (dd, *J* = 5.3, 2.2 Hz, 2H), 7.21 (td, *J* = 7.4, 1.3 Hz, 1H), 5.35 (ddd, *J* = 10.2, 7.1, 2.9 Hz, 1H), 3.75 – 3.57 (m, 3H), 3.38 – 3.30 (m, 2H), 3.17 – 3.14 (m, 1H), 3.13 – 3.10 (m, 2H), 2.90 – 2.80 (m, 1H), 2.63 – 2.58 (m, 1H), 2.49 – 2.46 (m, 1H), 2.37 (dd, *J* = 15.5, 3.0 Hz, 1H), 1.72 – 1.62 (m, 1H), 1.64 – 1.53 (m, 1H), 1.13 (d, *J* = 6.7 Hz, 3H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 175.93, 170.03, 169.33, 168.88, 140.21, 137.89, 137.63, 131.18, 129.42, 129.33, 129.16, 128.66, 127.72, 127.40, 125.93, 125.67, 51.25, 47.22, 40.98, 40.83, 39.17, 39.16, 38.99, 35.51, 29.13, 18.36.

HRMS (ESI): C₂₆H₃₀ClN₇NaO₄⁺ *calcd*: 562.1940, *found*: 562.1940.

MC NP12-AA118-TA622 2-22



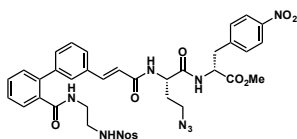
In a 10 mL flask **2-22f** (100.0 mg, 185.0 μmol , 1.0 eq.) and **TA622** (38.6 mg, 185.0 μmol , 1.0 eq.) were dissolved in DMSO (2 mL) and a solution of sodium ascorbate (22.0 mg, 111.0 μmol , 0.6 eq.) in H₂O (370 μL) was added. The solution was degassed by bubbling N₂ through the solution for 1 min. A solution of copper(II) sulfate pentahydrate (13.9 mg, 55.6 μmol , 0.3 eq.) in H₂O (278 μL) was added and the mixture was degassed again for 1 min. The reaction was stirred at RT for 14 h, after which UPLC-MS analysis showed full conversion to the desired product. The solution was directly purified by preparative RP-HPLC (Method A) and the product-containing fractions were lyophilized to yield the desired product as a white solid (129 mg, **93%**). Mixture of diastereomers in the ratio 57:31:11 (UPLC). NMR showed the protonated species of the tertiary amine sidechain.

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 9.77 (s, 1H), 9.09 – 9.00 (m, 1H), 8.81 (dd, *J* = 6.8, 3.8 Hz, 1H), 8.31 (d, *J* = 7.3 Hz, 1H), 8.14 (d, *J* = 5.7 Hz, 1H), 7.93 (t, *J* = 5.7 Hz, 1H), 7.87 (s, 1H), 7.47 – 7.42 (m, 1H), 7.42 – 7.40 (m, 1H), 7.40 – 7.38 (m, 1H), 7.38 – 7.35 (m, 1H), 7.32 – 7.23 (m, 4H), 5.40 (ddd, *J* = 10.3, 7.3, 3.2 Hz, 1H), 4.37 (t, *J* = 5.7 Hz, 2H), 4.18 (ddd, *J* = 14.2, 9.0, 5.6 Hz, 1H), 4.10 (ddd, *J* = 13.8, 8.9, 6.6 Hz, 1H), 3.87 (s, 2H), 3.80 – 3.78 (m, 1H), 3.69 – 3.65 (m, 1H), 3.52 – 3.43 (m, 1H), 3.34 – 3.30 (m, 2H), 3.29 – 3.24 (m, 1H), 3.24 – 3.20 (m, 1H), 3.19 – 3.11 (m, 1H), 2.92 – 2.82 (m, 1H), 2.64 (dd, *J* = 13.6, 6.5 Hz, 1H), 2.59 – 2.53 (m, 2H), 2.49 – 2.35 (m, 2H), 2.12 – 2.03 (m, Hz, 1H), 2.00 – 1.93 (m, 2H), 1.93 – 1.88 (m, 1H), 1.73 (d, *J* = 13.0 Hz, 1H), 1.15 (d, *J* = 6.8 Hz, 3H), 0.86 (d, *J* = 6.6 Hz, 6H), 0.76 (q, *J* = 12.2 Hz, 1H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ /ppm: 175.89, 169.77, 169.38, 168.95, 163.93, 143.62, 140.17, 137.85, 136.99, 131.19, 129.44, 129.40, 129.37, 128.65, 127.92, 127.64, 127.39, 125.80, 122.98, 57.75, 56.79, 51.03, 47.46, 46.23, 40.87, 40.81, 39.39, 39.18, 38.55, 35.78, 34.23, 30.73, 28.15, 18.26, 17.97.

HRMS (ESI): C₃₈H₅₁ClN₉O₅⁺ *calcd*: 748.3696, *found*: 748.3685.

Nos-Sc-NP01-TFL(N₃)-AA058-OMe 2-23d



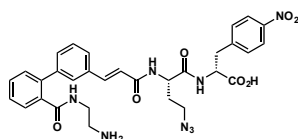
In a 50 mL flask **2-18c** (1.1 g, 1.8 mmol, 1.0 eq.) was dissolved in THF (16 mL) and H-*p*-Nitro-*D*-Phe-OMe hydrochloride (560.0 mg, 2.2 mmol, 1.2 eq.) was added, followed by DIPEA (936.0 μL , 5.4 mmol, 3.0 eq.) and HATU (815.0 mg, 2.1 mmol, 1.2 eq.). The mixture was stirred at RT for 1 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed *in vacuo* and the residue was taken up in EtOAc (100 mL) and H₂O (100 mL). The mixture was acidified with concentrated HCl to pH 1. The layers were separated and the aqueous phase was extracted with EtOAc (2 x 100 mL) and the combined organic layers were washed with diluted HCl solution (1%, 2 x 100 mL), brine (100 mL), dried over Na₂SO₄ and concentrated by rotary evaporation. The crude was purified by flash column chromatography on the ISOLERA (Silica, 120 g, DCM/MeOH, UV) to yield the desired product as a slightly yellow solid (1.3 g, **88%**).

^1H NMR (400 MHz, CD_3CN) δ /ppm: 8.14 – 8.03 (m, 2H), 8.01 – 7.94 (m, 1H), 7.92 – 7.73 (m, 3H), 7.64 – 7.54 (m, 1H), 7.54 – 7.44 (m, 4H), 7.44 – 7.37 (m, 4H), 7.38 – 7.31 (m, 2H), 7.19 (d, $J = 8.2$ Hz, 1H), 6.97 (d, $J = 7.8$ Hz, 1H), 6.63 (t, $J = 6.1$ Hz, 1H), 6.58 (d, $J = 15.7$ Hz, 1H), 6.11 (t, $J = 5.8$ Hz, 1H), 4.78 – 4.68 (m, 1H), 4.47 (dtd, $J = 13.4, 8.3, 5.3$ Hz, 1H), 3.66 (s, 3H), 3.38 – 3.18 (m, 5H), 3.09 (dt, $J = 13.9, 8.6$ Hz, 1H), 3.01 – 2.88 (m, 2H), 2.03 – 1.85 (m, 1H), 1.85 – 1.68 (m, 1H).

^{13}C NMR (101 MHz, CD_3CN) δ /ppm: 172.12, 171.96, 170.93, 166.48, 147.93, 145.95, 142.00, 141.08, 140.02, 137.29, 135.88, 135.11, 133.86, 133.65, 131.47, 131.46, 131.40, 131.02, 130.96, 130.94, 129.80, 128.85, 128.78, 128.57, 127.91, 126.02, 124.33, 122.16, 53.85, 53.02, 51.95, 48.62, 43.70, 40.09, 37.59, 31.78.

HRMS (ESI): $\text{C}_{38}\text{H}_{37}\text{N}_9\text{NaO}_{11}\text{S}^+$ *calcd*: 850.2225, *found*: 850.2225.

NH_2 -Sc-NP01-TFL(N_3)-AA058-OH **2-23e**



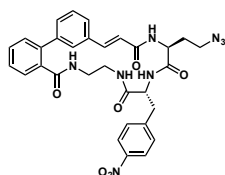
Under an inert (N_2) atmosphere **2-23d** (1.3 g, 1.6 mmol, 1.0 eq.) was dissolved in MeCN (15 mL) and DIPEA (2.7 mL, 15.5 mmol, 10.0 eq.) was added, followed by thiophenol (788.0 μL , 7.7 mmol, 5.0 eq.). The yellow solution was stirred at RT for 2 h, after which UPLC-MS analysis showed complete Nosyl deprotection. A solution of LiOH monohydrate (511.0 mg, 12.2 mmol, 8.0 eq.) in H_2O (5 mL) was added and the mixture was stirred for 1.5 h, after which UPLC-MS analysis showed complete ester deprotection. The mixture was acidified to pH 5-6 by the addition of concentrated HCl. The volatiles were removed by rotary evaporation and the crude was purified by reversed phase flash column chromatography on the ISOLERA (RP-Silica, 340 g, $\text{H}_2\text{O}/\text{MeCN} + 0.1\%$ TFA, UV (254/280 nm)). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (940 mg, **97%**). NMR and LC-MS analysis showed 25% epimerized product.

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ /ppm: 8.55 (d, $J = 8.5$ Hz, 1H), 8.33 (t, $J = 5.6$ Hz, 1H), 8.28 (dd, $J = 8.3, 2.9$ Hz, 1H), 8.13 (dd, $J = 8.8, 7.1$ Hz, 2H), 7.80 (brs, 3H), 7.64 – 7.50 (m, 6H), 7.50 – 7.41 (m, 4H), 7.37 (dt, $J = 7.6, 1.5$ Hz, 1H), 6.79 (d, $J = 15.9$ Hz, 1H), 4.56 (dddd, $J = 10.3, 8.0, 6.4, 3.7$ Hz, 1H), 4.48 (td, $J = 8.7, 5.5$ Hz, 1H), 3.37 – 3.10 (m, 4H), 3.10 – 2.99 (m, 2H), 2.80 – 2.62 (m, 2H), 1.95 – 1.68 (m, 1H), 1.66 – 1.50 (m, 1H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ /ppm: 172.29, 170.72, 169.57, 164.70, 146.29, 146.00, 140.82, 138.97, 138.68, 136.49, 134.79, 130.64, 130.57, 129.83, 129.75, 128.75, 127.86, 127.63, 127.36, 123.18, 123.13, 122.15, 52.65, 50.04, 47.22, 38.10, 36.73, 36.47, 31.45.

HRMS (ESI): $\text{C}_{31}\text{H}_{33}\text{N}_8\text{O}_7^+$ *calcd*: 629.2467, *found*: 629.2463.

MC-NP01-AA058 **2-23f**



In a 50 mL flask **2-23e** (15.0 mg, 23.9 μmol , 1.0 eq.) was dissolved in DMF (20 mL) and DIPEA (12.5 μL , 71.6 μmol , 3.0 eq.) was added, followed by HATU (18.0 mg, 47.7 μmol , 2.0 eq.). The mixture was stirred at RT for 2 h, after which UPLC-MS analysis showed full conversion of the

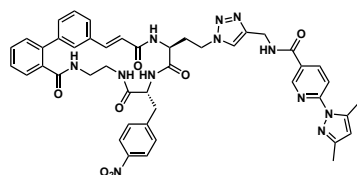
starting material. The mixture was concentrated by rotary evaporation and purified by preparative reversed phase HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (13.0 mg, **89%**).

^1H NMR (600 MHz, $\text{DMSO-}d_6$) δ /ppm: 8.61 – 8.59 (m, 1H), 8.57 (d, $J = 8.5$ Hz, 1H), 8.38 (d, $J = 9.2$ Hz, 1H), 8.13 – 8.12 (m, 2H), 7.61 – 7.58 (m, 1H), 7.57 – 7.54 (m, 2H), 7.53 – 7.52 (m, 2H), 7.48 – 7.47 (m, 4H), 7.43 – 7.41 (m, 2H), 7.41 – 7.39 (m, 1H), 6.47 (d, $J = 15.9$ Hz, 1H), 4.51 – 4.48 (m, 1H), 4.32 (q, $J = 7.6$ Hz, 1H), 3.28 – 3.26 (m, 1H), 3.19 – 3.17 (m, 2H), 3.17 – 3.14 (m, 2H), 3.13 – 3.11 (m, 1H), 2.95 (dd, $J = 13.8, 10.2$ Hz, 2H), 1.89 – 1.82 (m, 1H), 1.78 – 1.72 (m, 1H).

^{13}C NMR (600 MHz, 2D NMR, $\text{DMSO-}d_6$) δ /ppm: 170.52, 170.49, 169.97, 166.13, 146.76, 146.14, 140.0, 137.92, 137.82, 134.76, 130.30, 130.28, 129.25, 129.15, 128.90, 128.85, 128.54, 127.19, 126.84, 124.78, 122.89, 122.58, 53.53, 50.33, 47.32, 39.10, 38.11, 35.58, 28.75

HRMS (ESI): $\text{C}_{31}\text{H}_{30}\text{N}_8\text{NaO}_6^+$ *calcd*: 633.2181, *found*: 633.2176.

MC NP01-AA058-TA229 2-23



In a 1.5 mL Eppendorf tube **2-23f** (10.0 mg, 16.4 μmol , 1.0 eq.) and **TA229** (4.2 mg, 16.4 μmol , 1.0 eq.) were dissolved in DMSO (200 μL) and a solution of sodium ascorbate (2.0 mg, 9.8 μmol , 0.6 eq.) in H_2O (32.8 μL) was added. The solution was degassed by bubbling N_2 through the solution for 1 min. A solution of copper(II) sulfate pentahydrate (1.2 mg, 4.9 μmol , 0.3 eq.) in H_2O (24.6 μL) was added and the mixture was degassed again for 1 min. The reaction was agitated at RT for 12 h, after which UPLC-MS analysis showed full conversion to the desired product. The solution was directly purified by preparative RP-HPLC (Method A) and the product-containing fractions were lyophilized to yield the desired product as a white solid (5 mg, **35%**). Mixture of diastereomers in the ratio 61:39 (UPLC)/47:42:6:5 (NMR).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ 9.24 – 9.19 (m, 1H), 8.89 (d, $J = 3.3$ Hz, 1H), 8.64 (d, $J = 8.4$ Hz, 1H), 8.61 – 8.55 (m, 1H), 8.34 (dt, $J = 7.0, 4.3$ Hz, 1H), 8.27 (d, $J = 9.2$ Hz, 1H), 8.11 (t, $J = 8.3$ Hz, 2H), 8.00 (s, 1H), 7.91 – 7.89 (m, 1H), 7.65 (t, $J = 5.6$ Hz, 1H), 7.63 – 7.59 (m, 1H), 7.55 – 7.54 (m, 1H), 7.53 – 7.50 (m, 2H), 7.49 – 7.47 (m, 2H), 7.47 – 7.45 (m, 2H), 7.43 – 7.41 (m, 2H), 7.41 – 7.37 (m, 2H), 6.55 (d, $J = 15.8$ Hz, 1H), 6.16 (s, 1H), 4.56 – 4.53 (m, 2H), 4.52 – 4.47 (m, 1H), 4.28 – 4.23 (m, 1H), 4.23 – 4.18 (m, 2H), 3.20 – 3.17 (m, 2H), 3.13 – 3.08 (m, 2H), 2.96 – 2.87 (m, 2H), 2.61 (s, 3H), 2.21 (s, 3H), 2.10 – 2.02 (m, 1H), 2.00 – 1.92 (m, 1H).

^{13}C NMR (500 MHz, 2D NMR, $\text{DMSO-}d_6$) δ /ppm: 175.58, 169.91, 165.81, 165.69, 163.64, 154.18, 149.45, 146.76, 146.08, 145.84, 145.80, 144.42, 141.08, 139.93, 137.76, 137.62, 134.50, 130.17, 130.14, 129.53, 128.95, 128.47, 127.13, 126.85, 126.16, 124.76, 123.84, 122.84, 122.74, 122.59, 122.45, 113.83, 109.57, 53.29, 50.13, 45.89, 39.11, 36.52, 35.71, 34.54, 30.68, 14.30, 13.06.

HRMS (ESI): $\text{C}_{45}\text{H}_{44}\text{N}_{12}\text{NaO}_7^+$ *calcd*: 887.3348, *found*: 887.3346.

13.12 Differential Scanning Fluorimetry (DSF)

Stock Solutions:

MC 2-18 , 250 μ M (25% DMSO/PBS)	MC 2-22f , 500 μ M (PBS)
MC 2-19f , 500 μ M (PBS)	MC 2-22 , 500 μ M (PBS)
MC 2-19 , 500 μ M (25% DMSO/PBS)	MC 2-23f , 250 μ M (25% DMSO/PBS)
MC 2-20f , 500 μ M (PBS)	MC 2-23 , 250 μ M (25% DMSO/PBS)
MC 2-20 , 500 μ M (PBS)	AGP 10 mg/ml (PBS)
MC 2-21f , 500 μ M (5% DMSO/PBS)	HSA 10 mg/ml (PBS)

Sample Preparation:

The MC stock solution (97.5 μ L) was mixed with the protein solution (2.5 μ L), giving a final protein concentration of 0.25 mg/ml and 487.5 μ M MC concentration in the according buffer. For every buffer system, a reference sample (buffer + protein) was prepared. All MCs were tested against HSA and AGP (**2-23f** and **2-23** were not completely soluble). The measurements were performed in triplicates on a Prometheus NT.48 with a rate of 1.5°C per minute in a temperature range from 20°C - 95°C.

Remark: AGP (10 mg/ml, 1 mL) was dialyzed against PBS (3 x 1000 mL, 4°C, 3 x 18 h) prior to the binding assay. **2-21** could not be tested due to high fluorescence activity of the compound.

13.13 Isothermal Titration Calorimetry (ITC)

General:

For ITC measurements the protein solution (50.0 μ M in PBS, approx. 300 μ L) was placed in the sample cell and the ligand (macrocycle) solution (500.0 μ M in PBS, approx. 40 μ L) was placed in the syringe. The titrations were performed at 25°C or 10°C with a stirring rate of 750 rpm. Reference power was set to 6 μ cal/s. In total 17 injections were performed with an initial delay of 300 s. The added volume per injection was 2.3 μ L over a duration of 4.6 s (first injection 0.5 μ L in 1.0 s) with a spacing of 150 s between the additions and a filter time of 2 s. The data was analyzed by NitPick and Sedphat for fitting and K_D determinations.

2-22f versus AGP:

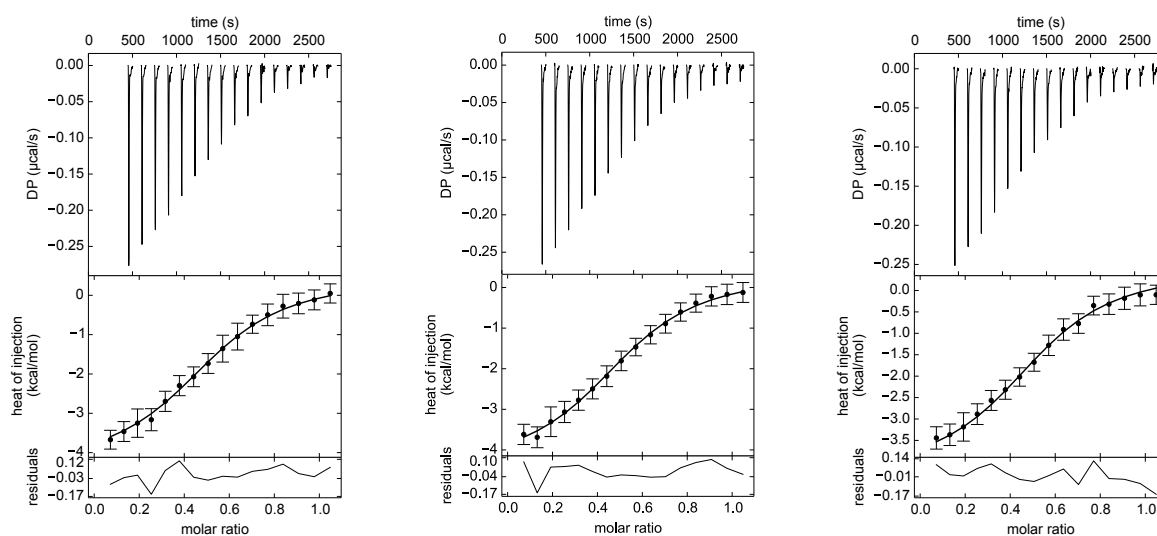


Figure 102. The three ITC titrations of **2-22f** versus AGP. All graphs were analyzed and integrated using NITPIC. For the physicochemical calculations all three measurements were fitted and processed using SEDPHAT.

For optimal binding constant analysis ligand **2-22f** was used as a 260 μM solution in PBS buffer with 50 μM AGP in PBS buffer at 25°C. Triplicate measurements were conducted as well as the reference titration (**2-22f** into PBS buffer). After subtraction of the reference titration from the binding measurements, the data was globally fitted to give the following calculated values:

$K_D = 4.1 \mu\text{M}$ with a confidence interval (95%) from 2.9 - 6.1 μM .

$\Delta K_D = 2.0 \mu\text{M}$

$\Delta H = -4.78 \text{ kcal/mol} = -20.01 \text{ kJ/mol}$ with a confidence interval (95%) from -5.51 - -4.28 kcal/mol.

$\Delta\Delta H = 0.73 \text{ kcal/mol} = 3.05 \text{ kJ/mol}$

2-22 versus AGP:

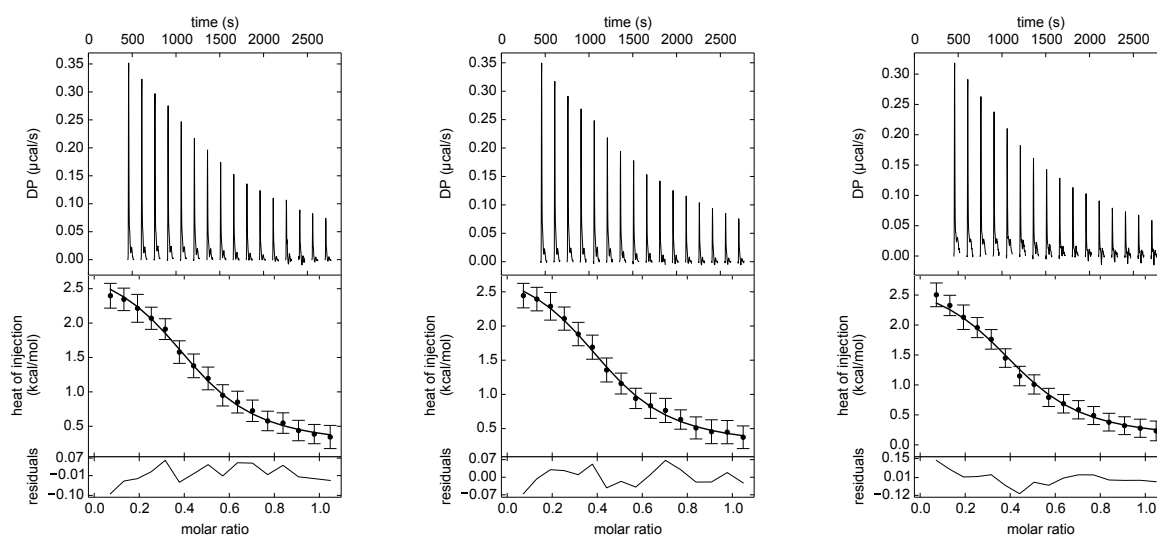


Figure 103. The three ITC titrations of **2-22** versus AGP. All graphs were analyzed and integrated using NITPIC. For the physicochemical calculations all three measurements were fitted and processed using SEDPHAT.

The optimal conditions were found to be 100 μM AGP and 520 μM **2-22** in PBS buffer. Triplicate measurements were performed with a reference titration. Due to the small binding enthalpies the temperature of the assay was lowered to 10°C. After subtraction of the reference titration from the binding measurements, the data was globally fitted to give the following calculated values:

$K_D = 7.0 \mu\text{M}$ with a confidence interval (95%) from 4.7 - 10.7 μM .

$\Delta K_D = 3.7 \mu\text{M}$

$\Delta H = 2.72 \text{ kcal/mol} = 11.39 \text{ kJ/mol}$ with a confidence interval (95%) from 2.43 - 3.14 kcal/mol.

$\Delta\Delta H = 0.42 \text{ kcal/mol} = 1.76 \text{ kJ/mol}$

2-20 and 2-21 versus AGP:

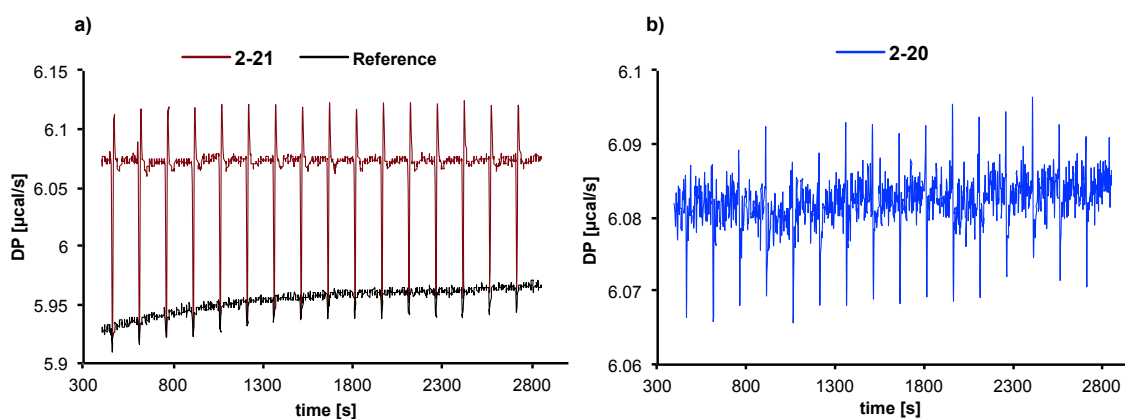


Figure 104. a) ITC titration of **2-21** versus AGP with the reference titration **2-21** vs PBS buffer. The data was not analyzed and integrated. b) ITC titration of **2-20** versus AGP. The data was not analyzed and integrated.

The titrations were performed at 50 μM AGP and 260 μM macrocycle ligand (5% DMSO in PBS for **2-21**, pure PBS for **2-20**) in single measurements at 25°C. A reference titration of **2-21** was conducted to evaluate the solvation enthalpy. The measurements were not further improved nor the data analyzed due to the very weak binding. We assumed no binding for **2-20** due to the very weak differential power (DP) changes that probably uniquely arose from the solvation enthalpy.

2-18, 2-20 and 2-22 versus HSA:

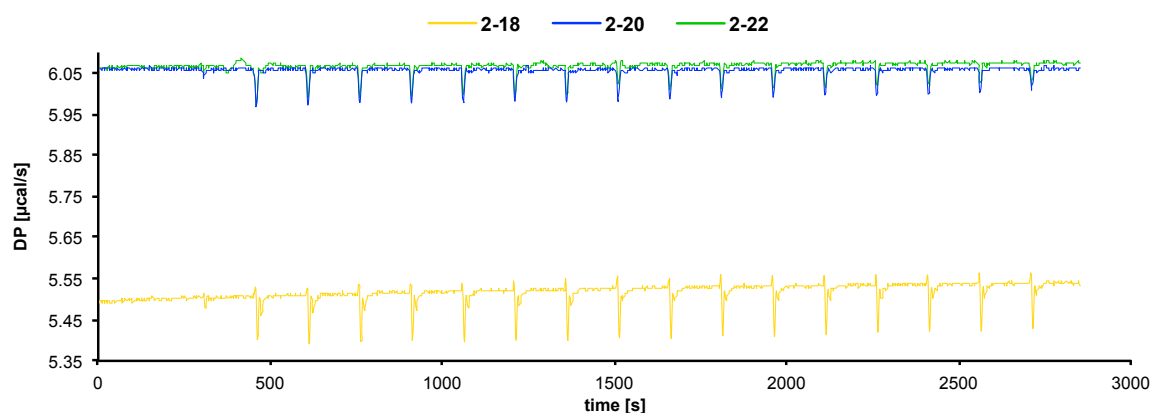


Figure 105. ITC titrations of **2-18**, **2-20** and **2-22** versus HSA. The data was not analyzed and integrated. Reference titrations not shown.

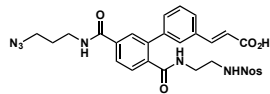
The **2-18** versus HSA assay was performed at 25 μM HSA in 25% DMSO/PBS and 250 μM macrocycle ligand in 25% DMSO/PBS at 25°C. Care was taken to exactly match the DMSO/PBS ratios in all samples and buffers.

2-20 and **2-22** versus HSA measurements were performed at 50 μM /500 μM in PBS buffer at 25°C.

In all three assays no strong binding events were observed. Therefore, no further measurement optimizations and analyses were conducted.

13.14 Building Block Synthesis

Synthesis of NP01



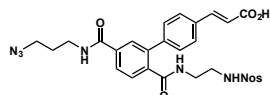
2-4 (43.0 mg, 71.5 μmol , 1.0 eq.), 3-(2-carboxyvinyl)phenylboronic acid (27.5 mg, 143.0 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (7.4 mg, 14.3 μmol , 20 mol%) and K_3PO_4 (45.5 mg, 215.0 μmol , 3.0 eq.) were suspended in EtOH (800 μL) and H_2O (400 μL). The mixture was stirred at 50°C for 1 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (28 mg, **63%**).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ /ppm: 12.43 (s, 1H), 8.65 (t, J = 5.7 Hz, 1H), 8.33 (t, J = 5.9 Hz, 1H), 8.14 (t, J = 6.0 Hz, 1H), 8.00 (dd, J = 5.8, 3.3 Hz, 1H), 7.96 – 7.92 (m, 1H), 7.92 – 7.85 (m, 4H), 7.69 – 7.66 (m, 1H), 7.64 – 7.55 (m, 2H), 7.53 (d, J = 8.4 Hz, 1H), 7.42 – 7.33 (m, 2H), 6.52 (d, J = 16.0 Hz, 1H), 3.41 (t, J = 6.7 Hz, 2H), 3.37 – 3.34 (m, 2H), 3.15 (dt, J = 7.9, 6.1 Hz, 2H), 2.77 (dt, J = 8.2, 6.1 Hz, 2H), 1.79 (p, J = 6.8 Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, $\text{DMSO-}d_6$) δ /ppm: 166.97, 168.24, 165.11, 147.52, 143.29, 139.92, 139.83, 139.56, 138.75, 138.04, 134.89, 133.88, 132.47, 129.89, 129.09, 128.43, 128.17, 127.70, 127.49, 127.04, 126.05, 124.31, 119.37, 48.22, 41.17, 38.46, 36.42, 28.00.

HRMS (ESI): $\text{C}_{28}\text{H}_{27}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 644.1534, *found*: 644.1534.

Synthesis of NP02



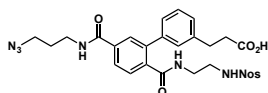
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), 4-(2-carboxyvinyl)phenylboronic acid (12.8 mg, 66.5 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K_3PO_4 (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H_2O (200 μL). The mixture was stirred at 50°C for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (16 mg, **77%**).

^1H NMR (400 MHz, Acetone- d_6) δ /ppm: 8.13 – 8.08 (m, 1H), 8.05 (t, J = 5.7 Hz, 1H), 7.99 – 7.94 (m, 1H), 7.93 – 7.86 (m, 4H), 7.71 – 7.64 (m, 3H), 7.59 (d, J = 8.5 Hz, 1H), 7.55 – 7.45 (m, 3H), 6.75 (t, J = 6.0 Hz, 1H), 6.54 (d, J = 16.1 Hz, 1H), 3.54 – 3.44 (m, 4H), 3.39 (q, J = 6.3 Hz, 2H), 3.13 (q, J = 6.1 Hz, 2H), 1.90 (p, J = 6.7 Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, Acetone- d_6) δ /ppm: 169.71, 167.35, 166.37, 148.91, 144.75, 142.56, 142.47, 139.76, 136.73, 134.80, 134.53, 133.85, 133.43, 131.21, 129.91, 129.19, 128.89, 128.79, 127.04, 125.66, 119.28, 49.60, 43.32, 39.89, 37.60, 29.35.

HRMS (ESI): $\text{C}_{28}\text{H}_{27}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 644.1534, *found*: 644.1534.

Synthesis of NP03



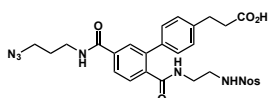
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), 3-(2-carboxyethyl)phenylboronic acid (12.9 mg, 66.5 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K_3PO_4 (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H_2O (200 μL). The mixture was stirred at 50°C for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (15 mg, **72%**).

^1H NMR (500 MHz, Acetone- d_6) δ /ppm: 8.13 – 8.07 (m, 1H), 7.99 – 7.94 (m, 1H), 7.95 – 7.90 (m, 2H), 7.88 (dd, $J = 7.9, 1.8$ Hz, 1H), 7.86 (dd, $J = 1.8, 0.6$ Hz, 1H), 7.56 (dd, $J = 7.8, 0.5$ Hz, 1H), 7.32 – 7.29 (m, 1H), 7.28 (dd, $J = 7.1, 0.7$ Hz, 1H), 7.26 – 7.21 (m, 2H), 3.48 (dt, $J = 15.6, 6.8$ Hz, 4H), 3.33 (t, $J = 6.5$ Hz, 2H), 3.05 (t, $J = 6.5$ Hz, 2H), 2.91 (t, $J = 7.7$ Hz, 2H), 2.61 (t, $J = 7.7$ Hz, 2H), 1.90 (p, $J = 6.7$ Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, Acetone- d_6) δ /ppm: 173.62, 169.79, 166.57, 148.92, 141.85, 141.25, 140.59, 140.40, 139.74, 136.55, 134.74, 133.42, 131.20, 129.31, 129.23, 129.02, 128.74, 128.29, 127.09, 126.57, 125.64, 49.57, 43.17, 39.78, 37.53, 35.51, 31.16, 29.32.

HRMS (ESI): $\text{C}_{28}\text{H}_{29}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 646.1691, *found*: 646.1695.

Synthesis of NP04



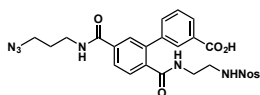
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), 4-(2-carboxyethyl)phenylboronic acid (12.9 mg, 66.5 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K_3PO_4 (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H_2O (200 μL). The mixture was stirred at 50°C for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (14 mg, **68%**).

^1H NMR (500 MHz, Acetone- d_6) δ /ppm: 8.14 – 8.09 (m, 1H), 8.00 – 7.95 (m, 1H), 7.96 – 7.90 (m, 2H), 7.87 (dd, $J = 7.9, 1.8$ Hz, 1H), 7.84 (d, $J = 1.6$ Hz, 1H), 7.55 (d, $J = 7.9$ Hz, 1H), 7.37 – 7.30 (m, 2H), 7.29 – 7.23 (m, 2H), 3.48 (dt, $J = 15.9, 6.8$ Hz, 4H), 3.34 (t, $J = 6.3$ Hz, 2H), 3.09 (t, $J = 6.3$ Hz, 2H), 2.91 (t, $J = 7.7$ Hz, 2H), 2.62 (t, $J = 7.7$ Hz, 2H), 1.90 (p, $J = 6.7$ Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, Acetone- d_6) δ /ppm: 173.60, 169.93, 166.46, 148.91, 141.23, 140.33, 140.22, 139.62, 138.36, 136.49, 134.78, 133.43, 131.22, 129.31, 129.19, 128.97, 128.81, 126.56, 125.66, 49.58, 43.30, 39.84, 37.56, 35.45, 30.89, 29.33.

HRMS (ESI): $\text{C}_{28}\text{H}_{29}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 646.1691, *found*: 646.1692.

Synthesis of NP05



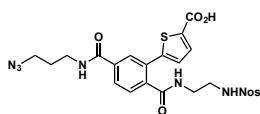
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), 3-carboxyphenylboronic acid (12.9 mg, 77.8 μmol , 2.3 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K_3PO_4 (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H_2O (200 μL). The mixture was stirred at 50°C for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (16 mg, **81%**).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ /ppm: 8.70 (t, $J = 5.7$ Hz, 1H), 8.39 (t, $J = 5.8$ Hz, 1H), 8.13 (t, $J = 6.0$ Hz, 1H), 8.02 – 7.97 (m, 2H), 7.97 – 7.93 (m, 1H), 7.92 – 7.85 (m, 5H), 7.61 (ddd, $J = 7.6, 1.9, 1.2$ Hz, 1H), 7.54 (d, $J = 7.8$ Hz, 1H), 7.46 (t, $J = 7.7$ Hz, 1H), 3.41 (t, $J = 6.7$ Hz, 2H), 3.34 (td, $J = 6.8, 5.5$ Hz, 2H), 3.14 (dt, $J = 7.9, 6.0$ Hz, 2H), 2.82 (dt, $J = 8.3, 6.1$ Hz, 2H), 1.79 (p, $J = 6.8$ Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, $\text{DMSO-}d_6$) δ /ppm: 168.29, 166.84, 165.12, 147.44, 139.64, 139.44, 138.90, 137.93, 135.09, 133.92, 132.57, 132.52, 130.52, 129.16, 129.04, 128.25, 128.14, 128.11, 127.69, 126.23, 124.35, 48.28, 41.25, 38.56, 36.48, 28.05.

HRMS (ESI): $\text{C}_{26}\text{H}_{25}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 618.1378, *found*: 618.1381.

Synthesis of NP06



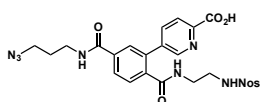
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), 5-carboxythiophene-2-boronic acid pinacol ester (16.9 mg, 66.5 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K_3PO_4 (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H_2O (200 μL). The mixture was stirred at 50°C for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (2 mg, **10%**).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ /ppm: 8.72 (t, $J = 5.6$ Hz, 1H), 8.59 (t, $J = 5.8$ Hz, 1H), 8.18 (t, $J = 6.0$ Hz, 1H), 8.03 – 7.94 (m, 3H), 7.92 – 7.85 (m, 3H), 7.64 (d, $J = 3.9$ Hz, 1H), 7.52 (d, $J = 7.9$ Hz, 1H), 7.26 (d, $J = 3.9$ Hz, 1H), 3.44 – 3.40 (m, 2H), 3.37 – 3.32 (m, 2H), 3.24 (dt, $J = 7.4, 6.1$ Hz, 2H), 2.98 (dt, $J = 7.8, 6.1$ Hz, 2H), 1.80 (p, $J = 6.8$ Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, $\text{DMSO-}d_6$) δ /ppm: 168.00, 164.76, 162.44, 147.44, 146.74, 138.67, 135.06, 134.38, 134.00, 133.23, 132.61, 132.20, 130.16, 129.27, 128.09, 127.96, 127.55, 127.15, 124.38, 48.33, 41.45, 38.75, 36.56, 28.06.

HRMS (ESI): $\text{C}_{24}\text{H}_{22}\text{N}_7\text{O}_8\text{S}_2^-$ *calcd*: 600.0977, *found*: 600.0985.

Synthesis of NP07



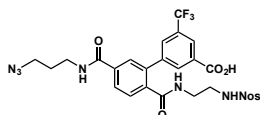
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), 2-methoxycarbonylpyridine-5-boronic acid (16.9 mg, 93.4 μmol , 2.8 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K_3PO_4 (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H_2O (200 μL). The mixture was stirred at 50°C for 2 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the impure methyl ester of the product as beige solid (7.0 mg, 35%). The material was dissolved in MeCN (0.5 mL) and a solution of LiOH monohydrate (4.8 mg, 115.0 μmol , 10.0 eq.) in H_2O (382 μL) was added. The solution was stirred at RT for 2 h, after which UPLC-MS analysis showed full conversion to the desired product. The reaction mixture was purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (6.0 mg, **88%**).

^1H NMR (500 MHz, $\text{DMSO}-d_6$) δ /ppm: 8.75 – 8.66 (m, 2H), 8.57 (t, J = 5.8 Hz, 1H), 8.18 (t, J = 6.0 Hz, 1H), 8.06 (dd, J = 8.0, 0.8 Hz, 1H), 8.02 – 7.92 (m, 5H), 7.91 – 7.84 (m, 2H), 7.64 (d, J = 7.9 Hz, 1H), 3.42 (t, J = 6.8 Hz, 2H), 3.35 (td, J = 6.8, 5.5 Hz, 2H), 3.19 (q, J = 6.4 Hz, 2H), 2.92 (dt, J = 7.5, 6.1 Hz, 2H), 1.80 (p, J = 6.8 Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, $\text{DMSO}-d_6$) δ /ppm: 167.81, 165.73, 165.03, 148.41, 147.44, 146.84, 138.34, 138.27, 136.70, 135.24, 134.97, 133.92, 132.56, 132.30, 129.23, 128.58, 128.06, 127.04, 124.32, 123.98, 48.26, 41.43, 38.76, 36.50, 28.03.

HRMS (ESI): $\text{C}_{25}\text{H}_{23}\text{N}_8\text{O}_8\text{S}^-$ *calcd*: 595.1365, *found*: 595.1371.

Synthesis of NP08



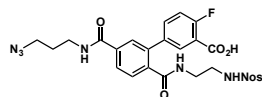
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), 3-methoxycarbonyl-5-trifluoromethylphenylboronic acid pinacol ester (22.0 mg, 66.5 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K_3PO_4 (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H_2O (200 μL). The mixture was stirred at 50°C for 1 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the impure methyl ester of the product (12.0 mg, 53%). The material was dissolved in MeCN (500 μL) and a solution of LiOH monohydrate (7.4 mg, 177.0 μmol , 10.0 eq.) in H_2O (590 μL) was added. The solution was stirred at RT for 2 h, after which UPLC-MS analysis showed full conversion to the desired product. The reaction mixture was purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as white solid (8.0 mg, **68%**).

^1H NMR (500 MHz, $\text{DMSO}-d_6$) δ /ppm: 13.62 (s, 1H), 8.72 (t, J = 5.7 Hz, 1H), 8.51 (t, J = 5.9 Hz, 1H), 8.24 – 8.20 (m, 1H), 8.17 (t, J = 6.0 Hz, 1H), 8.14 – 8.10 (m, 1H), 8.00 – 7.96 (m, 1H), 7.96 – 7.90 (m, 4H), 7.90 – 7.84 (m, 2H), 7.60 (d, J = 7.9 Hz, 1H), 3.42 (t, J = 6.7 Hz, 2H), 3.35 (td, J = 6.9, 5.6 Hz, 2H), 3.14 (dt, J = 8.0, 6.1 Hz, 2H), 2.81 (dt, J = 8.4, 6.0 Hz, 2H), 1.80 (p, J = 6.8 Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, $\text{DMSO-}d_6$) δ /ppm: 167.87, 165.38, 164.91, 147.22, 140.92, 138.64, 136.23, 135.33, 133.85, 132.77, 132.44, 129.00, 128.97, 128.47, 128.05, 127.79, 126.97, 124.30, 124.27, 122.17, 120.00, 48.23, 41.00, 38.61, 36.44, 28.02.

HRMS (ESI): $\text{C}_{27}\text{H}_{24}\text{F}_3\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 686.1251, *found*: 686.1262.

Synthesis of NP09



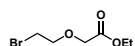
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), 3-carboxy-4-fluorophenylboronic acid (12.2 mg, 66.5 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K_3PO_4 (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H_2O (200 μL). The mixture was stirred at 50°C for 1 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as brownish solid (18.0 mg, **88%**).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ /ppm: 13.32 (s, 1H), 8.70 (t, $J = 5.6$ Hz, 1H), 8.44 (t, $J = 5.8$ Hz, 1H), 8.15 (t, $J = 6.0$ Hz, 1H), 8.01 – 7.95 (m, 2H), 7.92 – 7.83 (m, 5H), 7.61 (ddd, $J = 8.5, 4.5, 2.5$ Hz, 1H), 7.54 (d, $J = 7.9$ Hz, 1H), 7.30 (dd, $J = 10.7, 8.5$ Hz, 1H), 3.41 (t, $J = 6.7$ Hz, 2H), 3.34 (q, $J = 6.6$ Hz, 2H), 3.16 (dt, $J = 7.5, 6.0$ Hz, 2H), 2.88 (dt, $J = 8.1, 6.1$ Hz, 2H), 1.79 (p, $J = 6.8$ Hz, 2H).

^{13}C NMR (126 MHz, $\text{DMSO-}d_6$) δ /ppm: 168.44, 165.38, 164.84, 159.62, 147.66, 138.93, 137.17, 135.84, 135.41, 134.50, 134.11, 132.73, 132.60, 131.72, 129.37, 128.33, 127.94, 126.50, 124.54, 119.07, 116.94, 116.76, 48.54, 41.61, 38.89, 36.75, 28.34.

HRMS (ESI): $\text{C}_{26}\text{H}_{24}\text{FN}_7\text{NaO}_8\text{S}^+$ *calcd*: 636.1283, *found*: 636.1291.

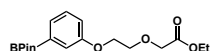
Synthesis of Ethyl 2-bromoethoxyacetate NP10a



Under an inert atmosphere 2-bromoethanol (709.0 μL , 10.0 mmol, 1.0 eq.) was dissolved in DCM (17 mL) and cooled in an ice bath. $\text{Rh}_2(\text{OAc})_4$ (44.2 mg, 100.0 μmol , 1.0 mol%) was added and the mixture was stirred for 5 min. The ice bath was removed and a solution of ethyl 2-diazoacetate (1.2 mL, 9.6 mmol, 0.96 eq.) in DCM (8 mL) was added dropwise while gas evolution occurred. The green mixture was stirred at RT for 2 h. The mixture was filtered over Celite and concentrated *in vacuo*. The crude product was purified by bulb-to-bulb distillation (80°C , 3.6×10^{-1} mbar) to yield the desired product as a colorless liquid (1.3 g, **61%**). Analytical data was in agreement with reported data.^[245]

^1H NMR (400 MHz, CDCl_3) δ /ppm: 4.23 (q, $J = 7.2$ Hz, 2H), 4.15 (s, 2H), 3.89 (t, $J = 6.2$ Hz, 2H), 3.51 (t, $J = 6.2$ Hz, 2H), 1.29 (t, $J = 7.1$ Hz, 3H).

Synthesis of 3-Ethylloxycarbonylmethoxyethoxyphenylboronic acid pinacol ester NP10b



Under an inert atmosphere 3-hydroxyphenylboronic acid pinacol ester (47.0 mg, 214.0 μmol , 1.0 eq.), **NP10a** (54.1 mg, 256.0 μmol , 1.2 eq.) and potassium carbonate (88.5 mg, 641.0 μmol , 3.0 eq.) were mixed in MeCN (1 mL). The suspension was stirred at 65°C for 39 h, after which UPLC-MS analysis showed 78% conversion to the desired product. Potassium carbonate

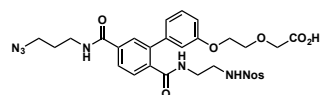
(29.5 mg, 213.4 μmol , 1.0 eq.) and **NP10a** (10.8 mg, 51.2 μmol , 0.2 eq.) were added and the mixture was stirred at 65°C for another 24 h. UPLC-MS analysis showed >95% conversion to the desired product. The solvent was removed by rotary evaporation and the residue was taken up in DCM (5 mL) and H₂O (5 mL). The organic layer was separated and the aqueous layer was extracted with DCM (2 x 5 mL). The combined organic layers were washed with brine (10 mL), dried over Na₂SO₄ and concentrated *in vacuo*. The residue was purified by flash column chromatography (Silica, 9 g, 10:1→5:1 cyclohexane: EtOAc, R_f= 0.29, KMnO₄) to yield the desired product as colorless oil (53 mg, **71%**).

¹H NMR (400 MHz, CDCl₃) δ /ppm: 7.40 (dt, *J* = 7.2, 1.0 Hz, 1H), 7.33 (dd, *J* = 2.8, 0.9 Hz, 1H), 7.28 (dd, *J* = 8.3, 7.3 Hz, 1H), 7.03 (ddd, *J* = 8.2, 2.8, 1.1 Hz, 1H), 4.24 – 4.18 (m, 6H), 3.97 – 3.91 (m, 2H), 1.34 (s, 12H), 1.28 (t, *J* = 7.1 Hz, 3H).

¹³C NMR (101 MHz, CDCl₃) δ /ppm: 170.57, 158.18, 129.11, 127.62, 119.72, 118.53, 83.98, 70.27, 69.09, 67.66, 61.04, 25.01, 14.35.

HRMS (ESI): C₁₈H₂₇BNaO₆⁺ *calcd*: 373.1793, *found*: 373.1793.

Synthesis of NP10



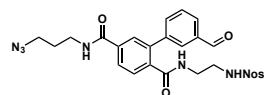
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.), **NP10b** (23.3 mg, 66.5 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (3.5 mg, 6.7 μmol , 20 mol%) and K₃PO₄ (21.2 mg, 99.8 μmol , 3.0 eq.) were suspended in EtOH (400 μL) and H₂O (200 μL). The mixture was stirred at 50°C for 1 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the product as an ester/acid mixture (12.0 mg, 52%). The material was dissolved in MeCN (0.5 mL) and a solution of LiOH monohydrate (7.2 mg, 177.0 μmol , 10.0 eq.) in H₂O (573 μL) was added. The solution was stirred at RT for 2 h, after which UPLC-MS analysis showed full conversion to the desired product. The reaction mixture was purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (8.0 mg, **70%**).

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 8.66 (t, *J* = 5.7 Hz, 1H), 8.24 (t, *J* = 5.9 Hz, 1H), 8.13 (t, *J* = 6.0 Hz, 1H), 8.03 – 7.98 (m, 1H), 7.98 – 7.94 (m, 1H), 7.93 – 7.87 (m, 2H), 7.86 – 7.82 (m, 2H), 7.49 (d, *J* = 8.4 Hz, 1H), 7.20 (t, *J* = 7.9 Hz, 1H), 6.98 (dd, *J* = 2.6, 1.6 Hz, 1H), 6.95 – 6.91 (m, 1H), 6.86 (ddd, *J* = 8.3, 2.6, 0.9 Hz, 1H), 4.12 – 4.06 (m, 4H), 3.82 – 3.79 (m, 2H), 3.41 (t, *J* = 6.7 Hz, 2H), 3.34 (q, *J* = 6.5 Hz, 2H), 3.15 (dt, *J* = 8.0, 6.1 Hz, 2H), 2.79 (dt, *J* = 8.3, 6.1 Hz, 2H), 1.79 (p, *J* = 6.8 Hz, 2H).

¹³C NMR (500 MHz, HMBC/HMQC, DMSO-*d*₆) δ /ppm: 171.35, 168.53, 165.26, 157.94, 147.49, 140.71, 138.87, 138.56, 134.89, 133.95, 132.60, 132.28, 129.20, 128.99, 128.06, 127.52, 125.95, 124.37, 120.64, 114.61, 113.04, 68.76, 67.38, 66.83, 48.29, 41.25, 38.50, 36.48, 28.06.

HRMS (ESI): C₂₉H₃₁N₇NaO₁₀S⁺ *calcd*: 692.1745, *found*: 692.1751.

Synthesis of the Aldehyde Precursor Building Block 2-25



2-4 (53.0 mg, 88.1 μmol , 1.0 eq.), 3-formylphenylboronic acid (26.4 mg, 176.0 μmol , 2.0 eq.), Pd (π -cinnamyl) chloride dimer (9.1 mg, 17.6 μmol , 20 mol%) and K₃PO₄ (56.1 mg, 264.0 μmol ,

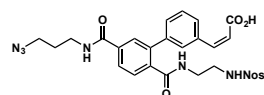
3.0 eq.) were suspended in EtOH (1 mL) and H₂O (500 μL). The mixture was stirred at 50°C for 1 h, after which UPLC-MS analysis showed complete conversion to the desired product. The black mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (34.0 mg, **67%**).

¹H NMR (400 MHz, CDCl₃) δ/ppm: 9.89 (s, 1H), 8.13 – 7.98 (m, 1H), 7.88 – 7.81 (m, 1H), 7.80 – 7.66 (m, 4H), 7.58 (d, *J* = 1.6 Hz, 1H), 7.55 – 7.39 (m, 3H), 7.38 – 7.27 (m, 3H), 6.08 (t, *J* = 5.9 Hz, 1H), 3.50 (q, *J* = 6.4 Hz, 2H), 3.43 (t, *J* = 6.5 Hz, 2H), 3.33 (q, *J* = 5.5 Hz, 2H), 3.06 (t, *J* = 5.6 Hz, 2H), 1.89 (p, *J* = 6.7 Hz, 2H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 192.55, 169.98, 167.53, 148.08, 140.04, 138.72, 137.89, 136.48, 136.01, 134.64, 133.92, 133.36, 132.97, 131.06, 130.20, 129.50, 128.86, 128.79, 128.41, 126.34, 125.44, 49.52, 42.97, 39.97, 38.15, 28.64.

HRMS (ESI): C₂₆H₂₅N₇NaO₇S⁺ *calcd*: 602.1428, *found*: 602.1438.

Synthesis of NP11



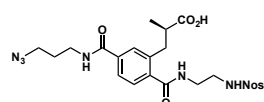
Under an inert atmosphere 18-crown-6 ether (3.0 mg, 11.2 μmol, 1.3 eq.) was dissolved in THF (500 μL) and KHMDS (2.1 mg, 10.4 μmol, 1.2 eq.) was added. The mixture was cooled to -78°C in a dry ice/acetone bath. Methyl 2-[bis(2,2,2-trifluoroethoxy)phosphoryl]acetate (2.0 μL, 9.5 μmol, 1.1 eq.) was added and the mixture was stirred at -78°C for 5 min. **2-25** (5.0 mg, 8.6 μmol, 1.0 eq.) was added and the mixture was stirred at -78°C for 3 h. The cooling bath was removed and the solution was slowly warmed to RT. After a total reaction time of 6 h, UPLC-MS analysis showed full conversion to the methyl ester product. A solution of LiOH monohydrate (3.6 mg, 86.3 μmol, 10.0 eq.) in H₂O (288 μL) was added and the mixture was continued stirring overnight. UPLC-MS analysis showed full conversion to the acid in a 3:1 isomeric ratio. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (2.5 mg, **47%**, 24:1 *cis:trans* mixture).

¹H NMR (500 MHz, DMSO-*d*₆) δ/ppm: 12.51 (s, 1H), 8.66 (t, *J* = 5.6 Hz, 1H), 8.32 (t, *J* = 5.9 Hz, 1H), 8.15 (t, *J* = 6.0 Hz, 1H), 8.03 – 7.99 (m, 1H), 7.98 – 7.94 (m, 1H), 7.92 – 7.88 (m, 2H), 7.87 – 7.83 (m, 2H), 7.59 – 7.55 (m, 2H), 7.51 (d, *J* = 7.8 Hz, 1H), 7.37 – 7.27 (m, 2H), 6.85 (d, *J* = 12.7 Hz, 1H), 5.94 (d, *J* = 12.7 Hz, 1H), 3.45 – 3.41 (m, 2H), 3.34 (q, *J* = 6.2 Hz, 2H), 3.16 (dt, *J* = 7.9, 6.0 Hz, 2H), 2.81 (dt, *J* = 8.2, 6.1 Hz, 2H), 1.79 (p, *J* = 6.8 Hz, 2H).

¹³C NMR (500 MHz, HMBC/HMQC, DMSO-*d*₆) δ/ppm: 168.31, 167.00, 165.26, 147.32, 139.93, 139.23, 138.75, 138.51, 135.11, 134.55, 133.96, 132.59, 130.73, 129.51, 129.19, 128.61, 128.28, 128.18, 127.64, 127.62, 125.90, 124.37, 121.19, 48.30, 41.28, 38.58, 36.50, 28.06.

HRMS (ESI): C₂₈H₂₇N₇NaO₈S⁺ *calcd*: 644.1534, *found*: 644.1540.

Synthesis of NP12



2-4 (20.0 mg, 33.3 μmol, 1.0 eq.) and Pd (π-cinnamyl) chloride dimer (1.7 mg, 3.3 μmol, 10 mol%) were placed in a dry 5 mL Schlenk tube, which was evacuated and backfilled with N₂ three times. THF (400 μL) was added and the yellow solution was cooled to 0°C. A solution of

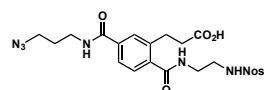
(S)-(3-methoxy-2-methyl-3-oxopropyl)zinc(II) bromide (99.8 μL , 49.9 μmol , 500 mM in THF, 1.5 eq.) was slowly added and the yellow solution was stirred at 0°C for 15 min. The solution was warmed to RT and stirred for another 45 min. UPLC-MS analysis showed full conversion to the methyl ester product. A solution of LiOH monohydrate (7.0 mg, 166.0 μmol , 5.0 eq.) in H₂O (333 μL) was added and the mixture was stirred overnight. UPLC-MS analysis showed complete conversion to the desired product. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (15 mg, **80%**).

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 8.55 (t, *J* = 5.7 Hz, 1H), 8.44 (t, *J* = 5.7 Hz, 1H), 8.20 (t, *J* = 5.9 Hz, 1H), 8.04 – 8.00 (m, 1H), 8.00 – 7.96 (m, 1H), 7.90 – 7.85 (m, 2H), 7.73 – 7.68 (m, 2H), 7.44 – 7.40 (m, 1H), 3.41 (t, *J* = 6.7 Hz, 2H), 3.32 (qd, *J* = 6.2, 3.0 Hz, 4H), 3.11 – 3.01 (m, 3H), 2.76 (dd, *J* = 13.5, 7.8 Hz, 1H), 2.65 (h, *J* = 7.1 Hz, 1H), 1.79 (p, *J* = 6.8 Hz, 2H), 0.97 (d, *J* = 7.0 Hz, 3H).

¹³C NMR (500 MHz, HMBC/HMQC, DMSO-*d*₆) δ /ppm: 176.50, 168.42, 165.38, 147.43, 138.92, 137.58, 137.34, 134.63, 133.88, 132.52, 129.23, 128.87, 127.14, 124.57, 124.28, 48.26, 41.67, 39.92, 38.74, 36.41, 35.68, 28.05, 16.50.

HRMS (ESI): C₂₃H₂₇N₇NaO₈S⁺ *calcd*: 584.1534, *found*: 584.1542.

Synthesis of NP13



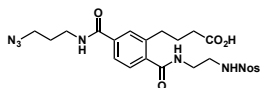
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.) and Pd (π -cinnamyl) chloride dimer (1.7 mg, 3.3 μmol , 10 mol%) were placed in a dry 5 mL Schlenk tube, which was evacuated and backfilled with N₂ three times. THF (400 μL) was added and the yellow solution was cooled to 0°C. A solution of (3-ethoxy-3-oxopropyl)zinc(II) bromide (99.8 μL , 49.9 μmol , 500 mM in THF, 1.5 eq.) was slowly added and the yellow solution was stirred at 0°C for 15 min. The solution was warmed to RT and stirred for another 45 min. UPLC-MS analysis showed full conversion to the ethyl ester product. A solution of LiOH monohydrate (14.0 mg, 333.0 μmol , 10.0 eq.) in H₂O (665 μL) was added and the mixture was stirred at RT for 2 h. UPLC-MS analysis showed complete conversion to the desired product. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (16 mg, **88%**).

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 8.55 (t, *J* = 5.7 Hz, 1H), 8.44 (t, *J* = 5.7 Hz, 1H), 8.22 (t, *J* = 6.0 Hz, 1H), 8.05 – 8.00 (m, 1H), 8.00 – 7.96 (m, 1H), 7.90 – 7.84 (m, 2H), 7.75 (d, *J* = 1.7 Hz, 1H), 7.69 (dd, *J* = 7.9, 1.8 Hz, 1H), 7.41 (d, *J* = 7.9 Hz, 1H), 3.41 (t, *J* = 6.8 Hz, 2H), 3.35 – 3.29 (m, 4H), 3.09 (q, *J* = 6.5 Hz, 2H), 2.91 (dd, *J* = 8.7, 7.1 Hz, 2H), 2.53 (dd, *J* = 8.7, 7.2 Hz, 2H), 1.79 (p, *J* = 6.8 Hz, 2H).

¹³C NMR (500 MHz, HMBC/HMQC, DMSO-*d*₆) δ /ppm: 173.51, 168.47, 165.50, 147.47, 138.64, 138.57, 135.03, 133.96, 132.63, 129.31, 128.31, 127.17, 126.62, 124.53, 124.38, 48.35, 41.79, 38.86, 36.48, 34.97, 28.15, 27.89.

HRMS (ESI): C₂₂H₂₅N₇NaO₈S⁺ *calcd*: 570.1378, *found*: 570.1385.

Synthesis of NP14



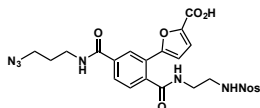
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.) and Pd (π -cinnamyl) chloride dimer (1.7 mg, 3.3 μmol , 10 mol%) were placed in a dry 5 mL Schlenk tube, which was evacuated and backfilled with N_2 three times. THF (400 μL) was added and the yellow solution was cooled to 0°C . A solution of (4-ethoxy-4-oxobutyl)zinc(II) bromide (99.8 μL , 49.9 μmol , 500.0 mM in THF, 1.5 eq.) was slowly added and the yellow solution was stirred at 0°C for 15 min. The solution was warmed to RT and stirred for another 45 min. UPLC-MS analysis showed full conversion to the ethyl ester product. A solution of LiOH monohydrate (14.0 mg, 333.0 μmol , 10.0 eq.) in H_2O (665 μL) was added and the mixture was stirred at RT for 2 h. UPLC-MS analysis showed complete conversion to the desired product. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (7 mg, **38%**).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ /ppm: 8.57 (t, J = 5.7 Hz, 1H), 8.40 (t, J = 5.7 Hz, 1H), 8.20 (t, J = 5.9 Hz, 1H), 8.06 – 8.00 (m, 1H), 8.01 – 7.96 (m, 1H), 7.90 – 7.84 (m, 2H), 7.71 (d, J = 1.7 Hz, 1H), 7.69 (dd, J = 7.9, 1.7 Hz, 1H), 7.38 (d, J = 7.9 Hz, 1H), 3.41 (t, J = 6.7 Hz, 2H), 3.32 (p, J = 6.3 Hz, 4H), 3.08 (q, J = 6.4 Hz, 2H), 2.73 – 2.67 (m, 2H), 2.19 (t, J = 7.5 Hz, 2H), 1.78 (dt, J = 10.2, 7.1 Hz, 4H).

^{13}C NMR (500 MHz, HMBC/HMQC, $\text{DMSO-}d_6$) δ /ppm: 173.93, 168.52, 165.48, 147.44, 139.23, 139.02, 138.80, 134.82, 133.91, 132.56, 129.27, 128.17, 127.01, 124.34, 124.30, 48.31, 41.77, 38.75, 36.44, 33.13, 31.63, 28.11, 25.95.

HRMS (ESI): $\text{C}_{23}\text{H}_{27}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 584.1534, *found*: 584.1533.

Synthesis of NP15



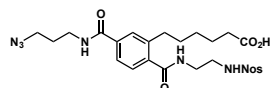
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.) and Pd (π -cinnamyl) chloride dimer (6.7 mg, 13.3 μmol , 40 mol%) were placed in a dry 5 mL Schlenk tube, which was evacuated and backfilled with N_2 three times. THF (400 μL) was added and the yellow solution was cooled to 0°C . A solution of (5-ethoxycarbonyl-2-furyl)zinc(II) bromide (99.8 μL , 49.9 μmol , 500.0 mM in THF, 1.5 eq.) was slowly added and the yellow solution was stirred at 0°C for 15 min. The solution was warmed to RT and stirred for another 45 min. UPLC-MS analysis showed 50% conversion to the ethyl ester product. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a yellow-brown solid (6 mg, 29%). The material was dissolved in MeCN (300 μL) and a solution of LiOH monohydrate (4.1 mg, 97.8 μmol , 10.0 eq.) in H_2O (196 μL) was added and the mixture was stirred at RT for 2 h. UPLC-MS analysis showed complete conversion to the desired product. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (3 mg, **52%**).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ /ppm: 8.74 (t, J = 5.6 Hz, 1H), 8.60 (t, J = 5.7 Hz, 1H), 8.23 – 8.18 (m, 2H), 8.04 – 8.01 (m, 1H), 8.01 – 7.97 (m, 1H), 7.91 – 7.85 (m, 3H), 7.50 (d, J = 7.9 Hz, 1H), 7.25 (d, J = 3.6 Hz, 1H), 6.85 (d, J = 3.6 Hz, 1H), 3.43 (t, J = 6.7 Hz, 2H), 3.34 (dq, J = 19.1, 6.6 Hz, 4H), 3.08 (dt, J = 7.6, 6.2 Hz, 2H), 1.81 (p, J = 6.8 Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, $\text{DMSO-}d_6$) δ /ppm: 168.20, 164.92, 153.68, 147.43, 144.15, 137.36, 135.03, 133.89, 132.57, 129.22, 127.92, 126.99, 126.00, 125.59, 124.30, 119.27, 110.50, 99.05, 48.27, 41.54, 38.86, 36.52, 28.03.

HRMS (ESI): $\text{C}_{24}\text{H}_{23}\text{N}_7\text{NaO}_9\text{S}^+$ *calcd*: 608.1170, *found*: 608.1167.

Synthesis of NP16



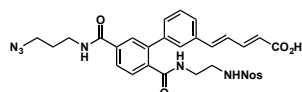
2-4 (20.0 mg, 33.3 μmol , 1.0 eq.) and Pd (π -cinnamyl) chloride dimer (1.7 mg, 3.3 μmol , 10 mol%) were placed in a dry 5 mL Schlenk tube, which was evacuated and backfilled with N_2 three times. THF (400 μL) was added and the yellow solution was cooled to 0°C . A solution of (6-ethoxy-6-oxohexyl)zinc(II) bromide (99.8 μL , 49.9 μmol , 500.0 mM in THF, 1.5 eq.) was slowly added and the yellow solution was stirred at 0°C for 15 min. The solution was warmed to RT and stirred for another 45 min. UPLC-MS analysis showed 16% conversion to the ethyl ester product plus a lot of byproduct. A solution of LiOH monohydrate (14.0 mg, 333.0 μmol , 10.0 eq.) in H_2O (665 μL) was added and the mixture was stirred at RT for 2 h. UPLC-MS analysis showed complete conversion to the desired product. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a brownish solid (3 mg, **15%**).

^1H NMR (500 MHz, $\text{DMSO-}d_6$) δ /ppm: 11.96 (s, 1H), 8.54 (t, $J = 5.7$ Hz, 1H), 8.38 (t, $J = 5.7$ Hz, 1H), 8.20 (t, $J = 6.0$ Hz, 1H), 8.04 – 8.00 (m, 1H), 8.00 – 7.96 (m, 1H), 7.91 – 7.85 (m, 2H), 7.70 (d, $J = 1.7$ Hz, 1H), 7.67 (dd, $J = 7.9, 1.8$ Hz, 1H), 7.37 (d, $J = 7.9$ Hz, 1H), 3.41 (t, $J = 6.7$ Hz, 2H), 3.33 – 3.28 (m, 4H), 3.10 – 3.03 (m, 2H), 2.72 – 2.64 (m, 2H), 2.16 (t, $J = 7.4$ Hz, 2H), 1.78 (p, $J = 6.8$ Hz, 2H), 1.56 – 1.42 (m, 4H), 1.32 – 1.20 (m, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, $\text{DMSO-}d_6$) δ /ppm: 174.21, 168.64, 165.59, 147.46, 139.96, 138.83, 138.74, 134.79, 133.92, 132.55, 129.16, 128.17, 126.95, 124.31, 124.16, 48.32, 41.76, 38.70, 36.44, 33.29, 32.21, 30.45, 28.25, 28.11, 23.99.

HRMS (ESI): $\text{C}_{25}\text{H}_{31}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 612.1847, *found*: 612.1846.

Synthesis of NP17



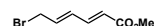
In a dry Schlenk tube triethyl-4-phosphonocrotonate (5.0 μL , 22.4 μmol , 1.3 eq.) was dissolved in THF (400 μL) and cooled to 0°C . $n\text{-BuLi}$ (16.2 μL , 25.9 μmol , 1.5 eq. 1.6 M in hexanes) was added and the yellow solution was stirred for 15 min at 0°C . **2-25** (10.0 mg, 17.3 μmol , 1.0 eq.) was added and the yellow solution was stirred at 0°C for 15 min. The ice bath was removed and the solution was stirred at RT overnight, after which UPLC-MS analysis showed complete conversion of the starting material. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the ethyl ester product as a white solid (5 mg, 43%). The material was dissolved in MeCN (500 μL) and a solution of LiOH monohydrate (6.2 mg, 148.0 μmol , 20.0 eq.) in H_2O (493 μL) was added and the mixture was stirred at RT for 2 h. UPLC-MS analysis showed complete conversion to the desired product. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (2 mg, **42%**).

^1H NMR (500 MHz, DMSO- d_6) δ /ppm: 12.29 (brs, 1 H), 8.66 (t, J = 5.7 Hz, 1H), 8.31 (t, J = 5.9 Hz, 1H), 8.14 (t, J = 6.0 Hz, 1H), 8.02 – 7.97 (m, 1H), 7.96 – 7.92 (m, Hz, 1H), 7.91 – 7.88 (m, 2H), 7.88 – 7.85 (m, 2H), 7.57 – 7.51 (m, 2H), 7.48 (dt, J = 7.2, 1.8 Hz, 1H), 7.38 – 7.28 (m, 3H), 7.14 – 7.00 (m, 2H), 6.01 (d, J = 15.2 Hz, 1H), 3.41 (t, J = 6.8 Hz, 2H), 3.34 – 3.31 (m, 2 H), 3.15 (dt, J = 8.0, 6.1 Hz, 2H), 2.76 (dt, J = 8.4, 6.2 Hz, 2H), 1.79 (p, J = 6.8 Hz, 2H).

^{13}C NMR (500 MHz, HMBC/HMQC, DMSO- d_6) δ /ppm: 168.31, 167.23, 165.19, 147.47, 144.03, 143.89, 139.95, 139.31, 138.81, 138.39, 135.71, 135.00, 133.95, 132.56, 129.17, 128.63, 128.36, 128.19, 127.54, 126.95, 126.79, 126.04, 125.97, 124.38, 122.26, 48.30, 41.24, 38.53, 36.50, 28.07.

HRMS (ESI): $\text{C}_{30}\text{H}_{29}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 670.1691, *found*: 670.1694.

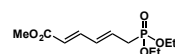
Synthesis of 6-Bromo methyl sorbate **NP18a**^[198]



Methyl sorbate (517.0 μL , 4.0 mmol, 1.0 eq.) and NBS (733.0 mg, 4.1 mmol, 1.0 eq.) were suspended in chlorobenzene (3.4 mL) and stirred at 100°C for 1 h. Benzoylperoxide (86.6 mg, 359.0 μmol , 9 mol%) was added and the mixture was refluxed for 5 h. The reaction was concentrated *in vacuo* and the residue was dissolved in Et₂O (8 mL). The organic layer was washed with aqueous NaOH (5%, 2 mL per wash) until the aqueous layer remained colorless. The organic layer was dried over Na₂SO₄, concentrated by rotary evaporation and the crude was purified by flash column chromatography (Silica, 60 g, cyclohexane:EtOAc, 30:1 \rightarrow 25:1 \rightarrow 20:1, R_f = 0.31, UV) to yield the desired product as yellow oil (254 mg, **31%**). Analytical data was in agreement with reported data.^[246]

^1H NMR (400 MHz, CDCl₃) δ /ppm: 7.29 – 7.23 (m, 1H), 6.42 – 6.35 (m, 1H), 6.28 – 6.20 (m, 1H), 5.94 (d, J = 15.2 Hz, 1H), 4.03 (d, J = 7.7 Hz, 2H), 3.75 (s, 3H).

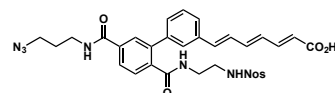
Synthesis of Methyl (2E,4E)-6-(diethoxyphosphoryl)hexa-2,4-dienoate **NP18b**^[198]



NP18a (100.0 mg, 488.0 μmol , 1.0 eq.) was dissolved in toluene (935 μL) and P(OEt)₃ (1.0 mL, 5.9 mmol, 12.0 eq.) was dropwise added. The mixture was refluxed for 2 h, after which the solvent was removed by rotary evaporation. The crude residue was purified by flash column chromatography (Silica, 15 g, petrol ether:EtOAc, 1:1 \rightarrow 1:4 \rightarrow 100% EtOAc, R_f = 0.25, UV) to yield the desired product as a colorless oil (104 mg, **81%**). Analytical data was in agreement with reported data.^[247]

^1H NMR (400 MHz, CDCl₃) δ /ppm: 7.29 – 7.22 (m, 1H), 6.34 – 6.26 (m, 1H), 6.10 – 6.00 (m, 1H), 5.85 (dd, J = 15.4, 2.4 Hz, 1H), 4.14 – 4.06 (m, 4H), 3.74 (s, 3H), 2.71 (dd, J = 23, 7.6 Hz, 2H), 1.31 (t, J = 7.1 Hz, 3H).

Synthesis of **NP18**^[198]



NP18b (10.9 mg, 41.4 μmol , 1.2 eq.) was dissolved in THF (600 μL) and cooled to -78°C. *n*-BuLi (32.4 μL , 51.8 μmol , 1.5 eq. 1.6 M in hexanes) was added and the mixture was stirred at -78°C for 15 min. A solution of **2-25** (20.0 mg, 34.5 μmol , 1.0 eq.) in THF (400 μL) was added and the mixture was stirred at -78°C for 2 h. The mixture was slowly warmed to RT and stirred for another 5.5 h, after which HMPA (14.4 μL , 82.8 μmol , 2.4 eq.) was added. The mixture was stirred at RT

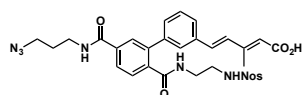
overnight, filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a yellowish solid (14 mg, 59%). The material was dissolved in MeCN (400 μ L) and a solution of LiOH monohydrate (8.5 mg, 203.7 μ mol, 10.0 eq.) in H₂O (310 μ L) was added. The mixture was agitated at RT for 1 h, after which UPLC-MS analysis showed complete conversion to the desired product. The solution was acidified to pH 3 with aqueous HCl. The crude was purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a yellowish solid (2.6 mg, **24%**).

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 12.26 (s, 1H), 8.70 – 8.61 (m, 1H), 8.30 (td, *J* = 5.8, 3.3 Hz, 1H), 8.17 – 8.09 (m, 1H), 8.03 – 7.97 (m, 1H), 7.96 – 7.92 (m, 1H), 7.92 – 7.81 (m, 4H), 7.67 – 7.43 (m, 3H), 7.42 – 7.24 (m, 3H), 7.23 – 6.99 (m, 1H), 6.90 – 6.72 (m, 1H), 6.62 – 6.45 (m, 1H), 6.40 – 6.13 (m, 1H), 5.95 (dd, *J* = 17.3, 15.1 Hz, 1H), 3.41 (t, *J* = 6.7 Hz, 2H), 3.34 – 3.31 (m, 2H), 3.15 (q, *J* = 6.7 Hz, 2H), 2.83 – 2.71 (m, 2H), 1.79 (p, *J* = 6.8 Hz, 2H).

¹³C NMR (500 MHz, HMBC/HMQC, DMSO-*d*₆) δ /ppm: 168.42, 167.36, 165.25, 147.44, 139.95, 138.88, 138.53, 136.81, 136.26, 135.83, 135.01, 133.94, 132.58, 130.62, 129.19, 128.61, 128.22, 128.20, 127.53, 127.08, 127.00, 126.51, 125.99, 125.95, 125.73, 124.38, 121.67, 48.31, 41.27, 38.55, 36.51, 28.07.

HRMS (ESI): C₃₂H₃₁N₇NaO₈S⁺ *calcd*: 696.1847, *found*: 696.1855.

Synthesis of NP19



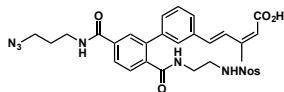
In a dry Schlenk tube triethyl 3-methyl-4-phosphono-2-butenolate (24.2 μ L, 89.8 μ mol, 1.3 eq.) was dissolved in THF (1.6 mL) and cooled to 0°C. n-BuLi (64.8 μ L, 103.6 μ mol, 1.5 eq. 1.6 M in hexanes) was added and the yellow solution was stirred at 0°C for 15 min. **2-25** (40.0 mg, 69.0 μ mol, 1.0 eq.) was added and the yellow solution was stirred at 0°C for 15 min. The ice bath was removed and the solution was stirred at RT for 38 h, after which UPLC-MS analysis showed about 60% conversion to the product. The mixture was filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the two isomeric ethyl ester products as white solids (major isomer, 26 mg, 54%)/(minor isomer, 6 mg, 13%). The major isomer was dissolved in MeCN (1 mL) and a solution of LiOH monohydrate (22.8 mg, 544.0 μ mol, 15.0 eq.) in H₂O (1.1 mL) was added and the mixture was stirred at RT for 22 h. UPLC-MS analysis showed complete conversion to the desired product. The mixture was acidified with conc. HCl, filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (15 mg, **63%**). NOE analysis confirmed the correct structure of the product.

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 12.15 (s, 1H), 8.67 (t, *J* = 5.6 Hz, 1H), 8.30 (t, *J* = 5.8 Hz, 1H), 8.14 (t, *J* = 6.0 Hz, 1H), 8.02 – 7.97 (m, 1H), 7.96 – 7.92 (m, 1H), 7.91 – 7.86 (m, 4H), 7.60 (d, *J* = 1.6 Hz, 1H), 7.54 – 7.50 (m, 2H), 7.34 – 7.27 (m, 2H), 7.08 – 6.98 (m, 2H), 5.96 (d, *J* = 1.4 Hz, 1H), 3.45 – 3.40 (m, 2H), 3.34 (q, *J* = 6.6 Hz, 2H), 3.15 (dt, *J* = 7.9, 6.0 Hz, 2H), 2.78 (dt, *J* = 8.4, 6.1 Hz, 2H), 2.31 (d, *J* = 1.2 Hz, 3H), 1.79 (p, *J* = 6.8 Hz, 2H).

¹³C NMR (500 MHz, HMBC/HMQC, DMSO-*d*₆) δ /ppm: 168.36, 167.51, 165.23, 150.72, 147.47, 139.95, 138.82, 138.51, 135.96, 134.99, 133.95, 133.22, 132.55, 132.08, 129.18, 128.35, 128.34, 128.24, 127.55, 127.00, 126.79, 126.01, 125.87, 124.39, 120.66, 48.32, 41.28, 38.56, 36.53, 28.09, 13.03.

HRMS (ESI): $C_{31}H_{31}N_7NaO_8S^+$ *calcd*: 684.1847, *found*: 684.1859.

Synthesis of NP20



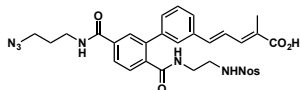
The isolated minor isomer from **NP19** synthesis (5.0 mg, 7.3 μ mol, 1.0 eq.) was dissolved in MeCN (200 μ L) and a solution of LiOH monohydrate (4.6 mg, 109.0 μ mol, 15.0 eq.) in H₂O (217 μ L) was added. The solution was stirred at RT for 22 h, after which UPLC-MS analysis showed full conversion. The solution was acidified with HCl (2 M) and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (3 mg, **63%**). NOE analysis confirmed the correct structure of the product.

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 12.15 (s, 1H), 8.67 (t, *J* = 5.4 Hz, 1H), 8.33 – 8.25 (m, 2H), 8.13 (dt, *J* = 9.9, 5.8 Hz, 1H), 8.03 – 7.97 (m, 1H), 7.93 (dd, *J* = 6.2, 3.2 Hz, 1H), 7.90 – 7.86 (m, 4H), 7.56 – 7.48 (m, 2H), 7.44 (dt, *J* = 7.6, 1.5 Hz, 1H), 7.35 – 7.28 (m, 2H), 7.02 (d, *J* = 16.1 Hz, 1H), 5.76 (s, 1H), 3.44 – 3.39 (m, 3H), 3.36 – 3.33 (m, 2H), 3.17 – 3.12 (m, 2H), 2.74 (dt, *J* = 8.3, 6.1 Hz, 2H), 2.08 (d, *J* = 1.3 Hz, 3H), 1.79 (td, *J* = 6.7, 1.3 Hz, 2H).

¹³C NMR (500 MHz, HMBC/HMQC, DMSO-*d*₆) δ /ppm: 168.28, 166.77, 165.23, 149.53, 147.30, 140.01, 138.82, 138.56, 136.23, 135.02, 134.40, 133.92, 132.52, 129.16, 128.46, 128.43, 128.11, 127.50, 127.34, 127.10, 126.00, 125.69, 125.54, 124.34, 118.65, 48.28, 41.22, 38.50, 36.48, 28.04, 20.18.

HRMS (ESI): $C_{31}H_{31}N_7NaO_8S^+$ *calcd*: 684.1847, *found*: 684.1857.

Synthesis of NP21



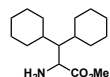
In a dry Schlenk tube 2-methyl-triethyl-4-phosphonocrotonate (44.5 mg, 135.0 μ mol, 1.3 eq.) was dissolved in THF (1 mL) and cooled to 0°C. *n*-BuLi (97.1 μ L, 155.3 μ mol, 1.5 eq. 1.6 M in hexanes) was added and the yellow solution was stirred at 0°C for 15 min. **2-25** (60.0 mg, 104.0 μ mol, 1.0 eq.) was added and the yellow solution was stirred at 0°C for 15 min. The ice bath was removed and the solution was stirred at RT for 25 h, after which UPLC-MS analysis showed about 66% conversion to the product. The solvent was removed by rotary evaporation and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the ethyl ester product as a white solid (41 mg, 57%). The material was dissolved in MeCN (1.5 mL) and a solution of LiOH monohydrate (24.9 mg, 594.0 μ mol, 10.0 eq.) in H₂O (743 μ L) was added and the mixture was stirred at RT for 23 h. UPLC-MS analysis showed complete conversion to the desired product. The mixture was acidified with conc. HCl, filtered and purified by reversed-phase preparative HPLC (Method A). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (30 mg, **76%**).

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 8.67 (t, *J* = 5.7 Hz, 1H), 8.34 (t, *J* = 5.9 Hz, 1H), 8.14 (t, *J* = 6.0 Hz, 1H), 8.01 – 7.97 (m, 1H), 7.96 – 7.92 (m, 1H), 7.91 – 7.85 (m, 4H), 7.62 (s, 1H), 7.57 – 7.49 (m, 2H), 7.33 (t, *J* = 7.5 Hz, 1H), 7.30 (dt, *J* = 7.6, 1.6 Hz, 1H), 7.27 – 7.18 (m, 2H), 6.97 (d, *J* = 13.4 Hz, 1H), 3.41 (t, *J* = 6.7 Hz, 2H), 3.37 – 3.32 (m, 2H), 3.16 (dt, *J* = 7.9, 6.0 Hz, 2H), 2.79 (dt, *J* = 8.4, 6.1 Hz, 2H), 1.98 (d, *J* = 1.2 Hz, 3H), 1.79 (p, *J* = 6.8 Hz, 2H).

^{13}C NMR (126 MHz, $\text{DMSO}-d_6$) δ /ppm: 169.11, 168.68, 165.50, 147.70, 140.20, 139.17, 138.78, 138.28, 137.63, 136.38, 135.22, 134.10, 132.72, 132.57, 129.37, 128.64, 128.57, 128.43, 128.06, 127.71, 126.98, 126.36, 126.19, 124.64, 124.57, 48.57, 41.52, 38.85, 36.78, 28.33, 12.86.

HRMS (ESI): $\text{C}_{31}\text{H}_{31}\text{N}_7\text{NaO}_8\text{S}^+$ *calcd*: 684.1847, *found*: 684.1854.

Synthesis of AA008^[198]



Under an inert atmosphere *rac*-H- β,β -Dicyclohexyl-Ala-OH (100.0 mg, 395.0 μmol , 1.0 eq.) was suspended in MeOH (3 mL) and cooled to 0°C. SOCl_2 (288 μL , 4.0 mmol, 10.0 eq.) was slowly added, whereby a clear solution formed. The solution was stirred in the ice bath for 10 min and was then warmed to RT. After 20 h another portion of SOCl_2 (288.0 μL , 4.0 mmol, 10.0 eq.) was added and the solution continued stirring. After 72 h more SOCl_2 (144.0 μL , 2.0 mmol, 5.0 eq.) was added. After in total 7 d UPLC-MS analysis showed >98% conversion to the product. The volatiles were removed by rotary evaporation and the residue was dissolved in DCM (20 mL) and washed with half-saturated NaHCO_3 solution (3 x 20 mL). The combined aqueous layers were extracted with DCM (2 x 20 mL) and the combined organic layers were washed with brine (20 mL), dried over Na_2SO_4 and concentrated *in vacuo* to yield the desired product as yellowish oil (94 mg, **89%**).

^1H NMR (400 MHz, CD_2Cl_2) δ /ppm: 3.66 (s, 3H), 3.61 (d, $J = 2.6$ Hz, 1H), 1.74 – 1.54 (m, 11H), 1.41 – 1.37 (m, 1H), 1.28 – 1.03 (m, 11H).

^{13}C NMR (101 MHz, CD_2Cl_2) δ /ppm: 178.61, 54.59, 52.37, 52.14, 38.83, 36.88, 32.89, 32.43, 32.42, 30.61, 27.66, 27.65, 27.60, 27.51, 27.32, 27.20.

HRMS (ESI): $\text{C}_{16}\text{H}_{30}\text{NO}_2^+$ *calcd*: 268.2271, *found*: 268.2272.

Synthesis of AA076^[198]



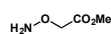
H-tBu-Gly-OH (100.0 mg, 762.0 μmol , 1.0 eq.) was suspended in MeOH (5 mL) under an inert atmosphere and cooled to 0°C. SOCl_2 (1.7 mL, 22.9 mmol, 30.0 eq.) was slowly added, whereby a clear solution was formed. The solution was stirred at 0°C for 10 min and was then heated to 45°C. After 63 h reaction time, more SOCl_2 (278.0 μL , 3.8 mmol, 5.0 eq.) was added. After 110 h reaction time, the volatiles were removed *in vacuo* and the desired product was obtained as a yellow solid (142 mg, **quant.**).

^1H NMR (400 MHz, CD_3OD) δ /ppm: 3.85 (s, 3H), 3.81 (s, 1H), 1.10 (s, 9H).

^{13}C NMR (126 MHz, CD_3OD) δ /ppm: 168.77, 61.53, 51.96, 32.89, 25.33.

HRMS (ESI): $\text{C}_7\text{H}_{16}\text{NO}_2^+$ *calcd*: 146.1176, *found*: 146.1177.

Synthesis of AA077

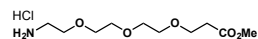


Aminoxyacetic acid (1.0 g, 11.0 mmol, 1.0 eq.) was suspended in MeOH (25 mL) and cooled to 0°C. SOCl_2 (4.0 mL, 54.9 mmol, 5.0 eq.) was slowly added and the solution was stirred at 0°C for 10 min. The cooling bath was removed and the solution was stirred at RT for 20 h. The solvent was removed by rotary evaporation and the residue was dissolved in half-saturated NaHCO_3 solution (25 mL). The mixture was extracted with EtOAc (3 x 25 mL) and the combined organic

layers were washed with brine (25 mL), dried over Na₂SO₄ and concentrated. The crude material was purified by bulb-to-bulb distillation (100°C, 90 mbar). The desired product was isolated as colorless liquid (300 mg, **26%**). Analytical data was in agreement with reported data.^[248]

¹H NMR (400 MHz, CDCl₃) δ/ppm: 4.48 (s, 2H), 3.80 (s, 3H).

Synthesis of AA078



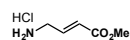
3-[2-[2-(2-aminoethoxy)ethoxy]ethoxy]propanoic acid (118.0 mg, 533.0 μmol, 1.0 eq.) was dissolved in MeOH (3 mL) and cooled in an ice bath. SOCl₂ (389.0 μL, 5.3 mmol, 10.0 eq.) was dropwise added and the mixture was stirred at 0°C for 10 min. The ice bath was removed and the mixture was stirred at RT for 14 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the residue was dissolved in MeOH (5 mL), filtered and concentrated to yield the desired product as yellow oil (121 mg, **84%**).

¹H NMR (500 MHz, CD₃OD) δ/ppm: 3.75 (t, *J* = 6.1 Hz, 2H), 3.73 – 3.70 (m, 2H), 3.69 – 3.67 (m, 3H), 3.67 – 3.65 (m, 3H), 3.66 – 3.59 (m, 5H), 3.15 – 3.11 (m, 2H), 2.60 (t, *J* = 6.1 Hz, 2H).

¹³C NMR (126 MHz, CD₃OD) δ/ppm: 173.91, 71.52, 71.34, 71.20, 71.18, 67.83, 67.60, 52.18, 40.68, 35.62.

HRMS (ESI): C₁₀H₂₂NO₅⁺ *calcd*: 236.1492, *found*: 236.1494.

Synthesis of AA087



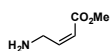
NaH in mineral oil (38.5 mg, 1.0 mmol, 1.6 eq. 60%) was suspended in THF (3 mL) and the mixture was cooled to 0°C. Methyl 2-diethoxyphosphorylacetate (173.0 μL, 942.0 μmol, 1.5 eq.) was added and the solution was stirred at 0°C for 15 min. 2-(*N*-Boc)acetaldehyde (100.0 mg, 628.0 μmol, 1.0 eq.) was added and the mixture was stirred for 15 min. The cooling bath was removed and the solution was stirred at RT for 3.5 h, after which TLC showed full conversion of the starting material. The reaction was quenched by the addition of EtOH (3 mL). The solvent was removed by rotary evaporation and the crude was purified by flash column chromatography (Silica, 37 g, 4:1 cyclohexane:EtOAc, R_f = 0.28, KMnO₄). Product-containing fractions were combined and concentrated to yield the Boc protected intermediate as colorless oil (77 mg, 57%). The material was dissolved in DCM (2 mL) and cooled in an ice bath. HCl in dioxane (895.0 μL, 3.6 mmol, 10.0 eq. 4.0 M) was slowly added and the solution was stirred at 0°C for 10 min. The cooling bath was removed and the mixture was stirred at RT for 16 h. The volatiles were removed by rotary evaporation and the residue was co-evaporated with DCM (5 mL) to yield the desired product as beige solid (47 mg, **86%**).

¹H NMR (500 MHz, DMSO-*d*₆) δ/ppm: 8.28 (brs, 3H), 6.87 (dt, *J* = 16.0, 5.6 Hz, 1H), 6.18 (dt, *J* = 16.0, 1.8 Hz, 1H), 3.71 (s, 3H), 3.70 – 3.66 (m, 2H).

¹³C NMR (126 MHz, DMSO-*d*₆) δ/ppm: 165.36, 140.61, 123.35, 51.71, 40.11.

HRMS (ESI): C₅H₁₀NO₂⁺ *calcd*: 116.0706, *found*: 116.0707.

Synthesis of AA088



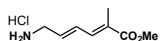
18-crown-6 ether (216.0 mg, 817.0 μmol , 1.3 eq.) was dissolved in THF (3 mL) and KHMDs (150.0 mg, 754.0 μmol , 1.2 eq.) was added. The yellow solution was cooled to -78°C , methyl 2-[bis(2,2,2-trifluoroethoxy)phosphoryl]acetate (147.0 μL , 691.0 μmol , 1.1 eq.) was added and the solution was stirred at 0°C for 5 min. 2-(*N*-Boc)acetaldehyde (100.0 mg, 628.0 μmol , 1.0 eq.) was added and the solution was stirred at -78°C for 1 h. The cooling bath was removed and the solution was stirred at RT for 1 h, after which TLC showed full conversion of the starting material. The reaction was quenched by the addition of EtOH (3 mL). The solvent was removed by rotary evaporation and the crude was purified by flash column chromatography (Silica, 37 g, 5:1 cyclohexane:EtOAc, $R_f = 0.23$, KMnO_4). Product-containing fractions were combined and concentrated to yield the Boc protected intermediate as a beige solid (60 mg, 44%). The material was dissolved in DCM (2 mL) and cooled in an ice bath. HCl in dioxane (696.0 μL , 2.8 mmol, 10.0 eq. 4.0 M) was slowly added and the solution was stirred at 0°C for 10 min. The cooling bath was removed and the mixture was stirred at RT for 15 h. The volatiles were removed by rotary evaporation and the residue was co-evaporated with DCM (5 mL) to yield the desired product as beige solid (41 mg, 96%).

^1H NMR (500 MHz, $\text{DMSO}-d_6$) δ /ppm: 8.21 (brs, 3H), 6.37 (dt, $J = 11.7, 5.9$ Hz, 1H), 6.08 (dt, $J = 11.5, 2.2$ Hz, 1H), 4.00 (s, 2H), 3.69 (s, 3H).

^{13}C NMR (126 MHz, $\text{DMSO}-d_6$) δ /ppm: 165.51, 141.74, 121.89, 51.50, 37.58.

HRMS (ESI): $\text{C}_5\text{H}_{10}\text{NO}_2^+$ *calcd*: 116.0706, *found*: 116.0708.

Synthesis of AA089



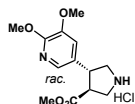
Under an inert atmosphere in a Schlenk tube methyl (*E*)-3-bromo-2-methylprop-2-enoate (250.0 mg, 1.4 mmol, 1.0 eq.), bis-(Boc)-allylamine (719.0 mg, 2.8 mmol, 2.0 eq.), $\text{Pd}(\text{OAc})_2$ (9.4 mg, 41.9 μmol , 3 mol%), tris(*o*-tolyl)phosphine (25.5 mg, 83.8 μmol , 6 mol%) and triethylamine (389.0 μL , 2.8 mmol, 2.0 eq.) were mixed and heated to 100°C and stirred at this temperature for 19 h, after which UPLC-MS analysis showed complete conversion. The black mixture was quenched with NaOH (10%, 10 mL) and extracted with MtBE (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The crude material was purified by flash column chromatography (Silica, 95g, cyclohexane:EtOAc 20:1 \rightarrow 10:1, $R_f = 0.20$, UV/ KMnO_4) to yield the desired product as yellow oil (341 mg, 69%). The material (287.0 mg, 807 μmol , 1.0 eq.) was dissolved in DCM (4 mL) and cooled in an ice bath. HCl in dioxane (4.0 mL, 16.1 mmol, 20.0 eq. 4.0 M in dioxane) was added and the solution was stirred at 0°C for 15 min. The mixture was warmed to RT and stirred for 19 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the residue was dissolved in MeOH (3 mL). The mixture was filtered by syringe filter, cooled in an ice bath to 0°C and Et_2O (5 x 3 mL) was added portionwise. The formed suspension was continued stirring at 0°C for 30 min and the precipitate was filtered off, washed with cold Et_2O (10 mL), dried *in vacuo* to yield the desired product as a white solid (100 mg, 65%).

^1H NMR (400 MHz, CD_3OD) δ /ppm: 7.19 (d, $J = 11.3$ Hz, 1H), 6.82 (ddd, $J = 15.2, 11.3, 1.7$ Hz, 1H), 6.11 (dt, $J = 15.2, 6.7$ Hz, 1H), 3.76 (d, $J = 1.7$ Hz, 3H), 3.71 (d, $J = 6.7$ Hz, 2H), 1.99 (s, 3H).

^{13}C NMR (101 MHz, CD_3OD) δ/ppm : 169.88, 137.43, 132.10, 131.96, 130.75, 52.52, 42.03, 12.93.

HRMS (ESI): $\text{C}_8\text{H}_{14}\text{NO}_2^+$ *calcd*: 156.1019, *found*: 156.1018.

Synthesis of AA106



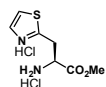
(*Trans-racemic*) 1-*tert*-butyl-3-methyl-4-(5,6-dimethoxypyridin-3-yl)pyrrolidine-1,3-dicarboxylate (250.0 mg, 682.0 μmol , 1.0 eq.) and triethylsilane (1.1 mL, 6.8 mmol, 10.0 eq.) were dissolved in DCM (3 mL) and cooled in an ice bath. HCl in dioxane (1.7 mL, 6.8 mmol, 10.0 eq. 4.0 M in dioxane) was added dropwise over 5 min and the solution was stirred in the ice bath for 10 min. The cooling bath was removed and the solution was stirred at RT for 50 min, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the crude was purified by reversed phase column chromatography on the ISOLERA (RP-Silica, 100 g, $\text{H}_2\text{O}:\text{MeCN}$, UV). Product-containing fractions were combined and lyophilized to yield the desired product as a white, hygroscopic solid (38 mg, **16%**).

^1H NMR (500 MHz, D_2O) δ/ppm : 7.65 (d, $J = 1.9$ Hz, 1H), 7.37 (d, $J = 2.1$ Hz, 1H), 3.97 (s, 3H), 3.91 (s, 3H), 3.87 (ddd, $J = 11.7, 8.5, 5.0$ Hz, 2H), 3.80 – 3.73 (m, 2H), 3.69 (s, 3H), 3.61 – 3.55 (m, 1H), 3.46 (t, $J = 11.4$ Hz, 1H).

^{13}C NMR (126 MHz, D_2O) δ/ppm : 172.81, 153.96, 144.19, 135.21, 126.19, 117.38, 55.71, 53.91, 52.95, 50.82, 48.60, 47.05, 44.36.

HRMS (ESI): $\text{C}_{13}\text{H}_{19}\text{N}_2\text{O}_4^+$ *calcd*: 267.1339, *found*: 267.1338.

Synthesis of AA107



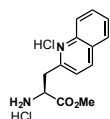
H- β -(2-thiazolyl)-Ala-OH (88.0 mg, 511.0 μmol , 1.0 eq.) was suspended in MeOH (4.5 mL) and cooled in an ice bath. SOCl_2 (373.0 μL , 5.1 mmol, 10.0 eq.) was dropwise added and the mixture was stirred at 0°C for 10 min. The ice bath was removed and the mixture was stirred at RT for 23 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the residue was dissolved in MeOH (10 mL), filtered over a G4 glass sintered funnel and cooled in an ice bath. Et_2O (15 mL) was slowly added and the mixture was stirred in the ice bath for 30 min. The formed precipitate was filtered off, washed with cold Et_2O (3 x 5 mL) and dried *in vacuo* to yield the desired product as a white solid (104 mg, **79%**).

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ/ppm : 8.75 (s, 4H), 7.77 (d, $J = 3.3$ Hz, 1H), 7.70 (d, $J = 3.3$ Hz, 1H), 4.56 – 4.48 (m, 1H), 3.69 (s, 3H), 3.63 (d, $J = 5.8$ Hz, 2H).

^{13}C NMR (101 MHz, $\text{DMSO}-d_6$) δ/ppm : 168.68, 163.23, 142.32, 120.87, 52.85, 51.27, 32.23.

HRMS (ESI): $\text{C}_7\text{H}_{11}\text{N}_2\text{O}_2\text{S}^+$ *calcd*: 187.0536, *found*: 187.0533.

Synthesis of AA108



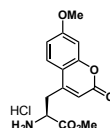
H- β -(2-quinoly)-Ala-OH (100.0 mg, 462.0 μ mol, 1.0 eq.) was suspended in MeOH (3 mL) and cooled in an ice bath. SOCl₂ (337.0 μ L, 4.6 mmol, 10.0 eq.) was dropwise added and the mixture was stirred at 0°C for 10 min. The ice bath was removed and the mixture was stirred at RT for 20 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation, the residue was dissolved in MeOH (2.5 mL), filtered by a syringe filter and cooled in an ice bath. Et₂O (3 mL) was slowly added and the mixture was stirred in the ice bath for 30 min. The formed precipitate was filtered off, washed with cold Et₂O (3 x 5 mL) and dried *in vacuo* to yield the desired product as a white solid (103 mg, **74%**).

¹H NMR (400 MHz, CD₃OD) δ /ppm: 9.08 (d, *J* = 8.6 Hz, 1H), 8.33 (ddd, *J* = 13.9, 8.5, 1.1 Hz, 2H), 8.16 (ddd, *J* = 8.6, 7.0, 1.4 Hz, 1H), 8.07 (d, *J* = 8.6 Hz, 1H), 7.96 (ddd, *J* = 8.1, 6.9, 1.1 Hz, 1H), 4.90 (t, *J* = 7.3 Hz, 1H), 3.93 (d, *J* = 7.3 Hz, 2H), 3.79 (s, 3H).

¹³C NMR (126 MHz, CD₃OD) δ /ppm: 169.07, 155.65, 147.49, 140.56, 136.04, 131.03, 130.29, 129.50, 124.15, 122.28, 54.25, 52.84, 35.65.

HRMS (ESI): C₁₃H₁₅N₂O₂⁺ *calcd*: 231.1128, *found*: 231.1131.

Synthesis of AA109



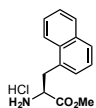
H- β -(7-methoxycoumarin-4-yl)-Ala-OH (87.0 mg, 330.0 μ mol, 1.0 eq.) was suspended in MeOH (3.5 mL) and cooled in an ice bath. SOCl₂ (241.0 μ L, 3.3 mmol, 10.0 eq.) was dropwise added and the mixture was stirred at 0°C for 10 min. The ice bath was removed and the mixture was stirred at RT for 21 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the residue was dissolved in MeOH (2.5 mL) and cooled in an ice bath. Et₂O (5 mL) was slowly added and the mixture was stirred in the ice bath for 30 min. The formed precipitate was filtered off, washed with cold Et₂O (3 x 5 mL) and dried *in vacuo* to yield the desired product as a white solid (82 mg, **79%**).

¹H NMR (250 MHz, DMSO-*d*₆) δ /ppm: 8.55 (s, 3H), 7.69 (d, *J* = 8.8 Hz, 1H), 7.06 (d, *J* = 2.5 Hz, 1H), 7.02 (dd, *J* = 8.7, 2.6 Hz, 1H), 6.32 (s, 1H), 4.37 (t, *J* = 7.2 Hz, 1H), 3.87 (s, 3H), 3.72 (s, 3H), 3.43 – 3.34 (m, 1H), 3.24 (dd, *J* = 14.4, 8.2 Hz, 1H).

¹³C NMR (126 MHz, DMSO-*d*₆) δ /ppm: 168.80, 162.52, 159.75, 155.21, 149.32, 125.87, 113.83, 112.29, 111.76, 101.19, 56.02, 52.94, 50.95, 31.68.

HRMS (ESI): C₁₄H₁₆NO₅⁺ *calcd*: 278.1023, *found*: 278.1024.

Synthesis of AA110



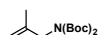
H-1-Nal-OH (100.0 mg, 465.0 μmol , 1.0 eq.) was dissolved in MeOH (2 mL) and cooled in an ice bath. SOCl_2 (74.6 μL , 1.0 mmol, 2.2 eq.) was added dropwise and the mixture was stirred at 0°C for 10 min. The ice bath was removed and the mixture was stirred at RT for 26 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the residue was dissolved in MeOH (500 μL) and cooled in an ice bath. Et_2O (4 mL) was slowly added and the mixture was stirred in the ice bath for 30 min. The formed precipitate was filtered off, washed with cold Et_2O (3 x 5 mL) and dried *in vacuo* to yield the desired product as a white solid (92 mg, **75%**).

^1H NMR (400 MHz, D_2O) δ /ppm: 8.11 (d, J = 8.4 Hz, 1H), 8.06 (d, J = 8.0 Hz, 1H), 7.99 (d, J = 8.2 Hz, 1H), 7.71 (t, J = 7.6 Hz, 1H), 7.66 (t, J = 7.4 Hz, 1H), 7.56 (t, J = 7.6 Hz, 1H), 7.50 (d, J = 7.0 Hz, 1H), 4.55 (dd, J = 8.6, 6.3 Hz, 1H), 3.94 (dd, J = 14.7, 6.4 Hz, 1H), 3.76 (s, 3H), 3.63 (dd, J = 14.6, 8.5 Hz, 1H).

^{13}C NMR (126 MHz, D_2O) δ /ppm: 170.55, 134.17, 131.51, 130.29, 129.57, 129.28, 128.86, 127.49, 126.85, 126.20, 123.19, 53.95, 53.77, 33.49.

HRMS (ESI): $\text{C}_{14}\text{H}_{16}\text{NO}_2^+$ *calcd*: 230.1176, *found*: 230.1176.

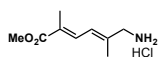
Synthesis of *N,N*-bis(Boc)-2-methyl-prop-2-enamine AA124a



NaH (212.0 mg, 5.5 mmol, 1.5 eq. 60% in mineral oil) was suspended in DMF (6 mL) and cooled to 0°C . A solution of Boc_2NH (800.0 mg, 3.7 mmol, 1.0 eq.) in DMF (6 mL) was added dropwise. The cooling bath was removed and the suspension was stirred at RT for 45 min. 3-Bromo-2-methylpropene (483.0 μL , 4.8 mmol, 1.3 eq.) was added and the solution was stirred at RT for 4 h. The reaction was quenched by the addition of H_2O (1 mL). The solvent was removed by rotary evaporation and the residue was dissolved in Et_2O (25 mL) and H_2O (20 mL). The layers were separated and the aqueous layer was extracted with Et_2O (25 mL). The combined organic layers were washed with brine (2 x 15 mL), dried over Na_2SO_4 and concentrated. The crude material was purified by flash column chromatography (Silica, 56 g, cyclohexane:EtOAc 20:1, R_f = 0.23, KMnO_4). Product-containing fractions were combined and concentrated *in vacuo* to yield the desired product as a white solid (864 mg, **87%**). Analytical data was in agreement with reported data.^[249]

^1H NMR (400 MHz, CDCl_3) δ /ppm: 4.82 (p, J = 1.5 Hz, 1H), 4.75 (qd, J = 1.6, 0.8 Hz, 1H), 4.12 (q, J = 0.8 Hz, 2H), 1.72 (t, J = 1.2 Hz, 3H), 1.49 (s, 18H).

Synthesis of AA124



Under an inert atmosphere in a Schlenk tube methyl (*E*)-3-bromo-2-methyl-prop-2-enoate (422.0 mg, 2.4 mmol, 1.6 eq.), **AA124a** (400.0 mg, 1.5 mmol, 1.0 eq.), $\text{Pd}(\text{OAc})_2$ (9.9 mg, 44.2 μmol , 3 mol%), tri(*o*-tolyl)phosphine (26.9 mg, 88.4 μmol , 6 mol%) and NEt_3 (411.0 μL , 3.0 mmol, 2.0 eq.) were mixed and heated to 100°C . After 19 h UPLC-MS showed full conversion to the desired product. The reaction was quenched with 10% NaOH (10 mL). The mixture was extracted with MtBE (3 x 10 mL) and the combined organic layers were washed with brine (10 mL), dried over Na_2SO_4 and concentrated by rotary evaporation. The crude was purified by

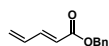
flash column chromatography (Silica, 67 g, 20:1→10:1 cyclohexane:EtOAc, R_f = 0.22, UV/KMnO₄). Product-containing fractions were combined and concentrated *in vacuo* to yield the desired product as a yellow waxy solid (351 mg, 64%). The material was dissolved in DCM (4 mL) and cooled in an ice bath. HCl in dioxane (4.7 mL, 18.8 mmol, 20.0 eq. 4.0 M in dioxane) was added and the solution was stirred at 0°C for 15 min. The mixture was warmed to RT and stirred for 6 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the residue was dissolved in MeOH (2 mL). The mixture was filtered by syringe filter, cooled in an ice bath to 0°C and Et₂O (5 x 2 mL) was added portionwise. The formed suspension was continued stirring at 0°C for 30 min and the precipitate was filtered off, washed with cold Et₂O (10 mL) and dried *in vacuo* to yield the desired product as a white solid (104 mg, **54%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 8.34 (s, 3H), 7.34 (dd, J = 11.6, 1.7 Hz, 1H), 6.50 (dq, J = 11.8, 1.6 Hz, 1H), 3.70 (s, 3H), 3.55 (s, 2H), 1.93 (s, 3H), 1.90 (s, 3H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 167.91, 139.00, 132.36, 127.46, 122.86, 51.84, 45.17, 15.64, 12.65.

HRMS (ESI): C₉H₁₆NO₂⁺ *calcd*: 170.1176, *found*: 170.1175.

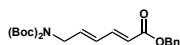
Synthesis of Benzyl (E)-penta-2,4-dienoate **AA125a**^[198]



Pentadienoic acid (200.0 mg, 2.0 mmol, 1.0 eq.) was suspended in DCM (3 mL) and benzylic alcohol (422.0 μL, 4.1 mmol, 2.0 eq.) was added. The mixture was cooled in an ice bath and EDC hydrochloride (782.0 mg, 4.1 mmol, 2.0 eq.) and DMAP (167.0 mg, 1.4 mmol, 0.7 eq.) were added. The mixture was warmed to RT and stirred for 5.5 h. HCl solution (1.0 M) was added to acidify the mixture and the phases were separated. The aqueous phase was extracted with DCM (2 x 8 mL) and the combined organic layers were washed with saturated NaHCO₃ solution, dried over Na₂SO₄ and concentrated by rotary evaporation. The crude was purified by column chromatography (Silica, 50 g, cyclohexane:EtOAc 1:1) to yield the desired product as a colorless liquid (271 mg, **71%**). Analytical data was in agreement with reported data.^[250]

¹H NMR (400 MHz, CDCl₃) δ/ppm: 7.41 – 7.27 (m, 6H), 6.46 (dt, J = 17.1, 10.1 Hz, 1H), 5.96 (d, J = 15.6 Hz 1H), 5.62 (d, J = 17.1 Hz, 1H), 5.50 (d, J = 10.1 Hz, 1H), 5.20 (s, 2H).

Synthesis of Benzyl (2E, 4E)-6-(bis-(Boc)amino)hexa-2,4-dienoate **AA125b**^[198]



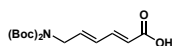
AA125a (100.0 mg, 531.0 μmol, 1.0 eq.) and bis-Boc allylamine (342.0 mg, 1.3 mmol, 2.5 eq.) were dissolved in DCM (1.5 mL). A solution of the Hoveyda-Grubbs catalyst (23.3 mg, 37.2 μmol, 7.0 mol%) in DCM (1.2 mL) was added and the reaction mixture was stirred at 40°C for 16 h. The reaction mixture was concentrated by rotary evaporation and purified by preparative reversed phase HPLC (Method A, no TFA) to yield the desired product as a greenish solid (70 mg, **32%**).

¹H NMR (400 MHz, CD₃CN) δ/ppm: 7.45 – 7.62 (m, 6H), 6.33 – 6.625 (m, 1H), 6.23 – 6.14 (m, 1H), 5.97 (d, J = 15.4 Hz, 1H), 5.16 (s, 2H), 4.25 (d, J = 5.5 Hz, 2H), 1.46 (s, 18H).

¹³C NMR (101MHz, CD₃CN) δ/ppm: 167.2, 153.1, 145.0, 139.8, 137.6, 129.7, 129.5, 129.0, 129.0, 121.8, 83.2, 66.7, 48.3, 28.2.

HRMS (ESI): C₂₃H₃₁NNaO₆⁺ *calcd*: 440.2044, *found*: 440.2051.

Synthesis of (2E, 4E)-6-(bis-(Boc)amino)hexa-2,4-dienoic acid **AA125c**^[198]



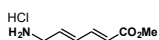
AA125b (60.0 mg, 144.0 μmol , 1.0 eq.) was dissolved in H_2O (1.2 mL) and MeCN (1.2 mL) and LiOH monohydrate (90.5 mg, 2.2 mmol, 15.0 eq.) was added. The mixture was stirred at RT for 18 h. The mixture was purified by preparative reversed phase HPLC (Method A, no TFA) to yield the desired product as a brownish solid (32 mg, **68%**).

^1H NMR (500 MHz, DMSO- d_6) δ /ppm: 6.71 – 6.63 (m, 1H), 6.18 – 6.10 (m, 1H), 5.78 – 5.70 (m, 2H), 4.11 (d, J = 6.4 Hz, 2H), 1.43 (s, 18H).

^{13}C NMR (126 MHz, DMSO- d_6) δ /ppm: 169.4, 151.8, 134.7, 134.4, 131.8, 130.1, 81.8, 47.4, 27.6.

HRMS (ESI): $\text{C}_{16}\text{H}_{25}\text{NNaO}_6^+$ *calcd*: 350.1574, *found*: 350.1576.

Synthesis of **AA0125**^[198]



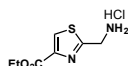
AA125c (32.0 mg, 97.7 μmol , 1.0 eq.) was suspended in MeOH (5.3 mL) and cooled in an ice bath. SOCl_2 (71.3 μL , 97.7 μmol , 10.0 eq.) was added and the mixture was stirred for 15 min at 0°C . The cooling bath was removed and the mixture was stirred at RT for 3 h. The volatiles were removed by rotary evaporation and the crude was recrystallized from MeOH/ Et_2O to yield the desired product as a brown solid (10 mg, **58%**).

^1H NMR (400 MHz, CD_3OD) δ /ppm: 7.31 (dd, J = 15.4, 10.9 Hz, 1H), 6.59 (dd, J = 15.4, 11.0 Hz, 1H), 6.19 (dt, J = 15.4, 6.4 Hz, 1H), 6.06 (d, J = 15.4 Hz, 1H), 3.75 (s, 3H), 3.69 (d, J = 6.4 Hz, 2H).

^{13}C NMR (126 MHz, CD_3OD) δ /ppm: 169.5, 144.0, 134.5, 133.5, 124.2, 52.2, 41.8.

HRMS (ESI): $\text{C}_7\text{H}_{12}\text{NO}_2^+$ *calcd*: 142.0863, *found*: 142.0863.

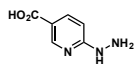
Synthesis of **AA0126**



Ethyl 2-(Boc-aminomethyl)thiazole-4-carboxylate (100.0 mg, 349.0 μmol , 1.0 eq.) was dissolved in DCM (1.2 mL) and cooled in an ice bath. HCl in dioxane (1.7 mL, 7.0 mmol, 20.0 eq. 4.0 M) was added and the mixture was stirred at 0°C for 15 min. The ice bath was removed and the mixture was stirred at RT for 2 h, after which UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the residue was dissolved in MeOH (1 mL) and cooled in an ice bath. Et_2O (6 mL) was slowly added and the mixture was stirred in the ice bath for 30 min. The formed precipitate was filtered off, washed with cold Et_2O (3 x 5 mL) and dried *in vacuo* to yield the desired product as a brownish solid (51 mg, **66%**). Analytical data was in agreement with reported data.^[251]

^1H NMR (400 MHz, DMSO- d_6) δ /ppm: 8.66 (s, 3H), 8.61 (s, 1H), 4.48 (s, 2H), 4.32 (q, J = 7.1 Hz, 2H), 1.31 (t, J = 7.1 Hz, 3H).

Synthesis of 6-hydrazinopyridine-3-carboxylic acid **TA229a**

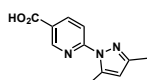


In a 10 mL screw-capped vial 6-chloropyridine-3-carboxylic acid (1.0 g, 6.4 mmol, 1.0 eq.) was mixed with hydrazine hydrate (6.7 mL, 88.9 mmol, 14.0 eq. 64% in H_2O) and the solution was

stirred at 100°C for 17 h, after which UPLC-MS analysis showed full conversion to the desired product. The mixture was evaporated to dryness and the white residue was dissolved in H₂O (20 mL). The mixture was acidified to pH 5.5 by addition of conc. HCl whereby precipitation occurred. The yellow suspension was stirred at RT for 15 min and the precipitate was filtered off. The yellow solid was washed with EtOH (20 mL), Et₂O (20 mL) and dried *in vacuo* to yield the desired product as yellow solid (774 mg, **80%**). Analytical data was in agreement with reported data.^[252]

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 8.53 (d, *J* = 2.1 Hz, 1H), 8.08 (dd, *J* = 9.0, 2.1 Hz, 1H), 6.93 (d, *J* = 9.0 Hz, 1H).

Synthesis of 6-(3,5-dimethyl-1H-pyrazol-1-yl)nicotinic acid **TA229b**



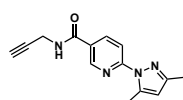
In a 50 mL flask **TA229a** (749.0 mg, 4.9 mmol, 1.0 eq.) was suspended in H₂O (35 mL) and acetylacetone (600 μL, 5.9 mmol, 1.2 eq.) and acetic acid (839.0 μL, 14.7 mmol, 3.0 eq.) were added subsequently. The mixture was stirred at RT for 1 h, after which UPLC-MS showed full conversion to the desired product. The mixture was diluted with EtOAc (150 mL) and H₂O (100 mL). The biphasic mixture was acidified with conc. HCl to pH 1. The layers were separated and the aqueous layer was extracted with EtOAc (2 x 100 mL). The combined organic layers were washed with 1% aqueous HCl (2 x 100 mL), brine (100 mL), dried over Na₂SO₄ and was concentrated *in vacuo*. The desired product was obtained as small white needles (966 mg, **91%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 13.32 (s, 1H), 8.90 (dd, *J* = 2.3, 0.8 Hz, 1H), 8.34 (dd, *J* = 8.7, 2.3 Hz, 1H), 7.93 (dd, *J* = 8.7, 0.8 Hz, 1H), 6.16 (s, 1H), 2.62 (s, 3H), 2.21 (s, 3H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 165.77, 155.42, 150.14, 149.07, 141.75, 139.65, 123.31, 114.23, 110.17, 14.78, 13.39.

HRMS (ESI): C₁₁H₁₂N₃O₂⁺ *calcd*: 218.0924, *found*: 218.0924.

Synthesis of **TA229**



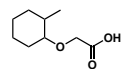
In a 50 mL flask **TA229b** (500.0 mg, 2.3 mmol, 1.0 eq.) was dissolved in THF (18 mL) and H₂O (2 mL) and propargylamine (177.0 μL, 2.8 mmol, 1.2 eq.) was added, followed by DIPEA (2.0 mL, 11.5 mmol, 5.0 eq.), HOAt (407.0 mg, 3.0 mmol, 1.3 eq.) and EDC hydrochloride (574.0 mg, 3.0 mmol, 1.3 eq.). The yellow mixture was stirred at RT 17 h. The volatiles were removed by rotary evaporation and the residue was dissolved in EtOAc (50 mL) and H₂O (50 mL). The mixture was acidified with conc. HCl to pH 1. The layers were separated and the aqueous layer was extracted with EtOAc (2 x 50 mL). The combined organic layers were washed with 1% aqueous HCl (2 x 50 mL), half-saturated bicarbonate solution (3 x 50 mL), brine (50 mL), dried over Na₂SO₄ and concentrated *in vacuo* to yield the desired product as white solid (477 mg, **82%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 9.12 (t, *J* = 5.5 Hz, 1H), 8.87 (dd, *J* = 2.4, 0.8 Hz, 1H), 8.33 (dd, *J* = 8.7, 2.4 Hz, 1H), 7.91 (dd, *J* = 8.7, 0.8 Hz, 1H), 6.16 (s, 1H), 4.09 (dd, *J* = 5.5, 2.5 Hz, 2H), 3.16 (t, *J* = 2.5 Hz, 1H), 2.62 (d, *J* = 0.9 Hz, 3H), 2.21 (s, 3H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 163.90, 154.58, 149.84, 147.02, 141.47, 137.90, 126.12, 114.15, 109.90, 80.96, 73.12, 28.49, 14.65, 13.38.

HRMS (ESI): C₁₄H₁₄N₄NaO⁺ *calcd*: 277.1060, *found*: 277.1063.

Synthesis of (2-methylcyclohexyl)oxyacetic acid **TA256a**



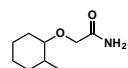
In a dry 100 mL flask under an inert atmosphere 2-methylcyclohexanol (5.7 mL, 46.6 mmol, 2.2 eq.) was dissolved in THF (30 mL) and lithium (375.0 mg, 54.0 mmol, 2.6 eq. in pellets) was added. The mixture was refluxed for 16 h, after which the suspension was filtered and a solution of 2-chloroacetic acid (2.0 g, 21.2 mmol, 1.0 eq.) in THF (10 mL) was slowly added. The solution was refluxed for 6 h, whereafter H₂O (35 mL) was added and the volatiles were removed by rotary evaporation. The residue was dissolved in H₂O (30 mL) and extracted with IPE (3 x 30 mL) to remove residual alcohol. The combined IPE layers were washed with H₂O (30 mL) and the combined aqueous layers were acidified to pH 1-2 with conc. HCl and extracted with Et₂O (3 x 30 mL). The combined Et₂O layers were washed with H₂O (30 mL), brine (30 mL), dried over Na₂SO₄ and were concentrated *in vacuo* to yield the desired product as yellow oil (3.3 g, **89%**).

¹H NMR (400 MHz, CDCl₃) δ/ppm: 9.01 (sbr, 1H), 4.24 – 4.03 (m, 2H), 2.94 (td, *J* = 9.7, 4.1 Hz, 1H), 2.10 – 1.95 (m, 1H), 1.88 – 1.66 (m, 2H), 1.65 – 1.55 (m, 1H), 1.55 – 1.42 (m, 1H), 1.31 – 1.09 (m, 3H), 1.08 – 0.82 (m, 4H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 174.27, 85.71, 65.89, 38.13, 33.78, 30.95, 25.40, 24.88, 18.90.

HRMS (ESI): C₉H₁₆NaO₃⁺ *calcd*: 195.0992, *found*: 195.0989.

Synthesis of 2-((2-methylcyclohexyl)oxy)acetamide **TA256b**



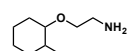
In a 50 mL flask **TA256a** (3.0 g, 17.4 mmol, 1.0 eq.) was dissolved in SOCl₂ (6.4 mL, 87.1 mmol, 5.0 eq.) and the solution was stirred at 50°C for 2 h, after which TLC showed full consumption of the SM. The generated gas during the reaction was washed and neutralized with conc. NaOH solution. Residual SOCl₂ was removed by distillation (60°C, water-jet vacuum). In a 50 mL flask the residue was slowly added to aqueous ammonia (7.2 mL, 122.0 mmol, 7.0 eq. 32%) while cooling in an ice bath. A white precipitate immediately formed. The mixture was stirred for 15 h while slowly warming to RT. The mixture was suspended in H₂O (100 mL) and extracted with Et₂O (3 x 100 mL). The combined ethereal layers were washed with H₂O (100 mL), brine (100 mL), dried over Na₂SO₄ and were concentrated *in vacuo* to yield the desired product as a white solid (2.3 g, **76%**).

¹H NMR (400 MHz, CDCl₃) δ/ppm: 6.60 (s, 1H), 6.20 (s, 1H), 4.05 (d, *J* = 15.5 Hz, 1H), 3.86 (d, *J* = 15.6 Hz, 1H), 2.87 (td, *J* = 9.6, 3.8 Hz, 1H), 2.13 – 1.91 (m, 1H), 1.87 – 1.67 (m, 2H), 1.64 – 1.54 (m, 1H), 1.52 – 1.40 (m, 1H), 1.31 – 1.07 (m, 3H), 1.06 – 0.89 (m, 4H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 173.93, 85.30, 67.89, 38.16, 33.71, 31.02, 25.43, 24.82, 19.00.

HRMS (ESI): C₉H₁₇NNaO₂⁺ *calcd*: 194.1151, *found*: 194.1154.

Synthesis of 2-((2-methylcyclohexyl)oxy)ethan-1-amine **TA256c**



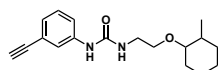
In a 100 mL flask under an inert atmosphere **TA256b** (2.0 g, 11.7 mmol, 1.0 eq.) was dissolved in THF (20 mL) and a suspension of LiAlH₄ (975.0 mg, 25.7 mmol, 2.2 eq.) in THF (15 mL) was slowly added while cooling in an ice bath. After addition, the cooling bath was removed and the grey mixture was warmed to RT, followed by refluxing for 22 h, after which TLC (5% MeOH in DCM, KMnO₄ and Ninhydrin) showed full consumption of the SM. The mixture was cooled in an ice bath and the reaction was quenched by the addition of H₂O (3 mL), NaOH (15%, 3 mL) and H₂O (3 mL). The solids were filtered off and washed with Et₂O (3 x 20 mL). The combined filtrates were washed with brine (50 mL), dried over Na₂SO₄ and concentrated by rotary evaporation to yield the desired product as a colorless liquid (1.7 g, **94%**).

¹H NMR (400 MHz, CDCl₃) δ/ppm: 3.65 – 3.58 (m, 1H), 3.37 – 3.30 (m, 1H), 2.83 (dd, *J* = 5.7, 4.7 Hz, 2H), 2.74 (td, *J* = 9.9, 4.0 Hz, 1H), 2.06 – 1.99 (m, 1H), 1.88 (s, 2H), 1.76 – 1.64 (m, 2H), 1.59 – 1.53 (m, 1H), 1.43 – 1.34 (m, 1H), 1.21 – 1.08 (m, 3H), 1.01 – 0.92 (m, 4H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 84.36, 70.64, 42.39, 38.52, 34.03, 31.41, 25.69, 25.10, 18.98.

HRMS (ESI): C₉H₂₀NO⁺ *calcd*: 158.1539, *found*: 158.1540.

Synthesis of **TA256**



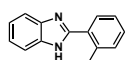
Under an inert atmosphere in a 50 mL flask triphosgene (101.0 mg, 341.0 μmol, 0.4 eq.) was dissolved in THF (8 mL) and cooled in an ice bath. A solution of 3-ethynylaniline (100.0 mg, 854.0 μmol, 1.0 eq.) and triethylamine (180.0 μL, 1.3 mmol, 1.5 eq.) in THF (2 mL) was added over a period of 1 h by syringe pump. The white suspension was stirred at 0°C for another 15 min, then the cooling bath was removed and the mixture was slowly warmed to RT. A solution of **TA256c** (201.0 mg, 1.3 mmol, 1.5 eq.) and triethylamine (180.0 μL, 1.3 mmol, 1.5 eq.) in THF (1 mL) was added dropwise over 5 min. The mixture was stirred at RT for 45 min, after which UPLC-MS analysis showed full consumption of the starting materials. The reaction was quenched by the addition of H₂O (5 mL) and the volatiles were removed by rotary evaporation. The residue was taken up in Et₂O (20 mL) and H₂O (20 mL). The mixture was acidified to pH 1 with conc. HCl and the layers were separated. The aqueous layer was extracted with Et₂O (2 x 20 mL) and the combined organic layers were washed with H₂O (20 mL), brine (20 mL), dried over Na₂SO₄ and concentrated *in vacuo*. The crude material was purified by flash column chromatography on the ISOLERA (Silica, 50 g, UV). Product-containing fractions were combined and concentrated to yield the desired product as colorless oil (192 mg, **75%**).

¹H NMR (400 MHz, CDCl₃) δ/ppm: 7.44 (t, *J* = 1.8 Hz, 1H), 7.35 (ddd, *J* = 8.0, 2.3, 1.3 Hz, 1H), 7.23 (t, *J* = 7.8 Hz, 1H), 7.17 (dt, *J* = 7.6, 1.4 Hz, 1H), 3.70 (ddd, *J* = 9.5, 6.2, 3.6 Hz, 1H), 3.48 (td, *J* = 6.4, 3.5 Hz, 1H), 3.42 (td, *J* = 6.3, 3.2 Hz, 2H), 3.04 (s, 1H), 2.80 (td, *J* = 9.9, 4.1 Hz, 1H), 2.07 – 1.98 (m, 1H), 1.77 – 1.71 (m, 1H), 1.70 – 1.65 (m, 1H), 1.62 – 1.54 (m, 1H), 1.41 – 1.32 (m, 1H), 1.24 – 1.04 (m, 4H), 1.02 – 0.86 (m, 5H).

¹³C NMR (101 MHz, CDCl₃) δ/ppm: 156.98, 138.54, 129.32, 127.58, 124.12, 123.09, 121.34, 84.94, 83.32, 77.54, 68.46, 41.54, 38.41, 33.92, 31.33, 25.55, 25.01, 18.91.

HRMS (ESI): C₁₈H₂₄N₂NaO₂⁺ *calcd*: 323.1730, *found*: 323.1726.

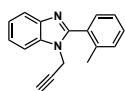
Synthesis of 2-(2-methylphenyl)-1H-benzimidazole **TA333a**



In a 50 mL flask phenylenediamine (300.0 mg, 2.8 mmol, 1.0 eq.) was dissolved in DMF (25 mL) and H₂O (2.5 mL) was added, followed by 2-methylbenzaldehyde (353.0 μ L, 3.1 mmol, 1.1 eq.). The solution was heated to 80°C and stirred under air (open flask) for 4.5 h, after which TLC and UPLC-MS analysis showed full conversion to the desired product. The volatiles were removed by rotary evaporation and the crude was directly purified by flash column chromatography on the ISOLERA (Silica, 50g, cyclohexane:EtOAc, UV). Product-containing fractions were combined and concentrated to yield the desired product as slightly yellow solid (437 mg, **76%**). Analytical data was in agreement with reported data.^[253]

¹H NMR (500 MHz, DMSO-*d*₆) δ /ppm: 12.60 (s, 1H), 7.74 (dt, *J* = 7.5, 1.1 Hz, 1H), 7.70 – 7.47 (m, 2H), 7.43 – 7.33 (m, 3H), 7.21 (dd, *J* = 6.4, 3.0 Hz, 2H), 2.61 (s, 3H).

Synthesis of **TA333**



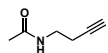
In a 50 mL flask **TA333a** (385 mg, 1.9 mmol, 1.0 eq.) was dissolved in DMF (15 mL) and propargyl bromide (339.0 μ L, 3.1 mmol, 1.7 eq.) and potassium carbonate (766.0 mg, 5.6 mmol, 3.0 eq.) were added subsequently. The mixture was stirred at RT for 4 h, after which UPLC-MS analysis showed full conversion to the desired product. The solids were filtered off and the volatiles were removed by rotary evaporation. The residue was purified by flash column chromatography on the ISOLERA (Silica, 50 g, cyclohexane:EtOAc, UV). Product-containing fractions were combined and concentrated to yield the desired product as a yellowish solid (439 mg, **96%**).

¹H NMR (400 MHz, CDCl₃) δ /ppm: 7.92 – 7.84 (m, 1H), 7.64 – 7.57 (m, 1H), 7.47 – 7.41 (m, 2H), 7.40 – 7.31 (m, 4H), 4.73 (d, *J* = 2.5 Hz, 2H), 2.38 (t, *J* = 2.5 Hz, 1H), 2.32 (s, 3H).

¹³C NMR (101 MHz, CDCl₃) δ /ppm: 152.61, 138.67, 134.09, 130.87, 130.52, 130.32, 128.53, 126.02, 123.46, 123.20, 119.91, 110.49, 76.73, 74.10, 34.25, 20.01.

HRMS (ESI): C₁₇H₁₅N₂⁺ *calcd*: 247.1230, *found*: 247.1230.

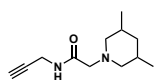
Synthesis of **TA607**



Under an inert atmosphere in a 10 mL flask but-3-ynamine (59.2 μ L, 724 μ mol, 1.0 eq.), DMAP (2.2 mg, 18.1 μ mol, 2.5 mol%) and triethylamine (101.0 μ L, 724.0 μ mol, 1.0 eq.) were dissolved in THF (3 mL) and cooled to 0°C. Acetic anhydride (137.0 μ L, 1.5 mmol, 2.0 eq.) was added dropwise and the cooling bath was removed. The solution was stirred at RT for 1 h, after which TLC (10% MeOH in DCM, KMnO₄/Ninhydrin, R_f = 0.58) showed complete consumption of the starting material. The volatiles were removed and the residue was dissolved in EtOAc (10 mL). The solution was washed with H₂O (3 x 10 mL) and the combined aqueous layers were extracted with DCM (7 x 10 mL). The combined organic layers were dried over Na₂SO₄ and concentrated *in vacuo* to yield the desired product as clear oil (66 mg, **82%**). Analytical data was in agreement with reported data.^[254]

¹H NMR (400 MHz, CDCl₃) δ /ppm: 5.79 (sbr, 1H), 3.41 (q, *J* = 6.2 Hz, 2H), 2.41 (td, *J* = 6.4, 2.6 Hz, 2H), 2.01 (t, *J* = 1.6 Hz, 1H), 2.01 (s, 3H).

Synthesis of TA622



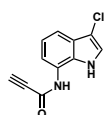
Under an inert atmosphere in a 100 mL flask propargylamine (407.0 μ L, 6.4 mmol, 1.0 eq.) and triethylamine (1.2 mL, 8.3 mmol, 1.3 eq.) were dissolved in DCM (30 mL) and cooled to 0°C. 2-Chloroacetyl chloride (557.0 μ L, 7.0 mmol, 1.1 eq.) was added dropwise over 5 min. The reaction was stirred at 0°C for 1 h, after which TLC (1:1 cyclohexane:EtOAc, R_f = 0.51, KMnO₄) showed full consumption of the starting material. The solution was mixed with H₂O (30 mL) and acidified with conc. HCl to pH 1. The layers were separated and the organic layer was washed with diluted HCl (30 mL, pH 1) and H₂O (30 mL). The combined aqueous layers were extracted with DCM (2 x 30 mL), the combined organic layers were washed with brine (30 mL), dried over Na₂SO₄ and concentrated by rotary evaporation. The residue was purified by flash column chromatography on the ISOLERA (Silica, 50 g, cyclohexane:EtOAc, KMnO₄) to yield 2-chloro-*N*-(propargyl)acetamide as a yellow oil (571 mg, 68%). In a 50 mL flask 2-chloro-*N*-(propargyl)acetamide (200.0 mg, 1.5 mmol, 1.0 eq.), 3,5-dimethylpiperidine (303.0 μ L, 2.3 mmol, 1.5 eq.) and potassium carbonate (420.0 mg, 3.0 mmol, 2.0 eq.) were suspended in THF (15 mL). The mixture was stirred at RT for 21 h, after which TLC (5% MeOH in DCM, R_f = 0.44, KMnO₄) showed full consumption of the starting material. The volatiles were removed by rotary evaporation and the crude material was purified by flash column chromatography on the ISOLERA (Silica, 50g, DCM:MeOH, KMnO₄) to yield the desired product as a yellow oil (141 mg, 45%).

¹H NMR (400 MHz, CDCl₃) δ /ppm: 7.60 (sbr, 1H), 4.06 (dd, J = 5.6, 2.5 Hz, 2H), 3.08 (sbr, 2H), 2.81 (sbr, 2H), 2.22 (t, J = 2.6 Hz, 1H), 1.86 – 1.68 (m, 4H), 0.91 (t, J = 6.8 Hz, 1H), 0.87 (s, 3H), 0.85 (s, 3H), 0.64 – 0.49 (m, 1H).

¹³C NMR (101 MHz, CDCl₃) δ /ppm: 160.72, 79.70, 71.34, 61.49, 42.64, 41.33, 31.22, 28.79, 19.40.

HRMS (ESI): C₁₂H₂₁N₂O⁺ *calcd*: 209.1648, *found*: 209.1649.

Synthesis of TA660



In a 5 mL flask 3-chloro-1*H*-indol-7-amine (50.0 mg, 300.0 μ mol, 1.0 eq.) was dissolved in EtOAc (1 mL) and propiolic acid (27.7 μ L, 450.0 μ mol, 1.5 eq.), DIPEA (157.0 μ L, 900.0 μ mol, 3.0 eq.) and T3P (447.0 μ L, 750.0 μ mol, 2.5 eq. 50% in EtOAc) were added subsequently. The black solution was stirred at RT for 3 h. Saturated NaHCO₃ solution (1 mL) was added and the mixture was stirred overnight. The biphasic mixture was diluted with H₂O (10 mL) and EtOAc (10 mL). The layers were separated and the aqueous layer was extracted with EtOAc (2 x 10 mL). The combined organic layers were washed with brine (10 mL), dried over Na₂SO₄ and were concentrated by rotary evaporation. The crude was purified by flash column chromatography (Silica, 10 g, cyclohexane:EtOAc 2:1, R_f = 0.26) to yield the desired product as a greenish solid (37 mg, 56%).

¹H NMR (500 MHz, CDCl₃) δ /ppm: 9.73 (s, 1H), 7.84 (s, 1H), 7.53 (dt, J = 8.0, 0.9 Hz, 1H), 7.22 (d, J = 2.6 Hz, 1H), 7.11 (t, J = 7.8 Hz, 1H), 6.82 (dd, J = 7.6, 1.0 Hz, 1H), 3.03 (s, 1H).

¹³C NMR (126 MHz, CDCl₃) δ /ppm: 149.72, 128.46, 127.00, 122.13, 121.79, 120.10, 116.97, 114.26, 106.76, 77.28, 75.57.

HRMS (ESI): C₁₁H₇ClN₂NaO⁺ *calcd*: 241.0139, *found*: 241.0138.

Synthesis of TA661

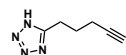


Propargyl bromide (1.9 mL, 18.0 mmol, 1.0 eq.) and sodium sulfite (2.9 g, 22.7 mmol, 1.3 eq.) were dissolved in MeOH (7 mL) and H₂O (7 mL) and the mixture was heated to 65°C for 7 h. After cooling, MeOH (120 mL) was added and the mixture was filtered over a G4 sintered glass funnel. The filtrate was concentrated to an approximate volume of 5 mL. Acetone (100 mL) was added and the precipitate was filtered off, washed with acetone (2 x 10 mL) and dried *in vacuo* to yield the desired product as white solid (2.4 g, **93%**). Analytical data was in agreement with reported data.^[255]

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 3.29 (d, *J* = 2.7 Hz, 2H), 2.92 (t, *J* = 2.7 Hz, 1H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 78.87, 73.24, 42.09.

Synthesis of TA662



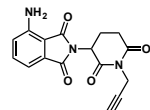
In a dry flask 5-hexynenitrile (1.1 mL, 10.7 mmol, 1.0 eq.) was dissolved in toluene (20 mL) and sodium azide (2.1 g, 32.2 mmol, 3.0 eq.) and triethylamine hydrochloride (4.4 g, 32.2 mmol, 3.0 eq.) were added subsequently. The mixture was heated to reflux for 9 h. After cooling the mixture was extracted with H₂O (4 x 10 mL) and the combined aqueous layers were acidified to pH 1 by addition of concentrated HCl. The aqueous phase was extracted with EtOAc (4 x 20 mL) and the combined EtOAc layers were dried over Na₂SO₄ and concentrated by rotary evaporation. The crude red oil was purified by flash column chromatography (Silica, 80 g, cyclohexane:EtOAc 2:1 + 1% formic acid→1:1 + 1% formic acid, R_f= 0.36, KMnO₄) to yield the desired product as yellowish oil that crystallized in the fridge (1.0 g, **69%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ/ppm: 15.89 (s, 1H), 2.96 (dd, *J* = 8.1, 7.2 Hz, 2H), 2.85 (t, *J* = 2.7 Hz, 1H), 2.26 (td, *J* = 7.0, 2.7 Hz, 2H), 1.88 (dq, *J* = 8.3, 7.0 Hz, 2H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ/ppm: 155.44, 83.44, 72.00, 25.90, 21.81, 17.21.

HRMS (ESI): not measurable on the ESI.

Synthesis of TA663



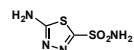
In a screw-capped vial pomalidomide (100.0 mg, 366.0 μmol, 1.0 eq.) was suspended in DMF (2 mL) and THF (5 mL). Propargyl bromide (59.2 μL, 549.0 μmol, 1.5 eq.) and potassium carbonate (152.0 mg, 1.1 mmol, 3.0 eq.) were added and the mixture was heated to 65°C for 15 h, after which UPLC-MS analysis showed full conversion to the desired product. The solvent was removed by rotary evaporation and the residue was dissolved in EtOAc (25 mL) and H₂O (25 mL). The layers were separated and the aqueous layer was extracted with EtOAc (2 x 25 mL). The combined organic layers were washed with brine (25 mL), dried over Na₂SO₄ and were concentrated. The crude material was purified by flash column chromatography (Silica, 20 g, DCM + 1% MeOH, R_f= 0.34 (2% MeOH in DCM), UV). Product-containing fractions were combined and concentrated *in vacuo* to yield the desired product as a yellow solid (116 mg, **quant.**).

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ /ppm: 7.48 (dd, $J = 8.5, 7.0$ Hz, 1H), 7.02 (dd, $J = 7.7, 5.6$ Hz, 2H), 6.55 (s, 2H), 5.20 (dd, $J = 13.0, 5.4$ Hz, 1H), 4.38 (t, $J = 2.4$ Hz, 2H), 3.11 (t, $J = 2.4$ Hz, 1H), 3.03 (ddd, $J = 17.3, 13.9, 5.4$ Hz, 1H), 2.80 (ddd, $J = 17.3, 4.4, 2.5$ Hz, 1H), 2.57 (td, $J = 13.3, 4.4$ Hz, 1H), 2.05 (dtd, $J = 12.9, 5.4, 2.4$ Hz, 1H).

^{13}C NMR (126 MHz, $\text{DMSO-}d_6$) δ /ppm: 170.84, 169.02, 168.45, 167.27, 146.77, 135.51, 131.95, 121.76, 111.01, 108.44, 79.06, 72.99, 49.01, 31.03, 29.05, 21.18.

HRMS (ESI): $\text{C}_{16}\text{H}_{13}\text{N}_3\text{NaO}_4^+$ *calcd*: 334.0798, *found*: 334.0796.

Synthesis of TA664a

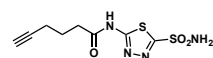


Azetazolamide (1.0 g, 4.5 mmol, 1.0 eq.) was suspended in aqueous HCl (8.0 mL, 24.0 mmol, 5.3 eq. 3.0 M) and the mixture was heated to reflux for 3 h, after which UPLC-MS analysis showed full conversion to the desired product. The solution was neutralized with NaOH (4.0 M, 7.0 mL), whereby precipitation occurred. The mixture was extracted with EtOAc (3 x 80 mL) and the combined organic layers were washed with brine (80 mL), dried over Na_2SO_4 and concentrated *in vacuo* to yield the desired product as a white solid (436 mg, **54%**). Analytical data was in agreement with reported data.^[256]

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ /ppm: 8.05 (s, 2H), 7.80 (s, 2H).

^{13}C NMR (101 MHz, $\text{DMSO-}d_6$) δ /ppm: 171.62, 157.84.

Synthesis of TA664



Under an inert atmosphere 5-hexynoic acid (91.9 μL , 832.0 μmol , 1.0 eq.) was dissolved in DCM (3.5 mL) and DMF (5.0 μL , 64.6 μmol , 8 mol%) was added. The solution was cooled in an ice bath while oxalyl chloride (65.0 μL , 757.0 μmol , 0.91 eq.) was added dropwise over 15 min. The cooling bath was removed and the solution was stirred at RT for 1 h. The mixture was concentrated by rotary evaporation, the residue was slowly added to a solution of **TA664a** (150.0 mg, 832.0 μmol , 1.0 eq.) and pyridine (134.0 μL , 1.7 mmol, 2.0 eq.) in DMF (1 mL). The yellow solution was stirred at RT for 3 h. The volatiles were removed by rotary evaporation and the crude material was purified by reversed phase column chromatography on the ISOLERA (100 g, $\text{H}_2\text{O}/\text{MeCN} + 0.1\%$ TFA, UV). Product-containing fractions were combined and lyophilized to yield the desired product as a white solid (131 mg, **57%**). Analytical data was in agreement with reported data.^[228]

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ /ppm: 13.02 (s, 1H), 8.31 (s, 2H), 2.82 (t, $J = 2.6$ Hz, 1H), 2.64 (t, $J = 7.4$ Hz, 2H), 2.23 (td, $J = 7.1, 2.7$ Hz, 2H), 1.80 (p, $J = 7.2$ Hz, 2H).

13.15 Building Block Encoding and Virtual Library Assembly

Virtual Library Assembly with SMILES

To generate our macrocycle library *in silico* we divided the macrocycle structure into the three diversity elements, the basic scaffold and the linker scaffold. These parts were depicted as SMILES strings and could therefore be easily connected with each other using our self-developed DECL-Gen software (see Chapter 8.2 for details about the software, the connection is based on the idea used in Smilib^[257]). Meanwhile, every building block of the three diversity elements was encoded by a unique DNA sequence (codon). Connecting the SMILES strings of the building blocks with the corresponding DNA codons generated the macrocycle structures along with their full DNA sequences. This methodology could also be used in the other direction to generate the macrocycle structure from the DNA coding information after next generation sequencing (see Chapters 9.2 and 9.3).

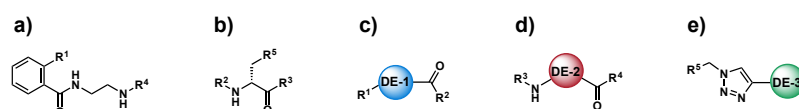


Figure 106. a) Basic scaffold with the linkage sites R¹ and R⁴. b) Trifunctional scaffold with the linkage sites R² and R³. c) Schematic representation of the DE-1 building blocks with linkage sites R¹ and R². d) Schematic representation of the DE-2 building blocks with linkage sites R³ and R⁴. e) Schematic representation of the DE-3 building blocks with linkage site R⁵. The triazole moiety from the click reaction is already included in this smiles element.

The SMILES string for the basic scaffold with the trifunctional linker scaffold included (**Figure 106 a/b**) has the following structure:

```
c1ccc(C(=O)NCCN[R4])c([R1])c1.N([R2])[C@@H](C[R5])C(=O)[R3]
```

The DNA strand sequence used for the generation of the DNA strands of the DEML was the following:

```
GGAGCTTGTGAATTCTGG{de1}GGACGTGTGTGAATTGTC{de2}GTGTGCGGATCCAAGTTCGGTGAATGGA  
{de3}ACTACGGATGGATACTCT
```

The {de1}, {de2} and {de3} parts represent the defined codons for the building blocks with 3, 4 and 6 DNA bases respectively. For the generation of the small macrocycle library (smDEML) we used the same SMILES structures as stated above and for the DE-3 elements we just defined one element for the sidechain azide with the SMILES code: [R5]C-N=[N+]=[N-]

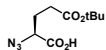
We used the following DNA sequence for the analysis of the smDEML selections:

```
GGAGCTTGTGAATTCTGG{de1}GGACGTGTGTGAATTGTC{de2}GTGTGCGGATCCAATAC
```

SMILES codons and DNA codons for the building blocks of DE1, DE-2 and DE-3 are found in **Table 16**, **Table 17** and **Table 18**.

13.16 Synthesis of the Small Molecule Library Scaffold

(S)-2-azido-5-(tert-butoxy)-5-oxopentanoic Acid **2-26a**



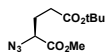
Under an inert atmosphere H-Glu(OtBu)-OH (100.0 mg, 492.0 μmol , 1.0 eq.) and **2-5a** (124.0 mg, 590.0 μmol , 1.2 eq.) were dissolved in MeOH (3 mL) and potassium carbonate (184.0 mg, 1.3 mmol, 2.7 eq.) was added followed by copper sulfate pentahydrate (1.2 mg, 4.9 μmol , 1 mol%). The blue suspension was stirred at RT for 2 h, after which TLC showed full consumption of the starting material. The solvent was removed by rotary evaporation and the residue was dissolved in H₂O (10 mL) and EtOAc (10 mL). The mixture was acidified with concentrated aqueous HCl to pH 1-2. The layers were separated and the aqueous phase was extracted with EtOAc (2 x 15 mL). The combined organic layers were washed with H₂O (2 x 20 mL), brine (20 mL), dried over Na₂SO₄ and concentrated *in vacuo*. The crude material was purified by flash column chromatography (Silica, 12 g, DCM/2% MeOH/0.25% formic acid, R_f= 0.21, Bromocresol Green stain) to yield the desired product as yellow oil (110 mg, **98%**).

¹H NMR (400 MHz, DMSO-*d*₆) δ /ppm: 13.42 (s, 1H), 4.14 (dd, *J* = 8.8, 4.8 Hz, 1H), 2.36 – 2.23 (m, 2H), 2.03 – 1.93 (m, 1H), 1.86 – 1.73 (m, 1H), 1.40 (s, 9H).

¹³C NMR (101 MHz, DMSO-*d*₆) δ /ppm: 171.45, 171.22, 60.64, 55.49, 31.15, 27.71, 26.29.

HRMS (ESI): C₉H₁₅N₃NaO₄⁺ *calcd*: 252.0955, *found*: 252.0950.

5-(tert-butyl) 1-methyl (S)-2-azidopentanedicarboxylate **2-26b**



Under an inert atmosphere **2-26a** (100.0 mg, 436.0 μmol , 1.0 eq.) was dissolved in MeOH (4 mL) and cooled in an ice bath. TMS-diazomethane (2.2 mL, 4.4 mmol, 10.0 eq. 2.0 M in hexane) was added dropwise over 10 min while cooling. Strong gas evolution was observed at the beginning. The yellow solution was stirred at 0°C for 10 min and was then slowly warmed to RT. After 17 h TLC showed complete conversion of the starting material. The volatiles were removed by rotary evaporation and the residue was dissolved in DCM (30 mL). The solution was washed with half-saturated bicarbonate solution (20 mL) and H₂O (2 x 20 mL). The combined aqueous layers were extracted with DCM (2 x 20 mL). The combined organic layers were washed with brine (20 mL), dried over Na₂SO₄ and concentrated *in vacuo* to yield the desired product as a yellow oil (101 mg, **95%**).

¹H NMR (400 MHz, CDCl₃) δ /ppm: 3.99 (dd, *J* = 8.8, 5.1 Hz, 1H), 3.80 (s, 3H), 2.44 – 2.31 (m, 2H), 2.15 (dtd, *J* = 14.2, 7.5, 5.1 Hz, 1H), 2.05 – 1.91 (m, 1H), 1.45 (s, 9H).

¹³C NMR (126 MHz, CDCl₃) δ /ppm: 171.64, 170.73, 81.07, 61.33, 52.85, 31.42, 28.23, 26.83.

HRMS (ESI): C₁₀H₁₇N₃NaO₄⁺ *calcd*: 266.1111, *found*: 266.1112.

(S)-4-azido-5-methoxy-5-oxopentanoic Acid; Linker Scaffold **2-26**



2-26b (289.0 mg, 1.2 mmol, 1.0 eq.) was dissolved in formic acid (7.6 mL, 202.0 mmol, 170 eq.) and stirred at RT for 17 h. The volatiles were removed by rotary evaporation and the crude material was purified by flash column chromatography (Silica, 10 g, DCM/2%MeOH/0.25% formic acid, R_f= 0.20, bromocresol green) to yield the desired product as yellow/orange oil (174 mg, **78%**).

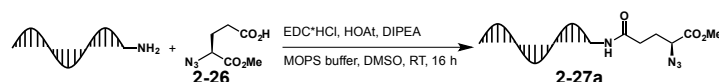
^1H NMR (400 MHz, CDCl_3) δ /ppm: 4.05 (dd, $J = 8.7, 5.1$ Hz, 1H), 3.82 (s, 3H), 2.59 – 2.48 (m, 2H), 2.20 (dtd, $J = 14.3, 7.5, 5.1$ Hz, 1H), 2.08 – 1.97 (m, 1H).

^{13}C NMR (126 MHz, CDCl_3) δ /ppm: 177.54, 170.48, 61.04, 52.94, 29.76, 26.33.

HRMS (ESI): $\text{C}_6\text{H}_9\text{N}_3\text{NaO}_4^+$ *calcd*: 210.0485, *found*: 210.0487.

13.17 Validation of the Small Molecule Library Chemistry

DNA Coupling with the Scaffold 2-26



In a 5 mL Eppendorf tube, 5'-amine modified DNA with the sequence 5'-GGAGCTTGTATATGCTGGAAATGCAATGTCACAGTCTTA-3' (825.0 μL , 100.0 μM in MOPS buffer, 1.0 eq.) was mixed with **2-26** (454.0 μL , 200.0 mM in DMSO, 1100.0 eq.), EDC hydrochloride (206.0 μL , 400.0 mM in DMSO, 1000.0 eq.), HOAt (206.0 μL , 400.0 mM in DMSO, 1000.0 eq.) and DIPEA (413.0 μL , 300.0 mM in DMSO, 1500.0 eq.). The yellow solution was agitated at RT for 23 h, after which HPLC analysis showed 76% conversion. The reaction was purified by EtOH precipitation. The DNA pellet was further purified by reversed-phase preparative HPLC (Method C). The product-containing fractions were combined and lyophilized. The white solid was dissolved in H_2O (500 μL) to yield a 90 μM solution.

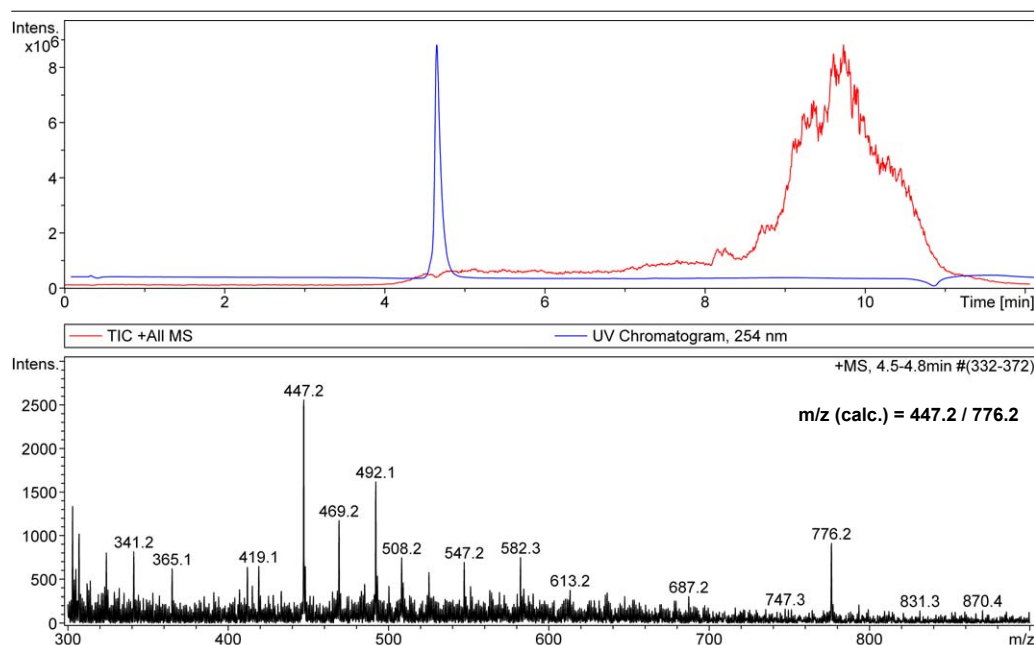
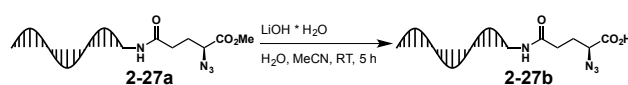


Figure 107. LC-MS analysis of the purified DNA-encoded scaffold **2-27a**.

Ester Hydrolysis of 2-27a



In a 5 mL Eppendorf tube **2-27a** (1.0 mL, 85.0 μM in H_2O , 1.0 eq.) was mixed with LiOH monohydrate (142.0 μL , 300.0 mM in H_2O , 500.0 eq.) and MeCN (571 μL). The solution was agitated at RT for 5 h. The reaction was purified by EtOH precipitation and the DNA pellet was redissolved in MOPS buffer (1.2 mL) to yield a 66 μM solution.

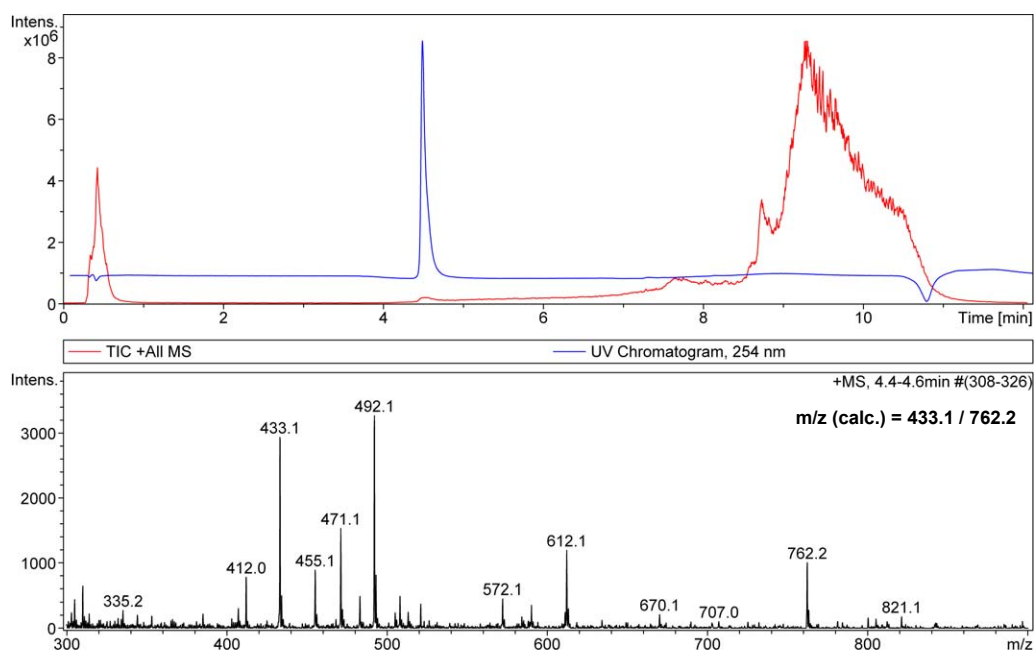
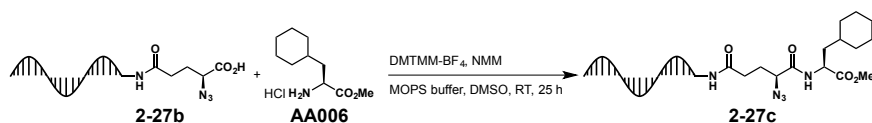


Figure 108. LC-MS analysis of the ester hydrolysis product **2-27b** after purification.

Amide Coupling with **AA006**



In a 1.5 mL Eppendorf tube **2-27b** (15.0 μ L, 50.0 μ M in MOPS buffer, 1.0 eq.) was mixed with **AA006** (1.5 μ L, 500.0 mM in DMSO, 1000.0 eq.), DMTMM-BF₄ (7.5 μ L, 100.0 mM in DMSO, 1000.0 eq.) and NMM (12.5 μ L, 180.0 mM in DMSO, 3000.0 eq.). The solution was agitated at RT for 25 h. The reaction was purified by EtOH precipitation and the DNA pellet was redissolved in TEAA buffer (15 μ L, 500 mM, pH 7.2) to yield a 55 μ M solution.

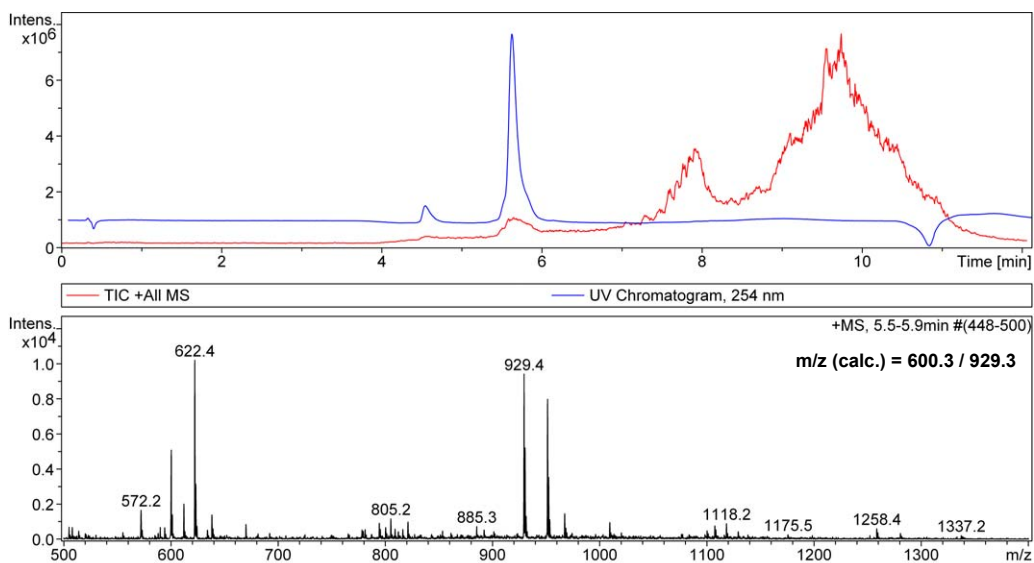
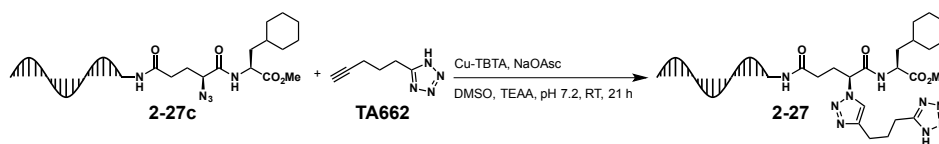


Figure 109. LC-MS analysis of the purified compound **2-27c**.

Click Reaction between **2-27c** and **TA662**



In a 0.5 mL Eppendorf tube **2-27c** (12.0 μL , 24.0 μM in TEAA, 1.0 eq.) was mixed with **TA662** (2.9 μL , 10.0 mM, 100.0 eq.), NaOAsc (2.9 μL , 20.0 mM in H_2O , 200.0 eq.) and DMSO (11 μL). The solution was degassed with N_2 for 30 s. Cu-TBTA complex (5.8 μL , 10.0 mM in 55% DMSO, 200.0 eq.) was added and the solution was degassed again with N_2 for 30 s. The solution was agitated at RT for 21 h. The reaction was purified by EtOH precipitation and the DNA pellet was dissolved in H_2O (10 μL) to yield a 29 μM solution.

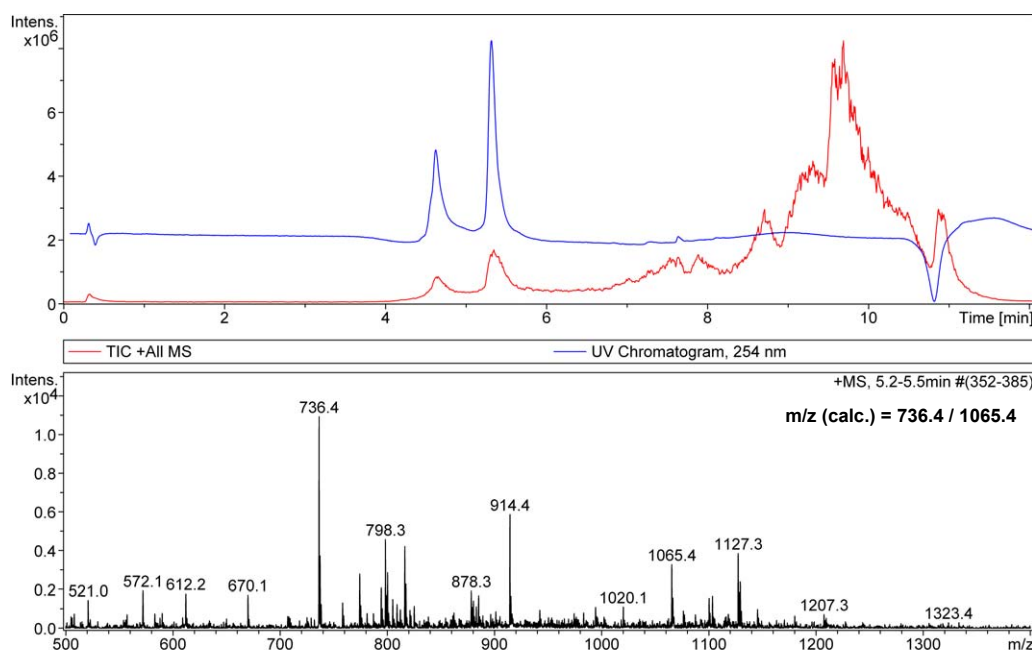
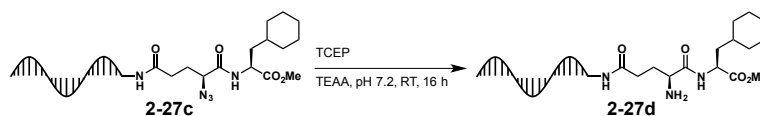


Figure 110. LC-MS of the clicked product **2-27**. The UV peak at 4.5 min arises from incomplete scaffold ester hydrolysis in a previous reaction. Optimizations showed that this peak could be avoided and only the desired product was found.

Staudinger Reduction with TCEP of **2-27c**



In a 1.5 mL Eppendorf tube **2-27c** (30.0 μL , 63.0 μM in 500 mM TEAA, 1.0 eq.) was mixed with TCEP (30.0 μL , 100.0 mM in H_2O , 1587.0 eq.) and the solution was agitated at RT for 16 h. The reaction was purified by EtOH precipitation and the DNA pellet was redissolved in MOPS buffer (50 μL) to yield a 34 μM solution.

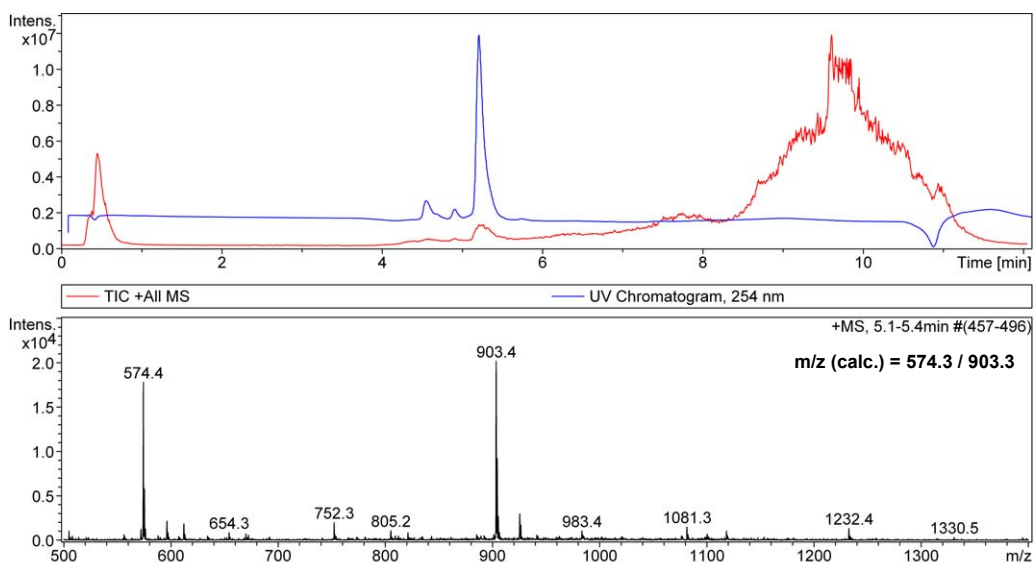
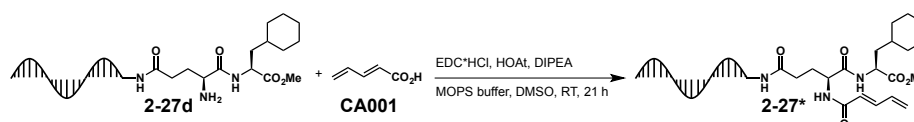


Figure 111. LC-MS analysis of the azide reduction of **2-27c**.

Carboxylic Acid Coupling with **2-27d**



In a 0.5 mL Eppendorf tube **2-27d** (4.5 μ L, 18.5 μ M in MOPS buffer, 1.0 eq.), **CA001** (0.8 μ L, 100.0 mM in DMSO, 1000.0 eq.), EDC hydrochloride (1.7 μ L, 50.0 mM in DMSO, 1000.0 eq.), HOAt (1.7 μ L, 50.0 mM in DMSO, 1000.0 eq.) and DIPEA (2.5 μ L, 50.0 mM in DMSO, 1500.0 eq.) were mixed and the solution was left standing at RT for 21 h. The mixture was purified by EtOH precipitation, followed by LC-MS analysis.

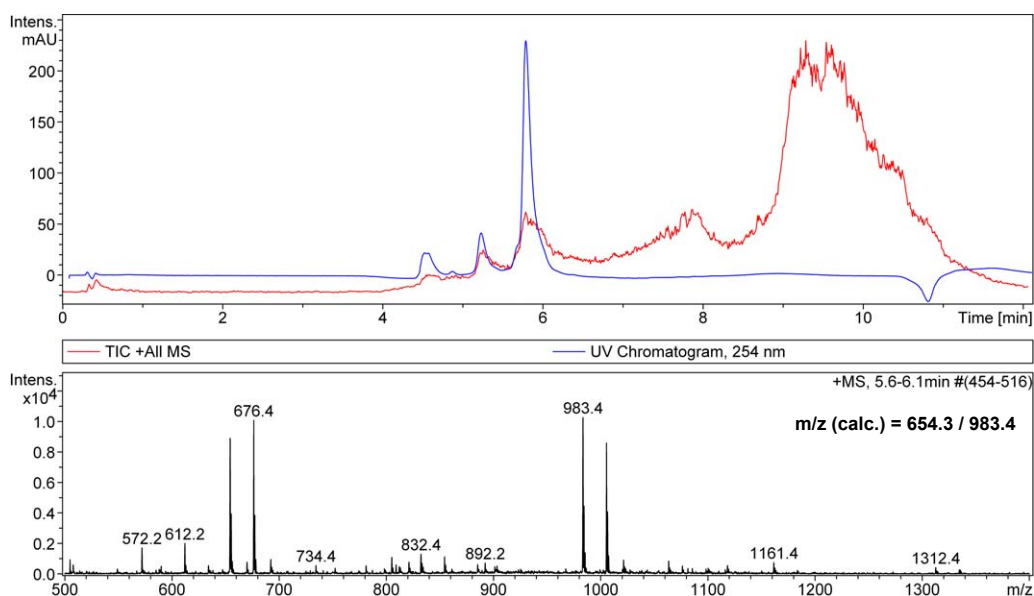
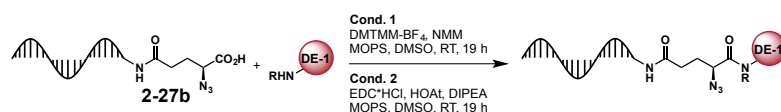


Figure 112. LC-MS analysis of the carboxylic acid coupling with **2-27d**.

13.18 Small Molecule Library Assembly and Building Block Validations

DE-1 Building Block Validation



In PCR 96-well plates, 126 amino acids 26 secondary amines and 26 primary amines were tested for their coupling efficiencies. Each reaction was conducted with the following amounts. **2-27b** (2.5 μ L, 50.0 μ M in MOPS buffer, 1.0 eq.) was mixed with the DE-1 building block (2.5 μ L, 50.0 mM in DMSO, 1000.0 eq.), NMM (2.1 μ L, 180.0 mM in DMSO, 3000.0 eq.) and DMTMM-BF₄ (2.1 μ L, 60.0 mM in DMSO, 1000.0 eq.). The reactions were left standing at RT for 9 h. The reactions were purified by EtOH precipitation and the DNA pellets were redissolved in H₂O (20 μ L each). All reactions were analyzed by LC-MS.







































Reactions with EDC/HOAt/DIPEA conditions were performed accordingly. EDC hydrochloride and HOAt were used at 120.0 mM concentration and 1.1 μ L each.





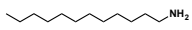


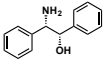
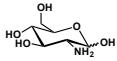
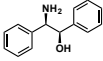
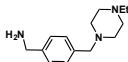
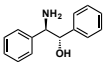





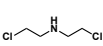






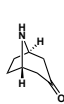
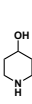



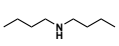

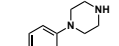






Table 14. DE-1 building blocks screening results. 126 amino acids, 26 primary amines and 26 secondary amines were tested for their reactivity in amide coupling. Red coloring indicates the building blocks that were excluded from the final DEML assembly due to insufficient purity. Blue coloring indicates building blocks that fulfilled the criteria but were not included into the final library. Cutoff: >80% purity.

No.	Structure	Calc. Mass [m/z]	Obs. Mass [m/z] ^[a]	Purity	No.	Structure	Calc. Mass [m/z]	Obs. Mass [m/z] ^[a]	Purity
AA001		518.2/ 847.2	540.2/ 869.3	>95%	AA064		672.1/ 1001.2	694.2/ 1001.2	>95%
AA002		518.2/ 847.2	540.2/ 869.3	>95%	AA065		560.2/ 889.3	582.3/ 889.4	>95%
AA003		518.2/ 847.2	540.2/ 869.3	>95%	AA066		560.2/ 889.3	582.3/ 889.4	>95%
AA004 ^[b]		552.1/ 881.2	510.2/ 839.2	0%	AA067		546.2/ 875.3	568.3/ 875.4	>95%
AA005		532.2/ 861.3	554.2/ 883.3	ca 90%	AA068 ^[f]		633.2/ 962.3	533.3/ 862.4	>95%
AA006		600.3/ 929.3	622.3/ 929.4	>95%	AA069 ^[f]		633.26/ 962.3	533.3/ 862.4	>95%
AA007		595.2/ 924.3	595.3/ 924.3	>95%	AA070 ^[f]		647.3/ 976.3	547.3/ 876.4	>95%
AA008 ^[c]		682.4/ 1011.4	704.4/ 1011.4	>95%	AA071		532.2/ 861.3	554.3/ 861.3	85- 90%
AA009		546.2/ 875.3	568.3/ 897.3	>95%	AA072		532.2/ 861.3	554.3/ 861.3	>95%
AA010		546.2/ 875.3	568.2/ 897.3	>95%	AA073		542.2/ 871.3	564.2/ 893.3	>95%
AA011		560.2/ 889.3	582.3/ 911.3	19%	AA074		560.2/ 889.3	582.3/ 911.4	72%
AA012		504.2/ 829.3	526.2/ 851.3	>95%	AA075		532.2/ 857.3	554.3/ 879.3	>95%

		833.3	833.3				861.3	883.3	
AA013		518.2/ 847.3	540.2/ 869.3	>95%	AA076		560.2/ 889.3	582.3/ 889.4	90- 95%
AA014		580.2/ 909.3	602.3/ 909.3	>95%	AA077		520.2/ 849.2	542.2/ 849.3	41%
AA015		594.2/ 923.3	616.3/ 923.4	>95%	AA078		650.3/ 979.3	672.3/ 979.4	88%
AA016 ^[d]		596.2/ 925.3	618.3/ 925.3	36% >95%	AA079		588.3/ 917.3	588.3/ 917.4	>95%
AA017 ^[e]		618.3/ 947.3	640.3/ 947.3	>95%	AA080		560.2/ 889.3	560.3/ 889.4	>95%
AA018		598.2/ 927.3	598.3/ 927.4	>95%	AA081		574.3/ 903.3	574.3/ 903.4	>95%
AA019		654.2/ 983.3	676.3/ 983.3	>95%	AA082		588.3/ 917.3	588.3/ 917.4	>95%
AA020 ^[d]		610.2/ 939.3	610.3/ 939.3	20% 93%	AA083 ^[b]		543.2/ 872.2	612.2/ 800.3	0%
AA021 ^[d]		610.2/ 939.3	610.3/ 939.3	30% 94%	AA084		611.2/ 940.2	633.3/ 940.3	>95%
AA022 ^[d]		624.2/ 953.3	624.3/ 953.4	24% 56%	AA085		600.2/ 929.3	622.3/ 929.3	88%
AA023 ^[d]		862.0/ 1191.1	862.1/ 1191.1	0% 92%	AA086		528.2/ 857.3	550.2/ 857.3	>95%
AA024		575.2/ 904.3	597.3/ 904.4	>95%	AA087		530.2/ 859.3	552.2/ 859.3	>95%
AA025		534.2/ 863.2	556.2/ 863.3	>95%	AA088		530.2/ 859.3	552.2/ 859.3	>95%
AA026		534.2/ 863.2	556.2/ 863.3	>95%	AA089		570.2/ 899.3	570.3/ 899.4	>90%
AA027		560.2/ 889.3	560.3/ 889.4	>95%	AA090		558.2/ 887.3	580.3/ 887.4	94%
AA028		574.3/ 903.3	574.3/ 903.4	68%	AA091		572.2/ 901.3	572.3/ 901.4	>95%
AA029		633.2/ 962.3	633.2/ 962.4	96%	AA092		572.2/ 901.3	572.3/ 901.4	>95%
AA030		633.2/ 962.3	633.3/ 962.4	95%	AA093		544.2/ 873.3	566.3/ 873.4	>95%
AA031		647.3/ 976.3	647.4/ 976.4	82%	AA094		530.2/ 859.3	552.2/ 859.3	89%

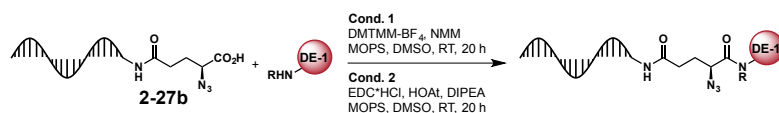
AA032		548.2/ 877.3	570.3/ 877.4	>95%	AA095		600.3/ 929.3	622.4/ 929.4	66%
AA033		548.2/ 877.3	570.2/ 899.3	>95%	AA096		544.2/ 873.3	544.3/ 873.4	95%
AA034		548.2/ 877.3	570.3/ 877.4	>95%	AA097		584.2/ 913.3	584.3/ 913.4	>95%
AA035		548.2/ 877.3	570.2/ 899.3	>95%	AA098		586.3/ 915.3	608.3/ 915.4	>90%
AA036		648.3/ 977.3	670.3/ 977.4	>95%	AA099		634.2/ 963.3	634.3/ 963.3	>95%
AA037		648.3/ 977.3	670.3/ 977.3	>95%	AA100		622.2/ 951.2	644.2/ 951.3	83%
AA038 ^[f]		661.3/ 990.3	561.3/ 890.4	>95%	AA101		666.3/ 995.4	688.4/ 995.4	>95%
AA039		560.2/ 889.3	582.3/ 889.4	>95%	AA102		584.2/ 913.3	606.3/ 913.3	>95%
AA040		560.2/ 889.3	582.3/ 911.4	>95%	AA103		600.3/ 929.3	622.3/ 929.4	>95%
AA041 ^[f]		675.3/ 1004.4	575.3/ 904.4	>95%	AA104 ^[b]		589.2/ 918.2	797.3/ 936.3	0%
AA042		709.3/ 1038.3	731.3/ 1038.4	>95%	AA105 ^[b]		608.2/ 937.2	572.2/ 1100.2	0%
AA043 ^[f]		675.3/ 1004.4	575.3/ 904.4	42%	AA106		681.3/ 1010.3	681.3/ 1010.3	94%
AA044		578.2/ 907.3	600.3/ 907.3	>95%	AA107		601.2/ 930.2	623.2/ 930.3	>95%
AA045		578.2/ 907.3	600.3/ 907.3	>95%	AA108		645.2/ 974.3	645.3/ 974.3	>95%
AA046 ^[e]		632.3/ 961.3	576.3/ 905.3	>95%	AA109		692.2/ 1021.3	714.3/ 1021.3	>95%
AA047		561.2/ 890.3	583.3/ 912.3	>95%	AA110		644.2/ 973.3	666.3/ 973.3	>95%
AA048		544.2/ 873.3	544.3/ 873.3	>95%	AA111		608.2/ 937.3	630.3/ 937.3	>95%
AA049		544.2/ 873.3	566.3/ 873.3	>95%	AA112		608.2/ 937.3	630.3/ 959.3	>95%
AA050		558.2/ 887.3	580.3/ 909.4	>95%	AA113		626.2/ 955.3	648.3/ 977.3	>95%
AA051		558.2/ 887.3	580.3/ 909.4	>95%	AA114		620.2/ 955.3	642.3/ 977.3	>95%

		887.3	909.4				949.3	949.4	
AA052		560.2/ 889.3	582.3/ 911.3	>95%	AA115		638.2/ 967.3	660.3/ 967.3	>95%
AA053		560.2/ 889.3	582.3/ 889.3	>95%	AA116		622.3/ 951.3	644.3/ 951.4	>95%
AA054		542.2/ 871.3	542.3/ 871.3	>95%	AA117		640.2/ 969.3	662.3/ 969.3	>95%
AA055		594.2/ 923.3	616.3/ 923.4	>95%	AA118		628.2/ 957.2	650.2/ 957.3	>95%
AA056		594.2/ 923.3	616.3/ 923.4	>95%	AA119		670.3/ 999.3	692.3/ 999.4	87%
AA057		608.2/ 937.3	630.3/ 937.4	>95%	AA120		658.3/ 987.3	680.3/ 987.4	>95%
AA058		639.2/ 968.3	661.3/ 968.3	>95%	AA121		586.2/ 915.2	608.2/ 937.3	>95%
AA059		639.2/ 968.3	661.3/ 968.3	>95%	AA122		584.2/ 913.3	584.3/ 913.4	>90%
AA060		608.2/ 937.3	608.3/ 937.4	63%	AA123		688.3/ 1017.3	688.4/ 1017.4	>95%
AA061		626.2/ 955.3	648.3/ 955.4	>95%	AA124		584.2/ 913.3	606.3/ 913.4	>95%
AA062		612.2/ 941.3	634.3/ 941.4	>95%	AA125		556.2/ 885.3	556.3/ 885.4	>95%
AA063		612.2/ 941.3	634.3/ 941.3	>95%	AA126		601.2/ 930.2	623.2/ 930.3	>95%
PAM001		474.2/ 803.3	474.3/ 803.3	>95%	PAM014		621.3/ 950.3	621.4/ 950.4	86%
PAM002		565.3/ 894.3	565.3/ 894.4	84%	PAM015		547.2/ 876.3	547.2/ 876.3	>95%
PAM003		567.2/ 896.2	567.3/ 896.4	>90%	PAM016		471.2/ 800.2	471.2/ 800.3	>95%
PAM004		516.2/ 845.3	516.3/ 845.4	>95%	PAM017		506.2/ 835.3	506.2/ 835.3	>95%
PAM005		504.2/ 833.3	504.3/ 833.3	94%	PAM018		543.2/ 872.2	543.2/ 872.3	>95%
PAM006		587.3/ 916.3	587.3/ 916.4	94%	PAM019^[b]		488.2/ 817.3	612.2/ 831.3	0%
PAM007		566.2/ 895.3	566.3/ 895.4	>95%	PAM020		586.2/ 915.3	586.3/ 915.4	92%

PAM008^[b]		547.2/ 876.2	572.2/ 1100.2	0%	PAM021		502.2/ 831.3	502.3/ 831.4	85%
PAM009		502.2/ 831.3	502.3/ 831.3	89%	PAM022		516.3/ 845.3	516.3/ 845.4	82%
PAM010		600.4/ 929.4	600.4/ 929.5	80%	PAM023		583.3/ 912.4	583.4/ 912.5	80%
PAM011		540.2/ 869.2	540.2/ 869.3	>95%	PAM024		628.3/ 957.3	628.3/ 957.4	>90%
PAM012		594.2/ 923.3	616.3/ 923.4	80%	PAM025		628.3/ 957.3	628.3/ 957.4	>90%
PAM013		648.3/ 977.4	648.4/ 977.5	>90%	PAM026		628.3/ 957.3	628.3/ 957.4	>90%
SAM001		518.2/ 847.2	518.2/ 847.3	>95%	SAM014^[b]		516.3/ 845.3	522.3/ 851.4	0%
SAM002		528.3/ 857.3	550.3/ 857.4	8%	SAM015		500.2/ 829.3	500.2/ 829.3	89%
SAM003^[b]		556.3/ 885.3	522.3/ 829.4	0%	SAM016^[b]		556.1/ 885.2	612.2/ 867.3	0%
SAM004^[b]		596.3/ 925.4	550.3/ 879.3	0%	SAM017		516.2/ 845.3	538.3/ 845.3	88%
SAM005		528.3/ 857.3	528.3/ 857.4	83%	SAM018		514.2/ 843.3	514.3/ 843.4	>95%
SAM006		515.2/ 844.3	515.3/ 844.4	>90%	SAM019		572.2/ 901.3	572.3/ 901.4	88%
SAM007		540.2/ 869.3	540.3/ 869.4	>95%	SAM020		516.2/ 845.3	516.3/ 845.4	>95%
SAM008		548.2/ 877.3	570.3/ 877.3	>95%	SAM021		578.1/ 907.2	578.2/ 907.3	>90%
SAM009		490.2/ 819.3	490.2/ 819.3	>95%	SAM022		544.3/ 873.3	544.4/ 873.5	64%
SAM010^[b]		488.2/ 817.3	492.1	0%	SAM023^[d]		593.2/ 922.3	593.3/ 922.3	5% 32%
SAM011		502.2/ 831.3	502.2/ 831.3	>95%	SAM024^[b]		512.2/ 841.3	635.4/ 1100.2	0%
SAM012^[b]		578.3/ 907.3	524.2/ 831.3	0%	SAM025		486.2/ 815.3	486.2/ 815.3	>95%
SAM013		517.3/ 846.3	517.3/ 846.4	61%	SAM026		484.2/ 813.2	484.2/ 813.3	>95%

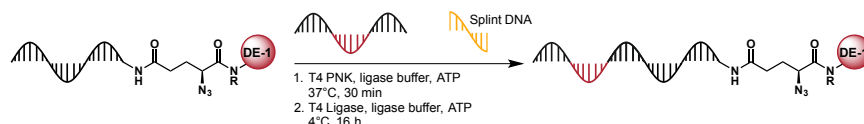
[a] Sodium adducts were often observed as main mass. [b] Unknown product(s). [c] Diastereomers were found. [d] upper value: DMTMM conditions, lower value: EDC/HOAt/DIPEA reaction conditions. [e] *t*Bu ester hydrolysis in ESI-MS also observed [f] Boc removal under ESI-MS conditions.

DE-1 Introduction into the Small Molecule Library, Split Synthesis



In low bind 96 U-shaped well-plates (Eppendorf) 150 evaluated building blocks (114 amino acids, 21 primary amines and 15 secondary amines) for library synthesis were treated according to the following reaction conditions. **2-27b** (7.7 μL , 65.5 μM in MOPS buffer, 1.0 eq.), the DE-1 building block (10.1 μL , 50.0 mM in DMSO, 1000.0 eq.), DMTMM-BF₄ (5.0 μL , 100.0 mM in DMSO, 1000.0 eq.) and NMM (8.4 μL , 180.0 mM in DMSO, 3000.0 eq.) were mixed and the plates were left standing at RT for 20 h. The reactions were purified by EtOH precipitation. The DNA pellets were redissolved in nuclease-free H₂O (50 μL per well) yielding a theoretical concentration of 10.1 μM . For building blocks containing phenolic groups (4 in total) DIPEA was used instead of NMM and EDC hydrochloride (200.0 mM, 2.6 μL) and HOAt (200.0 mM, 2.6 μL) replaced the DMTMM-BF₄.

Encoding of the DE-1 Building Blocks by Splint Ligation



DNA Phosphorylation:

In PCR 96-well plates (Eppendorf, Low-Bind) encoding DNA with the sequence 5'-GTAGTTGGA TCCGCACACYYYYGACAATTCACACACGTCC-3' with YYYY representing the building block codon (10.1 μL , 100.0 μM in H₂O, 2.0 eq.) was dissolved in DNase free H₂O (41 μL) and 10X ligase buffer (10.0 μL) was added, followed by ATP (20.0 μL , 10.0 mM in H₂O, 396.5 eq.) and the T4 PNK (19.0 μL , 1 U/ μL in 1X ligase buffer, 19 U). The mixture was incubated at 37°C for 30 min and the reaction was stopped by heating to 65°C for 20 min.

Ligation:

The phosphorylated DNA was mixed with the DNA-SM-library (50.0 μL , 10.1 μM in H₂O). H₂O (8 μL), splint DNA with the sequence 5'-ATCCAATACTAAGACTGTG-3' (9.5 μL , 80.0 μM in H₂O, 1.5 eq.) and 10X ligase buffer (10.0 μL) were added and the solution was incubated at 65°C for 5 min. After cooling to RT ATP (20.0 μL , 10.0 mM in H₂O, 396.5 eq.) was added and after further cooling on ice the reaction was started by the addition of T4 DNA ligase (2.5 μL , 400 U/ μL , 1000 U). The well-plates were kept in the fridge at 4°C for 16 h. The ligations were stopped by heating to 65°C for 10 min. The encoding performance was analyzed by 12% denaturing PAGE (150V, TBE, 65 min, SYBR Gold). All wells were combined and each well was subsequently washed with H₂O (40 μL). The combined reactions and washing phases were lyophilized and redissolved in H₂O (8 mL). The encoded sub-library was purified by EtOH precipitation. The obtained DNA pellet was dissolved in H₂O (3.6 mL), filtered by syringe filter and purified by preparative reversed-phase HPLC (Method C). The obtained fractions were analyzed by denaturing PAGE (12%, 150V, 65 min, SYBR gold) and library containing fractions were combined, lyophilized and redissolved in 500 mM TEAA buffer (2.1 mL) to yield a 16 μM encoded library solution.

Note: The used 10X ligase buffer was freshly prepared (no commercial ligase buffer due to DTT!) with the following final concentrations: 500 mM Tris-HCl, 100 mM MgCl₂, pH 7.5.

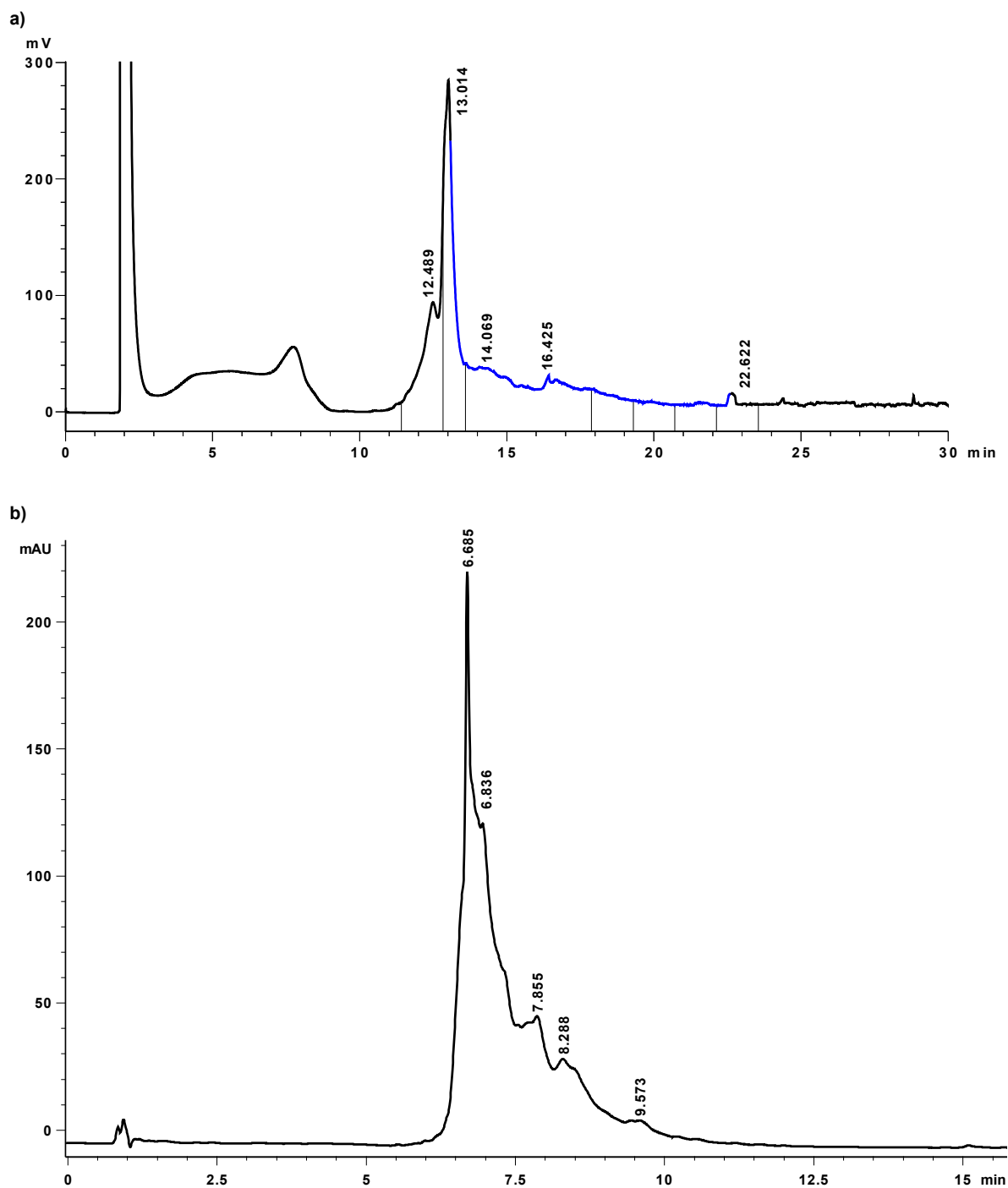
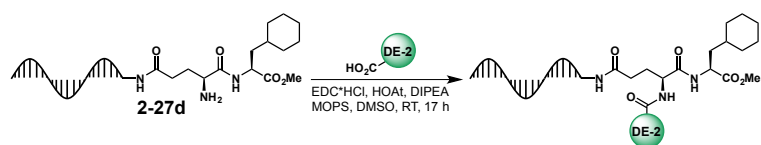


Figure 113. **a)** Chromatogram of the preparative small molecule library purification. Blue coloring indicates the library containing fractions. **b)** HPLC analysis of the purified small molecule library after DE-1 encoding.

DE-2 Carboxylic Acids Validation

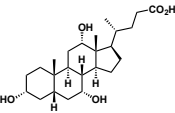












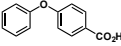
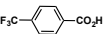

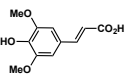
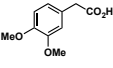







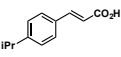

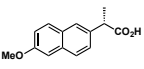

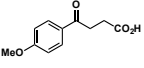





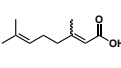





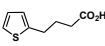




In a U-shaped 96 well-plate 39 carboxylic acids screening reactions were conducted with the following conditions. **2-27d** (1.5 μ L, 34.0 μ M in MOPS buffer, 1.0 eq.) was mixed with the carboxylic acid (0.5 μ L, 100.0 mM in DMSO, 1000.0 eq.), EDC hydrochloride (1.0 μ L, 50.0 mM in

DMSO, 1000.0 eq.), HOAt (1.0 μ L, 50.0 mM in DMSO, 1000.0 eq.) and DIPEA (1.5 μ L, 50.0 mM in DMSO, 1500.0 eq.). The reactions were left standing at RT for 17 h. The reactions were purified by EtOH precipitation and were analyzed by LC-MS. 19 carboxylic acids were considered unsuitable for library construction due to low yields and purities. 89 carboxylic acids were chosen for further library synthesis. Cut-off: >50% purity.

Table 15. List of all carboxylic acid DE-2 building blocks with the screening assay results. 39 representative members were screened for their coupling efficiency. Red coloring indicates insufficient reaction purity. Blue coloring indicates compounds that were excluded from library synthesis due to their similarity to low yielding building blocks. These compounds were supposed to yield similarly low conversions.

No.	Structure	Calc. Mass [m/z]	Obs. Mass [m/z] ^[a]	Purity	No.	Structure	Calc. Mass [m/z]	Obs. Mass [m/z] ^[a]	Purity
CA001		654.3/ 983.4	676.4/ 983.4	>90%	CA055				
CA002					CA056				
CA003		630.3/ 959.4	630.4/ 959.4	>90%	CA057				
CA004		679.3/ 1008.4	679.4/ 1008.4	>90%	CA058		762.2/ 1091.2	762.3/ 1091.3	79%
CA005		740.3/ 1069.4	740.4/ 1069.4	>90%	CA059 ^[d]		684.2/ 1013.3	574.4/ 903.5	0%
CA006		674.3/ 1003.4	674.5/ 1003.5	90%	CA060				
CA007		684.3/ 1013.3	684.4/ 1013.4	90%	CA061				
CA008					CA062				
CA009					CA063 ^[c]		706.3/ 1035.4	613.3/ 975.4	0%
CA010		682.3/ 1011.4	682.4/ 1011.5	70%	CA064		685.3/ 1014.4	685.4/ 1014.5	>90%
CA011					CA065				
CA012					CA066				
CA013		626.3/ 955.3	626.4	56%	CA067				
CA014		757.3/ 1086.3	757.4/ 1086.4	>90%	CA068				
CA015					CA069 ^[d]		784.3/ 1113.3	574.4/ 903.5	0%
CA016					CA070				

CA017 ^[b]		964.6/ 1293.6	932.7/ 1261.7	>90%	CA071				
CA018		738.4/ 1067.4	738.5/ 1067.5	78%	CA072				
CA019					CA073		784.4/ 1113.4	784.5/ 1113.5	48%
CA020		672.4/ 1001.4	672.5/ 1001.5	78%	CA074				
CA021					CA075				
CA022		658.3/ 987.4	658.4/ 987.5	82%	CA076				
CA023					CA077				
CA024					CA078				
CA025 ^[c]		780.3/ 1109.4	935.7/ 1264.6	0%	CA079				
CA026 ^[d]		688.3/ 1017.4	574.4/ 903.5	0%	CA080				
CA027					CA081				
CA028					CA082				
CA029					CA083				
CA030		778.3/ 1107.4	778.4/ 1107.5	28%	CA084				
CA031					CA085				
CA032					CA086				
CA033					CA087				
CA034 ^[d]		712.3/ 1041.3	574.4/ 903.5	0%	CA088				
CA035					CA089				
CA036					CA090		698.3/ 1027.3	698.4/ 1027.4	54%
CA037		779.3/ 1108.4	779.5/ 1108.5	16%	CA091				
CA038					CA092		726.3/ 1055.4	726.5/ 1055.4	48%

CA039		726.4/ 1055.5	726.6/ 1055.5	61%	CA093				
CA040					CA094				
CA041		748.3/ 1077.4	748.5/ 1077.5	<20%	CA095				
CA042		694.3/ 1023.4	694.5/ 1023.4	>90%	CA096		823.3/ 1152.3	823.4/ 1152.4	84%
CA043					CA097				
CA044		713.3/ 1042.3	713.4/ 1042.5	69%	CA098		785.3/ 1114.3	785.4/ 1114.5	>90%
CA045		736.4/ 1065.4	736.5/ 1065.5	86%	CA099				
CA046					CA100		747.3/ 1076.4	747.4/ 1076.5	>90%
CA047					CA101				
CA048 ^[d]		708.3/ 1037.4	574.4/ 903.5	0%	CA102		718.3/ 1047.4	718.5/ 1047.5	>90%
CA049					CA103				
CA050					CA104				
CA051		695.3/ 1024.4	695.5/ 1024.5	43%	CA105				
CA052					CA106 ^[d]		836.2/ 1165.3	574.4/ 903.5	0%
CA053					CA107				
CA054		721.4/ 1050.4	721.4/ 1050.5	67%	CA108				

[a] Sodium adducts were often observed. [b] Fragmentation in the ESI was observed. [c] Unknown product. [d] Starting material.

14. References

- [1] W. R. Algar, *Chemoselective and Bioorthogonal Ligation Reactions*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2017**.
- [2] O. Boutureira, G. J. L. Bernardes, *Chem. Rev.* **2015**, *115*, 2174–2195.
- [3] N. Voloshchuk, J. K. Montclare, *Mol. Biosyst.* **2010**, *6*, 65–80.
- [4] K. Lang, J. W. Chin, *Chem. Rev.* **2014**, *114*, 4764–4806.
- [5] A. Strømgaard, A. A. Jensen, K. Strømgaard, *ChemBioChem* **2004**, *5*, 909–916.
- [6] J. D. WATSON, F. H. C. CRICK, *Nature* **1953**, *171*, 737–738.
- [7] N. Krall, F. P. da Cruz, O. Boutureira, G. J. L. Bernardes, *Nat. Chem.* **2016**, *8*, 103–113.
- [8] J. Kalia, R. T. Raines, *Bioorganic Med. Chem. Lett.* **2007**, *17*, 6286–6289.
- [9] A. Pal, M. Bérubé, D. G. Hall, *Angew. Chemie Int. Ed.* **2010**, *49*, 1492–1495.
- [10] S. Saito, T. L. Massie, T. Maeda, H. Nakazumi, C. L. Colyer, *Sensors* **2012**, *12*, 5420–5431.
- [11] G. F. Whyte, R. Vilar, R. Woscholski, *J. Chem. Biol.* **2013**, *6*, 161–174.
- [12] B. Kong, A. Zhu, Y. Luo, Y. Tian, Y. Yu, G. Shi, *Angew. Chemie Int. Ed.* **2011**, *50*, 1837–1840.
- [13] K. E. Secor, T. E. Glass, *Org. Lett.* **2004**, *6*, 3727–3730.
- [14] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chemie Int. Ed.* **2001**, *40*, 2004–2021.
- [15] J. E. Moses, A. D. Moorhouse, *Chem. Soc. Rev.* **2007**, *36*, 1249–1262.
- [16] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chemie Int. Ed.* **2002**, *41*, 2596–2599.
- [17] R. Huisgen, *Angew. Chemie Int. Ed. English* **1963**, *2*, 565–598.
- [18] R. Huisgen, *Angew. Chemie Int. Ed. English* **1963**, *2*, 633–645.
- [19] C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057–3064.
- [20] M. Meldal, C. W. Tornøe, *Chem. Rev.* **2008**, *108*, 2952–3015.
- [21] J. Martell, E. Weerapana, *Molecules* **2014**, *19*, 1378–1393.
- [22] V. K. Tiwari, B. B. Mishra, K. B. Mishra, N. Mishra, A. S. Singh, X. Chen, *Chem. Rev.* **2016**, *116*, 3086–3240.
- [23] F. Amblard, J. H. Cho, R. F. Schinazi, *Chem. Rev.* **2009**, *109*, 4207–4220.
- [24] V. Hong, N. F. Steinmetz, M. Manchester, M. G. Finn, *Bioconjug. Chem.* **2010**, *21*, 1912–1916.
- [25] C. S. McKay, M. G. Finn, *Chem. Biol.* **2014**, *21*, 1075–1101.
- [26] V. O. Rodionov, S. I. Presolski, S. Gardinier, Y.-H. Lim, M. G. Finn, *J. Am. Chem. Soc.* **2007**, *129*, 12696–12704.
- [27] T. R. Chan, R. Hilgraf, K. B. Sharpless, V. V. Fokin, *Org. Lett.* **2004**, *6*, 2853–2855.
- [28] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.
- [29] J. Dommerholt, F. P. J. T. Rutjes, F. L. van Delft, *Top. Curr. Chem.* **2016**, *374*, 16.
- [30] H. Wang, M. Gauthier, J. R. Kelly, R. J. Miller, M. Xu, W. D. O'Brien, J. Cheng, *Angew. Chemie Int. Ed.* **2016**, *55*, 5452–5456.
- [31] J. Zayas, M. Annoual, J. K. Das, Q. Felty, W. G. Gonzalez, J. Miksovská, N. Sharifai, A. Chiba, S. F. Wnuk, *Bioconjug. Chem.* **2015**, *26*, 1519–1532.
- [32] C. J. Pickens, S. N. Johnson, M. M. Pressnall, M. A. Leon, C. J. Berkland, *Bioconjug. Chem.* **2018**, *29*, 686–701.
- [33] A. Darko, S. Wallace, O. Dmitrenko, M. M. Machovina, R. A. Mehl, J. W. Chin, J. M. Fox, *Chem. Sci.* **2014**, *5*, 3770–3776.
- [34] B. L. Oliveira, Z. Guo, G. J. L. Bernardes, *Chem. Soc. Rev.* **2017**, *46*, 4895–4950.
- [35] M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.
- [36] R. Rossin, S. M. van den Bosch, W. ten Hoeve, M. Carvelli, R. M. Versteegen, J. Lub, M. S. Robillard, *Bioconjug. Chem.* **2013**, *24*, 1210–1217.
- [37] M. R. Karver, R. Weissleder, S. A. Hilderbrand, *Bioconjug. Chem.* **2011**, *22*, 2263–2270.
- [38] N. K. Devaraj, R. Weissleder, *Acc. Chem. Res.* **2011**, *44*, 816–827.
- [39] E. M. Sletten, C. R. Bertozzi, *Angew. Chemie Int. Ed.* **2009**, *48*, 6974–6998.
- [40] J. M. Chalker, C. S. C. Wood, B. G. Davis, *J. Am. Chem. Soc.* **2009**, *131*, 16346–16347.
- [41] Z. Gao, V. Gouverneur, B. G. Davis, *J. Am. Chem. Soc.* **2013**, *135*, 13612–13615.
- [42] L. Lercher, J. F. McGouran, B. M. Kessler, C. J. Schofield, B. G. Davis, *Angew. Chemie - Int. Ed.* **2013**, *52*, 10553–10558.
- [43] H. Noda, G. Erős, J. W. Bode, *J. Am. Chem. Soc.* **2014**, *136*, 5611–5614.
- [44] L. Reguera, Y. Méndez, A. R. Humpierre, O. Valdés, D. G. Rivera, *Acc. Chem. Res.* **2018**, *51*, 1475–1486.
- [45] F. Saito, H. Noda, J. W. Bode, *ACS Chem. Biol.* **2015**, *10*, 1026–1033.
- [46] D. M. Patterson, L. A. Nazarova, J. A. Prescher, *ACS Chem. Biol.* **2014**, *9*, 592–605.
- [47] A. Holmberg, A. Blomstergren, O. Nord, M. Lukacs, J. Lundberg, M. Uhlén, *Electrophoresis* **2005**, *26*, 501–510.
- [48] W. P. Jencks, in *Prog. Phys. Org. Chem.*, **1964**, pp. 63–128.
- [49] N. J. Fina, J. O. Edwards, *Int. J. Chem. Kinet.* **1973**, *5*, 1–26.
- [50] E. Buncel, I.-H. Um, *Tetrahedron* **2004**, *60*, 7801–7825.
- [51] E. H. Cordes, W. P. Jencks, *J. Am. Chem. Soc.* **1962**, *84*, 4319–4328.

- [52] A. Dirksen, T. M. Hackeng, P. E. Dawson, *Angew. Chemie Int. Ed.* **2006**, *45*, 7581–7584.
- [53] M. Rashidian, M. M. Mahmoodi, R. Shah, J. K. Dozier, C. R. Wagner, M. D. Distefano, *Bioconjug. Chem.* **2013**, *24*, 333–342.
- [54] P. Crisalli, E. T. Kool, *J. Org. Chem.* **2013**, *78*, 1184–1189.
- [55] P. Crisalli, E. T. Kool, *Org. Lett.* **2013**, *15*, 1646–1649.
- [56] E. T. Kool, D. Park, P. Crisalli, *J. Am. Chem. Soc.* **2013**, *135*, 17663–17666.
- [57] D. K. Kölmel, E. T. Kool, *Chem. Rev.* **2017**, *117*, 10358–10376.
- [58] O. Dilek, A. Sorrentino, S. Bane, *Synlett* **2016**, *27*, 1335–1338.
- [59] S. Wang, G. N. Nawale, S. Kadekar, O. P. Oommen, N. K. Jena, S. Chakraborty, J. Hilborn, O. P. Varghese, *Sci. Rep.* **2018**, *8*, 2193.
- [60] P. Schmidt, L. Zhou, K. Tishinov, K. Zimmermann, D. Gillingham, *Angew. Chemie - Int. Ed.* **2014**, *53*, 10928–10931.
- [61] P. M. S. D. Cal, J. B. Vicente, E. Pires, A. V. Coelho, L. F. Veiros, C. Cordeiro, P. M. P. Gois, *J. Am. Chem. Soc.* **2012**, *134*, 10299–10305.
- [62] A. Bandyopadhyay, J. Gao, *Chem. - A Eur. J.* **2015**, *21*, 14748–14752.
- [63] P. Schmidt, C. Stress, D. Gillingham, *Chem. Sci.* **2015**, *6*, 3329–3333.
- [64] C. J. Stress, P. J. Schmidt, D. G. Gillingham, *Org. Biomol. Chem.* **2016**, *14*, 5529–5533.
- [65] O. Dilek, Z. Lei, K. Mukherjee, S. Bane, *Chem. Commun.* **2015**, *51*, 16992–16995.
- [66] H. Gu, T. I. Chio, Z. Lei, R. J. Staples, J. S. Hirschi, S. Bane, *Org. Biomol. Chem.* **2017**, *15*, 7543–7548.
- [67] P. Tschampel, H. R. Snyder, *J. Org. Chem.* **1964**, *29*, 2168–2172.
- [68] M. J. S. Dewar, R. C. Dougherty, *J. Am. Chem. Soc.* **1964**, *86*, 433–436.
- [69] P. D. Robinson, M. P. Groziak, L. Chen, *Acta Crystallogr. Sect. C Cryst. Struct. Commun.* **1998**, *54*, 71–73.
- [70] M. P. Groziak, L. Chen, L. Yi, P. D. Robinson, *J. Am. Chem. Soc.* **1997**, *119*, 7817–7826.
- [71] M. A. Grassberger, F. Turnowsky, J. Hildebrandt, *J. Med. Chem.* **1984**, *27*, 947–953.
- [72] M. C. Davis, S. G. Franzblau, A. R. Martin, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 843–846.
- [73] P. D. Robinson, M. P. Groziak, *Acta Crystallogr. Sect. C Cryst. Struct. Commun.* **1999**, *55*, 1701–1704.
- [74] D. Kanichar, L. Roppiyakuda, E. Kosmowska, M. A. Faust, K. P. Tran, F. Chow, E. Buglo, M. P. Groziak, E. A. Sarina, M. M. Olmstead, et al., *Chem. Biodivers.* **2014**, *11*, 1381–1397.
- [75] A. Bandyopadhyay, S. Cambray, J. Gao, *J. Am. Chem. Soc.* **2017**, *139*, 871–878.
- [76] D. Gillingham, *Org. Biomol. Chem.* **2016**, *14*, 7606–7609.
- [77] I. Sarkar, A. K. Mishra, *Appl. Spectrosc. Rev.* **2018**, *53*, 552–601.
- [78] V. I. Martynov, A. A. Pakhomov, N. V. Popova, I. E. Deyev, A. G. Petrenko, *Acta Naturae* **2016**, *8*, 33–46.
- [79] T. Terai, T. Nagano, *Pflügers Arch. - Eur. J. Physiol.* **2013**, *465*, 347–359.
- [80] B. Z. Tang, K. M. Solntsev, *J. Mater. Chem. C* **2016**, *4*, 2638–2639.
- [81] T. Kowada, H. Maeda, K. Kikuchi, *Chem. Soc. Rev.* **2015**, *44*, 4953–4972.
- [82] O. Dilek, S. Bane, *Chemosensors* **2016**, *4*, 5.
- [83] S. Hachiya, T. Inagaki, D. Hashizume, S. Maki, H. Niwa, T. Hirano, *Tetrahedron Lett.* **2010**, *51*, 1613–1615.
- [84] S. Cambray, A. Bandyopadhyay, J. Gao, *Chem. Commun.* **2017**, *53*, 12532–12535.
- [85] J. Kalia, R. T. Raines, *Angew. Chemie Int. Ed.* **2008**, *47*, 7523–7526.
- [86] B. E. Collins, S. Sorey, A. E. Hargrove, S. H. Shabbir, V. M. Lynch, E. V. Anslyn, *J. Org. Chem.* **2009**, *74*, 4055–4060.
- [87] A. Adamczyk-Woźniak, K. M. Borys, I. D. Madura, A. Pawełko, E. Tomecka, K. Żukowski, *New J. Chem.* **2013**, *37*, 188–194.
- [88] J. Yan, G. Springsteen, S. Deeter, B. Wang, *Tetrahedron* **2004**, *60*, 11205–11209.
- [89] A. J. J. Lennox, G. C. Lloyd-Jones, *J. Am. Chem. Soc.* **2012**, *134*, 7431–7441.
- [90] J. A. Gonzalez, O. M. Ogba, G. F. Morehouse, N. Rosson, K. N. Houk, A. G. Leach, P. H. Y. Cheong, M. D. Burke, G. C. Lloyd-Jones, *Nat. Chem.* **2016**, *8*, 1067–1075.
- [91] G. Vidarsson, G. Dekkers, T. Rispens, *Front. Immunol.* **2014**, *5*, 1–17.
- [92] H. W. Schroeder, L. Cavacini, *J. Allergy Clin. Immunol.* **2010**, *125*, S41–S52.
- [93] N. Jain, S. W. Smith, S. Ghone, B. Tomczuk, *Pharm. Res.* **2015**, *32*, 3526–3540.
- [94] Z. Gao, Y. Hao, M. Zheng, Y. Chen, *RSC Adv.* **2017**, *7*, 7604–7609.
- [95] W. C. Still, M. Kahn, A. Mitra, *J. Org. Chem.* **1978**, *43*, 2923–2925.
- [96] E. Sella, D. Shabat, *Org. Biomol. Chem.* **2013**, *11*, 5074.
- [97] P.-C. Lin, S.-H. Ueng, M.-C. Tseng, J.-L. Ko, K.-T. Huang, S.-C. Yu, A. K. Adak, Y.-J. Chen, C.-C. Lin, *Angew. Chemie Int. Ed.* **2006**, *45*, 4286–4290.
- [98] R. Robinson, M. L. Tomlinson, *J. Chem. Soc.* **1934**, *1*, 1524.
- [99] Y. Murakami, T. Watanabe, H. Takahashi, H. Yokoo, Y. Nakazawa, M. Koshimizu, N. Adachi, M. Kurita, T. Yoshino, T. Inagaki, et al., *Tetrahedron* **1998**, *54*, 45–64.
- [100] Y. Xia, K. Cao, Y. Zhou, M. R. K. Alley, F. Rock, M. Mohan, M. Meewan, S. J. Baker, S. Lux, C. Z. Ding, et al., *Bioorg. Med. Chem. Lett.* **2011**, *21*, 2533–2536.
- [101] M. Barbasiewicz, M. Michalak, K. Grela, *Chem. - A Eur. J.* **2012**, *18*, 14237–14241.

- [102] J. Cody, C. J. Fahmi, *Tetrahedron* **2004**, *60*, 11099–11107.
- [103] T. Kodadek, *Chem. Commun.* **2011**, *47*, 9757–9763.
- [104] R. Liu, X. Li, K. S. Lam, *Curr. Opin. Chem. Biol.* **2017**, *38*, 117–126.
- [105] H. M. Geysen, R. H. Meloen, S. J. Barteling, *Proc. Natl. Acad. Sci.* **1984**, *81*, 3998–4002.
- [106] R. A. Houghten, *Proc. Natl. Acad. Sci.* **1985**, *82*, 5131–5135.
- [107] K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, *354*, 82–84.
- [108] G. Smith, *Science (80-)*. **1985**, *228*, 1315–1317.
- [109] S. Brenner, R. A. Lerner, *Proc. Natl. Acad. Sci.* **1992**, *89*, 5381–5383.
- [110] N. van Hilten, F. Chevillard, P. Kolb, *J. Chem. Inf. Model.* **2019**, *in Press*, DOI 10.1021/acs.jcim.8b00737.
- [111] L. Hoffer, Y. V. Voitovich, B. Raux, K. Carrasco, C. Muller, A. Y. Fedorov, C. Derviaux, A. Amouric, S. Betzi, D. Horvath, et al., *J. Med. Chem.* **2018**, *61*, 5719–5732.
- [112] R. MacArron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, et al., *Nat. Rev. Drug Discov.* **2011**, *10*, 188–195.
- [113] J. R. Broach, J. Thorner, *Nature* **1996**, *384*, 14–16.
- [114] J. Inglese, R. L. Johnson, A. Simeonov, M. Xia, W. Zheng, C. P. Austin, D. S. Auld, *Nat. Chem. Biol.* **2007**, *3*, 466–479.
- [115] F. Fan, K. V. Wood, *Assay Drug Dev. Technol.* **2007**, *5*, 127–136.
- [116] B. N. G. Giepmans, S. R. Adams, M. H. Ellisman, R. Y. Tsien, *Science (80-)*. **2006**, *312*, 217–224.
- [117] C. M. Hammers, J. R. Stanley, *J. Invest. Dermatol.* **2014**, *134*, 1–5.
- [118] J. Bazan, I. Calkosiński, A. Gamian, *Hum. Vaccin. Immunother.* **2012**, *8*, 1817–1828.
- [119] D. Neri, R. A. Lerner, *Annu. Rev. Biochem.* **2018**, *87*, 479–502.
- [120] R. Liu, X. Li, W. Xiao, K. S. Lam, *Adv. Drug Deliv. Rev.* **2017**, *110–111*, 13–37.
- [121] K. Deyle, X. D. Kong, C. Heinis, *Acc. Chem. Res.* **2017**, *50*, 1866–1874.
- [122] S. S. Kale, C. Villequey, X. D. Kong, A. Zorzi, K. Deyle, C. Heinis, *Nat. Chem.* **2018**, *10*, 715–723.
- [123] R. A. Lerner, S. Brenner, *Angew. Chemie - Int. Ed.* **2017**, *56*, 1164–1165.
- [124] J. Nielsen, S. Brenner, K. D. Janda, *J. Am. Chem. Soc.* **1993**, *115*, 9812–9813.
- [125] M. C. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett, M. A. Gallop, *Proc. Natl. Acad. Sci.* **1993**, *90*, 10700–10704.
- [126] X. Li, D. R. Liu, *Angew. Chemie - Int. Ed.* **2004**, *43*, 4848–4870.
- [127] S. Melkko, J. Scheuermann, C. E. Dumelin, D. Neri, *Nat. Biotechnol.* **2004**, *22*, 568–574.
- [128] D. R. Halpin, P. B. Harbury, *PLoS Biol.* **2004**, *2*, 1015–1021.
- [129] N. Favalli, G. Bassi, J. Scheuermann, D. Neri, *FEBS Lett.* **2018**, *592*, 2168–2180.
- [130] Z. J. Gartner, D. R. Liu, *J. Am. Chem. Soc.* **2001**, *123*, 6961–6963.
- [131] A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493.
- [132] Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu, *Science (80-)*. **2004**, *305*, 1601–1605.
- [133] B. N. Tse, T. M. Snyder, Y. Shen, D. R. Liu, *J. Am. Chem. Soc.* **2008**, *130*, 15611–15626.
- [134] D. L. Usanov, A. I. Chan, J. P. Maianti, D. R. Liu, *Nat. Chem.* **2018**, *10*, 704–714.
- [135] Y. Li, P. Zhao, M. Zhang, X. Zhao, X. Li, *J. Am. Chem. Soc.* **2013**, *135*, 17727–17730.
- [136] M. H. Hansen, P. Blakskjær, L. K. Petersen, T. H. Hansen, J. W. Højfeldt, K. V. Gothelf, N. J. V. Hansen, *J. Am. Chem. Soc.* **2009**, *131*, 1322–1327.
- [137] Y. Li, R. De Luca, S. Cazzamalli, F. Pretto, D. Bajic, J. Scheuermann, D. Neri, *Nat. Chem.* **2018**, *10*, 441–448.
- [138] F. Buller, M. Steiner, K. Frey, D. Mircsof, J. Scheuermann, M. Kalisch, P. Bühlmann, C. T. Supuran, D. Neri, *ACS Chem. Biol.* **2011**, *6*, 336–344.
- [139] M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuzzo, et al., *Nat. Chem. Biol.* **2009**, *5*, 647–654.
- [140] G. Zimmermann, D. Neri, *Drug Discov. Today* **2016**, *21*, 1828–1834.
- [141] S. Melkko, Y. Zhang, C. E. Dumelin, J. Scheuermann, D. Neri, *Angew. Chemie - Int. Ed.* **2007**, *46*, 4671–4674.
- [142] M. Wichert, N. Krall, W. Decurtins, R. M. Franzini, F. Pretto, P. Schneider, D. Neri, J. Scheuermann, *Nat. Chem.* **2015**, *7*, 241–249.
- [143] C. Zambaldo, S. Barluenga, N. Winssinger, *Curr. Opin. Chem. Biol.* **2015**, *26*, 8–15.
- [144] F. V. Reddavid, W. Lin, S. Lehnert, Y. Zhang, *Angew. Chemie - Int. Ed.* **2015**, *54*, 7924–7928.
- [145] M. L. Malone, B. M. Paegel, *ACS Comb. Sci.* **2016**, *18*, 182–187.
- [146] R. M. Franzini, F. Samain, M. Abd Elrahman, G. Mikutis, A. Nauer, M. Zimmermann, J. Scheuermann, J. Hall, D. Neri, *Bioconjug. Chem.* **2014**, *25*, 1453–1461.
- [147] Y. Li, E. Gabriele, F. Samain, N. Favalli, F. Sladojevich, J. Scheuermann, D. Neri, *ACS Comb. Sci.* **2016**, *18*, 438–443.
- [148] Y. Ding, G. J. Franklin, J. L. DeLorey, P. A. Centrella, S. Mataruse, M. A. Clark, S. R. Skinner, S. Belyanskaya, *ACS Comb. Sci.* **2016**, *18*, 625–629.
- [149] F. Buller, L. Mannocci, Y. Zhang, C. E. Dumelin, J. Scheuermann, D. Neri, *Bioorganic Med. Chem. Lett.* **2008**, *18*,

- 5926–5931.
- [150] H. Li, Z. Sun, W. Wu, X. Wang, M. Zhang, X. Lu, W. Zhong, D. Dai, *Org. Lett.* **2018**, *20*, 7186–7191.
- [151] Y. Chen, A. S. Kamlet, J. B. Steinman, D. R. Liu, *Nat. Chem.* **2011**, *3*, 146–153.
- [152] L. Fan, C. P. Davie, *ChemBioChem* **2017**, *18*, 843–847.
- [153] D. K. Kölmel, R. P. Loach, T. Knauber, M. E. Flanagan, *ChemMedChem* **2018**, *13*, 2159–2165.
- [154] Y. Ruff, F. Berst, *Medchemcomm* **2018**, *9*, 1188–1193.
- [155] J. Wang, H. Lundberg, S. Asai, P. Martín-Acosta, J. S. Chen, S. Brown, W. Farrell, R. G. Dushin, C. J. O'Donnell, A. S. Ratnayake, et al., *Proc. Natl. Acad. Sci.* **2018**, *115*, E6404–E6410.
- [156] K. Pels, P. Dickson, H. An, T. Kodadek, *ACS Comb. Sci.* **2018**, *20*, 61–69.
- [157] K. Shu, T. Kodadek, *ACS Comb. Sci.* **2018**, *20*, 277–281.
- [158] N. Tran-Hoang, T. Kodadek, *ACS Comb. Sci.* **2018**, *20*, 55–60.
- [159] C. A. Lipinski, *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249.
- [160] J. W. Cuzzo, P. A. Centrella, D. Gikunju, S. Habeshian, C. D. Hupp, A. D. Keefe, E. A. Sigel, H. H. Soutter, H. A. Thomson, Y. Zhang, et al., *ChemBioChem* **2017**, *18*, 864–871.
- [161] Z. Zhu, A. Shaginian, L. C. Grady, T. O'Keeffe, X. E. Shi, C. P. Davie, G. L. Simpson, J. A. Messer, G. Evindar, R. N. Bream, et al., *ACS Chem. Biol.* **2018**, *13*, 53–59.
- [162] M. Leimbacher, Y. Zhang, L. Mannocci, M. Stravs, T. Geppert, J. Scheuermann, G. Schneider, D. Neri, *Chem. - A Eur. J.* **2012**, *18*, 7729–7737.
- [163] L. Mannocci, S. Melkko, F. Buller, I. Molnár, J. P. Gapian Bianké, C. E. Dumelin, J. Scheuermann, D. Neri, *Bioconjug. Chem.* **2010**, *21*, 1836–1841.
- [164] W. Decurtins, M. Wichert, R. M. Franzini, F. Buller, M. A. Stravs, Y. Zhang, D. Neri, J. Scheuermann, *Nat. Protoc.* **2016**, *11*, 764–780.
- [165] Y. Li, G. Zimmermann, J. Scheuermann, D. Neri, *ChemBioChem* **2017**, *18*, 848–852.
- [166] J. Bao, S. M. Krylova, L. T. Cherney, R. L. Hale, S. L. Belyanskaya, C. H. Chiu, A. Shaginian, C. C. Arico-Muendel, S. N. Krylov, *Anal. Chem.* **2016**, *88*, 5498–5506.
- [167] S. Kochmann, A. T. H. Le, R. Hili, S. N. Krylov, *Electrophoresis* **2018**, *39*, 2991–2996.
- [168] L. M. McGregor, D. J. Gorin, C. E. Dumelin, D. R. Liu, *J. Am. Chem. Soc.* **2010**, *132*, 15522–15524.
- [169] L. M. McGregor, T. Jain, D. R. Liu, *J. Am. Chem. Soc.* **2014**, *136*, 3264–3270.
- [170] K. E. Denton, C. J. Krusemark, *Medchemcomm* **2016**, *7*, 2020–2027.
- [171] B. Shi, Y. Deng, P. Zhao, X. Li, *Bioconjug. Chem.* **2017**, *28*, 2293–2301.
- [172] L. Mannocci, Y. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko, D. Neri, *Proc. Natl. Acad. Sci.* **2008**, *105*, 17670–17675.
- [173] G. Zimmermann, Y. Li, U. Rieder, M. Mattarella, D. Neri, J. Scheuermann, *ChemBioChem* **2017**, *18*, 853–857.
- [174] Z. Wu, T. L. Graybill, X. Zeng, M. Platchek, J. Zhang, V. Q. Bodmer, D. D. Wisnoski, J. Deng, F. T. Coppo, G. Yao, et al., *ACS Comb. Sci.* **2015**, *17*, 722–731.
- [175] R. M. Franzini, T. Ekblad, N. Zhong, M. Wichert, W. Decurtins, A. Nauer, M. Zimmermann, F. Samain, J. Scheuermann, P. J. Brown, et al., *Angew. Chemie - Int. Ed.* **2015**, *54*, 3927–3931.
- [176] A. Litovchick, C. E. Dumelin, S. Habeshian, D. Gikunju, M. A. Guié, P. Centrella, Y. Zhang, E. A. Sigel, J. W. Cuzzo, A. D. Keefe, et al., *Sci. Rep.* **2015**, *5*, 1–8.
- [177] M. Bigatti, A. Dal Corso, S. Vanetti, S. Cazzamalli, U. Rieder, J. Scheuermann, D. Neri, F. Sladojevich, *ChemMedChem* **2017**, *12*, 1748–1752.
- [178] R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai, D. R. Liu, *J. Am. Chem. Soc.* **2010**, *132*, 11779–11791.
- [179] G. Georghiou, R. E. Kleiner, M. Pulkoski-Gross, D. R. Liu, M. A. Seeliger, *Nat. Chem. Biol.* **2012**, *8*, 366–374.
- [180] W. H. Connors, S. P. Hale, N. K. Terrett, *Curr. Opin. Chem. Biol.* **2015**, *26*, 42–47.
- [181] M. Muroi, K. Haibara, M. Asai, T. Kishi, *Tetrahedron Lett.* **1980**, *21*, 309–312.
- [182] C. M. Ireland, A. R. Durso, R. A. Newman, M. P. Hacker, *J. Org. Chem.* **1982**, *47*, 1807–1811.
- [183] Š. Vítko, R. Margreiter, W. Weimar, J. Dantal, H. G. Viljoen, Y. Li, A. Jappe, N. Cretin, *Transplantation* **2004**, *78*, 1532–1540.
- [184] G. Rabbani, S. N. Ahn, *Int. J. Biol. Macromol.* **2019**, *123*, 979–990.
- [185] M. Fasano, S. Curry, E. Terreno, M. Galliano, G. Fanali, P. Narciso, S. Notari, P. Ascenzi, *IUBMB Life* **2005**, *57*, 787–796.
- [186] I. Petitpas, A. A. Bhattacharya, S. Twine, M. East, S. Curry, *J. Biol. Chem.* **2001**, *276*, 22804–22809.
- [187] A. Shibukawa, M. E. R. Rosas, T. Nakagawa, *Chromatography* **2001**, *22*, 25–31.
- [188] M. A. K. Azad, J. X. Huang, M. A. Cooper, K. D. Roberts, P. E. Thompson, R. L. Nation, J. Li, T. Velkov, *Biochem. Pharmacol.* **2012**, *84*, 278–291.
- [189] G. De Simone, C. T. Supuran, *Biochim. Biophys. Acta - Proteins Proteomics* **2010**, *1804*, 404–409.
- [190] C. Stress, B. Sauter, L. Schneider, T. Sharpe, D. Gillingham, *Angew. Chemie Int. Ed.* **2019**, DOI 10.1002/anie.201902513.
- [191] E. Jahnke, J. Weiss, S. Neuhaus, T. N. Hoheisel, H. Frauenrath, *Chem. - A Eur. J.* **2009**, *15*, 388–404.

- [192] J. E. Hein, J. C. Tripp, L. B. Krasnova, K. B. Sharpless, V. V. Fokin, *Angew. Chemie Int. Ed.* **2009**, *48*, 8018–8021.
- [193] P. G. M. Wuts, T. W. Greene, *Greene's Protective Groups in Organic Synthesis*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2006**.
- [194] K. Tishinov, K. Schmidt, D. Häussinger, D. G. Gillingham, *Angew. Chemie - Int. Ed.* **2012**, *51*, 12000–12004.
- [195] K. Tishinov, N. Fei, D. Gillingham, *Chem. Sci.* **2013**, *4*, 4401–4406.
- [196] D. Gillingham, R. Shahid, *Curr. Opin. Chem. Biol.* **2015**, *25*, 110–114.
- [197] D. R. Halpin, J. A. Lee, S. J. Wrenn, P. B. Harbury, *PLoS Biol.* **2004**, *2*, 1031–1038.
- [198] L. Schneider, Results and Analyses Were Adopted from the Work Performed during Wahlpraktikum and Master Thesis under the Supervision and Guidance of C. Stress, **2018**.
- [199] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **2012**, *64*, 4–17.
- [200] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.
- [201] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem.* **2002**, *45*, 2615–2623.
- [202] M. D. Shultz, *J. Med. Chem.* **2019**, *62*, 1701–1714.
- [203] A. Mullard, *Nat. Rev. Drug Discov.* **2018**, *17*, 777–777.
- [204] E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discov.* **2008**, *7*, 608–624.
- [205] E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. Porco, D. Kozakov, S. Vajda, A. Whitty, *Nat. Chem. Biol.* **2014**, *10*, 723–731.
- [206] A. Whitty, M. Zhong, L. Viarengo, D. Beglov, D. R. Hall, S. Vajda, *Drug Discov. Today* **2016**, *21*, 712–717.
- [207] P. Matsson, J. Kihlberg, *J. Med. Chem.* **2017**, *60*, 1662–1664.
- [208] B. Over, P. Matsson, C. Tyrchan, P. Artursson, B. C. Doak, M. A. Foley, C. Hilgendorf, S. E. Johnston, M. D. Lee, R. J. Lewis, et al., *Nat. Chem. Biol.* **2016**, *12*, 1065–1074.
- [209] P. Matsson, B. C. Doak, B. Over, J. Kihlberg, *Adv. Drug Deliv. Rev.* **2016**, *101*, 42–61.
- [210] B. Sauter, Software for the Generation of the in Silico Library Was Developed by Basilius Sauter and Was Used for the Generation of Physicochemical Analysis Plots., **2018**.
- [211] P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, et al., *Bioinformatics* **2009**, *25*, 1422–1423.
- [212] “RDKit: Open-source cheminformatics,” can be found under <http://www.rdkit.org>, **n.d.**
- [213] T. E. Oliphant, *Guide to NumPy*, Trelgol Publishing, USA, **2006**.
- [214] E. Jones, T. E. Oliphant, P. Peterson, “SciPy: Open Source Scientific Tools for Python,” can be found under <http://www.scipy.org/>, **2001**.
- [215] W. Mc Kinney, “pandas: Python Data Analysis Library,” can be found under <http://pandas.pydata.org>, **n.d.**
- [216] J. D. Hunter, *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- [217] A. H. El-Sagheer, T. Brown, *Chem. Sci.* **2014**, *5*, 253–259.
- [218] A. Shivalingam, A. E. S. Tyburn, A. H. El-Sagheer, T. Brown, *J. Am. Chem. Soc.* **2017**, *139*, 1575–1583.
- [219] H. L. Gahlon, W. B. Schweizer, S. J. Sturla, *J. Am. Chem. Soc.* **2013**, *135*, 6384–6387.
- [220] M. H. Rätz, H. R. Dexter, C. L. Millington, B. van Loon, D. M. Williams, S. J. Sturla, *Chem. Res. Toxicol.* **2016**, *29*, 1493–1503.
- [221] A. M. Maxam, W. Gilbert, *Proc. Natl. Acad. Sci.* **1977**, *74*, 560–564.
- [222] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci.* **1977**, *74*, 5463–5467.
- [223] K. Neveling, A. Hoischen, *medizinische Genet.* **2014**, *26*, 231–238.
- [224] B. Mayo, C. Rachid, A. Alegria, A. Leite, R. Peixoto, S. Delgado, *Curr. Genomics* **2014**, *15*, 293–309.
- [225] R. Kanagal-Shamanna, in *Methods Mol. Biol.*, **2016**, pp. 33–42.
- [226] R. Ke, M. Mignardi, T. Hauling, M. Nilsson, *Hum. Mutat.* **2016**, *37*, 1363–1367.
- [227] B. Sauter, Next Generation Sequencing Sample Handling, Data Processing and Results Plot Generation Was Performed by B. Sauter., **2018**.
- [228] N. Krall, F. Pretto, W. Decurtins, G. J. L. Bernardes, C. T. Supuran, D. Neri, *Angew. Chemie Int. Ed.* **2014**, *53*, 4231–4235.
- [229] D. E. Scott, C. Spry, C. Abell, in *Fragm. Drug Discov. Lessons Outlook*, Wiley-VCH Verlag GmbH & Co. KgaA, **2016**, pp. 139–172.
- [230] V. Pogacic, J. Schwaller, B. Marsden, S. Muller, P. Rellos, M. Sundstrom, O. Fedorov, A. N. Bullock, S. Knapp, *Proc. Natl. Acad. Sci.* **2007**, *104*, 20523–20528.
- [231] O. Fedorov, F. H. Niesen, S. Knapp, in *Kinase Inhib.*, Humana Press, **2012**, pp. 109–118.
- [232] D. Martinez Molina, P. Nordlund, *Annu. Rev. Pharmacol. Toxicol.* **2016**, *56*, 141–161.
- [233] M. M. Pierce, C. S. Raman, B. T. Nall, *Methods* **1999**, *19*, 213–221.
- [234] J. Li, S. G. Ballmer, E. P. Gillis, S. Fujii, M. J. Schmidt, A. M. E. Palazzolo, J. W. Lehmann, G. F. Morehouse, M. D. Burke, *Science (80-.)* **2015**, *347*, 1221–1226.
- [235] D. Morton, S. Leach, C. Cordier, S. Warriner, A. Nelson, *Angew. Chemie Int. Ed.* **2009**, *48*, 104–109.
- [236] E. D. Goddard-Borger, R. V. Stick, *Org. Lett.* **2007**, *9*, 3797–3800.

- [237] W. G. Kim, M. E. Kang, J. Bin Lee, M. H. Jeon, S. Lee, J. Lee, B. Choi, P. M. S. D. Cal, S. Kang, J.-M. Kee, et al., *J. Am. Chem. Soc.* **2017**, *139*, 12121–12124.
- [238] C. Sandford, V. K. Aggarwal, *Chem. Commun.* **2017**, *53*, 5481–5494.
- [239] J. W. B. Fyfe, A. J. B. Watson, *Chem* **2017**, *3*, 31–55.
- [240] S. Keller, C. Vargas, H. Zhao, G. Piszczek, C. A. Brautigam, P. Schuck, *Anal. Chem.* **2012**, *84*, 5066–5073.
- [241] H. Zhao, G. Piszczek, P. Schuck, *Methods* **2015**, *76*, 137–148.
- [242] E. Nordhoff, F. Kirpekar, P. Roepstorff, *Mass Spectrom. Rev.* **1996**, *15*, 67–138.
- [243] F. Landi, C. M. Johansson, D. J. Campopiano, A. N. Hulme, *Org. Biomol. Chem.* **2010**, *8*, 56–59.
- [244] T. Aubineau, J. Cossy, *Chem. Commun.* **2013**, *49*, 3303–3305.
- [245] K. D. Verma, A. Forgács, H. Uh, M. Beyerlein, M. E. Maier, S. Petoud, M. Botta, N. K. Logothetis, *Chem. - A Eur. J.* **2013**, *19*, 18011–18026.
- [246] M. H. Lacoske, J. Xu, N. Mansour, C. Gao, E. A. Theodorakis, *Org. Chem. Front.* **2015**, *2*, 388–393.
- [247] D. Amans, V. Bellosta, J. Cossy, *Chem. - A Eur. J.* **2009**, *15*, 3457–3473.
- [248] S. R. Woulfe, M. J. Miller, *J. Med. Chem.* **1985**, *28*, 1447–1453.
- [249] H. Wu, B. Yang, L. Zhu, R. Lu, G. Li, H. Lu, *Org. Lett.* **2016**, *18*, 5804–5807.
- [250] M. Zhang, K. Watanabe, M. Tsukamoto, R. Shibuya, H. Morimoto, T. Ohshima, *Chem. - A Eur. J.* **2015**, *21*, 3937–3941.
- [251] H. Wahyudi, W. Tantisantisom, X. Liu, D. M. Ramsey, E. K. Singh, S. R. McAlpine, *J. Org. Chem.* **2012**, *77*, 10596–10616.
- [252] Y. Joyard, L. Bischoff, V. Levacher, C. Papamicael, P. Vera, P. Bohn, *Lett. Org. Chem.* **2014**, *11*, 208–214.
- [253] Y. S. Lee, Y. H. Cho, S. Lee, J. K. Bin, J. Yang, G. Chae, C. H. Cheon, *Tetrahedron* **2015**, *71*, 532–538.
- [254] Y. Jeong, J. Lee, J. S. Ryu, *Bioorganic Med. Chem.* **2016**, *24*, 2114–2124.
- [255] K. Ouadahi, E. Allard, B. Oberleitner, C. Larpent, *J. Polym. Sci. Part A Polym. Chem.* **2012**, *50*, 314–328.
- [256] J. Broichhagen, J. A. Frank, N. R. Johnston, R. K. Mitchell, K. Šmid, P. Marchetti, M. Bugliani, G. A. Rutter, D. Trauner, D. J. Hodson, *Chem. Commun.* **2015**, *51*, 6018–6021.
- [257] A. Schüller, V. Hähnke, G. Schneider, *QSAR Comb. Sci.* **2007**, *26*, 407–410.

15. Acknowledgement

I would like to thank *Prof. Dr. Dennis Gillingham* for accepting me in his research group. During my projects in the course of Wahlpraktikum, Master thesis and PhD, I could strongly improve my skills and knowledge, and we always had a very nice working atmosphere.

I thank *Prof. Dr. Christof Sparr* for the co-examination of my thesis.

A big thank-you goes to all current and former PhDs, Post-Docs, scientific co-workers, Master and Bachelor students of the Gillingham group. We always had a great working atmosphere in the lab and I learned many new things from them. Apart from the life in the lab, we also spent many nice moments outside of the work environment at parties, BBQs etc.

My thesis corrections team, composed of *Basilius Sauter, Julia Hildesheim, David Lim* and *Anja Stampfli*, did a great job proof-reading this work. I highly appreciate the efforts they took to find mistakes I could not find anymore and to improve this thesis.

I would also like to thank *Jonathan James Lutabu Nduakulu, Simon Märk, Felix Raps, Lukas Schneider* and *Caspar Vogel* who helped me with my projects in the course of their practicals and Master thesis.

The people, who are running this whole department, were essential for bringing my research to a successful end. For that purpose, I would like to thank the *Werkstatt team* around *Markus Ast*, all the *secretaries, Markus Hauri, Oliver Ilg, Roy Lips* and *Bernhard Jung*.

NMR is one of the most important analytical tools in chemistry, so it is very important that these machines are always accessible and running. I would like to thank *Dr. Daniel Häussinger* and his group for providing the NMR service and for their help in NMR related problems.

I am very grateful for the MS service provided by the department of chemistry. I would like to thank *Heinz Nadig, Michael Pfeffer* and *Sylvie Mittelheiser* for conducting the high resolution MS measurements.

A big thank-you goes to my *friends*, who always support me in all situations of life. I am very grateful for the good times we have had and the ones that are going to come.

My family is the most important support in my life. I would like to thank my parents *Kurt* and *Claudia Stress* as well as my sister *Cynthia Stress* for their support during all this time. Without them, I could not have accomplished all this. They made it possible that I could cut my own path and reach the point where I am now.

Finally, I would like to thank my girlfriend *Julia Hildesheim* for everything she has done for me. I am looking forward to all the future "adventures" with her that will come. I love you.

Part III: Appendix

16. Building Blocks

16.1 Diversity Element 1

Table 16. SMILES codes and DNA codons for the building blocks of diversity element 1. Complete DNA sequence: 5'-hexyne-GGAGCTTGTGAATTCTGGXXX GGACGTGTGTGAATTGTC-3' with XXX representing the given codon sequence in the table.

No.	SMILES	Codon	No.	SMILES	Codon	No.	SMILES	Codon
NP01	[R2]C(/C=C/C1=CC([R1])=CC=C1)=O	AAA	NP08	[R1]C1=CC(C(F)(F)F)=CC(C([R2])=O)=C1	ATC	NP15	[R1]C1=CC=C(C([R2])=O)O1	ACG
NP02	[R2]C(/C=C/C1=CC=C([R1])C=C1)=O	AAT	NP09	[R1]C1=CC=C(F)C(C([R2])=O)=C1	AGA	NP16	[R2]C(CCCCC[R1])=O	ACC
NP03	[R2]C(CCC1=CC([R1])=CC=C1)=O	AAG	NP10	[R1]C1=CC=CC(OCCOCC([R2])=O)=C1	AGT	NP17	[R1]C1=CC=CC(/C=C/C=C/C([R2])=O)=C1	TAA
NP04	[R2]C(CCC1=CC=C([R1])C=C1)=O	AAC	NP11	[R1]C1=CC=CC(/C=C/C([R2])=O)=C1	AGG	NP18	[R1]C1=CC=CC(/C=C/C=C/C/C([R2])=O)=C1	TAT
NP05	[R1]C1=CC=CC(C([R2])=O)=C1	ATA	NP12	[R1]C[C@@H](C)C([R2])=O	AGC	NP19	[R1]C1=CC=CC(/C=C/C(C)=C/C([R2])=O)=C1	TAG
NP06	[R1]C1=CC=C(C([R2])=O)S1	ATT	NP13	[R1]CCC([R2])=O	ACA	NP20	[R1]C1=CC=CC(/C=C/C(C)=C/C([R2])=O)=C1	TAC
NP07	[R1]C1=CC=C(C([R2])=O)N=C1	ATG	NP14	[R2]C(CCC[R1])=O	ACT	NP21	[R1]C1=CC=CC(/C=C/C=C(C)/C([R2])=O)=C1	TTA

16.2 Diversity Element 2

Table 17. SMILES codes and DNA codons for the building blocks of diversity element 2. Complete DNA sequence: 5'-GTAGTTGGATCCGCACACYYYYGACAA TTCACACACGTCC-3' with YYYY representing the given codon sequence in the table.

No.	SMILES	Codon	No.	SMILES	Codon	No.	SMILES	Codon
AA001	C[C@@H](C([R4])=O)N[R3]	AAAA	AA043	CC(OC(N[C@@H](C(CCCN[R3])C([R4])=O)=O)(C)C	AGGG	AA085	[R4]C(CC(C(F)(F)F)N[R3])=O	TTTA
AA002	[R3]NCCC([R4])=O	AAAT	AA044	CSCC[C@@H](C([R4])=O)N[R3]	AGGC	AA086	[R3]NCC#CC([R4])=O	TTTT
AA003	C[C@H](C([R4])=O)N[R3]	AAAG	AA045	CSCC[C@H](C([R4])=O)N[R3]	AGCA	AA087	[R3]NC/C=C/C([R4])=O	TTTG
AA004	C[C@@H](C([R4])=O)N[R3]	AAAC	AA046	[R3]N[C@@H](CCC(OC(C)C(C)=O)C([R4])=O	AGCT	AA088	[R3]NC/C=C/C([R4])=O	TTTC
AA005	C[C@@H](C([R4])=O)N(C)[R3]	AATA	AA047	[R3]N[C@@H](CC(N)=O)C([R4])=O	AGCG	AA089	C/C(C([R4])=O)=C\C=C\CN[R3]	TTGA
AA006	[R3]N[C@@H](CC1C	AATT	AA048	[R3]N1[C@H](C([R4])	AGCC	AA090	CC(CC(C([R4])=O)N[TTGT

	CCCC1C([R4])=O			=O)CCC1			R3))=C	
AA007	[R3]N[C@H](CC1=C C=CN=C1)C([R4])=O	AATG	AA049	[R3]N1[C@@H](C([R 4])=O)CCC1	ACAA	AA091	O=C([C@@H]1CN([R 3])CCC1)[R4]	TTGG
AA008	[R3]NC(C1CCCC 1)C2CCCC2C([R4])=O	AATC	AA050	[R3]N1[C@H](C([R4])=O)CCC1	ACAT	AA092	O=C([C@H]1CN([R3])CCC1)[R4]	TTGC
AA009	CC(C)[C@@H](C([R 4])=O)N[R3]	AAGA	AA051	[R3]N1[C@@H](C([R 4])=O)CCC1	ACAG	AA093	[R3]NC1(C([R4])=O) CC1	TTCA
AA010	CC(C)[C@H](C([R4])=O)N[R3]	AAGT	AA052	O[C@H](C[C@H]1C([R4])=O)CN1[R3]	ACAC	AA094	[R4]C(C1CN(C1)[R3]) =O	TTCT
AA011	CC(C)[C@@H](C([R 4])=O)N(C)[R3]	AAGG	AA053	O[C@@H](C[C@H]1 C([R4])=O)CN1[R3]	ACTA	AA095	[R4]C(CC1(CCCCC1) N[R3])=O	TTCG
AA012	[R3]NCC([R4])=O	AAGC	AA054	[R3]N1[C@H](C([R4])=O)C=CC1	ACTT	AA096	[R3]NC1(CCC1)C([R 4])=O	TTCC
AA013	CN(CC([R4])=O)[R3]	AACA	AA055	[R3]N[C@@H](CC1= CC=CC=C1)C([R4])= O	ACTG	AA097	[R3]N[C@H]1[C@H](C([R4])=O)[C@@H]2 CC[C@H]1C2	TGAA
AA014	[R3]N[C@@H](C1=C C=CC=C1)C([R4])=O	AACT	AA056	[R3]N[C@H](CC1=C C=CC=C1)C([R4])=O	ACTC	AA098	[R3]NC1(C([R4])=O) CCCCC1	TGAT
AA015	[R3]N[C@H](C1=CC= CC=C1)C([R4])=O	AACG	AA057	CC(CC1=CC=CC=C1)C([R4])=O)N[R3]	ACGA	AA099	CS(C(CC1)(C([R4])= O)CCN1[R3])=O	TGAG
AA016	OC(C=C1)=CC=C1[C @H](C([R4])=O)N[R3]	AACC	AA058	[R3]N[C@H](CC1=C C=C([N+])([O-]))=O)C=C1)C([R4])= O	ACGT	AA100	O=S1(CCC(N[R3])(C([R4])=O)CC1)=O	TGAC
AA017	[R3]N[C@@H](CC(O C(C)C(C)=O)C([R4]) =O	ATAA	AA059	[R3]N[C@@H](CC1= CC=C([N+])([O-]))=O)C=C1)C([R4])= O	ACGG	AA101	O=C(C(CN[R3])C1[C @H]2C[C@@H]3C[C @@H](C[C@H]1C3) C2)[R4]	TGTA
AA018	CN(C=N1)C=C1C[C @@H](C([R4])=O)N[R3]	ATAT	AA060	CN([C@@H](CC1=C C=CC=C1)C([R4])=O)R3]	ACGC	AA102	[R3]NC(C1)CC21CC(C([R4])=O)C2	TGTT
AA019	[R3]N[C@@H](CSCC 1=CC=CC=C1)C([R4])=O	ATAG	AA061	FC(C=C1)=CC=C1C[C@@H](C([R4])=O)N [R3]	ACCA	AA103	[R3]NC1C(C([R4])=O))CCCCC1	TGTG
AA020	OC(C=C1)=CC=C1C[C@@H](C([R4])=O)N [R3]	ATAC	AA062	FC(C=C1)=CC=C1C[C@@H](C([R4])=O)N[R 3]	ACCT	AA104	[R3]NCC1=NC(C([R4])=O)CS1	TGTC
AA021	OC(C=C1)=CC=C1C[C@H](C([R4])=O)N[R 3]	ATTA	AA063	FC1=CC=CC(C[C@H)C([R4])=O)N[R3])=C 1	ACCG	AA105	O=S(CC(C([R4])=O)N 1[R3])(CC1)=O	TGGA
AA022	OC(C=C1)=CC=C1C[C@@H](C([R4])=O)N (C)[R3]	ATTT	AA064	BrC(C=C1)=CC=C1C [C@@H](C([R4])=O) N[R3]	ACCC	AA106	[R3]N1C[C@@H](C([R4])=O)[C@H](C2=C C(OC)=C(OC)N=C2) C1	TGGT
AA023	OC(C(I)=C1)=C(I)C= C1C[C@@H](C([R4]) =O)N[R3]	ATTG	AA065	CCCC(C([R4])=O)N [R3]	TAAA	AA107	[R3]N[C@@H](CC1= NC=CS1)C([R4])=O	TGGG

AA024	[R3]N[C@@H](CCC(N)=O)C([R4])=O	ATTC	AA066	CCCC[C@@H](C([R4])=O)N[R3]	TAAT	AA108	[R3]N[C@@H](CC1=CC=C(C=CC=C2)C2=N1)C([R4])=O	TGGC
AA025	OC[C@@H](C([R4])=O)N[R3]	ATGA	AA067	CCC[C@@H](C([R4])=O)N[R3]	TAAG	AA109	O=C(OC1=C2C=CC(OC)=C1)C=C2C[C@@H](C([R4])=O)N[R3]	TGCA
AA026	OC[C@H](C([R4])=O)N[R3]	ATGT	AA068	[R3]N[C@@H](CNC(OC(C)(C)C)=O)C([R4])=O	TAAC	AA110	[R3]N[C@@H](CC1=CC=CC=C1C=CC=C2)C([R4])=O	TGCT
AA027	CC[C@@H](C)[C@@H](C([R4])=O)N[R3]	ATGG	AA069	[R3]N[C@H](CNC(OC(C)(C)C)=O)C([R4])=O	TATA	AA111	[R4]C(CC(C1=CC=C(C=C1)N[R3])=O	TGCG
AA028	CC[C@@H](C)[C@@H](C([R4])=O)N(C)[R3]	ATGC	AA070	[R3]N[C@@H](CCN(C(OC(C)(C)C)=O)C([R4])=O	TATT	AA112	[R4]C(CC(C1=CC=C(C)C=C1)N[R3])=O	TGCC
AA029	[R3]N[C@@H](CC1=CNC2=C1C=CC=C2)C([R4])=O	ATCA	AA071	CC(C)(C([R4])=O)N[R3]	TATG	AA113	[R4]C(CC(C1=CC=C(F)C=C1)N[R3])=O	TCAA
AA030	[R3]N[C@H](CC1=CN2=C1C=CC=C2)C([R4])=O	ATCT	AA072	CC[C@@H](C([R4])=O)N[R3]	TATC	AA114	O=C([R4])[C@H]1[C@@H](N[R3])C2=CC=CC=C2C1	TCAT
AA031	CC(CC1=CNC2=C1C=CC=C2)C([R4])=O)N[R3]	ATCG	AA073	C#CC[C@@H](C([R4])=O)N[R3]	TAGA	AA115	[R4]C(CC(C1=CC(OC2)=C2C=C1)N[R3])=O	TCAG
AA032	C[C@@H](O)[C@@H](C([R4])=O)N[R3]	ATCC	AA074	[R3]NCCCCC([R4])=O	TAGT	AA116	[R4]C(CC(C1=C(C)C=C(C)C=C1)N[R3])=O	TCAC
AA033	C[C@H](O)[C@H](C([R4])=O)N[R3]	AGAA	AA075	[R3]NCCCC([R4])=O	TAGG	AA117	[R4]C(CC(C1=CC=C(SC)C=C1)N[R3])=O	TCTA
AA034	C[C@H](O)[C@@H](C([R4])=O)N[R3]	AGAT	AA076	CC(C)(C)[C@@H](C([R4])=O)N[R3]	TAGC	AA118	[R4]C(CC(C1=C(C)C=CC=C1)N[R3])=O	TCTT
AA035	C[C@@H](O)[C@H](C([R4])=O)N[R3]	AGAG	AA077	[R3]NOCC([R4])=O	TACA	AA119	O=C(CC(C1=CC=C(C2=CC=CC=C2)C=C1)N[R3])[R4]	TCTG
AA036	N=C(N[N+])([O-])=O)NCCC[C@@H](C([R4])=O)N[R3]	AGAC	AA078	[R3]NCCOCCOCCOCCC([R4])=O	TACT	AA120	O=C(C(CC1=C(C=C(C=C2)C2=CC=C1)C)N[R3])[R4]	TCTC
AA037	N=C(N[N+])([O-])=O)NCCC[C@H](C([R4])=O)N[R3]	AGTA	AA079	CC(C)CC(N[R3])CC([R4])=O	TACG	AA121	O=C(C(N[R3])C1=CC=CS1)[R4]	TCGA
AA038	[R3]N[C@@H](CCNC(OC(C)(C)C)=O)C([R4])=O	AGTT	AA080	[R4]C(C(C)(C)CN[R3])=O	TACC	AA122	CN1N=CC=C1C(N[R3])C([R4])=O	TCGT
AA039	CC(C)C[C@@H](C([R4])=O)N[R3]	AGTG	AA081	[R4]C(CC(C)(C)N[R3])=O	TTAA	AA123	CC1=NN(C2=CC=CC=C2)C(C)=C1C(N[R3])C([R4])=O	TCGG
AA040	CC(C)C[C@H](C([R4])=O)N[R3]	AGTC	AA082	[R4]C(CC(CCCC)N[R3])=O	TTAT	AA124	C/C(C([R4])=O)=C\C=C(C)CN[R3]	TCGC

AA041	[R3]N[C@@H](CCC CNC(OC(C)(C)C)=O) C([R4])=O	AGGA	AA083	[R3]NC(C#N)C([R4])= O	TTAG	AA125	O=C([R4])/C=C/C=C/ CN[R3]	TCCA
AA042	O=C(OCC1=CC=CC =C1)NCCCC[C@H](C([R4])=O)N[R3]	AGGT	AA084	[R3]NC(CS(=O)(N)=O)C([R4])=O	TTAC	AA126	[R3]NCC1=NC(C([R4)=O)=CS1	TCCT

16.3 Diversity Element 3

Table 18. SMILES codes and DNA codons for the building blocks of diversity element 3. Complete DNA sequence: 5'-GTTCAAGCCACTTACCTZZZZZTGATGCCTACCTATGAGA-3' with ZZZZZ representing the given codon sequence in the table. Note that the given SMILES code already includes the triazole, formed after the click reaction. The corresponding terminal alkyne structure was the actually used compound for the DEML synthesis.

No.	SMILES	Codon	No.	SMILES	Codon	No.	SMILES	Codon
TA001	[R5]CN1C=C(C(C) (C)NC(N)=O)N=N1	AAAAAA	TA222	[R5]CN1C=C(CNC (=O)c2cc(C3CC3)n c3onc(C)c23)N=N 1	AACTCT	TA443	[R5]CN1C=C(CNC (=O)CCc2c(C)nn(- c3ccccc3)c2)N=N 1	ATGCGG
TA002	[R5]CN1C=C(C(= O)O)N=N1	AAAAAT	TA223	[R5]CN1C=C(CNC (=O)C(C)n2nnc(- c3ccccc(F)c3)n2)N= N1	AACTCG	TA444	[R5]CN1C=C(c2cc cc(NC(=O)c3ccc4c (c3)C(=O)N(C)C4= O)c2)N=N1	ATGCGC
TA003	[R5]CN1C=C(CN2 CCOC[C@H]2C)N =N1	AAAAAG	TA224	[R5]CN1C=C(CN2 CCN(Cn3c(=O)oc4 cc(Cl)ccc43)CC2) N=N1	AACTCC	TA445	[R5]CN1C=C(CNC (=O)NC(c2ccccc2) c2ccc(C)cc2)N=N1	ATGCCA
TA004	[R5]CN1C=C(CN2 CCS(=O)CC2)N=N 1	AAAAAC	TA225	[R5]CN1C=C(CNC (=O)C2CCCN(c3n c4ccccc4s3)C2)N= N1	AACGAA	TA446	[R5]CN1C=C(CNC (=O)C=Cc2cnn(- c3ccccc3)c2)N=N1	ATGCCT
TA005	[R5]CN1C=C(CO[C@H](C)C(=O)OC)N=N1	AAAATA	TA226	[R5]CN1C=C(CN2 CCN(c3ccc4nnc(C (F)(F)F)n4n3)CC2) N=N1	AACGAT	TA447	[R5]CN1C=C(CNC (=O)c2c(C)oc3nnc(C)c(=O)c23)N=N1	ATGCCG
TA006	[R5]CN1C=C(CCC (=O)O)N=N1	AAAATT	TA227	[R5]CN1C=C(c2cc cc(NC(=O)N(Cc3c cccn3)C3CC3)c2) N=N1	AACGAG	TA448	[R5]CN1C=C(CNC (=O)NC(c2ccc(C)c c2)c2cccs2)N=N1	ATGCCC
TA007	[R5]CN1C=C(CN2 CCCCC2)N=N1	AAAATG	TA228	[R5]CN1C=C(c2cc cc(NCc3nc(- c4ccsc4)no3)c2)N =N1	AACGAC	TA449	[R5]CN1C=C(c2cc cc(NC(=O)Nc3cccc c3F)c2)N=N1	ATCAAA
TA008	[R5]CN1C=C(CSC CO)N=N1	AAAATC	TA229	[R5]CN1C=C(CNC (=O)c2ccc(- n3nc(C)cc3C)nc2) N=N1	AACGTA	TA450	[R5]CN1C=C(CNC (=O)c2nc3ccccc3c(=O)[nH]2)N=N1	ATCAAT
TA009	[R5]CN1C=C(C(= O)c2cccc(C(F)(F)F)c2)N=N1	AAAAGA	TA230	[R5]CN1C=C(CNC (=O)C2CCCN(c3cc c(Cl)nn3)C2)N=N1	AACGTT	TA451	[R5]CN1C=C(c2cc cc(NC(=O)Cn3nnc 4ccccc43)c2)N=N1	ATCAAG
TA010	[R5]CN1C=C(C(= O)O)N=N1	AAAAGT	TA231	[R5]CN1C=C(CN2 CCN(Cn3c(=O)oc4 cc(Cl)ccc43)CC2) N=N1	AACGTG	TA452	[R5]CN1C=C(c2cc cc(NC(=O)Cn3nnc 4ccccc43)c2)N=N1	ATCAAC

	O)CC)N=N1			CCN(Cn3nc4ccccn4c3=O)CC2)N=N1			cc(NC(=O)Nc3sc(C)c(C)c3C#N)c2)N=N1	
TA011	[R5]CN1C=C(CN2CC(=O)OC(=O)C2)N=N1	AAAAGG	TA232	[R5]CN1C=C(c2ccc(NC(=O)Cc3c(C)nc4nc(C(F)(F)F)nn4c3C)c2)N=N1	AACGTC	TA453	[R5]CN1C=C(CNC(=O)NC(C)c2ccc3c(c2)OCO3)N=N1	ATCATA
TA012	[R5]CN1C=C(Cn2cccc2C=O)N=N1	AAAAGC	TA233	[R5]CN1C=C(CNC(=O)c2cn(-c3cccc(Br)c3)nn2)N=N1	AACGGA	TA454	[R5]CN1C=C(CNC(=O)c2ccc(F)cc2)N=N1	ATCATT
TA013	[R5]CN1C=C(CN2CCNCC2=O)N=N1	AAAACA	TA234	[R5]CN1C=C(CNC(=O)CSc2nc(-c3cccc3)cs2)N=N1	AACGGT	TA455	[R5]CN1C=C(CNC(=O)NC(C)c2ccc(OCC)cc2)N=N1	ATCATG
TA014	[R5]CN1C=C(c2ccc(C(=O)O)cc2)N=N1	AAAACT	TA235	[R5]CN1C=C(CNS(=O)(=O)c2cccc(C(=O)NC34CC5CC(C(C5)C3)C4)c2)N=N1	AACGGG	TA456	[R5]CN1C=C(CNC(=O)C=Cc2cc(Cl)c(OCC)c(OC)c2)N=N1	ATCATC
TA015	[R5]CN1C=C(CC(N)CO)N=N1	AAAACG	TA236	[R5]CN1C=C(COC(=O)CCc2nc(-c3ccc(Cl)cc3)no2)N=N1	AACGGC	TA457	[R5]CN1C=C(c2ccc(NC(=O)C3CC4CCCC(C3)C4=O)c2)N=N1	ATCAGA
TA016	[R5]CN1C=C(C(O)c2ccc(Cl)cc2F)N=N1	AAAACC	TA237	[R5]CN1C=C(CN(CCC(N)=O)c2ccc(OC)cc2)N=N1	AACGCA	TA458	[R5]CN1C=C(c2ccc(NC(=O)CCN3C(=O)CCC3=O)c2)N=N1	ATCAGT
TA017	[R5]CN1C=C(c2cc(F)c(N)c(F)c2)N=N1	AAATAA	TA238	[R5]CN1C=C(CNC(=O)CCC2CCCC2)N=N1	AACGCT	TA459	[R5]CN1C=C(CNC(=O)c2ccc2C)N=N1	ATCAGG
TA018	[R5]CN1C=C(c2cccncn2)N=N1	AAATAT	TA239	[R5]CN1C=C(CNC(=O)c2nc(-c3cccs3)oc2C)N=N1	AACGCG	TA460	[R5]CN1C=C(CNC(=O)N(C)C(C)c2ccc(Cl)cc2Cl)N=N1	ATCAGC
TA019	[R5]CN1C=C([Si](CC)(CC)CC)N=N1	AAATAG	TA240	[R5]CN1C=C(CNC(=O)c2cn(-c3cccc3Br)nn2)N=N1	AACGCC	TA461	[R5]CN1C=C(CNC(=O)Cc2c(C)[nH]c3ccc(F)cc23)N=N1	ATCACAA
TA020	[R5]CN1C=C(Cn2c(C)ccc2C)N=N1	AAATAC	TA241	[R5]CN1C=C(CNC(=O)NCc2ccc(COC(C(F)(F)F)cc2)N=N1	AACCAA	TA462	[R5]CN1C=C(CNC(=O)Cc2c(C)[nH]c3ccc(C)cc23)N=N1	ATCACT
TA021	[R5]CN1C=C(COC)N=N1	AAATTA	TA242	[R5]CN1C=C(CNC(=O)Nc2ccc(F)c(F)c2)N=N1	AACCAT	TA463	[R5]CN1C=C(CNC(=O)NC(c2cccc2)C(C)C)N=N1	ATCACG
TA022	[R5]CN1C=C(C(C)(O)COC)N=N1	AAATTT	TA243	[R5]CN1C=C(CNC(=O)Cn2nnn(-c3cccc3Cl)c2=O)N=N1	AACCAG	TA464	[R5]CN1C=C(CNC(=O)C=Cc2csc(-c3ccc(OC)cc3)nn2)N=N1	ATCACC
TA023	[R5]CN1C=C(c2c(C)noc2C)N=N1	AAATTG	TA244	[R5]CN1C=C(COC(=O)CCNc2cccc2[AACCAC	TA465	[R5]CN1C=C(CNC(=O)c2cc(CC=C)c(ATCTAA

				N+](=O)[O-]N=N1			OC)c(OC)c2)N=N1	
TA024	[R5]CN1C=C(CNC(=O)c2ccc(Cl)nn2)N=N1	AAATTC	TA245	[R5]CN1C=C(CNC(=O)c2ccc(S(=O)(=O)NC)c2)N=N1	AACCTA	TA466	[R5]CN1C=C(CNC(=O)c2ccc3c(c2)CC(c2cccc2)OC3=O)N=N1	ATCTAT
TA025	[R5]CN1C=C(CCO S(C)(=O)=O)N=N1	AAATGA	TA246	[R5]CN1C=C(CN2C(=O)c3ccc(C(=O)Nc4ccc(F)cc4OC)c3S2(=O)=O)N=N1	AACCTT	TA467	[R5]CN1C=C(CNC(=O)c2cccc2C(=O)c2cccc2)N=N1	ATCTAG
TA026	[R5]CN1C=C(c2ccc(C3CCC3)c2)N=N1	AAATGT	TA247	[R5]CN1C=C(CNC(=O)c2cc(-c3cccc3)nc3c2c(C)nn3C)N=N1	AACCTG	TA468	[R5]CN1C=C(CNC(=O)c2ccc(OC)c3cccc23)N=N1	ATCTAC
TA027	[R5]CN1C=C(c2cc3nc(C)ccc3c2)N=N1	AAATGG	TA248	[R5]CN1C=C(CNC(=O)c2cc(C(C)C)nc3c2c(C)nn3C)N=N1	AACCTC	TA469	[R5]CN1C=C(CNC(=O)c2nn(-c3ccc(Cl)cc3)c(C)c2=O)N=N1	ATCTTA
TA028	[R5]CN1C=C(CN2CCS(=O)(=O)CC2)N=N1	AAATGC	TA249	[R5]CN1C=C(CNC(=O)c2ccc(OCC=C)c(OC)c2)N=N1	AACCGA	TA470	[R5]CN1C=C(CNS(=O)(=O)c2ccc(C(=O)OC(C)C)cc2)N=N1	ATCTTT
TA029	[R5]CN1C=C(c2cc3nc2)N=N1	AAATCA	TA250	[R5]CN1C=C(COC(=O)c2ccc(S(=O)(=O)NC(C)(C)C)c2)N=N1	AACCGT	TA471	[R5]CN1C=C(CNS(=O)(=O)c2ccc(C(=O)N3CCCC3)cc2)N=N1	ATCTTG
TA030	[R5]CN1C=C(CN2C[C@H](C)O[C@H](C)C2)N=N1	AAATCT	TA251	[R5]CN1C=C(CNC(=O)CN2CCN(S(=O)(=O)c3ccc4c(c3)OCCO4)CC2)N=N1	AACCGG	TA472	[R5]CN1C=C(CNC(=O)c2cccc(CS(C)(=O)=O)c2)N=N1	ATCTTC
TA031	[R5]CN1C=C(CCCO)N=N1	AAATCG	TA252	[R5]CN1C=C(CNC(=O)C2=NN(c3ccc(C)c(Cl)c3)C(=O)C2)N=N1	AACCGC	TA473	[R5]CN1C=C(CNC(=O)C(CC(C)C)NC(=O)c2cccc2C)N=N1	ATCTGA
TA032	[R5]CN1C=C(C(OCC)OCC)N=N1	AAATCC	TA253	[R5]CN1C=C(CNC(=O)c2nc(N3CCC3)nc2Cl)N=N1	AACCCA	TA474	[R5]CN1C=C(CNC(=O)CN2CCCC2C(=O)Nc2cccc2)N=N1	ATCTGT
TA033	[R5]CN1C=C(CSC)N=N1	AAAGAA	TA254	[R5]CN1C=C(CNC(=O)CCn2c(=O)n(CC)c3cccc32)N=N1	AACCTT	TA475	[R5]CN1C=C(CNC(=O)c2ccc(SCC(C)C)c([N+](=O)[O-])c2)N=N1	ATCTGG
TA034	[R5]CN1C=C(Cn2cccc2=O)N=N1	AAAGAT	TA255	[R5]CN1C=C(c2ccc(NC(=O)C)C)NC(=O)c3ccc3c2)N=N1	AACCCG	TA476	[R5]CN1C=C(CNC(=O)COC(=O)c2ccc([N+](=O)[O-])cn2C)N=N1	ATCTGC
TA035	[R5]CN1C=C(CC2(CO)CCC2)N=N1	AAAGAG	TA256	[R5]CN1C=C(c2ccc(NC(=O)NCCOC3CCCC3C)c2)N=N1	AACCCC	TA477	[R5]CN1C=C(c2ccc(NC(=O)c3ccc(C(C)=O)s3)c2)N=N1	ATCTCA
TA036	[R5]CN1C=C(c2cc	AAAGAC	TA257	[R5]CN1C=C(CNC	ATAAAA	TA478	[R5]CN1C=C(CNC	ATCTCT

	c(F)cn2)N=N1			(=O)c2cc(C(C)C)n c3c2cnn3C(C)C)N =N1			(=O)CSc2nc3sc(- c4ccccc4)c(C)c3c(=O)n2CC)N=N1	
TA037	[R5]CN1C=C(CCC CC(=O)O)N=N1	AAAGTA	TA258	[R5]CN1C=C(CNC (=O)c2ccc3[nH]c4c (c3c2)CCCC4)N=N 1	ATAAAT	TA479	[R5]CN1C=C(CNC (=O)NCc2ccc3c(c2)OCCO3)N=N1	ATCTCG
TA038	[R5]CN1C=C([C@ H](O)c2ccccc2)N= N1	AAAGTT	TA259	[R5]CN1C=C(CNC (=O)Cc2ccc(N3CC CC3=O)cc2)N=N1	ATAAAG	TA480	[R5]CN1C=C(CNC (=O)NCCOc2ccc(C)cc2)N=N1	ATCTCC
TA039	[R5]CN1C=C(C(C) =O)N=N1	AAAGTG	TA260	[R5]CN1C=C(CNC (=O)C(C)Nc2ccc3[nH]c(=O)[nH]c3c2) N=N1	ATAAAC	TA481	[R5]CN1C=C(CNC (=O)CSc2nnc3ccc cn23)N=N1	ATCGAA
TA040	[R5]CN1C=C(CON)N=N1	AAAGTC	TA261	[R5]CN1C=C(CNC (=O)NC(Cc2ccccc 2)c2ccccc2)N=N1	ATAATA	TA482	[R5]CN1C=C(COc 2ncccn2)N=N1	ATCGAT
TA041	[R5]CN1C=C(Cn2c ncn2)N=N1	AAAGGA	TA262	[R5]CN1C=C(c2cc cc(NC(=O)NCCn3c cnc3)c2)N=N1	ATAATT	TA483	[R5]CN1C=C(CN2 C(=O)c3ccc(C(=O) O)cc3S2(=O)=O)N =N1	ATCGAG
TA042	[R5]CN1C=C(c2cc c(F)nc2)N=N1	AAAGGT	TA263	[R5]CN1C=C(c2cc cc(NC(=O)NCCn 3ccnc3)c2)N=N1	ATAATG	TA484	[R5]CN1C=C(c2cn n(CCC)c2)N=N1	ATCGAC
TA043	[R5]CN1C=C(c2cc cs2)N=N1	AAAGGG	TA264	[R5]CN1C=C(CN2 CCCC2C(=O)NC c2ccccc2COC)N= N1	ATAATC	TA485	[R5]CN1C=C(CC2(O)CCSC2)N=N1	ATCGTA
TA044	[R5]CN1C=C(CNC)N=N1	AAAGGC	TA265	[R5]CN1C=C(CNC (=O)N2CCN(C(=O) c3ccc(C)cc3)CC2) N=N1	ATAAGA	TA486	[R5]CN1C=C(c2cn cc(C(=O)O)c2)N= N1	ATCGTT
TA045	[R5]CN1C=C(C(C) (N)CC)N=N1	AAAGCA	TA266	[R5]CN1C=C(c2cc cc(NCc3nc(- c4ccccc4)no3)c2)N =N1	ATAAGT	TA487	[R5]CN1C=C(CN2 CCCS(=O)(=O)CC 2)N=N1	ATCGTG
TA046	[R5]CN1C=C(CCC C=C)N=N1	AAAGCT	TA267	[R5]CN1C=C(CNC (=O)CCCC(=O)c2c (C)cc(C)cc2)N=N 1	ATAAGG	TA488	[R5]CN1C=C(c2cc cc(NC(=O)NC(=O) CCI)c2)N=N1	ATCGTC
TA047	[R5]CN1C=C(C2(O)CCCC2)N=N1	AAAGCG	TA268	[R5]CN1C=C(CN2 CCN(c3ccccc3OC) CC2)N=N1	ATAAGC	TA489	[R5]CN1C=C(C(= O)c2ccc(Cl)cc2)N= N1	ATCGGA
TA048	[R5]CN1C=C(c2cc c(N)cc2)N=N1	AAAGCC	TA269	[R5]CN1C=C(CN(Cc2nnc3n(C)c(=O) c4ccccc4n23)C2C Cc3ccccc32)N=N1	ATAACA	TA490	[R5]CN1C=C(CCC S(N)(=O)=O)N=N1	ATCGGT
TA049	[R5]CN1C=C(B2O C(C)C(C)C(O)O2)N=N1	AAACAA	TA270	[R5]CN1C=C(c2cc cc(NC(=O)NCC3cc nc(N(C)C)c3)c2)N =N1	ATAACT	TA491	[R5]CN1C=C(c2cc[nH]n2)N=N1	ATCGGG

TA050	[R5]CN1C=C(C2C C3(CC(C(=O)O)C3 C2)N=N1	AAACAT	TA271	[R5]CN1C=C(CN(Cc2nnc2Cl)C2CC c3cccc32)N=N1	ATAACG	TA492	[R5]CN1C=C(Cn2c cnc2C)N=N1	ATCGGC
TA051	[R5]CN1C=C(C(O) C2CC2)N=N1	AAACAG	TA272	[R5]CN1C=C(CNC (=O)Cc2c(C)nn(- c3nc(C)cc(C)n3)c2 C)N=N1	ATAACC	TA493	[R5]CN1C=C(c2cc cc(NC(=O)CCN3C CSC3=O)c2)N=N1	ATCGCA
TA052	[R5]CN1C=C(c2cc sc2)N=N1	AAACAC	TA273	[R5]CN1C=C(COc 2ccc(CCNC(=O)Nc 3nccs3)cc2)N=N1	ATATAA	TA494	[R5]CN1C=C(c2cc cc(NC(=O)CCc3c(C)noc3C)c2)N=N1	ATCGCT
TA053	[R5]CN1C=C(CCl) N=N1	AAACTA	TA274	[R5]CN1C=C(CN(Cc2ccc(C)cc2)C(= O)Nc2nccs2)N=N1	ATATAT	TA495	[R5]CN1C=C(CNS (=O)(=O)c2c(C)cc(C(C)(C)C)cc2C)N= N1	ATCGCG
TA054	[R5]CN1C=C(CCO)N=N1	AAACTT	TA275	[R5]CN1C=C(CN2 C(=O)C(=O)c3cc(B r)ccc32)N=N1	ATATAG	TA496	[R5]CN1C=C(CNC (=O)c2[nH]c3c(c2C)C(=O)CC(C)C3)N=N1	ATCGCC
TA055	[R5]CN1C=C(CCN C2CCCC2)N=N1	AAACTG	TA276	[R5]CN1C=C(CNC (=O)CNC(=O)C=C c2ccc3c(c2)OCCO 3)N=N1	ATATAC	TA497	[R5]CN1C=C(CNC (=O)c2csc3c2CCC C3)N=N1	ATCCAA
TA056	[R5]CN1C=C(CN(C)c2cccc2)N=N1	AAACTC	TA277	[R5]CN1C=C(CNC (=O)CCNC(=O)c2c c(C(C)C)nc3c2c(= O)[nH]c(=O)n3CC C)N=N1	ATATTA	TA498	[R5]CN1C=C(c2cc cc(NC(=O)c3ccc4[nH]nnc4c3)c2)N=N 1	ATCCAT
TA057	[R5]CN1C=C(CC2(C#N)CCCC2)N= N1	AAACGA	TA278	[R5]CN1C=C(CNC (=O)CCNC(=O)c2c c(F)cc(F)c2)N=N1	ATATTT	TA499	[R5]CN1C=C(CNC (=O)c2csc3cccc2 3)N=N1	ATCCAG
TA058	[R5]CN1C=C(c2cc c(Cl)nc2)N=N1	AAACGT	TA279	[R5]CN1C=C(CNC (=O)CCNC(=O)C2(c3ccc(F)cc3)CC2) N=N1	ATATTG	TA500	[R5]CN1C=C(c2cc cc(NC(=O)CCn3cn cn3)c2)N=N1	ATCCAC
TA059	[R5]CN1C=C(CC2(O)CCC2)N=N1	AAACGG	TA280	[R5]CN1C=C(c2cc cc(NCC(=O)N(C)C C(=O)NC(C)(C)C)c 2)N=N1	ATATTC	TA501	[R5]CN1C=C(CNC (=O)c2cc(C)sc2C) N=N1	ATCCTA
TA060	[R5]CN1C=C(C2C CCC2)N=N1	AAACGC	TA281	[R5]CN1C=C(CN(C(=O)c2ccc(CS(= O)(=O)c3cccc3)o 2)C2CCCC2)N= N1	ATATGA	TA502	[R5]CN1C=C(CNC (=O)c2cnn(- c3cccc3)n2)N=N1	ATCCTT
TA061	[R5]CN1C=C(CNC C#N)N=N1	AAACCA	TA282	[R5]CN1C=C(CNC (=O)c2ccc(S(=O)(= O)NOC)cc2)N=N1	ATATGT	TA503	[R5]CN1C=C(CNC (=O)CSc2nc3ccsc 3c(=O)n2CC)N=N1	ATCCTG
TA062	[R5]CN1C=C(CN2 C(=O)COCC2=O) N=N1	AAACCT	TA283	[R5]CN1C=C(CNC (=O)CN(C)c2ccc(F)c(F)c2)N=N1	ATATGG	TA504	[R5]CN1C=C(c2cc cc(NC(=O)c3sc(C) nc3C)c2)N=N1	ATCCTC
TA063	[R5]CN1C=C(CS(C)(=O)O)N=N1	AAACCG	TA284	[R5]CN1C=C(CNC (=O)c2ccc(- n3ccc(C(F)(F)F)n3	ATATGC	TA505	[R5]CN1C=C(CNC (=O)CSc2nc(C3CC 3)nc(C)c2C#N)N= N1	ATCCGA

)cc2)N=N1			N1	
TA064	[R5]CN1C=C(c2cc c(C(O)C(F)F)cc 2)N=N1	AAACCC	TA285	[R5]CN1C=C(CNC (=O)CCNC(=O)c2c cc3[nH]c4c(c3c2)C CCC4)N=N1	ATATCA	TA506	[R5]CN1C=C(c2cc cc(NC(=O)CN3CC CCCC3=O)c2)N=N 1	ATCCGT
TA065	[R5]CN1C=C(C=C C)N=N1	AATAAA	TA286	[R5]CN1C=C(CN(Cc2ccn(C)c(=O)c2)C2CCc3ccccc32) N=N1	ATATCT	TA507	[R5]CN1C=C(c2cc cc(NC(=O)c3cc4cc ccc4oc3=O)c2)N= N1	ATCCGG
TA066	[R5]CN1C=C([1C @H]2C[1C@H]2C O)N=N1	AATAAT	TA287	[R5]CN1C=C(CNC (=O)CN(Cc2ccccc 2)C2CCCC2)N= N1	ATATCG	TA508	[R5]CN1C=C(CNC (=O)CN2C(=O)NC 3(Cc4ccccc43)C 2=O)N=N1	ATCCGC
TA067	[R5]CN1C=C(CCC 2(CCC(=O)O)N=N 2)N=N1	AATAAG	TA288	[R5]CN1C=C(CNC (=O)CCCNC(=O)O Cc2ccccc2)N=N1	ATATCC	TA509	[R5]CN1C=C(CNC (=O)CCc2c(C)nn(C C(C)C)c2C)N=N1	ATCCCA
TA068	[R5]CN1C=C(c2cc c(F)cc2)N=N1	AATAAC	TA289	[R5]CN1C=C(CNC (=O)CN2CCN(CCc 3cccs3)CC2)N=N1	ATAGAA	TA510	[R5]CN1C=C(c2cc cc(NC(=O)C3(c4cc ccc4)CC3)c2)N=N 1	ATCCCT
TA069	[R5]CN1C=C(c2cn cn2C)N=N1	AATATA	TA290	[R5]CN1C=C(c2cc cc(NC(=O)NCC(C) N3CCOCC3)c2)N= N1	ATAGAT	TA511	[R5]CN1C=C(CNC (=O)NCc2ccc(OC(F)F)cc2)N=N1	ATCCCG
TA070	[R5]CN1C=C(C2C C2)N=N1	AATATT	TA291	[R5]CN1C=C(CNC (=O)c2ccc(C)c(NS(C)(=O)=O)c2)N=N 1	ATAGAG	TA512	[R5]CN1C=C(CNC (=O)C2CCCN(c3c nccn3)C2)N=N1	ATCCCC
TA071	[R5]CN1C=C(CCN C2COC2)N=N1	AATATG	TA292	[R5]CN1C=C(CNC (=O)CN2CCCC2C(=O)NCC(F)(F)F)N =N1	ATAGAC	TA513	[R5]CN1C=C(CNC (=O)c2cn3c(C)cccc 3n2)N=N1	AGAAAA
TA072	[R5]CN1C=C(CNC C2CC2)N=N1	AATATC	TA293	[R5]CN1C=C(CN2 CCN(Cc3ccc(OC)c (OC)c3OC)CC2)N =N1	ATAGTA	TA514	[R5]CN1C=C(c2cc cc(NC(=O)C3CSC 4(C)CCC(=O)N34) c2)N=N1	AGAAAT
TA073	[R5]CN1C=C(c2cc c(C=O)cc2)N=N1	AATAGA	TA294	[R5]CN1C=C(COC (=O)c2ccccc(NC(=O)NC3CC3)c2)N=N 1	ATAGTT	TA515	[R5]CN1C=C(CNC (=O)Cn2c(=O)c(C# N)cn(C3CC3)c2=O)N=N1	AGAAAG
TA074	[R5]CN1C=C(Cn2c c(Cl)cn2)N=N1	AATAGT	TA295	[R5]CN1C=C(CCC OC(=O)c2cc3c(C)n n(C)c3nc2C)N=N1	ATAGTG	TA516	[R5]CN1C=C(CNC (=O)NC(c2ccccc2) c2ccccc2)N=N1	AGAAAC
TA075	[R5]CN1C=C(C(C) (C)Br)N=N1	AATAGG	TA296	[R5]CN1C=C(CCO C(=O)c2ccc(CNC(=O)NC(C)C)cc2)N =N1	ATAGTC	TA517	[R5]CN1C=C(CNC (=O)c2ccc3c(c2)O CCCC3)N=N1	AGAATA
TA076	[R5]CN1C=C([C@ @H](O)c2ccccc2) N=N1	AATAGC	TA297	[R5]CN1C=C(COC (=O)c2[nH]c(C)c(C (C)=O)c2CCC)N= N1	ATAGGA	TA518	[R5]CN1C=C(CNC (=O)c2c(C)oc(- n3ccccc3)c2C#N)N =N1	AGAATT

TA077	[R5]CN1C=C(C(C)(C)C)N=N1	AATACA	TA298	[R5]CN1C=C(CNC(=O)c2cccc2NCC(=O)NC(=O)NC)N=N1	ATAGGT	TA519	[R5]CN1C=C(CNC(=O)CCc2c(C)nn(C)c2C)N=N1	AGAATG
TA078	[R5]CN1C=C(CNC(=O)OC)N=N1	AATACT	TA299	[R5]CN1C=C(CNC(=O)C2CC2c2cccc(Br)c2)N=N1	ATAGGG	TA520	[R5]CN1C=C(CNC(=O)c2sc(C)nc2-c2cccc2)N=N1	AGAATC
TA079	[R5]CN1C=C(CN2CCC[C@H]2C(=O)OC)N=N1	AATACG	TA300	[R5]CN1C=C(CNC(=O)c2sc(-c3ccc(CC)cc3)nc2C)N=N1	ATAGGC	TA521	[R5]CN1C=C(CNC(=O)c2sc3cccc3c2C)N=N1	AGAAGA
TA080	[R5]CN1C=C(CNS(C)(=O)=O)N=N1	AATACC	TA301	[R5]CN1C=C(CNC(=O)N(C)Cc2ccc(S C)c(OC)c2)N=N1	ATAGCA	TA522	[R5]CN1C=C(CNC(=O)Cn2cc([N+](=O)[O-])cn2)N=N1	AGAAGT
TA081	[R5]CN1C=C(C(C)(C)N)N=N1	AATTAA	TA302	[R5]CN1C=C(CCOc2ccc(C=CC(=O)c3cnn(C)c3)cc2)N=N1	ATAGCT	TA523	[R5]CN1C=C(CNC(=O)c2nc(SC)nc2Cl)N=N1	AGAAGG
TA082	[R5]CN1C=C(CCCCO)N=N1	AATTAT	TA303	[R5]CN1C=C(CN(CCC(=O)NC(N)=O)Cc2ccc(C)cc2)N=N1	ATAGCG	TA524	[R5]CN1C=C(CNC(=O)CN2CCC3(CC2)OCCO3)N=N1	AGAAGC
TA083	[R5]CN1C=C(c2ccc(C)c2)N=N1	AATTAG	TA304	[R5]CN1C=C(CNC(=O)CN2CCN(S(=O)(=O)CCC)CC2)N=N1	ATAGCC	TA525	[R5]CN1C=C(CNC(=O)CN(C)S(=O)(=O)c2cccs2)N=N1	AGAACA
TA084	[R5]CN1C=C(C(O)CC)N=N1	AATTAC	TA305	[R5]CN1C=C(CNC(=O)C(C)Sc2nc3cc c(F)cc3c(=O)[nH]2)N=N1	ATACAA	TA526	[R5]CN1C=C(CNC(=O)CCCC2CCCCC2)N=N1	AGAACT
TA085	[R5]CN1C=C(C(C)(C)NC)N=N1	AATTTA	TA306	[R5]CN1C=C(CNC(=O)CSC(C)c2c(F)cccc2F)N=N1	ATACAT	TA527	[R5]CN1C=C(c2ccc(NC(=O)CCc3c(C)[nH]c(=O)[nH]c3=O)c2)N=N1	AGAACG
TA086	[R5]CN1C=C(CCC(=O)CC(=O)OC)N=N1	AATTTT	TA307	[R5]CN1C=C(CN2CCC[C@H]2C(=O)O)N=N1	ATACAG	TA528	[R5]CN1C=C(CNC(=O)c2ccc(=O)n(CCOc)n2)N=N1	AGAACC
TA087	[R5]CN1C=C(CCN2CCOCC2)N=N1	AATTTG	TA308	[R5]CN1C=C(CNC(=O)Nc2cccc(C#N)c2)N=N1	ATACAC	TA529	[R5]CN1C=C(CN2C(=O)COc3ccc(Br)cc32)N=N1	AGATAA
TA088	[R5]CN1C=C(c2cc(F)cc(F)c2)N=N1	AATTC	TA309	[R5]CN1C=C(COc2cccc2O)N=N1	ATACTA	TA530	[R5]CN1C=C(CNS(=O)(=O)N2CCNC2)N=N1	AGATAT
TA089	[R5]CN1C=C(CN(CC)CC)N=N1	AATTGA	TA310	[R5]CN1C=C(Cn2nc(C)cc2C)N=N1	ATACTT	TA531	[R5]CN1C=C(CNC(=N)N)N=N1	AGATAG
TA090	[R5]CN1C=C(COc2cccc(N)c2)N=N1	AATTGT	TA311	[R5]CN1C=C(Cn2cnc2)N=N1	ATACTG	TA532	[R5]CN1C=C(CNC(=O)CC(=O)O)N=N1	AGATAC
TA091	[R5]CN1C=C(CC(C)C(=O)C)N=N1	AATTGG	TA312	[R5]CN1C=C(Cn2nc(C)c(C)c2C)N=N1	ATACTC	TA533	[R5]CN1C=C(CN2C(=O)CSCC2=O)N=N1	AGATTA

	=N1			1			=N1	
TA092	[R5]CN1C=C(C2(c3ccccc3)CC2)N=N1	AATTGC	TA313	[R5]CN1C=C(Cn2nc(C)nc2C)N=N1	ATACGA	TA534	[R5]CN1C=C(C2C CNCC2)N=N1	AGATTT
TA093	[R5]CN1C=C(c2ccc(OCC)c2)N=N1	AATTCA	TA314	[R5]CN1C=C(c2ccc(NC)c2)N=N1	ATACGT	TA535	[R5]CN1C=C(CN(CCO)CCO)N=N1	AGATTG
TA094	[R5]CN1C=C(CN(C)C(C)=O)N=N1	AATTCT	TA315	[R5]CN1C=C(CNC(=O)CN2CCCNCC2)N=N1	ATACGG	TA536	[R5]CN1C=C(C2(O)CC3CC2C3)N=N1	AGATTC
TA095	[R5]CN1C=C(c2csc(C)n2)N=N1	AATTGC	TA316	[R5]CN1C=C(CN2C(=O)C(=O)c3ccccc32)N=N1	ATACGC	TA537	[R5]CN1C=C(CN2Cc3ccccc3C2)N=N1	AGATGA
TA096	[R5]CN1C=C(CCC(C)=O)N=N1	AATTCC	TA317	[R5]CN1C=C(CNC(C)=O)N=N1	ATACCA	TA538	[R5]CN1C=C(CSc2nc(N)c3c4c(sc3n2)CCCC4)N=N1	AGATGT
TA097	[R5]CN1C=C(CN(C)C(C)C(C)C)N=N1	AATGAA	TA318	[R5]CN1C=C(CNC(=O)c2ccc3nc(C)c(C)nc3c2)N=N1	ATACCT	TA539	[R5]CN1C=C(CNC(=O)CCNC2=NS(=O)(=O)c3ccccc32)N=N1	AGATGG
TA098	[R5]CN1C=C(CCC(CCC(=O)O)N=N1	AATGAT	TA319	[R5]CN1C=C(c2ccc(NCc3cc(=O)n4c(C)csc4n3)c2)N=N1	ATACCG	TA540	[R5]CN1C=C(CNC(=O)C2=CC=CN3CCS(=O)(=O)N=C23)N=N1	AGATGC
TA099	[R5]CN1C=C(C(=O)N2CCC(OC)C2)N=N1	AATGAG	TA320	[R5]CN1C=C(Cn2c(C=Cc3ccc(C)cc3)nc3ccccc32)N=N1	ATACCC	TA541	[R5]CN1C=C(Cn2c(CCCO)nc3ccccc32)N=N1	AGATCA
TA100	[R5]CN1C=C(CNS(=O)(=O)C2CC2)N=N1	AATGAC	TA321	[R5]CN1C=C(CSc2nnnn2-c2cc(C)ccc2OC)N=N1	ATTAAA	TA542	[R5]CN1C=C(C(C)C(C)C)NC(=O)Cn2c(=O)c(C#N)cn(C)c2=O)N=N1	AGATCT
TA101	[R5]CN1C=C(C(O)c2ccccc2)N=N1	AATGTA	TA322	[R5]CN1C=C(CNC(=O)C2C(=O)N(c3ccc(C)cc3)C(=O)c3ccccc32)N=N1	ATTAAT	TA543	[R5]CN1C=C(C(C)C(C)C)NC(=O)CSc2nnnn2C2CC2)N=N1	AGATCG
TA102	[R5]CN1C=C(c2sc(C)c2)N=N1	AATGTT	TA323	[R5]CN1C=C(C(C)C(C)C)NC(=O)COc2ccccc2[N+](=O)[O-])N=N1	ATTAAG	TA544	[R5]CN1C=C(CSc2nnc(-c3sc4ccccc4c3Cl)o2)N=N1	AGATCC
TA103	[R5]CN1C=C(C(C)OS(C)(=O)O)N=N1	AATGTG	TA324	[R5]CN1C=C(C(C)C(C)C)NC(=O)COc2ccccc2C)N=N1	ATTAAC	TA545	[R5]CN1C=C(CSc2nc3sc(C)C(C)c3c(=O)[nH]2)N=N1	AGAGAA
TA104	[R5]CN1C=C(CNc2nc(Cl)nc(Cl)n2)N=N1	AATGTC	TA325	[R5]CN1C=C(CNC(=O)C2C(=O)N(CC)C(=S)N(C)C2=O)N=N1	ATTATA	TA546	[R5]CN1C=C(CSc2nc(C)cc(C)c2C#N)N=N1	AGAGAT
TA105	[R5]CN1C=C(CCN(C)C(=O)c2ccccc2)N=N1	AATGGA	TA326	[R5]CN1C=C(CNC(=O)C2C(=O)N(C)N(CCO)C2=O)N=N1	ATTATT	TA547	[R5]CN1C=C(CNC(=O)CSc2nc3nc4c(ccc4c-3n[nH]2)N=N1	AGAGAG
TA106	[R5]CN1C=C(c2c(AATGGT	TA327	[R5]CN1C=C(CNC	ATTATG	TA548	[R5]CN1C=C(CNC	AGAGAC

	F)cccc2F)N=N1			=C2C(=O)NC(=O)NC2=O)N=N1			(=O)Cn2[nH]c(=O)c3ccccc3c2=O)N=N1	
TA107	[R5]CN1C=C(CNC(=O)C(C)C)N=N1	AATGGG	TA328	[R5]CN1C=C(CNS(=O)(=O)c2cccc(C(=O)O)c2)N=N1	ATTATC	TA549	[R5]CN1C=C(CNC(=O)N2CCN(c3ncccn3)CC2)N=N1	AGAGTA
TA108	[R5]CN1C=C(CCN)N=N1	AATGGC	TA329	[R5]CN1C=C(CSc2nc(C)c(C(=O)O)c2C#N)N=N1	ATTAGA	TA550	[R5]CN1C=C(CN2CCN(S(=O)(=O)c3cccc(F)c3)CC2)N=N1	AGAGTT
TA109	[R5]CN1C=C(CCC(=O)NN)N=N1	AATGCA	TA330	[R5]CN1C=C(Cn2c(COc3ccc(Cl)cc3)n c3ccccc32)N=N1	ATTAGT	TA551	[R5]CN1C=C(CNC(=O)c2cc3c(s2)-c2ccccc2OC3)N=N1	AGAGTG
TA110	[R5]CN1C=C(c2cn c(Cl)nc2)N=N1	AATGCT	TA331	[R5]CN1C=C(Cn2c(-c3ccc(Cl)cc3)nc3ccccc32)N=N1	ATTAGG	TA552	[R5]CN1C=C(CNC(=O)NC2CC2)N=N1	AGAGTC
TA111	[R5]CN1C=C(CN2C(=O)c3ccccc3C2=O)N=N1	AATGCG	TA332	[R5]CN1C=C(Cn2c(-c3ccc(OC)cc3)nc3ccccc32)N=N1	ATTAGC	TA553	[R5]CN1C=C(CC2NC(=O)NC2=O)N=N1	AGAGGA
TA112	[R5]CN1C=C(C(C)(C)NS(C)(=O)=O)N=N1	AATGCC	TA333	[R5]CN1C=C(Cn2c(-c3ccccc3C)nc3ccccc32)N=N1	ATTACA	TA554	[R5]CN1C=C(COc2ccccc2[N+](=O)[O-])N=N1	AGAGGT
TA113	[R5]CN1C=C(c2ccc(N)c2)N=N1	AATCAA	TA334	[R5]CN1C=C(Cn2c(CCCNC(=O)c3ccc(O)c3nc3ccccc32)N=N1	ATTACT	TA555	[R5]CN1C=C(CNC(=O)CNC(=O)OC(C)(C)C)N=N1	AGAGGG
TA114	[R5]CN1C=C(Cn2c c(C)cn2)N=N1	AATCAT	TA335	[R5]CN1C=C(Cn2c(C=Cc3ccccc3)nc3ccccc32)N=N1	ATTACG	TA556	[R5]CN1C=C(CNC(=O)C(CC)CC)N=N1	AGAGGC
TA115	[R5]CN1C=C(Cn2c ccn2)N=N1	AATCAG	TA336	[R5]CN1C=C(COc2ccc(C=C3NC(=O)NC3=O)cc2OC)N=N1	ATTACC	TA557	[R5]CN1C=C(c2ccc(C#N)cc2)N=N1	AGAGCA
TA116	[R5]CN1C=C(Cn2c ccc2)N=N1	AATCAC	TA337	[R5]CN1C=C(CNC(=O)c2ccc3c(c2)[nH]c(=O)c(=O)n3CC)N=N1	ATTTAA	TA558	[R5]CN1C=C(COc2ccc(Br)nc2)N=N1	AGAGCT
TA117	[R5]CN1C=C(CNC(=O)N2CCNCC2)N=N1	AATCTA	TA338	[R5]CN1C=C(CNC(=O)c2cc(C3CC3)nn2-c2ccccc2)N=N1	ATTTAT	TA559	[R5]CN1C=C(CN2CCN(c3ccccc3)CC2)N=N1	AGAGCG
TA118	[R5]CN1C=C(CNS(=O)(=O)OC)N=N1	AATCTT	TA339	[R5]CN1C=C(CNC(=O)COC(=O)c2[nH]c(C)c(C(C)=O)c2C)N=N1	ATTTAG	TA560	[R5]CN1C=C(c2ccc(C(=O)OC)cc2)N=N1	AGAGCC
TA119	[R5]CN1C=C(CNC(=O)CC)N=N1	AATCTG	TA340	[R5]CN1C=C(CNC(=O)C2CCN2S(=O)(=O)c2ccccc2)N=N1	ATTTAC	TA561	[R5]CN1C=C(C(=O)N(C)C)N=N1	AGACAA

TA120	[R5]CN1C=C(c2ccc(NC(C)C)c2)N=N1	AATCTC	TA341	[R5]CN1C=C(CNC(=O)c2ccc(S(=O)(=O)NCCC#N)cc2)N=N1	ATTTTA	TA562	[R5]CN1C=C(CNC(=O)O)N=N1	AGACAT
TA121	[R5]CN1C=C(C(C)(C)O)N=N1	AATCGA	TA342	[R5]CN1C=C(CNC(=O)c2ccc(S(=O)(=O)N3CCCC3)cc2)N=N1	ATTTTT	TA563	[R5]CN1C=C(CC(=O)O)N=N1	AGACAG
TA122	[R5]CN1C=C(CC(C)O)N=N1	AATCGT	TA343	[R5]CN1C=C(CNC(=O)c2c(-c3cccc3)noc2C)N=N1	ATTTTG	TA564	[R5]CN1C=C(COc2ccc([N+](=O)[O-])cc2)N=N1	AGACAC
TA123	[R5]CN1C=C(CN(C)CCO)N=N1	AATCGG	TA344	[R5]CN1C=C(CNc2nnc3sc(C(=O)OCC)c(C)c23)N=N1	ATTTTC	TA565	[R5]CN1C=C(C(=O)c2cccc2)N=N1	AGACTA
TA124	[R5]CN1C=C(CN2CCOCC2)N=N1	AATCGC	TA345	[R5]CN1C=C(CNC(=O)CSc2n[nH]c3nc4cccc4n23)N=N1	ATTTGA	TA566	[R5]CN1C=C(CCCN)N=N1	AGACTT
TA125	[R5]CN1C=C(CC(C)O)c2cccs2)N=N1	AATCCA	TA346	[R5]CN1C=C(CNS(=O)(=O)c2ccc3ccc3c2)N=N1	ATTTGT	TA567	[R5]CN1C=C(CSC(=N)N)N=N1	AGACTG
TA126	[R5]CN1C=C(c2ccc(Cl)c2)N=N1	AATCCT	TA347	[R5]CN1C=C(CNC(=O)c2oc3ccc(OC)cc3c2C)N=N1	ATTTGG	TA568	[R5]CN1C=C(CSCC(=O)O)N=N1	AGACTC
TA127	[R5]CN1C=C(c2cc(F)ccc2F)N=N1	AATCCG	TA348	[R5]CN1C=C(CNS(=O)(=O)c2ccc3c(c2)CCC(=O)N3)N=N1	ATTTGC	TA569	[R5]CN1C=C(CNC(=O)c2ccnn2C)N=N1	AGACGA
TA128	[R5]CN1C=C(c2cncc(F)c2)N=N1	AATCCC	TA349	[R5]CN1C=C(CNS(=O)(=O)c2ccc3c(c2)CCN3C(C)=O)N=N1	ATTTCA	TA570	[R5]CN1C=C(CN2CCNC(=O)C2)N=N1	AGACGT
TA129	[R5]CN1C=C(c2cn(C)c2)N=N1	AAGAAA	TA350	[R5]CN1C=C(CNS(=O)(=O)c2ccc3c(c2)CCCN3C(C)=O)N=N1	ATTTCT	TA571	[R5]CN1C=C(CN2CCC(O)CC2)N=N1	AGACGG
TA130	[R5]CN1C=C(C(C)O)N=N1	AAGAAT	TA351	[R5]CN1C=C(CNS(=O)(=O)c2ccc(N3CCCC3=O)cc2)N=N1	ATTTCC	TA572	[R5]CN1C=C(COC(N)=O)N=N1	AGACGC
TA131	[R5]CN1C=C(CN2CCCC2=O)N=N1	AAGAAG	TA352	[R5]CN1C=C(CNC(=O)CSc2nnc3sc(C)c(C)c23)N=N1	ATTTCC	TA573	[R5]CN1C=C(CNC(=O)NC)N=N1	AGACCA
TA132	[R5]CN1C=C(CCC#N)N=N1	AAGAAC	TA353	[R5]CN1C=C(CNC(=O)Cn2nnn(-c3cccc3)c2=O)N=N1	ATTGAA	TA574	[R5]CN1C=C(COC(=O)CN2C(=O)c3ccc4cccc(c34)C2=O)N=N1	AGACCT
TA133	[R5]CN1C=C(CCOc2cccc2)N=N1	AAGATA	TA354	[R5]CN1C=C(COc2cccc2C=C(C#N)c2nc(C)cs2)N=N1	ATTGAT	TA575	[R5]CN1C=C(C#CCCCCCCCCCCC)N=N1	AGACCG

TA134	[R5]CN1C=C(c2ccc2F)N=N1	AAGATT	TA355	[R5]CN1C=C(CNC(=O)Cn2nc3sc4c(c3c2=O)CCC4)N=N1	ATTGAG	TA576	[R5]CN1C=C(CSc2nc(N)c3c(C)c(C)s c3n2)N=N1	AGACCC
TA135	[R5]CN1C=C(c2ccc2C)N=N1	AAGATG	TA356	[R5]CN1C=C(CNC(=O)CSc2nc3nc(C)cc(C)n3n2)N=N1	ATTGAC	TA577	[R5]CN1C=C(C2(C)CCC3C(O2)c2ccc cc2OC3(C)C)N=N1	AGTAAA
TA136	[R5]CN1C=C(CN2CCC(CN(CC)CC)C2)N=N1	AAGATC	TA357	[R5]CN1C=C(CNC(=O)N2CCN(c3ccc c(C)1c3)CC2)N=N1	ATTGTA	TA578	[R5]CN1C=C(CNC(=O)N2CCN(c3ccc (C(C)=O)cc3)CC2) N=N1	AGTAAT
TA137	[R5]CN1C=C(CNC(=O)CCCc2nc(C)no2)N=N1	AAGAGA	TA358	[R5]CN1C=C(CNC(=O)N2CCN(c3ccc cc3O)CC2)N=N1	ATTGTT	TA579	[R5]CN1C=C(CNS(=O)(=O)c2ccc(Br)cc2)N=N1	AGTAAG
TA138	[R5]CN1C=C(CN2CCCN(Cc3ccccc3)CC2)N=N1	AAGAGT	TA359	[R5]CN1C=C(CNC(=O)Cn2nc3sc(C)c(C)c3c2=O)N=N1	ATTGTG	TA580	[R5]CN1C=C(CNC(=O)CSc2nc3ccccc 3c(=O)n2C)N=N1	AGTAAC
TA139	[R5]CN1C=C(CNC(=O)C(C)Sc2ccccc 2F)N=N1	AAGAGG	TA360	[R5]CN1C=C(CNC(=O)c2ccc3ccccc3 c2)N=N1	ATTGTC	TA581	[R5]CN1C=C(C(C(C)C)CC)NC(=O)c2c cc(OC)c([N+](=O)[O-])c2)N=N1	AGTATA
TA140	[R5]CN1C=C(c2ccc(NC(=O)N3CCC C(O)C3)c2)N=N1	AAGAGC	TA361	[R5]CN1C=C(CSc2nnnn2-c2ccccc2OC)N=N1	ATTGGA	TA582	[R5]CN1C=C(CNS(=O)(=O)c2cc(OC)ccc2OC)N=N1	AGTATT
TA141	[R5]CN1C=C(CNC(=O)CN(C)C(C)c2c cc(F)cc2)N=N1	AAGACA	TA362	[R5]CN1C=C(CNC(=O)Cc2cc3c(C)c(C)ccc23)N=N1	ATTGGT	TA583	[R5]CN1C=C(COc2ccc(C=C(C#N)C# N)cc2OCC)N=N1	AGTATG
TA142	[R5]CN1C=C(CNC(=O)CN(CCC(=O)OCC)C2CC2)N=N1	AAGACT	TA363	[R5]CN1C=C(CNC(=O)COC(=O)Cn2ccc(=O)c3ccccc32) N=N1	ATTGGG	TA584	[R5]CN1C=C(COc2ccc(C=C(C#N)C(=O)OCC)cc2OCC) N=N1	AGTATC
TA143	[R5]CN1C=C(CNC(=O)CN(C)C(C)c2c cc(C)cc2)N=N1	AAGACG	TA364	[R5]CN1C=C(CNC(=O)c2ccc(COc3cc c(F)cc3)o2)N=N1	ATTGGC	TA585	[R5]CN1C=C(CNC(=O)c2ccc(Cl)cc2C l)N=N1	AGTAGA
TA144	[R5]CN1C=C(c2ccc(NCC(=O)NCc3c cco3)c2)N=N1	AAGACC	TA365	[R5]CN1C=C(c2ccc(NC(=O)NCc3nc (C)cs3)c2)N=N1	ATTGCA	TA586	[R5]CN1C=C(CN2CCCC2C(=O)Nc 2ccc3oc(C)nc3c2) N=N1	AGTAGT
TA145	[R5]CN1C=C(CNC(=O)CN(C)Cc2ccc(F)cc2F)N=N1	AAGTAA	TA366	[R5]CN1C=C(CNC(=O)Cc2csc(C(C)C)n2)N=N1	ATTGCT	TA587	[R5]CN1C=C(CNC(=O)c2ccc3c(=O)n 4c(nc3c2)CCCC4) N=N1	AGTAGG
TA146	[R5]CN1C=C(CNC(=O)CN(Cc2sccc2 C)C(C)C)N=N1	AAGTAT	TA367	[R5]CN1C=C(CNC(=O)N2CCN(Cc3nc(-c4ccccc4C)no3)C2)N=N1	ATTGCG	TA588	[R5]CN1C=C(COc2ccccc2C2NC(=O) c3ccccc3N2)N=N1	AGTAGC
TA147	[R5]CN1C=C(CNC(=O)CN(CC)C(C)c2ccc(F)cc2)N=N1	AAGTAG	TA368	[R5]CN1C=C(c2ccc(NC(=O)NCCC3C CS(=O)(=O)C3)c2)	ATTGCC	TA589	[R5]CN1C=C(CNC(=O)[C@@H]2CS CN2C(=O)NC2CC	AGTACA

				N=N1			CCC2)N=N1	
TA148	[R5]CN1C=C(CNC(=O)CN(C)C(C)c2cc(F)c(F)c2)N=N1	AAGTAC	TA369	[R5]CN1C=C(CNC(=O)COC(=O)c2[nH]c(C)c(C(=O)OC)c2C)N=N1	ATTCAA	TA590	[R5]CN1C=C(CNC(=O)C2C3CCC(C3)N2C(=O)C2CC2)N=N1	AGTACT
TA149	[R5]CN1C=C(CNC(=O)C(C)N(C)Cc2cc(Br)o2)N=N1	AAGTTA	TA370	[R5]CN1C=C(CNC(=O)CN2C(=O)NC(c3cccc3)(c3cccc3)C2=O)N=N1	ATTCAT	TA591	[R5]CN1C=C(CNC(=O)CN2C(=O)C3C4C=CC(C4)C3C2=O)N=N1	AGTACG
TA150	[R5]CN1C=C(CNC(=O)N(C)Cc2cccc2OC(F)(F)F)N=N1	AAGTTT	TA371	[R5]CN1C=C(CNC(=O)N2CCC(n3c(=O)[nH]c4cccc43)CC2)N=N1	ATTCAG	TA592	[R5]CN1C=C(CNC(N)=O)N=N1	AGTACC
TA151	[R5]CN1C=C(CN2CCC(C(=O)NCCN(C(C)C)C(C)C)CC2)N=N1	AAGTTG	TA372	[R5]CN1C=C(COC2cccc2C=C2C(=O)Nc3cccc32)N=N1	ATTCAC	TA593	[R5]CN1C=C(CNC(=O)C23CC4CC(C(C4)C2)C3)N=N1	AGTTAA
TA152	[R5]CN1C=C(CN(CC)S(=O)(=O)c2ccsc2C(=O)OC)N=N1	AAGTTC	TA373	[R5]CN1C=C(Cn2cnc3cccc32)N=N1	ATTCTA	TA594	[R5]CN1C=C(c2ccc(CO)cc2)N=N1	AGTTAT
TA153	[R5]CN1C=C(CNC(=O)N2CCC(C(=O)N3CCCCC3)CC2)N=N1	AAGTGA	TA374	[R5]CN1C=C(CNS(=O)(=O)c2ccc(OC)c(OC)c2)N=N1	ATTCTT	TA595	[R5]CN1C=C(c2cnc(C)cc2)N=N1	AGTTAG
TA154	[R5]CN1C=C(CNC(=O)CN2CCCN(c3nc(C)cs3)CC2)N=N1	AAGTGT	TA375	[R5]CN1C=C(C(C(C)CC)NC(=O)CSc2nc3cccc3o2)N=N1	ATTCTG	TA596	[R5]CN1C=C(c2ccc(C)cc2F)N=N1	AGTTAC
TA155	[R5]CN1C=C(CNC(=O)CN(C)Cc2ccc(C(F)(F)F)cc2)N=N1	AAGTGG	TA376	[R5]CN1C=C(CNC(=O)Cn2cnc3c(cnn3-c3cccc3)c2=O)N=N1	ATTCTC	TA597	[R5]CN1C=C(c2ccc(C)cc2)N=N1	AGTTTA
TA156	[R5]CN1C=C(c2ccc(NCC(=O)NCc3cccc3F)c2)N=N1	AAGTGC	TA377	[R5]CN1C=C(Cn2c(-c3cccs3)nc3cccc32)N=N1	ATTCGA	TA598	[R5]CN1C=C(CN2CCNC2=O)N=N1	AGTTTT
TA157	[R5]CN1C=C(CNC(=O)CN2C(=O)NC(C)(c3ccc(C)cc3)C2=O)N=N1	AAGTCA	TA378	[R5]CN1C=C(CNC(=O)CN2C(=O)C(=O)c3cc(C)ccc32)N=N1	ATTCGT	TA599	[R5]CN1C=C(CN2C(=O)C=CC2=O)N=N1	AGTTTG
TA158	[R5]CN1C=C(CNC(=O)c2cnn(-c3ccc(C)cc3)c2C(C)C)N=N1	AAGTCT	TA379	[R5]CN1C=C(CNC(=O)N2CCN(c3ccc(C)C)cc3F)CC2)N=N1	ATTCGG	TA600	[R5]CN1C=C(Cn2cnc3c(N)ncnc32)N=N1	AGTTTC
TA159	[R5]CN1C=C(c2ccc(NCC(=O)NCC3(c4cccc4)CC3)c2)N=N1	AAGTCG	TA380	[R5]CN1C=C(C(C(C)CC)NC(=O)Cn2cnc3cccc3c2=O)N=N1	ATTCGC	TA601	[R5]CN1C=C(COC2cccc(S(C)(=O)O)c2)N=N1	AGTTGA
TA160	[R5]CN1C=C(CNC(=O)COC(=O)[C@	AAGTCC	TA381	[R5]CN1C=C(Cn2c(C)CNC(=O)c3ccco	ATTCCA	TA602	[R5]CN1C=C(Cn2c(C)cc3cccc32)N=N1	AGTTGT

	@H](NS(=O)(=O)C=Cc2ccccc2)C(C)C)N=N1			3)nc3ccccc32)N=N1				
TA161	[R5]CN1C=C(C(C)C)(CC)NC(=O)Cc2cnn(-c3ccccc3)c2)N=N1	AAGGAA	TA382	[R5]CN1C=C(CNC(=O)c2cc3c(ccc4ccccc43)c2C)N=N1	ATTCCCT	TA603	[R5]CN1C=C(Cn2c3ccccc3c3ccccc32)N=N1	AGTTGG
TA162	[R5]CN1C=C(CNC(=O)COC(=O)c2cc(S(=O)(=O)N3CC(C)OC(C)C3)cc2)N=N1	AAGGAT	TA383	[R5]CN1C=C(COC2ccccc2C=C2SC(=S)N(CC)C2=O)N=N1	ATTCCG	TA604	[R5]CN1C=C(CCCC)N=N1	AGTTGC
TA163	[R5]CN1C=C(CNC(=O)COC(=O)c2cc(N)c([N+](=O)[O-])c2)N=N1	AAGGAG	TA384	[R5]CN1C=C(CNC(=O)c2ccc(-c3ccccc3(F)(F)F)c3o2)N=N1	ATTCCC	TA605	[R5]CN1C=C(C(C)C)n2ccccc2)N=N1	AGTTCA
TA164	[R5]CN1C=C(CNC(=O)c2ccccc(S(=O)(=O)NCC=C)c2)N=N1	AAGGAC	TA385	[R5]CN1C=C(c2ccc(NS(=O)(=O)c3c(C)cc(C)cc3C)c2)N=N1	ATGAAA	TA606	[R5]CN1C=C(CNC(=O)C(F)(F)F)N=N1	AGTTCT
TA165	[R5]CN1C=C(CNC(=O)COC(=O)CN2C(=O)c3ccccc3C2=O)N=N1	AAGGTA	TA386	[R5]CN1C=C(CNC(=O)C2c3ccccc3O c3ccccc32)N=N1	ATGAAT	TA607	[R5]CN1C=C(CCN(C)C(=O)N)N=N1	AGTTCCG
TA166	[R5]CN1C=C(CNC(=O)COC(=O)CCc2c(C)[nH]c(=O)[nH]c2=O)N=N1	AAGGTT	TA387	[R5]CN1C=C(c2ccc(NC(=O)C3ccsc3)c2)N=N1	ATGAAG	TA608	[R5]CN1C=C(CN2CCOC2=O)N=N1	AGTTCC
TA167	[R5]CN1C=C(CNC(=O)c2ccc(N3CCC(CC3)c([N+](=O)[O-])c2)N=N1	AAGGTG	TA388	[R5]CN1C=C(c2ccc(NC(=O)C3cc(O)C)cc(OC)c3)c2)N=N1	ATGAAC	TA609	[R5]CN1C=C(c2ccc(ccc2)N)N=N1	AGTGAA
TA168	[R5]CN1C=C(CNC(=O)CCCN(C)S(=O)(=O)c2cccs2)N=N1	AAGGTC	TA389	[R5]CN1C=C(c2ccc(NC(=O)C3cc(F)ccc3Cl)c2)N=N1	ATGATA	TA610	[R5]CN1C=C(COC(C)C(=O)O)N=N1	AGTGAT
TA169	[R5]CN1C=C(CNC(=O)CCn2c(S)nnc2-c2ccc(C)cc2)N=N1	AAGGGA	TA390	[R5]CN1C=C(CNC(=O)C=Cc2cc(Cl)c(OC)c(OCC)c2)N=N1	ATGATT	TA611	[R5]CN1C=C(CNC(=O)C=C)N=N1	AGTGAG
TA170	[R5]CN1C=C(CNC(=O)c2cc3ccccc3c2NC(=O)C2CC2)N=N1	AAGGGT	TA391	[R5]CN1C=C(CNC(=O)c2cc(C3CC3)n c3ccccc23)N=N1	ATGATG	TA612	[R5]CN1C=C(c2ccc(ccc2-c2ccccc2)N)N=N1	AGTGAC
TA171	[R5]CN1C=C(CNC(=O)CNC(=O)c2cnc3ccccc3c2O)N=N1	AAGGGG	TA392	[R5]CN1C=C(CNC(=O)CCc2cc(OC)c(C)OC)c2)N=N1	ATGATC	TA613	[R5]CN1C=C(c2ccc3c(C)CC3)N=N1	AGTGTA
TA172	[R5]CN1C=C(CNC(=O)COC(=O)c2[nH]c(=O)c3ccccc23)N=N1	AAGGGC	TA393	[R5]CN1C=C(CNC(=O)CSc2nc3ccccc3n2CC)N=N1	ATGAGA	TA614	[R5]CN1C=C(c2ccc(cc(-c3ccccc3)c2)N)N=N1	AGTGTT

TA173	[R5]CN1C=C(CNC(=O)C(=O)c2c(C)n n(-c3cccc3)c2C)N=N1	AAGGCA	TA394	[R5]CN1C=C(c2cc cc(NC(=O)c3cccc(OC)c3OC)c2)N=N1	ATGAGT	TA615	[R5]CN1C=C(CC(C)(O)CC)N=N1	AGTGTG
TA174	[R5]CN1C=C(CNC(=O)c2cnn(Cc3cccc3Cl)c2)N=N1	AAGGCT	TA395	[R5]CN1C=C(CNC(=O)c2ccc(C)c([N+](=O)[O-])c2)N=N1	ATGAGG	TA616	[R5]CN1C=C(CN2CCCCC2)N=N1	AGTGTG
TA175	[R5]CN1C=C(CNC(=O)COC(=O)C(C)NC2=NS(=O)(=O)c3cccc32)N=N1	AAGGCG	TA396	[R5]CN1C=C(CNC(=O)c2ccc(Cl)cc2[N+](=O)[O-])N=N1	ATGAGC	TA617	[R5]CN1C=C(CNC(C(F)(F)F)N=N1	AGTGGA
TA176	[R5]CN1C=C(CNC(=O)COC(=O)c2ccc(OC)c2O)N=N1	AAGGCC	TA397	[R5]CN1C=C(CNC(=O)c2ccc(N(C)C)c2)N=N1	ATGACA	TA618	[R5]CN1C=C(CC2(O)CCOCC2)N=N1	AGTGGT
TA177	[R5]CN1C=C(CNC(=O)CCNC(=O)c2ccc(F)cc2F)N=N1	AAGCAA	TA398	[R5]CN1C=C(Cn2c(=O)[nH]c3cccc3c2=O)N=N1	ATGACT	TA619	[R5]CN1C=C(c2cc nn2C)N=N1	AGTGGG
TA178	[R5]CN1C=C(CNC(=O)COC(=O)CN2C(=O)NC3(CCC(C(C)C)CC3)C2=O)N=N1	AAGCAT	TA399	[R5]CN1C=C(CNC(=O)C=Cc2ccc(OC(F)F)c(OCC)c2)N=N1	ATGACG	TA620	[R5]CN1C=C(CCN S(C)(=O)=O)N=N1	AGTGCC
TA179	[R5]CN1C=C(CNC(=O)c2cccc(S(=O)(=O)N3CCCC3)c2)N=N1	AAGCAG	TA400	[R5]CN1C=C(CNS(=O)(=O)c2ccc(C(=O)NC(C)CC)cc2)N=N1	ATGACC	TA621	[R5]CN1C=C(CN2CCC(N3CCCC3)CC2)N=N1	AGTGCA
TA180	[R5]CN1C=C(CNC(=O)c2cnc(S)n2-c2cccc2)N=N1	AAGCAC	TA401	[R5]CN1C=C(CNC(=O)c2cc(-c3cccc3)nc3cccc23)N=N1	ATGTAA	TA622	[R5]CN1C=C(CNC(=O)CN2CC(C)CC(C)C2)N=N1	AGTGCT
TA181	[R5]CN1C=C(CNC(=O)CCC(=O)c2cc c3c(c2)CCCC3)N=N1	AAGCTA	TA402	[R5]CN1C=C(c2cc cc(NS(=O)(=O)c3cccs3)c2)N=N1	ATGTAT	TA623	[R5]CN1C=C(CNC(=O)CN2CCCCC C2)N=N1	AGTGCG
TA182	[R5]CN1C=C(CN(CC)CC(=O)NCc2cc ccs2)N=N1	AAGCTT	TA403	[R5]CN1C=C(CNC(=O)c2cn(-c3cccc3)nc2-c2cccc2)N=N1	ATGTAG	TA624	[R5]CN1C=C(CN2 CCSCC2)N=N1	AGTGCC
TA183	[R5]CN1C=C(c2cc cc(NC(=O)c3ccc(-n4cncn4)cc3)c2)N=N1	AAGCTG	TA404	[R5]CN1C=C(c2cc cc(NC(=O)c3ccc(S(C)(=O)=O)c3)c2)N=N1	ATGTAC	TA625	[R5]CN1C=C(C2C CCCC2)N=N1	AGTCAA
TA184	[R5]CN1C=C(CNC(=O)COC2ccc(-n3cnnn3)cc2)N=N1	AAGCTC	TA405	[R5]CN1C=C(CNC(=O)C=Cc2ccc(OC(C)C)c(OC)c2)N=N1	ATGTTA	TA626	[R5]CN1C=C(COC(C(=O)OC)N=N1	AGTCAT
TA185	[R5]CN1C=C(c2cc cc(NC(=O)C(C)N3C(=O)CCC3=O)c2)N=N1	AAGCGA	TA406	[R5]CN1C=C(c2cc cc(NC(=O)c3ccc4nc(C)c(C)nc4c3)c2)N=N1	ATGTTT	TA627	[R5]CN1C=C(c2cncnc2)N=N1	AGTCAG

TA186	[R5]CN1C=C(CNC(=O)c2ccc(Br)cc2F)N=N1	AAGCGT	TA407	[R5]CN1C=C(c2ccc(NC(=O)c3ccc(C)ccc4n3)c2)N=N1	ATGTTG	TA628	[R5]CN1C=C(CCCN2CCOCC2)N=N1	AGTCAC
TA187	[R5]CN1C=C(c2ccc(NC(=O)CCOCC3CCCO3)c2)N=N1	AAGCGG	TA408	[R5]CN1C=C(c2ccc(NC(=O)c3cccc(C#N)c3)c2)N=N1	ATGTTC	TA629	[R5]CN1C=C(c2ccncc2)N=N1	AGTCTA
TA188	[R5]CN1C=C(c2ccc(NCC(=O)NC3(C#N)CCCC3)c2)N=N1	AAGCGC	TA409	[R5]CN1C=C(c2ccc(NCC(=O)NC3cccs3)c2)N=N1	ATGTGA	TA630	[R5]CN1C=C(C2(O)CCOCC2)N=N1	AGTCTT
TA189	[R5]CN1C=C(c2ccc(NCC(=O)NC3(C#N)CCCC3)c2)N=N1	AAGCCA	TA410	[R5]CN1C=C(CNC(=O)c2cc(-c3ccncc3)nc3cccc23)N=N1	ATGTGT	TA631	[R5]CN1C=C(c2ccc(OC(C)nc2)N=N1	AGTCTG
TA190	[R5]CN1C=C(c2ccc(NC(=O)CCc3nc(C(C)C)no3)c2)N=N1	AAGCCT	TA411	[R5]CN1C=C(CNC(=O)Cn2nc3c(C)cccc3c2=O)N=N1	ATGTGG	TA632	[R5]CN1C=C(C(C)(O)C=C)N=N1	AGTCTC
TA191	[R5]CN1C=C(c2ccc(NC(=O)CCc3nc(C(C)C)no3)c2)N=N1	AAGCCG	TA412	[R5]CN1C=C(c2ccc(NC(=O)c3cc(Cl)ccc3F)c2)N=N1	ATGTGC	TA633	[R5]CN1C=C(C(C)(C)C)N=N1	AGTCGA
TA192	[R5]CN1C=C(c2ccc(NCC(=O)N3CC(C)OC(C)C3)c2)N=N1	AAGCCC	TA413	[R5]CN1C=C(c2ccc(NC(=O)c3cccc(Cl)cc3F)c2)N=N1	ATGTCA	TA634	[R5]CN1C=C(CC(N)C(=O)NN)N=N1	AGTCGT
TA193	[R5]CN1C=C(CNC(=O)CN2CCOC3cccc32)N=N1	AACAAA	TA414	[R5]CN1C=C(CNC(=O)c2[nH]c(C)c(C)C(=O)c2C)N=N1	ATGTCT	TA635	[R5]CN1C=C(c2cc(F)c(F)cc2F)N=N1	AGTCGG
TA194	[R5]CN1C=C(CNC(=O)CSc2nccc(C(F)F)F)n2)N=N1	AACAAT	TA415	[R5]CN1C=C(c2ccc(NCC(=O)Nc3ccc(F)c(F)c3)c2)N=N1	ATGTCC	TA636	[R5]CN1C=C(C2(O)CCCN(C(=O)OC(C)C)CC2)N=N1	AGTCGC
TA195	[R5]CN1C=C(c2ccc(NC(=O)Cc3csc(C(C)C)n3)c2)N=N1	AACAAG	TA416	[R5]CN1C=C(CNC(=O)c2ccc(OC)c(O)C)c2)N=N1	ATGTCC	TA637	[R5]CN1C=C(CO)N=N1	AGTCCA
TA196	[R5]CN1C=C(c2ccc(NCc3nc(-c4ccc4)no3)c2)N=N1	AACAAC	TA417	[R5]CN1C=C(CNC(=O)c2ccc3c(c2)CCC3)N=N1	ATGGAA	TA638	[R5]CN1C=C(c2ccc(Br)cc2)N=N1	AGTCCT
TA197	[R5]CN1C=C(CNC(=O)C(C)Sc2nccc(C(F)F)F)n2)N=N1	AACATA	TA418	[R5]CN1C=C(CNC(=O)c2cc(C)cc(C)c2)N=N1	ATGGAT	TA639	[R5]CN1C=C(CCC(C#N)N)N=N1	AGTCCG
TA198	[R5]CN1C=C(COC2cccc(C#N)c2)N=N1	AACATT	TA419	[R5]CN1C=C(CNC(=O)c2ccc(C)cc2C)N=N1	ATGGAG	TA640	[R5]CN1C=C(CCCCO)N=N1	AGTCCC
TA199	[R5]CN1C=C(CNC(=O)CC#N)N=N1	AACATG	TA420	[R5]CN1C=C(CNC(=O)c2cccc(C)c2C)N=N1	ATGGAC	TA641	[R5]CN1C=C(C(=O)OCC)N=N1	AGGAAA

TA200	[R5]CN1C=C(CNC2CCS(=O)(=O)C2)N=N1	AACATC	TA421	[R5]CN1C=C(CNC(=O)c2nn(C)c(=O)c3ccccc23)N=N1	ATGGTA	TA642	[R5]CN1C=C(CCC)N=N1	AGGAAT
TA201	[R5]CN1C=C(CNC(=O)CCNS(=O)(=O)c2ccc(C)cc2C)N=N1	AACAGA	TA422	[R5]CN1C=C(CNC(=O)c2ccccc2)N=N1	ATGGTT	TA643	[R5]CN1C=C(CN)N=N1	AGGAAG
TA202	[R5]CN1C=C(c2ccc(NCC(=O)Nc3c(C)cccc3CC)c2)N=N1	AACAGT	TA423	[R5]CN1C=C(CNC(=O)c2cc3ccccc3s2)N=N1	ATGGTG	TA644	[R5]CN1C=C(CN(C)C)N=N1	AGGAAC
TA203	[R5]CN1C=C(CNC(=O)CSc2nnc(N)n2CC=C)N=N1	AACAGG	TA424	[R5]CN1C=C(CNS(=O)(=O)c2cccc(C(=O)N(CC)CC)c2)N=N1	ATGGTC	TA645	[R5]CN1C=C(c2ccc3ccccc23)N=N1	AGGATA
TA204	[R5]CN1C=C(CNC(=O)c2cc(C=CC)c(OC)c(OC)c2)N=N1	AACAGC	TA425	[R5]CN1C=C(CNC(=O)CSc2nc3ccccc3nc2C)N=N1	ATGGGA	TA646	[R5]CN1C=C(CCC)N=N1	AGGATT
TA205	[R5]CN1C=C(CNC(=O)Cn2nc(C)cc2C)N=N1	AACACA	TA426	[R5]CN1C=C(c2ccc(NC(=O)Cn3cccc3=O)c2)N=N1	ATGGGT	TA647	[R5]CN1C=C(Cc2cccc2)N=N1	AGGATG
TA206	[R5]CN1C=C(CNC(=O)N2CCCCC2)N=N1	AACACT	TA427	[R5]CN1C=C(CNC(=O)C2CCCC2)N=N1	ATGGGG	TA648	[R5]CN1C=C(CBr)N=N1	AGGATC
TA207	[R5]CN1C=C(CNC(=O)CSc2ccc(F)c(Cl)c2)N=N1	AACACG	TA428	[R5]CN1C=C(CNC(=O)c2c[nH]nc2-c2ccc(F)cc2)N=N1	ATGGGC	TA649	[R5]CN1C=C(CCCc2ccccc2)N=N1	AGGAGA
TA208	[R5]CN1C=C(c2ccc(NC(C)C(=O)Nc3sc3C#N)c2)N=N1	AACACC	TA429	[R5]CN1C=C(CNC(=O)c2nc(-c3ccccc3)oc2C)N=N1	ATGGCA	TA650	[R5]CN1C=C(c2ccc(CCCCC)cc2)N=N1	AGGAGT
TA209	[R5]CN1C=C(CNC(=O)c2nc(C)n(-c3ccc(Cl)cc3)n2)N=N1	AACTAA	TA430	[R5]CN1C=C(CNC(=O)c2cccc(S(C)(=O)=O)c2)N=N1	ATGGCT	TA651	[R5]CN1C=C(CCCc2ccccc2)N=N1	AGGAGG
TA210	[R5]CN1C=C(CNC(=O)N(C)Cc2ccc(Cl)c(Cl)c2)N=N1	AACTAT	TA431	[R5]CN1C=C(CNC(=O)N2CCC(c3c[nH]c4ccccc34)CC2)N=N1	ATGGCG	TA652	[R5]CN1C=C(COCc2ccccc2)N=N1	AGGAGC
TA211	[R5]CN1C=C(c2ccc(NC(=O)c3ncnc3N)c2)N=N1	AACTAG	TA432	[R5]CN1C=C(CNC(=O)c2ccc(C)C(C)C2)N=N1	ATGGCC	TA653	[R5]CN1C=C(c2cc(C(F)(F)F)cc(C(F)(F)F)c2)N=N1	AGGACA
TA212	[R5]CN1C=C(c2ccc(NC(=O)c3ccccc3F)c2)N=N1	AACTAC	TA433	[R5]CN1C=C(CNC(=O)CC23CC4CC(C)C(O)(C4)C2)C3)N=N1	ATGCAA	TA654	[R5]CN1C=C(c2ccc(C(C)(C)C)cc2)N=N1	AGGACT
TA213	[R5]CN1C=C(c2ccc(NCC(=O)N3CCOC(C)C3)c2)N=N1	AACTTA	TA434	[R5]CN1C=C(CNC(=O)C23CC4CC(C)C(C)C(C4)C2)C3)N=N1	ATGCAT	TA655	[R5]CN1C=C(CCB)N=N1	AGGACG

TA214	[R5]CN1C=C(CNC(=O)c2cccc(S(=O)(=O)N3CCCCC3)c2)N=N1	AACTTT	TA435	[R5]CN1C=C(CNC(=O)C2CC3CCCC(C2)C3=O)N=N1	ATGCAG	TA656	[R5]CN1C=C(c2cccc2C)N=N1	AGGACC
TA215	[R5]CN1C=C(CNC(=O)NC2(C(=O)O)CCc3cccc32)N=N1	AACTTG	TA436	[R5]CN1C=C(c2ccc(NC(=O)Cc3ccc4cc5c(cc34)CCC5)c2)N=N1	ATGCAC	TA657	[R5]CN1C=C(c2ccc(N(C)C)cc2)N=N1	AGGTAA
TA216	[R5]CN1C=C(CNC(=O)Nc2cc(C)on2)N=N1	AACTTC	TA437	[R5]CN1C=C(CNC(=O)c2cccc3c4c([nH]c23)CCCC4)N=N1	ATGCTA	TA658	[R5]CN1C=C(CCCC(C)CC)N=N1	AGGTAT
TA217	[R5]CN1C=C(CN(Cc2ccc(C(=O)NC)cc2)C2CC2)N=N1	AACTGA	TA438	[R5]CN1C=C(CNC(=O)C=Cc2ccc(OC)c(OC)c2)N=N1	ATGCTT	TA659	[R5]CN1C=C(c2ccc([N+](=O)[O-])cc2)N=N1	AGGTAG
TA218	[R5]CN1C=C(CN(CC)c2c(N)n(Cc3cccc3)c(=O)[nH]c2=O)N=N1	AACTGT	TA439	[R5]CN1C=C(CNC(=O)CSc2nc(N)c3c(C)c(C)sc3n2)N=N1	ATGCTG	TA660	[R5]CN1C=C(C(=O)Nc2cccc3c(Cl)c[nH]c23)N=N1	AGGTAC
TA219	[R5]CN1C=C(CNC(=O)C(C)N2CC=C(c3ccc(OC)cc3)CC2)N=N1	AACTGG	TA440	[R5]CN1C=C(CNC(=O)c2cnn(-c3ccc(Cl)cc3)c2C(C)C)N=N1	ATGCTC	TA661	[R5]CN1C=C(CS(=O)(=O)O)N=N1	AGGTAA
TA220	[R5]CN1C=C(CNC(=O)C(C)Nc2ccc3oc(C)nc3c2)N=N1	AACTGC	TA441	[R5]CN1C=C(c2ccc(NC(=O)Cc3cccc4ccnc34)c2)N=N1	ATGCGA	TA662	[R5]CN1C=C(CCCc2nnn[nH]2)N=N1	AGGTTT
TA221	[R5]CN1C=C(CNC(=O)NC2CCN(S(C)(=O)=O)CC2)N=N1	AACTCA	TA442	[R5]CN1C=C(CNC(=O)C2=Cc3cccc3OC2)N=N1	ATGCGT	TA663	[R5]CN1C=C(CN2C(=O)CCC(N3C(=O)c4cccc(N)c4C3=O)C2=O)N=N1	AGGTTG

TA664: O=C(CCCC1=CN(C[R5])N=N1)NC2=NN=C(S(=O)(N)=O)S2, AGGTTC